frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 28 August 2024
DOI 10.3389/fenvs.2024.1450354

OPEN ACCESS

EDITED BY
Tongwen Li,
Sun Yat-sen University, China

REVIEWED BY
Jamshid Piri,
Zabol University, Iran
Weifeng Gong,
Qufu Normal University, China

*CORRESPONDENCE
Ming Zhang,
yymingzhang2024@163.com

RECEIVED 18 June 2024
ACCEPTED 13 August 2024
PUBLISHED 28 August 2024

CITATION
Jiang J, Ban Y, Zhang M and Huang Z (2024) New discrete fractional accumulation Grey Gompertz model for predicting carbon dioxide emissions.
Front. Environ. Sci. 12:1450354.
doi: 10.3389/fenvs.2024.1450354

New discrete fractional accumulation Grey Gompertz model for predicting carbon dioxide emissions

Jianming Jiang[1], Yandong Ban[1], Ming Zhang[2]* and Zhongyong Huang[3]

[1]School of Public Health and Management Youjiang Medical University for Nationalities, Baise, China, [2]Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China, [3]College of science, Guangxi University of Science and Technology, Liuzhou, China

Predicting carbon dioxide emissions is crucial for addressing climate change and achieving environmental sustainability. Accurate emission forecasts provide policymakers with a basis for evaluating the effectiveness of policies, facilitating the design and implementation of emission reduction strategies, and helping businesses adjust their operations to adapt to market changes. Various methods, such as statistical models, machine learning, and grey prediction models, have been widely used in carbon dioxide emission prediction. However, existing research often lacks comparative analysis with other forecasting techniques. This paper constructs a new Discrete Fractional Accumulation Grey Gompertz Model (DFAGGM(1,1) based on grey system theory and provides a detailed solution process. The Whale Optimization Algorithm (WOA) is used to find the hyperparameters in the model. By comparing it with five benchmark models, the effectiveness of DFAGGM(1,1) in predicting carbon dioxide emissions data for China and the United States is validated.

KEYWORDS

carbon dioxide emissions forecasting, grey prediction model, DFAGGM(1,1) model, whale optimization algorithm, environmental sustainability

1 Introduction

1.1 Background and motivation

Predicting Carbon dioxide emissions levels is of utmost importance due to its profound implications for climate and environmental sustainability. Carbon dioxide is a primary greenhouse gas responsible for trapping heat in the Earth atmosphere, leading to global warming and climate disruption. Understanding and forecasting carbon dioxide concentrations are critical for several reasons. Firstly, accurate carbon dioxide predictions inform climate policymakers and governments about the effectiveness of current mitigation strategies and the urgency of implementing furthe measures to curb emissions. By forecasting future carbon dioxide levels, policymakers can design more effective policies aimed at reducing green house gas emissions, transitioning to renewable energy sources,and Promoting sustainable practices. Carbon dioxide forecasts are crucial for industries and businesses, particularly those reliant on fossil fuels and heavy carbon emissions. Anticipating future regulatory changes and market shifts related to carbon pricing and emissions trading enables companies to adapt their operations, invest in cleaner

technologies, and remain competitive in a rapidly evolving global economy. Thirdly, predicting carbon dioxide levels helps scientists and researchers better understand the complex dynamics of the Earth's climate system and assess the potential impacts of climate change on ecosystems, biodiversity, and human societies. This knowledge is essential for developing adaptation strategies to mitigate the adverse effects of climate change and safeguard vulnerable populations and ecosystems. Moreover, carbon dioxide forecasts play a vital role in international climate negotiations and agreements, providing countries with the necessary data to set emission reduction targets, monitor progress, and hold each other accountable for their commitments under global accords such as the Paris Agreement. In summary, forecasting carbon dioxide levels is indispensable for addressing climate change challenges effectively, guiding policy interventions, fostering sustainable development, and safeguarding the planet for future generations.

1.2 Literature review

To accurately predict carbon dioxide emissions, scholars have conducted related research using various prediction methods, mainly including statistical model predictions, machine learning predictions, and grey model predictions. Traditional statistical models primarily focus on time series models (Karakurt and Aydin, 2023; Javanmard et al., 2023; Ding and Zhang, 2023; Rao et al., 2023). For example, Cui et al. presents a SARIMA-based federated learning method to enhance privacy, accuracy, and efficiency in predicting industrial carbon emissions. By clustering clients and applying federated averaging, the method significantly improves prediction accuracy and convergence speed while protecting data privacy, offering a robust approach for utilizing multisource industrial data in environmental management (Cui et al., 2023). Kour M et al. examines the rise of carbon dioxide emissions in South Africa, a significant contributor to global greenhouse gases and ecological imbalance. Using annual data from 1980 to 2016, the researchers apply an ARIMA model to forecast carbon dioxide emissions from 2015 to 2027. The model predicts a steady increase in emissions over the next decade (Kour, 2023).

Due to the flexibility and adaptability of machine learning methods, as well as their ability to handle complex non-linear relationships, they have been widely applied in the field of carbon dioxide emission prediction (Bhatt et al., 2023; AlOmar et al., 2023; Khajavi and Rastgoo, 2023; Lin et al., 2021). For example, Qin et al. explores methods to calculate and reduce urban carbon dioxide emissions in China. Using data from 2000 to 2019, an inversion model for prefecture-level cities is created. Machine learning identifies GDP, financial revenue, and foreign investment as key emission factors. The study highlights the highest emissions in economically developed and resource-dependent regions and stresses the importance of continued low-carbon initiatives and focusing on high-emission areas for sustainable development (Qin and Gong, 2022). Zhang et al. examines carbon emissions in China's building and construction sector from 2005 to 2021, focusing on 30 provincial regions. It develops nine machine learning regression models to predict emission trends, using various economic, technological, and classification factors. The stacking ensemble regression model proved to be the most effective (Zhang et al., 2024). Faruque et al. believed that univariate predictions are difficult to capture sudden changes in time series, so they explored multivariate experiments based on four deep learning algorithms (Convolutional Neural Network (CNN), CNN Long Short-Term Memory (CNN-LSTM), Long Short-Term Memory (LSTM), and Dense Neural Network (DNN)) and multiple linear regression. The experimental results showed that DNN and multiple linear regression achieved first and second performance rankings with MAPE values of 3.678 and 5.541, respectively (Faruque et al., 2022).

