The development trend of China's aging population: a forecast perspective

Xuchong Liu¹ · Jianian Zhu² · Kai Zou²

Received: 19 October 2021 / Accepted: 28 January 2022 / Published online: 25 February 2022
© The Author(s) 2022

Abstract

To accurately predict the aging population in China, a novel grey prediction model (CFODGMW(1,1, α) model) is established in this study. The CFODGMW(1,1, α) model has all the advantages of the weighted least square method, combined fractional-order accumulation generation operation and grey prediction model with time power term, which makes it have excellent prediction performance. Compared with the traditional grey prediction model based on the least square method and the first-order accumulation operation, the CFODGMW(1,1, α) model has stronger adaptability. The proposed model and its competing models are used to analyze the aging population in five regions of China. The results show that the prediction performance of the CFODGMW(1,1, α) model is better than other models. Based on this, the CFODGMW(1,1, α) model is used to predict the aging population in China in the next 4 years, and some suggestions are given based on the development trend of the aging population.

Keywords Aging population · The weighted least square method · Grey prediction model · The whale optimizer · The combined fractional-order accumulation generation operation

Introduction

Population aging refers to the process in which the number of the elderly (≥ 65 years old) keeps increasing and the proportion of the total population keeps rising. Recently, China's population aging has become an arresting problem. Figure 1 shows the number and growth rate of China's total population and aging population (the data used for mapping comes from China's National Bureau of Statistics-NBS, <https://data.stats.gov.cn/>). It can be seen from Fig. 1 that the growth rate of China's total population is far less than that of the aging population. In the face of the growing aging population, a series of problems such as pension, medical care and assistance follow. Therefore, accurate prediction of China's aging population is helpful for the government to formulate corresponding policies.

Considering the exponential growth law of population and the applicable scope of the grey prediction model (GM model), this paper chooses the GM model to analyze and forecast the aging population in China. The GM model is a vital component of the grey system theory put forward by the famous Chinese scholar Deng [1]. Due to its excellent predictive performance, the GM model and its variants have been widely used in various areas of society [2–7]. The GM(1,1) model is the benchmark model of all grey variant models. The improvement work on it has never stopped. These improvement schemes mainly focus on the expansion of the model, accumulative order, initial value, background value and parameter estimation. Since the background value and initial value are irrelevant to the focus of this article, we will not cover the work of these two parts. The work of the other three parts is shown as follows:

✉ Jianian Zhu
201921000507@smail.xtu.edu.cn

¹ Key Laboratory of Network Crime Investigation of Hunan Provincial Colleges, Hunan Police Academy, Changsha 410136, China

² School of Public administration, Xiangtan University, Xiangtan 411100, China

- (1) Expansion of the model. To further expand the applicable scope of the GM model, Cui et al. put forward the NGM(1,1,k,c) model, which can simultaneously fit the time series with homogeneous exponential law and the time series with approximately non-homogeneous exponential series [8]. Qian et al. proposed the GM(1,1, t^α) model with time power term and proved that GM(1,1)

Fig. 1 The number and growth rate of China's aging population and total population from 2005 to 2019

and NGM(1,1,k) model are special expressions of this model [9]; Inspired by Qian et al., Luo and Wei put forward another grey model with time term, namely, GMP(1,1,N) model [10]. Similar to GM(1,1, t^α) model, GMP(1,1,N) model can also be transformed into other grey forecasting models by changing its parameters. Liu et al. analyzed the modeling process and mechanism of the GM(1,1, t^α) and GMP(1,1,N) model and found that these two models have certain defects. Based on this, Liu et al. proposed the OPGM(1,1, α) model with all the advantages of the two models [11]. Although the GMP(1,1,N) and OPGM(1,1, α) model have excellent predictive ability, their modeling mechanisms are too complex to be universal. Therefore, the GM(1,1, t^α) model with strong adaptability is the best choice for predictive analysis.

- (2) Accumulative order. The traditional GM model is built on the basis of the first-order cumulative generation operation (1-AGO), which aims to reduce the randomness of time series to achieve the purpose of accurate prediction. However, some scholars have found that the predictive accuracy of the GM model based on the 1-AGO is even lower than that of the GM model based on initial sequence modeling in some cases [12]. To solve this problem, Wu proposed the idea of the fractional-order accumulative generation operation (r-AGO) [13]. The idea of r-AGO is to establish a programming problem based on the model, and then use the intelligent optimization algorithms to solve the programming problem to obtain the accumulation order r that make the model have the minimum error. If the benefit brought by the accumulative operation is too small, the order of the r-AGO is 0, that is, the original time series is used for modeling. If the accumulative operation can greatly enhance the predictive accuracy of the predictive model, then the algorithms can find the best order to build the model. Since r-AGO can stably improve the performance of forecasting models, many grey system

models based on r-AGO have been developed [14,15]. In addition, inspired by Wu [13], many other fractional-order accumulation operations have come into being. For instance, inspired by the conformable fractional derivative, Ma et al. proposed the conformable fractional grey prediction model [16]. Subsequently, Chen et al. put forward the grey forecasting model with the fractional-order Hausdorff accumulation operation and proved the superiority of the proposed model by cases [17]. Based on Wu's research [13], Liu et al. proposed the weighted r-AGO [18] and so on.

- (3) Parameter estimation. Similar to other regression prediction models, GM models usually utilize the least square method (LSM) to estimate the structural parameters of the models. Indeed, the LSM is a good method because it is easy to understand, but it is antiquated and limits the predictive performance of the models to some extent. To further optimize the GM models, some scholars proposed to use other parameter estimation methods to estimate the structural parameters of the models. These methods include partial least square (PLS) [19], nonlinear least square (NLS) [20,21], and the weighted least square method (WLS) [22,23]. Among these methods, the weighted least square method has the simplest mechanism and powerful performance.

With the above discussion, we know that the GM(1,1, t^α) is an ideal model, the r-AGO and the WLS are measures that can stably enhance the performance of the GM models. To accurately predict China's aging population, this paper establishes an optimized GM(1,1, t^α) model based on the GM(1,1, t^α) model, r-AGO and WLS. Compared with previous studies, the contributions of this paper are as follows:

- (1) We use the midpoint formula to solve the problem that GM(1,1, t^α) model is difficult to be solved.

- (2) We propose a new combined fractional-order accumulation operator.
- (3) We use the weighted least square method with strong compatibility to estimate the parameters of the proposed model.
- (4) This paper uses the proposed model to predict China's aging population in the next 4 years, and gives corresponding suggestions based on the prediction results.

The other parts of this article are arranged as follows: the next section gives the necessary knowledge including the definition of the GM(1,1, t^α) model and combined r-AGO. The third section discusses the concepts, properties and solution methods of the proposed model. The fourth section proves the feasibility and effectiveness of the proposed model through five real cases. The fifth section describes the application of the proposed model in China's aging population. The conclusion is given in the last section.

Prerequisite knowledge

Traditional GM(1,1, t^α) model and its optimized discrete form

Supposing there is a time sequence $X^{(0)} = \{x^{(0)}(m), m = 1, 2, 3, \dots, n\}$, $n \geq 4$, then

$$X^{(1)} = \left\{ x^{(1)}(m) = \sum_{j=1}^m x^{(0)}(j), m = 1, 2, 3, \dots, n \right\} \quad (1)$$

is called the first-order accumulative generation operation (1-AGO) sequence of $X^{(0)}$.

