

碳索未来：中国碳排放预测模型与低碳发展策略

摘 要

本研究围绕中国实现 2030 年前碳达峰和 2060 年前碳中和的国家战略目标，开展了一系列分析和预测。通过综合评估影响二氧化碳排放的关键因素建立数学模型。探讨了能源消耗、工业生产、交通运输和人口规模等因素对二氧化碳排放的影响，并提出了实现低碳发展目标的技术路径。

针对问题一，本文首先从国家能源局发表的资源排放报表中利用灰色关联分析识别了与二氧化碳排放量关联度较高的因素，包括煤炭消耗量、人口数量和能源消耗等。随后，采用向量自回归（VAR）模型考虑变量之间的动态关系和时序特性对中国未来一段时间内的二氧化碳排放量进行预测。模型分析结果显示，2030 年中国的二氧化碳排放量预计将达到 19530.10 百万吨，而到 2060 年，这一数值预计将增长至 40694.18 百万吨。

针对问题二，本文着重分析 CCUS 技术及其他减排技术对中国二氧化碳排放量的影响。采用多元线性回归模型和长短期记忆（LSTM）神经网络模型预测自 2023 年起实施 CCUS 技术后中国的二氧化碳排放趋势。多元线性回归模型预测出 2030 年的二氧化碳排放量将降至 13,580.68 百万吨。而 LSTM 模型则指出到 2060 年，持续应用 CCUS 技术可将排放量进一步降低至 19,485.91 百万吨

针对问题三，本文通过分析自 1970 年来各清洁能源发展趋势，并基于此使用 ARIMA 模型预测了未来 50 年中国的发电量。预测结果显示，在实施清洁能源方案的情况下，2030 年和 2060 年中国的二氧化碳排放量将分别降至 19832.54 百万吨和 17415.79 百万吨。

针对问题四，本文分析了农地流转与农业低碳化发展的协同效应。利用熵权法和 TOPSIS 方法构建评价模型。结果表明农地流转能显著提升农业生产效率，推动低碳化发展。同时固定效应模型的评估结果进一步证实了农地流转对农业低碳化发展的正向影响。

针对问题五，本文将所用的模型和结果进行总结讨论，分析未来我国实现碳达峰、碳中和目标的可行技术路径，并撰写了一份建议报告给当地的能源管理部门。

最后，本文也对文中所建立的数学模型进行了评价和灵敏度分析。

本文综合运用多种数学模型和统计方法，全面评估了影响中国二氧化碳排放的多维度因素，并提出了切实可行的低碳技术发展路径。

关键词：碳达峰 碳中和 灰色关联分析 多元线性回归 VAR LSTM ARIMA

一、问题重述

1.1 问题背景

全球气候变化问题日益严峻，二氧化碳排放量的持续增加引发了严重的生态和环境问题。二氧化碳排放引起的温室效应导致全球气温升高、土地荒漠化、两极冰川融化、海平面上升等诸多生态变化，严重影响了人类的居住和生产活动。习近平在领导人气候峰会发表的重要讲话中指出：“中国将力争 2030 年前实现碳达峰、2060 年前实现碳中和。”为了实现碳达峰和碳中和的目标，需要对我国未来二氧化碳排放进行规划，分析煤炭高效开发和洁净化利用对减排的作用，以及洁净能源的替代效应，对未来进行预测。



图 1 中国双碳目标

1.2 问题重述

问题一：分析我国二氧化碳排放的主要来源，考虑能源消耗、钢铁产量以及人口等相关要素，建立数学模型分析现有数据，预测我国在未来一段时间内二氧化碳的排放量，给出我国在 2030 年碳达峰时二氧化碳的排放量峰值，以及 2060 年碳中和时的二氧化碳排放量。

问题二：分析通过哪些技术方法可以降低二氧化碳的排放量，分析得到技术方法对节能减排的作用效果。假设从 2023 年起开始使用这种技术方法，预测得到二氧化碳逐年减少的排放量，并计算 2030 年和 2060 年我国二氧化碳的排放量。

问题三：分析中国在核电、水电以及其他洁净能源方面的发展趋势，以及这些能源替代一部分石化能源后对于减少二氧化碳排放的效果。并且建立数学模型，预测在从 2023 年开始实施洁净能源方案的情况下，未来 50 年内中国的发电量及 2030 年、2060 年的二氧化碳排放量。

问题四：分析一种其他可以降低二氧化碳的可行方法，建立数学模型进行分析说明。

问题五：基于模型分析及其结果，探讨中国实现碳达峰和碳中和目标的技术途径，并撰写一份建议报告，为地方能源管理机构提供参考。

1.3 我们的工作

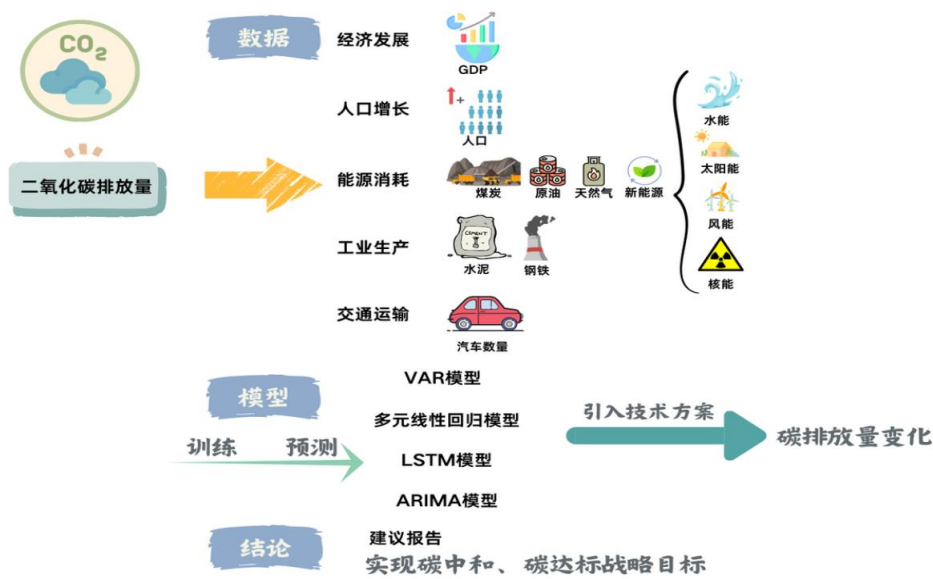


图 2 我们的工作

二、问题分析

2.1 问题一的分析

针对问题一，我们首先从国家统计局和相关学术论文中提取关键指标，包括直接排放源和间接影响因素。通过灰色关联分析识别出对排放量变化具有显著影响的因素，并进行了必要的数据处理和标准化。随后，我们使用向量自回归（VAR）模型构建了一个综合预测模型，捕捉这些因素的动态变化，预测了 2030 年和 2060 年中国的二氧化碳排放量。

2.2 问题二的分析

针对问题二，我们分析了近年来减少二氧化碳排放的主要技术方法，包括碳捕获、利用与储存（CCUS）技术、可再生能源技术、提高能效技术和电气化交通。假设从 2023 年起实施这些技术，我们使用多元线性回归模型和长短期记忆（LSTM）模型进行预测，综合得出在使用 CCUS 技术的情况下，2030 年和 2060 年我国二氧化碳的排放量。

2.3 问题三的分析

针对问题三，我们首先分析中国在核电、水电、光伏、风电、生物质能和氢能燃料电池等洁净能源领域的发展趋势。通过收集自 1970 年以来的相关历史数据，并利

用 ARIMA 模型对未来 50 年中国的发电量进行预测。最后通过长短期记忆（LSTM）模型，预测了在实施清洁能源方案的情况下 2030 年和 2060 年中国的二氧化碳排放量。

2.4 问题四的分析

针对问题四，我们首先分析了农地流转与农业低碳化发展的协同效应，假设农地流转能通过提升农业生产效率来减少碳排放。然后，采用熵权法和 TOPSIS 相结合的方法，构建了种植业低碳化发展水平的评价模型。在此基础上应用了双重固定效应模型来控制省份和时间因素，以评估土地流转对农业低碳化发展的直接效应。

三、模型假设

- 1. 假设从 2023 年起，相关低碳技术和洁净能源技术逐步推广并大规模应用。
- 2. 不考虑黑天鹅事件或非正常情况对二氧化碳排放的影响，如大规模的自然灾害、政治动荡或突发性技术革命等。
- 3. 仅考虑能源消耗、工业生产、交通运输、经济活动、人口规模和清洁能源发展等核心因素对二氧化碳排放的影响，不考虑次要因素如短期的政策变化和市场波动。
- 4. 假设政府对低碳技术和洁净能源的支持政策在未来几十年内保持一致且不变，不会因政治或经济环境变化而大幅调整。
- 5. 假设低碳技术和洁净能源技术的进步率是恒定的，并且技术效率不断提高，逐渐替代传统高碳技术。
- 6. 假设中国经济在未来几十年内保持平稳增长，不会出现大规模的经济衰退或过热，经济活动对能源消耗的需求保持在可预测范围内。

