A Novel Mutual Fractional Grey Bernoulli Model With Differential Evolution Algorithm and Its Application in Education Investment Forecasting in China

WANLI XIE, BIN PU, CHUNYING PEI, SHIN-JYE LEE, AND YAN KANG

1Institute of EduInfo Science and Engineering, Nanjing Normal University, Nanjing 210097, China
2College of Computer Science and Electronic Engineering, Hunan University, Changsha 410006, China
3Aerospace Piesat Research Institute, Piesat Information Technology Company Ltd., Beijing 100048, China
4Institute of Technology Management, National Chiao Tung University, Hsinchu 30010, Taiwan
5National Pilot School of Software, Yunnan University, Kunming 650091, China

Corresponding authors: Bin Pu (pubin@hnu.edu.cn) and Chunying Pei (peichunying@piesat.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61762092 and Grant 61762089, and in part by the Ministry of Science and Technology Research Grant in Taiwan under Grant MOST 108-2218-E-009-041.

ABSTRACT Nowadays, the nonlinear grey Bernoulli model [NGBM(1,1)] has been successfully applied to various fields. Its main advantage is that the power exponent can better reflect the non-linear characteristics of the original data. However, the parameters of the model (i.e., the order of accumulation, coefficient of background value, and power index) must be optimized to fit the development law of the system. In this study, a fractional non-linear grey Bernoulli model [MFNGBM (1,1)] is proposed to reduce the perturbation limit of the classical NGBM and further improve the accuracy of the model, which uses mutual fractional operators and a new optimization scheme with a differential evolution (DE) algorithm for forecasting education investment. In the scheme, the power exponent of the Bernoulli differential equation, coefficient of background, and cumulative order of the original sequence are taken as decision variables, and their optimal parameters obtained by iteratively adjusting fitness functions. The experimental evaluation is conducted on two types of open-source data, and the results show that the proposed method can be very competitive with the popular baselines. Finally, MFNGBM(1,1) is used to predict China’s education investment in 2020–2025.

INDEX TERMS Grey system, grey forecasting model, non-linear grey bernoulli model, fractional order, differential evolution algorithm, educational investment.

I. INTRODUCTION

With rapid economic development and an aging population, the cultivation and building of a talent reserve have become one of the development strategies of various countries. Achieving this goal largely relies on investment in education. Taking the allocation of educational resources in China as an example, as shown in Fig. 1, one can easily find that China’s investment in educational resources is increasing year by year. Second, the distribution of educational resources in southeast China has been far higher than that in northwest China. To balance and promote the development of the education industry in the central and western regions, the government is trying to change this situation. Forecasting education investment funds has profound guiding significance when making corresponding policies for the government and provides corresponding strategies for the government to allocate educational resources, which have practical application values.

However, the studies of education investment forecasting have been ignored in recent years, and forecasting educational investment funds face two main challenges. First, although forecasting education investment may be related to diverse data, such as population statistics, economy, transportation, and area range, the access required to obtain these data is limited, and obtaining these data can be time-consuming, labor intensive, and delayed. In other words, how to improve prediction accuracy on a limited small sample...
dataset is required. Second, although education investment is sustainable, it may remain unstable because of government policy and natural disasters. Therefore, how to make predictions while reducing the randomness and uncertainty and requiring a smaller amount of operational data should be analysed. Grey-based approaches are suitable to address the above challenges. A grey-based model is suitable for uncertain systems with poor information, and it requires fewer operational data, so it can be used for education funds prediction. A grey forecasting method as a novel prediction method has been applied in many fields [1]–[6], and a number of attempts have been made to carry out the related Grey system studies.

Many systems in the real world, such as energy, agriculture, and education, can be theoretically regarded as an energy system, and their characteristic values follow the Grey exponential law. Based on this consideration, a pioneering work [7] by Deng, was given for small-sample forecasting, which is a basic grey model called GM(1,1). The main idea of GM(1,1) is to use the solution of the differential equation to establish a time-response equation, and then discretize the differential equation to obtain the difference equation. The estimated parameters of the difference equation are brought into the time-response sequence for prediction. With this structure, GM(1,1) can effectively fit a time series with exponential growth. However, many actual systems are affected more or less by other relative factors, and their characteristic values never follow the grey exponential law completely [8]. To solve this problem, a non-homogeneous exponential law is given in literature [9] that achieves better performance. Xie et al. proposed a discrete grey prediction model that can directly derive the time-response equation through the difference equation and avoid the error caused by the transition from the differential equation to the difference equation [10].

Zeng et al. proposed a new multivariable grey prediction method by adding the dependent lag term, linear correction term, and random perturbation term to the traditional GM(1,N) model, which achieved good results [11]. Zeng et al. followed this study and continued to present a new grey Verhulst model by introducing a new non-homogeneous exponential function [12]. To overcome the shortage of a static value and improve effectiveness, Xia et al. presented a real-time rolling grey prediction method that can provide efficient and accurate prediction of the health of machinery when considering and analyzing the influence of operational load and other factors [13]. This method greatly expands the range of using a grey prediction model and greatly improves applicability and effectiveness. Other important studies on improved grey models and applications have been reported [14]–[18].

Although the currently grey model has achieved success in many areas, it must be further improved. For example, the time-response function of many grey models is always a monotonic sequence. Therefore, these basic models are not applicable to raw data with strong volatility, and cannot fit the non-linear characteristics of real-world data. NGBM(1,1) is proposed to solve this problem [19]. Compared with the traditional grey model, the power index of the background value of the grey derivative is added to the grey interaction, where the power index is an uncertain parameter. Thus, as long as the power index and other structural parameters can be accurately found, the fluctuation characteristics of the data can be reflected.

