



中国研究生创新实践系列大赛
“华为杯”第二十届中国研究生
数学建模竞赛

学 校 华东师范大学

参赛队号 23102690008

队员姓名 1.杨蔚昭

2.黄心清

3.万雨晗

中国研究生创新实践系列大赛

中国光谷·“华为杯”第二十届中国研究生 数学建模竞赛

题 目

实现区域双碳的目标与路径规划研究

摘 要：

2020 年 9 月，习近平主席在第七十五届联合国大会一般性辩论上阐明，应对气候变化《巴黎协定》代表了全球绿色低碳转型的大方向。同时宣布，中国将进一步提升自主贡献力度，采取更加有力的政策和措施，旨在确保二氧化碳排放在 2030 年前达到峰值，并努力争取在 2060 年前实现碳中和目标。2021 年 9 月，中共中央国务院正式发布了《关于完整、准确、全面贯彻新发展理念，做好碳达峰和碳中和工作的意见》，明确了中国双碳行动的战略规划。实现 2060 年的碳中和目标就成为了一项紧迫任务。因此，如何进行路径规划从而实现区域双碳目标是本文重点关注的问题。

为实现这一目标，本文对碳排放核算模型进行了构建与优化。利用某区域经济、人口以及能源消耗等指标，通过多元线性回归建立了不同指标之间的关联模型，并分别尝试使用不同的方法建立区域碳排放量以及经济、人口、能源消费量的预测模型。最后在此基础上，分别设计三种不同情境下的区域双碳（碳达峰与碳中和）目标与路径规划方法从而确定双碳路径。

对于问题一：通过分析附表中的各项数据与其之间的联系，对缺失值进行插值填充后利用公式对单位 GDP 能耗等参数进行计算。之后基于 kaya 模型对参数进行筛选后建立一个综合的指标体系。根据体系结构，分析区域碳排放量以及在 GDP、人口和能源消费量这些一级指标下二级指标之间的关系，对该区域碳排放量产生影响的各因素及其贡献进行总结，从而深入了解十二五与十三五碳排放的现状、趋势和影响因素。利用多元线性回归方法建立区域碳排放量以及经济、人口、能源消费量各指标及其关联模型，研判该区域实现碳达峰与碳中和需要面对的主要挑战，为该区域双碳路径规划中差异化的路径选择提供依据；

对于问题二：通过对经济、人口、能源消费量、能源消费结构以及区域碳排放量等数据进行分析，确定模型输入的各项指标。利用逻辑斯蒂曲线、多项式拟合以及 ARIMA 模型进行预测，获得 2010-2060 年经济、人口、能源消费结构以及非化石能源占比等指标的完整的时间序列。对比分析了多元线性回归和机器学习（BP 神经网络以及 LSTM 模型）对于能源消费量预测模型和区域碳排放量预测模型的解释度。通过建立多因素影响下的时间序列预测模型，构建出基于人口和经济变化的能源消费量模型以及基于多因素的区域碳排放量预测模型，有效的模拟出了未来长时间范围内的能源消耗量以及碳排放量的变化情况。在建立基于多因素的碳排放量预测模型中，分析了其变化的内在动因，重点关注了提高非化石能源消费占比对于碳排放量的影响，为实现区域双碳目标提供理论依据。

对于问题三：分别设计无人干预的自然情景、按时碳达峰与碳中和的基准情景与率先碳达峰与碳中和的雄心情景，并与能效提升和非化石能源消费比重提升相关联。利用问题二中的模型确定在实现双碳目标的前提下，2025 年、2030 年、2035 年、2050 年和 2060 年 GDP、人口和能源消费量的目标值，根据历史数据与政策发布建立能够按时达到碳达峰与碳中和的能源利用效率和提高非化石能源消费比重的模型，并对二者在不同时间节点的目标值进行预测。最后，为实现双碳目标对

能效提升、产业（产品）升级、能源脱碳和能源消费电气化这些举措进行定性与定量分析。

关键词：碳排放；碳达峰与碳中和；多元线性回归；ARIMA 模型；LSTM 模型；路径规划

目录

1 问题重述	6
1.1 问题背景	6
1.2 问题提出	6
2 模型假设及关键性符号说明	8
2.1 模型假设	8
2.2 关键性符号说明	8
3 问题一的建模与求解	9
3.1 问题分析	9
3.2 数据探索	9
3.2.1 建立指标与指标体系	9
3.2.2 分析区域碳排放量以及经济、人口、能源消费量的现状	13
3.2.3 区域碳排放量以及经济、人口、能源消费量各指标及其关联模型	18
4 问题二的建模与求解	22
4.1 问题分析及建模思路	22
4.1.1 人口和 GDP 增长模型构建思路	22
4.1.2 能源消耗量模型构建思路	22
4.1.3 碳排放量模型构建思路	22
4.2 模型简介	23
4.2.1 Logistic 曲线	23
4.2.2 多项式拟合模型	23
4.2.3 差分自回归移动平均模型 (ARIMA)	24
4.2.4 多元线性回归分析	24
4.2.5 长短期记忆递归神经网络模型 (LSTM)	24
4.2.6 BP 神经网络模型	25
4.3 模型建立	25
4.3.1 人口增长模型	25
4.3.2 GDP 增长模型	26
4.3.3 基于人口和经济变化的能源消费量预测模型	28
4.3.4 基于多因素的区域碳排放量预测模型	31
4.4 本章小结	44
5 问题三的建模与求解	44
5.1 问题分析	44
5.2 模型建立与求解	45
5.2.1 情景设计模型构建	45
5.2.2 多情景下碳排放量核算方法	48
5.2.3 双碳目标与路径确定	53
5.3 结果定性定量分析	56
9 模型评价与推广	57
10 参考文献	57
11 附录	59
11.1 问题一代码	59
11.2 问题二代码	60

11.3 问题三代码	62
------------------	----

1 问题重述

1.1 问题背景

长期以来，中国一直积极参与全球治理，将温室气体减排任务融入国家五年规划和2035远景目标之中。在政策实施方面，通过一系列措施，如产业结构调整、能源结构优化、能源效率提高、碳市场建设以及增加生态碳汇等，旨在达到显著的节能减排成果。在国家“3060”碳达峰碳中和的政策背景下，如何寻求经济-能源-环境的平衡有效发展是国家、省份、城市等不同级别经济体仍需探索的重要课题。

为了更深入地应对气候危机，并在全球碳减排进程中发挥更大作用，中国在第七十五届联合国大会和气候雄心峰会上宣布了新的国家自主贡献目标和长期愿景。相较于2015年提交的国家自主减排贡献方案，中国提高了碳强度目标。实现中国的碳中和目标要求全国在一定时限内，将由人为活动直接和间接排放的二氧化碳（CO₂）通过碳捕捉与封存、植树造林等固碳技术吸收，实现CO₂的“零排放”。相较于发达国家的历史进程，中国实现碳中和目标面临时间紧迫、任务繁重的严峻挑战。中国需要在更短的时间内实施更大规模的碳中和，因此需要制定适合中国国情的碳中和路线。

因此，碳达峰的行动方案必须在碳中和目标的指导和限制下进行综合规划。然而，目前关于中国实现碳达峰和碳中和目标路径的研究相对不足，尤其是在当前世界面临百年未有之大变局的情况下，后疫情时代社会经济发展存在较大不确定性，新技术如数字化、智能化和网络化对人们的生活和工作方式产生深远影响。

因此，如何在全面考虑社会、经济和技术领域的不确定性的前提下，提出适用于中国新发展格局的碳中和行动方案，是当前亟待解答的问题。本文研究区位于中国东南沿海，地势平坦，水陆交通便利，人口密集，经济发达，科教资源丰富，但能源及生态碳汇资源相对匮乏。在此条件下，进行数据挖掘分析，建立一系列数学模型，以分析、评估和预测各因素对于碳排放的影响，最终达到区域碳达峰与碳中和的目标与路径规划研究。

实现区域双碳目标具有多重意义，展示了中国积极应对全球气候变化问题的责任担当，有助于推动全面的绿色转型，加速构建清洁、高效、绿色和安全的现代治理体系。

1.2 问题提出

基于上述研究背景，本文需研究和解决如下问题：

问题 1：

通过建立一个综合的指标体系，对区域的碳排放量以及与经济、人口和能源消费量之间的关系进行全面分析，我们可以更深入地了解该区域碳排放的现状、趋势以及影响因素。同时，总结各因素对该区域碳排放量的影响程度和贡献，以揭示碳排放的根本原因。

在建立区域碳排放量、经济、人口和能源消费量等指标之间的关联模型的基础上，我们能够全面评估该区域实现碳达峰和碳中和所面临的主要挑战。这些模型可以帮助我们量化各因素对碳排放的影响，进而为制定区域双碳路径规划提供有针对性的建议和决策支持。通过深入分析，从而指定更适合研究区的碳减排策略，为可持续发展提供坚实的基础，确保碳中和目标的实现。

问题 2：

通过对区域碳排放量、经济、人口、能源消费等数据进行仔细分析和拟合，建立合理可靠的预测模型。这些模型将基于人口和经济变化等因素，构建出与能源消费量以及区域碳排放量相关的预测模型。我们将利用 2010 年至 2020 年的历史数据进行模型的训练和验证，以确保模型的精确性。通过回归分析和预测，我们能够更准确地反映这些因素对区域碳排放量的影响，同时帮助我们理解人口和经济变化如何影响能源消费和碳排放情况。通过科学的数据分析和模型构建，有助于帮助我们制定有效的减排政策和可持续发展战略，为问题三的解决夯实基础，为未来碳排放的趋势和可能的影响因素提供了重要的参考，

