UNDER PEER REVIEW

A novel neural-augmented grey system approach with its applications in carbon emission forecasting

Abstract

Grey system models (GMs) have achieved considerable progress in recent decades, yet their effectiveness is often limited when dealing with nonlinear data. In contrast, machine learning (ML) models can capture complex nonlinear relationships but generally require large datasets and lack interpretability. To address these limitations, this paper proposes a novel neural grey system model that embeds a neural network into the traditional GM framework. This integration enhances the model' s nonlinear learning capacity while maintaining the grey model's suitability for sparse and uncertain data. The model is optimized using the Adam algorithm, and hyperparameters are fine-tuned via GridSearch. To validate its effectiveness, we conduct carbon emission forecasting experiments for four countries, comparing the proposed model against eight benchmark models, including conventional GMs and ML-based approaches. Results demonstrate superior forecasting accuracy and generalization ability, confirming the proposed model' s potential for complex, nonlinear prediction tasks in environmental and energy domains.

Keywords: Grey system, Multilayer perceptron, Adaptive Moment Estimation, Carbon emission forecasting.

1 Introduction

Grey System Theory (GST) has undergone significant development over the past four decades. In contrast to conventional uncertainty modeling theories, GST is uniquely positioned to analyze systems with limited, incomplete, or low-quality data [1]. At the core of GST are Grey Models (GMs), which have found widespread applications across various domains such as agriculture [2], economic forecasting [3], energy systems [4], production planning [5], and traffic management [6]. Originating from Deng' s seminal work in 1984 [7], the theory was formally established through the introduction of the general grey modeling framework, GM(n, h) [8], with foundational models like GM$(1, 1)$ and GM$(1, N)$ applied early on to forecast China's grain production. Since then, numerous variants have emerged, including the Discrete Grey Models DGM$(1, 1)$

and DGM$(1, N)$ [9], which serve as the discrete analogues of their continuous counterparts.

Broadly, linear grey models can be divided into univariate and multivariate categories. The univariate models are further classified into continuous and discrete forms. Among the continuous univariate models derived from GM$(1, 1)$, several notable developments stand out. For instance, Wang proposed the DSGM$(1, 1)$ model [10], incorporating dynamic seasonal factors, while Wu introduced the FGM$(1, 1)$ model [11], which employs fractional-order accumulation to enhance predictive performance. Cui extended the framework to a generalized version, NGM$(1, 1, k)$ [12], and Chen further improved its adaptability through the NGM$(1, 1, k, c)$ model [13]. Jiang also contributed by formulating a version of NGM$(1, 1)$ [14] tailored for non-homogeneous and non-equidistant time series. Discrete univariate models, mostly derived from DGM$(1, 1)$, have also seen notable enhancements. Xie introduced the NGDM$(1, 1)$ model [15], based on a non-homogeneous index sequence, offering improved accuracy. To address seasonal variations, Xia proposed the SDGM model [16], utilizing a cyclic accumulation mechanism. Yang incorporated trigonometric functions into the DGM structure and introduced the DGM$(1, 1, T)$ model [17], which demonstrated effective drought prediction capabilities. While continuous multivariate models generally stem from GM$(1, N)$, discrete forms are typically based on DGM$(1, N)$ [18].

Despite the success of linear models, their limited capacity to capture nonlinear dynamics has prompted the development of nonlinear grey models. In this context, Ma introduced KGM$(1, n)$ [19], a nonlinear multivariate model based on kernel methods, and later proposed the GMW-KRGM model [20], which integrates kernel ridge regression. Further advancements include models embedding nonlinear elements such as $y^2(t)$ or $y^\gamma(t)$ to improve expressiveness. The Grey Verhulst Model (GVM$(1, 1)$), first used by Shaikh [21] for natural gas demand forecasting, marked an early attempt at modeling nonlinear growth. To counteract its 'drift phenomenon,' Zhou introduced the Generalized Grey Verhulst Model (GGVM) [22], which Xiao [23] later employed in dynamic traffic flow forecasting. Wang's NGM$(1, 1, \gamma)$ [24] leveraged biological metabolism concepts, and Chen's Nonlinear Grey Bernoulli Model (NGBM) [25] utilized Bernoulli equations. This line of research continued with Lu's ONGBM$(1, 1)$ [26] and Liu's weighted fractional NGBM [27]. Other important contributions include Xiao's Grey Riccati-Bernoulli Model (GRBM$(1, 1)$) [28] for energy demand forecasting and Wu's time-variant Bernoulli model NBGM$(1, 1, t^\alpha)$ [29] for solar power prediction. Wang developed a time-varying GM$(1, 1)$ [30] with adaptive structure parameters, and Wu's Grey Riccati Model (GRM) [31] integrated Riccati equations. Luo later applied a grey-Richards hybrid model for epidemic modeling [32], while Gatabazi [33, 34] proposed a Fractional Grey Lotka–Volterra Model (FGLVM) for cryptocurrency market analysis. Mao employed a classical Lotka–Volterra structure to study the impact of online payment systems in the banking sector [35].

Recent years have also witnessed the growing integration of grey models with neural networks, aiming to overcome the limitations of traditional models in handling nonlinearity. Neural networks, known for their strong approximation capabilities, have

been increasingly embedded into grey model structures. Examples include Lei's Neural ODE-based model (NODGM) [36], Li's DGNNM$(2,1)$ [37], which couples a discrete grey model with a feedforward neural network, and Hao's hybrid system combining grey prediction with a particle-swarm-optimized neural network [38]. While such combinations offer improved performance, they often result in complex architectures and challenging training procedures.

Motivated by the above literature, this study proposes a novel neural grey model that integrates neural network structures into the classical grey modeling framework for carbon emission forecasting. Specifically, a multilayer perceptron (MLP) is embedded into the whitenization equation of the GM(1,1) model to replace the grey action component. The model jointly optimizes the development coefficient and neural network weights via gradient descent. To enhance convergence and prediction performance, the Adam optimizer is employed, while Grid Search is used for systematic hyperparameter tuning. The model is applied to annual carbon emission datasets from multiple countries to evaluate its effectiveness in capturing nonlinear patterns under small-sample conditions. The main contributions of this paper are as follows:

- A novel and interpretable neural grey model is proposed, which combines the strengths of grey models in small-sample settings with the nonlinear learning capacity of neural networks.
- A dual optimization strategy is introduced, utilizing Adam for efficient convergence and Grid Search for optimal hyperparameter selection, thereby improving model accuracy and generalization.
- This study pioneers the application of neural grey models to national CO_2 emission forecasting, demonstrating strong adaptability and potential for supporting carbon peaking and neutrality initiatives.

The remainder of this paper is structured as follows: Section 2 introduces the general grey model formulation and solution process; Section 3 introduces the proposed neural grey model framework. Section 4 presents five real-world applications and the forecasting results; Section 5 discuss the forecasting results of the each carbon emission data; and Section 6 concludes the study.

2 Theoretical background

2.1 General formulation of multilayer perceptron

Formally, an MLP consists of an input layer, one or more hidden layers, and an output layer. Each layer is composed of numerous artificial neurons, also referred to as perceptrons or nodes, interconnected via weighted connections [39]. The primary function of the MLP is to transform input data through successive layers of nonlinear transformations, ultimately producing an output prediction. A simple mlp network structure with one-hidden-layer is shown in Fig.(1).

