

# 数学建模方法对碳排放、碳达峰问题的研究

李伟, 梁旭, 王宇峰

西安电子科技大学数学与统计学院, 陕西 西安

收稿日期: 2024年6月27日; 录用日期: 2024年7月24日; 发布日期: 2024年7月31日

## 摘要

以河北省碳排放为研究目标, 本文基于传统KAYA恒等式和STIRPAT模型, 提出了一个改进的STIRPAT模型, 并用于预测河北省碳排放以及碳达峰情况。首先使用ARIMA模型、灰色预测方法、长短期记忆(LSTM)网络预测模型分别对改进STIRPAT模型中的驱动指标进行了预测和比较, 得到了各指标的最优预测方法。之后, 结合时序型预测方法, 分别采用BP神经网络回归、Lasso回归和岭回归方法分别对STIRPAT模型种的碳排放进行了预测。结果表明, 随着第三产业占比的持续扩大和清洁能源占比的增加, 河北省的碳排放总量将呈现下降趋势。其中Lasso回归方法具有最小均方误差和最优拟合优度, 这种回归方法预测出河北省的碳排放总量在2035年将达到456.8百万吨, 并在2023年左右达到碳排放峰值。河北省的碳排放预测对同类省份的碳排放情况给出了参考, 这对于中国全国实现2035年碳达峰目标具有重要意义。

## 关键词

碳排放预测, 改进STIRPAT模型, 时序型预测方法, 回归型预测方法

# Research on Carbon Emissions and Carbon Peaking by Mathematical Modeling Methods

Wei Li, Xu Liang, Yufeng Wang

School of Mathematics and Statistics, Xidian University, Xi'an Shaanxi

Received: Jun. 27<sup>th</sup>, 2024; accepted: Jul. 24<sup>th</sup>, 2024; published: Jul. 31<sup>st</sup>, 2024

## Abstract

Taking carbon emissions in Hebei Province as the research objective, this paper proposes an improved STIRPAT model based on the traditional KAYA identity and STIRPAT model, and uses it to predict carbon emissions and carbon peaking for Hebei Province. Firstly, the ARIMA model, Grey prediction method, and Long-Short Term Memory (LSTM) network were used to predict and com-

pare the indicators in the improved STIRPAT model, and the optimal prediction method for each indicator was obtained. Afterwards, combined with time-series prediction methods, BP neural network regression, Lasso regression and ridge regression methods were used to predict carbon emissions in the STIRPAT model, respectively. The results indicate that with the continuous expansion of the proportion of the tertiary industry and the increase of the proportion of clean energy, the total carbon emissions in Hebei Province will show a downward trend. The Lasso regression method has the minimum mean-square error and optimal fitness. This regression method predicts that the total carbon emissions in Hebei Province will reach 456.8 million tons by 2035 and reach carbon peaking around 2023. The carbon emission forecast of Hebei Province provides a hint for the carbon emission of similar provinces, which is of great significance for China to achieve the carbon peaking target by 2035.

## Keywords

Carbon Emission Prediction, Improved STIRPAT Model, Time-Series Prediction Methods, Regression Prediction Methods

Copyright © 2024 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

习近平总书记在 2020 年联合国有关会议上提出[1]: 中国将力争于 2030 年之前实现碳达峰, 于 2060 年之前实现碳中和。国务院于 2021 年 10 月 26 日印发了《2030 年前碳达峰行动方案》[2], 将碳达峰目标列为 2021 年八大重点目标之一。可见, 碳达峰、碳中和已成为我国经济发展目标中的重中之重, 该目标的逐步实现也必将推动中国经济更加健康、稳定、安全的发展。

自第一次工业革命以来, 能源消费对环境的影响逐渐引起全球关注。绿色发展、绿色经济、减碳等概念逐渐普及。中国在上世纪经历了粗放式经济发展阶段, 导致严重的环境污染。进入 21 世纪, 随着“科学发展观”和“绿水青山就是金山银山”理念的提出, 绿色发展已成为社会和学术界关注的热点。

早在 2009 年, 林伯强等[3]就利用传统的环境库兹涅茨模拟与 LMDI 和 STIRPAT 模型研究了二氧化碳的达峰时间, 并预测出实现的时间约为 2020 年左右。朱永彬等[4]则在限定技术发展速度条件下预测中国碳达峰时间应在 2040 年之前。渠慎宁等[5]则基于 STIRPAT 模型对中国碳排放时间序列数据进行了回归, 并在限定经济社会发展速度情况下, 认为碳达峰时间会出现在 2020~2045 年之间。通过限制指标变化预测碳排放总量有一定意义, 但对于实际碳排放预测没有太大实际价值。杜强等[6]使用一个改进的 IPAT 模型对中国 2010~2050 年的碳排放总量进行了预测, 认为中国碳排放总量将于 2030 年达峰, 并在 2030~2050 年期间以 2.09%速度减少。进一步地, 杜强等[7]于 2013 年将全国各省分成五个群体, 针对每个群体分别使用 Logistic 回归模型研究了逐年累计碳排放量与人均累计排放量, 并对中国各省碳排放进行了预测。黄蕊等[8]基于 STIRPAT 模型, 结合多个经济社会指标, 在固定各指标变化速度情况下使用岭回归方法对江苏省碳排放进行了研究。王勇等[9]对 STIRPAT 模型进行了改进, 提出了门限-STIRPAT 模型, 并用其探究了中国超大城市碳达峰影响因素, 认为人口、GDP、能源强度对碳排放影响最大, 能源强度对碳排放影响还呈现阶段性特征, 这为本文中碳排放预测指标体系建立以及 STIRPAT 模型的改进提供了借鉴。还有很多相关的工作例如: 刘凯诚[10]使用改进之后的 KAYA 模型对碳达峰条件进行了探