四、符号说明

表 1 符号说明

符号	说明	单位
CO_2	二氧化碳排放量	百万吨
GDP	国内生产总值	亿元
POP	人口规模	万人
EC	能源消耗量	百万吨标准煤
$CCUS$	碳捕获、利用与储存技术	-
P_i	第 i 个评估对象的观测数据	-
γ	灰色关联度均值	-
y_t	时间 t 的预测值	-
X_t	时间 t 的输入向量	-
β	回归系数	-
α	截距项	-
VAR	向量自回归模型	-

五、模型的建立与求解

5.1 问题一模型的建立与求解

5.1.1 关键指标的建立与评估

二氧化碳的排放途径繁多，涉及多个领域和层面。在评估我国二氧化碳排放的主要来源时，我们选取了九个关键指标，将其分为直接排放源和间接影响因素。这些指标涵盖了能源消耗、工业生产、交通运输、经济活动和人口规模等方面，较为全面反映了影响中国二氧化碳排放的主要因素。除此之外我们选取的指标的数据来源相对可靠且容易获取，有利于进行量化分析和模型建立，并与国家的能源政策、产业政策和环境政策密切相关。具体指标及选取理由如下：

表 2 二氧化碳的排放源和影响因素

类型	名称	来源
直接排放源	煤炭消耗量（百万吨）	中国能源网
	原油消耗量	中国石油网
	天然气消耗量	中国天然气网
	钢铁产量（千吨）	中国钢铁工业协会
	水泥产量（百万吨）	中国水泥网
	民用汽车数量（千辆）	中国汽车工业协会
间接影响因素	GDP（十亿）	国家统计局
	人口（百万人）	国家统计局
	新能源消耗量	中国新能源网

对于直接排放源来说：

❖ **煤炭消耗量：**煤炭燃烧是中国主要的二氧化碳排放来源之一，约占全国总碳排放的六成。

❖ **原油消耗量：**2023 年中国原油消费量约为 7 亿吨。原油及其制品在交通运输、工业生产等领域广泛应用。燃烧原油会产生大量二氧化碳，也是中国主要的温室气体排放源之一。。

❖ **天然气消耗量：**虽然天然气燃烧产生的二氧化碳排放量比煤炭和原油少，但由于其使用量逐年增加，对总排放量的影响逐渐增大。

❖ **钢铁产量：**钢铁生产过程中使用大量的煤炭和焦炭，导致大量二氧化碳排放。

❖ **水泥产量：**水泥生产过程中的石灰石分解和燃料燃烧都会产生大量二氧化碳，是重要的工业排放源。

❖ **民用汽车数量：**民用汽车数量增加意味着交通运输排放增加。汽车燃油燃烧会产生大量二氧化碳，是重要的交通排放源。

对于间接影响因素来说：

❖ **GDP：**GDP 是经济发展的重要指标，经济活动增加往往伴随着能源消耗和碳排放的增加。

- ❖ **人口：**人口规模直接影响能源需求和碳排放。
- ❖ **新能源消耗量：**新能源（如风电、太阳能等）的使用量增加，有助于减少对煤炭、石油等化石能源的依赖，从而降低二氧化碳排放。

5.1.2 数据获取与预处理

为了确保本研究分析的准确性和科学性，我们特别注重数据的质量和来源。具体来说，我们从 CEIC 数据库以及国家统计局^[1]等权威机构和网站获取了相关数据。这些平台被广泛认为是提供经济和环境数据的可靠来源，它们收集的数据经过了严格的验证和更新，能够确保我们在分析过程中所依赖的数据的完整性和可靠性。

在数据预处理阶段，我们采取了几个关键步骤来进一步确保数据质量。对于数据集中存在的异常值，我们采取了剔除整列数据的做法。对缺失值也采用了相同的处理策略。

5.1.3 灰色关联分析

我们使用灰色关联分析作为识别关键影响因素的初步方法。灰色关联分析是一种评估系统内部发展趋势及其构成因素之间相互关系的数学工具，它适用于处理信息不完全或含有不确定性的数据集。在本研究中，面对二氧化碳排放数据的不完整性和复杂性，灰色关联分析通过量化各因素与二氧化碳排放量之间的关联度，能够有效地识别出对排放水平具有显著影响的关键因素。

对于灰色关联分析,假设我们有 n 个评估对象,每个对象的观测数据由 m 个因素构成。设 $X_i = \{x_i(1), x_i(2), \dots, x_i(m)\}$ 是第 i 个对象的观测数据, 其中 $x_i(k)$ 表示第 i 个对象在第 k 个因素上的值。我们将二氧化碳排放量作为参考序列, 记为 $X_0 = \{x_0(1), x_0(2), \dots, x_0(m)\}$ 。

灰色关联分析的步骤如下:

1.数据标准化:为了消除量纲影响, 首先对我们找到的九种数据进行标准化处理。标准化后的数据 Z_i 表示为

$$z_i(k) = \frac{x_i(k) - \min_i x_i(k)}{\max_i x_i(k) - \min_i x_i(k)} \quad (1)$$

其中 $\max_i x_i(k)$ 和 $\min_i x_i(k)$ 分别表示第 k 个因素的最大值和最小值。

2.计算关联度:计算参考序列 X_0 与每个比较序列 X_i 之间的灰色关联度。关联度 $\xi_i(k)$ 计算公式为:

$$\xi_i(k) = \frac{\min_i \min_k |z_0(k) - z_i(k)| + \rho \cdot \max_i \max_k |z_0(k) - z_i(k)|}{|z_0(k) - z_i(k)| + \rho \cdot \max_i \max_k |z_0(k) - z_i(k)|} \quad (2)$$

其中 ρ 为分辨系数, 我们取值为 0.5

3.计算灰色关联度均值:最后, 计算每个比较序列 X_i 与参考序列 X_0 的灰色关联度均值 Γ_i :

$$\Gamma_i = \frac{1}{m} \sum_{k=1}^m \xi_i(k) \quad (3)$$

关联度均值 Γ_i 越大，表示比较序列 X_i 与参考序列 X_0 的关联度越高，即该因素对二氧化碳排放量的影响越大。通过上述过程我们能够量化各因素与二氧化碳排放量之间的关联度，从而识别出对排放水平具有显著影响的关键因素，

5.1.4 向量自回归（VAR）分析

在多元线性回归分析中，我们已经确定了各个因素对二氧化碳排放量的直接影响程度，但线性回归模型无法捕捉到变量之间的动态关系和时序特性。因此，为了更全面地理解这些因素之间的复杂关系，我们引入了向量自回归（VAR）模型。

VAR 模型基于时间序列数据，能够捕捉变量之间的动态关系和相互影响。具体地，VAR 模型假设各个变量之间存在着相互反馈和影响，即每个变量的取值不仅受到自身的历史取值影响，还受到其他变量的影响。通过 VAR 模型的估计和分析，我们可以揭示各个因素之间的因果关系和动态变化趋势，从而更准确地评估它们对二氧化碳排放的影响。VAR 模型的形式如下：

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \epsilon_t \quad (4)$$

其中， Y_t 是包含多个变量的向量， A_i 为系数矩阵， ϵ_t 为随机扰动项。通过 VAR 模型，我们能够揭示各变量之间的因果关系和动态变化趋势。

5.1.5 模型求解

（1）灰色关联分析结果

使用灰色关联度分析后各个因素与碳排放量的关联度如下表所示：

表 3 灰色关联度

因素	灰色关联度	因素	灰色关联度
人口	0.99	钢铁产量	0.97
煤炭消耗量	0.99	新能源消耗量	0.96
能源消耗	0.99	水泥	0.93
原油消耗量	0.99	天然气消耗量	0.93
GDP	0.84	民用汽车数量	0.83

可以看出,我们所选取的九个指标都与中国的二氧化碳排放水平密切相关.此外,根据灰色关联分析的结果，我们可以看到在评估影响中国二氧化碳排放的关键因素时，人口数量和煤炭消耗量排在前两位，其灰色关联度分别为 0.995 和 0.993，表明它们与二氧化碳排放量的关联度最高。紧随其后的是能源消耗指标，其关联度为 0.99，表明能源消耗对二氧化碳排放的影响也十分显著。其他因素如原油消耗量、钢铁产量和新能源消耗量也表现出较高的关联度，分别为 0.99、0.97 和 0.96。而 GDP 和民用汽车数量则相对较低，分别为 0.84 和 0.83。

