Article

An Optimized Damping Grey Population Prediction Model and Its Application on China’s Population Structure Analysis

Xiaojun Guo 1,*, Rui Zhang 1,*, Houxue Shen 1 and Yingjie Yang 2

1 School of Science, Nantong University, Nantong 226019, China
2 Institute of Artificial Intelligence, De Montfort University, Leicester LE1 9BH, UK
* Correspondence: guoxj159@163.com (X.G.); 2002310007@stmail.ntu.edu.cn (R.Z.)

Abstract: Population, resources and environment constitute an interacting and interdependent whole. Only by scientifically forecasting and accurately grasping future population trends can we use limited resources to promote the sustainable development of society. Because the population system is affected by many complex factors and the structural relations among these factors are complex, it can be regarded as a typical dynamic grey system. This paper introduces the damping accumulated operator to construct the grey population prediction model based on the nonlinear grey Bernoulli model in order to describe the evolution law of the population system more accurately. The new operator can give full play to the principle of new information first and further enhance the ability of the model to capture the dynamic changes of the original data. A whale optimization algorithm was used to optimize the model parameters and build a smooth prediction curve. Through three practical cases related to the size and structure of the Chinese population, the comparison with other grey prediction models shows that the fitting and prediction accuracy of the damping accumulated–nonlinear grey Bernoulli model is higher than that of the traditional grey prediction model. At the same time, the damping accumulated operator can weaken the randomness of the original data sequence, reduce the influence of external interference factors, and enhance the robustness of the model. This paper proves that the new method is simple and effective for population prediction, which can not only grasp the future population change trend more accurately but also further expand the application range of the grey prediction model.

Keywords: population prediction; grey system; damping accumulated operator; grey Bernoulli model; whale optimization algorithm

1. Introduction

Population and environment are coupled to form a mutually restrictive and interdependent organic whole. The population has always been a key issue in the development of human society, which directly affects environmental change, economic development and social stability [1]. Only by properly handling the relationship between population, resources and environment can we promote the sustainable development of society and realize the comprehensive and coordinated development of man and nature [2]. At the same time, in the relationship between economic development, the population, as an index to measure the degree of social progress, can reflect the economic development level of a region. From an international perspective, the trend of world population development is an important basis for the United Nations to formulate a global population strategy. Similar to other countries, the formulation of basic state policies, the arrangement of labor and employment, the development of social welfare, and even the standards of national economic and social development strategies in China are profoundly affected by population change. Especially under the situations that China is facing, such as the problems of an unbalanced birth rate, insufficient labor supply, weakening demographic dividend and serious aging trend, it is of great significance to understand the future population development trend of China,
accurately grasp the population status, and scientifically predict the future population number for the scientific decision-making of population policy and the construction of a harmonious society.

A population forecast is a reliable and accurate forecast of the future population development trend based on the existing information, so as to make population activities and economic development move forward in a favorable direction. Choosing a reasonable model to accurately describe and forecast population and various population indicators has a decisive significance for formulating a suitable plan for population development which, coordinated with social and economic development, can provide scientific decision-making basis for economic and social development. Graunt and Erle et al. studied population prediction 300 years ago, focusing on the analysis and research from the perspective of age distribution and population evolution [3]. Later, Leslie et al. began to pay attention to the influence of deterministic models on population prediction and introduced population dynamics into population prediction [4]. With the rapid growth of the world population, the impact of population on society is increasing day by day. The population problem has become one of the major issues causing concern for the world. Common models for population prediction include the differential equation model, Leslie model, stochastic prediction model, mathematical statistics method, grey model, BP neural network model, linear time series model, deep learning models, etc. [5–8]. For the convenience of a comparative study, we summarize the mechanisms, assumptions, merits and drawbacks of mainstream population prediction methods, as shown in Table 1.

Table 1. Summary of mainstream population prediction methods.

| Method                | Mechanism                                      | Assumption                                                                 | Merits                                                                                                           | Drawbacks                                                                                                     |
|-----------------------|------------------------------------------------|---------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| Malthus model         | Exponential model of population growth.         | It is assumed that the change rate of the population is constant in an independent biological population. | The model has a deep historical background and it is suitable for short-term forecasting.                      | The model ignores the factors that have an important impact on the birth rate, such as limited living resources, space, productivity level and so on. |
| Leslie matrix model   | This model is a difference equation in vector form and a population growth model grouped by age. | Assume the female fertility rate and mortality rate of all ages.          | The method is classic and can clearly reflect the change in population development by age.                     | The model has high requirements for statistical data on age structure, and the changes in birth rate and mortality over time are not considered. |
| Logistic model        | The population is predicted by solving the differential equation of population growth rate and population size. | Assume the population growth rate and maximum population capacity of the environment. | Considering the influence of environmental factors on population growth, the model has a simple structure and high prediction accuracy. | It struggles to deal with difficult nonlinear problems and the error fluctuation of long-term prediction is relatively large. |
| Markov chain model    | A model of nondeterministic analysis in which the relationships between variables are given in the form of statistical values. | Under the condition that the present situation is known, the future situation of the system is only related to the present moment, not the past history. | The application range of the model is wide and the processing process is simple. | This model cannot make full use of existing information and is not suitable for long-term prediction. |
### Table 1. Cont.

| Method                     | Mechanism                                                                 | Assumption                                                                 | Merits                                                                 | Drawbacks                                                                                                                                             |
|----------------------------|---------------------------------------------------------------------------|----------------------------------------------------------------------------|--------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| Grey prediction model      | The randomness of the original sequence is weakened by the accumulation operator and constructs the equation with partial difference and partial differential characteristics. | Population system is an uncertain system with only partial information known. | Small data requirements, simple modeling process and high prediction accuracy.                                                                     | It is applicable to short-term and medium-term forecasts, and the data need to conform to the indexation trend.                                      |
| Mathematical statistics method | Use mathematics and statistics knowledge to establish population prediction equation. | A priori condition of population information is required.                  | The theoretical basis of the model is completed and the prediction accuracy is high.                                                                | The modeling conditions are strict, the model is complex and it requires prior assumptions.                                                           |
| Time series prediction model | Uses the continuous law of the development of objective things, through statistical analysis to further predict the future development trend of the market. | There is no need to make model assumptions.                              | The time series prediction method is suitable for short- and medium-term prediction. It is very convenient to implement the model by computer. | When the historical data fluctuate greatly, the prediction results tend to have large errors.                                                      |
| BP neural network model    | A computing system is established by simulating human brain neural tissue, and a large number of processing units are widely interconnected to form a network system. | There is no need to make model assumptions, just input historical data, since it has the basic characteristics of biological nervous system, it can learn data features and component models by algorithm. | The model has high self-learning and adaptive ability, strong generalization ability and high precision in the case of large amounts of data. | When the number of samples is not enough, the error of prediction is very large and the convergence speed of the algorithm is slow.               |

Each population prediction method has its own advantages and disadvantages, therefore it needs to make a choice according to the specific analysis of the problem. For example, although the prediction accuracy of the differential equation model is high, with the increase in variables, the equation will become more complicated and the difficulty of solving it will increase accordingly. The Leslie matrix model is more suitable for large time ranges and large time-span prediction problems. However, due to the different birth and death rates of different age groups, it is not suitable to analyze population growth under natural conditions, which limits the actual use of the Leslie matrix model. Although population development is a dynamic development process, the linear programming model is only suitable for the prediction of problems with a linear relationship, and the population problem inevitably has a nonlinear relationship, therefore the linear model has a limited role in population prediction. Although the grey prediction model reflects the regularity of data, it has the limitation that it cannot completely reflect the influence of various irregular social factors on the prediction index. The improvement of the grey prediction model, which not only gives full play to the regularity of data but also avoids the influence of irregular social factors on indicators, has become a research hotspot.

After professor Deng Julong first proposed the grey system theory in 1982, the grey system theory was gradually enriched and improved after 40 years of development [9]. As a new system analysis method to solve the problem of “few data, poor information and uncertain problem”, it has gradually become one of the main schemes to study uncertain systems. The inherent laws of the data sequence are fully mined to make scientific and
quantitative predictions, so as to reveal the future development trend of the system [10].

Since the grey prediction model is not strict about the sample size and the probability distribution of the original sequence, it has a unique advantage in the prediction of small sample data and effectively solves a large number of practical problems in production and life. At present, a large number of research results with practical value have been obtained in various dynamic systems such as public health [11,12], transportation [13–15], energy development [16–18], air quality [19,20] and so on. At the same time, many scholars have carried out abundant optimization studies on the traditional grey prediction model based on the characteristics of the grey prediction model [21], the construction of the initial value and background value of the model [22,23], the evolution of the accumulative generation operator [24,25], the expansion of the modeling equation [26], and the combined prediction with other models [27].

In the past, many scholars have undertaken a lot of exploration into the grey system theory. In particular, in the field of population prediction, most of the research is aimed at the traditional GM(1,1) model, which is formally expressed as a system with linear characteristics. However, in practical applications, most systems have mostly nonlinear characteristics and the dynamic changes of population time series data are the result of the synergistic effects of many factors, such as primary, secondary, direct, indirect, known, unknown, obvious and implicit, which are interrelated and mutually restrictive. These factors, which are partly known and partly unknown, interact with each other to determine the grey amount of reality—the total population. As a dynamic development process, population evolution inevitably has nonlinear characteristics, therefore only relying on a linear prediction model is not suitable for all systems. The Nonlinear Grey Bernoulli Model (referred to as NGBM), as a generalized nonlinear grey model, is an inevitable trend of research and development. Therefore, this paper considers using the NGBM model to model the population system, which is characterized by multi-factor control, complex evolution law, and uncertain data fluctuation, etc. Through a nonlinear model, the evolution law of the population system can be more effectively fitted.

As a general grey prediction model, the structure of the NGBM model changes with the power exponent. When the power exponent is zero, it is the traditional GM(1,1) model. When the power exponent is two, the model is a special grey Verhulst model [28]. Meanwhile, the power exponent of the NGBM model has the flexibility to change appropriately with the trend of the original sequence. When the value of the power exponent is appropriate, the NGBM model can achieve high prediction accuracy. Therefore, the NGBM model should have more ideal results than the traditional linear model to predict the future population. However, in practice, it is still found that the fitting and prediction effect of this model are unstable. In order to make the model capture the changing trend of different original behavior sequences more closely, make full use of the value of new information, and further improve the stability and computational efficiency of the model, a new accumulation operator can be applied in the NGBM model. A large number of studies were conducted on how to properly preprocess the original data and many cumulative operators suitable for different characteristics of the original data series were proposed, such as fractional order accumulation [29], complex order accumulation [30], seasonal accumulation [31,32], new information priority accumulation [33], etc.