Deng proposed the grey system theory in 1982, primarily to solve the problem of analysis, prediction, decision, and control of small data, poor information uncertain systems (JuLong, 1982). Guo et al. designed a grey prediction model based on exponential accumulation and applied it to the carbon dioxide emissions of the BRICS countries (Guo et al., 2021). Although the experiments showed that the model achieved good prediction performance, they did not consider a comparison with existing carbon dioxide emission models in their analysis. Therefore, the experiments lack persuasiveness. Qiao et al. designed a fractional-order grey GM(1,1) model and achieved the best performance in predicting carbon dioxide emissions in APEC member countries (Qiao et al., 2021). However, similar to the studies by Qiao et al. and Guo et al., the effectiveness of this method was not validated through comparative analysis with other carbon dioxide emission prediction techniques. Zeng et al. designed a fractional-order grey Verhulst model for predicting China's annual carbon dioxide emissions and achieved the best test performance on the dataset of China's annual carbon dioxide emissions (Gao et al., 2022). The aforementioned studies on carbon dioxide emission prediction hold significant promise for global carbon reduction efforts. Nevertheless, most of these studies require a substantial amount of data. Since policy interventions influence carbon dioxide emissions, the value of historical data for future forecasting is considerably diminished. Following the principle that new information should take precedence over old, it is more reasonable to use data from recent years. This results in a smaller sample size. Compared to mathematical statistics methods and big data technologies, grey prediction has greater advantages in handling systems characterized by "small data, poor information" uncertainty. Moreover, the carbon dioxide emissions system exhibits both regularity and randomness, typifying grey-cause, white-effect" uncertainty characteristics.

Reading through these studies, it becomes clear that constructing a reliable model for predicting annual carbon dioxide emissions is of paramount importance. While various methodologies, such as statistical model predictions, grey prediction models, deep learning techniques,have shown promise, a common shortcoming is the lack of comparative analysis with other prediction techniques. Accurate prediction models are crucial for effective policy-making, environmental planning, and achieving global climate goals. By addressing the current research gaps, particularly through robust comparative analyses and the integration of hyperparameters, future models can significantly enhance the accuracy and reliability of carbon dioxide emission predictions.

Based on the GM model, this paper constructs a New discrete Fractional Accumulation Grey Gompertz model (DFAGGM(1,1)),

provides the specific solution process for the model, and applies it to the prediction of carbon dioxide emissions in China and the United States. The subsequent sections of this paper are structured as follows: The section 2 mainly introduces the construction and solution process of the DFAGGM(1,1) model, the section3 uses the DFAGGM(1,1) model to predict and analyze carbon dioxide emissions data for China and the United States. The conclusion and summary are provided at the end.

2 Methods

2.1 Grey gompertz model

The Fractional Accumulation Grey Gompertz model (FAGGM(1,1)) is a grey forecasting technique proposed by Gao et al. (Zeng et al., 2023). For predicting carbon dioxide emissions. Its expression

$$\frac{dx^{(r)}(t)}{dt} + a \cdot x^{(r)}(t) = b \cdot x^{(r)}(t) \cdot ln\left(x^{(r)}(t)\right) \tag{1}$$

where a is the development coefficient, and $x^{(r)}(t)$ belongs to a primitive non-negative time series $\{x^{(0)}(1), \ldots, x^{(0)}(n)\}$'s r-th order cumulative generation sequence.

The parameter estimation values of FAGGM(1,1) are provided by

$$\hat{p} = \left(\hat{a}, \hat{b}\right)^T = \left(B^T B\right)^{-1} B^T Y, \tag{2}$$

The definitions of B and Y in Equation 2 are as follows: $Y = \begin{bmatrix} x^{(r-1)}(2) \\ x^{(r-1)}(3) \\ \vdots \\ x^{(r-1)}(n) \end{bmatrix}$ and $B = \begin{bmatrix} -z^{(r)}(2) & z^{(r)}(2)ln(z^{(r)}(2)) \\ -z^{(r)}(3) & z^{(r)}(3)ln(z^{(r)}(3)) \\ \vdots & \vdots \\ -z^{(r)}(n) & z^{(r)}(n)ln(z^{(r)}(n)) \end{bmatrix}$,

Respectively. $z^{(r)}(t) = o \cdot x^{(r)}(t) + (1-o) \cdot x^{(r)}(t-1), \quad t = 1, 2, \ldots, n.$

The predicted values of FAGGM(1,1) are given by

$$\hat{x}^{(0)}(t) = \left(\hat{x}^{(r)}(t)\right)^{(-r)}, \quad t = 1, 2, \ldots, n, \tag{3}$$

where solving equation 1 yields the following Equation 4

$$\hat{x}^{(r)}(t) = exp\left(\left(ln\left(x^{(0)}(1)\right) - \frac{\hat{a}}{\hat{b}}\right)e^{\hat{b}(t-1)} + \frac{\hat{a}}{\hat{b}}\right), \quad t = 1, 2, \ldots, n. \tag{4}$$

Based on the modeling process of FAGGM(1,1), the following points can be understood.