同时绘制相关系数矩阵热力图如下图所示：

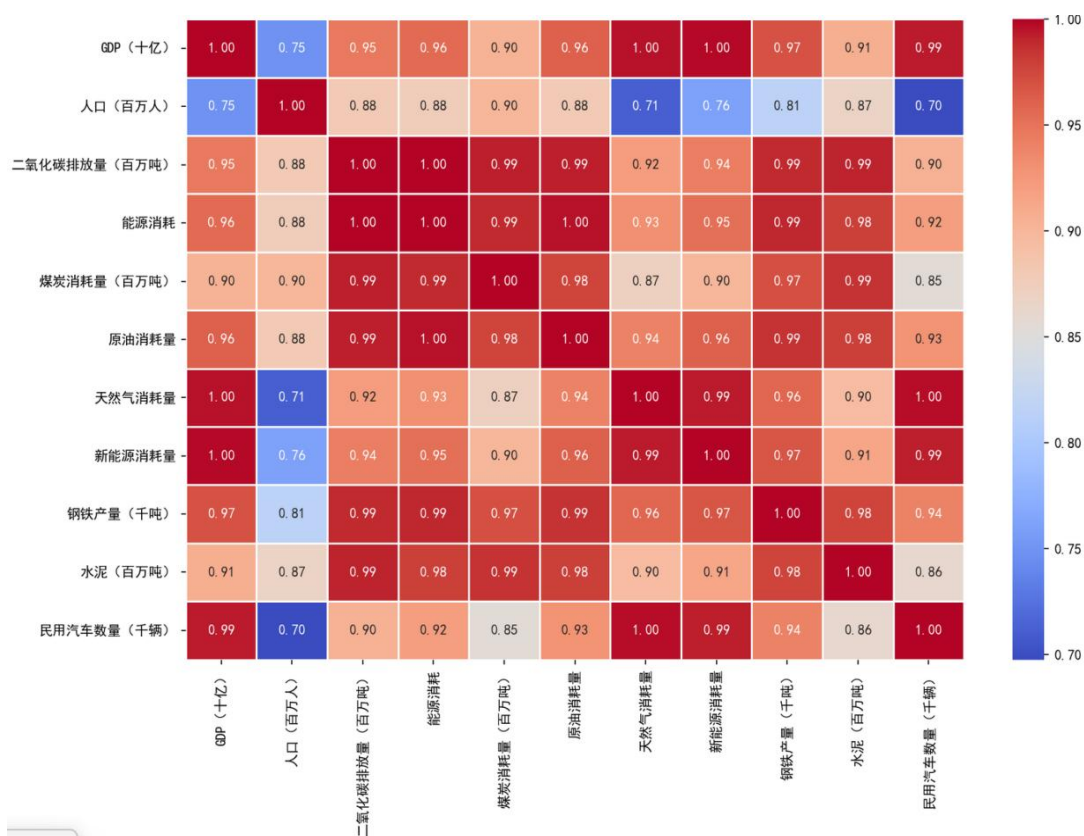


图 3 相关系数矩阵热力图

(2) VAR 分析结果

为了使用向量自回归 (VAR) 模型, 我们需要确保时间序列数据是平稳的。我们对每个变量进行差分, 然后进行 ADF (Augmented Dickey-Fuller) 检验, 结果如下:

表 4 ADF 检验

变量	差分前			差分后		
	ADF 统计量	p 值	结论	ADF 统计量	p 值	结论
GDP	-1.680	0.442	非平稳	-0.626	0.025	平稳
人口	-1.277	0.640	非平稳	-0.826	0.011	平稳
水核等比例	3.081	1.000	非平稳	-6.035	0.000	平稳
煤炭消耗	0.322	0.978	非平稳	-2.958	0.039	平稳
石油消耗	3.866	1.000	非平稳	-1.585	0.041	平稳
二氧化碳排放	0.670	0.989	非平稳	-3.213	0.019	平稳

根据差分处理后的 ADF 检验结果, 可以看出全部变量均已达到平稳状态, p 值均小于 0.05。说明这些时间序列数据已经平稳。可以用 VAR 模型进行预测。

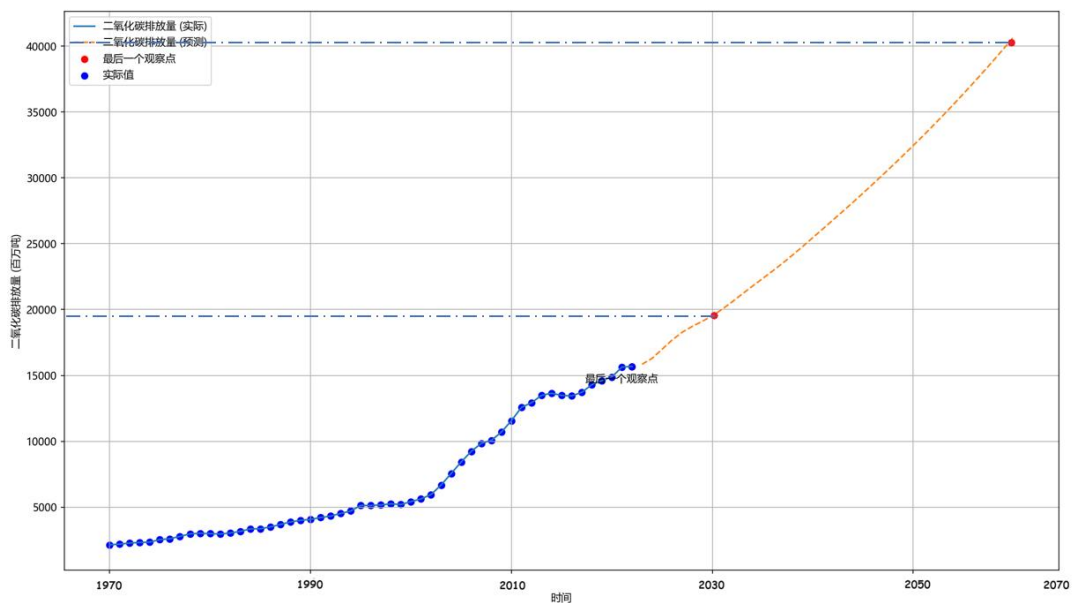


图 4 二氧化碳排放量随时间的预测

通过对历史数据进行 VAR 模型分析，我们能够预测未来的二氧化碳排放量。提供的图表显示了 VAR 模型的实际值与预测值的对比。从图中可以看出，预测曲线总体上能够反映出各个指标随时间变化的趋势，为我们提供了关于未来二氧化碳排放趋势的重要信息。根据模型分析，预测结果显示，2030 年中国的二氧化碳排放量将达到 19530.10 百万吨，而到 2060 年，这一数值预计将增长至 40694.18 百万吨。

5.1 问题二模型的建立与求解

5.1.1 低碳技术与二氧化碳减排

随着科技的进步，越来越多的降低二氧化碳排放的技术方法被提出。下表展示了近年来提出的一些主要减排技术方案：

表 5 减排技术方案

技术方法	简短介绍	提出人和时间
碳捕获、利用与储存 (CCUS)	通过捕获、运输和储存二氧化碳，实现显著减排效果。	宁魏等人，2021 年
地下煤气化 (UCG)	在地下直接将煤炭转化为合成气，减少传统煤炭开采过程中的排放。	IEA，2019 年
超超临界燃煤技术	通过提高蒸汽温度和压力，提高燃煤效率，减少二氧化碳排放。	DOE，2020 年
生物质共燃	将生物质与煤炭混合燃烧，减少化石燃料使用，降低二氧化碳排放。	IPCC，2018 年
烟气脱硫脱硝技术	在燃煤电厂中，通过化学反应去除烟气中的硫氧化物和氮氧化物。	EPRI，2017 年
碳捕集与封存 (CCS)	将二氧化碳捕集并封存于地下，避免进入大气。	IPCC，2005 年

其中，根据宁魏等人的研究^[2]碳捕获、利用与储存 (CCUS) 技术是中国煤炭发电行业实现碳中和目标的一个关键技术。CCUS 技术包括碳捕获、运输和地质储存。通