It is mentioned in the grey system theory that if the original data have an exponential growth trend, they are more in line with the requirements of grey system modeling. The population can be regarded as a kind of data with exponential properties and saturated growth trends, therefore it is very suitable to be processed by the NGBM model. However, in the process of predicting this kind of data, the traditional accumulation operator will lead the prediction result to increase too much in the short term and there will be a large deviation between the prediction result and the reality. At the same time, owing to the traditional accumulation operator, the model cannot accurately control the mutation points in the data series, resulting in it not being able to accurately describe the disadvantages of the population change trend. By using new accumulation methods, such as fractional order
or complex order accumulation, the order of data accumulation can be extended from the integer range to the real or complex number field, which can greatly increase the prediction accuracy of the model and enhance the goodness of fit with the real data. However, these methods are too complex in their calculations, which is not conducive to application and theoretical analysis. Therefore, in order to avoid the phenomenon of exponential growth or rapid decline of data change in the process of prediction and make the prediction result conform to the law of population growth as much as possible, Liu Lianyi et al. proposed a kind of damping accumulated operator with a damping trend, which makes it easy to be concise and easy to understand and popularize \[34\]. By giving different weights to the information at different time points, the proportion of new information in the system is higher than that of old information, so as to highlight the role of new information in the sequence and effectively enhance the smoothness of the prediction data sequence and the stability of the prediction model. Through research, it was found that the combination of the damping accumulated operator and the traditional GM(1,1) model provides better performance compared with the traditional grey prediction model.

In this paper, the damping accumulated generation operator is combined with a nonlinear grey Bernoulli model to form the damping accumulated nonlinear grey Bernoulli model (referred to as DA–NGBM). On the basis of further improving the prediction accuracy and universality of the nonlinear model, the stability of model prediction and the reliability of prediction results are enhanced effectively. After fully studying the model, three cases related to the population size and structure of China further proved that the damping accumulated operator can effectively optimize the prediction efficiency of the original NGBM model and it can not only grasp the changes of the original sequence of different trends more accurately but is also suitable for the study of population direction prediction. Thus, it will really expand the application scope of the grey prediction model.

The rest of this article is organized as follows. In the second section, the traditional NGBM model and its solving process are reviewed, as well as the damped accumulated generating operator and related properties. In the third part, a new DA–NGBM model is constructed and its parameter estimation method, error reduction and perturbation boundary analysis are discussed. In the fourth part, three cases about the population size and structure of China are selected to verify the validity of the new model by comparing it with other traditional models. The fifth part makes a visual analysis of the testing results of the four models and discusses the performance of six kinds of prediction models for three actual cases, and then analyzes the prediction advantages of the DA–NGBM model in detail. Finally, the sixth part gives the conclusion of this paper and the prospects for future work.

2. The Definition of the Damping Accumulated Generating Operator
2.1. Traditional Grey Accumulation Generating Operator and Its Generating Sequence

The real data, to a great extent, reflect the actual situation of a certain problem in the objective world. Even if the type of data is in a discrete form or there is no law that can be found intuitively, as a direct manifestation of the behavior of the system, it must contain the deep evolution law of the system. In order to mine valuable and usable information from discrete and chaotic original data, the concepts of accumulation operator and accumulation generating sequence are proposed. By applying the summation operator on the original sequence items of the system, the intuitive data with irregular and large fluctuation can be transformed into usable data with obvious regularity and a linearly increasing relationship. Such operations facilitate the subsequent model building and the data have a powerful ability to display the actual behavior of the system, so as to more clearly highlight the evolution process and internal laws of the system.

**Definition 1 ([35]).** Suppose \( X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)) \) is the non-negative original data sequence, \( X^{(0)}D = (x^{(0)}(1)d, x^{(0)}(2)d, \ldots, x^{(0)}(n)d) \) is the first-order accumulation sequence of \( X^{(0)} \), \( D \) is the sequence operator, where
\[ x^{(0)}(k) = \sum_{i=1}^{k} x^{(0)}(i), \quad k = 1, 2, \ldots, n \]  
(1)

\[ D \text{ is the first-order cumulative generating operator of } X^{(0)}, \text{ and is denoted as 1-AGO.} \]

\[ x^{(r)}(k) = \sum_{i=1}^{k} x^{(r-1)}(i), \quad k = 1, 2, \ldots, n \]  
(2)

is called the r-order accumulation generated sequence of the original sequence \( X^{(0)} \).

The traditional grey accumulation generating operator can effectively improve the smoothness of the sequence, reduce the fluctuation among the discrete data, enhance the regularity of the data, and ensure the accuracy and stability of the grey prediction model fitting results and prediction results. At the same time, as the operator of equal weight addition, it fails to distinguish between new and old data sequences at each time point, and the data at each time point has the same proportion in the system, which will lead to a decrease in the fitting accuracy of the prediction model. In addition, assigning the old and new information with the same weight to the fitting sequence cannot highlight the “new information priority principle” required by the grey system theory. Moreover, the new information with closer time distance and more accurate and timely practical significance cannot be effectively played, which limits the prediction effect of the model to a certain extent.

**Definition 2 ([35]).** Suppose \( X^{(0)} \) is the original data sequence, \( X^{(1)} \) is the first-order accumulative generated sequence of \( X^{(0)} \), \( Z^{(1)} = \{z^{(1)}(1), z^{(1)}(2), \ldots, z^{(1)}(n)\} \) is the action sequence of the adjacent mean value operator of \( X^{(1)} \),

\[ x^{(0)}(k) + az^{(1)}(k) = b \left[ z^{(1)}(k) \right]^\alpha \]  
(3)

is called the nonlinear grey Bernoulli model, where \( a \) is the development coefficient, \( b \) is the grey action quantity, \( \alpha \) is the power exponent, and

\[ z^{(1)}(k) = \frac{1}{2}(x^{(1)}(k-1) + x^{(1)}(k)), \quad k = 2, 3, \ldots, n. \]

The solution process of the NGBM model is described as follows:

**Step 1.** Suppose \( X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)) \) is the non-negative original data sequence, \( X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)) \) is the first-order accumulative generated sequence of \( X^{(0)} \) and \( Z^{(1)} = (z^{(1)}(1), z^{(1)}(2), \ldots, z^{(1)}(n)) \) is the action sequence as described in definition 2 above.

**Step 2.** Suppose \( X^{(0)}, X^{(1)} \) and \( Z^{(1)} \) are as described in Step 1, the grey differential equation of the NGBM model is

\[ x^{(0)}(k) + az^{(1)}(k) = b(z^{(1)}(k))^\alpha, \]  
(4)

and the whitening differential equation can be expressed as

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

(5)

By solving the above differential equations, the time response of the NGBM model can be obtained as

\[ x^{(1)}(t+1) = \left\{ e^{(a-1)at} \cdot \left[ 1 - a \right] \int b \cdot e^{(1-a)at} dt + c \right\} \frac{1}{1-a} = \left( \frac{b}{a} + c \cdot e^{(1-a)at} \right) \frac{1}{1-a} \]

(6)
Step 3. The parameter sequence \( \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} \) of the model can be estimated by least squares based on the principle of minimum error sum of squares

\[
\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = (B^T B)^{-1} B^T Y.
\]

(7)

where \( B = \begin{bmatrix} -z^{(1)}(2) & (z^{(1)}(2))^a \\ -z^{(1)}(3) & (z^{(1)}(3))^a \\ \vdots & \vdots \\ -z^{(1)}(n) & (z^{(1)}(n))^a \end{bmatrix} \), \( Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \).

Step 4. By substituting the parameters \( \hat{a} \) and \( \hat{b} \) into the time response function (6) and setting \( \hat{x}^{(1)}(1) = x^{(0)}(1) \), the time response of the original sequence can be obtained as

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

where \( k = 1, 2, \ldots, n \), \( \hat{x}^{(1)}(k+1) \) is the fitting value of the \( k+1 \) moment, from which the sequence \( \hat{X}^{(1)} \) can be obtained.

Step 5. The sequence \( \hat{X}^{(1)} \) is reduced by the first order, from which the fitting sequence of the original data can be obtained as \( \hat{X}^{(0)} = \left\{ \hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \ldots, \hat{x}^{(0)}(n) \right\} \), and the predicted value of the model is \( \hat{x}^{(0)}(n + 1), \hat{x}^{(0)}(n + 2), \ldots, \), where

\[
\hat{x}^{(0)}(k + 1) = \hat{x}^{(1)}(k + 1) - \hat{x}^{(1)}(k), \quad k = 1, 2, \ldots \tag{9}
\]

2.2. Damping Accumulated Generating Operator and Its Generating Sequence

The traditional grey cumulative generating operator (abbreviated as AGO) can solve the problem that the data in the original behavior sequence are discrete and have no intuitive rules. However, since the cumulative operator gives the same weight to the behavioral data at each time point in the system, AGO cannot distinguish the difference between the old and new information. This will lead to a large deviation between the prediction model and the real data in the process of fitting. In order to solve these problems and further improve the prediction accuracy of the grey prediction model, the damping accumulated operator is specially introduced. By setting damping parameters, the weights of old and new information in the system can be distinguished, which makes the model perfectly conform to the "new information priority principle", meaning more satisfactory prediction results can be obtained.

**Definition 3 ([34]).** Suppose \( X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)) \) is the non-negative original data sequence, its \( \zeta \)-order damping accumulation sequence is defined as \( X^{(\zeta)} = (x^{(\zeta)}(1), x^{(\zeta)}(2), \ldots, x^{(\zeta)}(n)) \), where

\[
x^{(\zeta)}(k) = \sum_{i=1}^{k} \frac{x^{(0)}(i)}{\zeta^{i-1}}, \quad k = 1, 2, \ldots, n \tag{10}
\]

is called the \( \zeta \)-order damping summation operator of the sequence \( X^{(0)} \), the damping coefficient \( \zeta \in (0, 1) \).

Due to the \( x^{(\zeta)}(n) = \sum_{i=1}^{n} \frac{x^{(0)}(i)}{\zeta^{i-1}} = x^{(0)}(1) + \frac{x^{(0)}(2)}{\zeta} + \cdots + \frac{x^{(0)}(n)}{\zeta^{n-1}} \) matrix form of the sequence generated by \( \zeta \)-order damping accumulation can be expressed as
\[
\left( x^{(\zeta)}(1), x^{(\zeta)}(2), \cdots, x^{(\zeta)}(n) \right) = \left( x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n) \right) \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1/\zeta & \cdots & 1/\zeta \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\zeta^{n-1}} \end{pmatrix}
\] (11)

When \( \zeta = 1 \), it can be found that the \( \zeta \)-order damping accumulated operator is equivalent to 1-AGO and the \( \zeta \)-order damping accumulation sequence of the original sequence is equivalent to the traditional first-order accumulation generation sequence proposed in definition 1.