Mathematically, the forward propagation process of an MLP can be expressed as follows: for each layer l, the output $\mathbf{x}^{(l)}$ is computed as the application of a nonlinear activation function σ to the linear transformation of the previous layer's output $\mathbf{x}^{(l-1)}$, incorporating weights $\mathbf{W}^{(l)}$ and biases $\mathbf{b}^{(l)}$:

$$\begin{aligned} \mathbf{z}^{(l)} &= \mathbf{W}^{(l)}\mathbf{x}^{(l-1)} + \mathbf{b}^{(l)} \\ \mathbf{x}^{(l)} &= \sigma\left(\mathbf{z}^{(l)}\right) \end{aligned} \tag{1}$$

During training, the parameters (weights and biases) of the MLP are optimized to minimize a predefined loss function, typically through backpropagation and gradient-based optimization techniques. Backpropagation involves the systematic calculation of gradients with respect to the parameters of the network, facilitating parameter updates in the direction that reduces the loss [40].

MLPs are characterized by their universal approximation capabilities, enabling them to approximate arbitrary functions with sufficient capacity and data. However, their effectiveness is contingent upon various factors, including network architecture design, activation functions, optimization algorithms, and hyperparameter tuning [41].

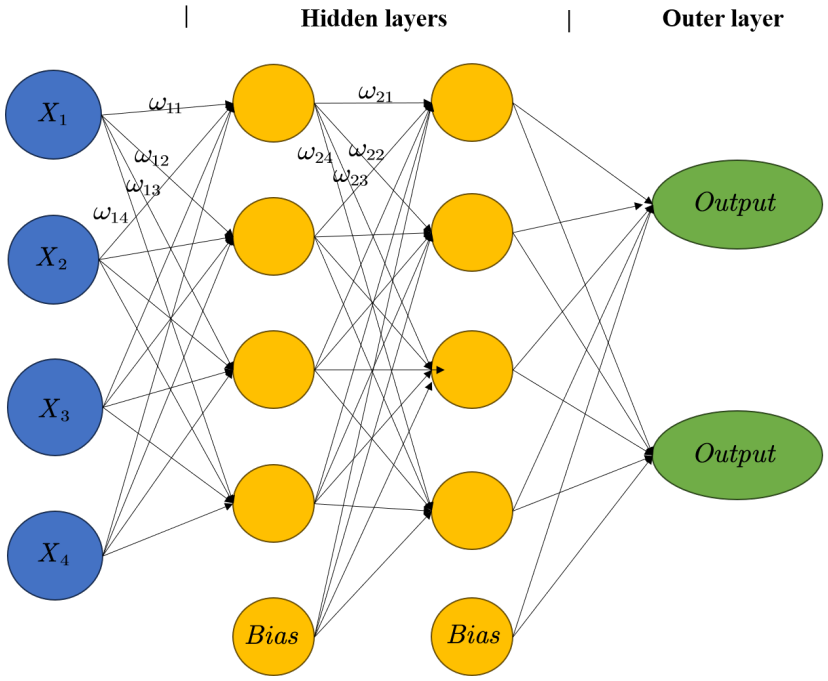

Fig. 1. The architecture of the Multi-Layer Perceptron (MLP)

2.2 General formulation of grey system models

In grey system theory, for original sequences $x_i^{(0)}(t)$ and $y^{(0)}(t)$ $(t = 1, 2, \ldots, n)$, their first-order accumulations $x_i^{(1)}(t)$ and $y^{(1)}(t)$ are defined as:

$$x_i^{(1)}(t) = \sum_{\tau=1}^{t} x_i^{(0)}(\tau), \quad y^{(1)}(t) = \sum_{\tau=1}^{t} y^{(0)}(\tau), \quad t = 1, 2, \ldots, n. \tag{2}$$

This form of accumulation is known as the first-order accumulated generating operation (1-AGO) [42].

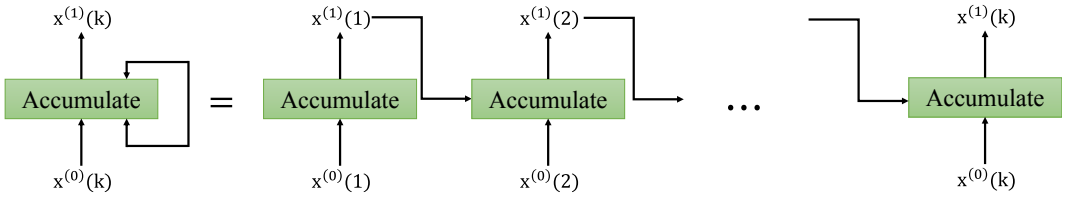

Fig. 2. The structure of 1-AGO

The general whitening differential equation of the grey system is expressed as [8]:

$$\frac{dy^{(1)}(t)}{dt} + ay^{(1)}(t) = f\big(x^{(1)}(t);\theta\big), \tag{3}$$

where $x^{(1)}(t)$ is a vector composed of $x_i^{(1)}(t)$:

$$x^{(1)}(t) = \big(x_1^{(1)}(t), x_2^{(1)}(t), \ldots, x_n^{(1)}(t)\big),$$

θ is a vector of parameters related to the input sequence, and a is referred to as the development coefficient. The function $f(\cdot)$ represents the system behavior.

By discretizing Eq. (3), we obtain the general discrete equation of the grey system:

$$y^{(0)}(k) + az^{(1)}(k) = f\left(\tfrac{1}{2}\big(x^{(1)}(k-1) + x^{(1)}(k)\big);\theta\right), \tag{4}$$

where $z^{(1)}(k)$ is called the *background value*, defined as:

$$z^{(1)}(k) = \tfrac{1}{2}\big(y^{(1)}(k-1) + y^{(1)}(k)\big).$$

For convenience, we let:

$$v_k = \tfrac{1}{2}\big(x^{(1)}(k-1) + x^{(1)}(k)\big),$$

so Eq. (4) can be rewritten as:

$$y^{(0)}(k) + az^{(1)}(k) = f(v_k;\theta). \tag{5}$$

Eq. (5) is typically used for parameter estimation. Once parameters a and θ are estimated, the forecasting phase begins.

