Open Access Library Journal
2022, Volume 9, e8416
ISSN Online: 2333-9721
ISSN Print: 2333-9705

Improving GM(1,1) Model Performance Accuracy Based on the Combination of Optimized Initial and Background Values in Time Series Forecasting

Mahdi Madhi[1]*, Norizan Mohamed[2]

[1]Department of Statistics, Faculty of Science, Sebha University, Libya
[2]Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, Malaysia
Email: *mah.arhuma@sebhau.edu.ly, norizan@umt.edu.my

How to cite this paper: Madhi, M. and Mohamed, N. (2022) Improving GM(1,1) Model Performance Accuracy Based on the Combination of Optimized Initial and Background Values in Time Series Forecasting. *Open Access Library Journal*, **9**: e8416.
https://doi.org/10.4236/oalib.1108416

Received: February 2, 2022
Accepted: April 21, 2022
Published: April 24, 2022

Abstract

The term grey forecasting model has been comprehensively utilized in numerous research arenas and discovered valid outcomes. Nevertheless, the model possesses certain possible problems that necessitate improvement. It has been proven that, part of the foremost issues distressing the prediction accurateness of the model are initial and background values. Henceforth, a new modified GM(1,1) model through the combination of optimized initial value and background value has been recommended in this study. The new initial value encompasses the median value of a sequence that has been generated using the first-order accumulative generating operation on the raw data sequence while the number of observations is odd or even. Meanwhile, in the standard model GM(1,1), the background value is rebuilt using an integral term to rectify the error term caused by the background value computation. An empirical study of real data was performed to validate the proposed model's prediction accuracy in this article. The real datasets were analyzed using R Code. The obtained findings showed that the modified model GM(1,1) has lower error as well as higher accuracy, which enriches grey model optimization theory and broadens the field of grey model implementation in time series forecasting.

Subject Areas

Mathematical Statistics, Applied Statistical Mathematics

Keywords

Grey Systems Theory, Standard Model GM(1,1), Background Value,

Initial Value, Prediction Accuracy

1. Introduction

The Grey system theory development is assigned or in other words accredited to Deng (1982) at the beginning of the 1980s as a method of quantitative predictions, which is recognized as a precisely multifaceted and universal theory. Theory has come to be relatively prevailing owing to its capability for handling systems with limited data samples and/or with deficient information. Within the previous three decades epoch, grey system theory is already grown rapidly and piqued the interest of many academics [1] [2] [3] [4].

Grey forecasting models served as an essential component of grey system theory. They possess favorable circumstances for dealing with tentative information by utilizing as few as four data points. Deng (1989) advocated the standard model GM(1,1), which is commonly widely utilized in the literature, which is alluded to as "Grey Model First Order One Variable". It has been broadly, productively and effectively modified and put into use by numerous systems [2] [5]. Nonetheless, the standard model GM(1,1) affirmed some limits that affect the model's application, as well as its forecasting accuracy. Several studies, notably in the recent three decades, have focused on the standard model GM(1,1) optimization research. In general, the research aimed to improve the model in two stages.

First, numerous studies concentrated on enhancements of the time response function's initial value. For example, Yaoguo *et al.* (2004) laid forward a procedure for upgrading standard model GM(1,1) utilizing the n^{th} item of $X^{(1)}$ as the grey differential model's initial value [6]. Xie and Liu (2009) recommended other procedures for individual grey forecasting models and consistent optimization of the parameter. The writers further clarified three kinds of grey forecasting models. These models involved the initial-point, the middle-point, and the ending-point, which fixed distinct grey model [7]. Juan (2011) proposed an improved grey model which removed old information, added new information, revised the exponent k of the time response sequence in the essential formula of the GM(1,1) model, and created a dynamic balance state. The exponent substituted k in the time response sequence in the essential formula of the GM(1,1) model, which can make quicker metabolism and lesser error [8]. Likewise, Wang *et al.* (2012) offered a different time response function's initial value, which is the average of $X^{(1)}$'s first and last items. They hired new bits of information in actual data, and used time response function's initial value of the standard model GM(1,1) [9]. Chen and Li (2015) also provided a novel procedure of the standard model GM(1,1) based on an optimum weighted mixture with a new initial condition [10]. Madhi and Mohamed (2017) proposed an innovative method for enhancing the capability of the standard model GM(1,1) by improving the initial value. To adjust the model's initial value, the authors utilized the lowest sum of squared errors [11]. Zeng *et al.* (2020) proposed a novel grey forecasting model that combines new initial values with actual data reprocessing. The accumulative order of the novel model is then configured using fraction accumulation creation [12].

Second, numerous studies reconstructed the background value of GM(1,1) model. For example, background value, according to Dai and Chen (2007), is a significant component of accuracy and flexibility in the standard model GM(1,1). As a result, the researchers listed above proposed a groundbreaking method for re-creating the standard model's traditional background value using the Gauss-Legendre technique [13]. Another research, piloted by Li *et al.* (2009), examined sample behavior using the trend and potency monitoring technique (TPTM). The authors used the trend and potency value to recreate the standard model's traditional background value after extracting the hidden information [14]. According to Li (2011), one of the important aspects influencing the model's precision is the background value. In order to improve the standard model GM(1,1)'s precision, the background value was restructured using the monotonic increase of the cumulative generating series. Instead of using the simple formula of the background value, the triangle field total was subtracted from the contracted trapezium area [15]. Yao and Wang's (2014) study correspondingly utilized an enhanced model considering the background value for prediction of electricity demand in east China [16]. Tsai and Lu (2015) put forward a novel method for improving the forecasting efficacy of a grey model. They modified background values for a new grey model optimization using the exponentially weighted moving average (EWMA) algorithm [17]. Cheng and Shi (2019) attempted to improve model parameter estimation using new background value optimization without modifying the model configuration, mostly using four approaches [18]. Li and Zhang (2019) suggested a novel optimized gray GM(1,1) model by transforming the raw dataset and then improving the background value measurement process [19]. Liu *et al.* (2019) used the golden segmentation optimization approach to refine the background value and Fourier-series principle to derive periodic information in the grey prediction model for residual error correction [20]. To address the shortcomings of standard grey prediction model, Wu *et al.* (2020) introduced a new background value modification nonlinear grey predictive model [21]. Yue and Diyi (2021) proposed a self-adaptive GM(1,1, λ) model determined by the structural method of background value that contained a self-adaptive component $\lambda \in (0,1)$ is incorporated into the traditional GM(1,1) model's formulation based on its feasibility, which is proven by the mean value theorem for integrals [22].

