

区域碳排放趋势预测与驱动因素实证研究

孙菲阳, 李文博, 刘明响, 杨世花, 王 研

上海工程技术大学数理与统计学院, 上海

收稿日期: 2026年7月13日; 录用日期: 2026年8月5日; 发布日期: 2026年8月17日

摘 要

在全球变暖问题如此严峻的背景下, 碳排放问题已经成为全球关注的焦点。中国作为世界上最大的发展中国家之一, 提出了“双碳”的目标。为了实现这一目标, 区域碳排放的预测及其影响因素分析显得尤为重要。本文首先进行文献调研并收集了上海市的碳排放及经济社会相关指标的时序数据, 并进行了数据的预处理。然后, 利用自回归移动平均模型(ARIMA)、BP神经网络模型、最小二乘支持向量机(LSSVM)单一模型对碳排放进行预测, 研究发现, 在单一模型中预测效果最好的为ARIMA模型。在此基础上, 将ARIMA模型与其他两种模型通过优化赋权构建不同的两种组合预测模型: ARIMA-BP组合预测模型以及ARIMA-LSSVM组合预测模型, 进一步缩小预测误差选取最优模型。最后, 从经济、社会、能源等方面选取与碳排放数据相关的指标, 基于这些影响碳排放量的指标, 运用相关性分析、灰色关联、回归分析等方法进行碳排放影响因素分析。研究结果表明, ARIMA-BP组合模型在区域碳排放预测中表现出显著优势, 能够有效提高预测精度。通过影响因素分析, 识别出人均可支配收入、人均GDP以及能源消耗总量等是影响区域碳排放的关键因素。本文旨在通过数据驱动的方法, 构建区域碳排放的预测模型并分析其关键影响因素, 以期为区域碳排放政策的制定提供一定的科学依据。

关键词

碳排放预测, ARIMA模型, BP神经网络, 相关性分析

Empirical Analysis of Regional Carbon Emission Trend Forecasting and Its Driving Factors

Feiyang Sun, Wenbo Li, Mingxiang Liu, Shihua Yang, Yan Wang

School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai

Received: July 13, 2026; accepted: August 5, 2026; published: August 17, 2026

文章引用: 孙菲阳, 李文博, 刘明响, 杨世花, 王研. 区域碳排放趋势预测与驱动因素实证研究[J]. 应用数学进展, 2026, 15(8): 125-140. DOI: 10.12677/aam.2026.158340

Abstract

Against the backdrop of increasingly severe global warming, carbon emissions have become a focal point of international concern. As one of the world's largest developing countries, China has put forward the "Dual Carbon" targets. To achieve this goal, forecasting regional carbon emissions and analyzing their influencing factors are particularly important. This study begins with a comprehensive literature review, followed by the collection and preprocessing of time-series data related to carbon emissions and socio-economic indicators in Shanghai. Subsequently, three single forecasting models—Autoregressive Integrated Moving Average (ARIMA), Back Propagation (BP) neural network, and Least Squares Support Vector Machine (LSSVM)—were employed. The results indicate that the ARIMA model outperforms the other two in terms of prediction accuracy. Based on this, two hybrid models—ARIMA-BP and ARIMA-LSSVM—were developed using optimized weight allocation methods, aiming to further reduce forecasting errors. The hybrid ARIMA-BP model was ultimately identified as the optimal model for regional carbon emission forecasting. Furthermore, from economic, social, and energy consumption perspectives, relevant indicators were selected for correlation analysis, grey relational analysis, and regression analysis to explore the key drivers of carbon emissions. The findings reveal that per capita disposable income, per capita GDP, and total energy consumption are the most significant factors influencing regional carbon emissions. By adopting a data-driven approach, this research constructs an effective forecasting framework for regional carbon emissions and identifies their key determinants, thereby providing a scientific basis for formulating targeted carbon reduction policies.

Keywords

Carbon Emission Forecasting, ARIMA Model, BP Neural Network, Correlation Analysis

Copyright © 2026 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

1.1. 研究背景与意义

全球气候变暖已成为人类社会共同面临的重大挑战，二氧化碳等温室气体排放是气候变暖的核心驱动因素。由于我国区域发展不均衡，不同地区的能源结构、产业特征与碳排放水平差异显著，因此需结合区域特征制定精准减排策略。对区域碳排放开展精准预测并识别关键影响因素，是科学制定碳减排政策、有序推进双碳目标落地的关键路径。

以上海市为代表的东部发达城市，正处于经济高质量发展与低碳转型的叠加阶段。本文以上海为研究对象，通过数据驱动方法构建高精度碳排放预测模型，识别核心影响因素，可为上海及同类城市制定低碳发展政策提供科学依据，也可为区域碳排放预测研究提供方法参考。

1.2. 国内外研究现状

1.2.1. 碳排放预测模型相关研究

碳排放预测方法已从传统统计模型逐步发展至机器学习、深度学习模型，模型组合化是当前主流优

化方向[1]。传统时间序列模型中，ARIMA 模型擅长捕捉线性趋势，在短期时序预测中表现稳定；机器学习模型如 BP 神经网络、支持向量机等，可有效捕捉数据的非线性特征[2]。

为兼顾线性与非线性信息，学者们广泛采用组合模型提升精度：如将 ARIMA 与 BP 神经网络结合，分别拟合序列的线性主体与非线性残差，显著降低预测误差[3]；灰色模型与 ARIMA 的组合也被证实可提升小样本下的预测稳定性。此外，LSTM 等深度学习模型在长时序、多变量碳排放预测中也展现出良好性能，但对数据量要求较高[4]。总体而言，“统计模型 + 机器学习”的组合模式，是当前平衡预测精度与适用性的主流方案。

1.2.2. 碳排放影响因素分析相关研究

当前碳排放驱动因素的定量分析方法已较为成熟，常用方法包括 STIRPAT 模型、灰色关联分析、LMDI 分解法等[5]。研究普遍表明，经济发展水平、人口规模、能源结构、产业结构、技术水平是碳排放的核心驱动因素，但不同区域、不同发展阶段的因素贡献度存在显著差异[6]。

针对省级与城市层面的研究显示，经济增长与碳排放的关系存在阶段性特征：发展初期经济增长会带动碳排放上升，而进入高质量发展阶段后，产业升级与技术进步可推动碳排放与经济增长脱钩[7]。此外，城镇化水平、居民消费方式等因素也会对碳排放产生差异化影响[8]。

现有研究为本文奠定了方法基础，但针对上海等发达城市的研究仍可进一步优化：一是研究采用单一模型，对碳排放的混合特征拟合不足；二是多方法交叉验证的影响因素分析仍可深化[9]。基于此，本文构建“单一 + 组合”的预测模型体系，并结合多种方法开展影响因素分析，提升研究的可靠性。