By solving Eq. (3) with the initial condition $y^{(1)}(1) = y^{(0)}(1)$, the continuous form of the response function is obtained:

$$y^{(1)}(t) = y^{(0)}(1)e^{-a(t-1)} + \int_1^t e^{-a(t-\tau)} f\big(x^{(1)}(\tau);\theta\big)\, d\tau. \tag{6}$$

However, this continuous formulation is difficult to use in practice and is typically approximated numerically. By discretizing the integral in Eq. (6), we get the following discrete response function:

$$\hat{y}^{(1)}(k) = y^{(0)}(1)e^{-a(k-1)} + \sum_{\tau=2}^{k} e^{-a(k-\tau+\frac{1}{2})} \cdot f\left(\tfrac{1}{2}\big(x^{(1)}(k-1) + x^{(1)}(k)\big);\theta\right), \tag{7}$$

which can also be written as:

$$\hat{y}^{(1)}(k) = y^{(0)}(1)e^{-a(k-1)} + \sum_{\tau=2}^{k} e^{-a(k-\tau+\frac{1}{2})} \cdot f(v_k;\theta). \tag{8}$$

After obtaining the estimated values of $\hat{y}^{(1)}(k)$ using Eq. (8), we can calculate the predicted values of the original sequence using the inverse accumulated generating operation (1-IAGO):

$$\hat{y}^{(0)}(k) = \hat{y}^{(1)}(k) - \hat{y}^{(1)}(k-1). \tag{9}$$

3 The proposed neural grey system model

3.1 The representation of grey system model and its solution

The previous section gave an overview of typical grey system models and their solutions. It is apparent that existing grey models encounter challenges when attempting to predict nonlinear time series data. This arises from the usual estimation of parameters a and θ in the function $f(\cdot)$ using the least squares method, resulting in the construction of a linear function.

In this work, we embed a single-hidden-layer neural network as the function $f(\cdot)$ to better capture nonlinear patterns in data. Accordingly, we rewrite the whitening differential equation (Eq. (3)) as:

$$\frac{dy^{(1)}(t)}{dt} + ay^{(1)}(t) = f(x_k^{(1)}; \theta), \tag{10}$$

where the neural network function $f(\cdot)$ is defined as:

$$f(x_k^{(1)}; \theta) = \sum_{j=1}^{L} \beta_j \cdot S(x_k^{(1)}; w_j, b_j) + b, \tag{11}$$

with L being the number of hidden units. Here, $S(\cdot)$ is the activation function, and β_j, w_j, and b_j are the parameters of the neural network.

In this study, we choose the sigmoid function due to its smoothness and differentiability:

$$S(x; w_j, b_j) = \frac{1}{1 + e^{-w_j^\top x + b_j}}. \tag{12}$$

To estimate the parameters, we solve Eq. (10) and derive its discrete version:

$$y^{(0)}(k) + az^{(1)}(k) = \sum_{j=1}^{L} \beta_j \cdot S(v_k; w_j, b_j) + b, \tag{13}$$

where the background value v_k is defined as:

$$v_k = \frac{1}{2}\left(x^{(1)}(k-1) + x^{(1)}(k)\right).$$

Eq. (13) is a general form of the grey system model with neural structure. If the neural network component on the right-hand side is removed and only the bias term b is retained, the model degenerates to the classical GM(1,1) [43].

We can also derive the continuous form of the response function by solving Eq. (10) with the initial condition $y^{(1)}(1) = y^{(0)}(1)$:

$$y^{(1)}(t) = y^{(0)}(1)e^{-a(t-1)} + \int_1^t e^{-a(t-\tau)} \left(\sum_{j=1}^{L} \beta_j \cdot S(x^{(1)}(\tau); w_j, b_j) + b \right) d\tau. \tag{14}$$

By discretizing the integral in Eq. (14), the discrete response function of the model is obtained as:

$$\hat{y}^{(1)}(k) = y^{(0)}(1)e^{-a(k-1)} + \sum_{\tau=2}^{k} e^{-a(k-\tau+\frac{1}{2})} \cdot \left(\sum_{j=1}^{L} \beta_j \cdot S(v_k; w_j, b_j) + b \right). \tag{15}$$

This final formulation characterizes the neural grey system model, where the dynamic response is regulated by a neural network embedded within the traditional grey framework.

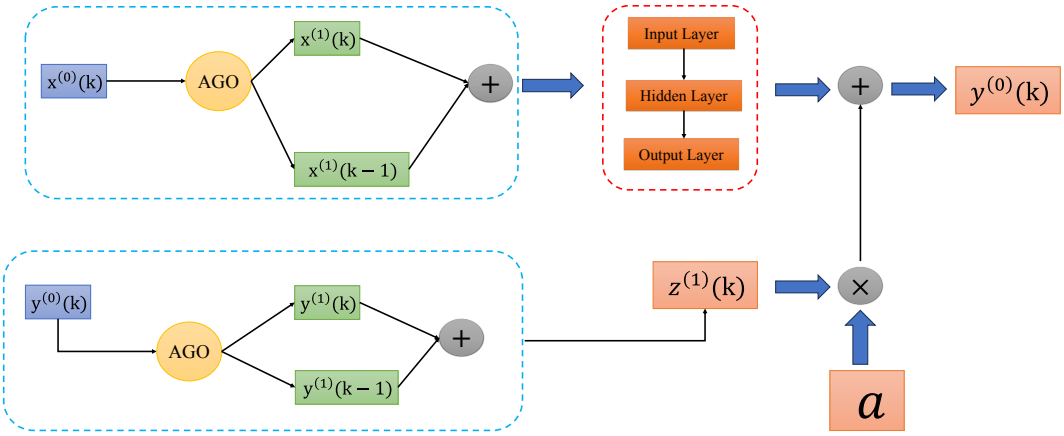

Fig. 3. The structure of the neural grey system model

3.2 Adam algorithm for training the neural grey system model

Typically, neural networks do not have analytical solutions, and thus optimization algorithms are required to estimate the parameters. Common algorithms include Gradient Descent (GD) [44], Stochastic Gradient Descent (SGD) [45], and Adam [46].

Gradient Descent computes the gradient over the entire training set, which is computationally expensive for large datasets. SGD uses only one sample per update, which reduces computation but introduces significant noise. Adam, however, introduces the concept of momentum and adaptively adjusts the learning rate using first and second moment estimates, which improves convergence speed and stability [46]. It is simple to implement and memory-efficient, making it suitable for training the proposed model. The complete procedure of training the neural grey model using the Adam optimizer is detailed in Algorithm 1. The specific algorithm is given as follows. We define the training error e_k at each step k as:

$$e_k = y^{(0)}(k) + az^{(1)}(k) - \sum_{j=1}^{L} \beta_j S(v_k; w_j, b_j) - b. \tag{16}$$

Then, the total training error is defined as:

$$E(a, \theta) = \frac{1}{N} \sum_{k=2}^{N} e_k^2 = \mathbf{e}^\top \mathbf{e}, \tag{17}$$

where $\theta = [\beta, w, b]$ is the parameter set of the neural network. The goal of the Adam optimizer is to minimize $E(a, \theta)$.

The gradient vector to be computed is:

$$J = \left[\frac{\partial E}{\partial a}, \frac{\partial E}{\partial \theta}\right], \tag{18}$$

where

$$\begin{aligned}
\frac{\partial E}{\partial a} &= \frac{2}{N}\sum_{k=2}^{N}\Big[y^{(0)}(k) + az^{(1)}(k) - \sum_{j=1}^{L}\beta_j S(v_k; w_j, b_j) - b\Big]z^{(1)}(k), \\
\frac{\partial E}{\partial \theta} &= -\frac{2}{N}\sum_{k=2}^{N}\Big[y^{(0)}(k) + az^{(1)}(k) - \sum_{j=1}^{L}\beta_j S(v_k; w_j, b_j) - b\Big]\sum_{j=1}^{L}\beta_j \frac{\partial S(v_k; w_j, b_j)}{\partial \theta}.
\end{aligned} \tag{19}$$

For standard GD, parameters are updated as:

$$\begin{bmatrix} a^{(k+1)} \\ \theta^{(k+1)} \end{bmatrix} = \begin{bmatrix} a^{(k)} \\ \theta^{(k)} \end{bmatrix} - \eta \cdot J, \tag{20}$$

where η is the learning rate.