To some extent, applying the strategies based on the literature stated above can reduce errors and improve forecast accuracy. But, it only employ one-side refinement of the initial value or background value, which are then utilized to simulate the values and predictions of the standard model GM(1,1).

The main task of this study is to offer a new technique of combining initial value optimization with background value optimization in the time response function. This technique is recommended to improve the existing model in time series prediction and interrelate it practically with the standard model GM(1,1) in terms of prediction performance assessment criteria. The new initial value is the median value computed using the first-order accumulative generating operation on the raw data series when the number of observations is even or odd. The new improved initial value can describe the new information priority principles highlighted in grey system theory. Whereas, an integration term is utilized to reconstruct the background value in order to remedy the error term occurred as a result of the standard model GM(1,1) background value calculation.

2. Methodology

2.1. The Standard Model GM(1,1)

Deng (1982) stipulated that the raw data be collected from a consecutive period in time and that the data be at least four in number. Furthermore, the model is central to grey system theory while standard model GM(1,1) is among the most extensively employed grey prediction models, with the sign GM(1,1) standing for "first order grey model in one variable." The following describes the construction mechanism of the standard model GM(1,1) [23]:

Step 1: Denote the raw data sequence.

$$X^{(0)} = \left\{ x^{(0)}{}_{(1)}, x^{(0)}{}_{(2)}, \cdots, x^{(0)}{}_{(n)} \right\}, n \geq 4,$$

Step 2: Uses the first-order Accumulative Generation Operator (1-AGO) on $X^{(0)}$ to create a new data set.

$X^{(1)} = \left\{ x^{(1)}{}_{(1)}, x^{(1)}{}_{(2)}, \cdots, x^{(1)}{}_{(n)} \right\}, n \geq 4$, where $x^{(1)}{}_{(1)} = x^{(0)}{}_{(1)}$ and

$$x^{(1)}{}_{(k)} = \sum_{i=1}^{k} x^{(0)}{}_{(i)}, k = 2, 3, \cdots, n \tag{1}$$

Step 3: Determine the background values $z^{(1)}$

$$z^{(1)}{}_{(k)} = 0.5x^{(1)}{}_{(k-1)} + 0.5x^{(1)}{}_{(k)}, \ k = 2, 3, \cdots, n \tag{2}$$

Step 4: The first-order differential equation and whitening equation of the standard model GM(1,1) are determined to be as follows:

$$x^{(0)}{}_{(k)} + az^{(1)}{}_{(k)} = b \tag{3}$$

$$\frac{\mathrm{d}x^{(1)}{}_{(k)}}{\mathrm{d}t} + ax^{(1)}{}_{(k)} = b \tag{4}$$

where a and b denote the development coefficient and grey action quantity, respectively.

Step 5 Estimate a and b values using the least square approach as

$$[a, b]^{\mathrm{T}} = \left(B^{\mathrm{T}} B\right)^{-1} B^{\mathrm{T}} Y, \tag{5}$$

where

$$Y = \left[x^{(0)}{}_{(1)}, x^{(0)}{}_{(2)}, \cdots, x^{(0)}{}_{(n)}\right]^{\mathrm{T}}$$

$$B = \begin{bmatrix} -z^{(1)}{}_{(2)} & 1 \\ -z^{(1)}{}_{(3)} & 1 \\ \vdots & \vdots \\ -z^{(1)}{}_{(n)} & 1 \end{bmatrix}$$

Step 6: whitenization equation's solution, which is also measured as time response function (Equation (4)) is depicted as

$$\hat{x}^{(1)}{}_{(t)} = \left(x^{(0)}{}_{(1)} - \frac{b}{a}\right)\mathrm{e}^{-a(t-1)} + \frac{b}{a} \tag{6}$$

By substituting $t = k$ into Equation (6), it follows that

$$\hat{x}^{(1)}{}_{(k)} = \left(x^{(0)}{}_{(1)} - \frac{b}{a}\right)\mathrm{e}^{-a(k-1)} + \frac{b}{a} \tag{7}$$

Step 7: The inverse accumulating generating operation (IAGO) can be used to obtain the recovered data

$$\hat{x}^{(0)}{}_{(k)} = \hat{x}^{(1)}{}_{(k+1)} - \hat{x}^{(1)}{}_{(k)} = \left(1 - \mathrm{e}^{a}\right)\left(x^{(0)}{}_{(1)} - \frac{b}{a}\right)\mathrm{e}^{-a(k-1)} \tag{8}$$

From Equation (7), a significant formula has been construed in which the precision of the standard model GM(1,1) is dependent on the coefficients a and b, nonetheless all of them be influenced by the raw data set and function of the background value. The trapezium formula is used in the basic background value calculation formula to accurately compute the area incorporated by $X^{(1)}{}_{(t)}$ and t axis on interval $[k-1, k]$. The dashed area in Figure 1 depicts the inaccuracy caused by the traditional background value. At the point when the raw data time interval is short and the series is flat, it is appropriate to utilize the traditional background value. In any case, if the trend of the raw data sequence changes quickly, the model's error becomes greater [24] [25]. In addition to that, the first value in a series produced using the 1-AGO on $X^{(0)}$ is the initial value in Equation (7). Whereas, this form of the time response function's initial value cannot exploit new pieces of information from the original data. In addition, this type of

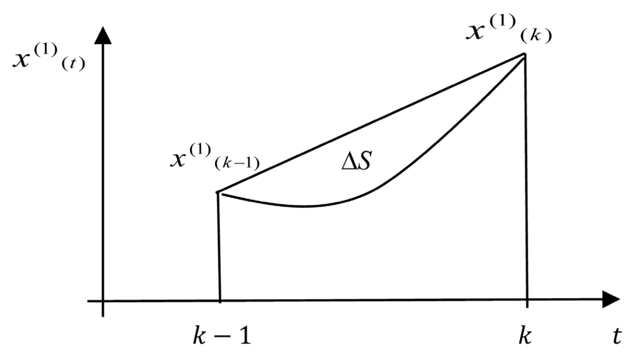

Figure 1. $z^{(1)}{}_{(k)}$ using the traditional background value.

initial value cannot be used to specify the priority principle for the most recent or new information [26]. Hence, the initial value and background value structure formula are two important parameters that cause simulation error $\varepsilon^{(0)}{}_{(k)} = x^{(0)}{}_{(k)} - \hat{x}^{(0)}{}_{(k)}$ and impact the stability and reliability of the standard model GM(1,1). As a result, as detailed in the following sections, this study suggests that the initial and background values be optimized to increase the performance of the existing model in time series forecasting.