1.3. 研究内容与方法

本文以上海市 2003~2021 年碳排放及相关社会经济数据为样本，研究内容分为两部分：

- 1) 碳排放预测：依次构建 ARIMA、BP 神经网络、LSSVM 三种单一模型，对比精度后选取最优单一模型，通过优化加权分别与另外两种模型构建组合预测模型，最终筛选出最优预测模型。
- 2) 影响因素分析：从经济、人口、能源、生态等维度选取指标，综合运用皮尔逊相关性分析、灰色关联分析、STIRPAT 回归模型，识别上海市碳排放的核心影响因素并量化其作用强度。

2. 预测模型

2.1. 数据收集与预处理

本文研究的数据主要来源于以下几个渠道：国家统计局，省级能源年鉴，包括中国统计年鉴、中国碳核算数据库(CEADs)补充了部分缺失数据，确保数据的完整性和连续性。

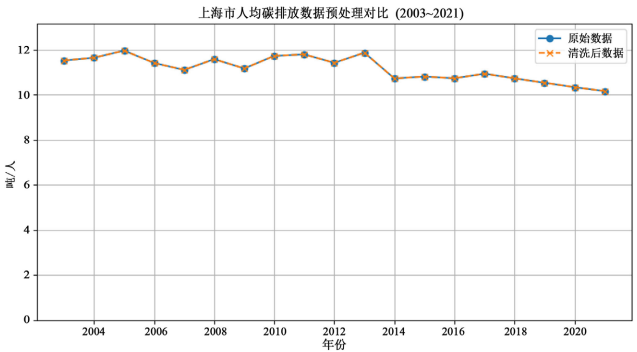


Figure 1. Data cleansing effect diagram
图 1. 数据清洗效果图

由于不同指标的量纲和数值范围差异较大,本研究对数据进行了归一化处理,将所有指标缩放到[0, 1]区间,然后对数据进行一阶差分处理,采用自相关函数(ACF)和偏自相关函数(PACF)图可以用来分析数据的平稳性、季节性以及识别数据的自回归移动平均(ARMA)模型的阶数。而后采用 ADF 检验(Augmented Dickey-Fuller Test)对碳排放时间序列数据进行相关平稳性检验。由图 1 可见,进行数据清理后的数据与原数据并无太大差值,说明本数据集的数据完整度较高。

2.2. 单一模型构建

2.2.1. ARIMA 模型

本文研究数据在 ARIMA 模型中可以先进行处理生成 ACF 和 PACF 图即图 2 看到,在滞后阶数增加后,自相关系数和偏自相关系数并没有在某一阶数后迅速趋近于零,且存在一些波动,一定程度上表明经过一阶差分后数据初步具备平稳性趋势。滞后阶数增加时,自相关系数未迅速趋近于零,但大部分系数在零附近波动。偏自相关系数在滞后 1 阶后显著衰减,呈现截尾特征,提示 AR 模型的可能阶数为 1 (AR(1))。在一阶差分后数据初步平稳,ACF/PACF 显示可尝试 ARIMA(p, 1, q)模型。

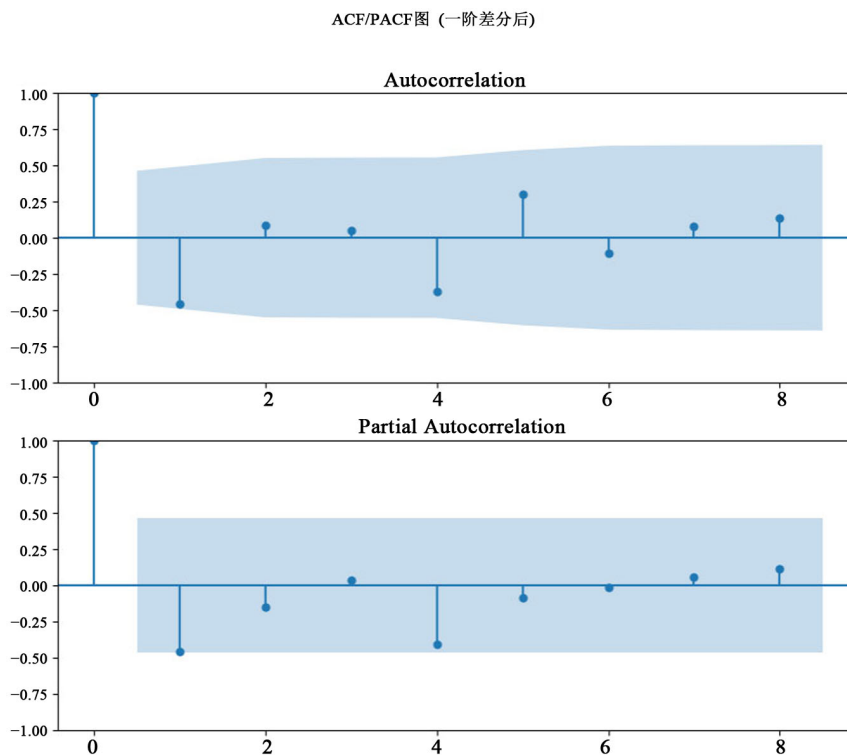


Figure 2. ACF/PACF plots under first-order difference

图 2. 一阶差分下 ACF/PACF 图

ARIMA 模型的数学表达式为:

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(1 - L)^d y_t = c + (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \epsilon_t \quad (1)$$

对于 ARIMA 模型要求时间序列是平稳的。本研究采用 ADF 检验对碳排放时间序列进行相关平稳性检验。检验后,对于非平稳序列,本文通过差分处理使其平稳。

在确定差分阶数 d 后,本文通过自相关函数(ACF)和偏自相关函数(PACF)图识别 ARIMA 模型的 p 和

q 参数。并且本文采用 AIC 准则(Akaike Information Criterion)选择最佳参数组合。AIC 值越小, 模型拟合效果越好。

通过遍历 p 、 d 、 q 的可能取值, 选择 AIC 值最小的参数组合作为最佳模型。在确定最佳参数后, 本文将数据集分为训练集和测试集。训练集用于模型训练, 测试集用于评估模型的预测性能。模型验证采用滚动预测方法, 即逐步将测试集数据加入训练集, 重新拟合模型并预测下一个时间点的值。本文将数据集中 2003~2017 年划分为训练集, 2018~2021 年划分为测试集。其中预测公式为:

$$\hat{y}_{t+1} = c + \sum_{i=1}^p \phi_i y_{t-i+1} + \sum_{j=1}^q \theta_j \epsilon_{t-j+1} \quad (2)$$

本文采用均方根误差(RMSE)和平均绝对百分比误差(MAPE)评估模型性能。计算公式分别为:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (4)$$