To improve stability and convergence, Adam uses first and second moment estimates. The biased moment estimates are:

$$m_k = \mu_1 \cdot m_{k-1} + (1 - \mu_1) \cdot J, \tag{21}$$

$$v_k = \mu_2 \cdot v_{k-1} + (1 - \mu_2) \cdot J^2, \tag{22}$$

where μ_1 and μ_2 are decay rates.

Then, bias-corrected estimates are computed as:

$$\hat{m}_k = \frac{m_k}{1 - \mu_1^k}, \tag{23}$$

$$\hat{v}_k = \frac{v_k}{1 - \mu_2^k}. \tag{24}$$

The final update rule is:

$$\begin{bmatrix} a^{(k+1)} \\ \theta^{(k+1)} \end{bmatrix} = \begin{bmatrix} a^{(k)} \\ \theta^{(k)} \end{bmatrix} - \eta \cdot \frac{\hat{m}_k}{\sqrt{\hat{v}_k} + \epsilon}, \tag{25}$$

where ϵ is a small constant to avoid division by zero.

Algorithm 1: Adam Algorithm for Training the Neural Grey System Model

Require: Objective function $E(a, \theta)$ (Eq. (17)), learning rate l, maximum number of epochs T

Ensure: Optimized parameters a, θ

Initialize parameters a_0, θ_0 randomly

Set $\mu_1 \leftarrow 0.9$, $\mu_2 \leftarrow 0.999$, $\epsilon \leftarrow 10^{-8}$

Initialize $m_0 \leftarrow 0$, $v_0 \leftarrow 0$

For each epoch k from 1 to T, perform:

  Compute gradient: $g_k \leftarrow \nabla E(a_{k-1}, \theta_{k-1})$

  First moment estimate: $m_k \leftarrow \mu_1 \cdot m_{k-1} + (1 - \mu_1) \cdot g_k$

  Second moment estimate: $v_k \leftarrow \mu_2 \cdot v_{k-1} + (1 - \mu_2) \cdot g_k^2$

  Bias-corrected first moment: $\hat{m}_k \leftarrow m_k / (1 - \mu_1^k)$

  Bias-corrected second moment: $\hat{v}_k \leftarrow v_k / (1 - \mu_2^k)$

  Update parameters:

    $a_k \leftarrow a_{k-1} - l \cdot \hat{m}_k / (\sqrt{\hat{v}_k} + \epsilon)$

    $\theta_k \leftarrow \theta_{k-1} - l \cdot \hat{m}_k / (\sqrt{\hat{v}_k} + \epsilon)$

return a_T, θ_T

3.3 Hyperparameter optimization via gridsearch

In Section 3.2, model parameters such as weights and biases $[\boldsymbol{\alpha}, \boldsymbol{\theta}]$ are optimized using the Adam algorithm. However, hyperparameters like the number of neurons L and learning rate l require separate tuning. To this end, the Grid Search algorithm systematically explores all candidate hyperparameter combinations, evaluates their performance via cross-validation, and selects the best set [47].

Let Θ denote the hyperparameter space, with each combination represented by a vector $\boldsymbol{\theta}$ including L, l, and other relevant parameters. The optimal combination $\boldsymbol{\theta}^*$ minimizes a validation loss function $f(\boldsymbol{\theta}, D_{\text{train}}, D_{\text{val}})$ after training on D_{train}:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta} \in \Theta} f(\boldsymbol{\theta}, D_{\text{train}}, D_{\text{val}}). \tag{26}$$

Here, f is the Negative Mean Squared Error (NMSE):

$$f(\boldsymbol{\theta}, D_{\text{train}}, D_{\text{val}}) = -\frac{1}{|D_{\text{val}}|} \sum_{i \in D_{\text{val}}} (y_i - \hat{y}_i)^2, \tag{27}$$

where $|D_{\text{val}}|$ is the validation set size, and y_i, $\hat{y}_i$ are true and predicted values.

As shown in Fig. 2, each parameter combination is denoted by PC_i, with corresponding model MD_i. The validation predictions VP_i are compared against true values using the Mean Absolute Percentage Error (MAPE) metric M_i:

$$M_i = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{\text{TrueValue}_t - VP_{i,t}}{\text{TrueValue}_t} \right|, \tag{28}$$

where n is the validation set length. Grid Search selects the combination minimizing M_i, thus optimizing hyperparameters to enhance model accuracy and generalization.

4 Applications

To verify the superiority of the proposed model, this study utilizes carbon emission data from five countries—China, the United States, India, Japan, and Canada—spanning the period from 1965 to 2022. For comparative analysis, four grey system models—GM, DGM, NGM, and BernoulliGM—along with four machine learning models—Support Vector Regression (SVR), Multi-Layer Perceptron (MLP), Random Forest (RF), and K-Nearest Neighbors (KNN)—are selected as benchmarks, as summarized in Table 1.

To evaluate the forecasting performance of these models, the Mean Absolute Percentage Error (MAPE) is employed as the evaluation metric. The MAPE is defined as follows:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100\% \tag{29}$$

where y_t denotes the actual value, $\hat{y}_t$ is the predicted value at time t, and n represents the total number of observations. MAPE measures the average percentage error between predicted and actual values, and is widely used in time series forecasting due to its scale-independence and interpretability.

Table 1: Summary of the eight benchmark models

Full Name	**Abbreviation**	**Reference**
Grey Model	GM	[8]
Discrete Grey Model	DGM	[9]
Nonlinear Grey Model	NGM	[21]
Bernoulli Grey Model	BernoulliGM	[48]
Support Vector Regression	SVR	[49]
Multilayer Perceptron	MLP	[39]
Random Forest	RF	[50]
K-Nearest Neighbors	KNN	[51]

4.1 Case I: Annual Carbon Emissions in China

The performance of various models in forecasting China's carbon emissions is illustrated in Fig. 4. The detailed MAPE values for both the training and testing datasets are presented in Table 2. As depicted in Fig. 4, China's carbon emissions exhibited a pronounced upward trend from 1965 to 2022, with an accelerated growth rate particularly after 2000. This trend reflects the rapid economic development and industrialization, which substantially increased energy demand. In recent years, the growth rate has moderated and exhibited fluctuations, indicating the influence of energy structure

adjustments and environmental policies. A thorough understanding of these dynamics is essential for accurate subsequent carbon emission forecasting.

In terms of prediction performance, it is evident that the proposed NGSMRF model demonstrates a significant advantage over the other eight models, achieving the lowest MAPE of 3.3141% on the testing dataset. In contrast, the GM model yields a testing MAPE of 20.3820%, and the DGM model reaches 17.2816%, highlighting the inherent limitations of traditional grey models that are fundamentally based on linear assumptions. Additionally, the MLP model reports a testing MAPE of 19.0688%, which suggests potential training deficiencies and indirectly underscores the superiority of the proposed NGSMRF framework that integrates MLP with GM to enhance generalization.