2.2. Optimization of Initial Value

Optimization of initial value was performed using medium value when the number of observations was odd and even as follows:

1) When the number of observations is odd, the median equals the middle value.

Let $x^{(1)}{}_{(i)}$ is the middle value of a sequence produced using the first-order accumulative generating operation on the raw data sequence and which corresponds to the arranged *i*.

Since,

$$x^{(1)}{}_{(t)} = c\mathrm{e}^{-at} + \frac{b}{a} \tag{9}$$

is the whitening equation's general solution (Equation (4)).

(c is a constant), we get the following Equation (10) by putting $t = i$ in Equation (9),

$$x^{(1)}{}_{(i)} = c\mathrm{e}^{-ai} + \frac{b}{a} \tag{10}$$

Solving Equation (10) gives

$$c = \left(x^{(1)}{}_{(i)} - \frac{b}{a} \right) \mathrm{e}^{ai} \tag{11}$$

The optimal time response function of the whitenization equation (Equation (4)) is then obtained as follows

$$\hat{x}^{(1)}{}_{(t)} = \left(x^{(1)}{}_{(i)} - \frac{b}{a} \right) \mathrm{e}^{-a(t-i)} + \frac{b}{a} \tag{12}$$

Using Equation (12) and $t = k$, we can deduce that

$$\hat{x}^{(1)}{}_{(k)} = \left(x^{(1)}{}_{(i)} - \frac{b}{a} \right) \mathrm{e}^{-a(k-i)} + \frac{b}{a} \tag{13}$$

The restored values can be defined with the use of the IAGO on $\hat{x}^{(1)}{}_{(k)}$ in Equation (13),

$$\hat{x}^{(0)}{}_{(k)} = \hat{x}^{(1)}{}_{(k)} - \hat{x}^{(1)}{}_{(k-1)} = \left(1 - e^{a}\right) \left(x^{(1)}{}_{(i)} - \frac{b}{a} \right) \mathrm{e}^{-a(k-i)} \tag{14}$$

2) The median is considered to be the average of the two values in the middle of the distribution if the number of observations is even.

Let $\dfrac{x^{(1)}{}_{(i)}+x^{(1)}{}_{(j)}}{2}$ is the average of the two values in the middle of a sequence produced using the first-order accumulative generating operation on the raw data sequence. And where i and j are arranged of $x^{(1)}{}_{(i)}$ and $x^{(1)}{}_{(j)}$ respectively ($i<j$).

Since,

$$x^{(1)}{}_{(t)}=c\mathrm{e}^{-at}+\frac{b}{a} \tag{15}$$

is the whitening equation's general solution Equation (4)).

(c is a constant), by putting

$$x^{(1)}{}_{(t)}\Big|_{t=i}=x^{(1)}{}_{(i)}$$

$$x^{(1)}{}_{(t)}\Big|_{t=j}=x^{(1)}{}_{(j)}$$

Into

$$x^{(1)}{}_{(i)}=c\mathrm{e}^{-ai}+\frac{b}{a} \tag{16}$$

$$x^{(1)}{}_{(j)}=c\mathrm{e}^{-aj}+\frac{b}{a} \tag{17}$$

In order to utilize the raw data with the latest information and keep the initial value in the standard model GM(1,1)'s function of time response, a new initial value has been set within function of time response equivalent to $0.5\left(x^{(1)}{}_{(i)}+x^{(1)}{}_{(j)}\right)$. Whereas, as shown in the following formula, the constant c may be computed using the aforementioned equations $x^{(1)}{}_{(i)}$ and $x^{(1)}{}_{(j)}$.

$$c=2\left(\mathrm{e}^{-ai}+\mathrm{e}^{-aj}\right)^{-1}\left(\frac{x^{(1)}{}_{(i)}+x^{(1)}{}_{(j)}}{2}-\frac{b}{a}\right) \tag{18}$$

The optimal time response function of the whitenization equation (Equation (4)) is then obtained as follows

$$\hat{x}^{(1)}{}_{(t)}=2\left(\mathrm{e}^{-ai}+\mathrm{e}^{-aj}\right)^{-1}\left(\frac{x^{(1)}{}_{(i)}+x^{(1)}{}_{(j)}}{2}-\frac{b}{a}\right)\mathrm{e}^{-at}+\frac{b}{a} \tag{19}$$

Using Equation (19) and $t=k$, we can deduce that

$$\hat{x}^{(1)}{}_{(k)}=2\left(\mathrm{e}^{-ai}+\mathrm{e}^{-aj}\right)^{-1}\left(\frac{x^{(1)}{}_{(i)}+x^{(1)}{}_{(j)}}{2}-\frac{b}{a}\right)\mathrm{e}^{-ak}+\frac{b}{a} \tag{20}$$

The restored values can be defined with the use of the IAGO on $\hat{x}^{(1)}{}_{(k)}$ in Equation (20),

$$\hat{x}^{(0)}{}_{(k)}=\hat{x}^{(1)}{}_{(k)}-\hat{x}^{(1)}{}_{(k-1)}=2\left(1-\mathrm{e}^{a}\right)\left(\mathrm{e}^{-ai}+\mathrm{e}^{-aj}\right)^{-1}\left(\frac{x^{(1)}{}_{(i)}+x^{(1)}{}_{(j)}}{2}-\frac{b}{a}\right)\mathrm{e}^{-ak} \tag{21}$$

2.3. Optimization of Background Value

Traditional background value denotes to the utilization of the trapezium formula to discover the area between curve and interval. If the raw data sequence slightly changes, the background value is suitable. However, if the raw data sequence significantly changes, the background value construction will prompt major error. Hence, for the error to be minimized, a new technique of background value reconstruction is put forward as follows.