(1) The parameter estimation formula of FAGGM(1,1) contains errors, which may affect the results of the model. Typically, in grey modeling processes, the integral form of the model is used as the parameter estimation formula. However, FAGGM(1,1) employs $z^{(r)}(t)$ instead of the definite integral of the $x^{(r)}(t)$, which seems unreliable

(2) The parameter estimation formula of FAGGM(1,1) is discrete, while the model's prediction formula is continuous, causing inconsistency that affects the model's rationality.

(3) There are no hyperparameters in the model that can adjust the model structure, which may limit the performance of the model.

2.2 New discrete fractional accumulation grey gompertz model (DFAGGM(1,1))

By utilizing first-order forward differencing to amend Equation 1,

$$x^{(r)}(t+1) = a' \cdot x^{(r)}(t) + b \cdot x^{(r)}(t) \cdot ln\left(x^{(r)}(t)\right) \tag{5}$$

can be obtained, where $a' = (1-a)$.

By introducing a bias c and a hyperparameter α, Equation 5 becomes

$$x^{(r)}(t+1) = a' \cdot x^{(r)}(t) + b \cdot \left(x^{(r)}(t)\right)^{\alpha} \cdot ln\left(x^{(r)}(t)\right) + c. \tag{6}$$

Based on Equation 6, the estimated value of $\xi = [a', b, c]^T$ in DFAGGM(1,1) can be obtained through

$$\hat{\xi} = \left(k^T k\right)^{-1} k^T \tau, \tag{7}$$

The matrix in Equation 7 is defined as follows:

$$k = \begin{bmatrix} x^{(r)}(1) & (x^{(r)}(1))^{\alpha} \cdot ln(x^{(r)}(1)) & 1 \\ \vdots & \vdots & \vdots \\ x^{(r)}(n-1) & (x^{(r)}(n-1))^{\alpha} \cdot ln(x^{(r)}(n-1)) & 1 \end{bmatrix}, \tau = \begin{bmatrix} x^{(r)}(2) \\ \vdots \\ x^{(r)}(n) \end{bmatrix}.$$

The prediction results of DFAGGM(1,1) are also obtained through Equation 3. It should be noted that

$$\hat{x}^{(r)}(t+1) = \left(\hat{a}'\right)^t \cdot \hat{x}^{(r)}(1) + \sum_{i=0}^{t-1} \left(\hat{a}'\right)^i \cdot \left[b \cdot \left(\hat{x}^{(r)}(t-i)\right)^{\alpha} \cdot ln\left(\hat{x}^{(r)}(t-i)\right) + \hat{c}\right]. \tag{8}$$

2.3 The method for solving the model

Due to the presence of unknown hyperparameters in the DFAGGM(1,1) model, it is challenging to apply the model directly. To address such issues, constructing an optimization model is a common approach. In this paper, we formulate an optimization problem using the mean absolute percentage error as the loss function, namely, $\min_{\alpha,r} fitness = Mean\left(\left|\frac{\hat{x}-x}{x}\right|\right) \times 100\%$

$$s.t. \begin{cases} \alpha \in [-3, 3], r \in [0, 2] \\ \hat{\xi} = (k^T k)^{-1} k^T \tau \\ k = \begin{bmatrix} x^{(r)}(1) & (x^{(r)}(1))^{\alpha} \cdot ln(x^{(r)}(1)) & 1 \\ \vdots & \vdots & \vdots \\ x^{(r)}(n-1) & (x^{(r)}(n-1))^{\alpha} \cdot ln(x^{(r)}(n-1)) & 1 \end{bmatrix}, \tau = \begin{bmatrix} x^{(r)}(2) \\ \vdots \\ x^{(r)}(n) \end{bmatrix}. \end{cases}$$

$\hat{x}^{(r)}(t+1) = (\hat{a}')^t \cdot \hat{x}^{(r)}(1) + \sum_{i=0}^{t-1} (\hat{a}')^i \cdot [b \cdot (\hat{x}^{(r)}(t-i))^{\alpha} \cdot ln(\hat{x}^{(r)}(t-i)) + \hat{c}]$. $\hat{x}^{(0)}(t) = (\hat{x}^{(r)}(t))^{(-r)}, \quad t = 1, 2, \ldots, n$

For nonlinear normalization models, ordinary mathematical solving methods require a substantial amount of time. In prediction problem solving, quick execution of predictions is essential. Therefore, utilizing swarm intelligence algorithms is an effective method for addressing such issues. This paper employs the classical Whale optimization Algorithm (WOA) to solve for the