Moreover, models such as RF and KNN, despite exhibiting strong performance on the training dataset, fail to maintain similar accuracy on the testing set. This indicates the presence of overfitting to some extent and further emphasizes the robustness and generalization capability of the NGSMRF model.

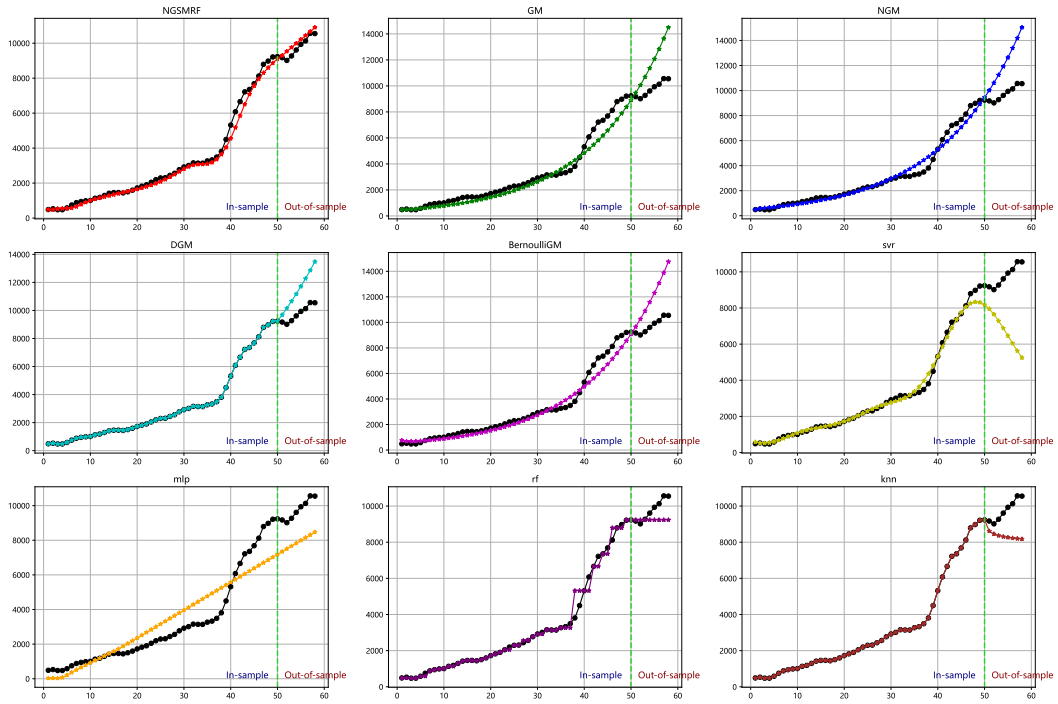

Fig. 4. Comparison of predicted annual carbon emissions for China

Table 2: MAPE for training and testing of models on China carbon emission

Metric	NN Grey	GM	NGM	DGM	Bernoulli GM	SVR	MLP	RF	KNN
Train MAPE (%)	6.5673	14.5445	9.8206	**1.0923**	14.2424	4.8864	33.9280	3.0225	1.2323
Test MAPE (%)	**3.3141**	20.3820	26.0647	17.2816	22.6771	31.3207	19.0688	6.0410	14.4995

4.2 Case II: Annual Carbon Emissions in US

As shown in Fig. 5, the annual carbon emissions of the United States demonstrate a generally high level with significant fluctuations from 1965 to 2022. While the emissions steadily increased during the 1960s to the early 2000s, peaking around the mid-2000s, a clear downward trend has been observed since approximately 2007. This decline reflects the combined effects of energy structure optimization, technological improvements, and environmental regulations. Unlike rapidly developing economies, the U.S. has entered a relatively stable stage of emissions, which makes accurate modeling and forecasting particularly dependent on capturing long-term non-linear patterns and structural changes.

As presented in Table 3, the proposed NGSMRF model outperforms all baseline models on the US dataset, achieving the lowest testing MAPE of 3.3139%, which highlights its robustness and strong generalization capability. In comparison, traditional grey models such as GM and NGM exhibit significantly higher testing errors (22.9998% and 28.9999%, respectively), largely due to their inherent linear assumptions. The DGM model, despite achieving perfect fitting on the training data, records a testing MAPE of 7.7398%, indicating poor generalization. Similarly, the MLP model also performs poorly on the testing set (25.9244%), suggesting instability during training.

Notably, RF achieves the best performance on the training set (1.0775%) but fails to maintain such accuracy on the testing data (7.0394%), suggesting signs of overfitting. A similar trend is observed with KNN. These results further demonstrate the advantages of NGSMRF, which integrates grey modeling principles with neural structures, in delivering stable and accurate predictions across datasets.

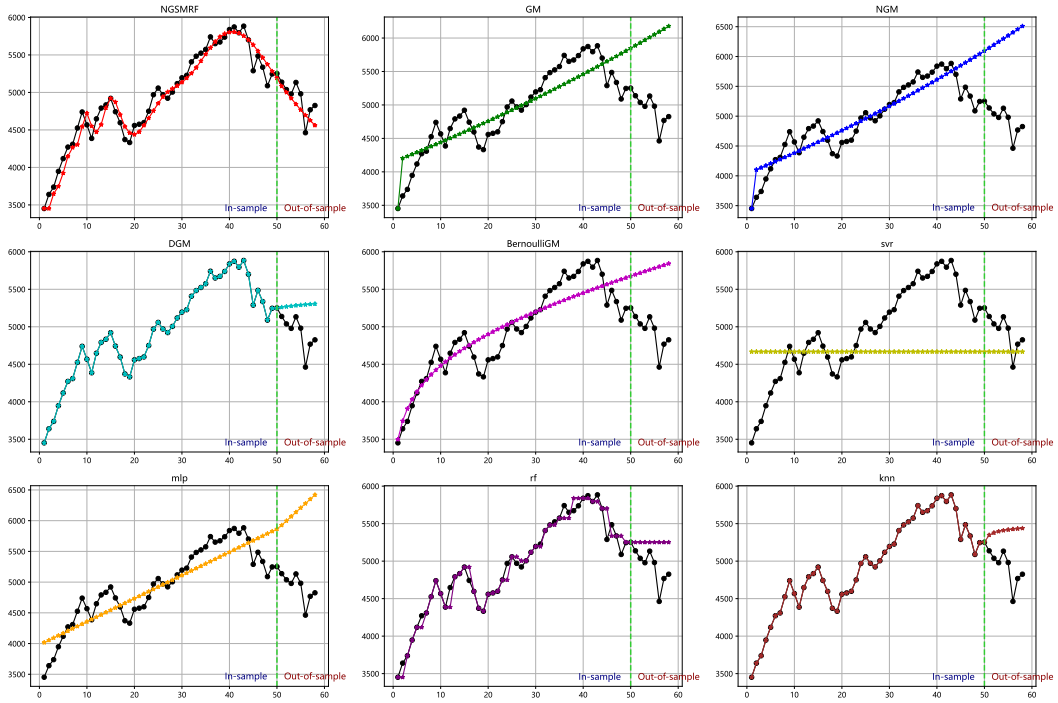

Fig. 5. Comparison of predicted annual carbon emissions for US

Table 3: MAPE for training and testing of models on the US dataset

Metric	NN Grey	GM	NGM	DGM	Bernoulli GM	SVR	MLP	RF	KNN
Train MAPE (%)	2.3073	4.9440	4.7762	**0.9822**	4.2191	11.0056	5.0608	1.0775	0.9892
Test MAPE (%)	**3.3139**	22.9998	28.9999	7.7398	17.6312	6.0064	25.9244	7.0394	10.2110