问题 3:

建立建立区域碳达峰和碳中和的目标与路径规划方法模型是一项复杂而全面的任务，需要综合考虑多个因素，包括地理信息、人口数据、经济因素等多个方面的要素。我们将在前两个问题的基础上进一步深入，从机理分析的角度出发，致力于建立区域碳达峰和碳中和的目标与路径规划方法模型。

在这一任务中，我们将借助地理信息、人口数据、经济因素等多个数据源，以及相关的科学理论和方法，通过分析碳排放的机理和影响因素，提出针对性的目标设定和路径规划策略。这些策略将有助于指导区域碳减排工作，推动可持续发展，减缓气候变化的速度，提高生态环境质量，同时促进经济增长和社会福祉的提升。通过科学的建模和规划，我们能够更加有效地达成区域碳达峰和碳中和的双碳目标。

2 模型假设及关键性符号说明

2.1 模型假设

- 1、数据真实可靠
- 2、所有预测指标是基于实际情况进行分析添加；

2.2 关键性符号说明

文中使用到的符号含义如下：

符号	符号含义	符号	符号含义
CO_2	二氧化碳排放量	δCO_2	二氧化碳相对变化率
P	人口	δP	人口相对变化率
GDP/P	人均 GDP	$\delta GDP/P$	人均 GDP 相对变化率
E/GDP	单位 GDP 能耗	$\delta E/GDP$	单位 GDP 能耗相对变化率
CO_2/E	单位能耗二氧化碳排放量	$\delta CO_2/E$	单位能耗二氧化碳排放量相对变化率

3 问题一的建模与求解

3.1 问题分析

本题通过建立一个综合的指标体系来分析区域的碳排放量以及与经济、人口和能源消费量的关系，通过数据的深入挖掘，对该区域碳排放量产生影响的各因素及其贡献进行总结。关注该区域的碳排放量与经济状况和人口、能源消费等数据，建立一个综合的指标体系以及区域碳排放量与各指标的关联模型，有助于我们了解环境和可持续发展方面的挑战，从而更深入地了解碳排放的现状、趋势和影响因素，研判研究区实现碳达峰与碳中和需要面对的主要挑战，为该区域双碳路径规划中差异化的路径选择提供依据。

3.2 数据探索

3.2.1 建立指标与指标体系

通过对数据附表中八个表格的分析与总结，基于能够描述某区域经济、人口、能源消费量和碳排放量的状况，能够描述各部门（能源供应部门、工业消费部门、建筑消费部门、交通消费部门、居民生活消费、农林消费部门）的碳排放状况，能够描述各主要指标之间的相互关系，并能够使部分指标的变化（同比或环比）成为碳排放量预测的基础这四个条件建立了以下指标体系。

该指标体系包含四个一级指标、三十一个二级指标构成，如表 3.1 所示。由于在二级指标中存在附件中没有提供的数据，因此需要根据概念和原理对其进行计算。

在计算之前首先要进行数据预处理，经过观察分析，发现在表《能源消费部门碳排放因子》中 2012 年建筑消费部门对应的煤炭碳排放因子为空值、表《能源供应部门碳排放因子》中 2012 年热力发电的碳排放因子为空值，但在其余年份对应数据均有数值且较为稳定。因此，决定利用其他年份数据对该年份的这两处异常值进行三次样条插值处理。异常值替换后分别为 2.651 和 2.943。对于因子表中其余“-”符号，将其替换为“0.000”。

异常值处理之后可以对于提出的指标进行计算。首先计算能源供应部门碳排放量。利用能源消耗结构表与对应的能源供应部门碳排放因子表格，根据注释 1，可以明确能源供应部门碳排放量的计算方法：

能源供应部门碳排放量=能源消费品种结构*对应的能源供应部门碳排放因子；

因此经过表格重排以及计算后，能源供应部门碳排放总量如图 3.1。

表 3.1 指标与指标体系

一级指标	二级指标
经济	GDP 总量
	人均 GDP
	一产农林消费部门增加值
	二产能源供应部门增加值
	二产工业消费部门增加值
	三产交通消费部门增加值
	三产建筑消费部门增加值
	能源消耗总量
	单位 GDP 能耗（能源消耗强度）
	非化石能源消费比重
能源消耗量	能源消耗总量同比变化率
	一产农林消费部门消耗量
	二产能源供应部门消耗量
	二产工业消费部门消耗量
	三产交通消费部门消耗量
	三产建筑消费部门消耗量
	居民生活消费
	常住人口
碳排放	碳排放总量
	一产农林消费部门碳排放量
	二产能源供应部门碳排放量
	二产工业消费部门碳排放量
	三产交通消费部门碳排放量
	三产建筑消费部门碳排放量
	居民生活消费
	一产农林消费部门碳排放因子
	二产能源供应部门碳排放因子
	二产工业消费部门碳排放因子
	三产交通消费部门碳排放因子
	三产建筑消费部门碳排放因子
	居民生活消费碳排放因子

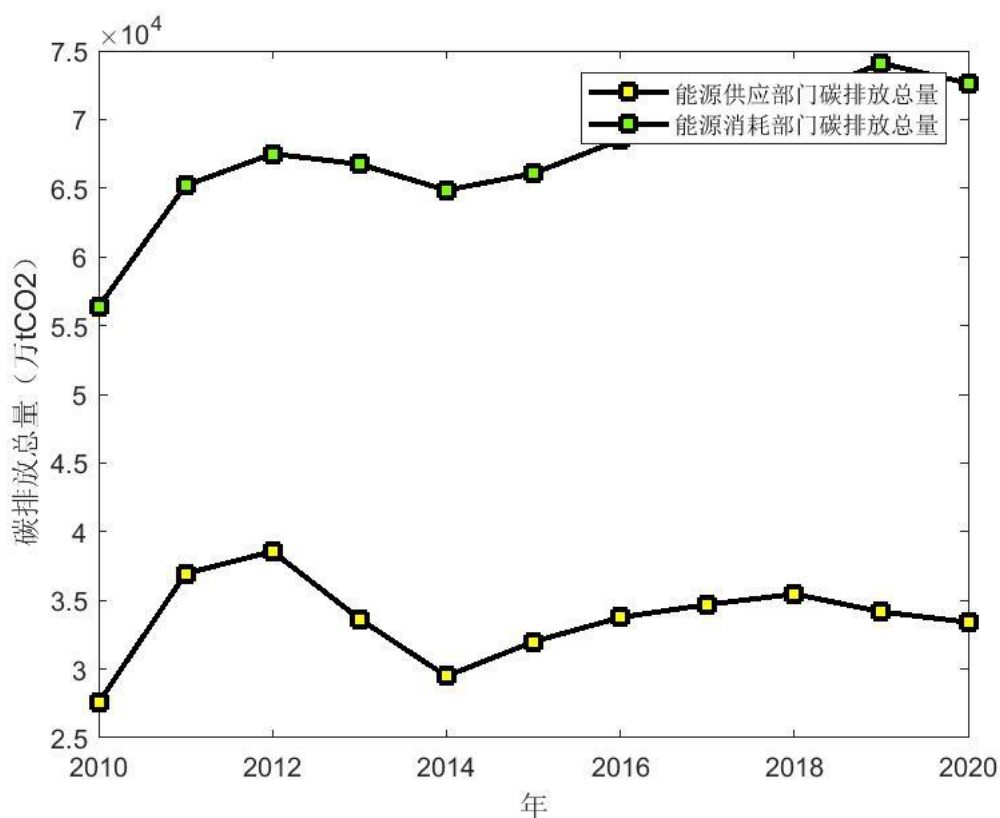


图3.1 能源供应部门与能源消耗部门碳排放总量结果对比

之后分别根据题目中的部分定义对人均 GDP、单位 GDP 能耗和人均能耗进行计算。其公式分别为：

$$\text{人均 GDP} = \text{GDP 总量} / \text{常驻人口} \quad (1)$$

单位 GDP 能耗 = (第一产业增加值权重*第一产业单位增加值能耗+第二产业增加值权重*第二产业单位增加值能耗+第三产业增加值权重*第三产业单位增加值能耗) / (1-生活能耗比重)

$$\text{单位 GDP 能耗} = \text{人均能耗} / \text{人均 GDP} \quad (2)$$

图 3.2 分别表现了 2010~2020 年间该三种指标的变化情况。

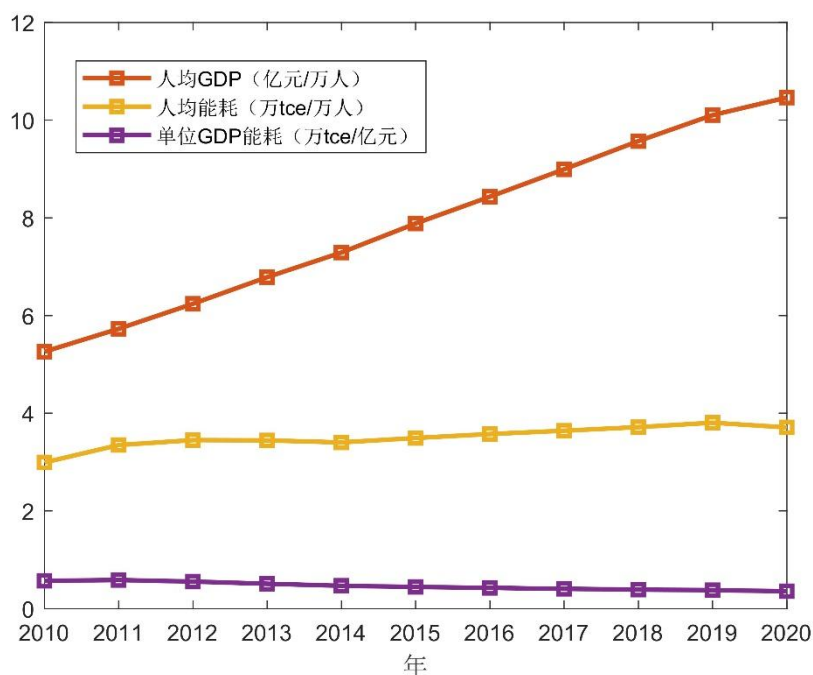


图3.2 2010-2020年人均GDP、单位GDP能耗和人均能耗变化情况

通过该图我们可以知道人均 GDP 在这 11 年里一直以较高的速度稳定上升，而单位 GDP 能耗在稳步下降，说明我国非化石能源的利用正在不断增加与优化。由于人均能耗与二者之间的关系，虽然单位 GDP 能耗在稳步下降，但由于收到快速增长的人均 GDP 的影像，人均能耗仍具有缓慢的上升趋势。