The damping accumulated operator can be used to optimize the traditional accumulation operator. Although 1-AGO makes the original discrete data with larger floating have a linear growth relationship easier to observe through data addition, the equal-weight addition of the data at each time in the original sequence ignores the difference of information between the old and new data and fails to make full use of the distance to study the practical value of the new data at a relatively recent time. Therefore, the damping accumulated operator is optimized in this aspect. By introducing the concept of damping parameters, the data information at each time point has its own weight so as to carry out differential treatment of the influence of the old and new information in the system. Therefore, the newer information in the system accounts for a larger proportion of influence, thus affecting the model fitting efficiency, and improving the prediction ability of the model.

**Property 1 ([34]).** Let \( x^{(\zeta)}(k) \) \((k = 1, 2, \cdots, n)\) be the \( \zeta \)-order damping accumulated operator, then the damping accumulation sequence \( X^{(\zeta)} = (x^{(\zeta)}(1), x^{(\zeta)}(2), \cdots, x^{(\zeta)}(n)) \) monotonically increases with respect to the time variable \( k \).

**Proof of Property 1.** For \( \forall k \in [1, n] \), we have
\[
x^{(\zeta)}(k+1) - x^{(\zeta)}(k) = \sum_{i=1}^{k+1} x^{(0)}(i) - \sum_{i=1}^{k} \frac{x^{(0)}(i)}{\zeta^{i-1}} = \frac{x^{(0)}(k+1)}{\zeta^k}
\]

Since \( x^{(0)} \) is a non-negative sequence, \( \zeta \in (0, 1] \), so \( x^{(\zeta)}(k+1) - x^{(\zeta)}(k) > 0 \) can be obtained. The proof is that the time variable is monotonically increasing. That is, \( X^{(\zeta)} \) is monotonically increasing with respect to the time variable \( k \). ☐

In the model, the more recent the research year, the more information contained in the new data, and the more research value. However, population systems generally have an evolutionary trend that cannot be mutated, and with the increase in time, the annual population generally shows a law of slow growth or decline. The connotation of this nature is basically consistent with the actual situation of the population system.

**Property 2 ([34]).** The value of \( x^{(\zeta)}(k) \) in the damping accumulation sequence \( X^{(\zeta)} \) is negatively correlated with the magnitude of the parameter \( \zeta \).

**Proof of Property 2.** For \( \forall \zeta_1, \zeta_2 \in (0, 1] \), set damping parameters \( \zeta_1 < \zeta_2 \), then \( \frac{1}{\zeta_1} > \frac{1}{\zeta_2} \), and \( \frac{1}{\zeta_1} > \frac{1}{\zeta_2^i}, i = 1, 2, \cdots, n \).

Due to the \( x^{(\zeta_1)}(k) - x^{(\zeta_2)}(k) = \sum_{i=1}^{k} x^{(0)}(i) (\frac{1}{\zeta_1^{i-1}} - \frac{1}{\zeta_2^{i-1}}) \), we can get to know \( x^{(\zeta_1)}(k) > x^{(\zeta_2)}(k), k = 1, 2, \cdots, n \).

That is, the value \( x^{(\zeta)}(k) \) is negatively correlated with the magnitude of the parameter \( \zeta \). ☐
By setting the value of damping parameters, the proportion of each item in the original sequence in the damping accumulation sequence can be controlled. It can be found that the equal weights of the original sequence items are no longer added in the damping accumulation term, and the smaller the value \( \zeta \) is, the greater the weight of the items in the original sequence in the new sequence will be, that is, it will occupy more proportion in the system research.

**Definition 4.** Suppose \( X^{(r)} = \{x^{(r)}(1), x^{(r)}(2), \ldots, x^{(r)}(n)\} \) is a non-negative time series, define

\[
\Delta(i) = \left| x^{(r)}(i + 1) - x^{(r)}(i) \right|
\]

(12)

is the information difference between \( x^{(r)}(i + 1) \) and \( x^{(r)}(i) \).

**Property 3.** Suppose \( X^{(\zeta)} \) is the damping accumulated with the original data to generate the sequence and the information difference \( \Delta(k) \) of the sequence \( X^{(\zeta)} \) is the increasing function of the time variable \( k \) and the decreasing function of the damping parameter \( \zeta \).

**Proof of Property 3.** According to the definition of information difference and damping accumulated generating operator,

\[
\Delta(k) = x^{(\zeta)}(k + 1) - x^{(\zeta)}(k) = \left( x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(k), x^{(0)}(k + 1) \right) \left( \begin{array}{c} 1 - 1 \\ \frac{1}{\zeta} - \frac{1}{\zeta^2} \\ \vdots \\ \frac{1}{\zeta^{k-1}} - \frac{1}{\zeta^k} \\ \frac{1}{\zeta^{k-1}} - 0 \end{array} \right) x^{(0)}(k + 1) \frac{1}{\zeta^k}.
\]

Then, the information difference \( \Delta(k) \) of \( X^{(\zeta)} \) can be expressed in matrix form as

\[
(\Delta(1), \Delta(2), \ldots, \Delta(k)) = \left( x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(k + 1) \right) \left( \begin{array}{ccc} 1 & 0 & \cdots & 0 \\ 0 & \frac{1}{\zeta} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\zeta^k} \end{array} \right).
\]

Obviously, when \( \zeta \in (0, 1) \), the information difference \( \Delta(k) \) of the accumulated damping sequence \( X^{(\zeta)} \) is the increasing function of the time variable \( k \) and the decreasing function of the damping parameter \( \zeta \). □

The larger the value of the time variable \( k \) is, the more recent the year represented by the data value is, which is consistent with the law of reality. This property can be expressed as the information difference between the newer data, which contains a more practical value and has more important research significance.

3. Damping Accumulated–Nonlinear Grey Bernoulli Model

3.1. Model Building

**Definition 5.** Suppose \( X(0) \) is original data sequence, \( X(\zeta) \) is the sequence generated by the damping accumulated of \( X(0) \), then

\[
x^{(\zeta)}(k) - x^{(\zeta)}(k - 1) + az^{(\zeta)}(k) = b \left[ z^{(\zeta)}(k) \right]^{a}, \quad k = 2, \ldots, n
\]

(13)

is called the damping accumulated–nonlinear grey Bernoulli model, where \( z^{(\zeta)}(k) = \frac{1}{2} \left( x^{(\zeta)}(k) + x^{(\zeta)}(k - 1) \right) \).
The modeling process of the DA–NGBM model is described as follows:

**Step 1.** Suppose \( X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)) \) is the non-negative sequence of the original data and the corresponding damped accumulative generating sequence \( X^{(g)} = (x^{(g)}(1), x^{(g)}(2), \ldots, x^{(g)}(n)) \) of the original data and the adjacent to the mean operator action sequence \( Z^{(g)} = (z^{(g)}(1), z^{(g)}(2), \ldots, z^{(g)}(n)) \) are defined in Section 2.

**Step 2.** Suppose \( X^{(0)}, X^{(g)} \) and \( Z^{(g)} \) are as described in Step 1, the grey differential equation of the DA–NGBM model is

\[
x^{(g)}(k) - x^{(g)}(k-1) + az^{(g)}(k) = b \left[ z^{(g)}(k) \right]^a
\]

where \( a \) is the development coefficient and \( b \) is the grey action quantity. The whitening differential equation can be expressed as

\[
\frac{dx^{(g)}(t)}{dt} + Qx^{(g)}(t) = b(x^{(g)}(t))^a
\]

After solving the above differential equation, the time response of the DA–NGBM model can be obtained as

\[
x^{(g)}(t+1) = \left( \frac{b}{a} + c \cdot e^{(a-1)t} \right)^{\frac{1}{a}}
\]

**Step 3.** The parameter sequence \([\hat{a}, \hat{b}]\) of the model can be estimated by least squares,

\[
[\hat{a}, \hat{b}] = \arg \min_{a,b} \sum_{k=1}^{n-1} \left[ (x^{(g)}(k+1) - x^{(g)}(k)) - \left( b \left[ z^{(g)}(k) \right]^2 - az^{(g)}(k) \right) \right]^2
\]

based on the principle of minimum sum of error squares, we can obtain

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

**Step 4.** Substituting parameters \( \hat{a} \) and \( \hat{b} \) into the time response Function (16) and letting \( \hat{x}^{(g)}(1) = x^{(0)}(1) \), the time response formula of the original sequence can be obtained as

\[
\hat{x}^{(g)}(k+1) = \left\{ \hat{b} + \left( \frac{b}{\hat{a}} \right)^{\frac{1}{a}} \left( x^{(0)}(1)^{1-a} - \frac{\hat{b}}{\hat{a}} \right) e^{-(1-a)kk} \right\}^{\frac{1}{a}}
\]

where \( k = 1, 2, \ldots, n \), \( \hat{x}^{(g)}(k+1) \) is the fitting value of the \( k+1 \) moment, thus obtaining the sequence \( \hat{X}^{(g)} \).

**Step 5.** After the first-order reduction of sequence \( \hat{X}^{(1)} \), the fitting sequence of the original data can be obtained as \( \hat{X}^{(0)} = \{ \hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \ldots, \hat{x}^{(0)}(n) \} \) and the predicted value of the model is \( \hat{x}^{(0)}(n+1), \hat{x}^{(0)}(n+2), \ldots \). The reductive formula of the model is

\[
\hat{x}^{(0)}(k) = \left\{ \begin{array}{ll}
\hat{x}^{(0)}(1) & , k = 1 \\
\left[ \hat{x}^{(g)}(k) - \hat{x}^{(g)}(k-1) \right] \delta^{k-1} & , k = 2, 3, \ldots
\end{array} \right.
\]

3.2. Parameter Estimation

Since the selection range of damping parameters is any real number within \((0, 1)\), it will consume a lot of resources and time to manually calculate the prediction accuracy.
of the model corresponding to each parameter value to determine the value of damping $\zeta$. For the indeterminate power exponent $\alpha$ in the DA–NGBM model, the optimal value still needs to be determined through a large number of calculations. Therefore, the estimation of an appropriate value can be optimized by using modern optimization algorithms. Considering the advantages of the heuristic algorithm, the whale optimization algorithm (WOA) currently has widely recognized computing power, therefore let the original sample data be divided into a training set and testing set, and the fitness function of the algorithm be set as the fitting prediction error of the DA–NGBM model. By finding the parameter corresponding to the global minimum of the objective function using the WOA algorithm, the damping parameter selection can be completed when the prediction model has the highest fitting accuracy.