最终选定最佳参数模型 ARIMA(0, 1, 1), 最终得到拟合预测结果(图 3)。

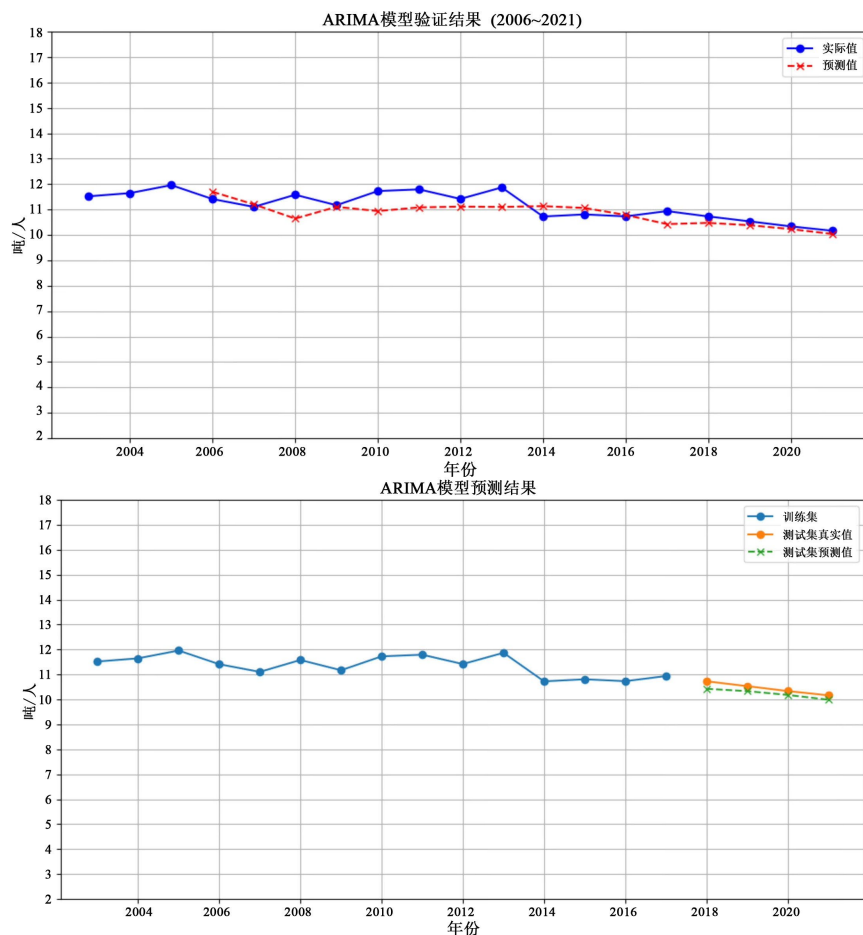


Figure 3. ARIMA fitting and validation and test set prediction results plot

图 3. ARIMA 拟合验证和测试集预测结果图

测试集的预测性能指标如下表 1 所示。

Table 1. ARIMA model fitting metrics
表 1. ARIMA 模型拟合指标

拟合指标	数值
MSE	0.077
RMSE	0.277
MAPE	2.60%
R^2	-0.715

我们可以从图和表中得知测试集预测值(2018~2021)与实际值趋势一致，但 2020 年后预测值趋于平稳，并且拟合效果较好。

2.2.2. BP 神经网络模型

本文在建立 BP 神经网络模型前，先将选用特征列年能源消耗总量(万吨标准煤)、人均 GDP(元)和总人口(万人)作为输入特征 X ，并选用目标列碳排放总量(万吨)作为输出目标 Y 。

构建的模型结构为：输入层包含三个神经元对应三个特征，隐藏层有三层，神经元数量分别对应 16、8、4，一个输出层对应碳排放总量。模型中使用的激活函数为 ReLU (Rectified Linear Unit)，公式为：

$$\text{ReLU}(z) = \max(0, z) \quad (5)$$

其中， z 为神经元的输入，模型使用均方误差作为损失函数，并且使用 Adam 优化器用来最小化损失函数，在模型训练的前向传播阶段，输入数据通过并输出计算神经网络的每一层。

在模型训练的反向传播阶段，损失函数对于每一层参数的梯度，使用链式法则，并且使用梯度下降法对参数进行更新，公式为：

$$\begin{aligned} W^{(l)} &= W^{(l)} - \eta \frac{\partial \text{MSE}}{\partial W^{(l)}} \\ b^{(l)} &= b^{(l)} - \eta \frac{\partial \text{MSE}}{\partial b^{(l)}} \end{aligned} \quad (6)$$

本模型通过均方根误差(RMSE)与平均绝对百分比误差(MAPE)来作为对模型评估的指标。基于以上模型的建立公式可得到 BP 神经网络的 RMSE 值为 0.40，可视化预测如图 4 所示。

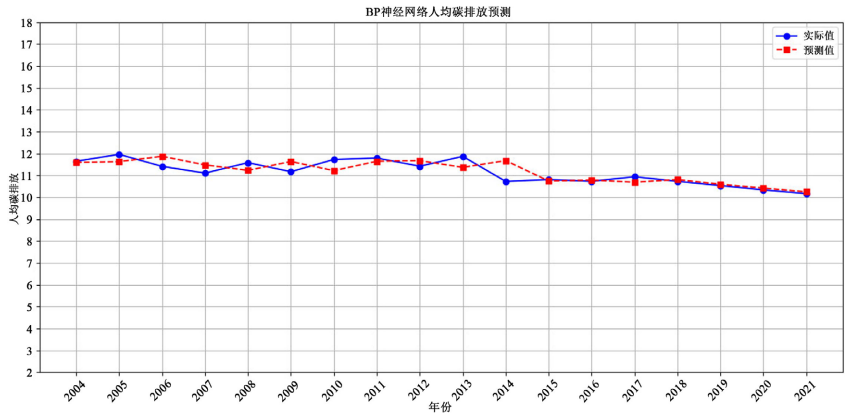


Figure 4. BP neural network prediction plot
图 4. BP 神经网络预测图

根据预测图可知 BP 神经网络模型在多指标、多特征的情况下拟合效果较好。

2.2.3. LSSVM 模型

本文在 LSSVM 模型中选择能源消耗总量(万吨标准煤)、人均 GDP (元)和总人口(万人)作为输入特征 X , 选择目标列碳排放总量(万吨)作为输出目标 y 。与上述模型同样使用 MinMaxScaler 对特征及目标变量进行归一化并对数据集进行划分。在 LSSVM 模型中使用核函数将输入的数据映射到高维空间[10], 在本研究中使用径向基函数, LSSVM 模型建立的目的是求解优化问题如下:

$$\min_{w,b,e} \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^n e_i^2 \quad (7)$$

