

基于GM-ARIMA耦合模型的碳排放动态预测 与绿色低碳转型路径优化研究

——以山东省的碳排放量为例进行预测分析

张海珊, 吴禹衡, 叶 础, 李 真*

广东财经大学, 统计与数学学院, 广东 广州

收稿日期: 2025年5月29日; 录用日期: 2025年6月10日; 发布日期: 2025年8月18日

摘 要

本文聚焦于农业生产材料投入能力与碳排放的研究, 以山东省为例, 对山东省碳排放量进行预测分析。首先, 建立了碳排放量的灰色时间序列预测模型, 并使用粒子群优化算法(PSO), 对该模型的灰作用量、发展系数两个参数进行优化。接着, 利用自回归积分滑动平均算法(ARIMA)分析预测模型的残差, 并通过网格搜索法优化ARIMA模型的 p 、 d 、 q 参数。最后, 利用ARIMA模型的残差预测结果对灰色预测模型进行补偿。本文创新性地提出了灰色理论-自回归滑动平均模型复合模型(GM-ARIMA), 旨在充分发挥灰色理论在处理不确定性和不完全信息方面的独特优势, 同时结合ARIMA模型在时间序列分析中的强大预测能力来消除模型残差, 从而为农业碳排放量的预测提供一个可靠的基础模型。

关键词

GM-ARIMA, 碳排放量测算, 耦合度分析, PSO, 网格搜索

Research on Carbon Emission Dynamic Prediction and Green and Low-Carbon Transformation Path Optimization Based on GM-ARIMA Coupling Model

—Conducting a Predictive Analysis Based on the Carbon Emissions of
Shandong Province

Haishan Zhang, Yuheng Wu, Chu Ye, Zhen Li*

School of Statistics and Mathematics, Guangdong University of Finance and Economics, Guangzhou Guangdong

*通讯作者。

文章引用: 张海珊, 吴禹衡, 叶础, 李真. 基于 GM-ARIMA 耦合模型的碳排放动态预测与绿色低碳转型路径优化研究[J]. 低碳经济, 2025, 14(3): 269-286. DOI: 10.12677/jlce.2025.143029

Abstract

This article focuses on the research of agricultural production materials input capacity and carbon emissions, using Shandong Province as a case study to predict and analyze the carbon emission levels in the region. Firstly, a gray time series prediction model for carbon emissions was established, and two parameters of the model—gray influence amount and development coefficient—were optimized using the Particle Swarm Optimization (PSO) algorithm. Subsequently, the Autoregressive Integrated Moving Average (ARIMA) algorithm was employed to analyze the residuals of the prediction model, followed by parameter optimization of the ARIMA model through grid search method. Finally, the residual predictions from the ARIMA model were used to compensate for the gray prediction model. This article innovatively proposes a composite model of Gray Theory and Autoregressive Moving Average (GM-ARIMA), aiming to fully leverage the unique advantages of gray theory in handling uncertainty and incomplete information, while integrating the powerful predictive capability of the ARIMA model in time series analysis to eliminate model residuals, thereby providing a reliable foundational model for predicting agricultural carbon emissions.

Keywords

GM-ARIMA, Carbon Emission Measurement, Coupling Degree Analysis, PSO, Grid Search

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

中国在应对气候变化和碳排放的区域性差异方面采取了一系列政策措施，在《国务院关于印发 2030 年前碳达峰行动方案的通知》中，明确指出我国将推动产业结构和能源结构的调整与优化旨在减少温室气体排放并促进绿色转型。

农业作为第二大碳排放源，占全球人为碳排放量的 23% [1]。作为农业大国，中国的农业碳排放量约占全国温室气体排放总量的 16%至 17% [2]。农业碳排放主要源自农业活动中使用的化肥、农药、燃料及其他生产材料的消耗。研究表明，农业领域的碳排放不仅对生态环境的可持续发展产生深远影响，也直接加剧了全球气候变化的挑战。在气候变化日益严峻的背景下，控制农业生产过程中的碳排放，预测碳排放量以制定相关政策，已成为各国政府和科学界关注的重点。

农业碳排放量的统计并不是直接得出，选择合适的指标对碳排放量进行测算至关重要。根据研究[3]的观点，在评估农业碳排放时，仅依赖直接的用碳强度计算并不足以全面反映真实的环境影响，因为这无法体现各要素之间的替代作用，能源的使用必须与其他生产要素结合。因此，单纯依靠碳的直接使用量来评估碳排放，可能会低估农业生产对环境的真实影响。随着研究的深入，农业碳源的界定逐渐完善，包括畜禽养殖[4]、农业薄膜[5]等也被纳入测算范围。特别是崔永福等[6]的研究，基于化肥、农药、农用塑料薄膜、农用柴油、土地翻耕和农业灌溉的六个碳源。

本文在以上研究基础上，采用农业生产材料的间接计算方法碳排放，系统分析农业生产活动的碳排放特征。同时，结合智能优化算法与复合预测模型(GM-ARIMA)，对 2000 年~2021 年间各省的农业碳排放进行深入研究。通过对国内 31 个省市的碳排放效率进行比较与评价，进一步探讨在“十四五”期间，

碳减排的约束性指标是否能够实现。这一研究不仅为国内部分地区的低碳发展提供了理论支持，也为政策制定者在推动绿色发展与可持续农业方面提供了科学依据。

2. 建立评价指标体系

2.1. 数据来源

本文数据均来自《中国农村统计年鉴》和国家统计局国家数据网 2000 年~2021 年统计数据，包含农业种植中使用的化肥、塑料薄膜、农用柴油、农药使用量以及种植面积与灌溉面积。本研究针对以上农业生产材料投入所产生的碳排放量作为农业碳排放量。

2.2. 农业生产要素的综合评价体系

2.2.1. 评价指标的选取

农业生产要素的评价指标主要反映农业生产材料的投入能力。为此，从使用量和作用面积两个方面选取了相关的评价指标，构建农业生产要素评价指标体系如下。其中，使用量方面的作用方向为“正向(+)”，作用面积的作用方向为“正向(+)”，见表 1。以化肥使用量、塑料薄膜使用量、农用柴油使用量、农药使用量、农作物播种面积、灌溉面积六大生产要素指标，对各省份农产品生产使用方式进行评价。其中，化肥、塑料薄膜、农用柴油、农药均以实际使用量为准；农作物以实际播种面积为准；灌溉均以有效灌溉面积计算。

Table 1. Agricultural production factor indicators

表 1. 农业生产要素指标

生产要素	二级指标	作用方向	单位
使用量	化肥	+	万吨
	塑料薄膜	+	吨
	农用柴油	+	万吨
	农药	+	吨
作用面积	农作物播种	+	千公顷
	灌溉	+	千公顷

2.2.2. 熵权-TOPSIS 综合评价

本文引入熵权法与 TOPSIS 法相结合的多指标综合评价方法。熵权法能够客观地从生产要素指标中提取各评价指标的信息量，合理分配权重，避免主观偏差，使得权重设置更具科学依据。结合 TOPSIS 法，则能基于指标权重，系统地评估各预测方案或模型的接近理想解的程度，实现对生产要素多维评价指标的排序和优劣判断。

首先，进行数据预处理除去异常数据，建立评价指标体系。接着，通过熵权法得到六个指标权重。将数据先正向化、再标准化后，用熵权法计算各指标的权重。

$$w_j = \frac{d_j}{\sum_{j=1}^m d_j}, j = 1, \dots, m.$$

式中， d_j 表示信息熵的冗余度差异值。由熵权法计算的结果如下，见表 2。

Table 2. The weights of each indicator calculated by entropy weight method

表 2. 熵权法计算所得各指标权重

	化肥	塑料薄膜	农用柴油	农药	播种面积	灌溉面积
权重	0.153257	0.158005	0.198704	0.168663	0.144699	0.176673

将熵权法和 TOPSIS 法相结合, 使用变权 TOPSIS 模型对各省市种植业碳排放量计算综合得分。建立目标方案评价模型的决策矩阵并进行相应的标准化处理, 得到矩阵 S 。

$$S = (s_{ij})_{m \times n} = \left(\frac{y_{ij}}{\sqrt{\sum_{i=1}^m y_{ij}^2}} \right)_{m \times n} .$$