最后对非化石能源消费比重进行计算。无碳排放的能源被称为化石能源，因此通过分析不同能源的碳排放因子，可将热力、风力与其他能源视为非化石能源，煤炭、油品、天然气为非化石能源。根据以下公式获得如图 3.3 所示的非化石能源比重变化情况。

$$\text{非化石能源消费比重} = \text{非化石能源消费量} / \text{能源消费量} \quad (3)$$

图 3.3 表现了该区域非化石能源消费比重在十二五（2011-2015 年）和十三五

（2016-2020 年）期间的变化情况。可以看出在十三五期间非化石能源消费比重的增加速度相比于十三五存在明显的提升，并始终处于稳定的增加状态。说明我国在实施"十三五"计划期间，政府采取的包括提供财政和税收激励、鼓励投资和研发非化石能源技术以及加强可再生能源发展等的一系列政策措施对该区域产生了积极的影响，并产生了显著的效果。

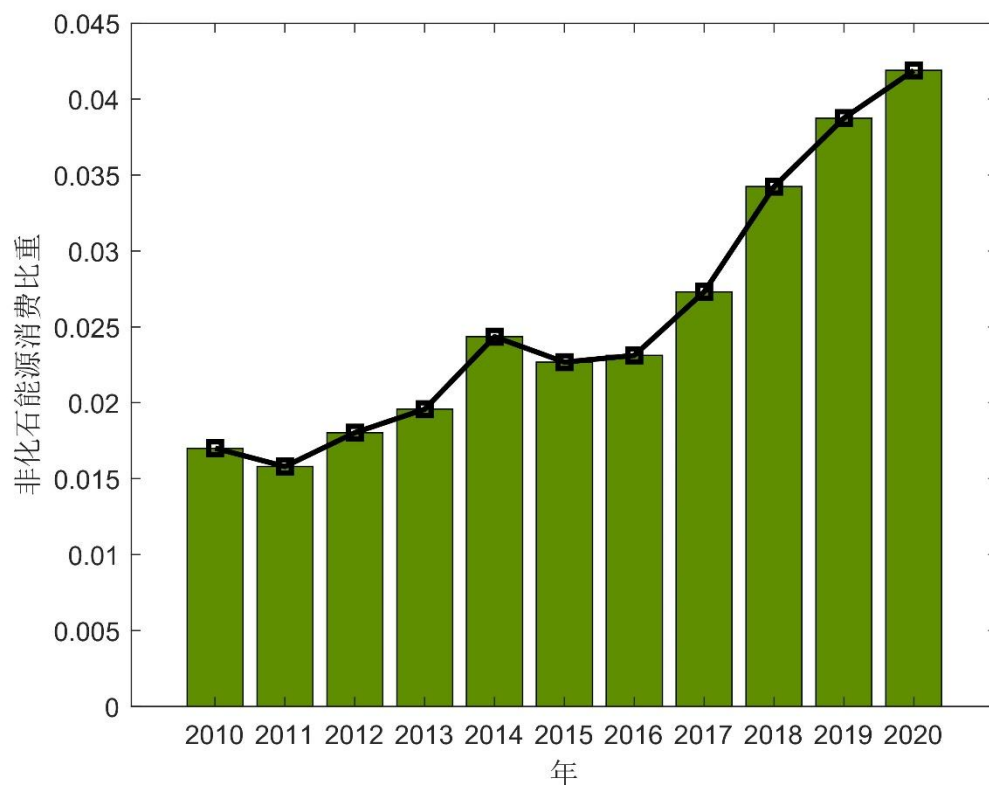


图 3.3 2010-2020 年非化石能源所占比重

3.2.2 分析区域碳排放量以及经济、人口、能源消费量的现状

区域碳排放量一直备受广泛关注，因为它直接塑造着气候变化的走势，深刻影响着环境质量。区域碳排放量与经济、人口和能源消费之间的错综复杂关系，构成了全球可持续发展所面临的一项巨大挑战。碳排放量与经济、人口和能源消费之间存在紧密纽带，它们相互交织，相互影响。随着经济的繁荣，人口的增长，以及能源消费的上升，碳排放量通常也会相应增加。

这种关联性需要深入研究，以找到减缓碳排放增长的方法，并推动可持续发展目标的实现。通过深入探讨这些关系，我们可以更好地理解碳排放问题的复杂性，并为制定有效的减排政策和可持续发展战略提供基础。在对研究区域的数据进行了详细分析和处理后，我们通过深入挖掘和分析数据，获得了关于区域碳排放量以及经济、人口和能源消费量的全面洞察，包括整体趋势和局部细节。这些深入的数据评估为我们建立后续模型提供了坚实的基础，有助于确保模型的精确性和可靠性。

通过对于研究区数据的精细评估，我们能够更好地理解区域碳排放问题的复杂性，发现潜在的关联性和趋势，为后续的研究工作提供了有力的支持。这些分析结果将为制定减排策略、规划碳中和路径以及可持续发展决策提供重要的参考和指导。

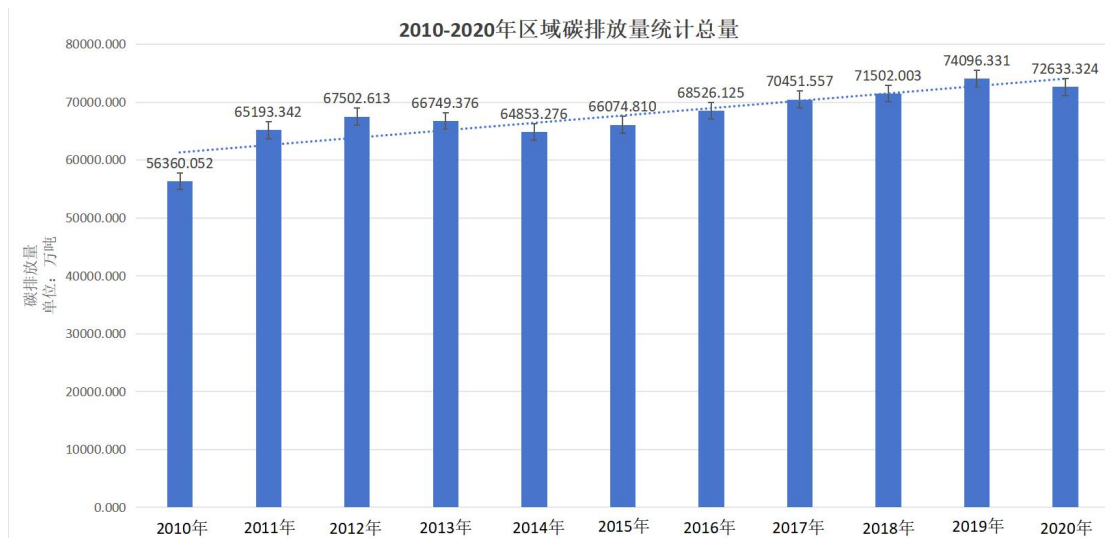


图3.4 2010-2020年碳排放总量统计

对 2010-2020 年碳排放总量进行数据统计分析，如图 3.1。通过深入挖掘，我们发现从 2010-2020 年这 11 年间，我们的研究区区域碳排放量总体呈上升趋势，在 2013-2015 年有了较小幅度的减少波动，随后仍呈上升趋势。最大排放量在 2019 年出现，最小排放量在 2010 年出现。