Some studies have applied NGBM(1,1) in analyzing and forecasting problem. For example, Chen et al. used the non-linear grey Bernoulli model for forecasting unemployment rate and trading partners [20], [21]. Then, some methods were designed to optimize NGBM(1,1). Particle swarm optimization and genetic algorithms were used to optimize the power index in the model, achieving better results [22], [23]. Wang et al. [24] established a non-linear model to simultaneously optimize the background value and power index, further enhancing the model’s predictive ability, and applied the optimization model to the prediction and early warning of the compliance rate of industrial wastewater discharge. In addition, several different methods for optimizing NGBM (1,1) have been proposed, such as Nash NGBM (1,1) [25] and NLS-based NGBM (1,1) [26].

From the above analysis, it can be seen that the non-linear Bernoulli model is very dependent on the choice of parameters. Moreover, the current non-linear Bernoulli models are mostly integer-order accumulation, which limits the scope of use for a certain extent. This is mainly because integer-order accumulation may impede the model’s ability to achieve the best performance. Meanwhile, the structure parameters of the model should be dynamically fitted according to the characteristics of the data themselves.

Based on such considerations, a novel fractional Bernoulli prediction method, abbreviated MFNGBM(1,1), is proposed, which uses a new optimization scheme with a differential evolution (DE) algorithm to optimize the power index, order, and background value at the same time, further improving the prediction accuracy. Previous grey Bernoulli models did not consider optimizing power index, exponents, background value at the same time. Finally, MFNGBM(1,1) is established.
to predict China’s education funds. The main contributions of this study are the following:

- A new non-linear Grey Bernoulli model is presented using a mutual fractional operator, which can reduce the original data perturbation bounds.
- A framework to optimize MFNGBM(1,1) using a DE algorithm is proposed, which can optimize the power, coefficient of background value, and order of original sequence.
- According to the characteristics of the X sequence dataset and China’s education investment funds data, the proposed new model, that is, MFNGBM, is more adaptive and intelligent based on the optimization mechanism of the DE algorithm, which achieves the best performance.
- According to the experimental analysis of China’s education funds, the new MFNGBM model is more accurate than the traditional classical grey GM, NGM, DGM, and FGM models. Based on this, China’s educational funds in 2020–2025 are forecasted. Based on the prediction results, providing some effective suggestions to balance education sources.

The remainder of this paper is organized as follows: In Section 2, we describe the detail of the NGBM(1,1) and MFNGBM, and parameter optimization with DE algorithm. In Section 3, we verify the performance of the MFNGBM (1, 1) on two datasets and give a discussion. In Section 4, we apply MFNGBM (1, 1) to forecast education funding from 2020 to 2025 in China. In section 5, we conclude and give future work.

II. MFNGBM METHOD

In this section, we show the detail of our approach. MFNGBM is an extension of NGBM. Firstly, previous traditional NGBM is introduced. Secondly, Based on a study of NGBM, some related definitions, theorems, and proofs of our approach are given. To demonstrate the advantages of our method, two data sets are used to verify the proposed method. We simultaneously optimize the accumulation order, background value, power index of our model (i.e., MAPE as the objective function), and compare it with grey models including the classic NGBM (1, 1). At the same time, the DE optimizer and the classic GA optimizer for the optimization are compared in this paper.

A. NGBM (1,1)

Previous grey models almost were based on linear methods and theories. Deng first proposed a non-linear Bernoulli model to solve the problem of non-linear prediction under poor information small samples [1].

Assuming

\[ X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\} \]  

(1)

is the original sequence, and the corresponding first-order accumulation generation sequence is

\[ X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\} \]  

(2)

Then \( x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) \). The equation

\[
\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b(x^{(1)}(t))^\alpha 
\]

(3)

is called whitening equation of NGBM(1,1), and \( \alpha \) represents the power of the model. If \( \alpha = 1 \), the NGBM (1,1) will becomes the GM (1,1) model. We can get the discrete form of the model through the trapezoidal formula calculation

\[
x^{(0)}(k) + \alpha z^{(1)}(k) = b(z^{(1)}(k))^\alpha 
\]

(4)

where \( z^{(1)}(k) \) indicates the background value and it is obtained as

\[
z^{(1)}(k) = 0.5 \left( x^{(1)}(k) + x^{(1)}(k-1) \right) 
\]

(5)

Let

\[
B = \begin{bmatrix}
-z^{(1)}(2) & z^{(1)}(2)^\alpha \\
-z^{(1)}(3) & z^{(1)}(3)^\alpha \\
\vdots & \vdots \\
-z^{(1)}(n) & z^{(1)}(n)^\alpha 
\end{bmatrix}, \\
Y = \begin{bmatrix}
x^{(0)}(2) \\
x^{(0)}(3) \\
\vdots \\
x^{(0)}(n)
\end{bmatrix}^T 
\]

(6)

The model parameters can be estimated by the least squares as

\[
(a, b)^T = (B^T B)^{-1} B^T Y 
\]

(7)

Therefore, the solution to Eq. (3) is

\[
\hat{x}^{(1)}(k) = \left( x^{(0)}(1) - \frac{b}{a} \right) e^{-\alpha(1-a)(k-1)} + \frac{b}{a} \sum_{i=1}^{k} \frac{\Gamma(r+k-i)}{\Gamma(k-i)} x^{(0)}(i), \quad k = 1, 2, \cdots, n, 
\]

(8)

By first-order difference calculation of \( \hat{x}^{(1)}(k) \), the prediction result can be obtained as

\[
\hat{x}^{(k)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1) 
\]

(9)

B. FRACTIONAL NONLINEAR GREY BERNOULLI MODEL

The definition of fractional accumulation is given.