When the integrals of both sides of Equation (4) are calculated from *k*–1 to *k*, the result is

$$\int_{k-1}^{k} \frac{\mathrm{d}x^{(1)}{}_{(t)}}{\mathrm{d}t}\mathrm{d}t + a\int_{k-1}^{k} x^{(1)}{}_{(t)}\mathrm{d}t = \int_{k-1}^{k} b\mathrm{d}t$$

$$\Rightarrow x^{(1)}{}_{(k)} - x^{(1)}{}_{(k-1)} + a\int_{k-1}^{k} x^{(1)}{}_{(t)}\mathrm{d}t = b$$

$$\Rightarrow x^{(0)}{}_{(k)} + a\int_{k-1}^{k} x^{(1)}{}_{(t)}\mathrm{d}t = b \tag{22}$$

When Equation (3) is compared to Equation (22), the parameters *a* and *b* computed using $\int_{k-1}^{k} x^{(1)}{}_{(t)}\mathrm{d}t$ as the background value are more adaptable to the whitenization equation. As can be shown, the error cause for the standard model GM(1,1) is the usage of $z^{(1)}{}_{(k)} = 0.5x^{(1)}{}_{(k)} + 0.5x^{(1)}{}_{(k-1)}$ rather than $\int_{k-1}^{k} x^{(1)}{}_{(t)}\mathrm{d}t$.

Then the background value is as follows:

$$z^{(1)}{}_{(k)} = \int_{k-1}^{k} x^{(1)}{}_{(t)}\mathrm{d}t \tag{23}$$

To avoid background value-related errors, we may suppose that,

$$x^{(1)}{}_{(t)} = C\mathrm{e}^{-At} + B \tag{24}$$

Substituting Equation (24) into Equation (23) gives

$$z^{(1)}{}_{(k)} = \int_{k-1}^{k} \left(C\mathrm{e}^{-At} + B\right)\mathrm{d}t = \frac{1}{-A}\left(x^{(1)}{}_{(k)} - x^{(1)}{}_{(k-1)}\right) + B = \frac{x^{(0)}{}_{(k)}}{-A} + B \tag{25}$$

Substitutions of *t* for *k*, *k*–1, *and k*–2 into (24) give

$$x^{(1)}{}_{(k)} = C\mathrm{e}^{-Ak} + B \tag{26}$$

$$x^{(1)}{}_{(k-1)} = C\mathrm{e}^{-A(k-1)} + B \tag{27}$$

$$x^{(1)}{}_{(k-2)} = C\mathrm{e}^{-A(k-2)} + B \tag{28}$$

Equation (26) - Equation (27)

$$x^{(1)}{}_{(k)} - x^{(1)}{}_{(k-1)} = C\mathrm{e}^{-Ak}\left(1 - \mathrm{e}^{A}\right) \tag{29}$$

Equation (27) - Equation (28)

$$x^{(1)}{}_{(k-1)} - x^{(1)}{}_{(k-2)} = C\mathrm{e}^{-Ak+A}\left(1 - \mathrm{e}^{A}\right) \tag{30}$$

The ratio of $x^{(0)}{}_{(k)}$ and $x^{(0)}{}_{(k-1)}$ based on Equation (29) and Equation (30) is

$$\frac{x^{(0)}{}_{(k-1)}}{x^{(0)}{}_{(k)}} = \frac{Ce^{-Ak+A}\left(1-e^{A}\right)}{Ce^{-Ak}\left(1-e^{A}\right)} = e^{A} \tag{31}$$

Then A can be written as

$$A = \ln\frac{x^{(0)}{}_{(k-1)}}{x^{(0)}{}_{(k)}} \tag{32}$$

Substituting Equation (32) into Equation (29) gives

$$C = \frac{\left(x^{(0)}{}_{(k)}\right)^{2}\left(x^{(0)}{}_{(k-1)}\right)^{k}}{\left(x^{(0)}{}_{(k)}\right)^{k+1} - \left(x^{(0)}{}_{(k)}\right)^{k} x^{(0)}{}_{(k-1)}} \tag{33}$$

Substituting $k=1$ into Equation (26)

$$x^{(1)}{}_{(1)} = Ce^{-A} + B \tag{34}$$

Equations (32) and (33) are substituted into Equation (34) to get

$$B = x^{(0)}{}_{(1)} - \frac{\left(x^{(0)}{}_{(k)}\right)^{2}\left(x^{(0)}{}_{(k-1)}\right)^{k-1}}{\left(x^{(0)}{}_{(k)}\right)^{k} - \left(x^{(0)}{}_{(k)}\right)^{k-1} x^{(0)}{}_{(k-1)}} \tag{35}$$

The new formula for calculating background value may then be created by substituting Equations (32) and (35) into Equation (25) as follows

$$\begin{aligned} z^{(1)}{}_{(k)} &= \frac{x^{(0)}{}_{(k)}}{\ln x^{(0)}{}_{(k)} - \ln x^{(0)}{}_{(k-1)}} + x^{(0)}{}_{(1)} \\ &\quad - \frac{\left(x^{(0)}{}_{(k)}\right)^{2}\left(x^{(0)}{}_{(k-1)}\right)^{k-1}}{\left(x^{(0)}{}_{(k)}\right)^{k} - \left(x^{(0)}{}_{(k)}\right)^{k-1} x^{(0)}{}_{(k-1)}}, k = 2,3,\cdots,n \end{aligned} \tag{36}$$

Through the use of the above precise description, the modified GM(1,1) model's prediction steps are presented below.

Step 1: $z^{(1)}{}_{(k)}$ is calculated using Equation (36) given the raw data sequence $X^{(0)}$.

Step 2: Equation (5) is used to calculate a and b.

Step 3: When the number of observations is odd, $x^{(1)}{}_{(i)}$, a and b are replaced into Equation (13) to yield $\hat{x}^{(1)}{}_{(k)}$, and the raw data sequence prediction values may be retrieved as $\hat{x}^{(0)}{}_{(k)}$ $\left(k = 1,2,\cdots,n+p\right)$ using Equation (14). Whereas, when the number of observations is even, $0.5\left(x^{(1)}{}_{(i)} + x^{(1)}{}_{(j)}\right)$, a and b are substituted into Equation (20) to get $\hat{x}^{(1)}{}_{(k)}$ and the raw data sequence prediction values may be retrieved as $\hat{x}^{(0)}{}_{(k)}$ $\left(k = 1,2,\cdots,n+p\right)$ using Equation (21), where p represents the size of prediction steps.