A nonlinear optimization problem was established and the mean percentage error (MAPE) of the prediction model was used as the objective function

$$\min _{\zeta} MAPE = \frac{1}{n} \sum _{i=1}^{n} \left| \frac{x^{(0)}(i) - \hat{x}^{(0)}(i)}{x^{(0)}(i)} \right| \times 100\% (20)$$

The whale optimization algorithm is mainly inspired by the predation behavior of the humpback whale population and realizes the optimization search by simulating the scenes of encirclement, hunting and attacking the prey of the whale population. It assumes that the whale population size is $N$, the search space is $d$-dimensional, and the position of the $i$-th whale in $d$-dimensional space can be expressed as $x_i = (x_{i1}, x_{i2}, \cdots, x_{id}), i = 1, 2, \cdots, N$.

The location of the prey corresponds to the global optimal solution of the problem and there is no prior condition for the global optimal position in the search space before solving the optimization problem. Therefore, it is assumed that the position of the optimal individual in the current population is the prey position and other individuals are surrounded by the optimal individual. Based on the random walk mechanism of whale group hunting, whale individuals update the position of the next generation according to the location information of each other in the group. According to the search results of the $t$ time, they update the predation behavior of the $t + 1$ time by using Equations (21) and (22).

$$\vec{D} = |\vec{C} \cdot \vec{x}^*(t) - \vec{x}(t)| (21)$$

$$\vec{x}(t + 1) = \vec{x}^*(t) - \vec{A} \cdot \vec{D} (22)$$

where $\vec{D}$ is any position of whales in the hunting group, $\vec{x}^*(t)$ is the prey position and $\vec{A}$ and $\vec{C}$ are the coefficient vector, which is defined as follows:

$$\vec{A} = 2 \vec{a} \cdot \vec{r}_1 - \vec{a} (23)$$

$$\vec{C} = 2 \cdot \vec{r}_2 (24)$$

where $\vec{r}_1$ and $\vec{r}_2$ are random vectors between $[0, 1]$, and $\vec{a}$ is the convergence factor, which linearly decreases from the form of Equation (25) to 0 as the number of iterations increases.

$$\vec{a} = 2 - \frac{2t}{t_{\text{max}}} (25)$$

where $t_{\text{max}}$ is the maximum number of iterations. When $A > 1$, the whales will randomly search for food $X_{\text{rand}}$, that is, individual whales will randomly search each other based on their location.

The position $(X, Y)$ of a searching individual can be updated to the current optimal position $(X^*, Y^*)$ according to Equation (22) and the optimal position can be reached at different positions around the best individual by adjusting the vector values of $\vec{A}$ and
There are two different approaches to the bubble net feeding behavior of whales, the constriction encircling mechanism and the spiral replacement position. Thus, the mathematical model of whale bubble feeding behavior is as follows:

$$\vec{X}(t + 1) = \begin{cases} \vec{X}^*(t) - \vec{A} \cdot \vec{D}, & p < 0.5 \\ \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t), & p \geq 0.5 \end{cases}$$

(26)

where $D'$ is the distance between the $i$-th whale and the prey, $b$ is the constant used to define the logarithmic spiral shape, and $l$ is the random number between $[-1, 1]$. In the optimization process, the selection of the enveloping mechanism and spiral position update probability $p$ are both 0.5.

The modeling flow chart of the DA–NGBM model is shown in Figure 1.

---

**Figure 1.** Modeling flow chart of DA–NGBM model.

---

### 3.3. Research on the Properties of the Model

The DA–NGBM model has the following two basic properties:

**Property 4.** When $\zeta = 1$, the damped accumulation operator degenerates into the traditional first-order accumulation operator and the DA–NGBM model degenerates into the NGBM model.

**Property 5.** When $\zeta = 1$, $\alpha = 0$, the DA–NGBM model degenerates into the traditional GM(1,1) model.

#### 3.3.1. Reduction Error Analysis

**Theorem 1.** Suppose that the non-negative sequence of the original data is $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\}$ and the corresponding damped accumulation generation sequence of the original data is $X^{(\zeta)} = \{x^{(\zeta)}(1), x^{(\zeta)}(2), \ldots, x^{(\zeta)}(n)\}$. The response value
sequence of the DA–NGBM model is \( \mathbf{x}^{(t)} = \left\{ x^{(t)}(1), x^{(t)}(2), \ldots, x^{(t)}(n) \right\} \) and the prediction sequence of the model is \( \mathbf{x}^{(0)} = \left\{ x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n) \right\} \). If the reduction error of the response value of the DA–NGBM model meets \( |x^{(t)}(k) - x^{(t)}(k)| \leq \varepsilon \), then the reduction error of the model meets \( |\hat{x}^{(0)}(k) - x^{(0)}(k)| \leq 2\xi^{k-1}\varepsilon \).

**Proof of Theorem 1.** It can be seen from the modeling results of the DA–NGBM model that

\[
(\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \ldots, \hat{x}^{(0)}(n)) = (x^{(t)}(1), x^{(t)}(2), \ldots, x^{(t)}(n)) \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\xi^{k-1}} \end{pmatrix}^{-1},
\]

Then, the reduction value of the model can be obtained by reduction as

\[
\hat{x}^{(0)}(k) = (x^{(t)}(k) - x^{(t)}(k - 1))\xi^{k-1}.
\]

Since it is known that the reduction error of the response value of the DA–NGBM model meets \( |\hat{x}^{(t)}(k) - x^{(t)}(k)| \leq \varepsilon \), then

\[
|\hat{x}^{(t)}(k) - x^{(t)}(k)| = |(x^{(t)}(k) - x^{(t)}(k - 1))\xi^{k-1} - (x^{(t)}(k) - x^{(t)}(k - 1))\xi^{k-1}| \\
= |(x^{(t)}(k) - x^{(t)}(k))\xi^{k-1} - (x^{(t)}(k) - x^{(t)}(k - 1))\xi^{k-1}| \\
\leq |(x^{(t)}(k) - x^{(t)}(k))\xi^{k-1}| + |(x^{(t)}(k) - x^{(t)}(k - 1))\xi^{k-1}| \\
\leq 2\xi^{k-1}\varepsilon.
\]

That is, when the reduction error of the response value of the DA–NGBM model is \( |\hat{x}^{(t)}(k) - x^{(t)}(k)| \leq \varepsilon \), the reduction error of the model is \( |\hat{x}^{(0)}(k) - x^{(0)}(k)| \leq 2\xi^{k-1}\varepsilon \). □

**Property 6.** The upper limit of reduction error of the predicted value of the DA–NGBM model is positively correlated with the damping parameter \( \xi \) and negatively correlated with the time variable \( k \).

3.3.2. Analysis of Prediction Results

**Property 7.** The exponential growth rate of the predicted results of the DA–NGBM model is positively correlated with the value of the damping parameter \( \xi \).

**Proof of Property 7.**

\[
\hat{x}^{(0)}(k + 1) = (x^{(t)}(k + 1) - x^{(t)}(k))\xi^k \\
= \left\{ \left[ \frac{b}{a} + \left( \left( x^{(0)}(1) \right)^{1-a} - \frac{b}{a} \right) e^{-(a-k)} \right]^{\frac{1}{a}} - \left[ \frac{b}{a} + \left( \left( x^{(0)}(1) \right)^{1-a} - \frac{b}{a} \right) e^{-(a-k)} \right]^{\frac{1}{a}} \right\} \cdot \xi^k.
\]

Considering the complexity of simplified calculation, the detailed proof of the determined parameters is ignored. Then, the above equation can be simplified as
\[ s^{(0)}(k + 1) = \left\{ \left[ \frac{b}{a} + c \cdot e^{-(1-a)t} \right]^n - \left[ \frac{b}{a} + c \cdot e^{-(1-a)(k+1)} \right]^n \right\} \cdot \varepsilon^k \]

\[ = \left\{ \left[ C_e^0 \cdot \left( c \cdot e^{-(1-a)t} \right)^n \cdot \left( \frac{b}{a} \right)^n + C^i \cdot \left( c \cdot e^{-(1-a)t} \right)^n \cdot \left( \frac{b}{a} \right)^{n-1} + \ldots + C^n \cdot \left( c \cdot e^{-(1-a)t} \right)^n \cdot \left( \frac{b}{a} \right) \right] \cdot \varepsilon^k \right\} \cdot C_e^0 \cdot \left( c \cdot e^{-(1-a)t} \right)^n \cdot \left( \frac{b}{a} \right)^n \cdot \left( 1 - e^{-(1-a)t} \right)^i \cdot \left( 1 - e^{-(1-a)t} \right)^j \cdot \left( c \cdot e^{-(1-a)t} \right)^n \cdot \left( \frac{b}{a} \right) \]

where \( C_i = C_e^i \cdot c^i \cdot \left( \frac{b}{a} \right)^{n-i} \cdot \left( 1 - e^{-(1-a)t} \right)^i \cdot \left( 1 - e^{-(1-a)t} \right)^j \), \( n = \frac{1}{1 - a} \).

Consider the above equation as a function of the time variable, equivalent to \( f(t + 1) = \sum_{i=1}^{n} C_i \cdot \eta^i \). It can be seen that the prediction results of the DA–NGBM model still have exponential growth form. According to the properties of the exponential function, the growth rate of model prediction results is controlled by the base number \( \eta \). The larger \( \eta \) is, the larger the value of the damping parameter \( \zeta \) is and the larger the exponential growth rate of the model prediction results is. □

### 3.3.3. Disturbance Bound Analysis

**Lemma 1** ([24]). \( A \in C^{m \times n}, b \in C^m \). Suppose \( A^+ \) is the generalized inverse of the matrix \( A \), \( B = A + E, c = b + k \in C^m \). Suppose \( x + h \) and \( x \) are solutions of the linear least squares problem \( \| Bx - c \|_2 = \text{min} \) and \( \| Ax - b \|_2 = \text{min} \), respectively. If \( \text{rank}(A) = \text{rank}(B) = n \) and \( \| A^+ \|_2 \| E \|_2 < 1 \), then we have

\[ \| h \| \leq \frac{\kappa^+}{\gamma^+} \left( \left\| \frac{\| E \|_2}{\| A \|} \| x \| + \frac{\| k \|}{\| A \|} + \frac{\kappa^+}{\gamma^+} \frac{\| E \|_2}{\| A \|} \right\| \right) \]

(27)

where \( \kappa^+ = \| A^+ \|_2 \| A \| _2, \gamma^+ = 1 - \| A^+ \|_2 \| E \|_2, r_s = b - A x \).