Definition 1 [27]: Let \( X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)) \) be a non-negative sequence. If the accumulated generating operator is applied \( r \) times on \( X^{(0)} \), \( r \in R^+ \), we obtain

\[
X^{(r)} = \left\{x^{(r)}(1), x^{(r)}(2), \cdots, x^{(r)}(n)\right\}
\]

(10)

where

\[
x^{(r)}(k) = \sum_{i=1}^{k} \frac{\Gamma(r+k-i)}{\Gamma(k-i+1) \Gamma(r)} x^{(0)}(i), \quad k = 1, 2, \cdots, n.
\]

(11)

The matrix expression of Eq. (11) is

\[
x^{(r)}(k) = \begin{bmatrix}
\Gamma(r+k-1) & \Gamma(r+k-2) & \cdots & \Gamma(r) \\
\Gamma(k) \Gamma(r) & \Gamma(k-1) \Gamma(r) & \cdots & \Gamma(1) \Gamma(r)
\end{bmatrix}
\]
\[
\begin{bmatrix}
    x^{(0)}(1) \\
    x^{(0)}(2) \\
    \vdots \\
    x^{(0)}(k)
\end{bmatrix}, \quad k = 1, 2, \cdots, n.
\] (12)

The definition of fractional subtraction operator is given.

**Definition 2** [27]: Set \(X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)), r \in R^+\), \(X^{(-r)} = (x^{(-r)}(1), x^{(-r)}(2), \ldots, x^{(-r)}(n))\) is the r-order decrementing sequence of \(X^{(0)}\), where

\[
x^{(-r)}(k) = \sum_{i=0}^{k-1} (-1)^{r} \frac{\Gamma(r+1)}{\Gamma(i+1)\Gamma(r-i+1)} x^{(0)}(k-i), \quad k = 1, 2, \cdots, n
\] (13)

The r-order cumulative generation and r-order decrementing generation have the following relationship,

\[
X^{(0)} = \left(x^{(r)}\right)^{(-r)} = \left(x^{(-r)}\right)^{(r)}
\] (14)

**Definition 3**: Establish a whitenization differential equation (WDE) for \(X^{(r)}\) as follows

\[
\frac{dx^{(r)}(t)}{dt} + ax^{(r)}(t) = b \left(x^{(r)}(t)\right)^{a}
\] (15)

Corresponding difference form of our model is

\[
x^{(r)}(k) - x^{(r)}(k-1) + ax^{(r)}(k) = b \left(x^{(r)}(k)\right)^{a}
\] (16)

Eq. (16) is shown the basic form of MFNGBM (1,1) model, where \(z^{(r)}(k) = \zeta x^{(r)}(k) + (1 - \zeta)x^{(r)}(k-1), \zeta \in [0, 1] \) is background value of \(X^{(r)}(k)\) in range of \([k - 1, k]\). In previous traditional Grey studies, the value of \(\zeta\) was always set as 0.5 [28]-[30]. There is a certain source of error for background value calculation, which uses right-angle trapezoid formula instead of the curved-edge trapezoid. As shown in Fig.2, \(x^{(1)}(k)\) is convex, concave and non-convex in each subinterval, respectively. The related definitions of estimation for the parameter are given.

**Definition 4**: Assuming that definition 1 and definition 2 exist, then the least-squares estimation matrix form of the parameters can be estimated as

\[
\hat{y} = [a, b]^T = (B^T B)^{-1} B^T Y
\] (17)

where

\[
B = \begin{bmatrix}
-z^{(r)}(2) & (z^{(r)}(2))^a \\
-z^{(r)}(3) & (z^{(r)}(3))^a \\
\vdots & \vdots \\
-z^{(r)}(n) & (z^{(r)}(n))^a
\end{bmatrix},
\]

\[
Y = \begin{bmatrix}
x^{(r)}(2) - x^{(r)}(1) \\
x^{(r)}(3) - x^{(r)}(2) \\
\vdots \\
x^{(r)}(n) - x^{(r)}(n-1)
\end{bmatrix}
\] (18)

because

\[
x^{(r-1)}(k) = x^{(r)}(k) - x^{(r)}(k-1)
\]

\[
= \sum_{i=1}^{k} \frac{\Gamma(r+k-i)}{\Gamma(k-i+1)\Gamma(r)} x^{(0)}(i) - \sum_{i=1}^{k-1} \frac{\Gamma(r+k-i-1)}{\Gamma(k-i)\Gamma(r)} x^{(0)}(i), \quad k = 2, 3, \cdots, n
\] (19)

when \(\zeta = 0.5\)

\[
z^{(r)}(k) = \frac{\sum_{i=1}^{k} \frac{\Gamma(r+k-i)}{\Gamma(k-i+1)\Gamma(r)} x^{(0)}(i)}{2} + \frac{\sum_{i=1}^{k-1} \frac{\Gamma(r+k-i-1)}{\Gamma(k-i)\Gamma(r)} x^{(0)}(i)}{2}, \quad k = 2, 3, \cdots, n
\] (20)

then (21), as shown at the bottom of this page.
For further derivation of $Y$, see Eq. (22), (23), as shown at the bottom of this page.

**Theorem 1:** [20]. In an initial condition $x^{(1)}(k)|_{k=1} = x^{(1)}(1)$, $k = 2, 3, \cdots, n$, the time response function can be written as

$$x^{(r)}(k - 1) = \left\{ \frac{b}{a} + \left[ \left( x^{(0)}(1) \right)^{1-a} - \frac{b}{a} \right] e^{-(1-a)ak} \right\}^{\frac{1}{1-a}} \quad (24)$$

$$B = \begin{bmatrix}
-1 2 \sum_{i=1}^{2} \frac{\Gamma(r + 2 - i)}{\Gamma(2 - i + 1) \Gamma(r)} x^{(0)(i)} + \frac{1}{2} \sum_{i=1}^{2} \frac{\Gamma(r + 1 - i)}{\Gamma(1 - i + 1) \Gamma(r)} x^{(0)(i)} \\
\vdots \vb \\
-1 2 \sum_{i=1}^{n} \frac{\Gamma(r + n - i)}{\Gamma(n - i + 1) \Gamma(r)} x^{(0)(i)} + \frac{1}{2} \sum_{i=1}^{n} \frac{\Gamma(r + n - 1 - i)}{\Gamma(n - i + 1) \Gamma(r)} x^{(0)(i)} \alpha
\end{bmatrix}
$$