索。李晓易等[11]针对交通运输领域碳达峰、碳中和实现路径进行了研究,提出“分类施策、远近结合、先易后难、控增量调存量、积极稳妥推进、梯次有序达峰”的路径策略。

从现有文献可以看出,目前关于碳排放问题的研究主要是针对中国整个国家,而单独省份或地区层面的碳排放研究较少。然而中国幅员辽阔,南北经济差异较大,基于中国整个国情的碳排放预测并不适合某个省份的碳排放计划,得到的预测结果也很难在具体的省份得到验证。为此本文将针对性地对特定省份的碳排放问题进行研究,研究结果可为相似经济发展水平的省份提供建设性的政策启示。除此之外,以上作者在预测上多选择根据以往经验人为限制社会影响指标,进而对未来的碳排放做出预测,主观因素较大,并不能准确预测出未来实际碳排放。

影响碳排放的因素有很多,并且可能相互关联,因此对多因素进行分解、研究各因素对碳排放的影响差异,是进行碳排放预测必不可少的首要任务。LMDI 分解模型以及 STIRPAT 模型被证明是处理多影响因素分解的有效方式之一。另外,碳排放预测模型的使用是研究“双碳”目标的重要组成,预测结果的正确性直接关系到该省份或地区的经济发展战略。因此本文将改进 STIRPAT 分解模型,提出多种预测模型,利用数学建模的理论与方法,对碳排放、碳达峰等问题进行系统研究。根据预测结果和分析,为具体省份未来碳排放政策提出合理的建议和展望。

## 2. 碳排放模型与指标

### 2.1. LMDI 分解模型和 STIRPAT 模型

LMDI (对数平均迪氏指数法, Logarithmic Mean Divisia Index)是基于指数分解法(IDA)发展而成的一种因素分解法,这种方法以对数的形式对影响碳排放的多种因素进行加权,同时消除分解残差,具有完全分解因子、无残差项等主要优势。LMDI 分解模型通常与 Kaya 恒等式[12]联合使用。Kaya 恒等式于 1989 年由日本经济学家 Yoichi Kaya [13]在 IPCC 会议上提出。Kaya 认为,碳排放总量主要取决于一个国家的能源消费、国民生产总值和人口数量,所以将碳排放量分解为这些影响因素互相关联的形式,其表达式为:

$$CE = \frac{CE}{PE} * \frac{PE}{GDP} * \frac{GDP}{POP} * POP \quad (1)$$

式中,CE——碳排放量(单位:百万吨),PE——能源总消费量(单位:万吨),GDP——国民生产总值(单位:亿元),POP——人口总数(单位:万人)。

另一方面,基于 1970 年 Ehrlich 和 Holden 提出的 IPAT 恒等式[14], Thomas Dieta 和 Eugene Rosa [15]提出了 STIRPAT (Stochastic Impacts by Regression on Population, Affluence, and Technology)模型,该模型常用于评估人类行为对环境产生的影响,特别是常用于反映人口、富裕程度、技术等因素对环境的综合影响。因该模型中引入了随机性,扩展了 IPAT 恒等式的应用范畴,使得该模型可以更加深入的探索不同自变量与因变量之间的关系,应用更具有广泛性和成熟度,因此成为环境压力评价的一个重要工具。其表达式如下:

$$I = aP^b \times A^c \times T^d \times \varepsilon \quad (2)$$

式中: $I$ ——环境影响, $P$ ——人口(population), $A$ ——财富(affluence), $T$ ——技术(technology), $a, b, c, d$ ——参数, $\varepsilon$ ——误差项。

### 2.2. 碳排放预测指标体系

基于 LMDI 分解思想和 STIRPAT 模型,本文将以河北省为例,以该省 1997~2021 年内能源消费的碳排放数据为主要依据[16],建立表 1 所示的预测指标体系。之后对 STIRPAT 模型进行改进,定性定量的

分析碳排放与各项因素的关联性，并根据改进的 STIRPAT 模型对河北省未来的碳排放量进行预测。改进的 STIRPAT 模型如下：

$$C = a \times \prod_{i=1}^8 \alpha_i^{\alpha_i} \times \bar{\varepsilon} \quad (3)$$

其中， $C$  为碳排放量， $\alpha_i$  为对应指标的待定指数系数， $a$  为未知常数项系数， $\bar{\varepsilon}$  为误差项。对(3)式两边同时取对数运算，将乘积幂次形式转变为求和形式，有

$$\ln C = \sum_{i=1}^8 \alpha_i \ln \alpha_i + \ln a + \ln \bar{\varepsilon} \quad (4)$$