根据确定的各评价指标权重系数, 构建加权标准决策矩阵 U :

$$U = (u_{ij})_{m \times n} = w_j (s_{ij})_{m \times n} .$$

确定正理想值 U^+ 和负理想值 U^- , 以正理想值为例:

$$U^+ = \left\{ \begin{array}{l} \left\{ \max_{1 \leq i \leq m} u_{ij} \mid j \in J^+, \left(\min_{1 \leq i \leq m} u_{ij} \mid j \in J^- \right) \right\} \\ \left\{ u_1^+, u_2^+, \dots, u_n^+ \right\} \end{array} \right\} .$$

式中, J^+ 为效益型评价指标集, J^- 为成本型评价指标集。

求任一解 y_{ij} 到正负理想的距离 L_i^+ 、 L_i^- , 以正理想的距离为例:

$$L_i^+ = \sqrt{\sum_{j=1}^n (u_{ij} - u_j^+)^2} .$$

式中, u^+ 、 u^- 分别表示正负理想解的第 j 个分量。

计算各方案到理想解的贴进度 G_i :

$$G_i = \frac{L_i^-}{L_i^- + L_i^+}, i = 1, 2, \dots, m .$$

根据相对贴进度 G_i 对各个方案进行排序, 选择贴进度最大的方案为较优方案。最终, 得到 2000 年~2021 年全国种植业碳排放量的最大的省是山东省, 其次为河北省、河南省。选定山东省为预测未来碳排放量的研究对象。

2.2.3. 建立耦合度模型

耦合度模型通常用来衡量系统内部子系统或变量间的关联和协调程度, 用于揭示变量间内在联系和动态互动, 而使用耦合度模型可以再次验证熵权-Topsis 模型中生产要素评价的可行性。

因此, 本文采用耦合度模型对农业生产要素的使用量与作用面积进行耦合分析, 以检验这两项指标之间的协同关系。本文依据王淑佳等[7]对耦合协调的定义, 将农业生产要素使用量与作用面积的耦合协调性定义为二者通过相互作用以达到良好的协同状态。构建的耦合度模型为:

$$C = \frac{1}{1 + e^{-k(A+B-\theta)}} .$$

式中, C 为耦合度, 表示两个模块之间的耦合程度, 取值范围在 (0,1] 之间。 A 为标准化后的生产要素使用量指标, B 为标准化后的作用面积指标。 k 为调节参数, 用于函数的斜率, 设置为 1。 θ 为阈值参数, 用于调整函数的平移, 设置为 0。

基于变权 TOPSIS 模型可得综合评价最高得分为山东省, 将其六个农业生产指标代入, 得到山东省使用量要素、面积要素及相应耦合度, 见表 3。

分析表 3 发现山东省 2000 年~2021 年耦合度均值为 0.957, 依据朱智滔等的耦合度类型[8]划分标准表, 见表 4。说明山东省农业生产使用量要素与作用面积要素耦合度高度协调, 证明了熵权-Topsis 法综合评价的可行性。

Table 3. Coupling degree analysis table of Shandong Province from 2000 to 2021

表 3. 山东省 2000 年~2021 年份耦合度分析表

年份	使用量 标准化均值	作用面积 标准化均值	耦合度	年份	使用量 标准化均值	作用面积 标准化均值	耦合度
2000	0.458	0.346	0.909	2011	0.389	0.832	0.984
2001	0.561	0.458	0.951	2012	0.457	0.806	0.984
2002	0.354	0.64	0.963	2013	0.422	0.769	0.981
2003	0.193	0.704	0.961	2014	0.444	0.702	0.976
2004	0.006	0.704	0.944	2015	0.494	0.656	0.974
2005	0.104	0.821	0.97	2016	0.876	0.615	0.985
2006	0.144	0.967	0.985	2017	0.77	0.513	0.973
2007	0.138	0.965	0.984	2018	0.788	0.376	0.956
2008	0.189	0.868	0.979	2019	0.709	0.236	0.914
2009	0.237	0.816	0.977	2020	0.696	0.143	0.877
2010	0.323	0.85	0.983	2021	0.742	0.073	0.855

Table 4. Coupling type classification standard table

表 4. 耦合度类型划分标准表

<i>D</i>	类型	<i>D</i>	类型
$0.900 \leq D \leq 1.000$	优质协调	$0.400 \leq D \leq 0.500$	濒临协调
$0.800 \leq D \leq 0.900$	良好协调	$0.300 \leq D \leq 0.400$	轻度协调
$0.700 \leq D \leq 0.800$	中级协调	$0.200 \leq D \leq 0.300$	中度协调
$0.600 \leq D \leq 0.700$	初级协调	$0.100 \leq D \leq 0.200$	严重失调
$0.500 \leq D \leq 0.600$	勉强协调	$0.000 \leq D \leq 0.100$	极度失调

2.3. 碳排放量测算

种植业碳排放主要分为农地利用相关的柴油、化肥、农药、农膜、灌溉、翻耕所生产的碳排放和农田土壤利用过程中农作物等温室气体排放。查阅相关文献确定种植业主要碳排放源和碳排系数，见表 5。

Table 5. Main carbon emission sources and carbon emission coefficients of the planting industry

表 5. 种植业主要碳排放源和碳排系数

碳源	碳排放系数	参考来源
柴油	0.59 kg/kg	IPCC2013 [9]
化肥	0.89 kg/kg	美国橡树岭国家实验室[10]
农药	4.93 kg/kg	美国橡树岭国家实验室[10]
农膜	5.18 kg/kg	南京农业大学农业资源与生态环境研究所[11]
灌溉	266.48 kg/hm ²	段华平等[12]
翻耕	312.60 kg/km ²	李波等[13]

借鉴联合国政府间气候变化专门委员会(IPCC)因子系数法，计算公式如下：

$$C = \sum C_i = \sum T_i * \varepsilon_i.$$

式中， C 表示种植业碳排放总量； C_i 表示第 i 类碳源的碳排放量； T_i 表示第 i 类碳源的投入量； ε_i 表示第 i 类碳源的排放系数。

将化肥使用量、塑料膜使用量、农用柴油使用量、农药使用量、农作物播种面积、灌溉面积，根据因子系数法转化为碳排放量。表 6 为使用六类不同的生产方式所转化为碳排放量的数值。

Table 6. Carbon emissions of various factors in Shandong Province from 2000 to 2021 (10,000 tons)

表 6. 山东省 2000 年~2021 年各因素的碳排放量(万吨)

年份	化肥	薄膜	柴油	农药	翻耕	灌溉
2000	376.6480	116.6122	92.6890	69.1684	3.4846	128.5729
2001	381.4540	133.4943	95.2850	71.4850	3.5218	128.8724
2002	386.1710	151.3467	99.7690	80.7233	3.4535	127.8422
...
2020	339.0010	137.6114	74.5760	56.3553	3.4039	141.0639
2021	330.1900	134.5132	72.0980	53.3574	3.4225	141.0639