**Theorem 2.** Suppose \( x \) is the solution of the DA–NGBM model satisfying the least square method \( \min ||Bx - y||_2 \). If \( \varepsilon = ||Ax| \) is the perturbation value of the initial sequence data \( x^{(0)}(i) \), \( (i = 1, 2, \ldots, n) \), the perturbation boundary of the solution of the model \( x \) increases with the increase in the data volume and decreases with the increase in the damping parameters.

**Proof of Theorem 2.** When only \( x^{(0)}(1) \) disturbs \( \varepsilon \), put \( \hat{x}^{(0)}(1) = x^{(0)}(1) + \varepsilon \) into Equation (11) and we can obtain

\[ \hat{B} + \Delta B = \begin{bmatrix} -z^{(\xi)}(2) & \left( z^{(\xi)}(2) \right)^a \\ -z^{(\xi)}(3) & \left( z^{(\xi)}(3) \right)^a \\ \vdots & \vdots \\ -z^{(\xi)}(n) & \left( z^{(\xi)}(n) \right)^a \end{bmatrix} + \begin{bmatrix} -\varepsilon & \varepsilon^a \\ -\varepsilon & \varepsilon^a \\ \vdots & \vdots \\ -\varepsilon & \varepsilon^a \end{bmatrix}, \hat{Y} + \Delta Y = \begin{bmatrix} x^{(\xi)}(2) - x^{(\xi)}(1) \\ x^{(\xi)}(3) - x^{(\xi)}(2) \\ \vdots \\ x^{(\xi)}(n) - x^{(\xi)}(n-1) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \]
Therefore, \( \| \Delta Y \|_2 = 0 \), \( \Delta B^T \Delta B = \begin{bmatrix} \frac{\sum_{i=1}^{n-1} \epsilon_{i}^2}{\gamma_{t} \| B \| \| B \|} & -\frac{\sum_{i=1}^{n-1} \epsilon_{i}^{a+1}}{\gamma_{t} \| B \| \| B \|} \\ -\frac{\sum_{i=1}^{n-1} \epsilon_{i}^{a+1}}{\gamma_{t} \| B \| \| B \|} & \frac{\sum_{i=1}^{n-1} \epsilon_{i}^{2a}}{\gamma_{t} \| B \| \| B \|} \end{bmatrix} \),

\( \| \Delta B \|_2 = \sqrt{(n-1)\epsilon_{2a}^2 + (n-1)\epsilon_{2}^2} \), according to Lemma 1, the disturbance bound of the model solution is

\[
L[x^{(0)}(1)] = \frac{\kappa_t}{\gamma_{t}} \sqrt{(n-1)\epsilon_{2a}^2 + (n-1)\epsilon_{2}^2} \left( \frac{1}{\| B \| \| B \|} \| x \| + \frac{\kappa_t}{\gamma_{t}} \frac{1}{\| B \| \| B \|} \| x_{n} \| \right).
\]

When only \( x^{(0)}(i) (i = 2, 3, \ldots, n-1) \) is disturbed \( \epsilon \),

\[
\| \Delta Y \|_2 = \frac{\epsilon}{\gamma_{t} \| B \| \| B \|},
\]

\[
\| \Delta B \|_2 = \sqrt{\frac{(\Delta B_{11} + \Delta B_{22}) + \sqrt{(\Delta B_{11} + \Delta B_{22})^2 - 4(n-i) \left( \frac{1}{\| B \| \| B \|} \| x \| + \frac{\kappa_t}{\gamma_{t}} \frac{1}{\| B \| \| B \|} \| x_{n} \| \right)^2}}{2}},
\]

where \( \Delta B_{11} = \left( \frac{1}{2} \cdot \frac{\epsilon}{\gamma_{t} \| B \| \| B \|} \right)^2 + (n-i) \left( \frac{\epsilon}{\gamma_{t} \| B \| \| B \|} \right)^2, \Delta B_{22} = \left( \frac{1}{2} \cdot \frac{\epsilon}{\gamma_{t} \| B \| \| B \|} \right)^2 + (n-i) \left( \frac{\epsilon}{\gamma_{t} \| B \| \| B \|} \right)^2 \).

By substituting the above results into Equation (27), the disturbance bound of the model solution can be obtained as

\[
L[x^{(0)}(i)] = \frac{\kappa_t}{\gamma_{t}} \left( \| \Delta B \|_2 \| x \| + \frac{\epsilon}{\| B \| \| B \|} \| x_{n} \| \right) \propto \sqrt{n}.
\]

Observe that only \( x^{(0)}(n) \) is disturbed \( \epsilon, x^{(0)}(n) = x^{(0)}(n) + \epsilon \),

\[
\| \Delta Y \|_2 = \frac{\epsilon}{\gamma_{t} \| B \| \| B \|},
\]

\[
\Delta B^T \Delta B = \begin{bmatrix} \frac{1}{2} \left( \frac{\epsilon}{\gamma_{t} \| B \| \| B \|} \right)^2 & \frac{\epsilon}{\gamma_{t} \| B \| \| B \|} \left( \frac{1}{2} \cdot \frac{\epsilon}{\gamma_{t} \| B \| \| B \|} \right)^{a+1} \\ \frac{\epsilon}{\gamma_{t} \| B \| \| B \|} \left( \frac{1}{2} \cdot \frac{\epsilon}{\gamma_{t} \| B \| \| B \|} \right)^{a+1} & \frac{1}{2} \left( \frac{\epsilon}{\gamma_{t} \| B \| \| B \|} \right)^{2a} \end{bmatrix},
\]

\[
\| \Delta B \|_2 = \sqrt{\frac{1}{2} \left( \frac{\epsilon}{\gamma_{t} \| B \| \| B \|} \right)^2 + \frac{1}{2} \left( \frac{\epsilon}{\gamma_{t} \| B \| \| B \|} \right)^{2a}}.
\]

According to Lemma 1, the disturbance bound of the model solution is

\[
L[x^{(0)}(n)] = \frac{\kappa_t}{\gamma_{t}} \left( \frac{1}{\| B \| \| B \|} \| x \| + \frac{1}{\| B \| \| B \|} \frac{\epsilon}{\gamma_{t} \| B \| \| B \|} \| x_{n} \| + \frac{\kappa_t}{\gamma_{t}} \frac{1}{\| B \| \| B \|} \| x_{n} \| \right),
\]

which conforms to the conclusion deduced above. To sum up, the disturbance bound of the solution of the model \( x \) increases with the increase in data quantity and decreases with the increase in the damping parameter value. \( \square \)

The matrix perturbation boundary theory proves that the essence of the grey prediction model of the DA–NGBM model remains unchanged despite the optimization of the original model. As a prediction model suitable for “small sample” data, the disturbance boundary of the DA–NGBM model increases with the increase in data sample size. To keep the information updated constantly and remove the old information in time, it is convenient to avoid the large disturbance of the predicted value of the model and also highlights that the new information has more important guiding significance for the prediction model. The selection of damping parameter value plays an important role in the stability of the predicted value of the model. Therefore, the influence of the change of damping parameter value on the prediction accuracy of the model should be fully considered in the prediction process.
3.4. Example Analysis

Based on the above research on the properties of the DA–NGBM model and the method of parameter selection, the following example is used to make a practical comparison between DA–NGBM and the classical population model—the Logistics model [36]. The number of the registered population in Shanghai from 2009 to 2018 is selected as the basic data. Through data fitting, the fitting value of the registered population in Shanghai from 2011 to 2018 is obtained and the number of registered population in Shanghai from 2019 to 2020 is predicted. The fitting error and prediction error are shown in Table 2. By calculating the MAPE value of the two prediction models, it can be found that the fitting accuracy and prediction accuracy of the DA–NGBM model were greatly improved compared with the logistic competitive biological model.

### Table 2. Comparison between Logistic model and DA–NGBM model of registered population of Shanghai.

| Year | Actual Data (Ten Thousand People) | Logistic Model | DA–NGBM Model |
|------|-----------------------------------|----------------|---------------|
|      | Fitting Data | MAPE (%) | Fitting Data | MAPE (%) |
| 2011 | 1419.36  | 1425.82  | 0.45     | 1419.57  | 0.01     |
| 2012 | 1426.93  | 1432.83  | 0.42     | 1426.10  | 0.06     |
| 2013 | 1432.34  | 1439.25  | 0.48     | 1432.29  | 0.00     |
| 2014 | 1438.69  | 1445.11  | 0.45     | 1438.29  | 0.03     |
| 2015 | 1442.97  | 1450.46  | 0.52     | 1444.18  | 0.08     |
| 2016 | 1450.00  | 1455.35  | 0.37     | 1449.99  | 0.00     |
| 2017 | 1455.13  | 1459.80  | 0.33     | 1455.74  | 0.04     |
| 2018 | 1462.38  | 1463.87  | 0.13     | 1461.46  | 0.06     |
|      | Simulation MAPE (%) | 0.39     |              | 0.04     |
| 2019 | 1469.30  | 1467.57  | 0.12     | 1468.84  | 0.15     |
| 2020 | 1475.63  | 1470.94  | 0.32     | 1475.29  | 0.19     |
|      | Prediction MAPE (%) | 0.22     |              | 0.17     |

4. Case Analysis of Population Prediction

The proposed DA–NGBM model provides a new idea for solving practical problems by using the grey prediction model. The behavior sequence of the model is constructed by using the damping accumulated method, which can not only satisfy the “new information first” criterion in grey modeling but also enhance the prediction accuracy and stability of the model to a certain extent. In order to verify the prediction effect of the DA–NGBM model after changing the accumulative method, three numerical calculation cases related to the prediction of population size and structure in China are selected in this section. Compared with the NGBM model, the original GM(1,1) model, the DA–GM (1,1) model [34], the deep learning model–LSTM [37] and the traditional statistical model–Exponential smoothing, the prediction ability of the improved DA–NGBM model was analyzed and evaluated.