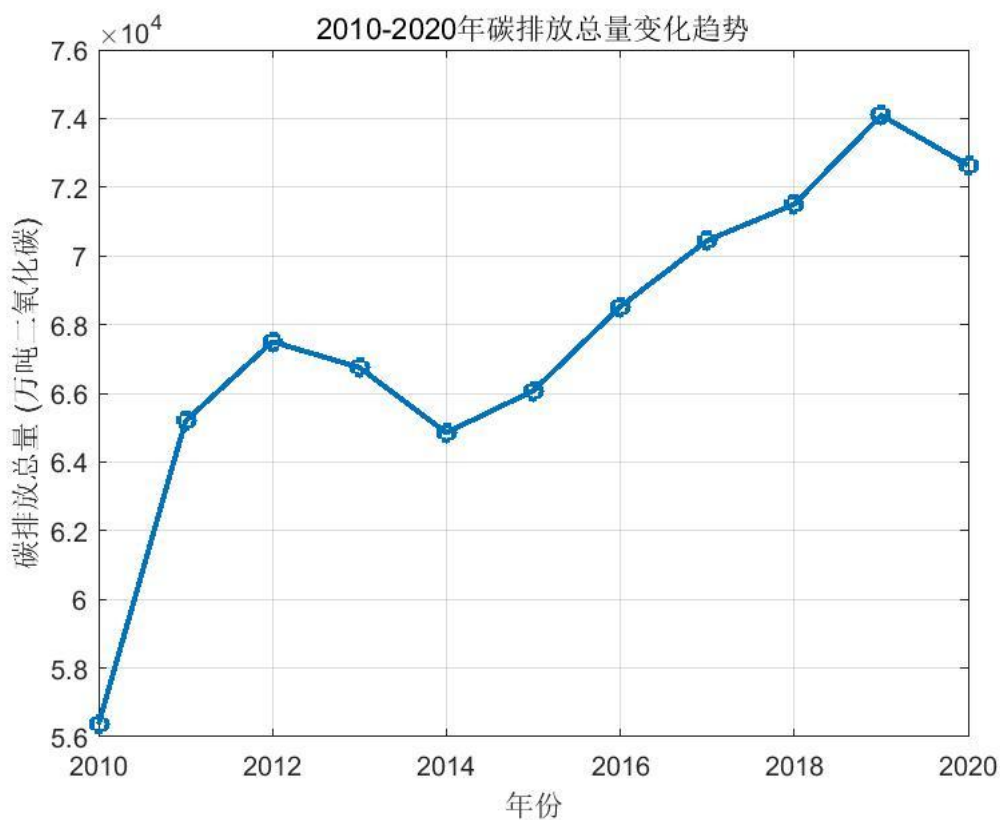


图 3.5 2010-2020 年碳排放总量变化趋势

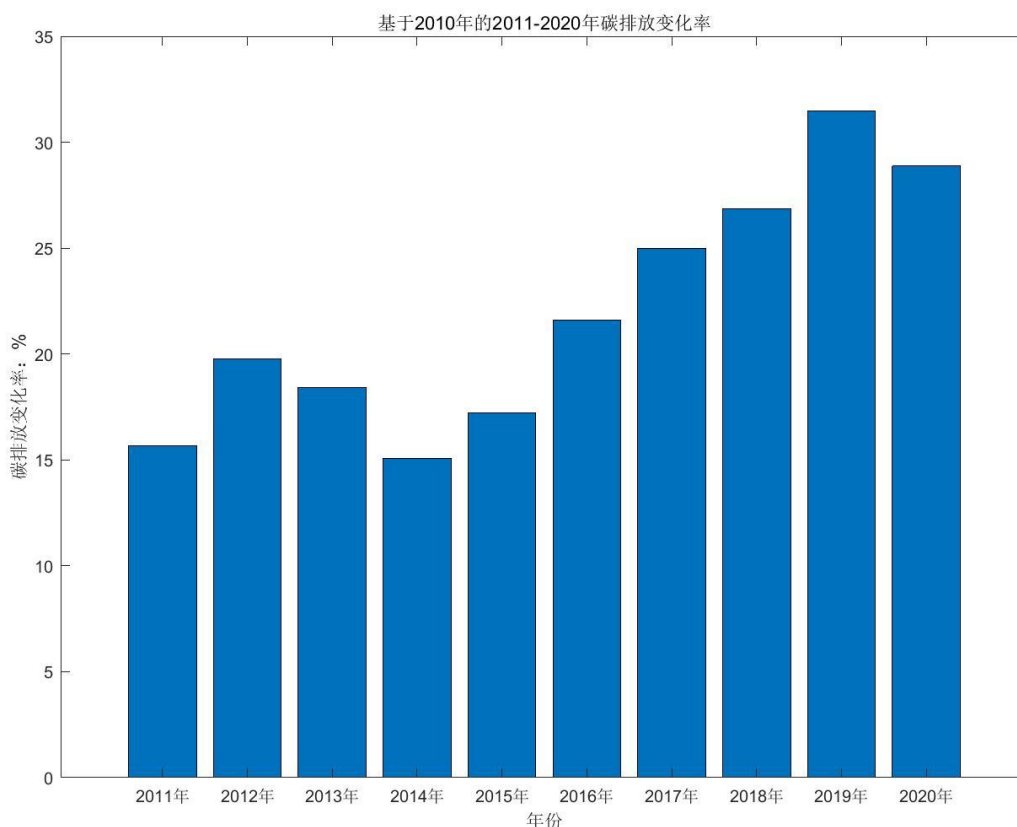


图 3.6 以 2010 年为基期的碳排放变化

随后以 2010 年为基期，对 2011-2020 年的碳排放量变化趋势进行分析，能够发现总体变化呈上升-下降-上升的趋势。在 2013 年我国启动清洁空气行动计划，成立第一个全国低碳日之后，碳排放量有下降趋势，在 2014 年变化率最低，随后开始增高。在 2019 年达到最高变化率之后开始下降。

分别以十二五（2011-2015 年）和十三五（2016-2020 年）为周期在软件中进行数据分析，在十二五期间，总碳排放量达到了 320658.66 万吨；在十三五期间，总碳排放量约 350655.83 万吨，十二五期间碳排放变化率约为 15.1%，十三五期间排放变化率约为 12.1%。

2013 年之后碳排放量有较为明显的下降，且十三五期间碳排放变化率下降，这与我国当年的国家政策有关。2013 年，我国采取了一系列减排措施，特别是风能和太阳能发电。旨在减少温室气体排放，改善环境质量，以及推动可持续发展。在此之后，我国经历了快速的工业化和城市化进程，在这个时期加大了基础设施建设的投资。随着城市化和经济增长，中国的交通运输需求也迅速增加。这也导致了更多的能源需求和城市基础设施建设，进而增加了碳排放。

中国政府已经采取了一系列措施来减少碳排放，包括提高能源效率、增加可再生能源比重、加强环保监管等。中国也在逐步推进碳市场建设，以鼓励减排。尽管碳排在某些年份上升，但政府的努力仍然旨在实现碳达峰并最终实现碳中和的目标。

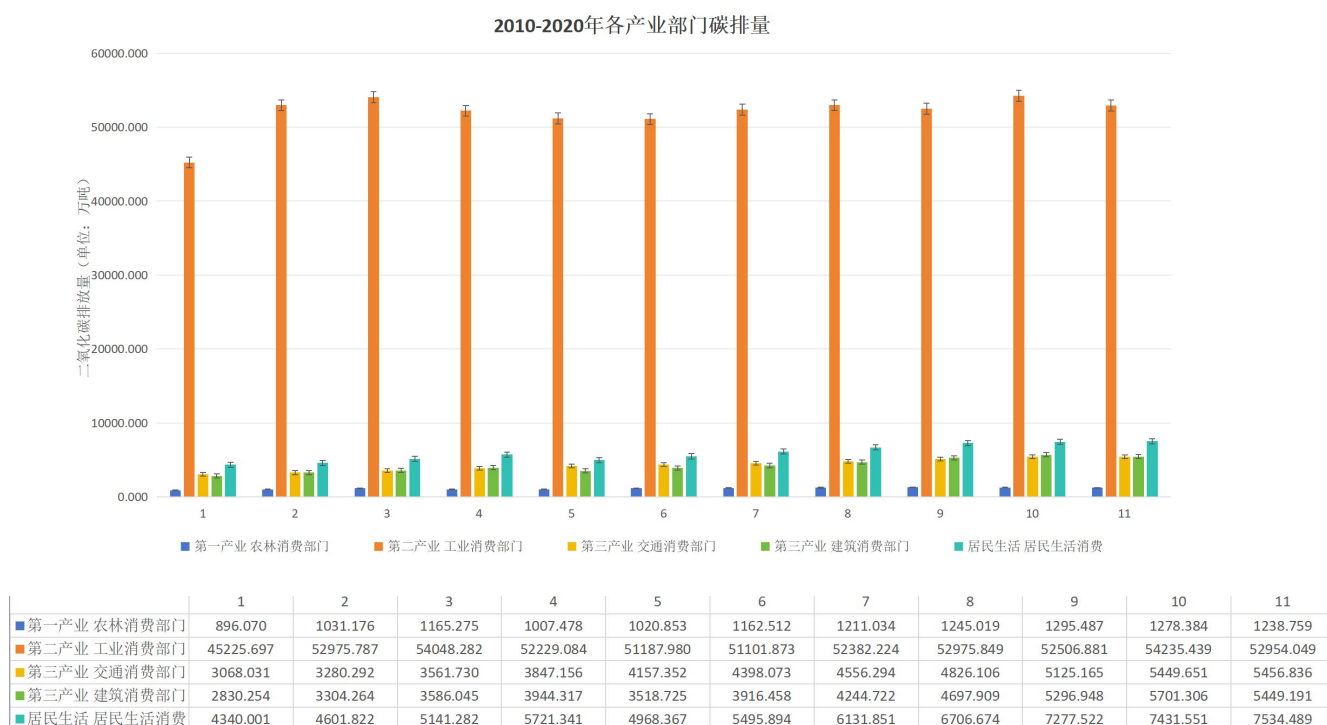


图 3.7 2010-2020 年各产业部分碳排放量

如图 3.4，对 2010-2020 年各产业部分总碳排放量进行数据统计分析。从图中可以很直观地看出，在该时期内，研究区的不同产业部门在碳排放中扮演着不同的角色。工业部门占据了相当大的份额，其碳排放量远远超过其他产业部门，总体来说呈现出稳定的年度变化趋势，没有出现较大的波动。这是因为中国是全球最大的制造业国家之一，工业和制造业的碳排放占据了较大的比重。在此时期，中国的工业和制造业快速持续发展，导致了多数碳排放的增加。