从表 6 可得，种植业碳排放主要以化肥碳排放量最大，依次为灌溉、塑料薄膜等。使用翻耕的生产方式其碳排放量最少。

3. 碳排放量的数据分析

3.1. 不同省份的农业碳排放量

在分析不同省份 2000 年~2021 年的农业碳排放数据时，以山东省、河北省和河南省为例，见图 1。

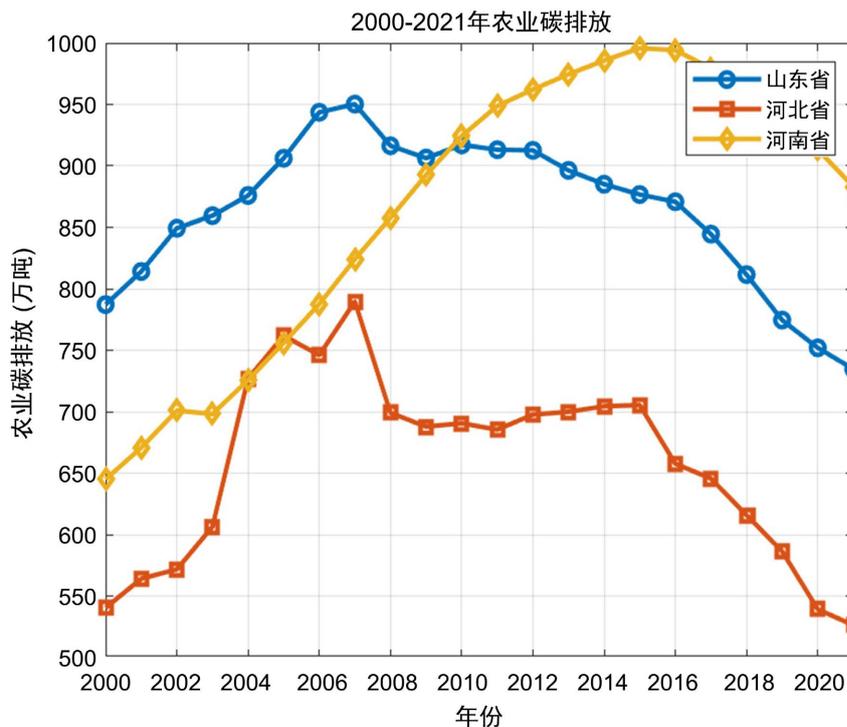


Figure 1. Agricultural carbon emissions time series

图 1. 农业碳排放时序图

山东省在过去 21 年中的农业碳排放呈现明显波动，整体排放水平较高，初期的增长随后趋向逐步下降，但近年来的减排趋势相对明显。河北省整体碳排放水平较低、涨幅不高；河南省表现出较高的排放水平和增长幅度。

3.2. 不同生产原料的碳排放量

基于上述分析，在分析不同生产原料的碳排放量数据时，选取综合评分最高的山东省对不同生产原料的碳排放量进行综合分析。

首先，从化肥排放方面，排放量介于 376.65~445.27 万吨，整体呈现出上升趋势。表明随着农业生产的增加以及对化肥使用的依赖加剧，碳排放量逐年攀升，反映出当前农业生产模式中对化肥的高依赖性，见图 2。

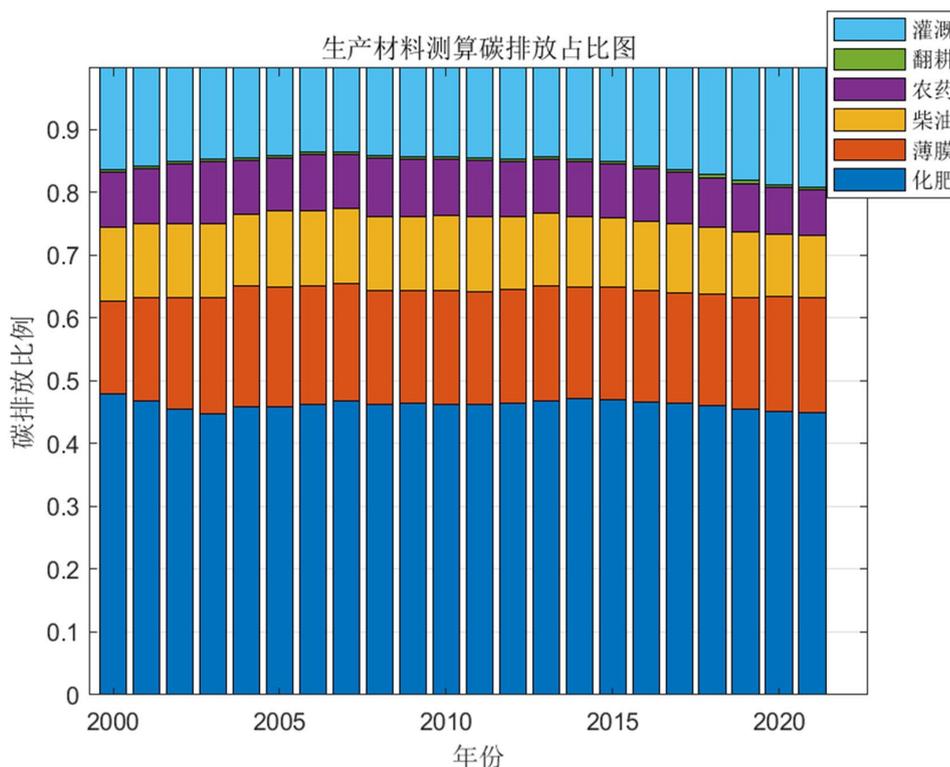


Figure 2. Annual trend chart of carbon emissions calculated by production materials

图 2. 生产材料测算碳排放占比年份趋势图

其次，在薄膜碳排放方面，排放量介于 116.61~176.74 万吨，波动幅度较大，整体趋势向上。这可能与农业中塑料薄膜使用的频率和种类增加有关，尤其是在保护性农业和温室种植中，薄膜的应用成为提升产量的重要手段。在柴油排放方面，排放量介于 72.10~113.58 万吨，显示出柴油在农业机械化进程中的重要性。随着农业机械化程度的提高，柴油需求日益增加，进而导致相关碳排放上升。在农药排放方面，排放量介于 53.36~84.46 万吨，表明农药的使用对碳排放也有一定的贡献，尤其是在病虫害防治中，农药的使用量直接影响到碳排放水平，见图 2。

最后，在翻耕排放方面，其碳排放量相对较小，介于 3.33~3.53 万吨，说明翻耕过程对整体碳排放的影响较小。在灌溉排放方面，排放量介于 127.03~141.06 万吨，显示出灌溉系统的能源消耗在农业生产中依然占据重要地位，尤其是在干旱地区或水资源匮乏的情况下，灌溉的能源需求会显著增加。

4. 碳排放量预测模型的建立

4.1. GM-ARIMA 时间序列预测算法

基于农业碳排放的特性，提出 GM-ARIMA 智能混合预测算法。以灰色理论为基础，灰色理论是一种适用于处理不确定性和不完全信息的数学模型，能够有效处理不确定性和不完全信息，可以为农业碳排放的动态变化提供一个合理的基础模型。同时，结合 ARIMA 模型在时间序列分析中的强大预测能力，通过系统地分析和消除模型残差，进一步提升预测的准确性。最后，结合 PSO 和网格搜索法，分别优化了灰色模型和 ARIMA 模型的参数，更大程度地捕捉数据中的趋势，为碳排放的预测提供更加精确的补充，见图 3。总的来说，GM-ARIMA 智能混合预测算法具有以下优势：

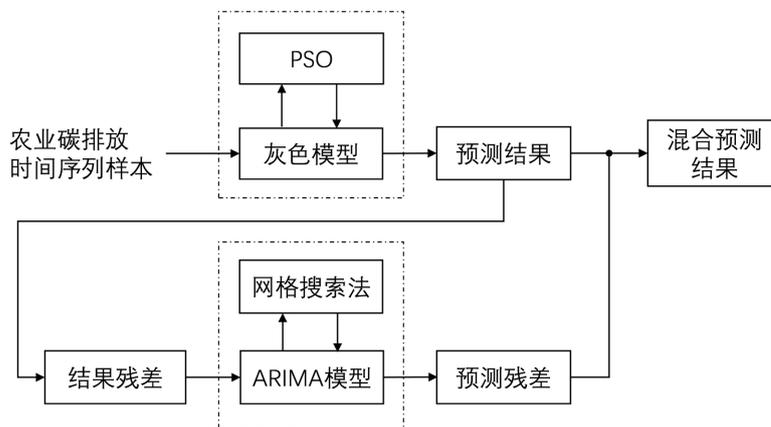


Figure 3. Diagram of the prediction algorithm principle
图 3. 预测算法原理图

- (1) 灰色模型用于捕捉农业碳排放数据的长期趋势和系统演化规律，提供趋势预测的基础；
- (2) ARIMA 模型针对灰色模型预测残差进行建模，拟合残余的随机波动，提高预测精度和稳定性；
- (3) 两者结合通过“趋势 + 波动”的分解思路，发挥各自优势可以实现更全面和准确的碳排放预测。