公式(4)即为本文提出的改进 STIRPAT 模型。

**Table 1.** Carbon emission projection indicator system

**表 1.** 碳排放预测指标体系

一级指标	二级指标	字符表示
人口	人口总量	$\alpha_1$
经济	GDP 全年总量	$\alpha_2$
	第三产业 GDP	$\alpha_3$
能源技术	能源消费总量	$\alpha_4$
	煤炭消费比例	$\alpha_5$
	石油消费比例	$\alpha_6$
	天然气消费比例	$\alpha_7$
	其他能源消费比例	$\alpha_8$
碳排放	全年碳排放总量	$C$

### 3. 碳排放预测模型

本文使用数据来自《河北统计年鉴》[16]，部分数据样本如表 2 所示。首先利用数据对改进 STIRPAT 模型中的各项指标进行预测。在数学建模理论中有多种预测模型和方法，每种方法均有自己独特的优势和适用的场景，对同一问题的分析也有可能得到相差各异的结果[17]。本文将采用多种预测模型对河北省 2035 年前碳排放量进行预测，并分析河北省碳达峰的预计时间。预测方法分为两种类型，分别对应两种预测思路。图 1 给出了两种不同类型预测模型的工作流程。

**Table 2.** Carbon emission driver indicator system data

**表 2.** 碳排放驱动因素指标体系数

年份	人口	GDP/亿元	第三产业/亿元	能源消费总量 万吨标准煤	煤炭	石油 比例	天然气比例	一次电力及其他 能源所占比例	CO <sub>2</sub> 排放 (mt)
2003	6769	6333.6	2346.8	15297.89	92.78	6.49	0.66	0.07	320.69
2004	6809	7588.6	2787.1	17347.79	91.14	8.01	0.75	0.1	365.68
2005	6851	8773.4	3304.2	19835.99	91.82	7.45	0.61	0.12	408.93

续表

2006	6898	10043	3869.3	21794.09	91.59	7.64	0.67	0.1	409.11
2007	6943	12152.9	4566.4	23585.13	92.36	6.87	0.68	0.09	453.98
2008	6989	14200.1	5261.8	24321.87	92.31	6.67	0.94	0.08	481.98
2009	7034	15306.9	6020.7	25418.79	92.51	6.21	1.21	0.07	511.06
2010	7193.6	18003.6	7060	26201.41	89.71	7.75	1.51	1.03	569.37
2011	7231.86	21384.7	8406.4	28075.03	89.09	8.12	1.66	1.13	623.46
2012	7262	23077.5	9243.8	28762.47	88.86	7.48	2.04	1.62	642.31
2013	7287.59	24259.6	9939.3	29664.38	88.69	7.22	2.23	1.86	657.72
2014	7322.9	25208.9	10567.3	29320.21	88.46	6.98	2.54	2.02	624.59
2015	7345.2	26398.4	11778.4	31036.73	88.83	5.99	3.13	2.05	639.37
2016	7374.99	28474.1	13059.3	31458.05	87.33	6.23	3.42	3.02	614.57
2017	7409.14	30640.8	14732.8	32082.56	86.05	6.14	3.94	3.87	541.90
2018	7426.37	32494.6	16252	32185.24	83.61	6.47	5.49	4.43	595.29
2019	7446.56	34978.6	18066.5	32545.43	81.96	5.86	6.61	5.57	593.43
2020	7463.84	36013.8	18368.4	32782.76	80.51	5.67	7	6.82	589.56
2021	7448	40391.3	19996.7	32590.07	76.58	6.63	7.59	9.2	579.36

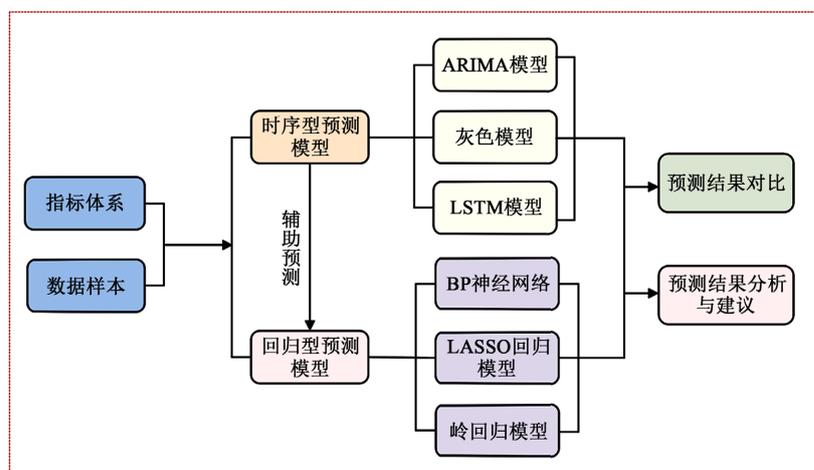


Figure 1. Ideas for total carbon emission forecast in Hebei Province

图 1. 河北省碳排放总量预测思路

### 3.1. 时序型预测方法

碳排放数据本身就是一列时序数据，于是可以考虑选用相应的时序模型进行预测。时序预测方法包括传统的数值类型预测方法以及机器学习预测方法。本文使用 ARIMA 模型、灰色预测方法和 LSTM 循环神经网络对碳排放进行预测。