$$= \begin{bmatrix}
-1 2 \left[ (r + 1)x^{(0)(1)} + x^{(0)(2)} \right] \vb \\
-1 2 \left[ \frac{r(r + 3)}{2} x^{(0)(1)} + (r + 1)x^{(0)(2)} + x^{(0)(3)} \right] \vb \\
\vdots \vb \\
-1 2 \left[ \frac{r(n - 1)}{2} x^{(0)(1)} + (r + 1)x^{(0)(2)} + \cdots + x^{(0)(n)} \right] \vb \\
\vdots \vb \\
-1 2 \left[ \frac{r(n - i - 1)}{2} x^{(0)(1)} + (r + 1)x^{(0)(2)} + \cdots + x^{(0)(n - i + 1)} \right] \vb \\
\vdots \vb \\
-1 2 \left[ \frac{r(n - n + 1)}{2} x^{(0)(1)} + (r + 1)x^{(0)(2)} + \cdots + x^{(0)(n - n + 1)} \right] \vb 
\end{bmatrix}
$$

$$= \begin{bmatrix}
a \\ x^{(r)(k)} = \sum_{i=1}^{k} \frac{(k - i + r - 1)(k - i + r - 2) \cdots (r + 1) r}{(k - i)!} x^{(0)(i)} , i = 1, 2, \cdots, n \\
\dot{x}^{(0)(k)} = \sum_{i=1}^{k - 1} \left( \frac{k - i - r}{k - i} \right) \dot{x}^{(r)(k - 1)} - \sum_{i=1}^{k - 1} \left( \frac{k - i - r - 1}{k - i - 1} \right) \dot{x}^{(r)(k - 1)} \\
\dot{x}^{(r)(k)} = \left( \frac{b}{a} + ce^{-(1-a)ak-1} \right)^{\frac{1}{1-a}} \\
z^{(r)(t)} = \zeta x^{(r)(k)} + (1 - \zeta) x^{(r)(k - 1)} \\
\dot{x}^{(0)(1)} = \dot{x}^{(r)(1)} \\
a = \sum_{k=2}^{n} \left( z^{(r)(k)} \right)^{a+1} \sum_{k=2}^{n} x^{(0)(k)} \left( z^{(r)(k)} \right)^{a} - \sum_{k=2}^{n} \left( z^{(r)(k)} \right)^{2a} \sum_{k=2}^{n} x^{(0)(k)} z^{(r)(k)} \\
b = \sum_{k=2}^{n} \left( z^{(r)(k)} \right)^{2a} \sum_{k=2}^{n} x^{(0)(k)} \left( z^{(r)(k)} \right)^{a} - \sum_{k=2}^{n} \left( z^{(r)(k)} \right)^{a+1} \sum_{k=2}^{n} x^{(0)(k)} z^{(r)(k)}
\end{bmatrix}
$$

s.t.
Proof: multiply \((x^{(r)}(t))^{-\alpha}\) on both sides of Eq. (15) and we get

\[
(x^{(r)}(t))^{-\alpha} \frac{dx^{(r)}(t)}{dt} + a(x^{(r)}(t))^{1-\alpha} = b
\]  

(25)

Let \(y^{(r)} = (x^{(r)}(t))^{1-\alpha}\), then

\[
\frac{dy^{(r)}(t)}{dt} = \frac{dx^{(r)}(t)}{dt}(1-\alpha)(x^{(r)}(t))^{-\alpha}
\]

(26)

Multiply 1 - \(\alpha\) on both sides of Eq. (26) and

\[
(1-\alpha)(x^{(r)}(t))^{-\alpha} \frac{dx^{(r)}(t)}{dt} + a(1-\alpha)(x^{(r)}(t))^{1-\alpha} = b(1-\alpha)
\]

(27)

Substitute (26) into Eq. (27) and we get

\[
\frac{dy^{(r)}(t)}{dt} + (1-\alpha)ay^{(r)}(t) = b(1-\alpha)
\]

(28)

Solve Eq. (28) and we obtain

\[
y^{(r)}(t+1) = e^{-(1-\alpha)a}[(1-\alpha)]be^{(1-\alpha)at}dt + c
\]

(29)

accordingly

\[
\hat{x}(t+1) = \left\{ e^{-(1-\alpha)a} \left[ (1-\alpha) \int be^{(1-\alpha)at}dt + d \right] \right\}^{1/(1-\alpha)}
\]

(30)

Let \(\hat{x}^{(r)}(0) = x^{(0)}(1)\) and we get

\[
\hat{x}^{(r)}(k) = \left\{ \frac{b}{a} + \left( x^{(0)}(1) \right)^{1-\alpha} - \frac{b}{a} \right\}^{1/(1-\alpha)}
\]

(31)

Finally, we get the prediction results by

\[
\hat{x}^{(0)}(k) = \begin{cases} 
\hat{x}^{(r)}(k), & k = 1 \\
\sum_{i=0}^{k-1} (-1)^i \Gamma(r+1) \Gamma(i+1) \Gamma(r-i+1) \hat{x}^{(r)}(k-i), & \text{else}
\end{cases}
\]

(32)

C. PARAMETER ESTIMATION BASED ON DE ALGORITHM

The power, order, and background value coefficients of the MFNGBM are not given by an exact value. In this paper, we can find the best parameters of MFNGBM by DE algorithm.