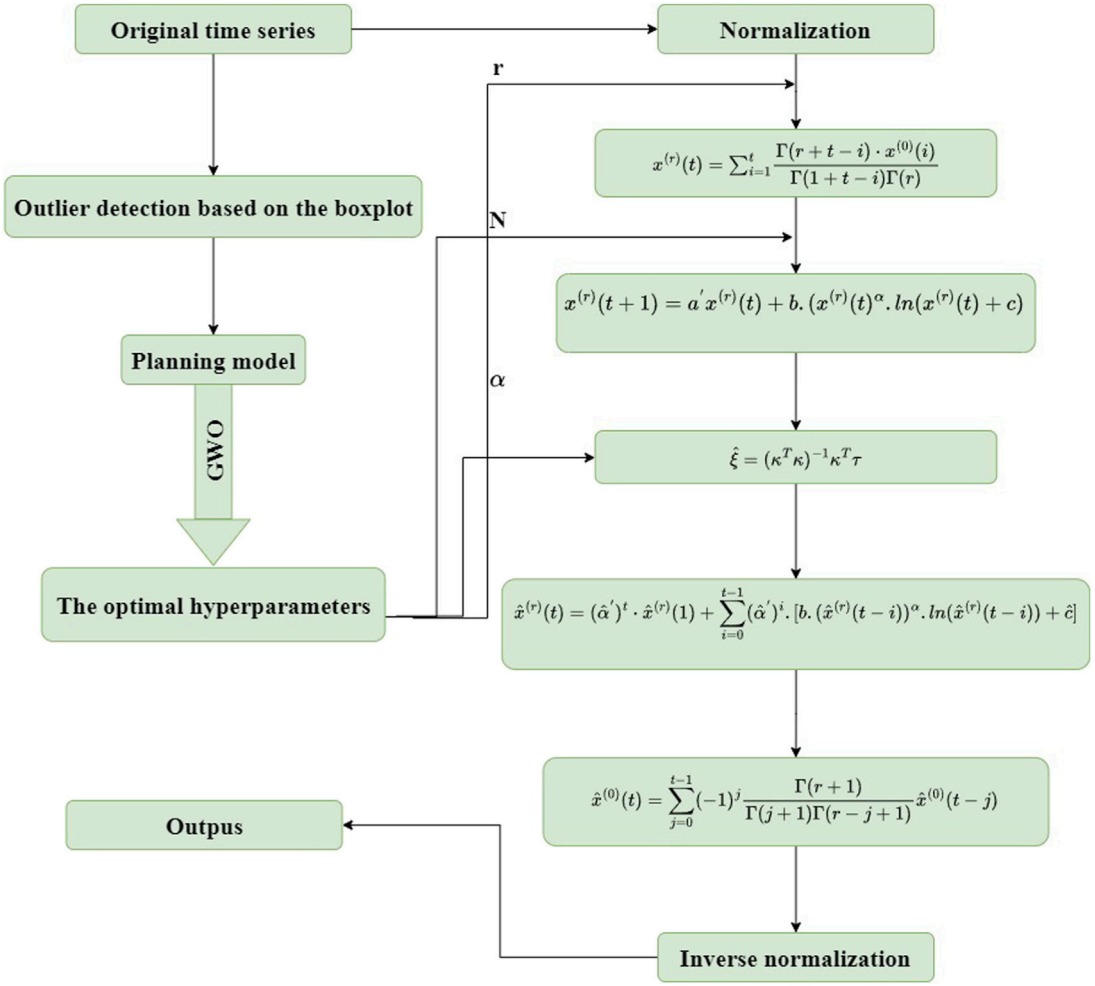

FIGURE 1
The application process of the proposed model.

optimal hyperparameters. The specific solving process of the WOA algorithm is as follow.

2.3.1 Initialize parameters

· Set algorithm parameters: This includes the number of whales N, maximum number of iterations T, dimension of the search spaced d,and the lower and upper bounds of the search space $[l_b, u_b]$.

· Initialize the positions of whale population: Randomly generate N solutions (whale positions) within the search space, denoted as $\vec{X}_i\,(i = 1, 2, \ldots, N)$.

2.3.2 Evaluate fitness

· Calculate the fitness value of each whale position to determine the best solution $\vec{X}^{\star}$.

2.3.3 Update positions

In each generation, update the whale positions using the following formulas:

· Encircling Prey.

When $p < 0.5$, whales simulate the behavior of encircling prey:

$\vec{D} = |\vec{C}\vec{X}^{\star} - \vec{X}(t)|, \vec{X}(t+1) = \vec{X}^{\star} - \vec{A}.\vec{D}$

Where.

$\vec{A} = 2\vec{a}.\vec{r} - \vec{a}, \vec{C} = 2.\vec{r}$

$\vec{a}$ linearly decreases from two to 0 over the course of iterations.

$\vec{r}$ is a random vector where each element is in the range [0,1] $\vec{X}^{\star}$ is the current best solution.

· Search for Prey.

When $p \geq 0.5$,whales simulate the search for prey:

$\vec{X}_{rand}$ is a randomly chosen position from the whale population.

$\vec{D} = |\vec{C}.\vec{X}_{rand} - \vec{X}(t)|, \vec{X}(t+1) = \vec{X}_{rand} - \vec{A}.\vec{D}$ ·

Spiral Update Position.

When $p < 0.5$, an alternative spiral update can be chosen:

$\vec{D}' = |\vec{X}^{\star} - \vec{X}(t)|$

$\vec{X}(t+1) = \vec{D}' cos(2\pi l) + \vec{X}^{\star}$

Where

b is a constant defining the shape of the logarithmic spiral

l is a random number in the range [-1,1]

2.3.4 Check constraints

After updating positions, check each solution to ensure it remains within the search space. If not, adjust the solution to bring it back within bounds.

TABLE 1 Carbon dioxide emissions from energy in China from 1993 to 2022.

Year	Data	Year	Data
1993	2769.6144	2008	7356.5599
1994	2921.3577	2009	7685.0202
1995	3009.1634	2010	8121.6879
1996	3156.6550	2011	8793.4911
1997	3142.3315	2012	8977.5597
1998	3139.8079	2013	9214.0861
1999	3268.4740	2014	9235.5298
2000	3327.9839	2015	9171.2830
2001	3489.7271	2016	9014.6386
2002	3809.2640	2017	9270.2826
2003	4494.1359	2018	9610.7037
2004	5317.1677	2019	9933.6878
2005	6079.2593	2020	10130.8681
2006	6659.9736	2021	10563.4736
2007	7217.1392	2022	10550.2477

2.3.5 Iterate

Repeat steps 2-4 until the maximum number of iterations TT is reached or a stopping criterion is met.

2.3.6 Output

Output the best solution $\vec{X}^*$ and its corresponding fitness value.

The solving process based on this algorithm is illustrated in Figure 1.