Based on the above conditions, the grey equation of the GM(1,1, t^α) model is expressed as

$$[x^{(1)}(t)]'_t + ax^{(1)}(t) = b \cdot t^\alpha + c. \quad (2)$$

Usually, the predictive formula of GM models is the solution of the grey differential equation. Obviously there is no exact solution for Eq. (1). When Qian et al. put forward the model, they did not provide a solving method for the model, but only used $\alpha = 2$ to establish the model for prediction, which greatly restricted the predictive ability of the GM(1,1, t^α) model. In the past, people usually used the forward difference formula [22] or the backward difference formula [24] to approximate the derivative to discretize the differential equation similar to Eq. (1). However, these two methods are not the best methods. According to the difference quotient method shown in Fig. 2, we can see that the accuracy of the midpoint formula is much higher than that of the forward difference formula and the backward difference

formula. Next, we estimate the errors of the three difference quotient formulas. Performing Taylor expansion of $f(a+h)$ at point a to get

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2}f''(\zeta), \zeta \in [a, a+h]. \quad (3)$$

Substituting the above equation into the forward difference formula $f'(a) = \frac{f(a+h)-f(a)}{h} + \varepsilon$, we can get

$$\begin{aligned} f'(a) &= \frac{f(a) + hf'(a) + \frac{h^2}{2}f''(\zeta) - f(a)}{h} + \varepsilon \\ &= f'(a) + \frac{h}{2}f''(\zeta) + \varepsilon, \end{aligned} \quad (4)$$

that is

$$\varepsilon = -\frac{h}{2}f''(\zeta). \quad (5)$$

Therefore, the truncation error of the forward difference formula is $O(h)$. Similarly, we can know that the truncation errors of the backward difference formula and the midpoint formula are $O(h)$ and $O(h^2)$. Obviously, the midpoint formula has the least error. Therefore, we use the midpoint formula to discretize Eq. (2), and then we can get

$$x^{(1)}(t+1) - x^{(1)}(t-1) = -2ax^{(1)}(t) + 2bt^\alpha + 2c. \quad (6)$$

Let $\mu_1 = -2a$, $\mu_2 = 2b$, $\mu_3 = 2c$, and then Eq. (6) can be expressed by

$$x^{(1)}(t+1) = x^{(1)}(t-1) + \mu_1 x^{(1)}(t) + \mu_2 t^\alpha + \mu_3. \quad (7)$$

Equation (7) is called the optimized discrete GM(1,1, t^α) model (ODGM(1,1, t^α) model). Similar to the regression prediction model, the parameters μ_1 , μ_2 , μ_3 of the ODGM(1,1, t^α) model are estimated by using the LSM, that is,

$$\hat{\delta} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3) = (B^T B)^{-1} B^T Y, \quad (8)$$

where

$$Y = \begin{pmatrix} x^{(1)}(3) - x^{(1)}(1) \\ x^{(1)}(4) - x^{(1)}(2) \\ \vdots \\ x^{(1)}(n) - x^{(1)}(n-2) \end{pmatrix}, B = \begin{pmatrix} x^{(1)}(2) & 2^\alpha & 1 \\ x^{(1)}(3) & 3^\alpha & 1 \\ \vdots & \vdots & \vdots \\ x^{(1)}(n-1) & (n-1)^\alpha & 1 \end{pmatrix}.$$

Once the estimated parameters $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3$ of the ODGM(1,1, t^α) are obtained by the least square method

Fig. 2 Schematic diagram of the difference quotient method

(LSM), the time response function the model can be expressed as

$$\hat{x}^{(1)}(t+1) = \hat{x}^{(1)}(t-1) + \hat{\mu}_1 x^{(1)}(t) + \hat{\mu}_2 t^\alpha + \hat{\mu}_3 c. \quad (9)$$

Since the ODGM(1,1, t^α) model is built using the 1-AGO, it is necessary to carry out the 1-order inverse AGO (1-AGO) on Eq. (9) to obtain the predicted values, that is,

$$\hat{x}^{(0)}(t) = \hat{x}^{(1)}(t) - \hat{x}^{(1)}(t-1), t = 2, 3, \dots \quad (10)$$

According to the introduction of the ODGM(1,1, t^α) model, we can still find some deficiencies.

- (1) The ODGM(1,1, t^α) model is built based on the 1-AGO. Although a sequence with weakened randomness can be obtained by 1-AGO, this does not mean that the prediction performance of the model is improved.
- (2) Similar to the traditional regression model, the ODGM(1,1, t^α) model uses the LSM to solve the structural parameters of the model. Indeed, the LSM is a good method, but it is too simple and limits the predictive performance of the model to some extent.

These two points provide ideas for the optimization scheme of this article. This article will discuss the optimization measures of the ODGM(1,1, t^α) model in detail in the next section.

The formula of three difference quotient methods for solving derivatives.

Forward difference: $f'(a) = \frac{f(a) - f(a-h)}{h} + \varepsilon.$

Backward difference: $f'(a) = \frac{f(a+h) - f(a)}{h} + \varepsilon.$

Midpoint difference: $f'(a) = \frac{f(a+h) - f(a-h)}{2h} + \varepsilon.$

where ε is the truncation error.

According to the definition of the slope, the slopes of the straight lines BA, AC and BC can be calculated as

$$k_{BA} = \frac{f(a) - f(a-h)}{h},$$

$$k_{AC} = \frac{f(a+h) - f(a)}{h},$$

$$k_{BC} = \frac{f(a+h) - f(a-h)}{2h}.$$

The combined fractional-order accumulative generation operation

The advantage of fractional-order accumulation generation operation (r-AGO) is that it can choose the most suitable order to build the model through algorithms. Compared with the traditional 1-AGO, it is more adaptable. Obviously, the r-AGO can solve the first deficiency described in the previous section. At present, there are two popular r-AGOs, namely the r-AGO proposed by Wu [13] and the fractional-order Hausdorff accumulation generation operation (δ -HAGO) proposed by Chen et al. [17]. To further enhance the predictive ability of the model, we combine these two measures and propose a new r-AGO, which is defined as follows.

Theorem 1 If $X^{(0)}$ is the original time series and

$$H^{(0)} = \left\{ h^{(0)}(k) = \sum_{p=1}^k x^{(0)}(p) \cdot [p^\delta - (p-1)^\delta], \delta \in R^+, k = 1, 2, 3, \dots, n \right\} \quad (11)$$

is the fractional-order Hausdorff accumulation generation sequence of $X^{(0)}$, then the combined fractional-order accumulative generation operator (CF-AGO) of $X^{(0)}$ is

$$\begin{aligned}
h^{(r)}(k) &= \sum_{l=1}^k \frac{\Gamma(r+k-l)}{\Gamma(k-l+1)\Gamma(r)} \cdot h^{(0)}(l) \\
&= \sum_{l=1}^k \frac{\Gamma(r+k-l)}{\Gamma(k-l+1)\Gamma(r)} \\
&\quad \cdot \sum_{p=1}^l x^{(0)}(p) \cdot [p^\delta - (p-1)^\delta]. \quad (12)
\end{aligned}$$

Proof According to the method proposed by Chen et al. [17], the δ -HAGO of $X^{(0)}$ can be given as

$$h^{(0)}(k) = \sum_{p=1}^k x^{(0)}(p) \cdot [p^\delta - (p-1)^\delta], k = 1, 2, 3, \dots \quad (13)$$

When $r = 1$, according to the 1-AGO, we can get

$$\begin{aligned}
h^{(1)}(k) &= \sum_{l=1}^k \frac{(k-l+1) \dots (k-l+r-1)}{(r-1)!} \cdot h^{(0)}(l) \\
&= \sum_{l=1}^k h^{(0)}(l). \quad (14)
\end{aligned}$$

Similarly, when $r = 2$, the 2-AGO of the δ -HAGO of $X^{(0)}$ is

$$\begin{aligned}
h^{(2)}(k) &= \sum_{l=1}^k \frac{(k-l+1)}{(2-1)!} \cdot h^{(0)}(l) \\
&= \sum_{l=1}^k (k-l+1) \cdot h^{(0)}(l). \quad (15)
\end{aligned}$$