4.2. 灰色预测模型

农业碳排放时间序列的波动性、非线性和空间异质性，以及通常较少的样本量，对预测方法提出了较高的要求。灰色系统理论由邓聚龙教授于 1982 年提出，是一种应对信息不完整的新数学方法，特别适用于“小样本”和“贫信息”系统，其核心在于通过建立灰色模型进行描述和预测。灰色系统的生成方法包括累加生成、累减生成和均值生成等。其中，累加生成能够反映变量的累积变化，每个新序列的元素是原序列中对应位置及之前元素的累加和，从而揭示原始数据中的隐藏特征。GM (1, 1)模型作为灰色预测理论的核心，基于累加生成的方法，通过对已知信息的处理，提取有价值的信息，准确把握系统的运行行为及演化规律。

4.2.1. 灰色预测模型 GM (1, 1)的实现

基于历史数据对时间序列数据进行预测，本文采用灰色预测模型 GM (1, 1)对 2000 年至 2021 年山东省种植业碳排放量和碳排放进行模拟[14]，计算步骤如下：

Step 1: 将碳排放量作为原始数据，计算原始数列 $x^{(0)}$ 的级比：

$$\lambda^{(0)}(t) = \frac{x^{(0)}(t-1)}{x^{(0)}(t)}$$

式中: $t=1,2,\dots,n$ 。若 $\lambda^{(0)}(t) \in \left(\frac{-2}{e^{n+1}} \right)$, 则利用原始数列建立灰色预测模型 GM(1, 1)。

Step 2: 构建碳排放量微分方程如下:

$$\frac{d(x^{(1)}(t))}{dt} + ax^{(1)}(t) = \mu.$$

式中: $x^{(1)}(t)$ 为累加数列, a 为发展系数, μ 为内生控制系数。利用最小二乘法求解灰色参数 a 和 μ , 式中 $\begin{bmatrix} a \\ \mu \end{bmatrix}$ 为代估计的参数向量:

$$\begin{bmatrix} a \\ \mu \end{bmatrix} = (B^T \times B)^{-1} \times B^T \times y_n.$$

Step 3: 将灰色参数代入时间函数:

$$x^{(1)}(t) = \left[x^{(0)}(t) - \frac{\mu}{a} \right] e^{-a(t-1)} + \frac{\mu}{a}.$$

碳排放量预测值如下, 式中 $t=1,2,\dots,n-1$:

$$x^{(0)}(t+1) = x^{(1)}(t+1) - x^{(1)}(t).$$

4.2.2. 基于 PSO 算法优化 GM(1, 1)模型参数

在传统的 GM(1, 1)算法中, 发展系数和灰色参数是通过最小二乘确定参数。然而, 最小二乘法在处理非线性问题或数据噪声比较大, 可能会导致参数估计的不准确。本文提出采用粒子群优化算法来优化 GM(1, 1)的模型的参数。PSO 是一种基于群体智能的优化算法, 通过模拟鸟群觅食行为来寻找最优解。通过引入 PSO, 可以在优化发展系数和灰作用量的过程中, 考虑到数据的非线性特征和不确定性, 从而提高模型的预测精度。

适应度函数代表模型适应的能力, 本文适应度函数以未来时刻碳排放量时间序列预测值与实际值的偏差值最小, 以下式为性能指标。

$$\begin{cases} |e(t)| = |v_s(u) - v| \\ J = \int_0^{t_\infty} |e(t)| dt \end{cases}.$$

式中, t_∞ 为稳态时间, $e(t)$ 为预测模型输出的未来时刻金融时间序列值应与实际的值偏差。 $v_s(u)$ 为该组参数下时间序列灰色模型拟合的预测值, v 为实际的测量值。PSO 的学习目标是使 J 最小。

4.2.3. PSO 算法中改进的学习因子和惯性权重

在传统的 PSO 模型中, 学习因子 c_1 、 c_2 决定了粒子向自身最优和全局最优靠近的权重。设置学习因子将大幅度地提升寻找最优策略的效率。文献[15]指出 c_1 是粒子自我学习能力的表现, c_2 是粒子向群体学习能力的表现。初期应关注个体自我学习能力(c_1), 后期则应该注重粒子向群体学习的能力(c_2)。改进的学习因子设定为:

$$\begin{cases} c_1 = \sin \omega \\ c_2 = 1 - \sin \omega \end{cases}.$$

在传统的 PSO 模型中, 惯性权重 ω 的大小反映当前粒子受上一代粒子的影响程度, 是 PSO 算法的重要参数之一。文献[16]中采用了非线性递减惯性权重, 对寻找最优策略效果良好。因此, 改进后惯性权重设定为非线性递减惯性权重:

$$\omega = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \cos \left(\tan \left(1 - \left(\frac{k}{T_{\max}} \right)^2 \right) \right)$$

式中, T_{\max} 为算法的最大迭代次数; k 为当前迭代次数。

4.3. 基于 ARIMA 的残差消除

ARIMA 模型是在原有 AR(自回归)模型、MA(移动平均)模型的基础上,通过差分运算来对平稳时间序列 $\{x_t\}$ 进行拟合的一种预测方法,其基本结构为:

$$\begin{cases} \varnothing(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0 \\ D(\varepsilon_t) = \sigma_\varepsilon^2 \\ E(\varepsilon_t, \varepsilon_s) = 0, s \neq t \\ E(x_t, \varepsilon_s) = 0, \forall s < t \end{cases}$$

式中, B 为延迟算子; $\varnothing(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$, 为平稳可逆 ARMA (p, q) 模型的自回归系数多项式; $\nabla^d = (1 - B)^d$; $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$, 为平稳可逆 ARMA (p, q) 模型的移动平均系数多项式。ARIMA 模型具有三个参数,分别为自回归项阶数 p 、差分阶数 d , 移动平均阶数 q , 因此模型可以记为 ARIMA (p, d, q) 。参数 p 和 q 分别决定了模型里的 AR 与 MA 的复杂度, d 决定得到一个平稳序列所需要的差分运算次数。

在时间序列预测中, GM(1, 1)模型常用于捕捉序列中的趋势特征。然而,由于传统 GM(1, 1)模型是基于“累计生成操作(AGO)”生成新序列的思想,对于可能存在着的复杂数据,拟合结果可能存在较大的残差,即实际观测值与模型预测值之间的差异。残差的存在可能表明模型未能充分捕捉数据中的某些特征,或者存在未建模的因素。因此,基于 ARIMA 的残差消除成为提高模型预测精度的重要步骤。

5. 实证分析

为了验证算法的有效性,分析算法结果,选取 2000 年~2021 年山东省农业材料换算碳排放量基础数据,用于碳排放量时间序列进行实证分析。

5.1. 确定 ARIMA 的参数值

在用粒子群优化算法对灰色预测模型进行寻优时,超参数的设置是寻优结果是否为全局最优的关键。为找到全局最优解,上述的超参数,与静态值相比,动态变化的超参数寻优效果会更好,根据文献[11]对惯性权值采用线性递减,个体学习因子 c_1 随搜索进程递减,社会学习因子 c_2 随搜索进程递增,得到最终的超参数,见表 7:

Table 7. Hyperparameter settings

表 7. 超参数的设置

惯性权重 ω	个体学习因子 c_1	社会学习因子 c_2
0.8	1.5	0.5

随后是对自回归积分滑动平均(ARIMA)模型中,自回归(AR)项的阶数 p ,反映了过去数据对当前数据的影响。差分阶数 d ,表示对原始时间序列进行几次差分以使其平稳。滑动平均(MA)项的阶数 q ,表示当前值与前 q 个误差项之间的关系。这三个参数共同决定了 ARIMA 模型的结构,将参数选择范围限定在 $[0, 5]$,通过遍历求解,得出最小的贝叶斯信息准则 BIC。