其次较高的是居民生活消费，碳排放量逐年增加，呈现逐年上升的趋势。交通和建筑部门的碳排放量相对较为接近，而农林消费部门在整体碳排放中所占比例较小。这一现象可能与我国经济水平逐年提升以及居民生活水平的改善密切相关。



图 3.8 2010-2020 年能源消费量

随后对 2010-2020 年该区域能源消费量作数据挖掘分析。2010-2020 年间，该区域能源消费量呈逐年上升趋势，趋势线显示高度一致。能源消费量逐年上升通常反映了在该时期内，研究区经济、人口呈增长态势，城市化进程加快，工业和制造业持续发展，生活水平提升。一定程度上反映了我国经济以及人口增长对于能源消费量的推动。

尽管能源消费量的上升在一定程度上与经济增长和生活水平提高有关，但也需要考虑能源效率的改进以及可再生能源的使用，以减轻能源消耗对环境的不利影响，实现可持续发展。因此，能源政策和技术创新也在努力寻找减少碳排放的途径。



图 3.9 2010-2020 年区域碳排放量与人口、经济、能源消费的相关性分析

最后对 2010 年-2020 年区域碳排放量与人口、经济、能源消费等指标进行相关性分析，综合前期数据，能够发现对于区域碳排放量，经济、人口、能源消费都有不同程度上的贡献。由图可知，其中最主要的是经济活动。经济活动是碳排放的主要来源之一。工业、制造、建筑、交通等经济部门的增长会导致更多的能源需求和碳排放。因此，经济活动对碳排放的贡献较大。其次是能源消费，各个部门产业的碳排放量也有一部分贡献。使用不同类型的能源会对碳排放产生不同的影响。依赖高碳能源如煤炭和石油的地区通常排放更多碳。提高可再生能源比例和能源效率可以减少碳排放。工业和制造业通常排放较多的碳，因为生产过程需要大量的能源。交通部门和建筑行业也会占一部分影响。此外，人口的贡献在该期间较为稳定，人口数量的增加也会增加对能源的需求，从而增加碳排放。

不同地区的碳排放因素可能会有所不同，具体的影响因素和贡献也会根据地理位置、产业结构、政策等因素而异。因此，在制定碳减排策略时，需要综合考虑这些因素，平衡好经济、人口增长与减少碳排放的相互关系，解决能源结构调整困难、非化石能源开发的限制等的矛盾，并根据具体情况制定相应的措施。

3.2.3 区域碳排放量以及经济、人口、能源消费量各指标及其关联模型

对各部门相关指标的同比变化进行计算，可以得到如图 3.10 和 3.11 的各部门能耗和碳排放同比变化结果图。通过两幅图的对比可以看出：不同部门的能耗与碳排放量变化情况

存在强相关性，各部门能耗与碳排放量之间存在很强相关性。

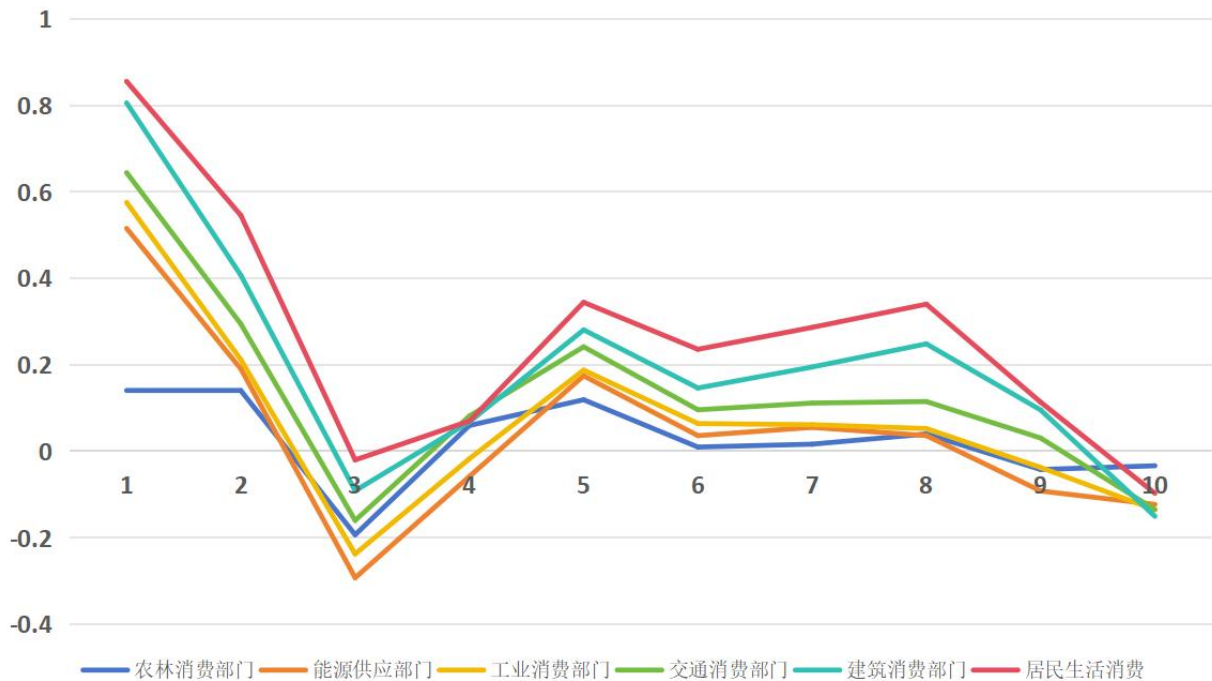


图 3.10 各部门能耗同比变化图

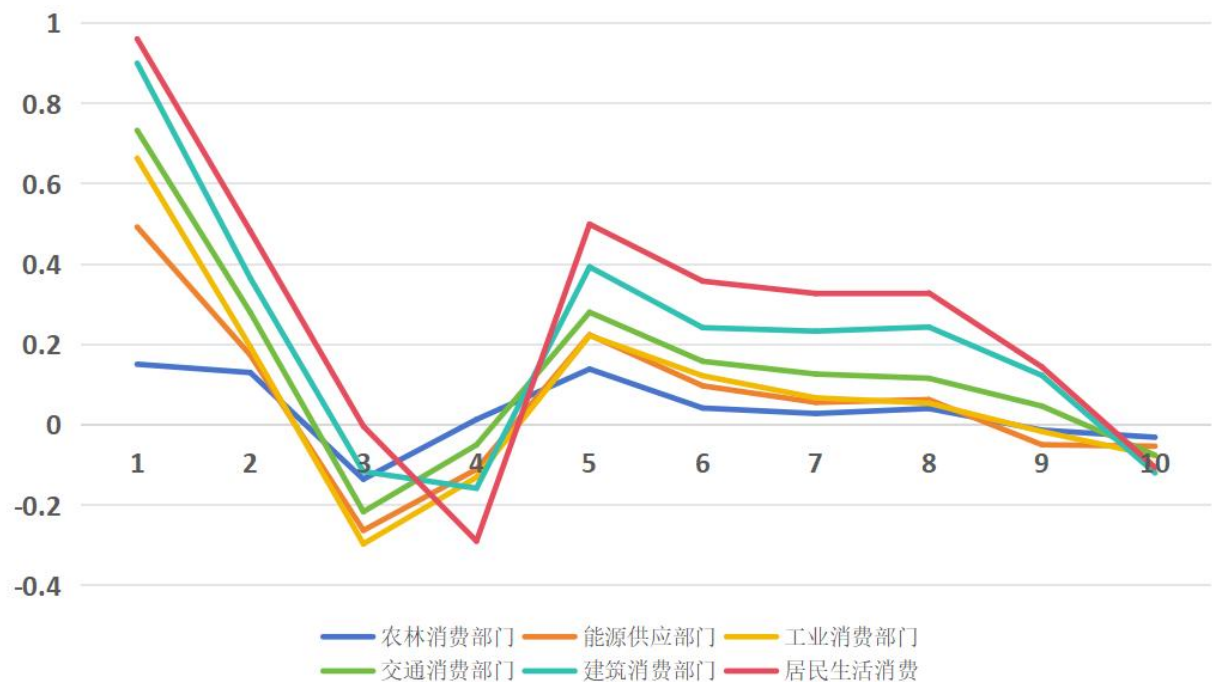


图 3.11 各部门碳排放总量同比变化图

从总体而言，无论是能耗还是碳排放，在 2011 和 2015 年的变化率均为最高，该现象可能表明了政策颁布开始，各部门从响应政策积极探索到找到正确方法合理降低能耗与碳排放的过程。对与不同部门而言，其逐年的同比变化趋势一致性较高，说明不同部门都受到了相同的政策约束，采取相似的措施进行节能减排。需要注意在 2013、2014 年存在的负增长可能是收到了经济不景气的影响因此减少。因此为探究区域碳排放量以及经济、

人口、能源消费量各指标之间的关系，进行多元线性回归分析。

根据题目介绍，Kaya 模型^[1]常常用来分析区域碳排放量和该区域人口、社会经济发展水平、能源利用效率以及碳排放因子的关系。其表达式为：

$$\text{二氧化碳排放量} = \text{人口} * \text{人均 GDP} * \text{单位 GDP 能耗} * \text{单位能耗二氧化碳排放量} \quad (4)$$

依据 Kaya 模型，我们可以认为对于二氧化碳排放量而言，其关联程度最大的主要指标包含模型中的人口、人均 GDP、单位 GDP 能耗以及单位能耗二氧化碳排放量。此外，非化石能源消费比重、能源消耗总量也是独立于以上几种指标并对二氧化碳排放量具有重要影响的指标之一。为探究这些指标之间的关联关系，首先以 2010 年数据为基准对数据进行初值化并画出散点图（图 3.12）进行分析。