Supposing the n -AGO of the δ -HAGO of $X^{(0)}$ is

$$h^{(n)}(k) = \sum_{l=1}^k \frac{(k-l+1) \dots (k-l+n-1)}{(n-1)!} \cdot h^{(0)}(l), \quad (16)$$

then, when $r = n + 1$, we have

$$\begin{aligned}
h^{(n+1)}(k) &= \sum_{l=1}^k \sum_{j=1}^l \frac{(l-j+1) \dots (l-j+n-1)}{(n-1)!} \\
&\quad \cdot h^{(0)}(j) \\
&= \sum_{j=1}^k \left[\frac{(1-j+2) \dots (2-j+n-1)(2-j+n)}{n!} \right]
\end{aligned}$$

$$\begin{aligned}
&\cdot h^{(0)}(j) + \dots + \sum_{j=1}^k \frac{(k-j+1) \dots (k-j+n-1)}{(n-1)!} \\
&\quad \cdot h^{(0)}(j) \\
&= \sum_{j=1}^3 \left[\frac{(3-j+1) \dots (3-j+n-1)(n+3-j)}{n!} \right] \\
&\quad \cdot h^{(0)}(j) + \dots + \sum_{j=1}^k \frac{(k-j+1) \dots (k-j+n-1)}{(n-1)!} \\
&\quad \cdot h^{(0)}(j) = \dots \\
&= \sum_{j=1}^k \frac{(k-j+1) \dots (k-j+n)}{n!} \cdot h^{(0)}(j). \quad (17)
\end{aligned}$$

Therefore, the r -AGO of the δ -HAGO of $X^{(0)}$ is

$$\begin{aligned}
h^{(r)}(k) &= \sum_{l=1}^k \frac{(k-l+1) \dots (k-l+r-1)}{(r-1)!} \\
&\quad \cdot h^{(0)}(l), r \in N^+. \quad (18)
\end{aligned}$$

According to the concept of Gamma function, Eq. (18) can be converted to

$$\begin{aligned}
h^{(r)}(k) &= \sum_{l=1}^k \frac{\Gamma(r+k-l)}{\Gamma(k-l+1)\Gamma(r)} \cdot h^{(0)}(l) \\
&= \sum_{l=1}^k \frac{\Gamma(r+k-l)}{\Gamma(k-l+1)\Gamma(r)} \cdot \sum_{p=1}^l x^{(0)}(p) \\
&\quad \cdot [p^\delta - (p-1)^\delta]. \quad (19)
\end{aligned}$$

By the broad definition of the combinatorial number [25], it can be seen that when $r \in R^+$, Eq. (19) is also true.

This completes the proof. $\square$

Theorem 2 If $H^{(0)}$ is the fractional-order Hausdorff accumulation generation sequence of $X^{(0)}$, and

$$H^{(-r)} = \{h^{(-r)}(1), h^{(-r)}(2), \dots, h^{(-r)}(n)\}, r \in R^+ \quad (20)$$

is the fractional-order inverse accumulative generation sequence of $H^{(0)}$, then the k th term in $H^{(-r)}$ can be represented by

$$\begin{aligned}
h^{(-r)}(k) &= \sum_{l=0}^{k-1} (-1)^l \cdot \frac{\Gamma(r+1)}{\Gamma(l+1)\Gamma(r+1-l)} \\
&\quad \cdot h^{(0)}(k-l), k = 1, 2, 3, \dots, n. \quad (21)
\end{aligned}$$

Proof Case 1: When $r = 1, k = 1$, according to the 1-order inverse AGO, we can get

$$h^{(-1)}(1) = \sum_{l=0}^{k-1} (-1)^l \cdot \frac{r!}{l!(r-l)!} \cdot h^{(0)}(k-l) = h^{(0)}(1). \quad (22)$$

At the same time, when $k = 2, 3, 4, \dots$, we can get

$$\begin{aligned} h^{(-1)}(k) &= \sum_{l=0}^r (-1)^l \cdot \frac{r!}{l!(r-l)!} h^{(0)}(k-l) \\ &= h^{(0)}(k) - h^{(0)}(k-1). \end{aligned} \quad (23)$$

Case 2: When $r = 2, k = 1$, we obtain

$$h^{(-2)}(1) = \sum_{l=0}^{k-1} (-1)^l \cdot \frac{r!}{l!(r-l)!} \cdot h^{(0)}(k-l) = h^{(0)}(1). \quad (24)$$

When $k = 2$, we know that

$$\begin{aligned} h^{(-2)}(2) &= \sum_{l=0}^{k-1} (-1)^l \cdot \frac{r!}{l!(r-l)!} \cdot h^{(0)}(k-l) \\ &= h^{(0)}(2) - 2h^{(0)}(1). \end{aligned} \quad (25)$$

When $k = 3, 4, \dots$, we can get

$$\begin{aligned} h^{(-2)}(k) &= \sum_{l=0}^r (-1)^l \cdot \frac{r!}{l!(r-l)!} \cdot h^{(0)}(k-l) \\ &= h^{(0)}(k) - 2h^{(0)}(k-1) + h^{(0)}(k-2). \end{aligned} \quad (26)$$

Case 3: Assuming that when $r = n$,

holds, then when $r = n + 1$, we can get

$$\begin{aligned} h^{(-n-1)}(k) &= h^{(-n)}(k) - h^{(-n)}(k-1) \\ &= \sum_{l=0}^n (-1)^l \cdot \frac{n!}{l!(n-l)!} \cdot h^{(0)}(k-l) \\ &\quad - \sum_{l=0}^n (-1)^l \cdot \frac{n!}{l!(n-l)!} \cdot h^{(0)}(k-1-l) \\ &= h^{(0)}(k) - \left[\frac{n!}{1!(n-1)!} + \frac{n!}{0!(n-0)!} \right] \\ &\quad \cdot h^{(0)}(k-1) + \left[\frac{n!}{2!(n-2)!} \right. \\ &\quad \left. + \frac{n!}{1!(n-1)!} \right] \cdot h^{(0)}(k-2) \\ &\quad - \left[\frac{n!}{3!(n-3)!} + \frac{n!}{2!(n-2)!} \right] \cdot h^{(0)}(k-3) + \dots \\ &\quad + (-1)^n \left[\frac{n!}{n!(n-n)!} + \frac{n!}{(n-1)!(n-n+1)!} \right] \\ &\quad \cdot h^{(0)}(k-n) \\ &\quad + (-1)^{n+1} \frac{n!}{n!(n-n)!} \cdot h^{(0)}(k-1-n), \end{aligned} \quad (28)$$

and

$$\begin{aligned} h^{(-n-1)}(k) &= h^{(-n)}(k) - h^{(-n)}(k-1) \\ &= \sum_{l=0}^{k-1} (-1)^l \cdot \frac{n!}{l!(n-l)!} \cdot h^{(0)}(k-l) \\ &\quad - \sum_{l=0}^{k-2} (-1)^l \cdot \frac{n!}{l!(n-l)!} \cdot h^{(0)}(k-1-l) \\ &= h^{(0)}(k) - \left[\frac{n!}{1!(n-1)!} + \frac{n!}{0!(n-0)!} \right] \cdot h^{(0)}(k-1) \\ &\quad + \left[\frac{n!}{2!(n-2)!} + \frac{n!}{1!(n-1)!} \right] \cdot h^{(0)}(k-2) \\ &\quad + \dots + (-1)^{k-1} \cdot \left[\frac{n!}{(k-1)!(n-k+1)!} \right. \\ &\quad \left. + \frac{n!}{(k-2)!(n-k+2)!} \right] \cdot h^{(0)}(1). \end{aligned} \quad (29)$$

It is easy to see that the coefficient of the l th term $h^{(0)}(k-l)$ in the expansion of $h^{(-n-1)}(k)$ is

$$h^{(-n)}(k) = \begin{cases} \sum_{l=0}^{k-1} (-1)^l \cdot \frac{\Gamma(n+1)}{\Gamma(l+1)\Gamma(n-l+1)} \cdot h^{(0)}(k-l), & k = 1, 2, 3, \dots, n \\ \sum_{l=0}^n (-1)^l \cdot \frac{\Gamma(n+1)}{\Gamma(l+1)\Gamma(n-l+1)} \cdot h^{(0)}(k-l), & k = n+1, n+2, \dots \end{cases} \quad (27)$$

$$(-1)^l \cdot \left[\frac{n!}{l!(n-l)!} + \frac{n!}{(l-1)!(n-l+1)!} \right] = (-1)^l \cdot \frac{(n+1)!}{l!(n-l+1)!}, \quad (30)$$

that is,

$$h^{(-n-1)}(k) = \begin{cases} \sum_{l=0}^{k-1} (-1)^l \cdot \frac{(n+1)!}{l!(n-l+1)!} \cdot h^{(0)}(k-l) \\ \sum_{l=0}^{n+1} (-1)^l \cdot \frac{(n+1)!}{l!(n-l+1)!} \cdot h^{(0)}(k-l) \end{cases}. \quad (31)$$