5.2. 模型求解

5.2.1. PSO 优化 GM 模型的实现步骤

Step 1: 设置参数 c_1 、 c_2 、 ω 随机产生粒子群初始位置和速度, 计算初始位置的适应度值, 将初始位置作为各粒子当前的最佳位置并挑选出群体的最佳位置。

Step 2: 更新惯性权重、学习因子、粒子位置及适应度值。将更新后的各粒子位置适应度值与群体最佳适应度进行比较, 若当前位置适应度值更小, 则保留为最佳位置。

Step 3: 使用交换操作对长期处于最佳位置的粒子进行扰动, 便于粒子跳出局部最优。

Step 4: 更新整个粒子群体的最佳位置、适应度值。判断当前迭代次数是否达到最大迭代次数。若达到最大迭代次数, PSO 算法中止。

Step 5: 若 PSO 算法终止, 将发展系数和灰作用量输入 GM (1, 1)。基于碳排放量数据 GM (1, 1) 利用最优发展系数和灰作用量进行拟合, 建立模型输出拟合效果。

5.2.2. ARIMA 的残差消除实现

Step 1: 对原始残差时间序列进行差分平稳化处理, 消除趋势和季节性因素的影响, 确保数据的均值和方差保持恒定。并作零均值化处理, 使数据围绕零均值分布。

Step 2: 利用自相关函数(ACF)和偏自相关函数(PACF)对处理后的数据进行分析。确定 ARIMA 模型中自回归(AR)和移动平均(MA)部分的阶数。

Step 3: 模型参数的优化。使用网格搜索方法对模型参数进行优化。通过计算不同参数配置下的赤池信息量准则(AIC)和贝叶斯信息量准则(BIC), 从中选择最小值, 以此来确保所选模型的拟合效果最佳。

Step 4: 构建残差预测模型。利用建立 ARIMA 模型, 导出残差预测模型。该模型将应用于灰色预测模型 GM (1, 1), 以实现农业碳排放的综合预测。

5.3. 模型检验

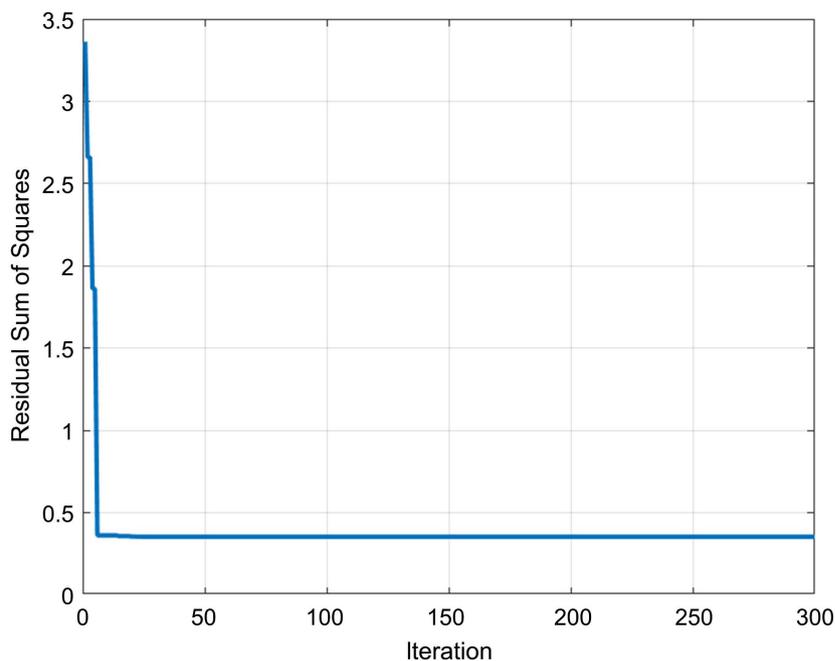
在对碳排放量样本数据进行 GM (1, 1)模型分析时, 发现小于 0.5 的数据占比高达 89.47%, 见图 4。这一结果表明, 大多数碳排放量数据均集中在较低的水平上, 说明通过准指数规律检验, 进一步验证了 GM(1, 1)模型的合理性, 同时, 在应用粒子群优化算法的过程中, 我们观察到当迭代次数达到第 6 次时, 粒子间的空间残差已经基本降低至最小, 见图 4。这一现象表明, 粒子群在搜索过程中逐渐聚集到最优解附近, 算法的收敛速度获得了显著提升。通过这种迭代优化, 粒子群算法有效地调整了发展系数和灰作用量, 使得 GM (1, 1)模型的预测能力得到了进一步增强。

由于碳排放量时间序列的非平稳性, 灰色预测模型难以做到残差消除, 采用网格搜索修正的 ARIMA 模型对灰色模型的残差进行预测, 见表 8。

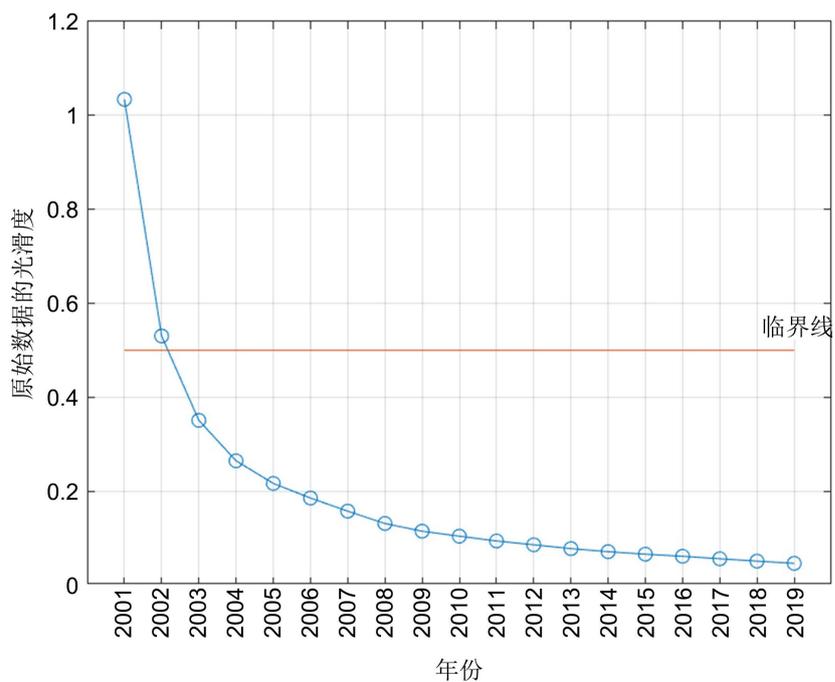
Table 8. BIC Values from grid search

表 8. 网格搜索 BIC 值

$p \backslash q$	0	1	2	3	4	5
0	212.2008	227.3911	236.1820	242.4822	247.3704	247.8580
1	229.1237	226.6218	240.6482	243.5876	247.2949	250.0321
2	254.4298	232.8678	231.7138	229.5962	232.2626	235.4297
3	259.4724	232.4877	234.8530	228.7861	233.0771	226.3210
4	266.0776	230.7611	225.9684	229.0294	228.1017	227.9684
5	267.2865	233.7641	225.8896	228.0892	238.8277	231.8677



(a)



(b)

Figure 4. Fitness Curve of Carbon Emissions (a) and Smoothness Curve of the Carbon Emission Sequence (b)
图 4. 碳排放量的适应度曲线图(a)和碳排放序列平滑度曲线图(b)

根据灰色模型的残差时间序列，进行 ARIMA 计算，确定差分阶为 $d = 1$ ，根据自相关协方差(ACF)和偏自相关协方差(PACF)分析。确定 BIC 为 226.321， $p = 5$ ， $q = 3$ 从而确定最终模型为 ARIMA(5, 1, 3)。

当自回归阶数 p 为 5，滑动平均阶数 q 为 3 时，PACF 和 ACF 的 95% 滞后自相关系数均位于置信区间内。结果表明，模型的残差展现出白噪声特性。残差序列并未表现出显著的自相关性，从而说明模型