#### 3.1.1. ARIMA (自回归综合移动平均模型)

是一种广泛应用于时间序列预测的统计模型，由 George E. P. Box 和 Gwilym M. Jenkins [18] 在 1970

年代初期提出，并在之后不断完善。

ARIMA 模型结合了自回归(AR)、差分(I)和移动平均(MA)的概念。ARIMA 模型包含三个参数( $p, d, q$ )，分别对应 AR 模型的项数  $p$ 、使数据变平稳的差分次数  $d$ 、以及移动平均项数  $q$ ，这些参数均可通过观察自相关图(ACF)和偏自相关图(PACF)来选择合适的  $p$ 、 $d$  和  $q$  值，对应的模型为：

$$W_t = c + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (5)$$

式中  $W_t$  为原始时间序列  $X_t$  经过  $d$  次差分后的序列， $c$  为常数项， $\phi_1, \phi_2, \dots, \phi_p$  为自回归部分的参数，表示过去值对当前值的影响。 $\theta_1, \theta_2, \dots, \theta_q$  为移动平均部分的参数，表示过去误差项对当前值的影响。 $\varepsilon_t$  为时间序列在时刻  $t$  的误差项，通常假设它们是独立同分布的随机变量。ARIMA 模型充分利用了数据本身的时序信息，用相对简单的线性形式进行预测，具有稳定性好、准确性高、适用性强等诸多优点。经过分析，河北省碳排放数据采用可确定三个参数值为 ARIMA(2, 2, 1)。

**Table 3.** Comparison of predicted effects

**表 3.** 预测效果对比

MSE 参数值灰色预测		ARIMALSTM	
人口总量/万人	2775	1845	<b>668</b>
GDP/亿元	1,366,853	<b>370,576</b>	5,206,193
第三产业	779,842	<b>116,558</b>	127,061
能源消费总量	5,894,940	796,383	<b>231,751</b>
煤炭	6.669	<b>4.29</b>	6.41
石油	<b>0.375</b>	0.65	0.42
天然气所占比例	1.04	0.476	<b>0.16</b>
其他能源所占比例	1.5	<b>0.96</b>	1.13

### 3.1.2. 灰色预测模型

灰色系统是由华中科技大学的邓聚龙教授[19]于 80 年代初创立，该系统作为新兴的横断学科，在短短的二十年里已得到了长足的发展。灰色预测方法(Grey Forecasting Method)是一种基于灰色系统理论的预测方法，用于处理具有少量数据、缺乏完整信息或非线性灰色机理性关系的问题。该方法多应用于经济、工程、管理和其他领域的预测和决策分析中。

### 3.1.3. LSTM 模型

长短时记忆网络(LSTM)预测方法是一种特殊的循环神经网络(RNN)，该模型于 1997 年由 Hochreiter S, Schmidhuber J [20]共同提出，并在数据预测、自然语言领域广泛应用。RNN 循环神经网络同样包括输入层、隐藏层、输出层，相比于传统神经网络不同的是，RNN 隐藏层还包含自传播过程，这使得 RNN 能很好地提取时序变量中的时序信息。但传统的 RNN 存在梯度消失或梯度爆炸的问题，导致对长序列的建模能力较弱。为了解决这个问题，本文使用 LSTM 预测方法进行预测。LSTM 通过在 RNN 中创建记忆门、输入门和输出门进而降低计算维度，从而完成对长序列的预测任务。

ARIMA(2, 2, 1)模型、灰色预测模型、LSMT 模型预测结果的均方误差见表 3，其中均方误差的 MSE 的计算公式为：

$$MSE = \frac{1}{n} \sum_{i=1}^n (\text{实际值}_i - \text{预测值}_i)^2 \quad (6)$$

这里我们取 1997~2021 年碳排放数据中前 80% 数据作为训练集, 后 20% 数据作为测试集, 表中加黑的数据是当前指标得到的最小均方误差, 对应的预测模型即为对应当前指标的最佳预测模型。例如驱动指标人口总数, 适合于该指标因素的最优预测模型为 LSMT 模型。

### 3.2. 回归型预测方法

完成各项驱动指标的预测后, 本文将基于改进的 STIRPAT 模型(4)采用回归型预测方法包括 BP 神经网络回归方法[21]、Lasso 回归方法[22]、岭回归方法[23]等对河北省 2022~2035 年的碳排放总量进行预测, 具体预测流程见图 2。为检验回归模型的有效性, 本文将指标预测数据的 75% 作为训练数据, 余下的 25% 作为检验数据进行拟合优度检验。