In summary, we can get

$$h^{(-r)}(k) = \begin{cases} \sum_{l=0}^{k-1} (-1)^l \cdot \frac{\Gamma(r+1)}{\Gamma(l+1)\Gamma(r-l+1)} \cdot h^{(0)}(k-l), & k = 1, 2, 3, \dots, r \\ \sum_{l=0}^r (-1)^l \cdot \frac{\Gamma(r+1)}{\Gamma(l+1)\Gamma(r-l+1)} \cdot h^{(0)}(k-l), & k = r+1, r+2, \dots \end{cases}, \quad r \in N^+. \quad (32)$$

Then, when $1 \leq k \leq r$,

$$h^{(-r)}(k) = \sum_{l=0}^{k-1} (-1)^l \cdot \frac{\Gamma(r+1)}{\Gamma(l+1)\Gamma(r-l+1)} \cdot h^{(0)}(k-l), \quad (33)$$

when $r+1 \leq k \leq n$, $l \leq k-1$, that is, $r+1 \leq l \leq k-1$, $\frac{r!}{l!(r-l)!} = 0$. (According to the nature of the Gamma function, when the function value is negative integer, it is infinite, and its reciprocal is 0).

Therefore, we have

$$\begin{aligned} h^{(-r)}(k) &= \sum_{l=0}^r (-1)^l \cdot \frac{r!}{l!(r-l)!} \cdot h^{(0)}(k-l) \\ &= \sum_{l=0}^r (-1)^l \cdot \frac{r!}{l!(r-l)!} \cdot h^{(0)}(k-l) \\ &\quad + \sum_{l=r+1}^{k-1} (-1)^l \cdot \frac{r!}{l!(r-l)!} \cdot h^{(0)}(k-l) \\ &= \sum_{l=0}^{k-1} (-1)^l \cdot \frac{\Gamma(r+1)}{\Gamma(l+1)\Gamma(r-l+1)} \cdot h^{(0)}(k-l). \end{aligned} \quad (34)$$

By generalizing Eq. (34), the k th term in $H^{(-r)}$ is

$$h^{(-r)}(k) = \sum_{l=0}^{k-1} (-1)^l \cdot \frac{\Gamma(r+1)}{\Gamma(l+1)\Gamma(r-l+1)} \cdot h^{(0)}(k-l), \quad k = 1, 2, 3, \dots, n. \quad (35)$$

By the broad definition of the combinatorial number [25], it can be seen that when $r \in R^+$, Eq. (35) is also true, and it completes the proof. $\square$

Theorem 2 tells us that

$$h^{(0)}(k) = \sum_{l=0}^{k-1} (-1)^l \cdot \frac{\Gamma(r+1)}{\Gamma(l+1)\Gamma(r-l+1)} \cdot h^{(r)}(k-l), \quad k = 1, 2, 3, \dots, n. \quad (36)$$

In this paper, Eq. (12) is used to generate accumulative sequence, while Eq. (36) is used to restore output sequence.

The proposed model

Based on the knowledge in Section “Prerequisite knowledge”, this section will give the definition of the CFODGMW(1,1, α) model.

For a given original time series $X^{(0)}$, if it satisfies Theorems 1 and 2, then the following equation:

$$h^{(r)}(k+1) = h^{(r)}(k-1) + \mu_1 \cdot h^{(r)}(k) + \mu_2 \cdot k^\alpha + \mu_3 \quad (37)$$

is called the combined fractional ODGM(1,1, α) based on weighted least square (WLS) (abbreviated as CFODGMW(1,1, α)), where μ_1 is the development modulus, and $\mu_2 \cdot k^\alpha + \mu_3$ is grey action.

Parameters' estimation of the CFODGMW(1,1, α) model

Let $k = 2, 3, \dots, n$ in Eq. (37), and we get the following system:

$$\begin{cases} h^{(r)}(3) - h^{(r)}(1) = \mu_1 \cdot h^{(r)}(2) + \mu_2 \cdot 2^\alpha + \mu_3 \\ h^{(r)}(4) - h^{(r)}(2) = \mu_1 \cdot h^{(r)}(3) + \mu_2 \cdot 3^\alpha + \mu_3 \\ \vdots \\ h^{(r)}(n) - h^{(r)}(n-2) = \mu_1 \cdot h^{(r)}(n-1) + \mu_2 \cdot (n-1)^\alpha + \mu_3 \end{cases} \quad (38)$$

and its matrix form is

$$X\psi = \Upsilon, \quad (39)$$

where

$$\Upsilon = \begin{pmatrix} h^{(r)}(3) - h^{(r)}(1) \\ h^{(r)}(4) - h^{(r)}(2) \\ \vdots \\ h^{(r)}(n) - h^{(r)}(n-2) \end{pmatrix},$$

$$X = \begin{pmatrix} h^{(r)}(2) & 2^\alpha & 1 \\ h^{(r)}(3) & 3^\alpha & 1 \\ \vdots & \vdots & \vdots \\ h^{(r)}(n-1) & (n-1)^\alpha & 1 \end{pmatrix}.$$

According to linear algebra knowledge, there is no exact solution to Eq. (39). Let the residual vector is $\tau = [\varepsilon(2), \dots, \varepsilon(n)]^T$, and we have

$$\tau = \Upsilon - X\psi.$$

Then, the estimated value of the parameter set ψ is

$$\hat{\psi} = \arg \min_{\psi} L = \tau^T W \tau = (\Upsilon - X\psi)^T W (\Upsilon - X\psi). \quad (40)$$

According to the extremum existence condition, it is easy to know that the WLS estimation of the structural parameters of the CFODGMW(1,1, α) model is

$$\hat{\psi} = [\hat{\mu}_1 \hat{\mu}_2 \hat{\mu}_3]^T = (X^T W X)^{-1} X^T W \Upsilon. \quad (41)$$

where W is the weight matrix.

The forecasting formula of the CFODGMW(1,1, α) model

According to Eq. (37), we can get

$$\begin{aligned} h^{(r)}(k+1) &= h^{(r)}(k-1) + \mu_1 h^{(r)}(k) + \mu_2 k^\alpha + \mu_3 \\ &= (\mu_1^2 + 1)h^{(r)}(k-1) + \mu_1 h^{(r)}(k-2) \\ &\quad + \mu_2 k^\alpha + \mu_1 \mu_2 (k-1)^\alpha + (\mu_1 + 1)\mu_3 \\ &= (\mu_1^2 + 1)h^{(r)}(k-3) \\ &\quad + [\mu_1(\mu_1^2 + 1) + \mu_1]h^{(r)}(k-2) \\ &\quad + (\mu_1^2 + 1)\mu_2(k-2)^\alpha \\ &\quad + \mu_1 \mu_2 (k-1)^\alpha + \mu_2 k^\alpha + [(\mu_1^2 + 1) + (\mu_1 + 1)]\mu_3 \\ &= \{\mu_1[\mu_1(\mu_1^2 + 1) + \mu_1] + (\mu_1^2 + 1)\}h^{(r)}(k-3) \\ &\quad + [\mu_1(\mu_1^2 + 1) + \mu_1]h^{(r)}(k-4) + \mu_2 k^\alpha \\ &\quad + \mu_1 \mu_2 (k-1)^\alpha + (\mu_1^2 + 1)\mu_2(k-2)^\alpha \\ &\quad + \mu_2[\mu_1(\mu_1^2 + 1) \\ &\quad + \mu_1](k-3)^\alpha + \mu_3[\mu_1(\mu_1^2 + 1) + \mu_1] \end{aligned}$$