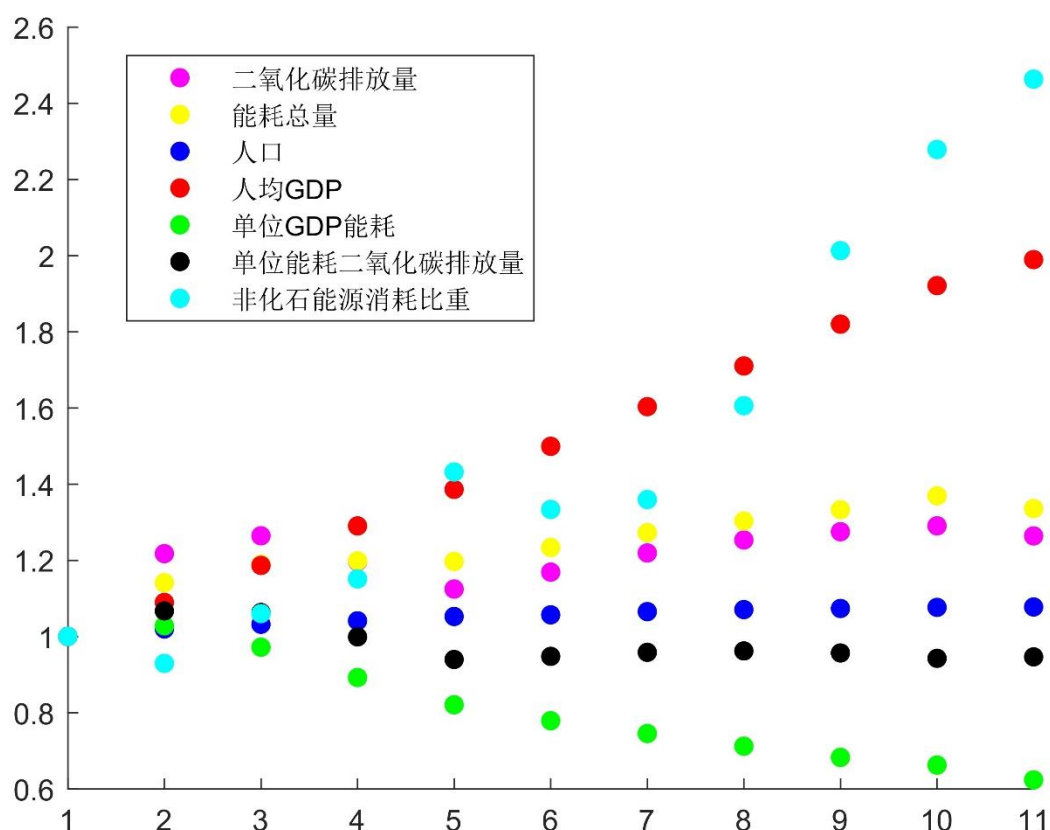


图 3.12 关联指标散点图

图3.12中，不同类型的散点在11年间的趋势大致呈现线性变化，因此这些指标之间很有可能存在线性关系。尝试利用除碳排放总量与能耗总量以外的其他指标对能耗总量进行多元线性回归。其拟合结果如下：

$$\begin{aligned} \text{能耗总量} = & -115406.24 + 13.12 * \text{人口} + 2350.20 * \text{人均GDP} + 51922.12 * \text{单位GDP能耗} + \\ & (-1501.47) * \text{单位能耗二氧化碳排放量} + (-47113.61) * \text{非化石能源消耗比重} \end{aligned} \quad (5)$$

其中人口的单位为万人，人均GDP的单位为亿元/万人，单位GDP能耗的单位为万tce/亿元，单位能耗二氧化碳排放量的单位为万tCO₂/万tce，非化石能源消耗比重的单位为1。该拟合结果的决定系数R²为0.9954，p值为7.8*10⁻⁶<0.05，因此拟合结果可信度较高。对碳排放总量的拟合结果如下：

碳排放总量 $= -91373.69 + 0.49 \times \text{人口} + (-1555.82) \times \text{人均GDP} + (-28075.50) \times \text{单位GDP能耗} + 30660.82 \times \text{单位能耗二氧化碳排放量} + 41483.72 \times \text{非化石能源消耗比重} + 3.64 \times \text{能耗总量}$ (6)

式中单位同上，能耗总量的单位为万tce。该拟合结果的决定系数 R^2 为1， p 值为 $8.25 \times 10^{-10} < 0.05$ ，因此拟合结果较好。图3.7分别展示了能耗和碳排放的拟合结果与原始数据对比的结果。

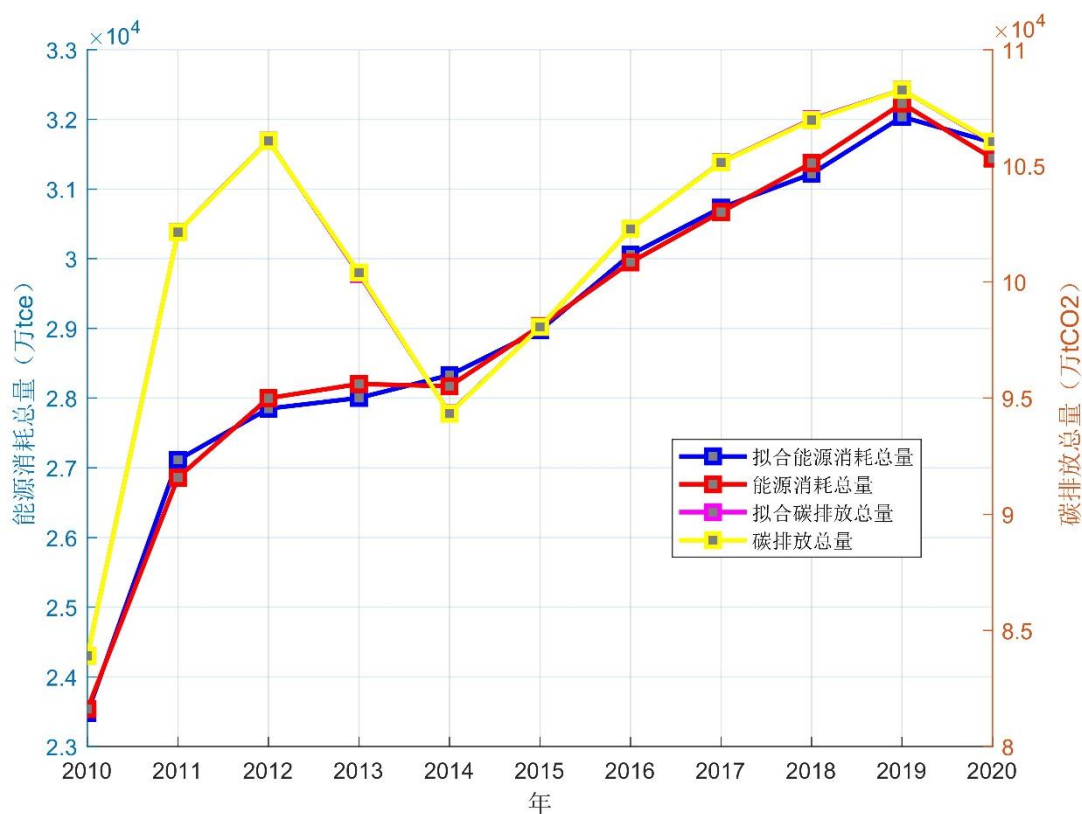


图 3.13 能耗和碳排放拟合结果与原始数据对比图

通过拟合结果图可以看出，碳排放与能耗和人口、人均GDP、单位GDP能耗以及单位能耗二氧化碳排放量和非化石能源消费比重的关联性可以通过回归分析模型来确定，并可以得到可信度很高的拟合结果。同时通过回归方程可以确定能源利用效率提升（单位GDP能耗）和非化石能源消费比重（降低单位能耗碳排放）等指标对于碳排放量的影响参数。

根据双碳政策，碳达峰是指2030年之前要求碳排放量对时间的偏导数为0，碳中和是指碳排放量与碳汇（生态碳汇+工程碳汇+碳交易）消纳量相平衡，要求非化石能源消费比重 $>80\%$ 。因此基于双碳政策以及未来可能存在的技术进步或政策颁布，该模型未来可以基于现有模型添加更多的参数进行预测。

4 问题二的建模与求解

4.1 问题分析及建模思路

为实现双碳目标，必须解决发展与碳减排之间的矛盾。深入研究人口、经济和能源消耗之间的内在联系，以及对未来能源需求的预测至关重要。同时，碳排放的预测作为实现双碳目标的理论依据和研究基础，其准确性直接影响着该地区政策和目标的设定。

在前文中，我们已经分析了该地区在 2010-2020 年间的碳排放量、经济状况、人口数据以及能源消耗情况。接下来的工作是深入研究这些因素之间的相互关系，以及它们对经济发展和碳排放的影响，寻找内在的动因。我们需要预测该地区在十四五规划期间（2021-2025 年）至二十一五年（2056-2060 年）的人口、经济、能源消耗和碳排放的变化趋势。

在建立能源消耗预测模型时，必须考虑人口和经济对能源需求的影响。同时，在构建碳排放预测模型时，需要充分考虑人口、经济和能源消耗对碳排放的影响，并进一步研究各个能源消耗领域和能源供应领域的能源消耗水平以及不同能源类型对碳排放的影响。