The three cases selected in this paper contain ten groups of data and the first eight groups of data in the sample are used as the training set for the model. First, the eight data are used to fit the parameters of the model through the whale optimization algorithm. The whale optimization algorithm is used to find the optimal power exponent and the optimal damping parameters to minimize the fitting error of the model, and the optimal solution can be found by setting the searching times to 50 times. The last two groups of data in the sample are taken as the testing set for the model. On the premise of obtaining model parameters, the DA–NGBM model and the traditional grey prediction model are used for two-step prediction. The error analysis is made between the obtained prediction results and the data in the testing set to compare the prediction ability of the model. The specific division of sample data is shown in Figure 2.
4.1. China’s Birth Rate Forecast from 2011 to 2020

According to the birth rate data of China from 2011 to 2020 in the China Statistical Yearbook 2021 [38], the DA–NGBM model and traditional NGBM model, traditional GM(1,1) model, DA–GM(1,1) model, LSTM and Exponential smoothing are used to conduct comparative experiments to fit and predict the birth rate of the Chinese population, respectively. The birth rate from 2011 to 2018 was used to fit the model, the birth rate from 2019 to 2020 was used as the testing set, and the average absolute error percentage (MAPE) was used as the absolute standard to judge the quality of the prediction model and then the prediction accuracy of the four grey prediction models, one deep learning model and one traditional statistical model, was compared. It can be found that after adding the damping accumulated factor and intelligent optimization algorithm, the fitting and prediction accuracy of the DA–NGBM model is still greatly improved from the original model when the predicted data has obvious floating.

![Figure 2. Division of sample data.](image)

The comparison and prediction results of the six models are shown in Table 3. It can be found that the LSTM prediction model has the highest fitting accuracy after 5000 times of training. In terms of prediction, since this is a small sample problem, the LSTM model lacks sufficient training samples, therefore the prediction effect of it is not so good. To some extent, deep learning algorithms that require large sample data are not suitable for the prediction of small sample problems. The average fitting error of the DA–NGBM model is 4.63%, the prediction error for 2019 is 2.40%, and the prediction error for 2020 is 17.96%. China’s birth rate fluctuated due to the COVID-19 outbreak in 2020. However, under such data conditions with obvious fluctuations, the prediction effect of the DA–NGBM model is still optimal. It can be proved that by introducing the damping accumulated factor into the NGBM model, the prediction ability of the model is further improved and that this method is an effective improvement.

In this example, the optimal power index and damping parameter values of the DA–NGBM are calculated by MATLAB software and the model parameters obtained by least square are \( a = -0.2446 \), \( b = 10.5439 \), respectively.

| Model       | Training set (2011-2018) | Testing set (2019) | Forecasting data (2020) |
|-------------|---------------------------|--------------------|-------------------------|
| LSTM        | 11.87 (4.54%)             | 12.42 (4.70%)      | 12.93 (0.00%)           |
| Exponential | 11.87 (4.54%)             | 12.42 (4.70%)      | 12.93 (0.00%)           |
| GM(1,1)     | 11.87 (4.54%)             | 12.42 (4.70%)      | 12.93 (0.00%)           |
| DA–GM(1,1)  | 11.87 (4.54%)             | 12.42 (4.70%)      | 12.93 (0.00%)           |
| DA–NGBM     | 11.87 (4.54%)             | 12.42 (4.70%)      | 12.93 (0.00%)           |
| NGBM        | 11.87 (4.54%)             | 12.42 (4.70%)      | 12.93 (0.00%)           |

4.2. Prediction of Registered Population in Jiangsu Province from 2011 to 2020

According to the data of the registered total population in Jiangsu Province from 2011 to 2020 in the Jiangsu Statistical Yearbook 2021 [39], six kinds of prediction models are used for prediction, the total registered population of Jiangsu from 2011 to 2018 is taken as the fitting data, and the total registered population of Jiangsu from 2019 to 2020 is taken as the testing data. The comparison and prediction results of the six models are shown in Table 4. It can be seen that the total registered population in Jiangsu is a kind of data that has an overall trend of monotonically increasing. In the fitting calculation of this type of data series, the average fitting error of the DA–NGBM model is the first among six types of prediction models with a small error of 0.06%, which is further improved compared with the original NGBM’s average fitting error of 0.08%.

| Model       | Training set (2011-2018) | Testing set (2019) | Forecasting data (2020) |
|-------------|---------------------------|--------------------|-------------------------|
| LSTM        | 11.87 (4.54%)             | 12.42 (4.70%)      | 12.93 (0.00%)           |
| Exponential | 11.87 (4.54%)             | 12.42 (4.70%)      | 12.93 (0.00%)           |
| GM(1,1)     | 11.87 (4.54%)             | 12.42 (4.70%)      | 12.93 (0.00%)           |
| DA–GM(1,1)  | 11.87 (4.54%)             | 12.42 (4.70%)      | 12.93 (0.00%)           |
| DA–NGBM     | 11.87 (4.54%)             | 12.42 (4.70%)      | 12.93 (0.00%)           |
| NGBM        | 11.87 (4.54%)             | 12.42 (4.70%)      | 12.93 (0.00%)           |
Table 3. Comparison of fitting and predicted birth rate in China from 2011 to 2020.

| Year | Actual Data (%) | NGBM Model | DA–NGBM Model | GM (1,1) Model | DA–GM (1,1) Model | LSTM | Exponential Smoothing |
|------|-----------------|------------|---------------|---------------|------------------|------|-----------------------|
|      | Fitting Data (%) | MAPE (%)   | Fitting Data (%) | MAPE (%)   | Fitting Data (%) | MAPE (%)   | Fitting Data (%) | MAPE (%)   |
| 2011 | 13.27           | 0.00       | 13.27           | 0.00       | 13.27           | 0.00       | -          | -          |
| 2012 | 14.57           | 0.00       | 14.57           | 0.00       | 14.25           | 2.17       | -          | -          |
| 2013 | 13.03           | 5.07       | 14.33           | 9.98       | 13.79           | 5.81       | 14.00      | 7.44       |
| 2014 | 13.83           | 4.99       | 13.83           | 0.00       | 13.34           | 3.58       | 13.46      | 2.68       |
| 2015 | 11.99           | 6.16       | 13.22           | 10.26      | 12.93           | 7.84       | 11.99      | 0.99       |
| 2016 | 13.57           | 8.65       | 12.58           | 7.30       | 12.48           | 8.07       | 12.43      | 8.40       |
| 2017 | 12.64           | 4.16       | 11.93           | 5.62       | 12.07           | 4.54       | 11.95      | 5.46       |
| 2018 | 10.86           | 9.27       | 11.28           | 3.87       | 11.67           | 7.46       | 11.48      | 5.71       |

Simulation MAPE (%)  5.47  4.63  5.60  4.70  0.06  7.25

Table 4. Comparison of fitting and predicted values of registered population in Jiangsu Province from 2011 to 2020.

| Year | Actual Data (Ten Thousand People) | NGBM Model | DA–NGBM Model | GM(1,1) Model | DA–GM(1,1) Model | LSTM | Exponential Smoothing |
|------|-----------------------------------|------------|---------------|---------------|------------------|------|-----------------------|
|      | Fitting Data (%) | MAPE (%)   | Fitting Data (%) | MAPE (%)   | Fitting Data (%) | MAPE (%)   | Fitting Data (%) | MAPE (%)   |
| 2011 | 7514.25               | 0.00       | 7514.25       | 0.00       | 7514.25           | 0.00       | -          | -          |
| 2012 | 7553.48               | 0.00       | 7553.20       | 0.00       | 7574.25           | 0.28       | 7576.12    | 0.30       |
| 2013 | 7616.84               | 0.08       | 7620.62       | 0.05       | 7619.25           | 0.03       | 7619.24    | 0.03       |
| 2014 | 7684.69               | 0.11       | 7672.74       | 0.16       | 7664.52           | 0.26       | 7662.61    | 0.29       |
| 2015 | 7717.59               | 0.05       | 7717.60       | 0.00       | 7710.05           | 0.10       | 7706.22    | 0.15       |
| 2016 | 7775.66               | 0.18       | 7788.19       | 0.22       | 7755.86           | 0.25       | 7750.08    | 0.33       |
| 2017 | 7794.19               | 0.07       | 7796.00       | 0.02       | 7801.94           | 0.10       | 7794.19    | 0.00       |
| 2018 | 7831.86               | 0.05       | 7831.86       | 0.00       | 7848.29           | 0.21       | 7838.56    | 0.09       |

Simulation MAPE (%)  0.08  0.06  0.18  0.15  0.18  0.52

In the table, ‘-’ indicates that the value is missing in the calculation of LSTM model and the MAPE value in the table may have a slight error due to the reservation of decimals. The minimum error of the fitting results and the predicted results have been shown in bold and underlined.

For the prediction data, compared with the prediction results of 2019, the prediction error of the original NGBM model is 0.14%, the prediction error of the original GM(1,1) model is 0.47%, and the prediction error of the DA–GM(1,1) model is about 0.2% lower than the original model. The DA–NGBM model still ranks first among the four models with a prediction error of only 0.1%. Similarly, the prediction data of 2020 were observed and the prediction error of the DA–NGBM model was 0.29% as the optimal prediction model. The performance of the deep learning model and the traditional statistical method in this case also has a certain gap with that of the new grey forecasting method. Therefore, in this kind of monotonically increasing data prediction, the prediction ability of the DA–NGBM should be affirmed.

In this case, the optimal power index and damping parameter values of the DA–NGBM model $\alpha = 0.0114$, $\xi = 0.9458$ are calculated by MATLAB software and the model parameters obtained by least square are $a = -0.0590$, $b = 6573.73$, respectively.

4.3. Prediction of the Male Population of Nanjing City from 2011 to 2020

According to the data of the male population of Nanjing city from 2011 to 2020 in the Nanjing Statistical Yearbook 2021 [40], four kinds of grey prediction models, a
deep learning model and a traditional statistical method, are used for prediction. The total male population of Nanjing city from 2011 to 2018 is taken as the fitting data and the male population of Nanjing city from 2019 to 2020 is taken as the testing data. The comparison and prediction results of the six models are shown in Table 5. It can be seen that the total male population of Nanjing city is a kind of data with an overall trend that is monotonically increasing.