1) DIFFERENTIAL EVOLUTION ALGORITHM

DE algorithms, which were first introduced in [31], depend on a few input parameters and require a very modest computational compared with the most traditional genetic algorithm (GA). Let \(i = 1, 2, \ldots, N\), \(j = 1, 2, \ldots, L\), \(rand \in [0,1]\), \(t\) is the current generation, \(j \in [1, L]\), and \(j_i\) is an integer, \(f\) is a fitness function, the main steps of DE algorithm are described as follows:

(1) Initialization

Encode population in an interval of \([U_{min}, U_{max}]\) and the initialization is described as

\[
X_{ij} = U_{min} + rand \times (U_{max} - U_{min})
\]

(33)

where \(X_{ij}\) represents the \(j - th\) gene on \(i - th\) chromosome.

(2) Mutation

Two different individuals in the population are randomly selected and the vector difference is scaled to make vector synthesis with the individual to be mutated.

\[
MX_{ij}^t = X_{id1}^t + F \times (X_{nd2}^t - X_{md3}^t)
\]

(34)

where \(F\) is the individual variation rate, and \(X_{nj}^t\) represents the \(rd - th\) individual in the \(t - th\) generation population.

(3) Crossover

If the random rate of individuals in the population is less than the crossover rate \(CR\), perform crossover operation.

\[
CX_{ij}^t = \begin{cases} 
MX_{ij}^t, & \text{if rand} (0,1) \leq CR \text{ or } j = j_{raud} \\
X_{ij}^t, & \text{otherwise}
\end{cases}
\]

(35)

where \(CR\) is individual crossover rate.

(4) Selection

The greedy algorithm is used to select individuals to enter the next generation of population.

\[
X_{i+1}^t = \begin{cases} 
CX_{ij}^t, & \text{if f} (CX_{ij}^t) < f (X_{ij}^t) \\
X_{ij}^t, & \text{otherwise}
\end{cases}
\]

(36)

FIGURE 3. Optimization process of MFNGBM (1,1) with DE algorithm.

2) OPTIMIZATION OF THE HYPERPARAMETERS WITH DIFFERENTIAL EVOLUTION ALGORITHM

Three parameters are optimized: power exponent of Bernoulli differential equation \(\alpha\), coefficient of background value \(r\) and cumulative order of original sequence \(\zeta\). As explained in Eq. (37), MAPE (mean absolute percentage error) is chose as objective function. Hyperparameters corresponding to the minimum MAPE will be obtained after optimization process. Optimization process is shown in Fig.3. We construct the
mathematical programming model for optimization, which is formulated as Eq. (23).
\[
\min \frac{1}{r \cdot \alpha \cdot \zeta} \times 100\% (37)
\]
where \( \alpha, r \) and \( \zeta \) represent power exponent of Bernoulli differential equation, production coefficient of background and cumulative order of original sequence, respectively. \( x^{(0)}(k) \) and \( \hat{x}^{(0)}(k) \) are ground truth and the predicted series, respectively. Parameters optimization steps of MFNGBM (1,1) are described as follows:

**Step 1:** Calculate the order accumulating generation sequence of an original data sequence.

**Step 2:** If iterations of algorithm reach the limit, then the procedure goes to step 6, otherwise, it goes to step 3.

**Step 3:** Perform mutation on three individuals randomly according to Eq. (33).

**Step 4:** Perform crossover operation according to Eq. (35).

**Step 5:** Construct a new population by selecting new individuals according to Eq. (36). If the algorithm satisfies the termination condition, it goes to step 6, if not then go to step 3.

**Step 6:** Output the optimal individual and get the optimal solution after optimization. Correspondingly, pseudo-code of the above optimization process is described in Algorithm 1.

Algorithm 1 Framework of Optimization for Solving the Constrained Optimization by DE Optimizer

**Input:**
- The given dataset of historical sequence \((x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})\); Dimension: D; Generation: T;

**Output:**
- The optimized vector \((r, \alpha, \zeta) - \Delta\);
- Initialize: the DE population M; the error \( \varepsilon \);
1: \( (r, \alpha, \zeta) \leftarrow M \);
2: \( (x_1^{(r)}, x_2^{(r)}, \ldots, x_n^{(r)}) \) \leftarrow Eq.(11);
3: \( [\hat{a}, \hat{b}] \) \leftarrow Eq.(18);
4: \( f \leftarrow \) objective function in Eq.(37);
5: \( t \leftarrow 1 \) (initialization);
6: for \( i = 1 \) to \( M \) do
7: \( (x^{(l)}, x^{(i)}) \leftarrow Eq.(33); \)
8: end for
9: while \(((f(\Delta) \geq \varepsilon) \lor (t \leq T))\) do
( Mutation and Crossover ) \leftarrow Eq.(33) and Eq.(34);
10: end while
11: return the best vector \( \Delta \)

### III. SIMULATION EXPERIMENTS

In this section, two datasets are used to evaluate the proposed approach and a discussion is given. Through the above analysis, it indicates that MFNGBM (1,1) can set appropriate parameters according to the law of the data itself, and can fit different types of data to characterize the development rules of different systems. The performance of MFNGBM (1,1) is evaluated on two datasets, which is \( X \) sequence data [32, 33] and the education investment data in real-world come from National Bureau of Statistics of China. \( X \) data contains six type characteristics of time series, which is widely used for evaluating Grey methods. The experiment on Chinese investment funds dataset is to verify the performance of our method in real-world dataset.

#### A. DATASET AND SETTING

The MFNGBM is evaluated with the best performance on two datasets, containing \( X \) sequence data and Chinese education funds. The investment funds dataset can be obtained at http://www.stats.gov.cn. The \( X \) data contains six sequences, which contains the characteristics of rising convex, rising concave, falling convex, falling concave, strictly non-homogeneous, and approximate non-homogeneous. These six type characteristics represent the diversity of \( X \) sequence. Chinese education funds are real-world dataset, which comes from the official website of the National Bureau of Statistics of China. All experiments in this paper were deployed on a PC with configurations as follows: Intel Core i7-2600k 3.40GHz (8 cores), 8GB of RAM, Windows 10 of Operating System. Matlab libraries and program are used to build our model. The max iteration, scaling factor, and crossover factor are set to 30, 0.5 and 0.8, respectively.