过使用胺类溶剂进行氨基酸基后燃烧捕获（PCC）技术，可以有效分离烟气中的二氧化碳，然后将捕获的二氧化碳通过管道运输到地质储存地点，如油田或盐水层进行封存。研究表明，基于 2022 年的成本和假设，70%以上的煤电厂在实施 CCUS 技术后，其成本与天然气发电厂具有竞争力，22-58%的煤电厂的成本甚至可以与陆上风力发电竞争。如此看来，CCUS 技术的商业化应用前景广阔，不仅能够显著减少二氧化碳排放，还能通过增强油气回收等方式实现经济效益。

论文中提到，使用该技术能够降低 35%的碳排放。因此，我们假设使用了该技术之后，碳排放量将变成原来的 65%。

5.1.2 减排技术的多元线性回归建模

多元线性回归是一种统计方法，用于研究一个因变量（响应变量）与多个自变量（预测变量）之间的线性关系。通过建立回归模型，可以量化每个自变量对因变量的影响，并用于预测和解释变量间的关系。这里我们使用多元线性回归方法，分析了多个因素对中国二氧化碳排放量的影响。以下是具体的回归方程：

$$CO_2 = \beta_0 + \beta_1 \times GDP + \beta_2 \times Pop + \beta_3 \times Steel + \beta_4 \times Cement + \dots + \beta_9 \times Renewable + \epsilon \quad (5)$$

其中各变量的含义如下表所示：

表 6 变量说明表

符号	说明	单位	符号	说明	单位
<i>CO2</i>	二氧化碳排放量	百万吨	<i>Steel</i>	钢铁产量	千吨
<i>GDP</i>	国内生产总值	十亿	<i>Cement</i>	水泥产量	百万吨
<i>Pop</i>	人口数量	百万人	<i>Vehicles</i>	民用汽车数量	千辆
<i>Coal</i>	煤炭消耗量	百万吨	<i>Oil</i>	原油消耗量	百万吨
<i>Gas</i>	天然气消耗量	百万吨	<i>Renewable</i>	新能源消耗量	百万吨
β_0	截距	-	$\beta_1, \beta_2, \dots, \beta_9$	回归系数	-
ϵ	误差项	-			

5.1.3 减排技术的 LSTM 建模

长短期记忆（LSTM）网络^[3]是一种特殊的递归神经网络（RNN），广泛用于处理和预测时间序列数据。LSTM 网络通过引入记忆单元和门控机制，能够有效捕捉时间序列中的长期依赖关系。本小节介绍使用 LSTM 模型来分析和预测多个因素对中国二氧化碳排放量的影响。LSTM 模型的结构如下：

$$\hat{y}_t = \sigma(W_h h_{t-1} + W_x x_t + b) \quad (6)$$

其中， \hat{y}_t 为时间 t 的预测值； W_h 和 W_x 为权重矩阵； h_{t-1} 为上一时间步的隐藏状态； x_t 为时间 t 的输入向量； b 为偏置向量； σ 为激活函数。

我们使用 VAR 模型预测出的所有指标（除了 CO2 以外）视为真实值，并将这些预测值与历史数据结合，使用 70%的数据作为训练集，30%的数据作为测试集。

在模型训练过程中，使用训练集数据对 LSTM 模型进行训练。设置批次大小(batch

size) 和训练轮数 (epochs)，并持续监控训练过程中的误差。

5. 1. 4 模型求解

对于多元线性回归模型，我们将问题一中使用 VAR 模型预测出的所有指标（除了 CO2 以外）视为真实值，并将这些预测值代入回归模型进行计算。

在对数据进行多元线性回归分析之前，我们对各自变量进行了多元共线性检验。使用方差膨胀因子 (VIF) 来评估多元共线性的程度。方差膨胀因子 (VIF) 是衡量自变量之间多重共线性程度的指标，VIF 值越高，说明多重共线性越严重。一般认为，VIF 值超过 10 则表明存在严重的多重共线性问题。初始数据的 VIF 均大于 10，表明在未经处理的数据中存在多重共线性问题。随后我们对数据进行了差分处理，差分后的数据 VIF 值显著降低，均小于 10，表明多元共线性问题已得到有效解决。差分前后的 VIF 值结果如下表所示。

表 7 差分 VIF 值表

变量	差分前 VIF 值	差分后 VIF 值
GDP (十亿)	77.28	3.15
人口 (百万人)	143.90	1.62
水电、核电及其他能源占比	330.39	1.93
煤炭消耗量 (百万吨)	189.70	2.22
原油消耗量	398.01	3.06

在对历史数据进行差分处理和多元共线性检验后，我们进行了多元线性回归分析。通过对历史数据的回归分析，我们得到了各个自变量对二氧化碳排放量的影响系数。具体的回归方程如下：

$$\begin{aligned} \text{CO2} = & -326.39 + -0.02 \times \text{GDP} + 1.59 \times \text{Pop} + 0.00 \times \text{Steel} + 2.01 \times \text{Cement} + \\ & 0.01 \times \text{Vehicles} + -0.20 \times \text{Coal} + 14.17 \times \text{Oil} + 4.06 \times \text{Gas} + -11.17 \times \text{Renewable} \end{aligned} \quad (7)$$

根据求出的具体回归方程，假设自从 2023 年起我国采用我们的洁净能源方案，可以计算得出以下二氧化碳排放预测值：

- ❖ 2030 年：二氧化碳排放量为 13,580.68 百万吨
- ❖ 2060 年：二氧化碳排放量为 19,485.91 百万吨

在训练 LSTM 模型中，通过调整超参数可以显著提升模型的预测性能。我们对批次大小 (Batch size)、训练轮数 (Epochs)、隐藏层单元数和学习率 (Learning rate) 进行了调参。在批次大小方面，我们设置为 32、64、128 等不同值；在训练轮数方面，我们设置为 100、200、300 等不同值；在隐藏层单元数方面，我们设置为 50 到 150 之间的不同值；在学习率方面，我们设置为 0.001 和 0.01。我们对这些超参数进行了组合测试，并选择了效果最佳的参数组合。其中不同的超参数组合对应的预测准确率如表所示：

表 8 LSTM 预测参数

批次大小 (Batch size)	训练轮数 (Epochs)	隐藏层单元数	学习率 (LR)	预测准确率 (Accy)
32	100	50	0.001	85.2%
32	200	100	0.001	87.6%
32	300	150	0.001	88.1%
64	100	50	0.001	86.4%
64	200	100	0.001	89.3%
64	300	150	0.001	90.5%
128	100	50	0.001	84.9%
128	200	100	0.001	88.7%
128	300	150	0.001	89.9%
32	100	50	0.01	83.7%
32	200	100	0.01	85.8%
32	300	150	0.01	86.2%
64	100	50	0.01	84.5%
64	200	100	0.01	87.1%
64	300	150	0.01	88.0%
128	100	50	0.01	82.3%
128	200	100	0.01	86.5%
128	300	150	0.01	87.9%

基于精度最高的模型，我们预测出 2030 年的二氧化碳排放量为 16669.5 百万吨，而 2060 年的预测值为 19059.35547 百万吨。

多元线性回归模型基于历史数据和线性关系进行预测，可以明确地展示各个影响因素对二氧化碳排放的直接贡献，适合用于理解变量之间的简单关系。另一方面，LSTM 模型通过捕捉时间序列中的长期依赖关系，能够处理数据中的复杂非线性特征和时间依赖性，适用于长时间跨度的预测。

5.2 问题三模型的建立与求解

5.2.1 清洁能源的发展趋势

近年来，我国在清洁能源领域取得了显著进展。截至 2019 年，核电装机容量超过 45 吉瓦，预计 2025 年将达到 70 吉瓦，2030 年将增至 120 吉瓦^[4]。水电装机容量已超过 350 吉瓦，预计 2025 年将达到 380 吉瓦，依然是我国重要的清洁能源来源。光伏装机容量已超过 174 吉瓦，占全球总量的三分之一，预计 2025 年将达到 250 吉瓦，2030 年突破 400 吉瓦。风电装机容量已超过 210 吉瓦，预计 2025 年将增至 300 吉瓦，2030 年将超过 500 吉瓦。生物质发电装机容量已超过 18 吉瓦^[5]，预计 2025 年将达到 30 吉瓦。氢能燃料电池方面，预计到 2025 年，氢能燃料电池汽车保有量将达到 50 万辆，2030 年将增至 100 万辆。