2.4. Practical Performance Evaluation

Two time series datasets have been prepared with regard to test the accuracy of the modified GM(1,1) model (MGM(1,1)) versus the standard model GM(1,1) (GM(1,1)), the method of Guan-Jun (2000) [27] (GGM(1,1)), and the method of Yao and Wang (2014) [14] (YWGM(1,1)). The models' predictive accuracy is examined based on the mean absolute percentage error (MAPE) labelled by

$$\text{MAPE} = \frac{1}{n-1}\sum_{k=2}^{n} \text{APE}(k) \times 100\% = \frac{1}{n-1}\sum_{k=2}^{n} \frac{\left| x^{(0)}_{(k)} - \hat{x}^{(0)}_{(k)} \right|}{x^{(0)}_{(k)}} \times 100\%$$

The forecasting accuracy level can be classified into four grades based on the MAPE of each model as indicated in **Table 1** Lewis (1982) [28]:

Table 1. Categorizing the grade of predicting accuracy.

Grade Level	Highly accurate prediction	Good prediction	Reasonable prediction	Inaccurate prediction
MAPE	<10%	10% - 20%	20% - 50%	>50%

3. Empirical Analysis

In such section, two examples have been prepared with respect to determining the accuracy of the modified GM(1,1) model (MGM(1,1)), the standard model GM(1,1) (GM(1,1)), the method of Guan-Jun (2000) (GGM(1,1)), and the method of Yao and Wang (2014) (YWGM(1,1)) when the number of observations is odd and even, respectively.

3.1. Example 1

This section considers the Chinese educational funds between 2007 and 2017 used in reference [29]. The eleven observations in the data sequence were all utilized to determine the accuracy prediction of the standard model GM(1,1), the method of Guan-Jun (2000), the method of Yao and Wang (2014), and the modified GM(1,1) model recommended in the current work. The four models were structured using the first nine observations from the raw data (in-sample data). Meanwhile, Predictive evaluation was performed using the last two observations in the raw data (out-of-sample data). This information is depicted in **Table 2** and **Figure 2**.

By the standard model GM(1,1), prediction equation of the raw data sequence is

$$\hat{x}^{(0)}_{(k)} = \left(1 - e^{-0.1239}\right)\left(x^{(0)}_{(1)} + \frac{133445910}{0.1239}\right)e^{0.1239(k-1)},\ k = 2,3,\cdots,11$$

By the method of Guan-Jun (2000), the prediction equation of the raw data sequence is

$$\hat{x}^{(0)}_{(k)} = \left(1 - e^{-0.1258}\right)\left(x^{(0)}_{(1)} + \frac{135616950}{0.1258}\right)e^{0.1258(k-1)},\ k = 2,3,\cdots,11$$

By the method of Yao and Wang (2014), the prediction equation of the raw data sequence is

$$\hat{x}^{(0)}_{(k)} = \left(1 - e^{-0.1238}\right)\left(x^{(0)}_{(1)} + \frac{134417775}{0.1238}\right)e^{0.1238(k-1)}, k = 2,3,\cdots,11$$

By the modified GM(1,1) model, the prediction equation of the raw data sequence is

$$\hat{x}^{(0)}_{(k)} = \left(1 - e^{-0.1103}\right)\left(x^{(0)}_{(5)} + \frac{140141263}{0.1103}\right)e^{0.1103(k-5)}, k = 2,3,\cdots,11$$

Table 2. Performance of the above-mentioned models in terms of prediction accuracy in China's educational funds (10,000 CNY).

Raw Data		The standard model GM(1,1)		Method of Guan-Jun (2000)		Method of Yao and Wang (2014)		Modified GM(1,1) Model	
Year	Actual Values	Model Values	APE (%)	Model Values	APE (%)	Model Values	APE (%)	Model Values	APE (%)
2007	121,480,663								
2008	145,007,374	158,094,556	9.03	160,801,850	10.89	159,093,730	9.71	160,254,933	10.52
2009	165,027,065	178,952,574	8.44	182,358,386	10.50	180,053,227	9.11	178,938,473	8.43
2010	195,618,471	202,562,470	3.55	206,804,715	5.72	203,773,992	4.17	199,800,259	2.14
2011	238,692,936	229,287,310	3.94	234,528,234	1.74	230,619,804	3.38	223,094,244	6.54
2012	286,553,052	259,538,060	9.43	265,968,272	7.18	261,002,366	8.92	249,103,991	13.07
2013	303,647,182	293,779,907	3.25	301,623,052	0.67	295,387,620	2.72	278,146,119	8.40
2014	328,064,609	332,539,411	1.36	342,057,586	4.27	334,302,893	1.90	310,574,163	5.33
2015	361,291,927	376,412,605	4.19	387,912,634	7.37	378,344,983	4.72	346,782,876	4.02
MAPE In-sample			5.40		6.04		5.58		7.30
2016	388,883,850	426,074,156	9.56	439,914,850	13.12	428,189,314	10.11	387,213,030	0.43
2017	425,620,069	482,287,746	13.31	498,888,301	17.21	484,600,290	13.86	432,356,788	1.58
MAPE Out-sample			11.44		15.17		11.98		1.01

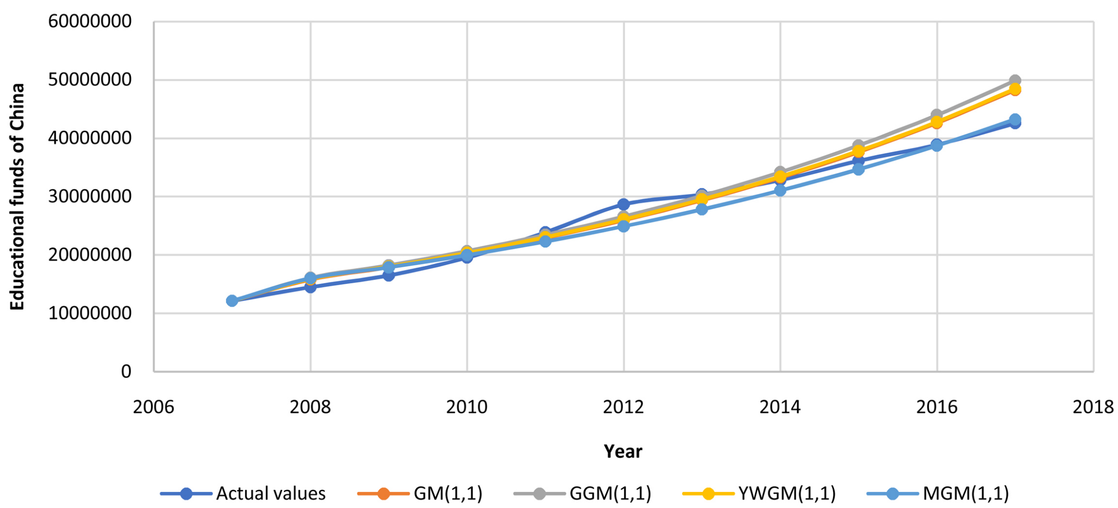

Figure 2. Performance of the above-mentioned models in terms of prediction accuracy in China's educational funds (10,000 CNY).