3 Application

3.1 Raw data collection and comparison model

The Carbon Dioxide Emissions from Energy in China from 1993 to 2022 and the energy-related carbon dioxide emissions from the total of states in the United States from 1970 to 2021 are taken as the subjects of the study in this article. These two datasets are sourced from https://www.energyinst.org/statistical-review and https://www.eia.gov/environment/emissions/state/, as shown in Table 1, 2. For each dataset, the last nine data points are used as the test set.

To validate the effectiveness of the models presented in this article, five recent carbon dioxide emission forecasting techniques were selected, including FAGGM (Gao et al., 2022), *FN_Verhulst* (Zeng et al., 2023), ELM (Rahman et al., 2024), FANGBM (Yang and Wu, 2023), CFDGM (Zhu et al., 2024) and ESN (Wang et al., 2018). It is particularly important to note that the FAGGM model is the conceptual origin of the model presented in this article. By performing a comparative analysis between the method proposed in this article and FAGGM, the rationality of the model presented in this article can be further verified.

3.2 Evaluation indices of the modelling accuracy

To evaluate the training and prediction performance of the model, this paper adopts five commonly used evaluation metrics, namely.

The mean absolute percentage error (MAPE), which is defined as follows

$$MAPE = \frac{1}{n-1}\sum_{i=2}^{n}\left|\frac{\hat{x}^{(0)}(t)-x^{(0)}(t)}{x^{(0)}(t)}\right| \times 100\%. \tag{9}$$

Mean Squared Error, MSE

$$MSE = \frac{1}{n-1}\sum_{i=2}^{n}\left(\hat{x}^{(0)}(t)-x^{(0)}(t)\right)^2. \tag{10}$$

Root Mean Squared Error, RMSE

$$RMSE = \sqrt{\frac{1}{n-1}\sum_{i=2}^{n}\left(\hat{x}^{(0)}(t)-x^{(0)}(t)\right)^2}. \tag{11}$$

Mean Absolute Error, MAE

$$MAE = \frac{1}{n-1}\sum_{i=2}^{n}\left|\hat{x}^{(0)}(t)-x^{(0)}(t)\right|. \tag{12}$$

Index of agreement (IA)

$$IA = 1-\frac{\sum_{i=1}^{n}\left(x^{(0)}(i)-\hat{x}^{(0)}(i)\right)^2}{\sum_{i=1}^{n}\left(\left|x^{(0)}(i)-\bar{x}\right|+\left|\hat{x}^{(0)}(i)-\bar{\hat{x}}\right|\right)^2}. \tag{13}$$

TABLE 2 Energy-related carbon dioxide emissions from the total of states in the United States from 1970 to 2021.

Year	Data	Year	Data	Year	Data	Year	Data
1970	4254.944	1983	4377.932	1996	5505.307	2009	5395.696
1971	4300.108	1984	4602.973	1997	5577.345	2010	5585.185
1972	4520.093	1985	4607.575	1998	5613.136	2011	5449.900
1973	4717.033	1986	4610.094	1999	5677.636	2012	5229.593
1974	4545.822	1987	4760.180	2000	5867.593	2013	5361.092
1975	4422.045	1988	4989.158	2001	5757.735	2014	5417.286
1976	4691.655	1989	5064.486	2002	5795.657	2015	5267.386
1977	4830.290	1990	5024.335	2003	5866.708	2016	5179.781
1978	4871.458	1991	4973.488	2004	5965.352	2017	5143.669
1979	4931.041	1992	5065.670	2005	5990.726	2018	5295.586
1980	4756.542	1993	5168.096	2006	5907.949	2019	5159.726
1981	4632.965	1994	5242.648	2007	5998.173	2020	4594.716
1982	4391.312	1995	5307.259	2008	5803.539	2021	4911.228

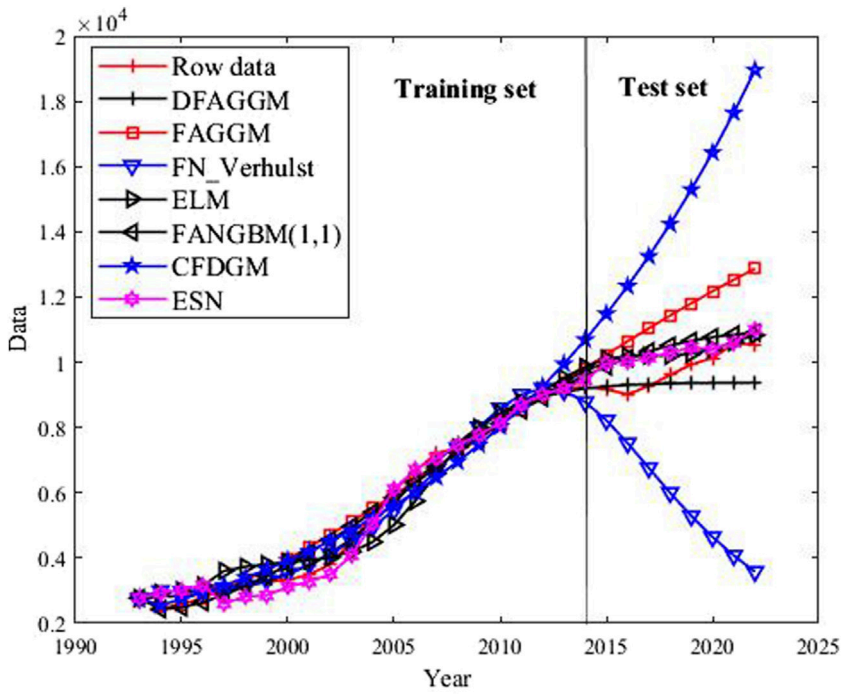

FIGURE 2
Fitting curves of the models for the case of China.