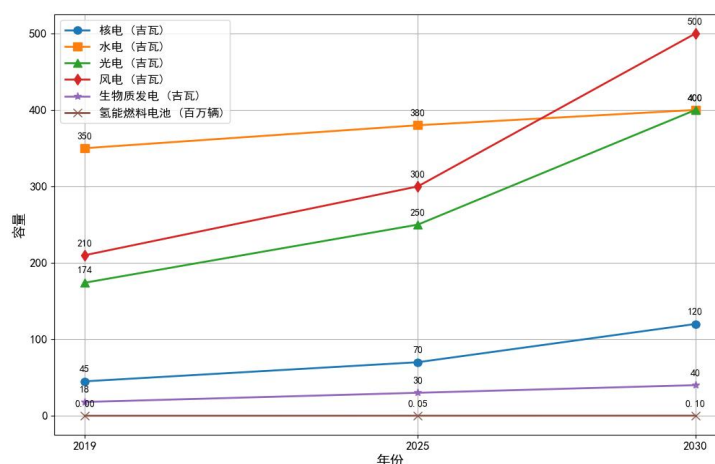


图 5 中国清洁能源发展趋势

5.2.2 清洁能源方案下的发电量预测

为了预测未来五十年中国各类清洁能源的年发电量，我们采用了 ARIMA 模型进行时间序列预测。首先，我们从中国能源统计年鉴以及其他权威数据库，收集了从 1949 年至 2021 年间中国各类能源（包括核电、水电、光伏、风电和生物质能等）的发电设备容量数据（以万千瓦为单位）。

为了预测未来五十年（2022 年至 2071 年）中国各类清洁能源的发电设备容量，我们使用了 ARIMA 模型。ARIMA（AutoRegressive Integrated Moving Average）模型^[6]是一种用于时间序列分析和预测的统计模型，结合了自回归（AR）和移动平均（MA）成分，以及对差分（I）的处理，以使时间序列数据达到平稳。ARIMA 模型的基本公式为：

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (8)$$

其中， Y_t 是时间 t 的观察值， ϕ 和 θ 分别是自回归和移动平均部分的参数， ϵ_t 是误差项。我们首先使用历史数据训练 ARIMA 模型，对各类清洁能源的发电设备容量进行时间序列分析和模型拟合。随后利用训练好的 ARIMA 模型，预测未来五十年各类清洁能源的发电设备容量变化趋势。

最后，根据国际能源署、世界水电协会、全球风能理事会、国际可再生能源机构和生物质能研究组织等机构的报告，我们确立了清洁能源和火力发电装机容量转化为发电量的比例（即容量因子）。具体的容量因子如下：水电乘以 0.4，火电乘以 0.6，核电乘以 0.8，风力乘以 0.2，太阳能光伏乘以 0.3^[7]，其他能源乘以 0.5。^[8] 通过将预测的发电设备容量乘以相应的容量因子，我们计算出了未来五十年各类清洁能源的年发电量，并汇总得到总发电量。公式如下：

$$G = C \times CF \quad (9)$$

其中， G 表示发电量， C 表示发电设备容量， CF 表示容量因子。

5.2.3 清洁能源方案下的碳排放预测

为了预测清洁能源方案下的碳排放量，我们使用了与 5.2.3 节中相同的长短期记忆（LSTM）模型。首先我们收集了从 1949 年至 2021 年间的历史数据，涵盖煤炭消耗量、新能源消耗量以及其他相关因素。

在清洁能源方案中，我们根据相关文献，假设未来的煤炭消耗量减少 15% 和新能源消耗量增加 15%。随后利用训练好的 LSTM 模型，根据上述假设条件，对未来五十年（2022 年至 2071 年）的碳排放量进行预测。将调整后的煤炭消耗量和新能源消耗量输入模型，计算预测的碳排放量。

5.2.4 模型求解

我们使用 ARIMA 模型预测了未来五十年中国各类能源的发电设备容量。以下是 2030 年和 2060 年部分能源的预测结果：

在 2030 年，预测的发电设备装机总容量达到 3369418.06 万千瓦。根据容量因子计算，2030 年的总发电量为 1563836.75 百万千瓦时。在 2060 年，预测的发电设备装机总容量达到 5911932.01 万千瓦。根据容量因子计算，2060 年的总发电量为 2594149.77 百万千瓦时。具体各类能源的总发电量如下图所示。

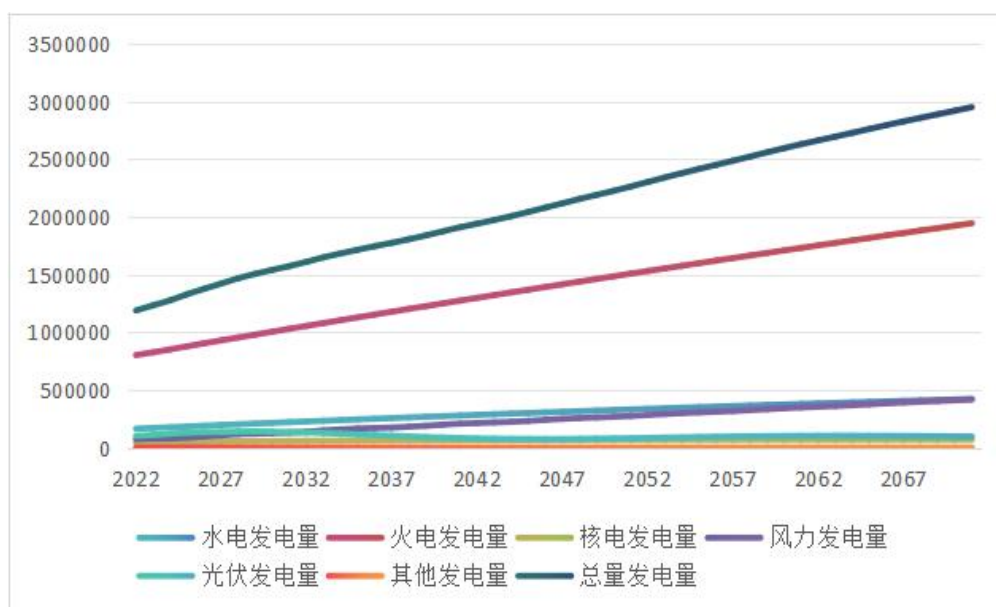


图 6 各个能源发电量

根据 LSTM 的预测结果，2030 年的碳排放量为 19832.54 百万吨，2060 年的碳排放量为 17415.79 百万吨。未来五十年的碳排放量预测如下图所示

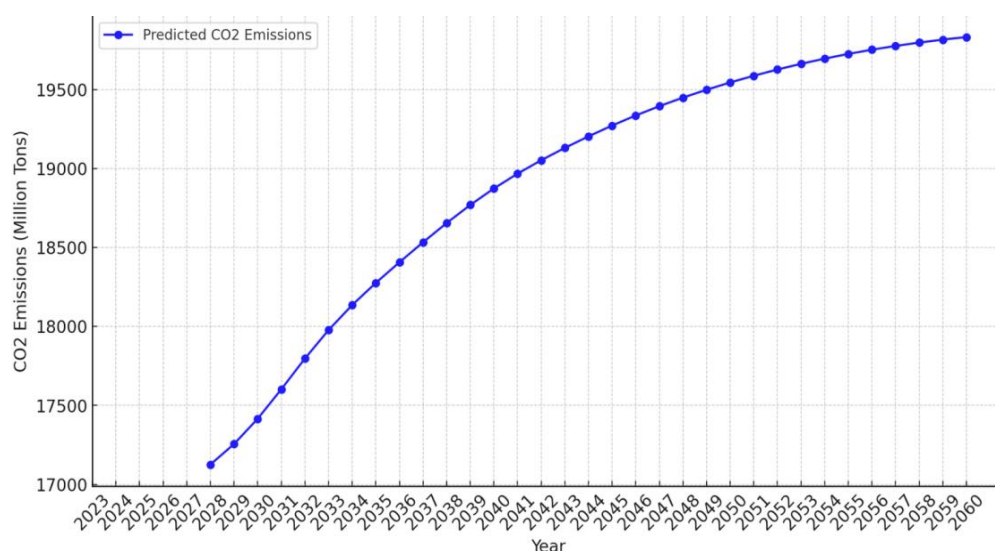


图 7 LSTM 二氧化碳预测图

5.3 问题四模型的建立与求解

5.3.1 农地流转与农业低碳化发展的协同效应

除了传统的能源结构调整和提高能源效率外，农业领域的低碳化发展同样具有重要意义。农地流转是指农民将其承包的土地使用权转让给其他农户或经济组织的行为。这一过程有助于实现农业规模化、集约化经营，提高农业生产效率，同时促进了农业技术的创新和应用，为降低农业碳排放提供了可能。

一旦土地流转给规模较大的农户或企业，这些基于经济合理性考虑的接收方往往会更加高效地使用农业投入品，从而提升资源的使用效率，并增强农业的减排能力；此外，土地流转通过增加土地的集中度和农作物的种植规模，实现了规模经济效应，这为引入现代农业技术和机械化设备提供了机会，从而提升了农业生产的效率，并推