Table 5. Comparison of fitting and predicted values of the male population of Nanjing city from 2011 to 2020.

| Year | Actual Data (Ten Thousand People) | NGBM Model | DA–NGBM Model | GM(1,1) Model | DA–GM(1,1) Model | LSTM | Exponential Smoothing |
|------|----------------------------------|-------------|----------------|---------------|------------------|------|----------------------|
|      |                                  | Fitting Data MAPE (%) | Fitting Data MAPE (%) | Fitting Data MAPE (%) | Fitting Data MAPE (%) | Fitting Data MAPE (%) | Fitting Data MAPE (%) | Fitting Data MAPE (%) | Fitting Data MAPE (%) | Fitting Data MAPE (%) |
| 2011 | 320.90                           | 320.90 0.00 | 320.90 0.00 | 320.90 0.00 | 320.90 0.00 | - - | 320.90 0.00 |
| 2012 | 321.39                           | 321.39 0.00 | 321.39 0.00 | 318.16 1.00 | 318.16 1.00 | - - | 321.39 0.15 |
| 2013 | 322.90                           | 321.66 0.38 | 321.49 0.44 | 322.17 0.23 | 322.17 0.23 | 322.95 0.01 | 321.39 0.47 |
| 2014 | 325.04                           | 324.52 0.16 | 324.33 0.22 | 326.22 0.36 | 326.22 0.36 | 326.13 0.33 | 322.90 0.66 |
| 2015 | 326.78                           | 328.64 0.57 | 328.48 0.52 | 330.33 1.09 | 330.33 1.09 | 327.01 0.07 | 325.04 0.53 |
| 2016 | 330.86                           | 333.53 0.81 | 333.47 0.79 | 334.49 1.10 | 334.49 1.10 | 330.38 0.15 | 326.78 1.23 |
| 2017 | 339.27                           | 338.99 0.08 | 339.05 0.06 | 338.70 0.17 | 338.70 0.17 | 339.88 0.18 | 330.86 2.48 |
| 2018 | 346.85                           | 344.90 0.56 | 345.10 0.51 | 342.97 1.12 | 342.97 1.12 | 345.40 0.42 | 339.27 2.19 |

In the table, '-' indicates that the value is missing in the calculation of LSTM model and the MAPE value in the table may have a slight error due to the reservation of decimals. The minimum error of the fitting results and the predicted results have been shown in bold and underlined.

In the fitting calculation of this type of data series, the fitting effect of the LSTM method is the most accurate. However, the prediction effect of the LSTM model is worse than the DA–NGBM model in the testing stage and the prediction results of the DA–NGBM for two years are closest to the real value, with prediction errors of only 0.36% and 0.09%, which has the best prediction ability among the six predicted models. By comparison, the grey method shows superior ability in solving small sample prediction problems. It should be noted that the GM(1,1) model and the DA–GM(1,1) model obtained the same result in this case because the optimal value of the damping parameter of the DA–GM(1,1) model was selected as 1, therefore the DA–GM (1,1) model degenerated into the classical GM(1,1) model.

In this example, the optimal power index and damping parameter values of the DA–NGBM $\alpha = -0.0468$, $\zeta = 0.9677$ are calculated by MATLAB software and the model parameters obtained by least square are $a = -0.0558$, $b = 407.1498$, respectively.

5. Discussion

In order to show the fitting and prediction ability of the DA–NGBM model more vividly, this example is divided into two experimental areas. The birth rate of China from 2011 to 2018 is regarded as the fitting data and the birth rate from 2019 to 2020 is regarded as the testing data. The birth rate of China in 2021–2022 predicted by the six types of comparison models in this example is regarded as the forecasting data, and the fitting, testing and forecasting results of the six types of models are visualized by the MATLAB program, as shown in Figure 3.

The data, in this case, fluctuated greatly and showed a downward trend as a whole. It can be found that the GM(1,1) model and the NGBM model, as traditional grey prediction methods, can only match the real value at a few points and that these models do not have a strong perception ability for the changing trend of a data series. The optimized GM(1,1) model has a certain improvement in data fitting but there is still a certain deviation from the real value in the test results. The performance of the traditional statistical method, in this case, is even less than satisfactory. Their prediction results are difficult to correctly
correspond with population changes, the description of the changing trend is not clear, and there are large errors in the fitting and testing stages, therefore the accuracy of future prediction results may not be too high. From the results curve of the LSTM model, it can be found that in the fitting area, the curve of the LSTM model is almost exactly consistent with the real data. However, in the testing area, the prediction curve of the LSTM model is significantly different from the real value and shows an opposite trend of change. Therefore, it can be confirmed that the LSTM model still has excellent fitting ability in small sample problems but it is not completely suitable for the prediction problem of small sample cases.

Figure 3. Comparison of the fitting and forecast curves of China’s birth rate from 2011 to 2022.

It is not difficult to find that the DA–NGBM model performs better than the other three grey models in fitting and predicting the original behavior sequences with large fluctuations. It can also be seen that this model has good stability and will not affect the overall prediction result because of the anomaly of a single point. Not only in the fitting stage is the data curve smooth and conforms to the law of population growth, which is suitable for the prediction of population system, but also in the prediction process, the multi-point data can be consistent with the real value, and also in the test results, it is close to the real trend of population change.

The registered population of Jiangsu province from 2011 to 2018 is the fitting data and the registered population of Jiangsu Province from 2019 to 2020 is taken as the testing data. The registered population of Jiangsu Province from 2021 to 2022 is regarded as the forecasting data of the six prediction models. The fitting, testing and forecasting results of the six types of models are visualized by the MATLAB program. The result is shown in Figure 4.

The data fluctuation of this case is small and shows an overall upward trend. It can be found that the fitting and test coincidence of several types of grey models are quite ideal cause the original sequence changes regularly. The traditional statistical method can also describe the changing trend of the population but there is a large deviation from the actual value. The LSTM model can show more accurate results in the fitting and testing area but the predicted value in the prediction area has changed significantly and the change mode does not conform to the law of population change in reality. Although the results obtained by the four types of grey prediction models are satisfactory, careful observation shows that, compared with the GM(1,1) model, the NGBM model with nonlinear characteristics has better performance in grasping data changes. It is more consistent with the real data in the fitting stage and has significantly less errors in the test stage than the GM(1,1) model.
deviation between the predicted curve and the real curve is small. This also reflects that the linear model has certain disadvantages. In order to make use of the advantages of the nonlinear model in mastering data changes, this paper has optimized the NGBM model.

![Figure 4. Comparison of the fitting and prediction curves of registered population in Jiangsu Province from 2011 to 2022.](image)

After the improvement of the original model, the prediction curve of the DA–NGBM model coincides highly with the actual data curve in the fitting stage. Compared with the DA–GM (1,1), the NGBM model, LSTM and Exponential smoothing, the DA–NGBM model has obvious improvement and the best fitting ability. In the testing stage, the deviation between the predicted results and the real data is the smallest among the six models. This shows that the DA–NGBM model has high prediction accuracy and the optimization in the NGBM model has a good effect.

Next, we took the male population of Nanjing city from 2011 to 2018 as the fitting data and the male population of Nanjing city from 2019 to 2020 as the testing data. The male population of Nanjing city from 2021 to 2022 is taken as the forecasting data of the four types of grey prediction models, LSTM and Exponential smoothing. The fitting, testing and forecasting results of the six types of models are visualized through the MATLAB program. The result is shown in Figure 5.

Through the six images, in this case, the superior fitting and prediction ability of the DA–NGBM model can be demonstrated more clearly. It can be seen that although the NGBM model without optimization also has a good fitting ability, the prediction result of the DA–NGBM model is closer to the real value curve in the test stage. The predicted curves of the GM(1,1) model and the improved GM(1,1) model were quite different from the real value, especially in the testing phase, and could not accurately describe the changing trend of this population data. The LSTM model shows superior ability in the fitting region, but compared with the grey prediction model, the prediction deviation of the LSTM and the traditional statistical models is higher than the DA–NGBM and the results have less reliability.

Based on the above three cases, compared with the other five comparison models, the DA–NGBM model can deeply explore the fluctuation form of the population system and closely capture the overall evolution trend and individual change of this dynamic system, no matter if the original data have obvious fluctuation characteristics or the data series have monotonous growth (decline) trends. It is more reliable in fitting and prediction accuracy,
especially in its prediction effect. This work, therefore, has great practical significance in population prediction. Promoting a country’s effective grasp of the trend of population change can provide guiding opinions for the country to formulate relevant population policies and play an important role in strengthening social governance, therefore this issue is suitable for further study and application in population system prediction analysis.

![Figure 5](image-url) Comparison of the fitting and prediction curves of the male population of Nanjing city from 2011 to 2022.

After verifying the effectiveness of the DA–NGBM model, the DA–NGBM was used to predict the birth rate of China, the number of registered residents in Jiangsu Province, and the male population of Nanjing city from 2021 to 2015. The prediction results are shown in Figure 6.

![Figure 6](image-url) Projections of the size and structure of China’s population from 2021 to 2025.
6. Conclusions and Future Work

Correctly understanding and handling the relationship between population development and resources, environment and economy, and implementing the strategy of sustainable development, which is a major issue related to the rise and fall of a country and the survival of a nation, is also the eternal theme of mankind. As a direct means of adjusting and controlling population indicators, population decision-making plays an important role in economic and social development, and the accurate prediction of the population is directly related to the scientific decision-making of population issues. Therefore, selecting a reasonable and effective population prediction model, grasping the size of the population accurately, and understanding its development situation, have far-reaching significance for the formulation of the national economic plan and social development strategy.

6.1. Conclusions

Taking into account the importance attached to population prediction by countries around the world, the need for population prediction in formulating population policies and social development strategies, and the improvement of social welfare facilities in order to promote sustainable development of society, and in order to promote the sustainable development of society, perfect social welfare facilities, a kind of grey Bernoulli model with damping accumulation is established in this paper. As an optimization of the traditional grey prediction model, this model has an obvious improvement effect on the reasonable prediction of population system change. The NGBM model is a nonlinear grey model essentially, whose flexible power index enables it to have a broader ability to solve nonlinear feature problems and a strong universality compared with the traditional GM(1,1) model. In this paper, the damping accumulated operator is integrated into the data sequence construction of the NGBM model, and the intelligent advantage of the whale optimization algorithm is used to find the parameter value of the model in order to make the prediction model more sensitive to the changes of the original data and enhance the accuracy and stability of the prediction model, as well as reduce the complexity of calculating and use convenient research.

Combined with three cases of birth rate in China, the total registered population in Jiangsu Province and the male population of Nanjing city, this paper studies the DA–NGBM model compared with three kinds of grey prediction methods and two traditional statistical methods. The result shows that the new prediction model can fully adapt to the original series with different development trend characteristics and has the best fitting and prediction performance. The experimental results show that the predicted curves obtained by the DA–NGBM model are in better agreement with the real data curves in both the fitting and testing stages. At the same time, the addition of the damping accumulated operator can effectively adjust the exponential growth rate of the predicted results and also ensure the reliability and stability of the predicted results for the fluctuating original data series. As a result, the DA–NGBM model can effectively predict future population changes reliably. Therefore it is believed that the DA–NGBM model will have a broad research background and application space in population prediction and it is worth promoting to the application research of relevant population problems.