$$\begin{aligned} &= \{\mu_1[\mu_1(\mu_1^2 + 1) + \mu_1] + (\mu_1^2 + 1)\}h^{(r)}(k-5) \\ &\quad + \left\{ \frac{\mu_1\{\mu_1[\mu_1(\mu_1^2 + 1) + \mu_1] + (\mu_1^2 + 1)\}}{(\mu_1^2 + 1)} + [\mu_1(\mu_1^2 + 1) + \mu_1] \right\}h^{(r)}(k-4) \\ &\quad + \mu_2 k^\alpha + \mu_1 \mu_2 (k-1)^\alpha + (\mu_1^2 + 1)\mu_2(k-2)^\alpha \\ &\quad + \mu_2[\mu_1(\mu_1^2 + 1) \\ &\quad + \mu_1](k-3)^\alpha + (\mu_1^2 + 1)\mu_2(k-4)^\alpha \\ &\quad + \mu_3 \left\{ \mu_1[\mu_1(\mu_1^2 + 1) + \mu_1] + (\mu_1^2 + 1) \right\} \\ &\quad + [\mu_1(\mu_1^2 + 1) + \mu_1] \Big\} \\ &= \dots = S_{k-1}h^{(r)}(2) + S_{k-2}h^{(r)}(1) \\ &\quad + \mu_2 \left[\sum_{i=1}^{k-2} S_i(k-i+1)^\alpha \right] + Z_{k-1}\mu_3. \quad (42) \end{aligned}$$

where $S_1 = \mu_1$, $S_2 = (1 + \mu_1^2)$, $S_n = \mu_1 S_{n-1} + S_{n-2}$, $n \geq 3$, $Z_1 = 1$, $Z_i = S_{i-1} + Z_{i-1}$.

Then, we substitute the $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3$ obtained by the WLS into Eq. (42), and the recursive formula of the model can be shown by

$$\begin{aligned} \hat{h}^{(r)}(k+1) &= \hat{S}_{k-1}\hat{h}^{(r)}(2) + \hat{S}_{k-2}\hat{h}^{(r)}(1) \\ &\quad + \hat{\mu}_2 \left[\sum_{i=1}^{k-2} \hat{S}_i(k-i+1)^\alpha \right] + \hat{Z}_{k-1}\hat{\mu}_3. \quad (43) \end{aligned}$$

According to Theorem 2, we can get

$$\begin{aligned} \hat{h}^{(0)}(k) &= \sum_{l=0}^{k-1} (-1)^l \cdot \frac{\Gamma(r+1)}{\Gamma(l+1)\Gamma(r-l+1)} \\ &\quad \cdot \hat{h}^{(r)}(k-l), \quad k = 2, 3, \dots, \\ \hat{h}^{(0)}(1) &= h^{(0)}(1). \end{aligned} \quad (44)$$

Therefore, the prediction formula of CFODGMW(1,1, α) is

$$\hat{x}^{(0)}(k) = \frac{\hat{h}^{(0)}(k) - \hat{h}^{(0)}(k-1)}{k^\delta - (k-1)^\delta}, \quad k = 2, 3, \dots. \quad (45)$$

Properties of the CFODGMW(1,1, α) model

Convertibility

1. When the weight matrix W and structural parameters of the CFODGMW(1,1, α) model takes different values, we have the following conclusions.

- (1) When $W = I$, the criterion for parameter estimation of the CFODGMW(1,1, α) model is the least square criterion, that is,

$$\hat{\vartheta} = \arg \min_{\vartheta} \frac{1}{n-1} \sum_{k=2}^n [h^{(r)}(k) - \hat{h}^{(r)}(k)]^2. \quad (46)$$

- (2) When $W = \text{diag}\{\frac{1}{[h^{(r)}(2)]^2}, \frac{1}{[h^{(r)}(3)]^2}, \dots, \frac{1}{[h^{(r)}(n)]^2}\}$, the criterion for parameter estimation of the CFODGMW(1,1, α) model is the minimum mean square relative error criterion, that is,

$$\hat{\vartheta} = \arg \min_{\vartheta} \frac{1}{n-1} \sum_{k=2}^n \left[\frac{h^{(r)}(k) - \hat{h}^{(r)}(k)}{h^{(r)}(k)} \right]^2. \quad (47)$$

- (3) When $\delta = 1, r = 0, \alpha = 0, W = I$, the CFODGMW(1,1, α) model degenerates to the DGM(1,1) model [24] based on midpoint difference.
- (4) When $\delta = 1, r = 0, \alpha = 1, W = I$, the CFODGMW(1,1, α) model degenerates to the NDGM(1,1,k) model [26] based on midpoint difference.

Unbiasedness

Unbiasedness means that the forecasting model can completely fit a time series. Since the modeling mechanism of the CFODGMW(1,1, α) model is too complex, it is very difficult to directly prove the unbiasedness of the model using the formula. Therefore, we use examples to prove that the model is unbiased.

Theorem 3 *The CFODGMW(1,1, α) model is unbiased for time series satisfying the homogeneous or non-homogeneous laws.*

Proof Case 1: When $\delta = 1, r = 0, \alpha = 0, W = I$, Eq. (37) (CFODGMW(1,1, α)) is transformed into

$$x^{(1)}(k+1) = x^{(1)}(k-1) + \mu_1 x^{(1)}(k) + \mu_2. \quad (48)$$

Obviously, Eq. (48) is the expression of the DGM(1,1) model based on midpoint difference. When the time series satisfies the homogeneous law, that is, $X_{\text{Fit}} + X_{\text{Pre}} = \{2 \cdot 3^k, k = 1, 2, 3, 4, 5\} + \{2 \cdot 3^k, k = 6, 7\}$. The X_{Fit} is applied to establish the model, and X_{Pre} is applied to test the predictive ability of the model.

According to Eq. (41), we can get

$$\begin{aligned} \hat{\psi} &= [\hat{\mu}_1, \hat{\mu}_2]^T \\ &= [2.666666666666667, 7.999999999999950]^T. \end{aligned}$$

Then, we bring $\hat{\mu}_1, \hat{\mu}_2$ into Eq. (48), and then according to the 1-IAGO, we can obtain

$$\hat{x}^{(0)}(k) = 2 \cdot 3^k = x^{(0)}(k), k = 1, 2, \dots, 7. \quad (49)$$

Case 2: When $\delta = 1, r = 0, \alpha = 1, W = I$, Eq. (37) is turned into

$$x^{(1)}(k+1) = x^{(1)}(k-1) + \mu_1 \cdot x^{(1)}(k) + \mu_2 \cdot k^\alpha + \mu_3. \quad (50)$$

Then, Eq. (50) is the expression of the NDGM(1,1,k) model. When the time series satisfies the non-homogeneous law, that is, $X_{\text{Fit}} + X_{\text{Pre}} = \{2 \cdot 3^k + 2, k = 1, 2, 3, 4, 5\} + \{2 \cdot 3^k + 2, k = 6, 7\}$. The X_{Fit} is applied to establish the model, and X_{Pre} is applied to test the predictive ability of the model.

According to Eq. (41), we can get

$$\begin{aligned} \hat{\psi} &= [\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3]^T = [2.666666666666660, \\ &\quad -5.333333333333168, 11.999999999998451]^T. \end{aligned} \quad (51)$$

Then, we bring $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3$ into Eq. (50), and then according to the 1-IAGO, we can obtain

$$\hat{x}^{(0)}(k) = 2 \cdot 3^k + 2 = x^{(0)}(k), k = 1, 2, \dots, 7. \quad (52)$$

Therefore, this complete the proof. $\square$

Note that since there are many time series that conform to the homogeneous exponential law or approximate non-homogeneous exponential law in real life, the CFODGMW(1,1, α) model has a wide application background.