改进的 STIRPAT 模型(4)可改写为

$$\ln C = f(\ln \alpha_1, \ln \alpha_2, \dots, \ln \alpha_8) \quad (7)$$

式中  $\alpha_i (i = 1, \dots, 8)$  分别对应公式(4)中的驱动指标因素。

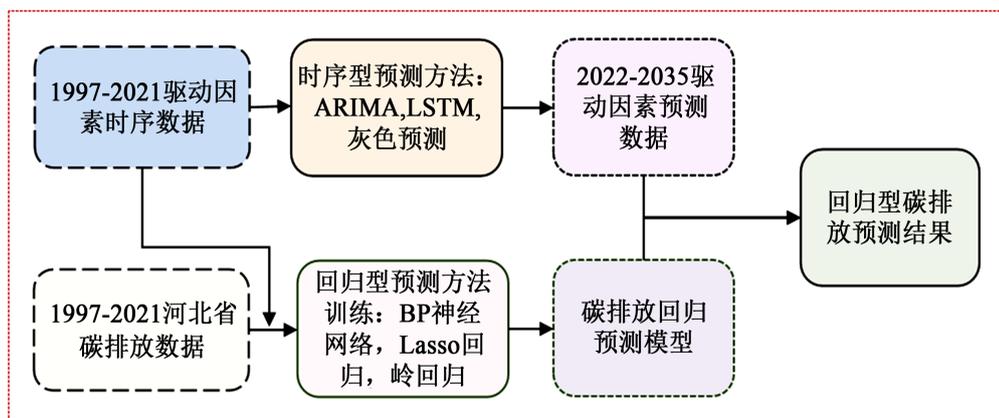


Figure 2. Flowchart for total carbon emissions projection

图 2. 碳排放总量预测流程图

#### 3.2.1. BP 神经网络回归

BP 神经网络回归是一种使用反向传播算法(Backpropagation)进行训练的神经网络模型, 由于具有结构简单、工作状态稳定、易于硬件实现等优点在众多的神经网络模型中 BP 网络的应用最为广泛[21], 可用于回归问题。它可以学习输入与输出之间的非线性映射关系, 并用于对新的输入数据进行回归预测。BP 神经网络回归的基本结构包括输入层、隐藏层和输出层。输入层接受输入数据的特征, 隐藏层包含一层或多层神经元, 输出层产生回归结果。图 3 给出了 BP 神经网络的具体结构。

以单层 BP 神经网络为例, 预测模型可表示为下式

$$\hat{y} = g_2(\text{net}_2) = g_2(v^T g_1(w^T x + b_1) + b_2) \quad (8)$$

其中  $\hat{y}$  为模型输出,  $g_1, g_2$  分别为输入层到隐藏层和隐藏层到输出层的激活函数,  $w^T, v^T$  分别为输入层到隐藏层和隐藏层到输出层的权重矩阵,  $b_1, b_2$  为偏置常数。模型会根据输出层结果计算损失函数对模型参数进行前馈调整。

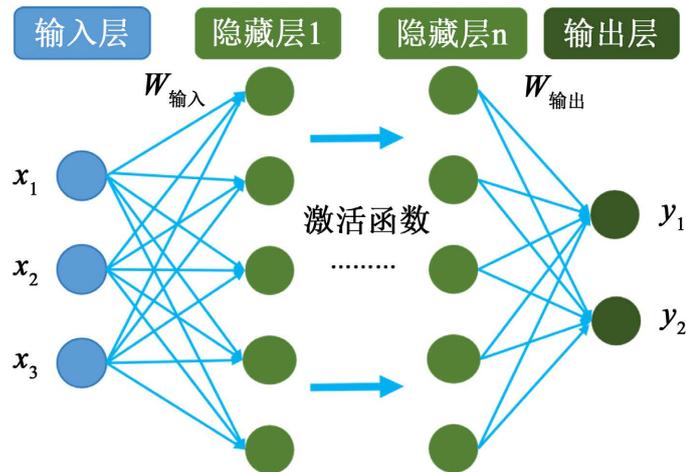


Figure 3. BP neural network prediction structure  
图 3. BP 神经网络预测结构

### 3.2.2. 岭回归

岭回归分析是一种专门用于共线性数据分析的有偏估计方法，它实际上是一种改良的最小二乘法 [22]。在岭回归中，损失函数由两部分组成：平方误差项和正则项。平方误差项用于拟合数据，而正则项则对模型的系数进行惩罚，使得系数的平方和尽可能小。

当输入特征之间存在高度相关性时，岭回归方法可以发挥其优势处理多重共线性，岭回归通过约束系数的平方和，能很好地避免这一问题，提高了模型的稳定性。

### 3.2.3. Lasso 回归

Lasso 回归(Least Absolute Shrinkage and Selection Operator Regression)是一种线性回归的方法，可用于特征选择，由 Tibshirani R [23]于 1996 年提出，并在机器学习与统计学中有广泛的应用。它通过在损失函数中加入正则项来实现这一点。Lasso 回归更倾向于选择与目标变量有强相关的特征变量，这种特点会使得 Lasso 回归方程中一些特征的系数为零。相比较岭回归，Lasso 回归倾向于对系数较小的变量特征进行惩罚，进而优化回归模型，使其变得简单。

## 4. 预测结果对比与分析

### 4.1. 碳排放驱动因素指标时序预测结果与分析

前面我们根据均方误差值最小分别选择了驱动因素的最佳预测模型，以方便我们在之后对碳排放总量进行回归型预测。图 4 与图 5 给出了八个驱动因素的预测结果。

由图 4 可以看出，与碳排放正相关的人口总量以及能源消费总量随时间增长逐渐达到峰值，呈现典型的 S 型曲线特征。值得注意的是，河北省的第三产业 GDP 以及其在总体 GDP 中所占比例在持续增长。根据预测：2035 年河北省第三产业 GDP 比例将达到 61.4%。