4.1.1 人口和 GDP 增长模型构建思路

本实验的解决方法为，首先采用 Logistic 曲线建立该区域的人口增长模型。其次，在选取 GDP 增长模型方面，考虑到训练样本不足，以及需要对未来进行长时间的预测，我们对对比了指数增长模型，ARIMA 模型和多项式拟合模型的拟合优度以及适用性，最后选择了多项式拟合模型用来预测 GDP 未来发展态势。通过上述建立的人口增长模型和 GDP 增长模型，最后得到 2010-2060 年该区域完整的人口及 GDP 年度变化数据，为之后建立能源消耗模型以及碳排放模型奠定基础。

4.1.2 能源消耗量模型构建思路

在建立能源消费量预测模型中，需要考虑（1）人口和 GDP 对能源消费量的影响。（2）能源消费量时间序列自相关。（3）人口增长率和 GDP 增长率对能源消费量的影响。我们采用两种方式对未来该地区的能源消费量进行预测分析，分别探讨了多元线性回归以及机器学习的方法对于能源消费量变化模型的建立效果。

方案一，选取 2010-2020 年人口和 GDP 数据，利用多元线性回归模型建立该地区的能源消费量模型，以预测 2021-2060 年该地区的能源消费量。

方案二，考虑到人口和 GDP 数据对能源消费量的解释并非线性关系，我们拟采用机器学习方法对其进行解释。但由于机器学习方法需要较大的训练样本，我们计算了 2010-2050 年的人口增长率和 GDP 增长率时间序列，并利用方案一中得到 2010-2050 年的能源消费量年度变化时间序列数据作为训练的输入样本，去预测 2050-2060 年区域的能源消费量变化。机器学习算法方面，我们选用了长短期记忆递归神经网络（LSTM）和 BP 神经网络模型探求人口、GDP 因素对于能源消费量的影响。

4.1.3 碳排放量模型构建思路

在得到人口、GDP 和能源消费量的 2010-2060 年完整的时间序列变化数据后，我们将探讨其对碳排放量的影响，以期实现对碳排放量的解释，为后续的工作计划制定提供充分的理论依据。在建立碳排放量模型中，需要考虑以下三个条件，并量化为具体指标：

（1）碳排放量与人口、GDP 和能源消费量相关联。（2）碳排放量与各能源消费部门以及能源供应部门的能源消费量相关联，其中具体包括工业消费部门、建筑消费部门、交通消费部门、居民生活消费、农林消费部门以及能源供应部门六个部门。（3）碳排放量与各能源消费部门以及能源供应部门的能源消费品种相关联。其中关于条件 2，本实验量化计算 2010-2020 年区域各部门的能源消费量变化时间序列；关于条件 3，本实验量化计算各 2010-2020 年区域各部门的非化石能源消费比重变化时间序列。

在模型构建方面，采用差分自回归移动平均模型（ARIMA）对 2020-2060 年各部门的能源消费量以及其非化石能源消费比重进行预测。在数据处理方面，通过差值方法得到能源供应部门的能源消费量，比值方法计算各部门的非化石能源消费比重。

最后利用预测得到的 2020-2060 年各影响指标因子的时间序列，通过对比多元线性回归模型，BP 神经网络以及 LSTM 模型构建 2010-2060 年碳排放量时间序列变化模型。

4.2 模型简介

4.2.1 Logistic 曲线

1845 年，比利时数学家 Pierre François Verhulst 提出了 logistic 方程，这是一个用于数学描述 S 型曲线的模型。这个方程已经在多个领域被广泛应用，用于建立和预测单位面积内某种生物的数量、社会经济指标如人口数量、某种商品的普及率等特殊情境。该模型考虑了社会环境中的人口最大承载能力，其人口增长率一开始上升，然后逐渐减小并趋于零，这与我国人口数量的发展趋势基本吻合。因此，本实验采用 Logistic 曲线来预测该地区从 2020 年到 2060 年的人口数量。

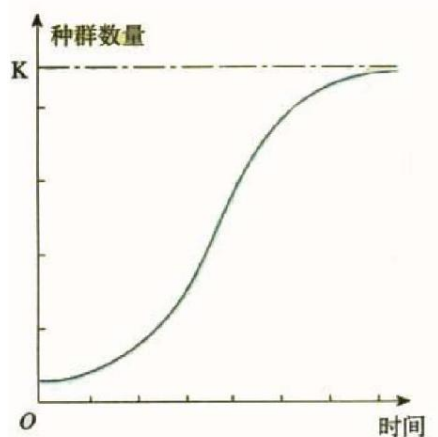


图 4.1 逻辑斯蒂增长模型 S 形曲线

4.2.2 多项式拟合模型

多项式拟合是一种常用的统计学方法，用于研究模型参数与数据之间的关系，并建立参数与数据的统计模型。它的基本思想是通过分析一系列数据，将其拟合成一个多项式方程。通过解决这个多项式方程，可以迅速且准确地获取模型的参数值。

假设有 m 对数据，如下所示：

$$\{(x_1, y_1)(x_2, y_2) \dots (x_m, y_m)\}$$

将其拟合成某 n 次多项式，求解出 a_0, a_1, \dots, a_n 这 $n+1$ 个拟合系数。

$$\hat{y} = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n$$

4.2.3 差分自回归移动平均模型 (ARIMA)

常用的时间序列模型有四种：自回归模型 AR(p)、移动平均模型 MA(q)、自回归移动平均模型 ARMA(p, q)、差分自回归移动平均模型 ARIMA(p, d, q)，可以说前三种都是 ARIMA(p, d, q) 模型的特殊形式。[1]ARIMA(p,d,q)中，AR 是"自回归"，p 为自回归项数;MA 为"滑动平均"，q 为滑动平均项数，d 为使之成为平稳序列所做的差分次数(阶数)。具体模型建立步骤为：1、获取被观测系统时间序列数据；2、对数据绘图，观测是否为平稳时间序列；对于非平稳时间序列要先进行 d 阶差分运算，化为平稳时间序列；3、经过第二步处理，已经得到平稳时间序列。要对平稳时间序列分别求得其自相关系数 ACF 和偏自相关系数 PACF，通过对自相关图和偏自相关图的分析，得到最佳的阶数 p 和阶数 q。4、由以上得到的 d、q、p，得到 ARIMA 模型。之后进行模型检验以及对未来结果进行预测。

4.2.4 多元线性回归分析

用回归方程定量地刻画一个应变变量与多个自变量间的线性依存关系，称为多元回归分析 (multiple linear regression)，简称多元回归(multiple regression)。回归分析的基本思想是：虽然自变量和因变量之间没有严格的、确定性的函数关系，但可以设法找出最能代表它们之间关系的数学表达形式。

值得注意的是，在建立多元线性回归模型中，需要考虑到自变量之间的多重共线性。本实验采用 VIF 方法进行检验，利用逐步回归方法选取合适的自变量因子，避免较高的多重共线性。

4.2.5 长短期记忆递归神经网络模型 (LSTM)

LSTM 网络是一种循环神经网络 (RNN)，它通过遍历时间步并更新 RNN 状态来处理输入数据。RNN 状态包含在所有先前时间步中记住的信息。我们可以使用 LSTM 神经网络，通过将先前的时间步作为输入来预测时间序列或序列的后续值。要为时间序列预测训练 LSTM 神经网络，请训练具有序列输出的回归 LSTM 神经网络，其中响应（目标）是将值移位了一个时间步的训练序列。也就是说，在输入序列的每个时间步，LSTM 神经网络都学习预测下一个时间步的值。作为循环神经网络 (RNN) 的一种改进，LSTM 网络结构中新建立了在时间方向上含有反馈连接的记忆单元，具体是添加了 3 个门结构：输入门、遗忘门和输出门，用于控制网络状态信息的输入、删除和输出，其核心公式如下^[2]：

$$\text{输入门控制器: } i^{(t)} = \sigma(w_{ix}x^{(t)} + w_{ih}h^{(t-1)} + b_i)$$

$$\text{遗忘门控制器: } f^{(t)} = \sigma(w_{fx}x^{(t)} + w_{fh}h^{(t-1)} + b_f)$$

$$\text{输出门控制器: } o^{(t)} = \sigma(w_{ox}x^{(t)} + w_{oh}h^{(t-1)} + b_o)$$

式中， $x^{(t)}$ — t 时刻的输入变量； σ —sigmoidal 函数； w 、 b —各门的权重、偏置； h 、 S —隐藏、单元状态。

4.2.6 BP 神经网络模型

BP 神经网络是一种按误差反向传播训练的多层前馈网络，其算法称为 BP 算法，它的基本思想是梯度下降法，利用梯度搜索技术，以期使网络的实际输出值和期望输出值的误差均方差为最小。基本 BP 算法包括信号的前向传播和误差的反向传播两个过程。即计算误差输出时按从输入到输出的方向进行，而调整权值和阈值则从输出到输入的方向进行。通过调整输入节点与隐层节点的联接强度和隐层节点与输出节点的联接强度以及阈值，使误差沿梯度方向下降，经过反复学习训练，确定与最小误差相对应的网络参数。