The solution method of the CFODGMW(1,1, α) model

From the expression of the CFODGMW(1,1, α) model, it can be seen that the CFODGMW(1,1, α) model is built on the basis of the r, α, δ . Then, the most important task is to determine the r, α, δ of the CFODGMW(1,1, α) model. To solve the parameters r, α, δ , we consider establishing a programming model with error minimization as the objective function

and the modeling process of the model as constraints. This programming model can be expressed by

$$\min_{r, \alpha, \delta} MAPE = \frac{1}{n-1} \sum_{k=2}^n \left| \frac{\hat{x}^{(0)}(k) - \hat{x}^{(0)}(k)}{\hat{x}^{(0)}(k)} \right|^2 \times 100\%$$

$$s.t. \begin{cases} \hat{\psi} = [\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3]^T = (X^T W X)^{-1} X^T W Y. \\ \hat{h}^{(r)}(k+1) = \hat{S}_{k-1} \hat{h}^{(r)}(2) + \hat{S}_{k-2} \hat{h}^{(r)}(1) \\ + \hat{\mu}_2 [\sum_{i=1}^{k-2} \hat{S}_i (k-i+1)^\alpha] + \hat{Z}_{k-1} \hat{\mu}_3. \\ \hat{h}^{(0)}(k) = \sum_{l=0}^{k-1} (-1)^l \frac{\Gamma(r+1)}{\Gamma(l+1)\Gamma(r-l+1)} \hat{h}^{(r)}(k-l), k=2, 3, \dots, \\ \hat{h}^{(0)}(1) = h^{(0)}(1). \\ \hat{x}^{(0)}(k) = \frac{\hat{h}^{(0)}(k) - \hat{h}^{(0)}(k-1)}{k^\delta - (k-1)^\delta}, k=2, 3, \dots. \end{cases} \quad (53)$$

Obviously, the traditional method is difficult to solve the above-mentioned planning problem. Here, we use the whale optimization algorithm (WOA) [27] to solve the above planning problem to obtain the parameters r, α, δ that make the model have the smallest error. The calculation process of the CFODGMW(1,1, α) model is shown in Fig. 3.

Validity

In this section, the aging population of five regions in China is taken as cases to verify the validity of the proposed model. The five data sets are from the statistical yearbook for each region, as shown in Table 1. The models used for comparison include exponential curve model (ECM), GM(1,1) model [1], polynomial regression model (PR) [28], FGM(1,1) model [29], OGM(1,1) model [30] and GMP(1,1,N) model with excellent predictive performance [10]. The aging population from 2010 to 2016 are applied to build models, and the last three data are applied to test the predictive ability of the models. The MAPEs of the seven predictive models are shown in Table 2. In addition, to facilitate readers to observe the gap between the models, we visualized the data in Table 2, as shown in Fig. 4.

It can be seen from Fig. 4 that in the fitting stage, the CFODGMW(1,1, α) model has a winning rate of 60%, which is much higher than that of other models. Therefore, the fitting performance of CFODGMW(1,1, α) model is better than that of its competitors. According to Fig. 4 (right) we can clearly see that the winning rate of the CFODGMW(1,1, α) model is 100%, which means that the prediction performance of the CFODGMW(1,1, α) model is the best among all models. Therefore, the validity of the CFODGMW(1,1, α) model is confirmed.

Fig. 3 Schematic diagram of the calculation process of the CFODGMW(1,1, α) model

Table 1 The number of aging populations in five regions of China from 2010 to 2019 (10,000)

Year	Liaoning	Shanghai	Anhui	Guangxi	Beijing
2010	438.35	221.00	608.69	476.69	171.00
2011	453.16	235.21	680.92	511.06	181.70
2012	473.72	245.27	723.35	510.38	190.40
2013	488.22	256.63	738.05	510.24	194.50
2014	513.55	270.07	712.31	529.43	212.30
2015	542.25	283.38	720.65	550.14	222.80
2016	571.74	299.02	743.46	555.11	230.40
2017	602.20	317.67	774.34	557.20	237.60
2018	635.91	336.90	820.17	563.64	241.40
2019	680.07	361.66	886.77	569.50	246.00

Application

Collection of raw data

The original data on China's aging population from 2005 to 2019 are collected from the NBS, and it is shown in Table 3.

Table 2 MAPEs of seven predictive models in five cases

Case		GM(1,1)	ECM	PR	GMP(1,1,N)	OGM(1,1)	FGM(1,1)	CFODGMW(1,1, α)
Liaoning	MAPE (Fit) (%)	0.6434	0.6717	0.2556	0.2695	2.5575	0.2801	0.1479
	MAPE (Pre) (%)	2.3405	3.2601	0.5221	1.8600	1.2173	1.5381	0.4581
Shanghai	MAPE (Fit) (%)	0.2856	0.3247	0.2112	0.0703	2.7400	0.0799	0.2196
	MAPE (Pre) (%)	2.9789	2.7006	1.2786	1.6930	1.2358	1.8366	0.4512
Anhui	MAPE (Fit) (%)	1.9340	3.4918	0.6301	1.4750	2.0296	1.5695	0.8896
	MAPE (Pre) (%)	8.3352	3.7361	14.8668	11.7406	9.7400	10.6855	1.6402
Guangxi	MAPE (Fit) (%)	1.0890	1.2608	1.2354	0.6161	1.4122	0.8835	0.5641
	MAPE (Pre) (%)	2.2303	3.4559	2.8330	5.1263	4.2690	8.2154	1.9184
Beijing	MAPE (Fit) (%)	1.0669	0.9806	1.1390	0.9852	3.0655	0.9449	0.7797
	MAPE (Pre) (%)	6.0440	6.2135	5.9933	5.9993	10.9449	3.8396	0.9837

Fig. 4 The rank of the seven predictive models in the five cases (the larger the value, the better the performance)**Table 3** China's aging population (10,000) from 2005 to 2019

Year	Data	Year	Data	Year	Data	Year	Data	Year	Data
2005	10,055	2008	10,956	2011	12,288	2014	13,755	2017	15,831
2006	10,419	2009	11,307	2012	12,714	2015	14,386	2018	16,658
2007	10,636	2010	11,894	2013	13,161	2016	15,003	2019	17,603

Evaluation index of model performance

It is an important part of data analysis to evaluate the predictive performance of the model [31]. The MAPE and APE are applied to judge the predictive performance of the models. Their definitions are as follows

$$MAPE = \frac{1}{n-1} \cdot \sum_{j=2}^n \left| \frac{\hat{x}^{(0)}(j) - x^{(0)}(j)}{x^{(0)}(j)} \right| \times 100\%,$$

$$APE = \left| \frac{\hat{x}^{(0)}(j) - x^{(0)}(j)}{x^{(0)}(j)} \right| \times 100\%.$$

Numerical results

The parameters $r = 0$, $\delta = 1.00492011229591$ and $\alpha = 3.49253638869618$ of the CFODGMW(1,1, α) model are obtained by using the whale optimizer. The table containing the calculation time of each part of the algorithm is placed in the Appendix. The predicted results and prediction errors of the CFODGMW(1,1, α) model and its competing models are shown in Tables 4 and 5, respectively. Similarly, to facilitate review, we visualized the data in Tables 4 and 5, as shown in Fig. 5.