3.2. Example 2

This subsection examines the number of new students enrolled in China's regular institutions of higher education from 2005 to 2018, as cited in [29]. The standard model GM(1,1), the method of Guan-Jun (2000), the method of Yao and Wang (2014), and the modified GM(1,1) model provided in the current work were all evaluated using the fourteen observations in the data sequence to establish their prediction accuracy.The four models were structured using the first ten observations from the raw data (in-sample data). Meanwhile, Predictive evaluation was performed using the last four observations in the raw data (out-of-sample data). This information is depicted in **Table 3** and **Figure 3**.

By standard model GM(1,1), the prediction equation of the raw data sequence is

$$\hat{x}^{(0)}{}_{(k)} = \left(1-\mathrm{e}^{-0.0332}\right)\left(x^{(0)}{}_{(1)} + \frac{537.4467}{0.0332}\right)\mathrm{e}^{0.0332(k-1)}, k = 2,3,\cdots,14$$

By the method of Guan-Jun (2000), the prediction equation of the raw data sequence is

$$\hat{x}^{(0)}{}_{(k)} = \left(1-\mathrm{e}^{-0.0333}\right)\left(x^{(0)}{}_{(1)} + \frac{539.7144}{0.0333}\right)\mathrm{e}^{0.0333(k-1)}, k = 2,3,\cdots,14$$

By the method of Yao and Wang (2014), the prediction equation of the raw data sequence is

$$\hat{x}^{(0)}{}_{(k)} = \left(1-\mathrm{e}^{-0.0331}\right)\left(x^{(0)}{}_{(1)} + \frac{538.3021}{0.0331}\right)\mathrm{e}^{0.0331(k-1)}, k = 2,3,\cdots,14$$

By the modified GM(1,1) model, the prediction equation of the raw data sequence is

$$\hat{x}^{(0)}{}_{(k)} = 2\left(1-\mathrm{e}^{-0.0322}\right)\left(\mathrm{e}^{0.1610}+\mathrm{e}^{0.1932}\right)^{-1} \times\left(\frac{x^{(1)}{}_{(5)}+x^{(1)}{}_{(6)}}{2} + \frac{538.5056}{0.0322}\right)\mathrm{e}^{0.0322k}, k = 2,3,\cdots,14$$

Table 3. Performance of the above-mentioned models in terms of prediction accuracy in the new students enrolled in China's regular institutions of higher education (10,000 Persons).

Raw Data		The standard model GM(1,1)		Method of Guan-Jun (2000)		Method of Yao and Wang (2014)		Modified GM(1,1) Model	
Year	Actual Values	Model Values	APE (%)	Model Values	APE (%)	Model Values	APE (%)	Model Values	APE (%)
2005	504.5								
2006	546.1	563.49	3.18	565.90	3.63	564.27	3.33	563.82	3.25
2007	565.9	582.51	2.93	585.07	3.39	583.24	3.06	582.28	2.89
2008	607.7	602.17	0.91	604.90	0.46	602.85	0.80	601.35	1.05
2009	639.5	622.49	2.66	625.39	2.21	623.13	2.56	621.03	2.89
2010	661.8	643.50	2.77	646.58	2.30	644.08	2.68	641.37	3.09
2011	681.5	665.21	2.39	668.49	1.91	665.74	2.31	662.37	2.81

Continued

2012	688.8	687.66	0.16	691.14	0.34	688.12	0.10	684.05	0.69
2013	699.8	710.87	1.58	714.56	2.11	711.26	1.64	706.45	0.95
2014	721.4	734.86	1.87	738.77	2.41	735.18	1.91	729.58	1.13
MAPE In-sample			2.05		2.08		2.04		2.08
2015	737.8	759.66	2.96	763.80	3.52	759.90	3.00	753.46	2.12
2016	748.6	785.30	4.90	789.68	5.49	785.46	4.92	778.13	3.94
2017	761.5	811.80	6.61	816.43	7.21	811.87	6.61	803.61	5.53
2018	791.0	839.20	6.09	844.09	6.71	839.17	6.09	829.92	4.92
MAPE Out-sample			5.14		5.73		5.16		4.13

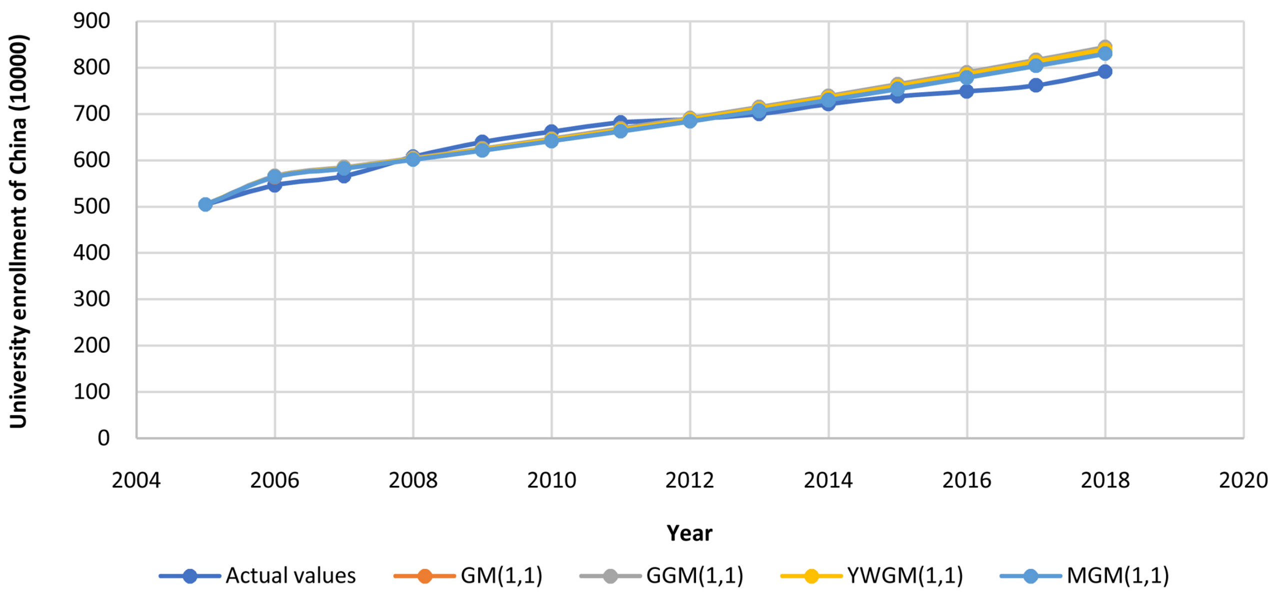

Figure 3. Performance of the above-mentioned models in terms of prediction accuracy in the new students enrolled in China's regular institutions of higher education (10,000 Persons).