#### B. ASSESSMENT CRITERIA

MAPE is used to evaluate predicting performance of baselines and our approach, which is defined as follows:
\[
\text{MAPE} = \frac{1}{n} \sum_{k=1}^{n} \left| \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \right| \times 100\% (38)
\]
where \( x^{(0)}(k) \) is ground truth, \( \hat{x}^{(0)}(k) \) is the predicted series. Lewis’ [34] benchmark of accuracy evaluation is applied in this paper, which is listed in Table 1.

| \( \text{MAPE} \) | \( \leq 10\% \) | \( 10\%-20\% \) | \( 0\%-50\% \) | \( \geq 50\% \) |
|------------------|-----------------|-----------------|-----------------|-----------------|
| evaluation       | accurate        | good            | reasonable      | inaccurate      |

#### C. EXPERIMENTAL RESULTS

1) RESULTS ON \( X \) DATA

Five popular Grey models are chosen baselines, which is GM(1,1) [7], NGM(1,1, k) [9], GM(1, 1, \( r^2 \)), DGM(1, 1) [10], FGM(1, 1, k) [35] and NGBM (1, \( k, k \)). Six baselines and our approach are employed to predict on six sequences and their predictive errors are shown in Table 2. As seen in Table 2, the predictive accuracy of MFNGBM wins the smallest error compared with those of the classical models. The predictive accuracy of MFNGBM is significantly improved compared with all baselines on six different sequence data, which indicates that the proposed approach adapts to different conditions and shows the robustness of MFNGBM. Our approach
achieves the best performance in X-sequence data prediction, which demonstrates the overall effectiveness of MFNGBM. Compared with NGBM (1,1) and MFNGBM (1,1) on X sequence data, MFNGBM (1,1) achieve lower error, which further demonstrates the extension version of NGBM (1,1) [i.e., MFNGBM (1,1)] is effective. It mainly because that classic grey Bernoulli model is integer order accumulation, which is not necessarily the best accumulation of the model.

In addition, we compared the performance of MFNGBM optimization using GA and DE algorithms on the X sequence dataset in 10 times independently. As shown in Fig.4, compared with DE, the GA algorithm achieves the small errors (average of 10 optimization results) at the beginning of the iteration. And then, their errors meet at an uncertain iteration, after the encounter, the errors of the DE algorithm decrease rapidly with the number of iterations compared with GA algorithm. As seen from Fig.4 (a)-(e), the DE algorithm reaches the smallest prediction errors after 30 iterations compared with DA, which denotes the advantages of the DE.

2) RESULTS ON EDUCATION FUNDS IN CHINA
The above chapter shows that the MFNGBM optimized by DE is suitable for small sample modeling without additional data. We conduct the experimental to predict on education investment data in China. The first experimental setting is to show the stability of MFNGBM. We observe the convergence of MFNGBM under different population sizes. As shown in Fig.5 (a), MFNGBM is always convergent when the population size is set 30 to 90. Meanwhile, the running fitness curve of each MFNGBM is convergent after oscillating within a certain range, as depicted in Fig.5 (b). The first and
second experiment indicates the better stability performance of MFNGBM. Next, to better understand the parameter change process, the correlation between power, order and background value coefficient is visualized under a 20 population size and a different iteration. The iteration of Fig.6 (a) to (c), (d) to (f), (g) to (i), and (j) to (l) are set as 1, 4, 10, 25, respectively. In Fig.6 (a) to (c), the MAPE is larger than others (Fig. 6 (d) to (l)), which indicates that there are a large number of solutions in the solution space. With the number of iterations increases, the poorer solutions are gradually eliminated and the better solutions are in the solution space, which indicates a better model with a smaller error (depicted Fig. 6 (j) to (l)).

We used historical data from 2006 to 2015 to train the model. The education funds of 2016 and 2017 are used as tests. Fig. 7 shows the curve of raw data, the prediction of our approach and baselines. Predictive errors of China’s
TABLE 3. Numerical results of prediction on China’s education investment funds data by the MFNGBM (1,1), DGM (1,1), DGM (2,1), GM (1,1) and NGBM (1,1).

| Year | Raw data | MFNGBM(1,1) | DGM(1,1) | DGM(2,1) | GM(1,1) | NGBM(1,1) |
|------|----------|-------------|----------|----------|--------|-----------|
|      | Fund     | MAPE        | Pre.     | MAPE     | Pre.   | MAPE     | Pre.     | MAPE     | Pre.   |
| 2007 | 121480663| 0.000       | 121480663| 0.000    | 121480663| 0.000    | 121480663| 0.000    | 121480663| 0.000 |
| 2008 | 145007374| 1.969       | 158563461| 9.715    | 134749259| 7.074    | 158094555| 9.025    | 13869522| 4.160 |
| 2009 | 165027065| 4.174       | 179455843| 3.868    | 161788249| 1.962    | 178952574| 8.438    | 169527898| 2.720 |
| 2010 | 195618471| 4.619       | 203101013| 6.007    | 189595963| 3.078    | 202562470| 3.550    | 203607329| 4.080 |
| 2011 | 238692936| 0.153       | 229861679| 3.321    | 218194255| 8.587    | 229287310| 3.940    | 238457939| 0.090 |
| 2012 | 286530532| 5.204       | 260148340| 1.324    | 247605602| 13.59    | 259538059| 9.428    | 272463256| 4.910 |
| 2013 | 303671812| 0.002       | 294425543| 5.206    | 277853119| 8.494    | 293779906| 3.250    | 304518886| 0.280 |
| 2014 | 328064609| 1.715       | 333219209| 6.994    | 308960578| 5.823    | 332539411| 1.364    | 333884729| 1.770 |
| 2015 | 361291927| 0.002       | 377124298| 4.853    | 340952428| 5.629    | 376412604| 4.185    | 360128742| 4.160 |
|      | Average of MAPE | 2.230 | 5.392 | 6.027 | 5.398 | 2.297 |
| 2016 | 388883850| 0.684       | 426814337| 9.750    | 373855813| 3.860    | 426074156| 9.560    | 383042247| 1.502 |
| 2017 | 425620069| 4.059       | 483051555| 13.400   | 407690589| 4.212    | 482287746| 13.300   | 402612309| 5.405 |
|      | Average of MAPE | 2.371 | 11.823 | 4.038 | 11.438 | 3.454 |

FIGURE 7. The scatter diagrams with prediction results of the all models.