According to the MAPE shown in Fig. 5, we can see that the GMP(1,1,N) model and FGM(1,1) model have better fitting performance than the CFODGMW(1,1, α) model dur-

Table 4 The predicted results of the proposed model and its comparative models

Year	Data	GM(1,1) In-sample	ECM	PR	GMP(1,1,N)	OGM(1,1)	OGM(1,1)	CFODGMW(1,1, α)
2005	10055.00	10055.00	9919.61	10050.50	10055.00	10055.00	10055.00	10055.00
2006	10419.00	10210.61	10287.22	10431.68	10396.60	10386.89	10477.54	10419.00
2007	10636.00	10601.51	10668.44	10783.00	10663.72	10793.66	10658.44	10583.78
2008	10956.00	11007.38	11063.80	10853.00	10986.58	11216.36	10962.28	11088.45
2009	11307.00	11428.79	11473.80	11276.00	11359.11	11655.61	11333.01	11313.39
2010	11894.00	11866.33	11899.00	11658.00	11775.86	12112.07	11751.51	11859.10
2011	12288.00	12320.62	12339.96	12481.00	12232.00	12586.41	12209.39	12152.29
2012	12714.00	12792.30	12797.25	12682.00	12723.24	13079.31	12702.60	12756.14
2013	13161.00	13282.04	13271.50	13140.00	13245.73	13591.53	13229.15	13137.11
2014	13755.00	13790.53	13763.32	13608.00	13796.06	14123.80	13788.24	13823.03
2015	14386.00	14318.49	14273.36	14349.00	14371.20	14676.91	14379.73	14316.93
2016	15003.00	14866.66	14802.31	15017.00	14968.42	15251.69	15003.96	15113.67
MAPE		0.6284%	0.7930%	0.6790%	0.3344%	2.1834%	0.3452%	0.4891%
Year	Data	Out-of-sample						
2017	15831.00	15435.82	15350.85	15620.00	15585.31	16469.66	15661.58	15750.35
2018	16658.00	16026.77	15919.73	16237.00	16219.73	17114.64	16353.47	16691.15
2019	17603.00	16640.34	16509.69	16854.00	16869.76	17784.89	17080.71	17505.01
MAPE		3.9181%	4.5586%	2.7050%	2.7828%	2.6029%	1.9551%	0.4217%

Table 5 The errors of the proposed model and its comparative models (APE and MAPE)

Year	Data	GM(1,1) In-sample	ECM	PR	GMP(1,1,N)	OGM(1,1)	FGM(1,1)	CFODGMW(1,1, α)
2005	10055.00	0.0000%	1.3465%	0.0448%	0.0000%	0.0000%	.0000%	0.0000%
2006	10419.00	2.0001%	1.2648%	0.1217%	0.2150%	0.3082%	0.5618%	0.0000%
2007	10636.00	0.3243%	0.3050%	1.3821%	0.2606%	1.4823%	0.2110%	0.4909%
2008	10956.00	0.4690%	0.9839%	0.9401%	0.2792%	2.3764%	0.0573%	1.2090%
2009	11307.00	1.0771%	1.4752%	0.2742%	0.4608%	3.0832%	0.2300%	0.0565%
2010	11894.00	0.2327%	0.0420%	1.9842%	0.9933%	1.8335%	1.1980%	0.2935%
2011	12288.00	0.2654%	0.4228%	1.5706%	0.4557%	2.4284%	0.6397%	1.1044%
2012	12714.00	0.6159%	0.6548%	0.2517%	0.0727%	2.8733%	0.0897%	0.3314%
2013	13161.00	0.9197%	0.8396%	0.1596%	0.6438%	3.2712%	0.5178%	0.1815%
2014	13755.00	0.2583%	0.0605%	1.0687%	0.2985%	2.6812%	0.2416%	0.4946%
2015	14386.00	0.4693%	0.7830%	0.2572%	0.1029%	2.0222%	0.0436%	0.4802%
2016	15003.00	0.9087%	1.3377%	0.0933%	0.2305%	1.6576%	0.0064%	0.7376%
MAPE		0.6284%	0.7930%	0.6790%	0.3344%	2.1834%	0.3452%	0.4891%
Year	Data	Out-of-sample						
2017	15831.00	2.4962%	3.0329%	1.3328%	1.5519%	4.0342%	1.0702%	0.5095%
2018	16658.00	3.7894%	4.4319%	2.5273%	2.6310%	2.7413%	1.8282%	0.1990%
2019	17603.00	5.4687%	6.2109%	4.2550%	4.1654%	1.0333%	2.9671%	0.5567%
MAPE		3.9181%	4.5586%	2.7050%	2.7828%	2.6029%	1.9551%	0.4217%

Fig. 5 The curves, APE (%) and MAPE (%) of the seven models in the case of China's aging population

ing the training phase, which means that these two models are also tools with excellent fitting performance. Although the GMP(1,1,N) model and FGM(1,1) model have better fitting performance than CFODGMW(1,1, α) model, we can see that there is little difference in their fitting performance. It is worth noting that the pink bar representing the CFODGMW(1,1, α) model is the lowest in the test phase, meaning that its predictive performance is better than the other models. From the first figure shown in Fig. 5, we can see that the curve of CFODGMW(1,1, α) model is very close to the curve of the real value, which indicates that the development trend of CFODGMW(1,1, α) model conforms to the actual situation. The APE in Fig. 5 also implies that the performance of the CFODGMW(1,1, α) model is better than other models. In summary, compared with other models, the CFODGMW(1,1, α) model is more suitable for analyzing China's aging population. Beside, the running time of WOA in this case is shown in Table 6.

Forecasts and suggestions

Since the CFODGMW(1,1, α) model has won the competition with other models, this section will use it to forecast China's aging population from 2020 to 2023 and give some corresponding policy recommendations according to the predicted results. Inspired by the idea of metabolism [32], based on the parameters r , δ and α of the CFODGMW(1,1, α) model described in Section "Numerical results", the original data from 2008 to 2019 are used to establish the CFODGMW(1,1, α) model to predict China's aging population from 2020 to 2023. The number of the aging population in the past 4 years is shown in Fig. 6.

According to the data shown in Fig. 6, it can be seen that the aging population in China will further increase in the next few years, which is bound to have a huge impact on the development of Chinese society.

On the one hand, the sustained growth of the aging population and the declining birth rate indicates that China will

completely become an "aging country". The main reason for this phenomenon can be attributed to the fact that young people in China are under great pressure in life, and their salary cannot support them to raise a newborn. In addition, the early "family planning" in China also contributed to this phenomenon. Based on this, this article gives two related suggestions

1. To curb the surge in housing prices, the biggest living pressure of Chinese young people comes from the continuous increase in housing prices. Lower housing prices will inevitably lead to social unrest in China, but the Chinese government can control housing prices to stabilize, which can reduce the pressure on the younger generation.
2. Promoting the birth subsidy program. China should encourage young people to have children, and provide certain financial assistance to parents with multiple newborns.

On the other hand, with the increase in the aging population, their huge medical expenditure directly promotes the increase in the total dependency ratio of the society, thereby reducing the growth rate of GDP, which has a certain hindering effect on the economic development of Chinese society. Given this situation, the Chinese government should speed up the coverage of insurance business, expand the applicable scope of the insurance business, so that more aging population can enjoy the benefits of insurance business.

Conclusion

Considering the exponential growth law of population and the applicable scope of the grey forecasting model, this paper proposes an optimized discrete grey forecasting model based on the WLS and combined fractional-order accumulation operation. By analyzing the modeling mechanism of the proposed model, it is found that the CFODGMW(1,1, α) model

Fig. 6 The growth rate and number of China's aging population in the next 4 years

can be transformed into other types of GM models by changing its structural parameters, and it is unbiased for the time series satisfying homogeneous or non-homogeneous laws. These two aspects prove that CFODGMW(1,1, α) model is a kind of prediction model with excellent predictive performance. Subsequently, the CFODGMW(1,1, α) model and its comparative models are used to study China's aging population. Numerical results show that the prediction performance of the CFODGMW(1,1, α) model is better than other prediction models. At the end of the article, the CFODGMW(1,1, α) model is used to forecast the aging population in China in the next 4 years, and some corresponding recommendations are given according to the forecast results.

In addition, although the CFODGMW(1,1, α) model is a forecasting model with excellent predictive performance, it still has some defects. According to its modeling steps, we can see that the information of the first data in the time series is ignored when the CFODGMW(1,1, α) model is built, and the information is hard-won and it is a shame to waste it. How to make full use of the first data to build a more reasonable CFODGMW(1,1, α) model is still a problem.

Acknowledgements This work was supported by foundation of Hunan high-tech industry science and technology innovation leadership program Grant (No. 2020GK2029) and key R&D program of Hunan province (No. 2022SK2109) and Hunan Province Science and Technology Major Project (No. 2017SK1040) and open research fund of

Hunan Provincial Key Laboratory of network investigational Technology, Grant (No.2017WLZC009)

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

See Table 6.