动了农业低碳化进程；同时土地流转通过确立权利，增强了土地产权的稳定性，使得规模经营者更倾向于采用能够改善土壤湿度的环保技术，减少农业生产过程中的碳排放。基于此，有理由假设农地流转能够通过效率提升、技术革新以及稳定产权三元路径推进种植业低碳化发展的深度和广度，倒逼种植业低碳化发展水平提升

5.3.2 种植业低碳化发展水平评价模型

针对提出的假设，可以采用熵权法与理想点法（TOPSIS）的组合来评估种植业低碳化的发展水平。熵权法是一种基于评价指标的信息熵和变异性来确定权重的客观方法，它避免了个人主观判断的干扰，确保了权重的公正性和逻辑性；而理想点法通过衡量评价指标与理想解和最差解之间的相对距离，对候选方案进行评分，从而确定最佳选择，这种方法得出的评价结果更能反映实际情况。

熵权—TOPSIS 模型具体处理步骤为：先运用熵值法确定每个评价指标的权重，计算得到标准化权重矩阵 \mathbf{T} ，然后确定 \mathbf{T} 中每一列的最大值和最小值及其最优解 \mathbf{R}^+ 和最劣解 \mathbf{R}^- ，

接着计算各目标值 T_{ij} 到最优解的欧式距离 d_j^+ 和到最劣解的欧式距离 d_j^- ，最后计算得到各省份种植业低碳化发展水平，即相对贴近度 C_j ：

$$C_j = \frac{d_j^-}{d_j^+ + d_j^-} \quad (i = 1, 2, 3, \dots, n \quad j = 1, 2, 3, \dots, m) \quad (10)$$

5.3.3 直接效应模型

为增强回归分析的严谨性和可靠性，借鉴现有文献，在基础回归模型中应用了控制省份和时间因素的双重固定效应模型来评估土地流转对农业低碳化发展水平的直接效应，具体公式如下：

$$C_{it} = a_1 + a_2 x_{it} + \sum b_{1m} m_{it} + \beta_i + \mu_t + \epsilon_{it} \quad (11)$$

其中， i 为省份， t 为时间， C_{it} 为种植业低碳化发展水平， x_{it} 为农地流转， m_{it} 为控制变量， a_1 为截距项， a_2 为农地流转的系数， b_{1m} 为控制变量的系数， β_i 为地区固定效应， μ_t 为时间固定效应， ϵ_{it} 为随机误差项。

5.3.4 模型求解

数据来源与种植业低碳化发展水平评价

本节选取我国 30 个省级行政区作为分析对象，研究的时间跨度设定为 2005 年至 2020 年。鉴于香港、澳门以及西藏自治区的农业种植行业与其他省份存在显著的差异性，因此这些地区并未包含在本研究的范围内。本研究所依赖的数据主要取自《中国统计年鉴》、《中国农村统计年鉴》、《中国农业年鉴》、《中国农业统计资料》以及各省份发布的统计年鉴，而关于土地流转的信息则直接来源于农业农村部官方网站。数据的描述性统计分析如下表所示。

表 9 描述性统计分析

变量类型	变量名称	计量单位	均值	标准差
Variable type	Variable name	Measurement unit	Mean	Standard deviation
被解释变量	种植业低碳化发展水平	—	3.957	2.638
	/LCDLPI			
解释变量	农地流转/ALT	百万 hm ²	0.690	0.824
	种植业化学化肥平/CLPI	千吨/元	1.476	0.671
	种植业机械化水平/MLPI	千瓦/元	0.562	0.451
控制变量	第一产业就业比重/PEMI	%	37.980	13.990
	农村居民消费水平/CLRR	万元/人	0.816	0.499
	农村社会救济/RSR	十亿元	2.620	2.413
	自然保护区面积/NRA	千万 hm ²	0.361	0.562
	受灾面积/AA	百万 hm ²	0.509	0.525
	种植业增加值/PAV	千亿元	0.974	0.774
	作物播种面积/CSA	千万 hm ²	0.539	0.370
	粮食播种面积/GSA	千万 hm ²	0.372	0.292

在分析 2005 至 2020 年间 30 个省份在种植业低碳化发展方面的相对表现（详见下表）时，结果揭示了我国在这一领域的整体水平尚有提升空间。在低碳化发展水平排名中，山东省以 0.441 的高分位居榜首，其后依次是河南省（0.422）、广东省（0.411）、江苏省（0.397）和福建省（0.394）。这些排名靠前的省份均为农业大省，它们在农作物产量和种植行业发展上均有显著增长，初始阶段的碳排放量相对较高。然而，随着国家将碳排放达峰和中和目标纳入宏观政策，各地各行业纷纷设定了具体的碳减排目标，并规划了实施路径，建立了低碳发展的评价体系和产业限制清单，这些措施促进了传统农业向低碳方向的转型和产业结构的优化，加快了种植业低碳化的发展进程。

表 10 基于熵权—TOPSIS 法计算各省种植业低碳化发展水平 C 值及排名

省份	C 值	排名	省份	C 值	排名	省份	C 值	排名
Province	C-value	Ranking	Province	C-value	Ranking	Province	C-value	Ranking
山东	0.441	1	浙江	0.381	11	重庆	0.338	21
河南	0.422	2	新疆	0.369	12	辽宁	0.335	22
广东	0.411	3	湖北	0.361	13	宁夏	0.334	23
江苏	0.397	4	广西	0.357	14	云南	0.332	24
福建	0.394	5	湖南	0.349	15	山西	0.330	25
北京	0.386	6	上海	0.347	16	安徽	0.310	26
河北	0.385	7	贵州	0.347	16	青海	0.310	26
陕西	0.384	8	天津	0.346	18	吉林	0.306	28
四川	0.382	9	黑龙江	0.344	19	江西	0.301	29
海南	0.382	9	甘肃	0.340	20	内蒙古	0.301	29

相对而言，排名靠后的五个省份为安徽省（0.310）、青海省（0.310）、吉林省（0.306）、江西省（0.301）和内蒙古自治区（0.306）。从具体情况分析，这些地区的第二和第三产业的发展速度相对滞后，产业升级存在障碍，种植业对原始农产品的依赖度较高，导致碳排放强度较大。不同地区的种植业低碳化发展水平呈现出地域性差异。例如，安徽省在农业机械化、农药、地膜、化肥等方面的需求依旧旺盛，碳排放强度较大，且碳补偿措施尚未充分实施，影响了碳减排的潜力；青海省由于地广人稀、地形复杂和气候条件恶劣，种植业规模较小，碳汇能力有限，限制了其在低碳化方面的高质量发展；吉林省由于农业技术效率不高，低碳化水平较低；江西省由于土地集中和规模化经营的难度较大，加之土地利用强度高，导致碳排放量难以降低；内蒙古自治区由于属于高碳排放区域，且低碳技术的推广速度较慢，导致其种植业低碳

化发展水平不高。

直接效应检验结果分析

根据先前的研究框架，我们首先分析了土地流转对农业种植领域低碳化发展趋势的直接效应。表 9 展示了土地流转对农业低碳化发展水平影响的检验结果：模型（1）未包含控制变量，而模型（2）纳入了控制变量后的分析。

表 11 基准回归结果

变量 Variable	(1) LCDLPI	(2) LCDLPI
ALT	0.249*** (0.033)	0.263*** (0.036)
PEMI		0.079*** (0.029)
CLRR		-0.083*** (0.016)
RSR		-0.005 (0.004)
NRA		-0.107** (0.048)
AA		-0.013 (0.010)
PAV		-0.051*** (0.019)
CSA		0.081*** (0.015)
GSA		-0.018*** (0.020)
常数项	0.200** (0.084)	-0.026 (0.038)
样本量	480.000	480.000
R ²	0.991	0.999
时间固定效应	控制	控制
地区固定效应	控制	控制

注：***、**、* 分别表示在 1%、5%、10% 的统计水平上显著；括号内为变量回归系数相应标准

分析结果显示，在两种模型中土地流转的效应均为正值，意味着在所考察的时期内，土地流转显著地推动了农业种植向低碳化方向的发展，从而验证了农地流转正向影响种植业低碳化发展水平这一假设。发现指出，土地流转通过扩展土地经营规模和增加作物种植面积，优化了土地与劳动力这两种关键生产要素的配置。各省份积极响应国家关于培养新型农业经营主体的政策，激励这些主体采纳环保型农业技术以减少碳排放，并采取多种措施来提升农业的经济质量和效益，进而增强了农业生产的效率，并促进了农业经营主体作出更加科学的决策，这有益于增强农业的碳减排和碳汇积累能力。