FIGURE 8. The overall MAPE of all approaches.

Accurate prediction of educational funds is an important theoretical basis for scientific decision-making and balanced education resources, which is related to the overall development of the educational cause [36]. The above sections have shown the advantages of our approach. Education funds from 2020 to 2025 are predicted using the proposed approach in China. Prediction results are shown in Fig.9 (a). It is a continuous growth between 2020 and 2025, which probably indicates the government has been devoted to investing more funds in developing education.

FIGURE 9. (a) is China’s education funds between 2020 and 2025 based on prediction of our approach. (b) is growth rate of China’s education funds between 2020 and 2025.

We can see a slight decrease in the rate of Chinese investment funds between 2020 to 2025 in Fig.9 (b). This may indicate that China’s education infrastructure has gradually improved.

IV. CONCLUSION
Forecasting education investment funds is an important part of developing a good education system and growing the talent reserve. Scientific, reasonable, and accurate forecasting of China’s education funds can provide effective suggestions to solve the balanced education investment problem. Based on the features of grey prediction theory and the fractional extension operator, a fractional non-linear grey Bernoulli model [MFNGBM(1,1)] is proposed in this study to reduce...
the perturbation limit of the classic NGBM model and further improve the accuracy of the model, which uses mutual fractional operators and a new optimization scheme with a differential evolution (DE) algorithm for forecasting education investment. Although the MFNGBM (1,1) model achieves excellent performance, there is some improvement work to be done in the future.

- The proposed approach is a relatively complex structure based on the mutual fractional order theory and the non-linear Bernoulli model, which may have some limits. In the next work, the MFNGBM should be simplified.
- Although we realize the optimization of parameters of our approach and obtain better performance, a more efficient optimization framework should be constructed in the future.