已成功捕捉了数据中的时间序列特征。

为了进一步确认模型在残差预测方面的能力，采用 Ljung-Box 检验，见下表 9。

Table 9. Model statistical data
表 9. 模型统计数据

模型	模型拟合度统计			Ljung-Box statistic			离群值数
	平稳 R^2	R^2	BIC	统计	DF	显著性	
ARIMA	0.565284	0.890071	226.321	14.08341	10	0.169221	0

通过对残差序列的深入分析，发现检验结果的 p 值为 0.16，显著高于常用的显著性水平 0.05。这一结果表明，可以接受残差为白噪声的假设，进一步增强了模型有效性。此外，Ljung-Box 检验的 R^2 为 0.89。显示出模型的良好拟合程度，表明模型能够有效解释数据的变异性。ARIMA (5, 1, 3)模型对碳排放量残差具有很好的拟合效果，见图 5。

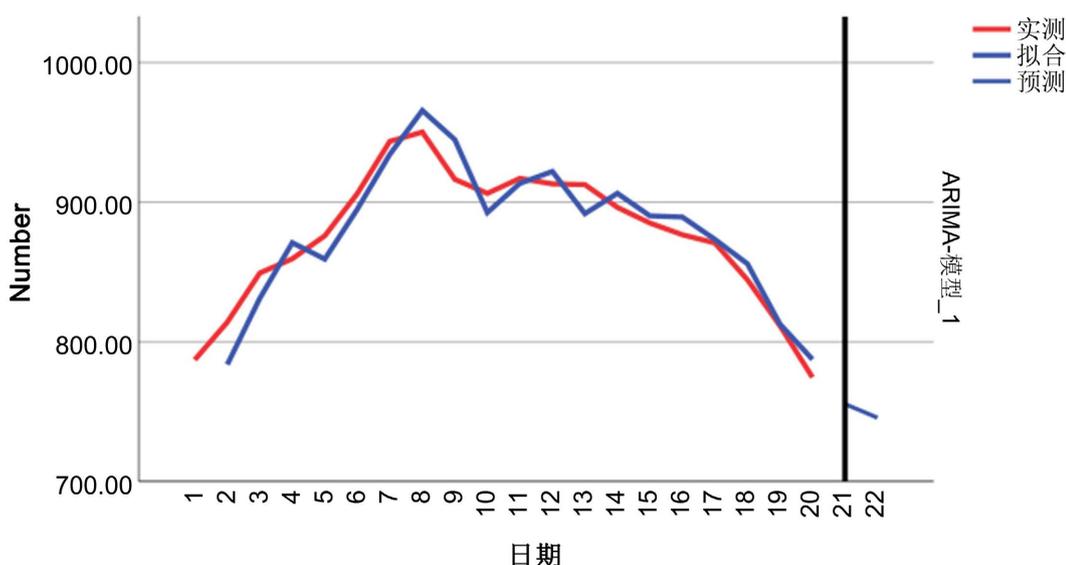


Figure 5. ARIMA prediction residual plot
图 5. ARIMA 预测残差图

结合 ARIMA 的灰色预测模型最终的拟合图，见图 6。可见得对残差预测时，基于 ARIMA 方法能根据为数不多的样本数据进行有效辨识，建立精度较高的预测模型，尤其是序列后期，拟合值几乎贴合了预测值。其中模型拟合度达到 85.2%，残差平均绝对值为 0.52，说明 GM-ARIMA 模型在碳排放量的预测有着十足的潜力。

5.4. 实验结果对比分析

为了说明基于 GM-ARIMA 的碳排放量时间序列智能混合预测算法的优势，在选用相同的样本和预测样本的条件下，分别采用单纯的 ARIMA 滑动平均自回归模型、单纯的灰色预测 GM (1, 1)模型、支持向量机(SVM)、神经网络(BP)进行对比试验，检验结果分别见图 7、图 8 所示。

由图 7(a)可知，传统的 ARIMA 模型对不平稳序列具有良好的拟合能力，ARIMA 模型通过自回归和滑动平均的组合，能够有效捕捉碳排放量数据中的趋势和季节性特征，且其预测结果与实际观测值之间

的偏差较小，反映出模型对数据的适应性。

由图 7(b)可知，GM (1, 1)模型在处理小样本和不确定性方面具有优势。但在面对复杂的碳排放数据时，单独使用 GM (1, 1)模型其预测精度相对较低，尤其是在数据波动较大时，模型的拟合效果明显不足，导致预测结果存在一定的偏差。

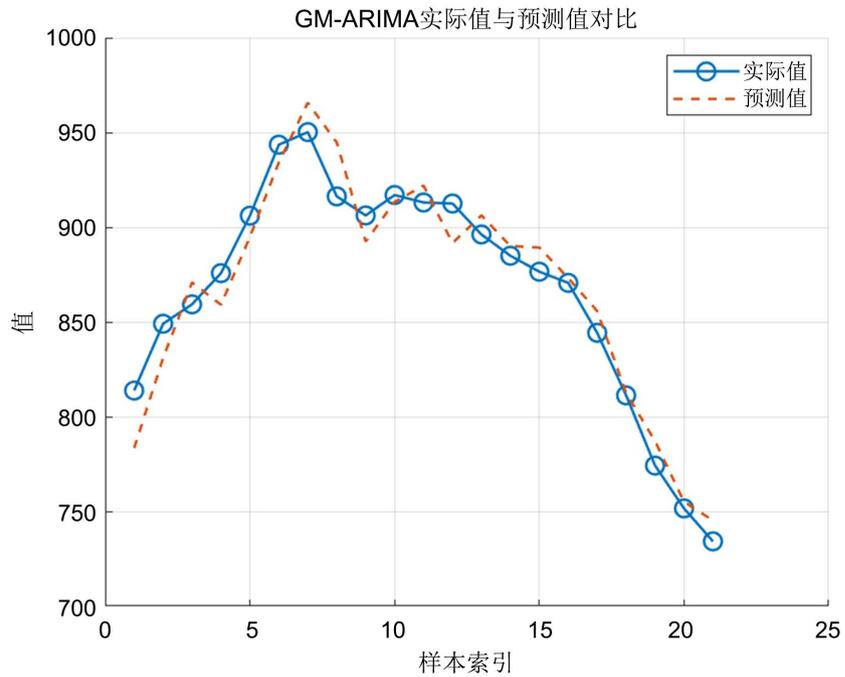
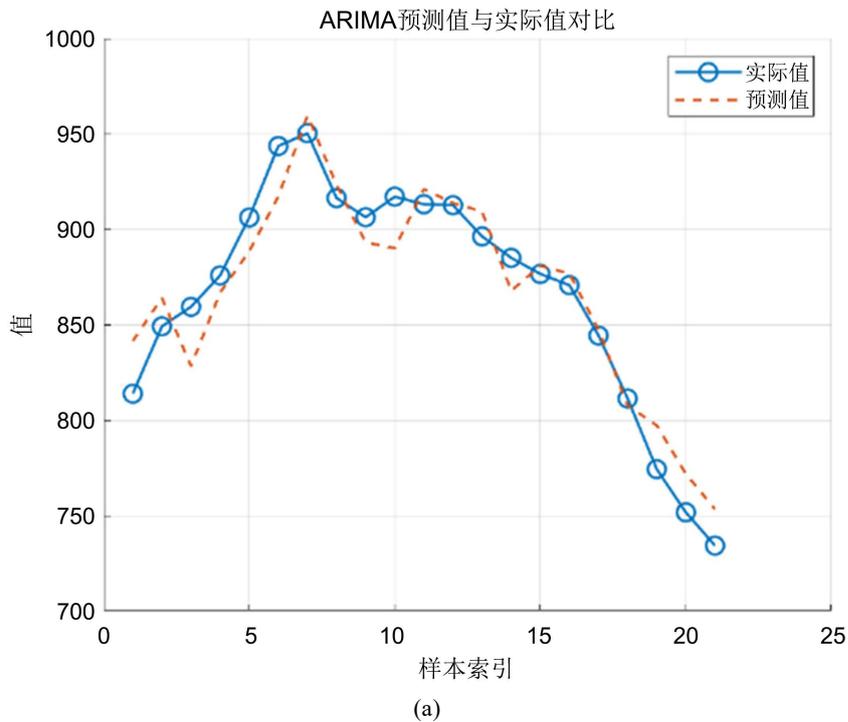