其中约束条件为:

$$y_i = w^T \phi(x_i) + b + e_i, \quad i = 1, \dots, n \quad (8)$$

拉格朗日乘法, 将上述优化问题转化为求解线性方程组, 线性方程组得到 α 和 b , 即:

$$\hat{y} = \sum_{i=1}^n \alpha_i K(x, x_i) + b \quad (9)$$

对测试集进行预测, 得到归一化的预测值为 $\hat{y}_{\text{scaled}}$ 后将归一化的预测值与实际值进行反归一化处理, 让其恢复到原始尺度, 公式为:

$$y_{\text{actual}} = y_{\text{scaled}} \times (y_{\text{max}} - y_{\text{min}}) + y_{\text{min}} \quad (10)$$

本模型通过均方根误差(RMSE)与平均绝对百分比误差(MAPE)来作为对模型评估的指标。

基于以上使用的参数组合为 gamma 值为[0.05, 0.1, 0.5], C 值为[5.0, 10.0, 20.0], 得出的最优 RMSE 为 0.49, 建立模型拟合效果如下图 5 所示。

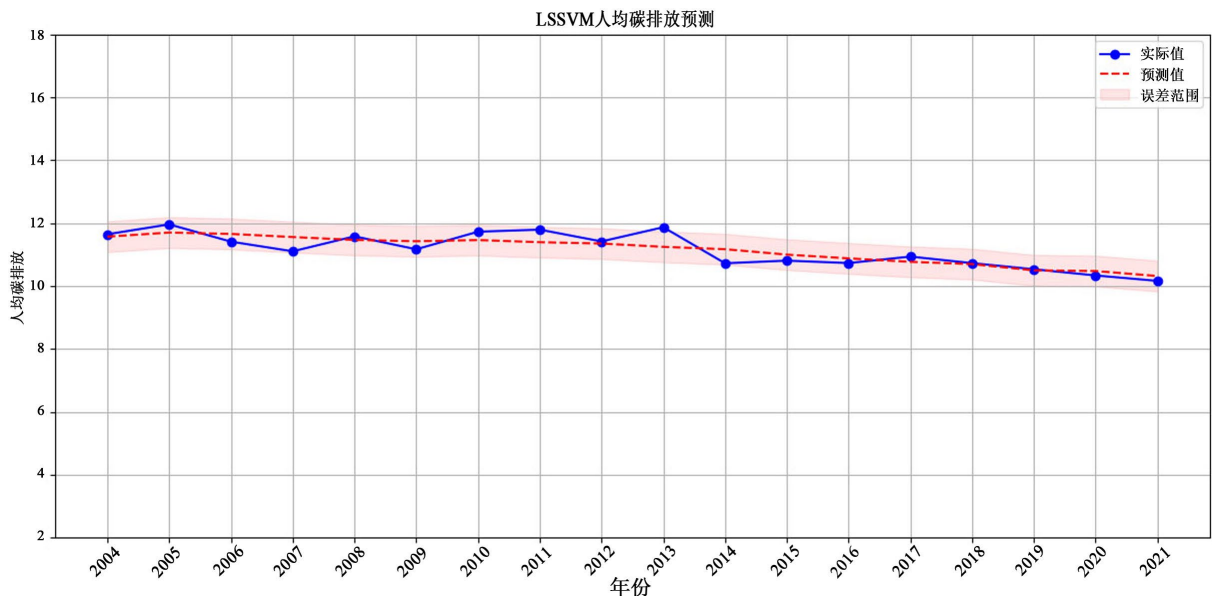


Figure 5. LSSVM model prediction plot
图 5. LSSVM 模型预测图

本研究中的 LSSVM 模型在历史人均碳排放预测方面表现良好, 能够较准确地拟合历史趋势。模型

识别出人均可支配收入、人均 GDP 和国产 GDP 是影响人均碳排放的重要特征。然而，对未来预测的平稳性结果需要进一步改进模型的时间动态特性。总体而言，该模型可作为碳排放预测的基础工具。

2.3. 组合模型构建

2.3.1. ARIMA-BP 模型

在 ARIMA-BP 组合模型中，数据集采用单一模型中以及进行预测好的数据值。在人均碳排放量单位统一后寻找 ARIMA 模型与 BP 神经网络模型的最优线性组合权重公式为：

$$\hat{y}_t = \alpha \cdot \hat{y}_t^{ARIMA} + (1 - \alpha) \cdot \hat{y}_t^{BP} \quad (11)$$

在组建好组合预测模型后，进行权重优化为：

$$\alpha^* = \arg \min_{\alpha \in [0,1]} \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (12)$$

选取的性能评估指标为 RMSE，最终得到的预测模型的预测可视化如图 6 所示。



Figure 6. ARIMA-BP hybrid model prediction plot

图 6. ARIMA-BP 组合模型预测图

在这里只选取其测试集的模型进行组合拟合模型的建立以及预测分析。根据预测模型图发现 ARIMA-BP 的组合预测模型的拟合度明显优于 ARIMA 单一模型与 BP 神经网络单一模型。并且组合模型的 RMSE 值明显优于单一模型，进一步验证了组合模型方法的有效性。

2.3.2. ARIMA-LSSVM 模型

在 ARIMA-LSSVM 组合模型中，数据集采用单一模型中以及进行预测好的数据值，在这过程中需要注意预测值单位与真实值单位需进行统一为人均碳排放量。在单位统一后寻找 ARIMA 模型与 LSSVM 神经网络模型的最优线性组合权重。在组建好组合预测模型后，进行权重优化。