REFERENCES

[1] S. Liu, Y. Yang, and J. Forrest, Grey Data Analysis. Springer, 2017.
[2] N. Xie, R. Wang, and N. Chen, “Measurement of shock effect following change of one-child policy based on grey forecasting approach,” *Kybernetes*, vol. 47, no. 3, pp. 559–568, Mar. 2018.
[3] C.-C. Chuang, T.-T. Chen, and C.-C. Chen, “Application of grey theory in the construction of impact criteria and prediction model of Players’ salary structure,” *Math. Problems Eng.*, vol. 2018, pp. 1–9, Feb. 2018.
[4] N. Zhang, C. Xiao, B. Liu, and X. Liang, “Groundwater depth predictions by GSM, RBF, and ANFIS models: A comparative assessment,” *Arabian J. Geosci.*, vol. 10, no. 8, pp. 189, Apr. 2017.
[5] X. Sun, W. Sun, J. Wang, Y. Zhang, and Y. Gao, “Using a Grey–Markov model optimized by cuckoo search algorithm to forecast the annual foreign tourist arrivals to China,” *Tourism Manage.*, vol. 52, pp. 369–379, Feb. 2016.
[6] A. Bezuglov and G. Comert, “Short-term freeway traffic parameter prediction: Application of grey system theory models,” *Expert Syst. Appl.*, vol. 62, pp. 284–292, Nov. 2016.
[7] J. L. Deng, “Introduction to grey system theory,” *J. Grey Syst.*, vol. 1, no. 1, pp. 1–24, 1989.
[8] H.-W.-V. Tang and M.-S. Yin, “Forecasting performance of grey prediction for education expenditure and school enrollment,” *Econ. Edu. Rev.*, vol. 31, no. 4, pp. 452–462, Aug. 2012.
[9] P.-Y. Chen and H.-M. Yu, “Foundation settlement prediction based on a novel NGM model,” *Math. Problems Eng.*, vol. 2014, pp. 1–8, Mar. 2014.
[10] N.-M. Xie and S.-F. Liu, “Discrete grey forecasting model and its optimization,” *Appl. Math. Model.*, vol. 33, no. 2, pp. 1173–1186, Feb. 2009.
[11] B. Zeng, H. Duan, and Y. Zhou, “A new multivariable grey prediction model with structure compatibility,” *Appl. Math. Model.*, vol. 75, pp. 385–397, Nov. 2019.
[12] B. Zeng, M. Tong, and X. Ma, “A new-structure grey verhulst model: Development and performance comparison,” *Appl. Math. Model.*, vol. 81, pp. 522–537, May 2020.
[13] T. Xia, X. Jin, L. Xi, Y. Zhang, and J. Ni, “Operating load based real-time rolling grey forecasting for machine health prognosis in dynamic maintenance schedule,” *J. Intell. Manuf.*, vol. 26, no. 2, pp. 269–280, Apr. 2015.
[14] B. Zeng, X. Ma, and J. Shi, “Modeling method of the grey GM(1,1) model with interval grey action quantity and its application,” *Complexity*, vol. 2020, pp. 1–10, Jan. 2020.
[15] S.-H. Yang, W.-J. Huang, J.-F. Tsai, and Y.-P. Chen, “Symbiotic structure learning algorithm for feedforward Neural-Network-Aided grey model and prediction applications,” *IEEE Access*, vol. 5, pp. 9378–9388, 2017.
[16] D. Zheng, H. Wang, J. An, J. Chen, H. Pan, and L. Chen, “Real-time estimation of battery state of charge with metabolic grey model and LabVIEW platform,” *IEEE Access*, vol. 6, pp. 13170–13180, 2018.
[17] Z. Ma, Y. Dong, H. Liu, X. Shao, and C. Wang, “Forecast of non-equal interval track irregularity based on improved grey model and PSO-SVM,” *IEEE Access*, vol. 6, pp. 34812–34818, 2018.
[18] Y. Zhang, H. Sun, and Y. Guo, “Wind power prediction based on PSO-SVR and grey combination model,” *IEEE Access*, vol. 7, pp. 136254–136267, 2019.
[19] T. Xia and L. Xi, “Manufacturing paradigm-oriented PHM methodologies for cyber-physical systems,” *J. Intell. Manuf.*, vol. 30, no. 4, pp. 1659–1672, Apr. 2019.
[20] C.-I. Chen, “Application of the novel nonlinear grey Bernoulli model for forecasting unemployment rate,” *Chaos, Solitons Fractals*, vol. 37, no. 1, pp. 278–287, Jul. 2008.
[21] C.-I. Chen, H. L. Chen, and S.-P. Chen, “Forecasting of foreign exchange rates of Taiwan’s major trading partners by novel nonlinear grey Bernoulli model NGBM(1,1),” *Commun. Nonlinear Sci. Numer. Simul.*, vol. 13, no. 6, pp. 1194–1204, Aug. 2008.
[22] J. Zhou, R. Fang, Y. Li, Y. Zhang, and B. Peng, “Parameter optimization of nonlinear grey Bernoulli model using particle swarm optimization,” *Appl. Math. Comput.*, vol. 207, no. 2, pp. 292–299, Jan. 2009.
[23] L.-C. Hsu, “A genetic algorithm based nonlinear grey Bernoulli model for output forecasting in integrated circuit industry,” *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4318–4323, Jun. 2010.
[24] Z.-X. Wang, K. W. Hipel, Q. Wang, and S.-W. He, “An optimized NGBM(1,1) model for forecasting the qualified discharge rate of industrial wastewater in China,” *Appl. Math. Model.*, vol. 35, no. 12, pp. 5524–5533, Dec. 2011.
[25] C.-I. Chen, P.-H. Hsin, and C.-S. Wu, “Forecasting Taiwan’s major stock indices by the Nash nonlinear grey Bernoulli model,” *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7557–7562, Dec. 2010.
[26] L. Pei, Q. Li, and Z. Wang, “The NLS-based nonlinear grey Bernoulli model with an application to employee demand prediction of high-tech enterprises in China,” *Grey Syst., Theory Appl.*, vol. 8, no. 2, pp. 133–143, Apr. 2018.
[27] B. Zeng and S. Liu, “A self-adaptive intelligence grey prediction model with the optimal fractional order accumulating operator and its application,” *Math. Methods Appl. Sci.*, vol. 40, no. 18, pp. 7843–7857, Dec. 2017.
[28] M. Hassan Mad and N. Mohamed, “Optimized GM(1,1) model based on the modified initial condition,” *J. Appl. Sci.*, vol. 17, no. 2, pp. 90–96, Jan. 2017.
[29] Z. F. Chen and W. H. Huang, “GM (1,1) model predicting elevator spare parts based on trigonometric function of variable coefficient,” *J. Discrete Math. Sci. Cryptogr.*, vol. 21, no. 6, pp. 1183–1187, Aug. 2018.
[30] K. Zhang and Q. Pinpin, “Multivariate discrete grey model base on dummy drivers,” in *Proc. IEEE Int. Conf. Grey Syst. Intell. Services (GSIS)*, Aug. 2015, pp. 266–271.
[31] R. Storn and K. Price, “Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces,” *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, 1997.
WANLI XIE received the B.S. degree in computer science and technology from the WenZheng College Soochow University, Suzhou, China, in 2015, and the M.S. degree from the School of Marine Science and Engineering, Nanjing Normal University, Nanjing, China, in 2019. He is currently pursuing the Ph.D. degree with the Institute of EduInfo Science and Engineering, Nanjing Normal University. His current research interests include Grey system theory, machine learning, and fractional difference theory.

BIN PU received the M.S. degree in software engineering from the National Pilot School of Software, Yunnan University, Kunming, China, in 2018. He is currently pursuing the Ph.D. degree in computer science with Hunan University, China. His research interests include deep learning on big data and data mining.

CHUNYING PEI received the B.S. degree in cartography and geographic information system from the Xi'an University of Science and Technology, Xi'an, China, in 2018. From 2016 to 2017, he was a Research Assistant with Institute of Geography, Chinese Academy, Beijing, China. He currently works at the Aerospace Piesat Research Institute, Piesat Information Technology Company Ltd. His team focuses on remote sensing, navigation, and Grey model.

SHIN-JYE LEE is currently an Associate Professor with the Institute of Technology Management, National Chiao Tung University, Taiwan, and the M.Sc. (Eng) degree from the Department of Computer Science, University of Sheffield, U.K., in 2001, the M.Phil. degree from the Judge Business School, University of Cambridge, U.K., in 2012, and the Ph.D. degree from the School of Computer Science, University of Manchester, U.K., in 2011. Before that, he was the Professor of National Pilot School of Software, Yunnan University, China. He also made his academic career in Poland, in 2012 winter. In addition, he also had practical experiences in Fujitsu and Microsoft, from 2002 to 2005. Further, his research interests primarily comprise machine learning, computational intelligence and decision support systems, operational research, and technology policy, especially for the climate change issues, and energy prediction.

YAN KANG received the Ph.D. degree in computer software and theory from the Institute of Software Chinese Academy of Sciences, Beijing, China, in 2003. She is currently a Senior Software Architect, an Associate Professor with the Software Institute of Yunnan University, and a Core Member of the Key Laboratory of Software Engineering and Data Science of Yunnan Province. Her research interests include software engineering, system optimization, big data processing, and mining.