4.3 Case III: Annual Carbon Emissions in Indian

As shown in Fig. 6, India's carbon emissions have exhibited a continuous and steady upward trend from 1965 to 2022. Starting from only 167.5 million tonnes in 1965, emissions have risen to over 2500 million tonnes by 2022. This consistent increase reflects India' s rapid population growth, industrialization, and expanding energy consumption over the decades. Unlike some developed countries, there is no significant plateau or decline in the emission curve, indicating that India is still in a phase of accelerating carbon output. Understanding this long-term trend is essential for accurate modeling and forecasting of future emissions.

As shown in Table 4, the proposed NGSMRF model achieves the best overall performance on the Indian dataset, with a testing MAPE of 5.0274%, outperforming all other models. Grey models such as GM and NGM show moderate accuracy, with testing MAPE values of 9.3371% and 9.0342%, respectively. DGM, while achieving a perfect fit on the training set (0.9812%), suffers from generalization issues, resulting in a test MAPE of 13.1059%.

The SVR and KNN models demonstrate poor generalization capabilities, with testing errors exceeding 25%, and SVR in particular reaching 51.6649%, indicating significant overfitting. MLP also struggles with a high test error of 17.2737%, which may reflect training instability or a mismatch with the dataset' s temporal dynamics. These results further confirm the robustness and predictive power of the NGSMRF model across different emission patterns and national profiles.

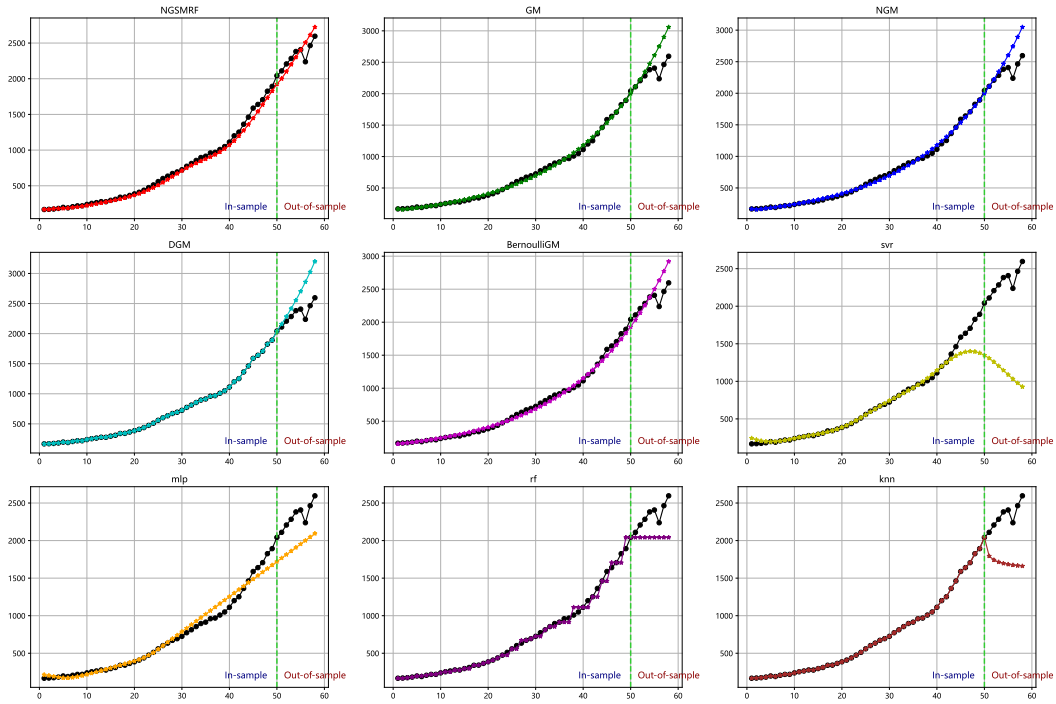

Fig. 6. Comparison of predicted annual carbon emissions for India

Table 4: MAPE for training and testing of models on India dataset

Metric	NN Grey	GM	NGM	DGM	Bernoulli GM	SVR	MLP	RF	KNN
Train MAPE (%)	4.8203	3.6871	3.7141	**0.9812**	4.3328	6.0619	7.5976	2.2730	1.0343
Test MAPE (%)	**5.0274**	9.3371	9.0342	13.1059	6.8464	51.6649	17.2737	12.2401	26.6656

4.4 Case IV: Annual Carbon Emissions in Canada

As shown in Fig. 7, Canada's annual carbon emissions have exhibited a relatively moderate upward trend from 1965 to 2022. The emissions increased steadily during the 1970s through the early 2000s, after which they began to stabilize, fluctuating within a narrower range. This pattern suggests that, following a period of industrial expansion, Canada has entered a phase of controlled emissions growth, likely influenced by environmental regulations, technological advancement, and energy structure optimization. Understanding this transitional trend is essential for developing accurate and reliable emission forecasting models.

As shown in Table 5, the proposed NGSMRF model achieves the best overall performance on the Canadian dataset, with a testing MAPE of 4.2154%, outperforming all baseline models. In contrast, traditional grey models such as GM and NGM result in high testing errors of 15.8169% and 18.5485%, respectively, due to their inherent linear structure and limited generalization capability.

Although the DGM model fits the training data perfectly (0.9833%), its test MAPE of 5.1286% suggests limited robustness. Models like SVR and RF exhibit good performance, with test errors close to that of NGSMRF (5.2571% and 4.7953%, respectively), but slightly inferior in terms of generalization. The MLP model again suffers from

overfitting, with a high test MAPE of 16.4517%. These results demonstrate the superiority of the NGSMRF model in capturing both the temporal dynamics and nonlinear patterns of Canada' s carbon emissions.