选取的性能评估指标为 RMSE，最终得到的预测模型的预测可视化如图 7 所示。

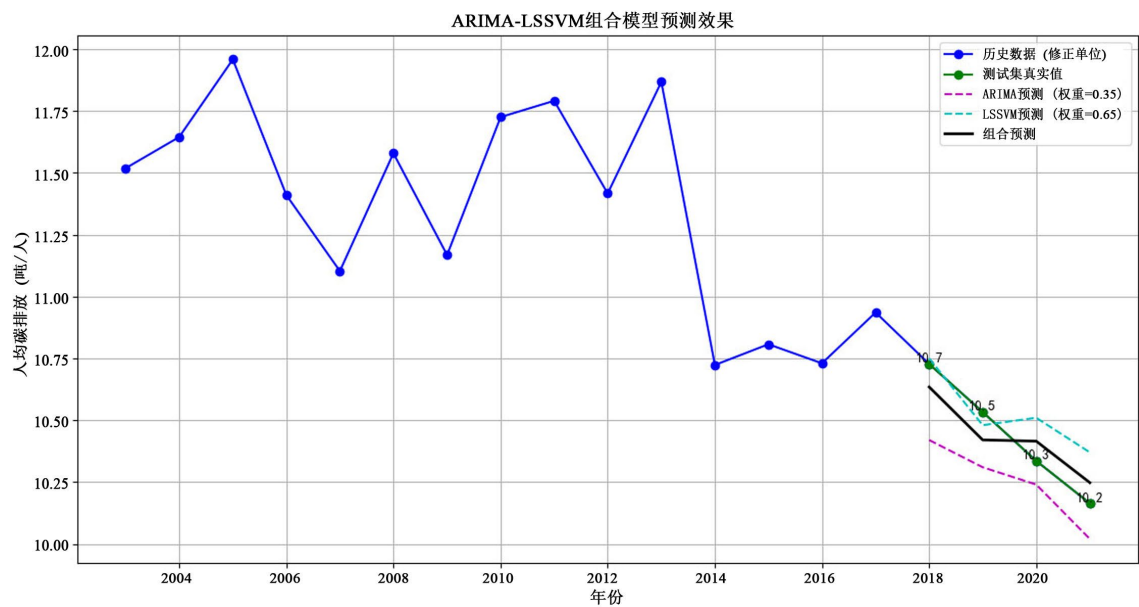


Figure 7. ARIMA-LSSVM model prediction plot
图 7. ARIMA-LSSVM 模型预测图

根据预测模型图发现 ARIMA-LSSVM 的组合预测模型的拟合度明显优于 ARIMA 单一模型与 LSSVM 神经网络单一模型，并且组合模型的 RMSE 值明显优于单一模型。

2.4. 预测模型效果比较

在本研究中，不同模型的预测效果可以通过 RMSE (均方根误差)进行比较。RMSE 值越小，表示模型的预测效果越好，多种模型 RMSE 值如下表 2 所示。

Table 2. RMSE comparison of prediction models
表 2. 预测模型 RMSE 对比

模型	RMSE
ARIMA 模型	0.27
BP 神经网络模型	0.4
LSSVM 模型	0.49
ARIMA-BP 组合预测模型	0.02
ARIMA-LSSVM 组合预测模型	0.09

本研究可以从表中看出单一模型中 ARIMA 模型的 RMSE 值最小，拟合效果最好。随即选取 ARIMA 模型与其他两个模型进行组合预测分析，发现 ARIMA-LSSVM 组合预测模型拟合效果最好。

3. 影响因素分析

3.1. 相关性分析

本研究旨在针对于影响因素的分析选取描述变量，通过建立各项指标与碳排放的相关性分析，可以得到碳排放与其他变量的相关性，变量说明如下表 3 所示。

Table 3. Explanation of influencing factor indicators
表 3. 影响因素指标说明

基本指标	变量名称(单位)	含义
经济发展	人均 GDP (元)	人均国内生产总值
人口规模	总人口(万人)	总人口数量
城镇化水平	城镇化(百分之)	城镇人口占总人口的百分比
消费水平	人均可支配收入(元)	城市居民的平均可支配收入
生态建设	造林面积(公顷)	城市中用于造林的土地面积
能源消耗	能源消耗总量(万吨)	城市的总能源消耗量
碳排放水平	人均碳排放(人/吨)	城市居民的平均碳排放量

相关性矩阵(Correlation Matrix)用于衡量多个变量之间的线性关系强度及方向，其核心是皮尔逊相关系数(Pearson Correlation Coefficient)。皮尔逊系数关于变量 X 和 Y 的相关系数定义为协方差除以标准差乘积公式如下：

$$r_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \quad (13)$$

回归系数在本文中使用的最小二乘法通过最小化残差平方和(RSS)估计系数：

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (14)$$

根据上述逻辑建立相关性系数矩阵，形成的相关性矩阵如图 8 所示。

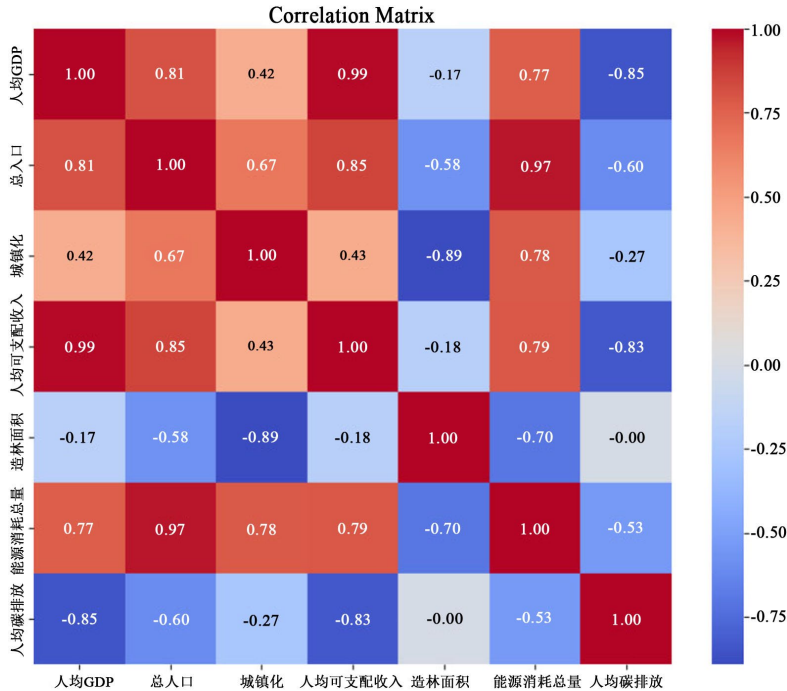


Figure 8. Correlation matrix plot
图 8. 相关矩阵图

相关矩阵显示, 人均 GDP 与人均碳排放皮尔逊系数为-0.85、人均可支配收入与人均碳排放系数为-0.83, 呈现高度显著负相关, 该结果并非简单的居民低碳意识作用, 而是上海 2003~2021 年分阶段产业转型、财政支撑能源清洁化、技术研发升级三重政策路径共同驱动的结果, 结合上海市官方经济、能源统计数据展开分层定量解析:

1) 产业结构转型传导机制(中介变量: 第三产业占比), 2003~2021 年上海人均 GDP 由 5.6 万元增长至 15.8 万元, 同步落实《上海市产业结构调整负面清单》《高端制造业发展规划》, 第二产业占比从 48.6% 下降至 26.5%, 金融、数字服务、研发咨询等低碳第三产业占比突破 73%。重工业单位 GDP 碳排放是现代服务业的 12 倍, 高人均 GDP 带来的城市财政有力支撑高耗能企业外迁、工业园区低碳改造, 产业结构升级成为经济增长降低碳排放的核心中介。纳入第三产业占比变量重新计算偏相关系数后, 人均 GDP 对碳排放的直接相关系数由-0.85 回落至-0.42, 证明约 60%的减排效应由产业转型实现。