Figure 6. Fitting chart of the grey prediction model

图 6. 灰色预测模型拟合图



(a)

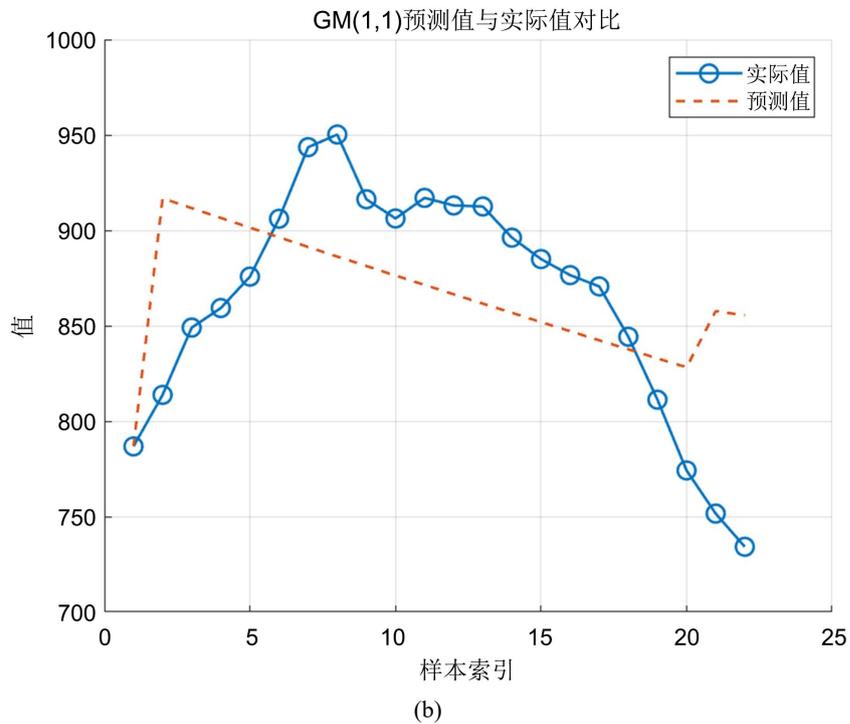
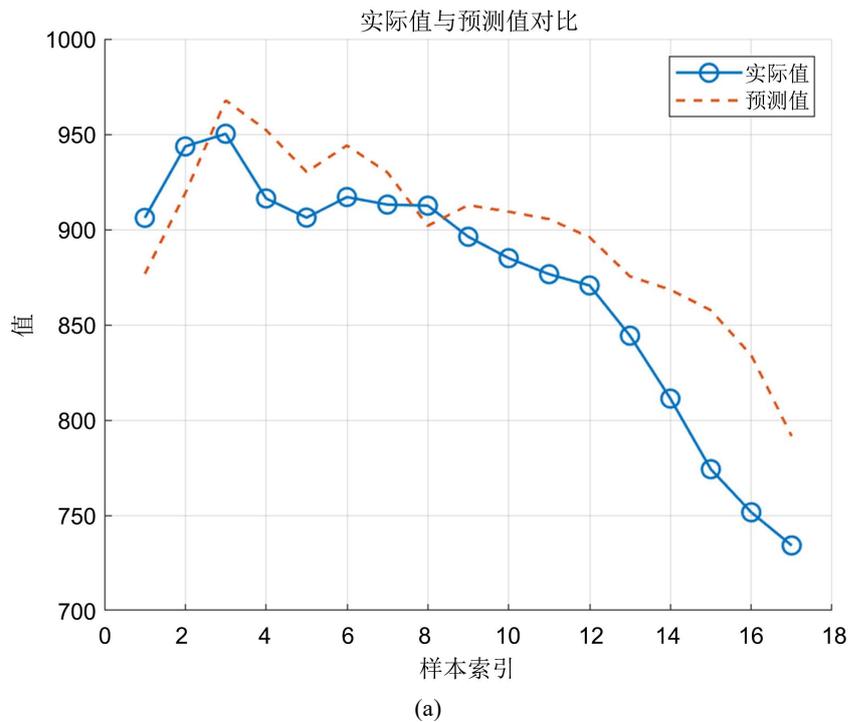


Figure 7. ARIMA prediction fitting plot (a) and GM (1, 1) prediction fitting plot (b)
 图 7. ARIMA 预测拟合图(a)和 GM (1, 1)预测拟合图(b)

支持向量机(SVM)常用于分类和回归问题，通过构造一个超平面来区分不同类别的数据，并利用核函数将数据映射到高维空间，从而捕捉复杂的非线性关系。在碳排放预测中，可以得出 SVM 模型对碳排放的预测量偏高，拟合度不及其他模型，见图 8(a)。



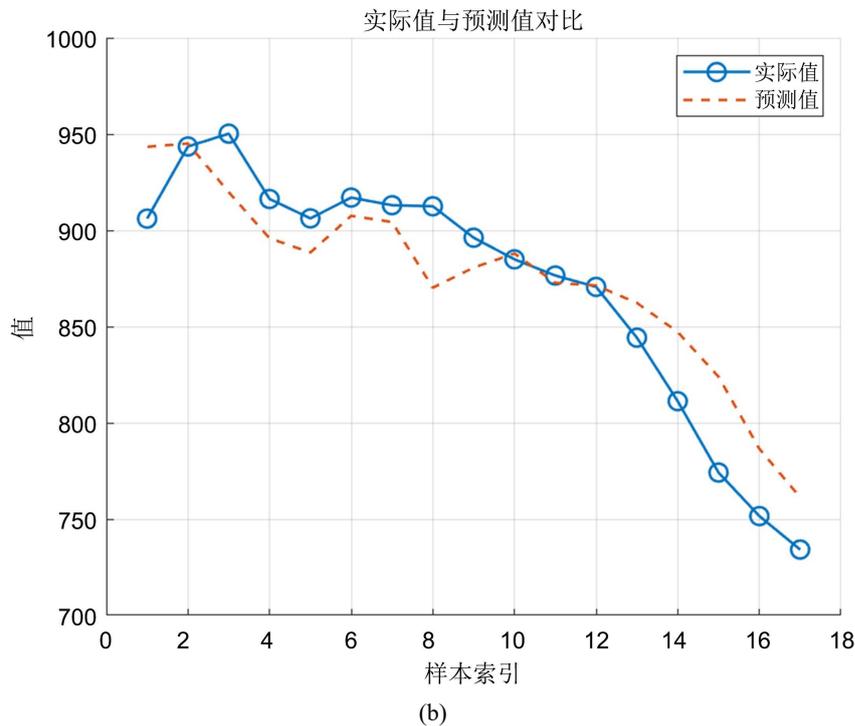


Figure 8. SVM prediction fitting plot (a) and neural network prediction fitting plot (b)
 图 8. SVM 预测拟合图(a)和神经网络预测拟合图(b)

BP 神经网络常用于处理复杂的非线性问题，通过多个层次的神经元连接组成，可以有效捕捉数据中的深层特征。BP 神经网络能够通过学习历史数据中的模式，提供高精度的预测结果，但对于小样本数据可能会出现过拟合现象。在碳排放预测中，得出 BP 神经网络也能较好拟合原本数据，但精确度不及 GM-ARIMA 模型，见图 8(b)。

通过结合 GM (1, 1)模型的灰色预测能力和 ARIMA 模型的自回归特性，混合模型在拟合效果和预测准确性上均优于单独的 ARIMA 和 GM (1, 1)模型。混合模型不仅有效地减少了残差，还提高了对未来碳排放趋势的预测能力，充分利用了两种模型的优势。为了加强验证 GM-ARIMA 算法的有效性，引入预测领域常用的平均绝对误差 MAPE，均方根误差 RMSE，趋势准确率 F 三种方法对模型进行评价。