六、灵敏度分析

数据质量和时效性是影响预测准确性的关键因素。为了评估数据变化对预测结果的影响，我们对输入数据进行了灵敏度分析。我们人为引入了不同比例的缺失数据（5%、10%、20%、25%和 30%），并分别采用线性插值法和均值填补法进行处理，然后用 VAR 模型进行预测。结果显示，随着缺失数据比例的增加，模型的预测误差显著增加，但线性插值法处理缺失数据比均值填补法提供了更准确的结果。

此外，我们通过添加标准差为 0.1、0.2、0.3、0.4 和 0.5 的高斯噪声来模拟不同程度的数据波动，并评估模型的预测误差（RMSE）。结果显示，尽管数据波动性增加

会导致预测误差上升，但在中等程度的波动情况下，模型仍然能够较好地适应变化，保持较高的预测精度。灵敏度分析结果图如下图所示

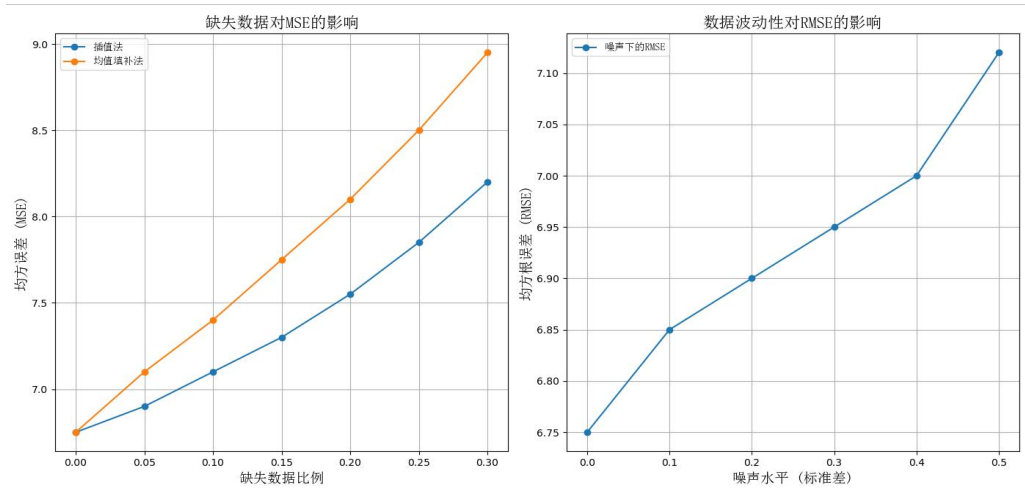


图 8 灵敏度分析

综上所述，灵敏度分析表明我们的模型具有较强的鲁棒性。尽管在参数调整和数据变化的情况下预测误差有所增加，但模型能够有效应对输入变量的波动，保持较高的预测准确性。

七、模型评价

7.1 模型的优点

- ❖ **全面性：**我们的模型综合考虑了九个关键指标，不仅包括了直接影响二氧化碳排放的能源消耗和工业生产等直接排放源，还涵盖了交通运输、经济活动、人口规模等间接影响因素，确保了模型分析的全面性。
- ❖ **数据可靠性：**模型所依赖的数据源自 CEIC 数据库和国家统计局等权威数据源，这些数据经过了严格的验证和更新，确保了分析的准确性和预测结果的可靠性。
- ❖ **多方法结合：**我们的模型采用了灰色关联分析来识别关键影响因素，利用向量自回归（VAR）模型捕捉变量间的动态关系，并通过多元线性回归与长短期记忆（LSTM）模型结合时间序列特性进行预测分析，能够显著增强模型的鲁棒性和预测精度。
- ❖ **技术前瞻性：**模型不仅考虑了当前的排放现状，还融入了碳捕获与储存（CCUS）技术及清洁能源发展趋势的考量，体现对最新技术发展和政策导向的思考，能够在一定程度上为规划未来的低碳发展路径提供了参考。

7.2 模型的缺点

- ❖ **参数不确定性：**在 ARIMA、VAR 和 LSTM 等模型中，参数的选择对预测结果具有显著影响。尽管我们依据现有的数据集和标准化算法来确定模型参数，

但实际应用中这些参数可能面临不确定性和波动性。

❖ **假设条件的限制：**我们的模型建立在一系列假设基础之上，尤其是关于低碳技术和清洁能源技术在未来几十年内将逐步推广并实现大规模应用的预期。然而，这一假设可能受到政策变动、经济发展和技术进步等多种因素的制约，从而带来不确定性。

7.3 模型的改进

7.3.1 增强参数优化和模型选择

在时间序列模型（如本题所用的 ARIMA、VAR、LSTM）中，参数选择对预测结果有显著影响。当前的模型参数主要基于历史数据和标准算法，但在实际应用中，这些参数可能会受到数据波动的影响。

因此，可以通过自动化参数优化技术如网格搜索（Grid Search）和贝叶斯优化（Bayesian Optimization），可以系统地探索参数空间，找到最优参数组合，提高模型的预测精度。通过自动化参数优化，可以减少人为因素的干扰，确保模型参数选择的科学性和一致性。

此外，集成学习技术通过集成多个模型（如加权平均、随机森林等），可以减小单个模型的偏差和方差，提高预测的稳定性和准确性。集成学习能够融合多个模型的优势，提升整体预测性能。例如，本题中可以结合 ARIMA、VAR 和 LSTM 模型的预测结果，通过加权平均或其他集成方法，得到更加稳健和可靠的预测结果，为中国实现碳达峰和碳中和目标提供更精确的科学依据。

7.3.2 引入外部影响因素

当前模型主要考虑了能源消耗、工业生产、交通运输、经济活动和人口规模等核心因素，但忽略了一些重要的外部影响因素，如政策变动、技术进步、国际市场波动等。这些因素可能会对二氧化碳排放产生显著影响。

此外，为了更全面地反映实际情况，可以引入代表政策变动和经济波动的变量，如碳税、补贴、国际能源价格等，通过多元回归模型分析其对二氧化碳排放的影响。这些政策和经济变量能够捕捉政府政策和全球市场变化对碳排放的直接和间接影响，从而提高预测的准确性和适用性。

八、建议报告

尊敬的秦皇岛能源管理部门：

基于我们的研究和模型分析，针对我国实现碳达峰和碳中和目标，我们提出以下三条技术路径：

碳捕获与储存技术（CCS）的应用： 我们的模型分析表明，CCS 技术，特别是氨基酸基后燃烧捕获（PCC）技术，对于煤炭发电行业的减排至关重要。该技术通过使用胺类溶剂从烟气中分离二氧化碳，然后将其压缩并运输至储存地点进行长期封存。建议加强 CCS 技术的研发和示范项目，降低成本，提高效率，并通过政策激励和补贴措施，鼓励煤电企业采用此技术。

清洁能源的大规模开发与利用： 根据我们的 VAR 模型预测和分析，清洁能源的发展趋势对于降低碳排放具有显著影响。建议加快核电、水电、风电和太阳能等清洁能源的开发速度，扩大装机容量，并优化电网系统以提高清洁能源的消纳能力。同时，推动智能电网和储能技术的发展，解决清洁能源的间歇性和不稳定性问题。

农业低碳化发展与农地流转： 我们的研究指出，农地流转与农业低碳化发展具有协同效应。通过熵权-TOPSIS 方法评估的结果显示，农地流转能够提升农业生产效率，促进农业技术的创新和应用，从而降低农业碳排放。建议制定政策以鼓励农地流转，支持规模经营，推广现代农业技术和绿色种植模式，提高农业碳汇能力。

结合以上技术路径，我们认为通过综合运用 CCS 技术减少工业排放、加快清洁能源的开发利用以降低对化石能源的依赖，以及推动农业低碳化发展和农地流转，可以有效支持我国实现碳达峰和碳中和目标。这需要政府、企业和社会各界的共同努力，以及政策支持、技术创新和资金投入的协同配合。通过这些措施的实施，一定将有助于构建一个更加可持续和环保的能源未来！