3.3 Numerical results

Based on the data provided, we can compare the performance of various forecasting methods (DFAGGM, FAGGM, *FN_Verhulst*, ELM, ESN, FANGBM, and CFDGM) against the raw data for the years 2014–2022. The evaluation metrics include Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Index of Agreement (IA). The simulation results and evaluation metrics of the models in the case study are shown in Table 2, and the fitting curves are shown in Figure 2.

TABLE 3 Simulation results and evaluation metrics of the models in the case of China.

Year	Raw data	DFAGGM	FAGGM	*FN_Verhulst*	ELM	FANGBM	CFDGM	ESN
2014	9235.530	9190.906	9816.882	8780.679	9858.872	9586.740	10686.717	9454.535
2015	9171.283	9258.588	10233.568	8220.048	10107.556	9873.037	11480.419	9964.732
2016	9014.639	9302.673	10642.179	7519.318	10188.015	10122.944	12333.070	10016.107
2017	9270.283	9331.000	11041.655	6758.092	10201.514	10336.400	13249.046	10138.529
2018	9610.704	9348.960	11431.047	6000.484	10199.185	10514.200	14233.052	10301.402
2019	9933.688	9360.172	11809.523	5288.615	10264.771	10657.852	15290.141	10452.807
2020	10130.868	9367.026	12176.362	4644.602	10402.715	10769.419	16425.739	10388.388
2021	10563.474	9371.080	12530.951	4076.062	10614.635	10851.365	17645.678	10597.115
2022	10550.248	9373.349	12872.784	3581.702	10837.431	10906.403	18956.221	11003.432
	MAPE (%)	4.863	17.077	36.103	6.152	7.149	47.944	5.674
	MSE	431560.106	3059406.978	18335050.130	458231.666	548771.933	27197966.706	383743.009
	RMSE	656.932	1749.116	4281.945	676.928	740.791	5215.167	619.470
	MAE	494.342	1674.915	3623.457	577.109	681.960	4757.708	537.370
	IA	0.512	0.483	0.100	0.668	0.681	0.230	0.733

In Table 3, DFAGGM shows the best performance in terms of MAPE (4.863%) and has relatively low MSE (431560.106), RMSE (656.932), and MAE (494.342). However, its IA is moderate (0.512). ESN performs well with a MAPE of 5.674%, MSE of 383743.009, RMSE of 619.470, MAE of 537.370. and IA of 0.733. ELM also performs well with a MAPE of 6.152%, MSE of 458231.666, RMSE of 676.928, and MAE of 577.109. Its IA is the highest among all models at 0.668, indicating a good agreement with the raw data. FANGBM has a reasonable MAPE of 7.149%, moderate MSE (548771.933), RMSE (740.791), and MAE (681.960). Its IA is also relatively high at 0.681. CFDGM has the highest MAPE (47.944%), MSE (27197966.706), RMSE (5215.167), and MAE (4757.708), making it the least accurate model. *FN_Verhulst* also performs poorly with a high MAPE (36.103%) and the lowest IA (0.100), indicating poor predictive accuracy and agreement with the raw data. According to Figure 2, it can be seen that the differences in the fitting abilities of the models during the training phase are not significant. However, in the testing phase, the overall trends of the red curve and the black curve are the closest. This phenomenon indicates that the model presented in this paper has strong generalization ability. DFAGGM and ESN are the most effective models based on the given metrics, with DFAGGM having slightly better error metrics and ESN showing the highest index of agreement. FANGBM also shows reasonable performance but is slightly less accurate than DFAGGM and ELM. CFDGM and *FN_Verhulst* are the least effective models for forecasting in this context. Although the ELM performance closely approaches the model proposed in this paper, it is important to note that the structure of machine learning algorithms is much more complex than the method presented in this paper. Therefore, the practicality of the method in this paper is stronger.

In Table 4, DFAGGM demonstrates the strongest performance with a MAPE of 2.674%, the lowest MSE (27790.780), RMSE (166.706), and MAE (135.296), along with a high Index of Agreement (IA) of 0.885. This indicates that DFAGGM closely matches the actual CO2 emissions data, making it the most accurate model among those evaluated. ESN also performs well, with a MAPE of 3.544%, MSE of 54806.406, RMSE of 234.108, and MAE of 176.529. Its IA of 0.652 suggests a strong agreement with the raw data, though slightly less accurate than DFAGGM. Despite the good performance, ESN complexity as a machine learning model contrasts with the simpler yet effective DFAGGM. ELM and ESN performed slightly worse, with a MAPE of 3.681, MSE of 59236.875, RMSE of 243.386, MAE of 181.468, and IA of 0.655. In contrast, FAGGM, FANGBM, and CFDGM exhibit higher errors and lower IA values. FAGGM has a MAPE of 7.842% and an IA of 0.372, indicating moderate performance. FANGBM and CFDGM show significant deviations with MAPEs of 21.426% and 20.143%, respectively, and low IAs, reflecting poor predictive accuracy. According to Figure 3, the training phase differences in model fitting are minimal, but during the testing phase, the red and black curves, representing DFAGGM and ELM, closely follow the actual data trends. This suggests that these models have robust generalization abilities. Overall, DFAGGM and ELM are the most effective models, with DFAGGM being slightly superior in terms of error metrics and practical simplicity. In fact,The main source of error in the carbon emission prediction process is the model's expressive capability. The reason DFAGGM outperforms FAGGM lies in its structural expression, which includes an additional adaptive

TABLE 4 Simulation results and evaluation metrics of the models in the case of the United States.