从表 10 的结果可以看出，针对提出的 MAPE、RMSE、F 值，GM-ARIMA 值分别为 12.305、14.403、85.20%均优于传统的 GM (1, 1)和 ARIMA 模型。对于预测值而言，2020 年与 2021 年的实际值分别是 752.01、734.6，显然 GM-ARIMA 的预测值更贴近于实际值，对未来有很好的判断能力。

Table 10. Statistical data values
 表 10. 统计数据值

	MAPE	RMSE	F	2020 预测	2021 预测
GM-ARIMA	12.305	14.403	85.20%	755.49	745.66
GM (1, 1)	45.922	55.084	61.90%	857.91	855.72
ARIMA	14.345	16.852	81.20%	772.17	753.65
SVM	59.991	79.382	75%	820.99	782.04
BP 神经网络	18.956	27.287	81.25%	770.28	758.85

与较为热门的机器学习算法 SVM 和 BP 神经网络相比, SVM 的 MAPE、RMSE、F 值为 59.991、79.382、75%; BP 神经网络的为 18.956、27.287、81.25%。两者的输出参数均逊色于 GM-ARIMA 模型。

5.5. 碳排放量预测结果

根据山东、河北和河南在 2025 年至 2029 年间的碳排放预测数据, 见表 11, 可以看出各省在低碳发展方面的表现和发展趋势。具体预测数据显示, 山东省的碳排放量从 2025 年的 647.98 万吨下降至 2029 年的 604.62 万吨, 整体呈现出明显的下降趋势, 这表明该省在实施减排政策方面取得了显著成效, 显示出其在绿色转型过程中迈出的坚实步伐。

Table 11. Forecast Table for Shandong, Hebei, and Henan (2025~2029)

表 11. 山东、河北、河南 2025~2029 预测表

	山东	河北	河南
2025	647.98	804.74	535.61
2026	644.46	791.53	543.39
2027	635.88	779.39	552.88
2028	619.66	770.22	547.62
2029	604.62	761.76	534.12

相较之下, 河北省在 2025 年的碳排放为 804.74 万吨, 到 2029 年降至 761.76 万吨, 有一定的降幅, 但整体减少幅度相对缓慢。因此, 河北省应持续加大减排力度, 尤其是在高碳排放行业, 确保政策有效落实。而河南省的碳排放量从 2025 年的 535.61 万吨波动后小幅上升至 534.12 万吨, 整体变化幅度不大, 这表明该省在控制碳排放方面同样取得了一定的成绩, 反映出其在农业和能源结构调整上有积极成效。

综上所述, 三个省份在碳排放控制方面各有特点, 山东省的减排成效最为显著, 河北省需加强后续措施, 而河南省则应继续保持现有成果, 推动进一步的减排策略。各省在未来的发展中, 推动低碳经济的实现, 为实现国家的碳达峰和碳中和目标贡献力量。

6. 研究成果与应用价值

6.1. 研究成果

本文提出的 GM-ARIMA 智能混合预测算法, 相较于单独的 ARIMA 和 GM (1, 1)模型, 采用了智能优化和消除残差方法, 拟合效果和预测准确性均有较大提升。针对 GM (1, 1)的参数进行优化, 采用了随搜索进程非线性动态变化的算法参数及对处于最佳位置的粒子进行轻微扰动的策略, 可以提高模型的预测精度。针对 ARIMA 预测残差的参数, 通过网格搜索方法确定最小的 BIC, 从而确定 ARIMA 模型的参数。最终实现对预测结果的残差消除, 使农业碳排放的预测结果更贴近未来实际的发展趋势, 从而提供了一种新的分析农业碳排放的有效途径。

6.2. 应用价值

本研究构建的预测模型为政府部门制定环境友好型农业政策提供了坚实的数据支持, 而且在推动农业绿色转型和实现可持续发展方面发挥着重要作用。通过了解未来农业碳排放的可能趋势, 可以更加精准实施相关政策。例如, 提高低碳农业技术的补贴或鼓励减少高碳材料的使用、评估不同省市的农业投入, 能够为各地量身定制不同的农业减排政策, 优化农业补贴机制, 提升政策的针对性和有效性, 促进

区域农业绿色发展。对全国各地未来的农业碳排放量进行预测,可以提供农业碳排放检测和预警,政府和企业可以及早识别出高碳排放的时间和区域,进而实施针对性的减排措施,调整农业碳排放结构,有助于实现“碳达峰”和“碳中和”目标。

更为重要的是,该预测模型的方法和框架具备良好的通用性和扩展性。除了农业碳排放领域外,该模型还可以推广应用于环境科学中的其他污染物排放预测、生态环境质量评价等方面;在金融学领域,可用于对相关环保投资风险和收益的动态分析;在经济学中,也可辅助进行绿色经济发展趋势分析和政策效果评估。通过跨学科的应用推广,模型能够支持更大范围的实证研究和决策分析,助力构建多元化、智能化的环境治理和可持续发展体系。

基金项目

广东省教育厅科研项目(特色创新项目)(2023KTSCX034);

广东省高等教育学会关于“十四五”规划2023年度高等教育研究课题(23GYB22);

高等学校大学数学教学研究与发展中心2024年教学改革项目(CMC20240605)。

参考文献

- [1] Jia, G.S., Shevliakova, E., Artaxo, P., *et al.* (2019) Land-Climate Interactions. IPCC Special Report on Climate Change and Land.
- [2] 黄杰, 孙自敏. 中国种植业碳生产率的区域差异及分布动态演进[J]. 农业技术经济, 2022(7): 109-127.
- [3] 杨红亮, 史丹. 能效研究方法和中国各地区能源效率的比较[J]. 经济理论与经济管理, 2008(3): 12-20.
- [4] 朱舰伟, 刘卫柏, 刘金丹. 内蒙古自治区农业碳排放时序变化与影响因素研究[J]. 湖南师范大学自然科学学报, 2023, 46(1): 20-28.
- [5] 田云, 李波, 张俊飏. 我国农地利用碳排放的阶段特征及因素分解研究[J]. 中国地质大学学报(社会科学版), 2011, 11(1): 59-63.
- [6] 崔永福, 高策, 王俊凤, 等. 河北省县域农业碳排放空间演化及对策[J]. 中国农机化学报, 2023, 44(5): 241-248, 256.
- [7] 王淑佳, 孔伟, 任亮, 等. 国内耦合度模型的误区及修正[J]. 自然资源学报, 2021, 36(3): 793-810.
- [8] 朱智谔, 徐婕, 李红艳, 等. “双碳”背景下能源消费碳排放效率与生态韧性耦合协调研究[J]. 资源与产业, 2024, 26(3): 21-35.
- [9] IPCC (2013) Climate Change 2013: The Physical Science Basis Technical Summary.
- [10] 马涛. 上海农业碳源碳汇现状评估及增加碳汇潜力分析[J]. 可持续展, 2011, 28(5): 38-41.
- [11] 王宝义, 张卫国. 中国农业生态效率测定及时空差异研究[J]. 中国人口·资源与环境, 2016, 26(6): 11-19.
- [12] 段华平, 张悦, 赵建波, 卞新民. 中国农田生态系统的碳足迹分析[J]. 水土保持学报, 2011, 25(5): 203-208.
- [13] 李波, 张俊飏. 基于投入视角的我国农业碳排放与经济发展脱钩研究[J]. 经济经纬, 2012(4): 27-31.
- [14] 赵洪亮, 陈思越, 谢立勇. 辽宁省种植业碳排放影响因素及预测分析[J]. 中国生态农业学报(中英文), 2024, 32(11): 1818-1828.
- [15] 谢金燕, 刘丽星, 杨欣, 等. 改进粒子群优化算法的果园割草机作业路径规划[J]. 中国农业大学学报, 2023, 28(11): 182-191.
- [16] 刘振超, 苑迎春, 王克俭, 等. 融合特征权重与改进粒子群优化的特征选择算法[J]. 计算机工程与科学, 2024, 46(2): 282-291.