The new approach's prediction efficiency (*i.e.*, modified GM(1,1) models) is thoroughly discussed. Figure 2 and Figure 3 demonstrate that the technique used in this analysis outperformed the standard approaches (*i.e.*, the standard model GM(1,1), the method of Guan-Jun (2000), and the method of Yao and Wang (2014)). In this analysis, it was revealed that altering the standard model GM(1,1)'s initial and background values reduced the prediction errors of the model significantly.

As seen in Table 2's in-sample predicting errors, the MAPEs of the standard model GM(1,1), the method of Guan-Jun (2000), the method of Yao and Wang (2014), and the modified GM(1,1) model from 2007 to 2015 were 5.40%, 6.04%, 5.58% and 7.30% respectively. Given the results of MAPEs in Table 2, it can be seen that the in-sample's MAPEs for all models were below 10%, which signifies highly prediction ability of the models.

In order to comparisons of out-of-sample predicting errors, the MAPEs of the standard model GM(1,1), the method of Guan-Jun (2000), the method of Yao and Wang (2014), and the modified GM(1,1) model from 2016 to 2017 were 11.44%, 15.17%, 11.98%, and 1.01% respectively. According to these findings, the MAPEs out-sample of the modified GM(1,1) model is less than 10%. Whereas, it can be observed that the MAPEs out-sample of the standard model GM(1,1), the method of Guan-Jun (2000), and the method of Yao and Wang (2014) are decreased to the next accuracy level (10% - 20%, *i.e.* good forecasting accuracy). As per the assessment standards presented in **Table 1**, this value revealed that the modified GM(1,1) model obtained high accurate forecasting compared with other proposed models. This shown that the accuracy of the modified GM(1,1) model's prediction is much higher than other models'. Significant differences existed between the outcomes of the three models as shown in **Table 2**. Those differences were caused mainly because the raw data is a fast-growth sequence.

As seen in **Table 3**'s in-sample predicting errors, the MAPEs of the standard model GM(1,1), the method of Guan-Jun (2000), the method of Yao and Wang (2014), and the modified GM(1,1) model between 2005 and 2014 were 2.05%, 2.08%, 2.04%, and 2.08% respectively. Lower values of MAPEs demonstrate a slight difference between expected and actual values. Given the findings of MAPEs in **Table 3**, it is observed that the in-sample's MAPEs for all models were below 10%, which signifies highly prediction ability of the models.

Concerning comparisons of the out-of-sample prediction errors, the MAPEs of the standard model GM(1,1), the method of Guan-Jun (2000), the method of Yao and Wang (2014), and the modified GM(1,1) model between 2015 and 2018 were 5.14%, 5.73%, 5.16%, and 4.13% respectively. According to these results, the performance of these models is 10% less than that of MAPEs. As per the assessment standards presented in **Table 1**, this degree of prediction accurateness achieves a high level of predicting accuracy. For the reason that the predictive accuracy of the standard model GM(1,1), the method of Guan-Jun (2000), the method of Yao and Wang (2014), and the modified GM(1,1) model is high, because the raw data is a low-growth sequence. Even though the predictive accuracy of these models is high, it can be noticeably observed that the modified GM(1,1) model has smaller values of MAPEs when compared to the standard model GM(1,1), the method of Guan-Jun (2000), and the method of Yao and Wang (2014).

In an effort to round up the above mentioned analyses, it can be recognized that the modified GM(1,1) model outperforms the standard model GM(1,1), the method of Guan-Jun (2000), and the method of Yao and Wang (2014) in terms of prediction accuracy in most of the cases. The modified GM(1,1) model exhibits high forecasting ability in equally the in-sample data and out-of-sample data according to the MAPE criteria. Therefore, the modified GM(1,1) model is able to increase model prediction accuracy efficiently, simulate low-growth rate raw data sequences, and similarly simulate high-growth rate raw data sequences. According to the prediction accuracies obtained for out-of-sample data (Table 2 and Table 3), the performance of the modified GM(1,1) model of two periods and four periods remains high in all the cases. Thus, the new approach is promising tool for short-term limited time series data forecasting.

4. Conclusion

The present work sets forward a modified model for improving predictive performance of the standard model GM(1,1) by optimizing initial and background values. The new initial value comprises the median value that has been generated using the first-order accumulative generating operation on the raw data while the number of observations is odd or even. Meanwhile, the background value is reconstructed using an integral term to correct the error term caused by the background value calculation in the standard model GM(1,1). To assess the predictive precision of the models in this article, two time series datasets were used. The findings illustrated that the modified GM(1,1) model considerably enhanced the accuracy of the grey prediction models not only for low-growth rate raw data sequences, but also for high-growth rate raw data sequences, and the new approaches could be appropriate tools for limited time series data forecasting. Therefore, it is of tremendous theoretical and practical importance, enriching the grey model's optimization theory and broadening its application reach. Further experiments on other data by the utilization of the modified GM(1,1) model reported in the present work might be conducted to analyze limited time series data.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Deng, J.L. (1982) Control Problems of Grey Systems. *Systems & Control Letters*, **1**, 288-294. https://doi.org/10.1016/S0167-6911(82)80025-X

[2] Deng, J.L. (1989) Introduction to Grey System Theory. *Journal of Grey System*, **1**, 1-24.

[3] Li, Q. and Lin, Y. (2014) A Briefing to Grey Systems Theory. *Journal of Systems Science and Information*, **2**, 178-192. https://doi.org/10.1515/JSSI-2014-0178

[4] Liu, S. and Forrest, J.Y.L. (2010) Grey Systems: Theory and Applications. Springer Science & Business Media, Berlin.

[5] Liu, S. and Lin, Y. (2006) Grey Information: Theory and Practical Applications. Springer Science & Business Media, Berlin.