Table 6 The running time of each part of the algorithm

Function name	Call frequency	The total time	Self-use time
WOA	1	14.070 s	0.536 s
Cost function	4000	13.481 s	13.182 s
mean	4000	0.299 s	0.299 s
Newplot	1	0.030 s	0.009 s
Initialization	1	0.022 s	0.022 s
Newplot>observe axes next plot	1	0.019 s	0.004 s
Cla	1	0.015 s	0.007 s
Graphics\private\cla notify	1	0.005 s	0.005 s
Graphics\private\clo	1	0.003 s	0.003 s
Gobjects	2	0.002 s	0.002 s
Ishold	1	0.001 s	0.001 s
clear notify	1	0.001 s	0.001 s
Newplot>observe figure next plot	1	0.001 s	0.001 s

References

- Ju-Long D (1982) problems of grey systems. Syst Control Lett 1(5):288–294. [https://doi.org/10.1016/s0167-6911\(82\)80025-x](https://doi.org/10.1016/s0167-6911(82)80025-x)
- Wang Z-X, Li D-D, Zheng H-H (2020) Model comparison of GM(1,1) and DGM(1,1) based on Monte-Carlo simulation. Phys A Stat Mech Appl 542:123341. <https://doi.org/10.1016/j.physa.2019.123341>
- Wang Z-X, Li Q (2019) Modelling the nonlinear relationship between CO2 emissions and economic growth using a PSO algorithm-based grey verhulst model. J Clean Prod 207:214–224. <https://doi.org/10.1016/j.jclepro.2018.10.010>
- Kiran M, Shanmugam PV, Mishra A, Mehendale A, Sherin HN (2021) A multivariate discrete grey model for estimating the waste from mobile phones, televisions, and personal computers in India. J Clean Prod 293. <https://doi.org/10.1016/j.jclepro.2021.126185>
- Yousuf MU, Al-Bahadly I, Avci E (2021) A modified GM(1,1) model to accurately predict wind speed. Sustain Energy Technol Assess 43. <https://doi.org/10.1016/j.seta.2020.100905>
- Şahin U (2021) Future of renewable energy consumption in France, Germany, Italy, Spain, turkey and UK by 2030 using optimized fractional nonlinear grey Bernoulli model. Sustain Prod Consump 25:1–14. <https://doi.org/10.1016/j.spc.2020.07.009>
- Javed SA, Zhu B, Liu S (2020) Forecast of biofuel production and consumption in top CO2 emitting countries using a novel grey model. J Clean Prod 276:123997. <https://doi.org/10.1016/j.jclepro.2020.123997>
- Cui J, Dang YG, Liu SF (2009) Novel grey forecasting model and its modeling mechanism. Control Decis 24: 1702–1706. <https://doi.org/10.1360/972009-754>. http://en.cnki.com.cn/Article_en/CJFDTOTAL-KZYC200911020.htm
- Qian WY, Dang YG, Liu SF (2012) Grey gm(1,1, r^α) model with time power and its application. Syst Eng Theory Pract 32(5): 2247–2252. http://en.cnki.com.cn/Article_en/CJFDTOTAL-XTLL201210018.htm
- Luo D, Wei BL (2017) Grey forecasting model with polynomial term and its optimization. J Grey Syst 29(7): 58–69. https://www.researchgate.net/publication/318967400_Grey_forecasting_model_with_polynomial_term_and_its_optimization
- Liu C, Xie W, Lao T, ting Yao Y, Zhang J (2020) Application of a novel grey forecasting model with time power term to predict china's GDP. Grey Syst Theory Appl 11(3):343–357. <https://doi.org/10.1108/gs-05-2020-0065>
- Wu LF, Liu SF, Yao L (2015) Grey model with caputo fractional order derivative. Syst Eng Theory Pract 35(3):1311–1316. http://en.cnki.com.cn/Article_en/CJFDTOTAL-XTLL201505023.htm
- Wu LF (2015) Fractional order grey forecasting models and their application
- Şahin U (2020) Projections of Turkey's electricity generation and installed capacity from total renewable and hydro energy using fractional nonlinear grey Bernoulli model and its reduced forms. Sustain Prod Consump 23:52–62. <https://doi.org/10.1016/j.spc.2020.04.004>
- Hu Y-C, Jiang P, Tsai J-F, Yu C-Y (2021) An optimized fractional grey prediction model for carbon dioxide emissions forecasting. Int J Environ Res Public Health 18(2):587. <https://doi.org/10.3390/ijerph18020587>
- Ma X, Wu W, Zeng B, Wang Y, Wu X (2020) The conformable fractional grey system model. ISA Trans 96:255–271. <https://doi.org/10.1016/j.isatra.2019.07.009>
- Chen Y, Lifeng W, Lianyi L, Kai Z (2020) Fractional Hausdorff grey model and its properties. Chaos Sol Fract 138:109915. <https://doi.org/10.1016/j.chaos.2020.109915>
- Liu C, Xie W, Wu W-Z, Zhu H (2021) Predicting Chinese total retail sales of consumer goods by employing an extended discrete grey polynomial model. Eng Appl Artif Intell 102:104261. <https://doi.org/10.1016/j.engappai.2021.104261>
- Liu JX (2017) The demand forecast of natural gas based on grey and partial least squares combination model
- Wei B, Xie N (2021) Parameter estimation for grey system models: a nonlinear least squares perspective. Commun Nonlinear Sci Numer Simul 95:105653. <https://doi.org/10.1016/j.cnsns.2020.105653>
- Pei L, Li Q, Wang Z (2018) The NLS-based nonlinear grey Bernoulli model with an application to employee demand prediction of high-tech enterprises in China. Grey Syst Theory Appl 8(2):133–143. <https://doi.org/10.1108/gs-11-2017-0038>
- Luo D, Wei BL (2019) A unified treatment approach for a class of discrete grey forecasting models and its application. Syst Eng Theory Pract 8:451–462. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2019&filename=XTLL201902016&uniplatform=NZKPT&v=iCcoRfaeRoxTERxgkxSEvF6wZuRJraPn-9nkbMwaW2dhnD6MuGn-pmOydReE_55e
- Tang L, Lu Y (2020) Study of the grey Verhulst model based on the weighted least square method. Phys A Stat Mech Appl 545:123615. <https://doi.org/10.1016/j.physa.2019.123615>

24. ming Xie N, feng Liu S (2009) Discrete grey forecasting model and its optimization. *Appl Math Model* 33(2):1173–1186. <https://doi.org/10.1016/j.apm.2008.01.011>
25. Qu WL (1989) Combinatorial mathematics
26. Xie N-M, Liu S-F, Yang Y-J, Yuan C-Q (2013) On novel grey forecasting model based on non-homogeneous index sequence. *Appl Math Model* 37(7):5059–5068. <https://doi.org/10.1016/j.apm.2012.10.037>
27. Qiao W, Khishe M, Ravakhah S (2021) Underwater targets classification using local wavelet acoustic pattern and multi-layer perceptron neural network optimized by modified whale optimization algorithm. *Ocean Eng* 219:10841. <https://doi.org/10.1016/j.oceaneng.2020.108415>
28. Saville D-J, Wood G-R (1991) Polynomial Regression
29. Xie Y (2020) Research on performance evaluation of comprehensive two-child policy based on fractional-order GM(1,1) model. <https://doi.org/10.27713/d.cnki.gcqgs.2020.000185>
30. Tian Z, Ji G, Liu M (2021) Analysis and prediction of total population in Xinjiang based on improved grey gm(1,1) model. *J Math Pract Theory* 51:258–264
31. Alsmirat MA, Jararweh Y, Obaidat I, Gupta BB (2016) Automated wireless video surveillance: an evaluation framework. *J Real Time Image Process* 13(3):527–546. <https://doi.org/10.1007/s11554-016-0631-x>
32. Liu C, Wu W-Z, Xie W, Zhang J (2020) Application of a novel fractional grey prediction model with time power term to predict the electricity consumption of India and China. *Chaos Sol Fract* 141:110429. <https://doi.org/10.1016/j.chaos.2020.110429>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.