Year	Raw data	DFAGGM	FAGGM	*FN_Verhulst*	ELM	FANGBM	CFDGM	ESN
2013	5361.092	5493.747	5497.584	5171.038	5265.826	6041.233	5990.483	5184.389
2014	5417.286	5420.340	5508.715	5054.103	5385.694	6087.905	6033.670	5294.469
2015	5267.386	5337.928	5519.358	4928.648	5436.180	6134.977	6077.169	5356.937
2016	5179.781	5246.674	5529.533	4795.680	5300.532	6182.450	6120.980	5228.111
2017	5143.669	5146.812	5539.258	4656.272	5219.786	6230.326	6165.108	5135.780
2018	5295.586	5038.637	5548.551	4511.531	5186.189	6278.607	6209.554	5087.149
2019	5159.726	4922.513	5557.430	4362.579	5326.292	6327.296	6254.320	5231.882
2020	4594.716	4798.860	5565.910	4210.526	5201.154	6376.394	6299.409	5114.471
2021	4911.228	4668.160	5574.009	4056.445	4652.939	6425.904	6344.823	4568.102
	MAPE (%)	2.674	7.842	9.955	3.681	21.426	20.143	3.554
	MSE	27790.780	219328.060	310699.642	59236.875	1294374.911	1,150,627.816	54806.406
	RMSE	166.706	468.325	557.404	243.386	1137.706	1072.673	234.108
	MAE	135.296	389.986	509.294	181.468	1083.847	1018.338	176.529
	IA	0.885	0.372	0.529	0.655	0.210	0.220	0.652

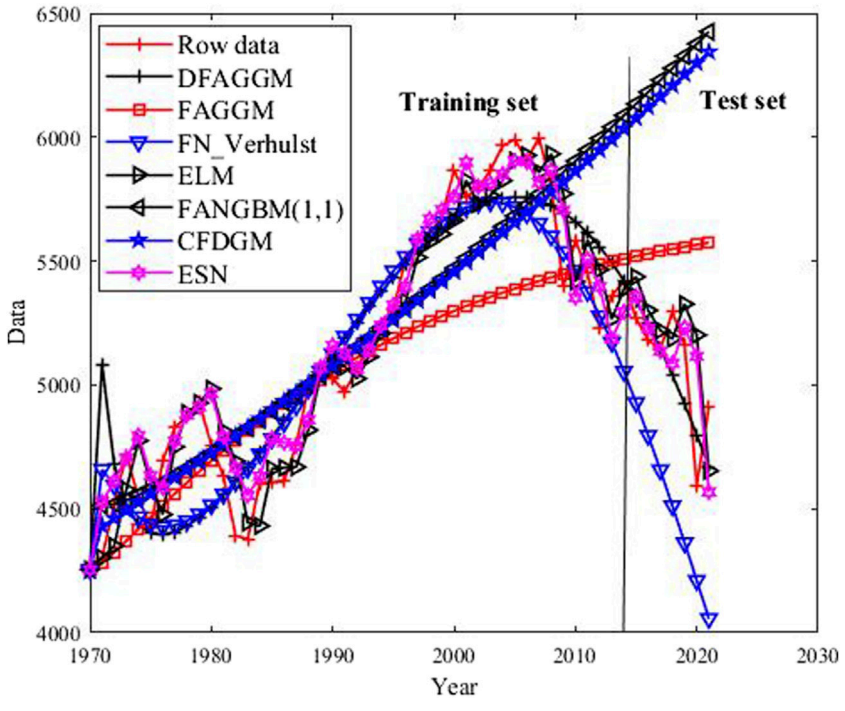

FIGURE 3
Fitting curves of the models for the case of the United States.

hyperparameter α. This allows it to better adjust its structure to fit the time series, resulting in improved prediction accuracy. The inferior performance of *FN_Verhulst* and CFDGM compared to DFAGGM can be anticipated due to the overly simplistic structures of these two models. Although FANGBM exhibits good nonlinear fitting capabilities, its grey action quantity is relatively simple, indicating that there is room for further improvement in its fitting ability. ELM and ESN performed worse compared to DFAGGM, possibly because the dataset is too small to allow for perfect training.

4 Conclusion and future work

To achieve sustainable development, countries around the world are striving to reach carbon peak and carbon neutrality goals while developing their economies. Accurate prediction of carbon emissions provides valuable references for decision-makers to manage effectively. However, due to the complexity of carbon emission data series, achieving high-precision carbon emission forecasts remains challenging. Through experiments and comparative analysis, the DFAGGM model proposed in this paper demonstrates high accuracy in annual carbon emission predictions and strong interpretability, making it a reliable model for decision-makers.

At the same time, it is worth noting that DFAGGM still has some shortcomings and room for improvement. Firstly, the model has a univariate limitation as it is a univariate grey prediction model and does not consider the impact of other factors on the main variable. Secondly, while the model performs well in handling short-term or medium-term time series data, it may lack the ability to capture cyclical changes and respond to sudden changes in the time series. Lastly, this study mainly uses carbon emission data from China and the United States to validate the model. Whether the model is applicable to data from other countries and regions still needs further verification. In the future, this research proposes several directions for further study. Firstly, the model proposed in this paper can be combined with other grey prediction model optimization methods to enhance the model's adaptability. Additionally, due to the complexity of carbon emission predictions, it is necessary to consider more influencing factors, such as economic scale, industrial structure, and carbon emission intensity. A multivariate grey prediction model can be constructed based on the analysis of these influencing factors and the model proposed in this paper. Finally, it is necessary to apply the DFAGGM model and its optimized model to carbon emission predictions in more countries and regions to test its generalization capability.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JJ: Conceptualization, Writing–original draft. YB: Writing–original draft, Data curation, Resources. MZ: Project administration, Writing–review and editing. ZH: Software, Validation, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

AlOmar, M. K., Hameed, M. M., Al-Ansari, N., Mohd Razali, S. F., and AlSaadi, M. A. (2023). Short-medium-and long-term prediction of carbon dioxide emissions using wavelet-enhanced extreme learning machine. *Civ. Eng. J.* 9 (4), 815–834. doi:10.28991/cej-2023-09-04-04