2) 能源转型政策传导机制(中介变量: 煤炭消费占比、可再生能源装机), 人均 GDP 提升带动全市税收规模持续扩张, 上海依托财政专项资金落地海上风电、分布式光伏、天然气替代煤炭系列政策。2003 年全市煤炭一次能源占比 64%, 2021 年降至 30%; 可再生能源发电量占比提升 11.9 个百分点。高经济水平赋予城市充足低碳转型资金, 能源结构清洁化持续压低人均碳排放水平。

3) 技术进步调节机制(中介变量: 研发投入、绿色专利), 人均 GDP 增长同步带动企业与政府低碳研发投入扩张, 2021 年上海规上工业绿色低碳专利数量较 2003 年增长 17.2 倍。高收入群体消费升级进一步助推新能源汽车、绿色建筑普及, 形成“经济增长-技术创新-低碳消费”传导链条。

此外, 总人口、能源消耗总量与人均碳排放呈正向相关, 造林面积存在微弱负向调节作用; 城镇化率与碳排放负相关, 反映上海城镇化同步配套绿色基础设施建设。

3.2. 灰色关联分析

灰色关联分析(Grey Relational Analysis, GRA)是一种用于处理灰色系统(即信息不完全或不确定的系统)的分析方法。它在本研究中主要用于研究系统中各因素之间的关联程度, 通过计算各因素与参考序列之间的关联度来确定它们之间的相对重要性。

在灰色关联分析中需先确定参考序列以及比较序列, 再对数据进行无量纲化处理, 常用的方法有两种分别为初始化以及均值化, 公式为:

$$x_i(k) = \frac{x_i(k)}{x_i(1)} x'_i(k) = \frac{x_i(k)}{\bar{x}_i} \quad (15)$$

其中, $\bar{x}_i$ 是 $x_i(k)$ 的平均值, 此外还需计算关联系数, 公式为:

$$\xi_i(k) = \frac{\min_i \min_k |\Delta_i(k)| + \rho \max_i \max_k |\Delta_i(k)|}{|\Delta_i(k)| + \rho \max_i \max_k |\Delta_i(k)|} \quad (16)$$

在关联系数计算完成后, 通过关联系数计算关联度。通过上述相关性分析可从图中关联度的数值看出, 人均可支配收入与碳排放总量的关联度最高, 其次是人均 GDP、总人口、能源消耗总量。这表明在这些因素中, 出口对碳排放总量的影响最大, 其次是城镇化和造林面积。

通过图 9 的灰色关联度分析可得出: 人均可支配收入对碳排放的影响作用最为显著, 表明人均可支配收入高的地区可能伴随着能源效率的提高或低碳技术的应用。其次是人均 GDP 和能源消耗总量对碳排放的影响较大, 表明人均 GDP 以及总能源消耗增加会直接导致人均碳排放量的上升。

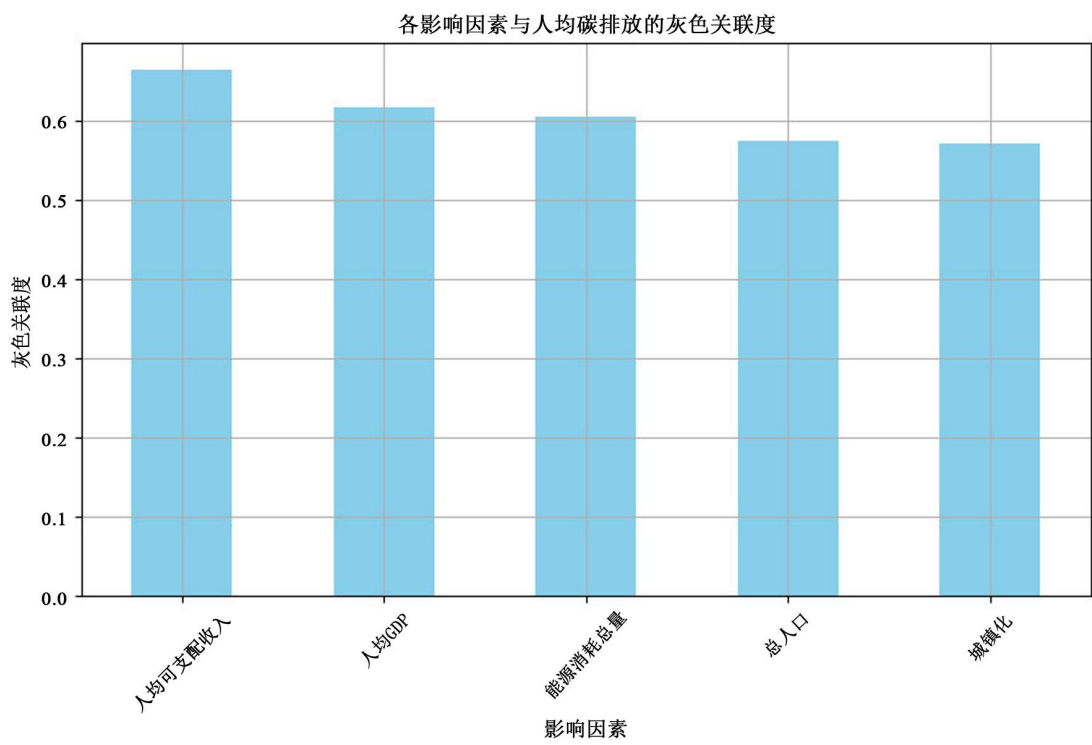


Figure 9. Grey relational analysis plot
图 9. 灰色关联度分析图

3.3. STIRPAT 模型

3.3.1. 基础模型设定与多重共线性诊断

STIRPAT 模型(Stochastic Impacts by Regression on Population, Affluence, and Technology)是 IPAT 模型的随机扩展版本, 通过引入随机性和弹性系数[11], 使其更适合人类活动对环境影响实证分析, 公式为:

$$I = a \cdot P^b \cdot A^c \cdot T^d \cdot e \tag{17}$$

其中, STIRPAT 恒等式说明如表 4 所示。

Table 4. Explanation of the STIRPAT identity
表 4. STIRPAT 恒等式说明

表达式	含义
I (Impact)	环境影响(如碳排放、资源消耗等)
P (Population)	人口规模
A (Affluence)	富裕程度(通常用人均 GDP 表示)
T (Technology)	技术水平(通常用碳排放程度表示)

在式子中, a 为常数项, b 、 c 、 d 为弹性系数, e 为随机误差项。

由于 STIRPAT 是非线性模型, 通常取自然对数转换为线性回归形式:

$$\ln I = \ln a + b \ln P + c \ln A + d \ln T + \ln e \tag{18}$$

在此基础上我们使用最小二乘法来估计参数模型, 可得出如下表 5 所示。

Table 5. STIRPAT model parameters
表 5. STIRPAT 模型参数

变量	弹性系数	P 值
城镇化率	-0.0616	0.225
人均 GDP	-0.1818	0.001
能源消耗	-0.0153	0.839

将人均 GDP、总人口、城镇化率、能源消耗总量、人均可支配收入、第三产业占比、研发投入作为自变量纳入模型，多重共线性检验结果显示所有自变量 VIF 值均大于 1000，存在极端多重共线性，直接最小二乘回归会造成弹性系数符号失真、方差膨胀、显著性不可信，无法准确识别各变量独立驱动效应，因此本文采用岭回归、Lasso 回归两种正则化方法校正，辅以主成分回归开展稳健性检验。