6.2. Future Work

As an important part of grey system theory, the grey prediction model reduces the randomness of the original sequence by cumulation operator and then introduces the data into the construction equation with biased and partial differential characteristics. Through this method, the inherent law of data series is fully mined, scientific and quantitative prediction is made, and the future development trend of the system is revealed. Grey system theory is a new system science method first proposed by Julong Deng in 1982 and has only been developed for 40 years. Therefore, some theories in this discipline are still not mature and many valuable theories have not yet been deeply explored. The DA–NGBM model, as a kind of grey prediction model, naturally has the advantages of grey system...
theory, but also there are some aspects to be improved. For example, the discussion of spatial stratified heterogeneity and temporal nonstationarity in grey forecasting models has not been presented. In the future, this issue will become a direction for further research and in-depth thinking.

In addition, in the case analysis of this paper, although the LSTM model shows excellent fitting ability, for data series with small or insufficient sample size, under the condition of low training times, the LSTM model does not have the same advantage as the grey model in predicting the ability of small samples. Deep learning models such as LSTM are generally considered suitable for dealing with problems that have big data characteristics, while the grey model in this paper is aimed at uncertain systems with small samples. Although more population information can be obtained from China’s long-term population series, such behavior is not appropriate due to China’s population data being easily affected by national policies, such as the family planning policy, the two-child policy or the three-child policy. Therefore, the longer the population series of China is, the less reliable it is. Additionally, deep learning models that require big data seem to be unable to deal with the problem of population prediction in China well. However, as a current research hotspot, the deep learning model deserves us to seriously think about its integration and development with grey system theory in future research and it is a worthy direction to combine grey prediction models with some deep learning models to obtain more advanced results. In addition, in order to better solve the grey multivariate problem, it is also a valuable thought to extend the new cumulative operator to the high-dimensional grey model.

Author Contributions: X.G. and Y.Y. conceived and designed the study; X.G. and R.Z. performed the modeling process and wrote the MATLAB code; R.Z. and H.S. performed the calculations and provided the recommendations. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Social Science Fund of China (grant no. 21BGL302).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets of this paper are available from the corresponding author upon reasonable request.

Acknowledgments: We would like to thank the anonymous referees for their constructive comments and suggestions that helped to improve the clarity and completeness of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Shi, H.P.; Ke, J.H.; Sun, C.; Lin, Z.Q. Rules of the population distribution of China and its evolution mechanism. Acta Phys. Sin. 2009, 58, 1–8.
2. Liang, Z.; Li, Z.; Ma, Z.D. Changing Patterns of the Floating Population in China, 2000–2010. Popul. Dev. Rev. 2014, 40, 695–716. [CrossRef] [PubMed]
3. Wu, Q.; Ye, D.Q.; Pan, H.F. The father of demography: John Graunt. Chin. J. Dis. Control. Prev. 2020, 24, 617–620. [CrossRef]
4. Allen, L.J. A density-dependent Leslie matrix model. Math. Biosci. 1986, 85, 179–187. [CrossRef]
5. Li, E.L. Analysis on the Influencing Factors of Aging Population in China. Aging Res. 2016, 32002, 11–18. [CrossRef]
6. Xie, N.M.; Wang, R.Z.; Chen, N.L. Measurement of shock effect following change of one-child policy based on grey forecasting approach. Kybernetes 2018, 47, 559–586. [CrossRef]
7. Khan, S.D.; Salih, Y.; Zafar, B.; Noorwali, A. A Deep-Fusion Network for Crowd Counting in High-Density Crowded Scenes. Int. J. Comput. Intell. Syst. 2021, 14, 168. [CrossRef]
8. Xu, L.; Magar, R.; Farimani, A.B. Forecasting COVID-19 new cases using deep learning methods. Comput. Biol. Med. 2022, 144, 105342. [CrossRef]
9. Deng, J.L. Control problems of grey systems. Syst. Control. Lett. 1982, 1, 288–294.
10. Liu, S.F.; Forrest, J.; Yang, Y.J. Advances in Grey Systems Research. J. Grey Syst. 2013, 25, 1–18.
11. Guo, X.J.; Shen, H.X.; Liu, S.F.; Xie, N.M.; Yang, Y.J.; Jin, J.L. Predicting the trend of infectious diseases using grey self-memory system model: A case study of the incidence of tuberculosis. Public Health 2021, 201, 108–114. [CrossRef] [PubMed]
12. Pan, H.; Hu, J.Y.; Wu, H.Y.; Tao, F.F.; Sun, X.D.; Zheng, Y.X.; Li, C.S.; Zheng, Y. Comparative study on the application of GM (1,1) grey model and ARIMA model in the incidence prediction of hand-foot-mouth disease in Shanghai. *Chin. J. Dis. Control Prev.* 2011, 5, 445–448.

13. Ding, S.; Li, R.J. Forecasting the sales and stock of electric vehicles using a novel self-adaptive optimized grey model. *Eng. Appl. Artificial Intell.* 2021, 100, 104148. [CrossRef]

14. Alkharbasheh, A.; Moslem, S.; Oubahman, L.; Duleba, S. An Integrated Approach of Multi-Criteria Decision-Making and Grey Theory for Evaluating Urban Public Transportation Systems. *Sustainability* 2021, 13, 2740. [CrossRef]

15. Xie, M.; Wu, L.F.; Li, B.; Li, Z.C. A novel hybrid multivariate nonlinear grey model for forecasting the traffic-related emissions. *Appl. Math. Model.* 2020, 77, 1242–1254. [CrossRef]

16. Xu, Q.W.; Xu, K.L. Mine safety assessment using grey relational analysis and bow tie model. *PLoS ONE* 2018, 13, e0193576.

17. Li, C.H.; Xu, W. Grey Multi-Level Risk Evaluation of China’s Oil Import. In Proceedings of the 2013 International Conference on Management Science and Engineering (ICMSE), Harbin, China, 17–19 July 2013.

18. Li, H.; Li, B.L.; Qin, J. Dynamic overcapacity modelling of early warning system for phosphorus fertilizer industry in China. In Proceedings of the 2020 International Conference on Energy, Environment and Bioengineering, Xi’an, China, 7–9 August 2020; Volume 185.

19. Wang, J.Z.; Du, P. Quarterly PM2.5 prediction using a novel seasonal grey model and its further application in health effects and economic loss assessment: Evidences from Shanghai and Tianjin, China. *Nat. Hazards* 2021, 107, 889–909. [CrossRef]

20. Wu, L.F.; Xu, Z.C. Analyzing the air quality of Beijing, Tianjin, and Shijiazhuang using grey Verhulst model. *Air Qual. Atmos. Health* 2019, 12, 1419–1426. [CrossRef]

21. Xie, N.M.; Liu, S.F. Discrete grey forecasting model and its optimization. *Appl. Math. Model.* 2009, 33, 1173–1186. [CrossRef]

22. Yang, X.L.; Zhou, M.; Zeng, B. A new method of background value construction for grey prediction model. *Stat. Decis.* 2018, 34, 14–18.

23. Yao, T.X.; Liu, S.F.; Dang, Y.G. Discrete grey prediction model based on optimized initial value. *Syst. Eng. Electron.* 2009, 31, 2394–2398.

24. Wu, L.F.; Liu, S.F.; Yao, L.G.; Yan, S.L.; Liu, D.L. Grey system model with the fractional order accumulation. *Commun. Nonlinear Sci. Numer. Simul.* 2013, 18, 1775–1785. [CrossRef]

25. Wu, L.F.; Liu, S.F.; Cui, W.; Liu, D.L.; Yao, T.X. Non-homogenous discrete grey model with fractional-order accumulation. *Neural Comput. Appl.* 2014, 25, 1215–1221. [CrossRef]

26. Ding, S.; Dang, Y.G.; Xu, N.; Wei, L.; Ye, J. Multi-variable time-delayed discrete grey model. *Control Decis.* 2017, 32, 1997–2004.

27. Guo, X.J.; Liu, S.F.; Wu, L.F.; Yang, Y.J.; Yang, Y.J. A multi-variable grey model with a self-memory component and its application on engineering prediction. *Eng. Appl. Artif. Intell.* 2015, 42, 92–93. [CrossRef]

28. Wang, Z.X. GM(1, 1) power model with time-varying parameters and its application. *Control Decis.* 2014, 29, 1828–1832.

29. Xie, W.L.; Liu, C.; Wu, W.Z. The fractional non-equidistant grey opposite-direction model with time-varying characteristics. *Soft Computing.* 2020, 24, 6603–6612. [CrossRef]

30. Wu, Z.P.; Chen, J.K.; Chai, J.P.; Zhang, F.Y. Grey System Model with Complex Order Accumulation. *J. Grey Syst.* 2021, 33, 98–117.

31. Xiao, X.P.; Yang, J.W.; Mao, S.H.; Wen, J.H. An improved seasonal rolling grey forecasting model using a cycle truncation accumulated generating operation for traffic flow. *Appl. Math. Model.* 2017, 51, 386–404. [CrossRef]

32. Zhou, W.J.; Ding, S. A novel discrete grey seasonal model and its applications. *Commun. Nonlinear Sci. Numer. Simul.* 2020, 93, 105493. [CrossRef]

33. Xia, J.; Ma, X.; Wu, W.Q.; Huang, B.L.; Li, W.P. Application of a new information priority accumulated grey model with time power to predict short-term wind turbine capacity. *J. Clean. Prod.* 2020, 244, 118573. [CrossRef]

34. Liu, L.Y.; Chen, Y.; Wu, L.F. The damping accumulated grey model and its application. *Commun. Nonlinear Sci. Numer. Simul.* 2021, 95, 105665. [CrossRef]

35. Liu, S.F.; Lin, Y. *Grey Systems: Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2010.

36. Chen, L.X.; Mu, T.Y.; Li, X.T.; Dong, J.C. Population Prediction of Chinese Prefecture-Level Cities Based on Multiple Models. *Sustainability* 2022, 14, 8. [CrossRef]

37. Khan, S.D.; Alarabi, L.; Basalamah, S. Toward Smart Lockdown: A Novel Approach for COVID-19 Hotspots Prediction Using a Deep Hybrid Neural Network. *Computers* 2020, 9, 99. [CrossRef]

38. National Bureau of Statistics China. *China Statistical Yearbook 2021*; China Statistics Press: Beijing, China, 2022.

39. Jiangsu Provincial Bureau of Statistics. *Jiangsu Statistical Yearbook 2021*; China Statistics Press: Beijing, China, 2021.

40. Nanjing Statistical Bureau. *Nanjing Statistical Yearbook 2021*; China Statistics Press: Beijing, China, 2022.