Bhatt, H., Davawala, M., Joshi, T., Shah, M., and Unnarkat, A. (2023). Forecasting and mitigation of global environmental carbon dioxide emission using machine learning techniques. *Clean. Chem. Eng.* 5, 100095. doi:10.1016/j.clce.2023.100095

Cui, T., Shi, Y., Lv, B., Ding, R., and Li, X. (2023). Federated learning with SARIMA-based clustering for carbon emission prediction. *J. Clean. Prod.* 426, 139069. doi:10.1016/j.jclepro.2023.139069

Ding, S., and Zhang, H. (2023). Forecasting Chinese provincial CO2 emissions: a universal and robust new-information-based grey model. *Energy Econ.* 121 (2023), 106685. doi:10.1016/j.eneco.2023.106685

Faruque, Md O., Rabby, Md A. J., Hossain, Md A., Islam, Md R., Rashid, Md M.Ur, and Muyeen, S. M. (2022). A comparative analysis to forecast carbon dioxide emissions. *Energy Rep.* 8, 8046–8060. doi:10.1016/j.egyr.2022.06.025

Gao, M., Yang, H., Xiao, Q., and Goh, M. (2022). A novel method for carbon emission forecasting based on Gompertz's law and fractional grey model: evidence from American industrial sector. *Renew. Energy* 181, 803–819. doi:10.1016/j.renene.2021.09.072

Guo, J., Liu, W., Tu, L., and Chen, Y. (2021). Forecasting carbon dioxide emissions in BRICS countries by exponential cumulative grey model. *Energy Rep.* 7, 7238–7250. doi:10.1016/j.egyr.2021.10.075

Javanmard, M. E., Tang, Y., Wang, Z., and Tontiwachwuthikul, P. (2023). Forecast energy demand, CO2 emissions and energy resource impacts for the transportation sector. *Appl. Energy* 338, 120830. doi:10.1016/j.apenergy.2023.120830

JuLong, D. (1982). Control problems of grey systems. *Syst. Control Lett.* 1 (5), 288–294. doi:10.1016/s0167-6911(82)80025-x

Karakurt, I., and Aydin, G. (2023). Development of regression models to forecast the CO2 emissions from fossil fuels in the BRICS and MINT countries. *Energy* 263, 125650. doi:10.1016/j.energy.2022.125650

Khajavi, H., and Rastgoo, A. (2023). Predicting the carbon dioxide emission caused by road transport using a Random Forest (RF) model combined by Meta-Heuristic Algorithms. *Sustain. Cities Soc.* 93, 104503. doi:10.1016/j.scs.2023.104503

Kour, M. (2023). Modelling and forecasting of carbon-dioxide emissions in South Africa by using ARIMA model. *Int. J. Environ. Sci. Technol.* 20 (10), 11267–11274. doi:10.1007/s13762-022-04609-7

Lin, X., Zhu, X., Feng, M., Han, Y., and Geng, Z. (2021). Economy and carbon emissions optimization of different countries or areas in the world using an improved Attention mechanism based long short term memory neural network. *Sci. Total Environ.* 792, 148444. doi:10.1016/j.scitotenv.2021.148444

Qiao, Z., Meng, X., and Wu, L. (2021). Forecasting carbon dioxide emissions in APEC member countries by a new cumulative grey model. *Ecol. Indic.* 125, 107593. doi:10.1016/j.ecolind.2021.107593

Qin, J., and Gong, N. (2022). The estimation of the carbon dioxide emission and driving factors in China based on machine learning methods. *Sustain. Prod. Consum.* 33, 218–229. doi:10.1016/j.spc.2022.06.027

Rahman, M., Rashid, F., Roy, S., and Habib, M. A. (2024). Application of extreme learning machine (ELM) forecasting model on CO2 emission dataset of a natural gas-fired power plant in Dhaka, Bangladesh. *Data Brief* 54, 110491. doi:10.1016/j.dib.2024.110491

Rao, C., Huang, Q., Chen, L., Goh, M., and Hu, Z. (2023). Forecasting the carbon emissions in Hubei Province under the background of carbon neutrality: a novel STIRPAT extended model with ridge regression and scenario analysis. *Environ. Sci. Pollut. Res.* 30 (20), 57460–57480. doi:10.1007/s11356-023-26599-w

Wang, L., Hu, H., Ai, X. Y., and Liu, H. (2018). Effective electricity energy consumption forecasting using echo state network improved by differential evolution algorithm. *Energy* 153, 801–815. doi:10.1016/j.energy.2018.04.078

Yang, J., and Wu, Z. (2023). Modelling and forecasting nonrenewable energy consumption and carbon dioxide emissions in China using a PSO algorithm based fractional nonlinear grey Bernoulli model. *Environ. Sci. Pollut. Res.* 30, 69651–69665. doi:10.1007/s11356-023-27189-6

Zeng, B., Zheng, T., Yang, Y., and Wang, J. (2023). A novel grey Verhulst model with four parameters and its application to forecast the carbon dioxide emissions in China. *Sci. Total Environ.* 899, 165648. doi:10.1016/j.scitotenv.2023.165648

Zhang, X., Sun, J., and Wang, F. (2024). Assessment and regression of carbon emissions from the building and construction sector in China: a provincial study using machine learning. *J. Clean. Prod.* 450, 141903. doi:10.1016/j.jclepro.2024.141903

Zhu, P., Zhang, H., Shi, Y., Xie, W., and Pang, M. (2024). A novel discrete conformable fractional grey system model for forecasting carbon dioxide emissions. *Environ. Dev. Sustain.*, 1–29. doi:10.1007/s10668-024-04479-8