从图 5 中可以发现，与河北省碳排放正相关的煤炭、石油在之后的预测中都有明显的下降，预测结果显示：至 2035 年，煤炭使用在全部的能源消费中所占比例由 2010 年的 89.7% 降至 2035 年的 64.5%，石油所占比例也从 7.8% 降至 4.6%，而与碳排放负相关的绿色能源天然气和一次能源及其它能源，在能源消费中的使用比例分别从 2010 年的 1.5%、1.0% 跃升至 5.7% 和 16.2%。这表明从能源消费角度看，河北省的能源结构正向更加绿色更加健康的方向调整。

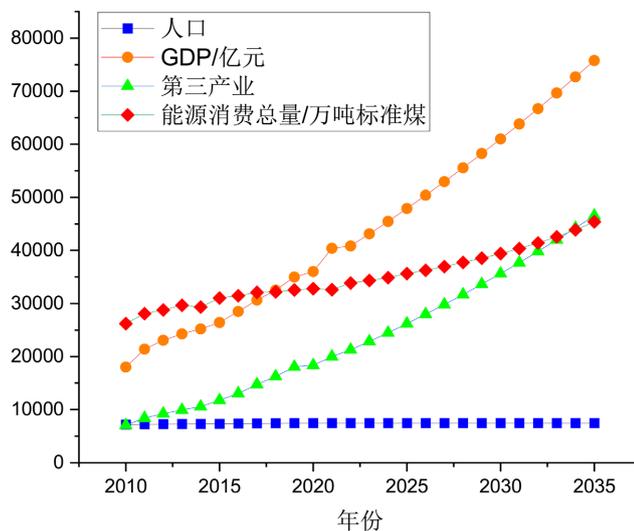


Figure 4. Population, GDP, tertiary sector, energy consumption forecasts  
图 4. 人口、GDP、第三产业、能源消费量预测图

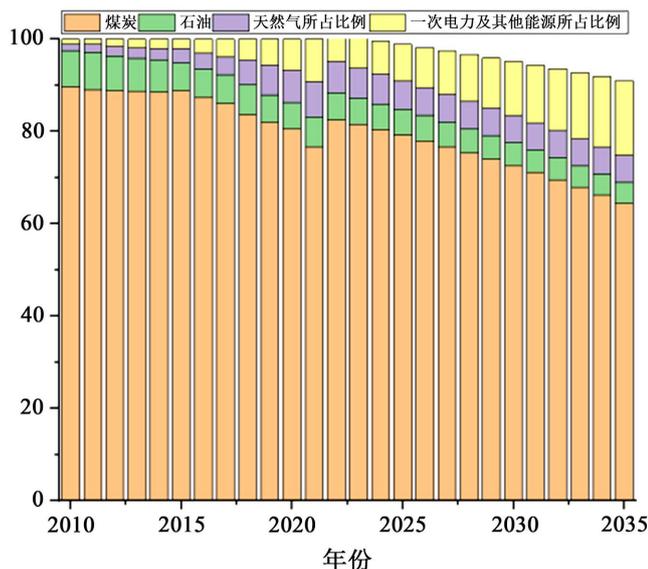


Figure 5. Projected energy consumption structure  
图 5. 能源消费结构预测图

#### 4.2. 河北省碳排放总量预测结果与分析对比

本文使用两种类型共六种方法对河北省 2022~2035 年的碳排放情况做了预测，并使用其中未参与训练的 20% 数据对六种预测模型进行了评估。表 4 集中展示了六种预测方法对应的检验均方根误差 RMSE 和拟合优度  $R^2$ 。

参数反映在所有预测方法中预测效果最好的为 Lasso 回归，其均方根误差 RMSE 和拟合优度  $R^2$  均表现最优，分别为 16.06 和 0.98，其中均方根误差 RMSE 和拟合优度  $R^2$  计算公式为：

$$\text{RMSE} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}, R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \quad (9)$$

**Table 4.** Comparison of predicted effects  
**表 4.** 预测效果对比

预测方法类型		时序预测方法		RMSE R <sup>2</sup>	
时序类型预测方法		ARIMA 时序预测		35.42	0.97
		LSTM 时序预测		18.13	0.82
		灰色预测		79.54	0.79
回归类型预测方法		BP 神经网络回归		23.11	0.98
		<b>Lasso 回归</b>		<b>16.06</b>	<b>0.98</b>
		岭回归		22.01	0.97

需要指出的是，这里的预测效果并不能完全反应十五年后结果的预测精度。事实上，随着预测时间的增加，预测精度会越来越低。但由于未来信息是未知的，所以本文使用 25% 的碳排放数据进行检验，从而对几种预测方法做出比较。回归类型预测方法中，三种回归方法都得到了相应的回归参数和回归方程。

BP 神经网络回归没有一个明确的回归方程来表示输入与输出之间的关系，而是通过训练网络的参数来学习这种关系。最终训练得到的模型隐藏层层数为 2 层，第一层有 5 个神经元，第二层有 10 个神经元。