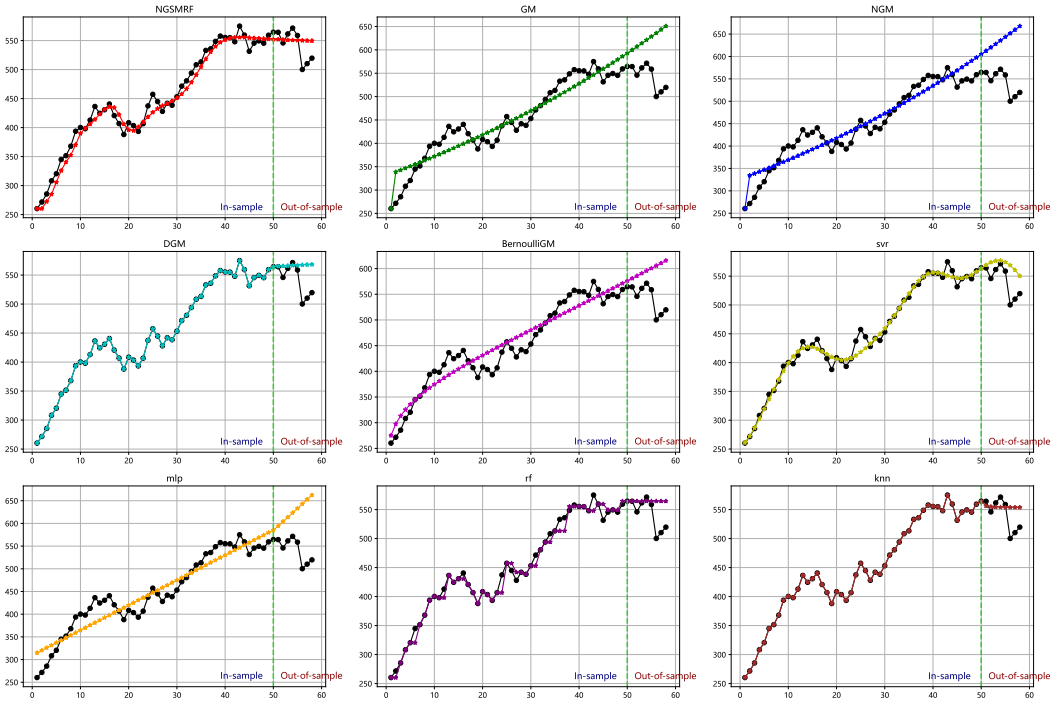

Fig. 7. Comparison of predicted annual carbon emissions for Canada

Table 5: MAPE for training and testing of models on the Canada dataset

Metric	NN Grey	GM	NGM	DGM	Bernoulli GM	SVR	MLP	RF	KNN
Train MAPE (%)	2.5621	5.4283	5.4256	0.9833	4.8105	1.4855	5.5699	1.1771	**0.2323**
Test MAPE (%)	**4.2154**	15.8169	18.5485	5.1286	10.7924	5.2571	16.4517	4.7953	4.2556

5 Discussion

The analysis of historical carbon emissions data from China, the United States, India, and Canada reveals distinct national-level trends that reflect their economic development stages and energy structures. China's emissions have shown a dramatic increase since 2000, driven by rapid industrialization, although recent years indicate a stabilization due to environmental policy interventions. The United States exhibits a plateauing and slightly declining trend after peaking in the early 2000s, suggesting a successful shift toward cleaner energy. India, on the other hand, continues on a sharp upward trajectory, with no signs of slowing, reflecting ongoing industrial expansion and energy demand. Canada shows a more moderate and stable emissions profile, with fluctuations in recent years indicating the complex balance between energy production and climate policy. These national patterns provide important context for interpreting model predictions and inform the need for country-specific forecasting strategies.

From a modeling perspective, the proposed NGSMRF model consistently achieves superior performance across all four national datasets, demonstrating both low training and testing MAPE values. Unlike traditional grey models such as GM and NGM, which suffer from limited accuracy due to their linear assumptions, NGSMRF effectively captures nonlinear temporal dependencies by integrating neural components with grey forecasting theory. While some machine learning models like RF and SVR achieve competitive training accuracy, they frequently exhibit signs of overfitting, as seen in increased testing errors. MLP-based models also demonstrate unstable generalization, likely due to difficulties in training on small or non-stationary datasets. In contrast, NGSMRF maintains robust performance across different countries and emission patterns, indicating its strong adaptability and generalization capability. These findings highlight the model' s potential as a reliable and accurate tool for long-term carbon emission forecasting.

6 Conclusion

In this study, we propose a novel neural-grey system model by integrating neural networks with the traditional grey system framework through a neural-grey whitening equation. The model is trained using the Adam optimizer with hyperparameters tuned via GridSearchCV. To evaluate its effectiveness, we compare it against four classical grey models (GM, DGM, NGM, Bernoulli GM) and four machine learning models (SVR, MLP, RF, KNN) on annual carbon emission datasets from China, the United States, India, and Canada. Results show that our model consistently achieves the lowest testing MAPE across all datasets, outperforming traditional grey models hindered by linear assumptions and machine learning models prone to overfitting. This demonstrates the model' s superior ability to capture nonlinear temporal dynamics, offering a robust and accurate tool for long-term carbon emission forecasting.

References

[1] Deng Ju-Long. Control problems of grey systems. *Systems & control letters*, 1(5):288–294, 1982.

[2] Bingjun Li, Shuhua Zhang, Wenyan Li, and Yifan Zhang. Application progress of grey model technology in agricultural science. *Grey Systems: Theory and Application*, 12(4):744–784, 2022.

[3] Song Ding. A novel discrete grey multivariable model and its application in forecasting the output value of china' s high-tech industries. *Computers & Industrial Engineering*, 127:749–760, 2019.

[4] Wenqing Wu, Xin Ma, Bo Zeng, Yong Wang, and Wei Cai. Forecasting short-term renewable energy consumption of china using a novel fractional nonlinear grey bernoulli model. *Renewable energy*, 140:70–87, 2019.

[5] Jianyu Long, Zhenzhong Sun, Panos M Pardalos, Yun Bai, Shaohui Zhang, and Chuan Li. A robust dynamic scheduling approach based on release time series forecasting for the steelmaking-continuous casting production. *Applied Soft Computing*, 92:106271, 2020.

[6] Anton Bezuglov and Gurcan Comert. Short-term freeway traffic parameter prediction: Application of grey system theory models. *Expert Systems with Applications*, 62:284–292, 2016.

[7] Deng Julong. Grey fuzzy forecast and control for grain. *J. Huazhong Univ. Sci. Technol. Med. Sci*, 2:1–8, 1983.

[8] Naiming Xie and Ruizhi Wang. A historic review of grey forecasting models. *Journal of grey System*, 29(4), 2017.

[9] Nai-ming Xie and Si-feng Liu. Discrete grey forecasting model and its optimization. *Applied mathematical modelling*, 33(2):1173–1186, 2009.

[10] Zheng-Xin Wang, Zhi-Wei Wang, and Qin Li. Forecasting the industrial solar energy consumption using a novel seasonal gm (1, 1) model with dynamic seasonal adjustment factors. *Energy*, 200:117460, 2020.

[11] Lifeng Wu, Sifeng Liu, Ligen Yao, Shuli Yan, and Dinglin Liu. Grey system model with the fractional order accumulation. *Communications in Nonlinear Science and Numerical Simulation*, 18(7):1775–1785, 2013.

[12] Jie Cui, Si-feng Liu, Bo Zeng, and Nai-ming Xie. A novel grey forecasting model and its optimization. *Applied Mathematical Modelling*, 37(6):4399–4406, 2013.

[13] Peng-Yu Chen and Hong-Ming Yu. Foundation settlement prediction based on a novel ngm model. *Mathematical Problems in Engineering*, 2014(1):242809, 2014.