[6] Yaoguo, D., Sifeng, L. and Kejia, C. (2004) The GM Models That $x(n)$ Be Taken as Initial Value. *Kybernetes*, **33**, 247-254. https://doi.org/10.1108/03684920410514175

[7] Xie, N.M. and Liu, S.F. (2009) Discrete Grey Forecasting Model and Its Optimization. *Applied Mathematical Modelling*, **33**, 1173-1186. https://doi.org/10.1016/j.apm.2008.01.011

[8] Juan, L. (2011) Application of Improved Grey GM (1, 1) Model in Tourism Revenues Prediction. 2011 *International Conference on Computer Science and Network Technology (ICCSNT)*, Vol. 2, Harbin, 24-26 December 2011, 800-802. https://doi.org/10.1109/ICCSNT.2011.6182084

[9] Wang, Y., Tang, J. and Cao, W. (2012) Grey Prediction Model-Based Food Security Early Warning Prediction. *Grey Systems: Theory and Application*, **2**, 13-23. https://doi.org/10.1108/20439371211197631

[10] Chen, Q. and Li, J. (2015) Research on Optimum Weighted Combination GM(1, 1) Model with Different Initial Value. *International Conference on Intelligent Computing* 2015, Fuzhou, 20-23 August 2015, 354-362. https://doi.org/10.1007/978-3-319-22053-6_39

[11] Madhi, M. and Mohamed, N. (2017) An Initial Condition Optimization Approach for Improving the Prediction Precision of a GM (1, 1) Model. *Mathematical and Computational Applications*, **22**, Article No. 21.

[12] Zeng, B., Zhou, M., Liu, X. and Zhang, Z. (2020) Application of a New Grey Prediction Model and Grey Average Weakening Buffer Operator to Forecast China's Shale Gas Output. *Energy Reports*, **6**, 1608-1618. https://doi.org/10.1016/j.egyr.2020.05.021

[13] Dai, W. and Chen, Y. (2007) Research of GM(1, 1) Background Value Based on Gauss-Legendre Quadrature and Its Application. 2007 *IEEE International Conference on Integration Technology*, Shenzhen, 20-24 March 2007, 100-102. https://doi.org/10.1109/ICITECHNOLOGY.2007.4290439

[14] Li, D.C., Yeh, C.W. and Chang, C.J. (2009) An Improved Grey-Based Approach for Early Manufacturing Data Forecasting. *Computers & Industrial Engineering*, **57**, 1161-1167. https://doi.org/10.1016/j.cie.2009.05.005

[15] Li, C. (2011) The Improved Background Value of Ameliorating GM (1, 1) Model and Its Application. 2011 2*nd International Conference on Intelligent Control and Information Processing*, Vol. 1, Harbin, 25-28 July 2011, 131-134. https://doi.org/10.1109/ICICIP.2011.6008214

[16] Yao, M. and Wang, X. (2014) Electricity Consumption Forecasting Based on a Class of New GM (1, 1) Model. In: Wang, W., Ed., *Mechatronics and Automatic Control Systems*, Springer, Cham, 947-953. https://doi.org/10.1007/978-3-319-01273-5_107

[17] Tsai, C.F. and Lu, S.L. (2015) The Exponential Grey Forecasting Model for CO_2 Emissions in Taiwan. *Grey Systems: Theory and Application*, **5**, 354-366. https://doi.org/10.1108/GS-05-2015-0026

[18] Cheng, M. and Shi, G. (2019) Improved Methods for Parameter Estimation of Gray Model GM (1, 1) Based on New Background Value Optimization and Model Application. *Communications in Statistics-Simulation and Computation*, **51**, 647-669. https://doi.org/10.1080/03610918.2019.1657450

[19] Li, K. and Zhang, T. (2019) A Novel Grey Forecasting Model and Its Application in Forecasting the Energy Consumption in Shanghai. *Energy Systems*, **12**, 357-372. https://doi.org/10.1007/s12667-019-00344-0

[20] Liu, D., Li, G., Chanda, E.K., Hu, N. and Ma, Z. (2019) An Improved GM(1.1) Model with Background Value Optimization and Fourier-Series Residual Error Correction and Its Application in Cost Forecasting of Coal Mine. *Gospodarka Surowcami Mineralnymi-Mineral Resources Management*, **35**, 75-98.

[21] Wu, L.Z., Li, S.H., Huang, R.Q. and Xu, Q. (2020) A New Grey Prediction Model and Its Application to Predicting Landslide Displacement. *Applied Soft Computing*, **95**, Article ID: 106543. https://doi.org/10.1016/j.asoc.2020.106543

[22] Yue, Z. and Diyi, S. (2021) Oil-Gas Cost in The Declining Period Predicting Model Based on Self-adaptive GM (1, 1, λ). 2021 6*th International Conference on Intelligent Computing and Signal Processing*, Xi'an, 9-11 April 2021, 69-72. https://doi.org/10.1109/ICSP51882.2021.9408831

[23] Sifeng, L., Yingjie, Y. and Jeffrey, F. (2016) Grey Data Analysis: Methods, Models and Applications. Springer, Berlin.

[24] Wei, B., Xie, N. and Hu, A. (2018) Optimal Solution for Novel Grey Polynomial Prediction Model. *Applied Mathematical Modelling*, **62**, 717-727. https://doi.org/10.1016/j.apm.2018.06.035

[25] Ye, J., Dang, Y. and Li, B. (2018) Grey-Markov Prediction Model Based on Background Value Optimization and Central-Point Triangular Whitenization Weight Function. *Communications in Nonlinear Science and Numerical Simulation*, **54**, 320-330. https://doi.org/10.1016/j.cnsns.2017.06.004

[26] Xu, N., Dang, Y. and Gong, Y. (2017) Novel Grey Prediction Model with Nonlinear Optimized Time Response Method for Forecasting of Electricity Consumption in China. *Energy*, **118**, 473-480. https://doi.org/10.1016/j.energy.2016.10.003

[27] Guan-Jun, T.A.N. (2000) The Structure Method and Application of Background Value in Grey System GM (1, 1) Model (I). *Systems Engineering-Theory & Practice*, **4**, 18.

[28] Lewis, C.D. (1982) Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting. Butterworth-Heinemann, Oxford.

[29] Xie, W., Wu, W.Z., Zhang, T. and Li, Q. (2020) An Optimized Conformable Fractional Non-Homogeneous Gray Model and Its Application. *Communications in Statistics-Simulation and Computation*, 1-16. https://doi.org/10.1080/03610918.2020.1788588