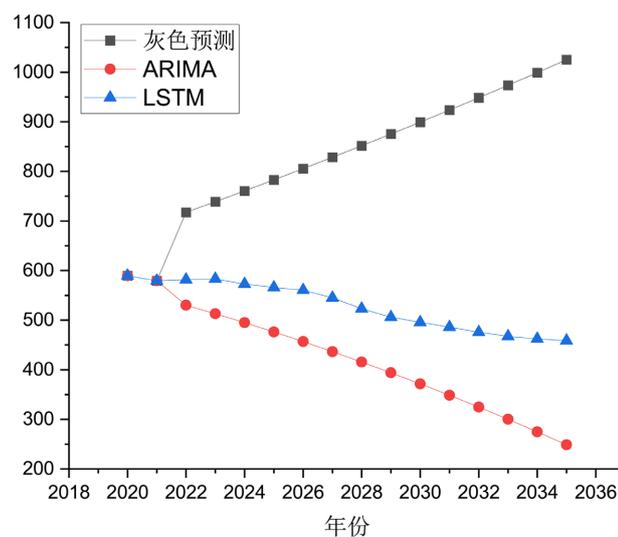
Lasso 回归方法最终得到回归方程为：

$$\ln C = 0 \ln \alpha_1 + 690.93 \ln \alpha_2 - 317.99 \ln \alpha_3 + 54.06 \ln \alpha_4 + 197.85 \ln \alpha_5 + 72.01 \ln \alpha_6 + 0 \ln \alpha_7 + 151.87 \ln \alpha_8 - 25.13 \quad (10)$$

岭回归方法最终得到回归方程为：

$$\ln C = 117.54 \ln \alpha_1 + 113.27 \ln \alpha_2 + 80.53 \ln \alpha_3 + 145.26 \ln \alpha_4 + 144.04 \ln \alpha_5 + 92.01 \ln \alpha_6 + 2.64 \ln \alpha_7 + 59.22 \ln \alpha_8 + 19.13 \quad (11)$$

图 6 与图 7 分别给出了两种类型预测方法的碳排放预测结果。



**Figure 6.** Projected carbon emission profiles by time series type  
**图 6.** 时序类型碳排放预测曲线图

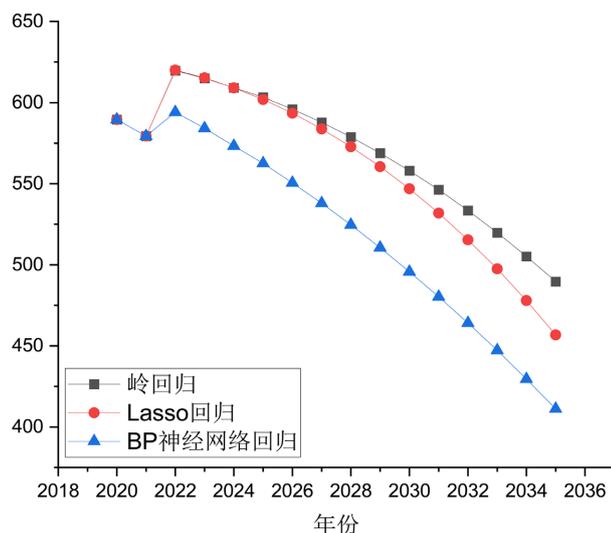


Figure 7. Regression type carbon emission projection curves  
图 7. 回归类型碳排放预测曲线

时序类型的三种方法在预测趋势上呈现出很大的不同。灰色预测指出 2022~2036 年碳排放将呈现近似线性增长态势，并于 2035 年河北省全省碳排放量将达到 1025 百万吨；LSTM 预测结果显示河北省碳排放总量将于 2023 年左右达到峰值，为 589.6 百万吨，之后逐年下降，至 2035 年碳排总量降低至 458.6 百万吨；而 ARIMA 预测结果显示河北省碳排总量将在之后 15 年逐年降低，但下降趋势逐渐放缓，2035 年最终降至 248.9 百万吨。灰色预测的结果呈逐年上升趋势与灰色预测的特点有关，灰色预测只能提取最近一段时间的变化趋势，对于短时间的预测有良好的效果，对于长时间的预测则有明显的缺陷。

回归类型的三种方法在预测趋势上呈现出很强的相似性，均显示河北省碳排放总量正在逐年下降，其中 BP 神经网络回归预测至 2035 年河北省碳排放总量将达到 411.2 百万吨；岭回归与 Lasso 回归结果相近，预测至 2035 年河北省碳排放总量分别为 489.7 百万吨与 456.8 百万吨。三种回归方法在峰值出现的时间上基本一致，均预测河北省将在 2023 年附近出现碳达峰现象。

### 4.3. 结论与建议

本文通过改进的 STIRPAT 模型，对河北省的碳排放及碳达峰进行了预测分析。主要研究发现如下：

1) 碳排放趋势：在六种预测模型的预测结果中，除了灰色预测对碳排放的预测结果呈逐渐递增趋势外，其余模型均显示河北省的碳排放总量将呈现下降趋势。

2) 碳达峰时间：预测结果显示，河北省的碳排放总量预计在 2023 年达到峰值，随后逐渐下降。河北省碳排放能在 2035 年之前完成达峰要求。

3) 在碳排放预测方法中回归预测方法效果明显优于时序预测方法，其中 Lasso 回归预测效果最佳。灰色预测方法仅能提取最近一段时间的变化趋势，对于较长时间的预测效果不够理想。