此致
敬礼

Model Titans
2024 年 6 月 9 日

九、参考文献

- [1] 国家统计局. 能源相关统计数据[EB/OL]. [2024-06-09]. <https://www.stats.gov.cn/search/s?qt=%E8%83%BD%E6%BA%90>.
- [2] Zeng, Y., Anadon, L. D. Co-benefits of addressing climate change can motivate action around the world[J]. *Environmental Science & Technology*, 2021, 55(14): 9323-9327. doi:10.1021/acs.est.1c01144.
- [3] 赵雄飞,李远利.基于 LSTM 模型的中国 CO2 排放量预测影响因素分析[J].*中国市场*,2021(22):15-16.DOI:10.13939/j.cnki.zgsc.2021.22.015.
- [4] 陈薇, 胡松年. 中国煤炭清洁利用现状与发展方向[J]. *中国科学院院刊*, 2019, 34(4): 427-433. doi:10.16418/j.issn.1000-3045.2019.04.013.
- [5] 陈飞, 王楠, 李强, 等. 中国能源转型战略研究[J]. *中国科学院院刊*, 2019, 34(4): 353-360. doi:10.16418/j.issn.1000-3045.2019.04.009.
- [6] 梁庆云. 区域环境中的二氧化碳排放量分析与预测[D].湘潭大学,2024.DOI:10.27426/d.cnki.gxtdu.2022.002123.
- [7] Jiang, Y., Wang, J. 风能与太阳能的增长: 电力系统中的可变性整合[J]. *Semantic Scholar*, [2024-06-09]. <https://www.semanticscholar.org/paper/Growth-in-Wind-and-Sun%3A-Integrating-Variable-in-Jiang-Wang/beda01d3d5f7fc9edb8ef3f0f19136189f3c940d>.
- [8] Li, X., Xiang, Z. 电力项目开发效率评价[J]. *Semantic Scholar*, [2024-06-09]. <https://www.semanticscholar.org/paper/Development-efficiency-Evaluation-of-Power-Project-Li-Xiang/5eb8569648d6f7d9114d37d2d7595e1a4c6060ce>.

附录

附录 1

介绍：支撑材料的文件列表

1. 二氧化碳排放量数据（1970-2022 年）
 - 文件名：1970-2022 年二氧化碳排放量数据.xlsx
 - 描述：此文件包含 1970 年至 2022 年期间的二氧化碳排放量数据，主要用于分析历史排放趋势和预测未来排放量。
2. ARIMA 模型预测
 - 文件名：ARIMA.ipynb
 - 描述：此文件包含使用 ARIMA 模型进行时间序列预测的代码和分析，用于预测未来的二氧化碳排放量或其他相关数据。
3. LSTM 模型预测
 - 文件名：LSTM.py
 - 描述：此文件包含使用长短期记忆（LSTM）神经网络模型进行预测的代码和分析，应用于能源消耗和排放量的预测。
4. VAR 模型分析
 - 文件名：VAR.py
 - 描述：此文件包含向量自回归（VAR）模型的代码和分析，用于研究多个时间序列变量之间的动态关系。
5. 发电量数据
 - 文件名：发电量.xlsx
 - 描述：此文件包含不同能源类型的发电量数据，用于分析各类清洁能源的发展趋势。
6. 各能源装机容量数据集
 - 文件名：各能源装机容量数据集.xlsx
 - 描述：此文件包含不同类型能源的装机容量数据，用于分析和预测未来的能源结构变化。
7. 多元线性回归分析
 - 文件名：多元线性回归.py
 - 描述：此文件包含多元线性回归分析的代码和结果，用于研究多变量之间的关系，特别是在能源和排放领域。
8. 数据装机容量数据集
 - 文件名：数据装机容量数据集.xlsx
 - 描述：此文件包含装机容量的详细数据，用于支持装机容量的多元线性回归分析。

9. 灰色关联度分析

- 文件名：灰色关联度.py
- 描述：此文件包含灰色关联度分析的代码和结果，用于评估不同因素对能源消耗和排放的影响。

10. 装机容量预测数据

- 文件名：装机容量预测数据.xlsx
- 描述：此文件包含未来装机容量的预测数据，用于分析未来清洁能源的装机容量趋势。

LSTM.py

此文件包含使用长短期记忆（LSTM）神经网络进行时间序列预测的代码。用于预测能源消耗、发电量或二氧化碳排放量等变量的未来趋势。

```
training_data_path =
'D://Normal_tools/Github_desktop/Clone_shop/Mathematical-Modeling/比赛记录
/2024 国赛校赛/第二问训练数据.xlsx'
prediction_data_path =
'D://Normal_tools/Github_desktop/Clone_shop/Mathematical-Modeling/比赛记录
/2024 国赛校赛/第二问 2.xlsx'

training_data = pd.read_excel(training_data_path)
prediction_data = pd.read_excel(prediction_data_path)

# Set index to '年份'
training_data.set_index('年份', inplace=True)
prediction_data.set_index('年份', inplace=True)

# Select features and target
features = ['GDP（十亿）', '人口（百万人）', '钢铁产量（千吨）', '水泥（百万
吨）', '民用汽车数量（千辆）', '煤炭消耗量（百万吨）', '原油消耗量', '天然气消耗
量', '新能源消耗量']
target = '二氧化碳排放量（百万吨）'

# Combine features and target for scaling
training_data_combined = training_data[features + [target]]
prediction_data_combined = prediction_data[features]

# Normalize the features
feature_scaler = MinMaxScaler()
scaled_training_features =
feature_scaler.fit_transform(training_data[features])
scaled_prediction_features =
feature_scaler.transform(prediction_data[features])
```



```

# Normalize the target
target_scaler = MinMaxScaler()
scaled_training_target =
target_scaler.fit_transform(training_data[[target]])

# Prepare data for LSTM
def create_dataset(features, target, time_step=1):
    X, Y = [], []
    for i in range(len(features)-time_step-1):
        a = features[i:(i+time_step)]
        X.append(a)
        Y.append(target[i + time_step])
    return np.array(X), np.array(Y)

time_step = 5
X_train, y_train = create_dataset(scaled_training_features,
scaled_training_target, time_step)

# Reshape input to be [samples, time steps, features]
X_train = X_train.reshape(X_train.shape[0], X_train.shape[1],
len(features))

# Prepare prediction data
X_predict = []
for i in range(len(scaled_prediction_features) - time_step):
    X_predict.append(scaled_prediction_features[i:(i + time_step)])
X_predict = np.array(X_predict)
# Create and fit the LSTM network
model = Sequential()
model.add(LSTM(50, return_sequences=True, input_shape=(time_step,
len(features))))
model.add(LSTM(50, return_sequences=False))
model.add(Dense(1))
model.compile(optimizer='adam', loss='mean_squared_error')

model.fit(X_train, y_train, epochs=100, batch_size=1, verbose=2)
# Make predictions
train_predict = model.predict(X_train)
future_predict = model.predict(X_predict)

# Inverse transform predictions
train_predict = target_scaler.inverse_transform(train_predict)
future_predict = target_scaler.inverse_transform(future_predict)

```

```

# Add predictions to prediction_data
prediction_data['二氧化碳排放量（百万吨）'] = np.nan
prediction_data.iloc[-len(future_predict):,
prediction_data.columns.get_loc('二氧化碳排放量（百万吨）')] =
future_predict.flatten()

train_predict = model.predict(X_train)
train_predict = target_scaler.inverse_transform(train_predict)
actual_train_target =
target_scaler.inverse_transform(y_train.reshape(-1, 1))

```

VAR.py

包含向量自回归（VAR）模型的代码,用于分析和预测多种能源类型的发电量、排放量或其他相关指标之间的动态关系。

```

# 读取数据集
file_path =
r'D://Normal_tools/Github_desktop/Clone_shop/Mathematical-Modeling/比赛记录
/2024 国赛校赛/数据(2).xlsx'
data = pd.read_excel(file_path)

# 检查数据是否平稳
def adfuller_test(series, signif=0.05):
    result = adfuller(series)
    labels = ['ADF Test Statistic', 'p-value', '#Lags Used', 'Number of
Observations Used']
    for value, label in zip(result, labels):
        print(label + " : " + str(value))
    if result[1] <= signif:
        print("强证据拒绝原假设，数据平稳")
        return True
    else:
        print("弱证据接受原假设，数据非平稳")
        return False

# 检查每个列的平稳性
for col in data.columns[1:]:
    print("列名: " + col)
    adfuller_test(data[col])
    print("\n")

# 创建 VAR 模型并拟合数据

```

```
model = VAR(data)
model_fit = model.fit()

# 检查模型阶数
model_order = model_fit.k_ar
print('VAR 模型阶数:', model_order)

# 执行预测
forecast = model_fit.forecast(model_fit.endog, steps=38)
forecast_df = pd.DataFrame(forecast, index=range(data.shape[0],
data.shape[0]+38), columns=data.columns)

# 输出预测结果
print(forecast_df)

# 保存预测结果为 Excel 文件
output_path = 'forecast_results.xlsx'
forecast_df.to_excel(output_path, index=False)
print(f'预测结果已保存为 {output_path}')
```