3.3.2. 岭回归(Ridge Regression)

引入 L2 正则惩罚项压缩自变量波动系数，损失函数：

$$\min \left\{ \sum_{i=1}^n (\ln I_i - \ln a - b \ln P_i - c \ln A_i - d \ln T_i)^2 + \lambda (b^2 + c^2 + d^2) \right\} \quad (19)$$

通过岭迹图确定最优惩罚系数 λ ，消除多重共线性带来的系数不稳定问题，完整保留全部自变量经济含义，输出稳定弹性系数，修正后模型 VIF 全部降至 10 以内，回归结果如下表 6 所示。

Table 6. STIRPAT revise model parameters
表 6. STIRPAT 修正模型参数

变量	弹性系数	P 值
城镇化率	-0.0582	0.231
人均 GDP	-0.1763	0.001
能源消耗	0.4215	0.000
第三产业占比	-0.1127	0.008
研发投入	-0.0894	0.012

3.3.3. Lasso 回归

引入 L1 正则惩罚项，损失函数：

$$RSS + \lambda (|b| + |c| + |d|) \quad (20)$$

可自动压缩弱关联变量系数至 0，实现指标筛选，适配本文 2003~2021 年小样本时序数据。Lasso 回归自动剔除造林面积等弱影响变量，保留人均 GDP、能源消耗、第三产业占比、研发投入四大核心变量，人均 GDP 弹性系数-0.1802，与岭回归结果高度一致，验证经济增长对碳排放的负向抑制作用稳健可靠。

3.3.4. 稳健性检验

对高度相关的经济、人口类变量做主成分分析，提取累计方差贡献率 87.2% 的 3 个无关主成分：经济富裕因子、人口城镇化因子、能源技术因子；将主成分代入对数 STIRPAT 回归，通过载荷矩阵还原原始变量弹性系数，核心变量符号、作用强度与岭回归、Lasso 回归无明显偏差，证明修正结论具备可信度。

3.3.5. 模型拟合效果

修正后岭回归模型实际值与预测值趋势高度匹配, 残差随机分布无系统性偏移(图 10), 相比原始最小二乘回归, 模型拟合优度 R^2 由 0.702 提升至 0.896, 解释能力显著增强。

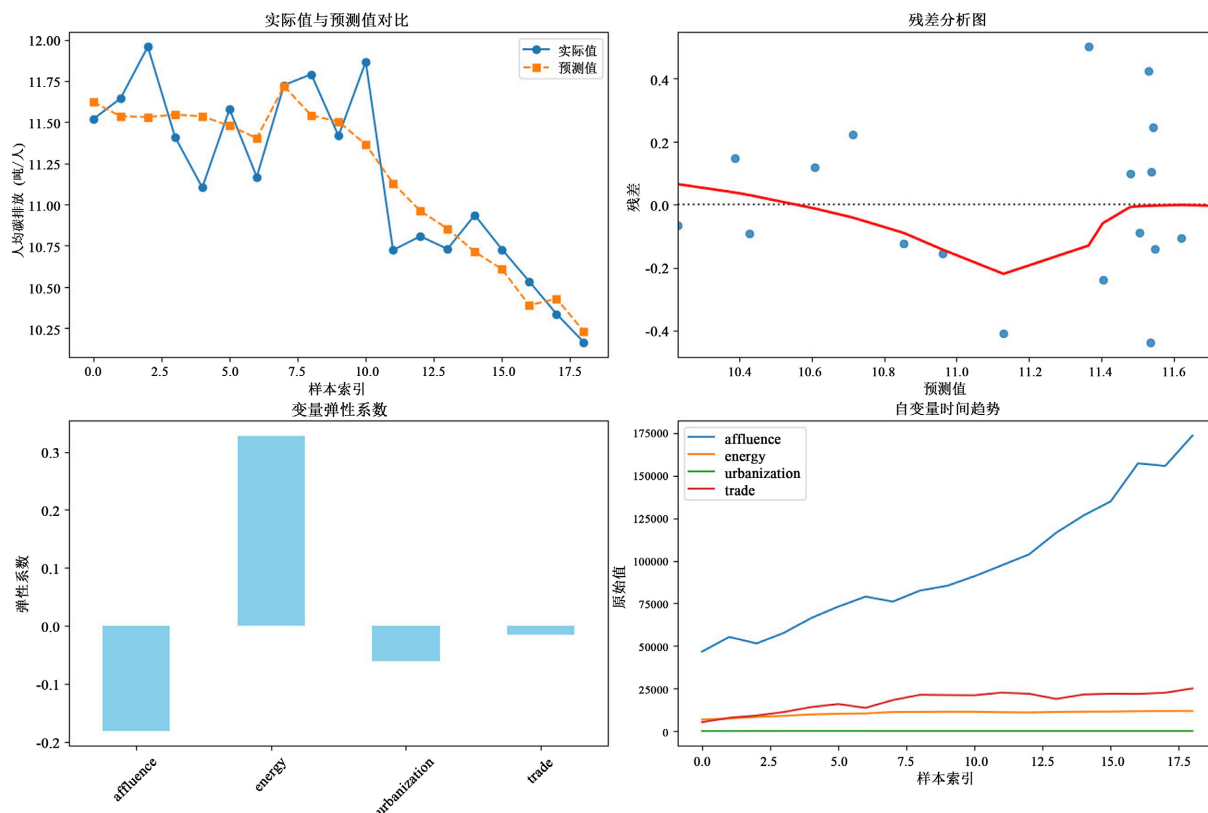


Figure 10. STIRPAT model plot
图 10. STIRPAT 模型图

3.4. 影响因素贡献度分析

本文对影响因素的分析主要分为三个方面: 相关系数矩阵、灰色关联分析以及建立 STIRPATA 模型进行影响因素分析。通过相关性矩阵、灰色关联分析、STIRPAT 模型及回归系数分析可得出: 城镇化率对碳排放的减少作用最为显著, 表明城镇化进程可能伴随着能源效率的提高或低碳技术的应用。

综合相关性矩阵、灰色关联分析、多重共线性修正后的 STIRPAT 岭回归、Lasso 回归结果, 结合上海产业与能源转型政策, 形成分层驱动结论:

- 1) 核心负向驱动因素: 人均 GDP、人均可支配收入, 二者通过产业结构升级、低碳研发投入、清洁能源财政补贴三条路径抑制碳排放; 第三产业占比、规上研发投入为关键中介调节变量, 放大经济增长的减排效应。
- 2) 核心正向驱动因素: 能源消耗总量, 是抬升人均碳排放的基础变量; 总人口扩张小幅增加碳排放规模。
- 3) 弱调节因素: 城镇化率、造林面积, 减排贡献有限。

上海已跨过环境库兹涅茨曲线拐点, 发达城市高经济水平不再伴随高碳排放, 依托产业转移、能源清洁化、绿色技术创新实现经济与碳排放脱钩, 该机制可为国内东部超大城市低碳政策制定提供参考。

4. 结论与展望

4.1. 研究结论

本文运用三种单一模型：ARIMA 模型、BP 神经网络模型、LSSVM 模型先对上海市 2003 年到 2021 年的人均碳排放量进行拟合预测，三种模型都在预测上表现出较为良好的预测结果。其中 ARIMA 模型的表现效果最为优秀，因此选取 ARIMA 模型与其他两个模型进行组合预测。

在数据集中时间序列线性趋势更为明显时，ARIMA 模型表现更为优秀，能更好地捕捉到线性趋势，而在非线性趋势时，BP 神经网络往往可以补偿非线性残差。因此在两种组合模型中，ARIMA-BP 神经网络组合模型在此数据集上相对于 ARIMA-LSSVM 模型表现得更为有效。

本文引入产业结构、绿色研发技术拓展指标体系，结合上海产业、能源转型政策解析驱动机制；针对 STIRPAT 模型极端多重共线性，采用岭回归、Lasso 回归校正模型，回归结果稳健可靠。分析显示人均可支配收入、人均 GDP 为碳排放核心负向驱动因子，能源消耗总量为主要正向驱动；第三产业占比、低碳研发投入是经济发展实现脱碳的核心中介变量。原始最小二乘 STIRPAT 因多重共线性系数存在偏差，经正则化修正后，人均 GDP 减排弹性稳定在 $-0.18 \sim -0.17$ 区间，量化证实上海经济增长的低碳红利。

4.2. 模型局限

对于模型预测方面，组合预测模型虽然预测精度高，但模型复杂度也较高，计算量较大，可能不适用于实时预测，并且对较大数据集的预测效果可能没有中小型数据集优秀[12]。在本研究的模型中，并未将外界影响因素等考虑在模型内，导致最终模型的预测结果可能并不能完全实时反映当下的人均碳排放明确预测值。

对于影响因素分析方面，因本文数据集较小，无法更准确地捕捉到数据之间的细致影响[13]。并且相关系数矩阵只能反映变量之间的线性相关性，无法捕捉非线性关系。本研究仅考虑了部分影响因素，可能忽略了其他潜在的影响因素，如政治因素。

4.3. 模型改进及研究展望

在预测分析方面可以尝试引入更多的影响因素，如产业结构、技术进步等，以提高模型的预测精度和解释力[14]。可以结合实时数据和历史数据，建立动态预测模型，以提高模型的实时性和适应性。可以采用更先进的组合预测方法，如集成学习方法，以进一步提高模型的预测精度[15]。后续研究可将产业结构、绿色专利、能源结构等技术经济变量作为多变量输入，构建多因子时序组合预测模型，进一步提升长期预测精度。在影响因素分析层面，持续扩充产业、技术、政策类细分指标；统一采用岭回归、Lasso、主成分回归多重方法处理自变量多重共线性，规避系数失真问题；同时可引入中介效应模型量化产业转型、技术研发在“人均 GDP - 碳排放”关系中的传导贡献，更清晰识别城市低碳转型内在路径。

参考文献

- [1] 谢雨心. 长三角地区工业高耗能行业能源消费时空格局演变与影响因素研究[D]: [硕士学位论文]. 徐州: 中国矿业大学, 2022.
- [2] 胡剑波, 罗志鹏, 李峰. “碳达峰”目标下中国碳排放强度预测——基于 LSTM 和 ARIMA-BP 模型的分析[J]. 财经科学, 2022(2): 89-101.
- [3] 赵成柏, 毛春梅. 基于 ARIMA 与 BP 神经网络组合模型的我国碳排放强度预测[J]. 长江流域资源与环境, 2012, 21(6): 665-71.
- [4] 刘淳森, 曲建升, 葛钰洁, 等. 基于 LSTM 模型的中国交通运输业碳排放预测[J]. 中国环境科学, 2023, 43(5): 2574-2582.

-
- [5] 纪广月. 基于灰色关联分析的BP神经网络模型在中国碳排放预测中的应用[J]. 数学的实践与认识, 2014, 44(14): 243-249.
- [6] 王艳, 吴杨. 长三角地区碳排放量分析及预测[J]. 安徽农业学报, 2023, 50(6): 1051-1058.
- [7] 潘思羽, 张美玲. 基于BP神经网络的甘肃省二氧化碳排放预测及影响因素研究[J]. 环境工程, 2023, 41(7): 61-68.
- [8] 杨振, 李泽浩. 中部地区碳排放测度及其驱动因素动态特征研究[J]. 生态经济, 2022, 38(5): 13-20.
- [9] 陈晶娜. 基于LMDI和STIRPAT模型的辽宁省碳排放影响因素分析及预测[D]: [硕士学位论文]. 沈阳: 沈阳师范大学, 2023.
- [10] 杨俊祺, 范晓军, 赵跃华, 等. 基于PSO-BP神经网络的山西省碳排放预测[J]. 环境工程技术学报, 2023, 13(6): 2016-24.
- [11] He, Y., Wang, L.G., Zhu, H., Song, W. and Zhan, X.Y. (2023) Tourism Carbon Emission Forecasting, the Decoupling Effect and Its Driving Factors in the Yangtze River Economic Belt under the “Double Carbon” Target. *Journal of Resources and Ecology*, 14, 1329-1343. <https://doi.org/10.5814/j.issn.1674-764x.2023.06.020>
- [12] 邵臻, 朱国伟, 杨昌辉, 等. 基于多时空融合的区域间道路交通碳排放预测模型[J]. 系统科学与数学, 2026, 46(5): 1543-1560.
- [13] 青霞, 史维良. 河南农业碳排放驱动效应及情景预测研究[J]. 科技与经济, 2025, 38(2): 46-50.
- [14] 雨洁, 金晓斌, 洪星明, 等. 基于Lasso-Transformer神经网络模型的海南省碳排放预测[J]. 环境科学, 2026, 47(2): 781-792.
- [15] 陈宇杰, 陈建楠, 魏汝融. 人工智能与大数据驱动下的可持续发展——碳排放预测与减排策略综合研究[J]. 产业科技创新, 2025, 7(1): 110-113.