[14] JIANG Ai-ping and ZHANG Qi-min. Methods and optimum of grey modeling for approximation non-homogenous and non-equidistant series. *Systems Engineering-Theory & Practice*, 34(12):3199–3203, 2014.

[15] Nai-Ming Xie, Si-Feng Liu, Ying-Jie Yang, and Chao-Qing Yuan. On novel grey forecasting model based on non-homogeneous index sequence. *Applied Mathematical Modelling*, 37(7):5059–5068, 2013.

[16] Min Xia and Wai Keung Wong. A seasonal discrete grey forecasting model for fashion retailing. *Knowledge-Based Systems*, 57:119–126, 2014.

[17] BH Yang and JS Zhao. Optimized discrete grey power model and its application. *Chinese Journal of Management Science*, 24(2):162–168, 2016.

[18] Naiming Xie. A summary of grey forecasting models. *Grey Systems: Theory and Application*, 12(4):703–722, 2022.

[19] Xin Ma and Zhi-bin Liu. The kernel-based nonlinear multivariate grey model. *Applied Mathematical Modelling*, 56:217–238, 2018.

[20] Xin Ma, Yanqiao Deng, and Minda Ma. A novel kernel ridge grey system model with generalized morlet wavelet and its application in forecasting natural gas production and consumption. *Energy*, 287:129630, 2024.

[21] Faheemullah Shaikh, Qiang Ji, Pervez Hameed Shaikh, Nayyar Hussain Mirjat, and Muhammad Aslam Uqaili. Forecasting china' s natural gas demand based on optimised nonlinear grey models. *Energy*, 140:941–951, 2017.

[22] Weijie Zhou and Lingling Pei. The grey generalized verhulst model and its application for forecasting chinese pig price index. *Soft Computing*, 24(7):4977–4990, 2020.

[23] Xinping Xiao and Huiming Duan. A new grey model for traffic flow mechanics. *Engineering Applications of Artificial Intelligence*, 88:103350, 2020.

[24] Qiang Wang and Xiaoxin Song. Forecasting china's oil consumption: a comparison of novel nonlinear-dynamic grey model (gm), linear gm, nonlinear gm and metabolism gm. *Energy*, 183:160–171, 2019.

[25] Chun-I Chen, Hong Long Chen, and Shuo-Pei Chen. Forecasting of foreign exchange rates of taiwan' s major trading partners by novel nonlinear grey bernoulli model ngbm (1, 1). *Communications in Nonlinear Science and Numerical Simulation*, 13(6):1194–1204, 2008.

[26] Jianshan Lu, Weidong Xie, Hongbo Zhou, and Aijun Zhang. An optimized nonlinear grey bernoulli model and its applications. *Neurocomputing*, 177:206–214, 2016.

[27] Chong Liu, Tongfei Lao, Wen-Ze Wu, Wanli Xie, and Hegui Zhu. An optimized nonlinear grey bernoulli prediction model and its application in natural gas production. *Expert Systems with Applications*, 194:116448, 2022.

[28] Qinzi Xiao, Mingyun Gao, Xinping Xiao, and Mark Goh. A novel grey riccati–bernoulli model and its application for the clean energy consumption prediction. *Engineering Applications of Artificial Intelligence*, 95:103863, 2020.

[29] Wenqing Wu, Xin Ma, Bo Zeng, Yuanyuan Zhang, and Wanpeng Li. Forecasting short-term solar energy generation in asia pacific using a nonlinear grey bernoulli model with time power term. *Energy & Environment*, 32(5):759–783, 2021.

[30] ZX Wang. Gm (1, 1) power model with time-varying parameters and its application. *Control Decis*, 29(10):1828–1832, 2014.

[31] Wenqing Wu, Xin Ma, Yong Wang, Wei Cai, and Bo Zeng. Predicting china' s energy consumption using a novel grey riccati model. *Applied Soft Computing*, 95:106555, 2020.

[32] Xilin Luo, Huiming Duan, and Kai Xu. A novel grey model based on traditional richards model and its application in covid-19. *Chaos, Solitons & Fractals*, 142:110480, 2021.

[33] Paul Gatabazi, Jules Clement Mba, and Edson Pindza. Fractional gray lotka-volterra models with application to cryptocurrencies adoption. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(7), 2019.

[34] Paul Gatabazi, Jules Clement Mba, Edson Pindza, and Coenraad Labuschagne. Grey lotka–volterra models with application to cryptocurrencies adoption. *Chaos, Solitons & Fractals*, 122:47–57, 2019.

[35] Shuhua Mao, Min Zhu, Xianpeng Wang, and Xinping Xiao. Grey–lotka–volterra model for the competition and cooperation between third-party online payment systems and online banking in china. *Applied Soft Computing*, 95:106501, 2020.

[36] Dajiang Lei, Kaili Wu, Liping Zhang, Weisheng Li, and Qun Liu. Neural ordinary differential grey model and its applications. *Expert Systems with Applications*, 177:114923, 2021.

[37] Junfeng Li, Aiping Yang, and Wenzhan Dai. Modeling mechanism of grey neural network and its application. In *2007 IEEE International Conference on Grey Systems and Intelligent Services*, pages 404–408. IEEE, 2007.

[38] Hao Hao, Qian Zhang, Zhiguo Wang, and Ji Zhang. Forecasting the number of end-of-life vehicles using a hybrid model based on grey model and artificial neural network. *Journal of cleaner production*, 202:684–696, 2018.

[39] Rudolf Kruse, Sanaz Mostaghim, Christian Borgelt, Christian Braune, and Matthias Steinbrecher. Multi-layer perceptrons. In *Computational intelligence: a methodological introduction*, pages 53–124. Springer, 2022.

[40] Barry J Wythoff. Backpropagation neural networks: a tutorial. *Chemometrics and Intelligent Laboratory Systems*, 18(2):115–155, 1993.

[41] Fatima Zahrae El-Hassani, Meryem Amri, Nour-Eddine Joudar, and Khalid Haddouch. A new optimization model for mlp hyperparameter tuning: modeling and resolution by real-coded genetic algorithm. *Neural Processing Letters*, 56(2):105, 2024.

[42] Tong Fa Deng, Yong Gui, and Ji Yang Yan. Prediction and analysis of tunnel crown settlement based on grey system theory. *Advanced Materials Research*, 490:423–427, 2012.

[43] Sifeng Liu and Jeffrey Yi Lin Forrest. *Grey systems: theory and applications*. Springer, 2010.

[44] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[45] Shun-ichi Amari. Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4-5):185–196, 1993.

[46] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[47] Petro Liashchynskyi and Pavlo Liashchynskyi. Grid search, random search, genetic algorithm: a big comparison for nas. *arXiv preprint arXiv:1912.06059*, 2019.

[48] Chun-I Chen. Application of the novel nonlinear grey bernoulli model for forecasting unemployment rate. *Chaos, Solitons & Fractals*, 37(1):278–287, 2008.

[49] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[50] Steven J Rigatti. Random forest. *Journal of Insurance Medicine*, 47(1):31–39, 2017.

[51] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. Knn model-based approach in classification. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, pages 986–996. Springer, 2003.