根据预测，河北省作为能源消费大省，能够率先通过调整产业结构，不断提高第三产业比例，优化能源消费结构，提高绿色能源使用比例，并提前实现碳排放达峰任务，这对于全国实现 2035 碳达峰目标意义重大。

基于以上发现和所做工作提出以下合理建议：

1) 河北省应继续优化产业结构，减少高能耗、高污染产业的比重，积极发展第三产业和高技术产业，以实现经济与环境的双赢。

2) 增加清洁能源和可再生能源的使用比例, 减少对煤炭等传统能源的依赖, 推动能源结构向绿色、低碳方向转型。

3) 未来研究可以对比更多预测模型, 增加更多变量, 以提高预测的效果。同时, 开展不同省份间的对比研究, 以制定更精准的减碳策略。特别是针对经济发展水平和能源消费结构相似的省份, 进行深入研究有助于推广最佳实践经验。

通过这些措施, 河北省不仅能够提前实现碳达峰目标, 还将为全国的碳减排工作提供宝贵经验和示范效应。

## 参考文献

- [1] 习近平出席第七十六届联合国大会一般性辩论并发表重要讲话[N]. 光明日报, 2021-09-22(001).
- [2] 2030年前碳达峰行动方案[N]. 人民日报, 2021-10-27(007).
- [3] 林伯强, 蒋竺均. 中国二氧化碳的环境库兹涅茨曲线预测及影响因素分析[J]. 管理世界, 2009(4): 27-36.
- [4] 朱永彬, 王铮, 庞丽. 基于经济模拟的中国能源消费与碳排放高峰预测[J]. 地理学报, 2009, 64(8): 935-944.
- [5] 渠慎宁, 郭朝先. 基于 STIRPAT 模型的中国碳排放峰值预测研究[J]. 中国人口·资源与环境, 2010, 20(12): 10-15.
- [6] 杜强, 陈乔, 陆宁. 基于改进 IPAT 模型的中国未来碳排放预测[J]. 环境科学学报, 2012, 32(9): 2294-2302.
- [7] 杜强, 陈乔, 杨锐. 基于 Logistic 模型的中国各省碳排放预测[J]. 长江流域资源与环境, 2013, 22(2): 143-151.
- [8] 黄蕊, 王铮, 丁冠群, 等. 基于 STIRPAT 模型的江苏省能源消费碳排放影响因素分析及趋势预测[J]. 地理研究, 2016, 35(4): 781-789.
- [9] 王勇, 许子易, 张亚新. 中国超大城市碳排放达峰的影响因素及组合情景预测——基于门限-STIRPAT 模型的研究[J]. 环境科学学报, 2019, 39(12): 4284-4292.
- [10] 刘铠诚, 何桂雄, 郭炳庆, 等. 基于改进 KAYA 算法的碳排放达峰条件分析和路径水平预测[J]. 计算机工程与应用, 2019(55): 351-354.
- [11] 李晓易, 谭晓雨, 吴睿, 等. 交通运输领域碳达峰\碳中和路径研究[J]. 中国工程科学, 2021, 23(6): 15-21.
- [12] Yang, J., Cai, W., Ma, M., Li, L., Liu, C., Ma, X., et al. (2020) Driving Forces of China's CO<sub>2</sub> Emissions from Energy Consumption Based on Kaya-LMDI Methods. *Science of the Total Environment*, **711**, Article ID: 134569. <https://doi.org/10.1016/j.scitotenv.2019.134569>
- [13] Wang, C., Chen, J. and Zou, J. (2005) Decomposition of Energy-Related CO<sub>2</sub> Emission in China: 1957-2000. *Energy*, **30**, 73-83. <https://doi.org/10.1016/j.energy.2004.04.002>
- [14] Ehrlich, P.R. and Holdren, J.P. (1971) Impact of Population Growth. *Science*, **171**, 1212-1217. <https://doi.org/10.1126/science.171.3977.1212>
- [15] Dietz, T. and Rosa, E.A. (1994) Rethinking the Environmental Impacts of Population, Affluence and Technology. *Human Ecology Review*, **1**, 277-300.
- [16] 河北省统计局, 国家统计局河北调查总队. 河北统计年鉴[M]. 北京: 中国统计出版社, 2022.
- [17] 李伟, 肖亚宁. 数学建模竞赛中数据驱动的预测方法之比较[J]. 建模与仿真, 2022, 11(6): 1539-1545.
- [18] Box, G.E.P., Jenkins, G.M., Reinsel, G.C., et al. (2015) *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Hoboken.
- [19] 邓聚龙. 灰色控制系统[J]. 华中工学院学报, 1982(3): 9-18.
- [20] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [21] 周政. BP 神经网络的发展现状综述[J]. 山西电子技术, 2008(2): 90-92.
- [22] 杨楠. 岭回归分析在解决多重共线性问题中的独特作用[J]. 统计与决策, 2004(3): 14-15.
- [23] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **58**, 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>