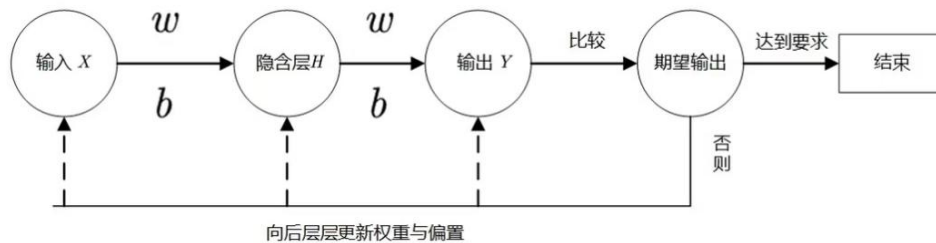


图 4.2 BP 神经网络算法流程图

BP 神经网络算法常用的激活函数：1) Sigmoid (logistic) 2) Tanh 函数（双曲正切函数）3) relu 函数是一个通用的激活函数，针对 Sigmoid 函数和 Tanh 的缺点进行改进的，目前在大多数情况下使用。BP 网络由输入层、隐藏层、输出层组成。输入层就是读入我们输入的数据；隐藏层为信息的处理端；输出层则输出我们需要的结果。

4.3 模型建立

4.3.1 人口增长模型

采用逻辑斯蒂模型对 2010-2020 年区域人口总量进行回归分析，建立模型，以预测 2020-2060 年区域人口总量变化时间序列。由图可知，该区域人口增长趋于平缓，人口增长率逐年递减。其中模型的 RMSE 为 14.72，拟合良好。

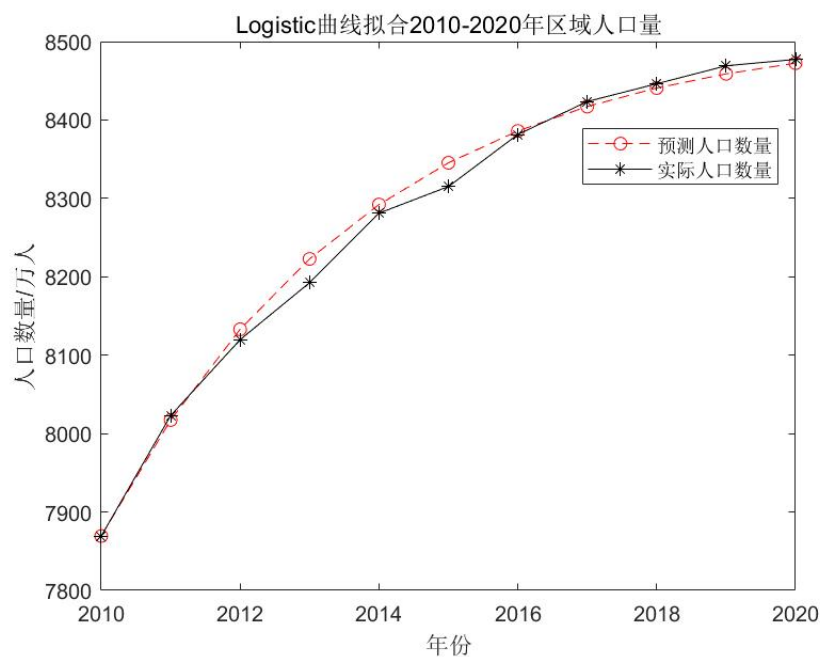


图 4.3 Logistic 曲线拟合 2010-2020 年区域人口量

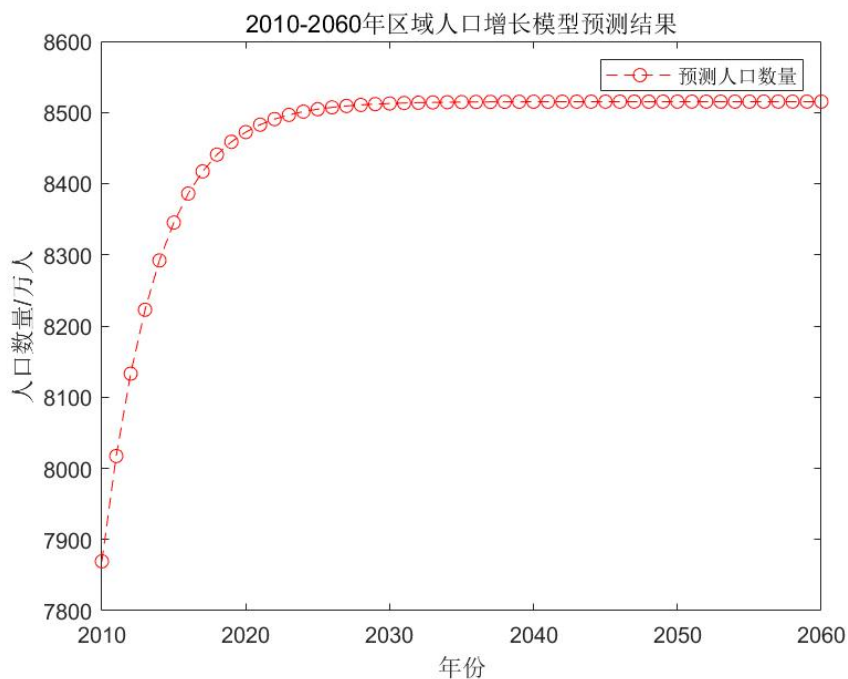


图 4.4 2010-2060 年区域人口增长模型预测结果

4.3.2 GDP 增长模型

在建立 GDP 增长模型方面，我们选取了指数模型，多项式拟合模型以及 ARIMA 模型进行建模分析，对比各个模型的拟合精度。其中多项式拟合模型的 RMSE 为 500.6，指数增长模型的 RMSE 为 1962.6。通过 ADF 检验以及 AIC, BIC 准则确定 ARIMA 模型的 p, d, q 参数，最后选用 ARIMA (0, 2, 0) 进行拟合，模型的 AIC 为 143.1。

考虑到国家总体规划目标中要求 2035 年的 GDP 比基期（2020 年）翻一番，2060 年比

基期翻两番。2020 年 GDP 总值约为 88683 亿元,那么预计在 2035 年 GDP 总值约为 177366 亿元,2060 年 GDP 总值约为 354732 亿元。对比这三个模型的未来变化情况,多项式拟合模型主要是呈现线性增长态势,较为符合国家总体规划目标中的 GDP 总值要求。指数模型预测的 GDP 总值呈现指数型“爆炸”增长,并不符合实际情况。ARIMA 模型预测的 GDP 总值呈现先增长后减少的发展态势,也不符合实际情况。综上,我们认为在预测 2020-2060 年的区域 GDP 总值变化时间序列上,多项式拟合模型表现较好,最后也选取该模型作为预测模型。

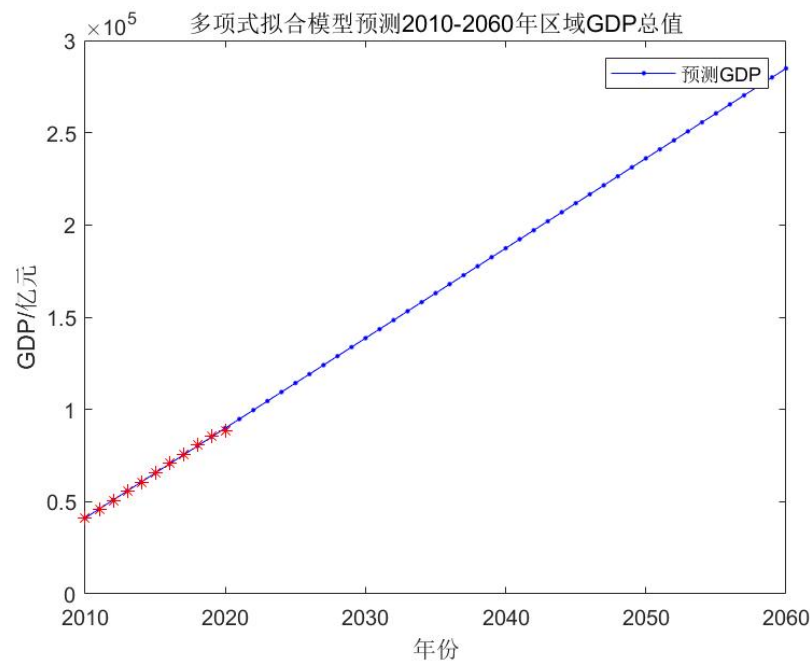


图 4.5 多项式拟合模型预测 2010-2060 年区域 GDP 总值

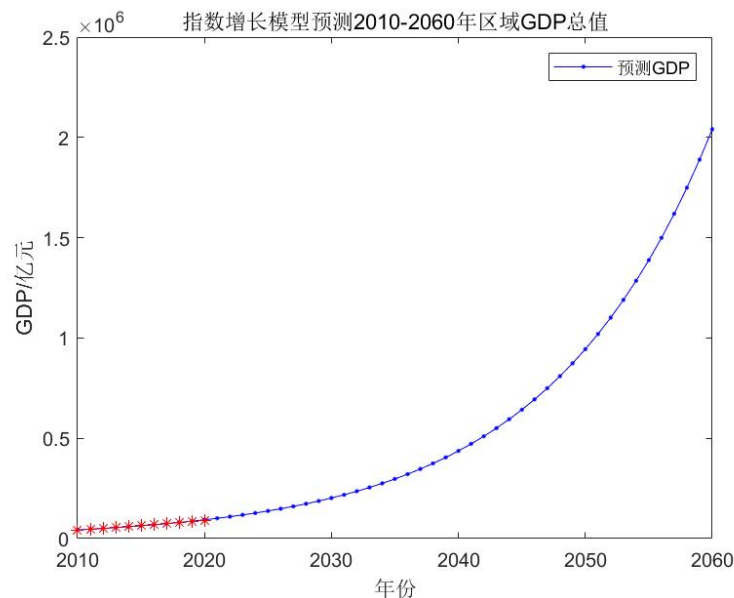


图 4.6 指数增长模型预测 2010-2060 年区域 GDP 总值

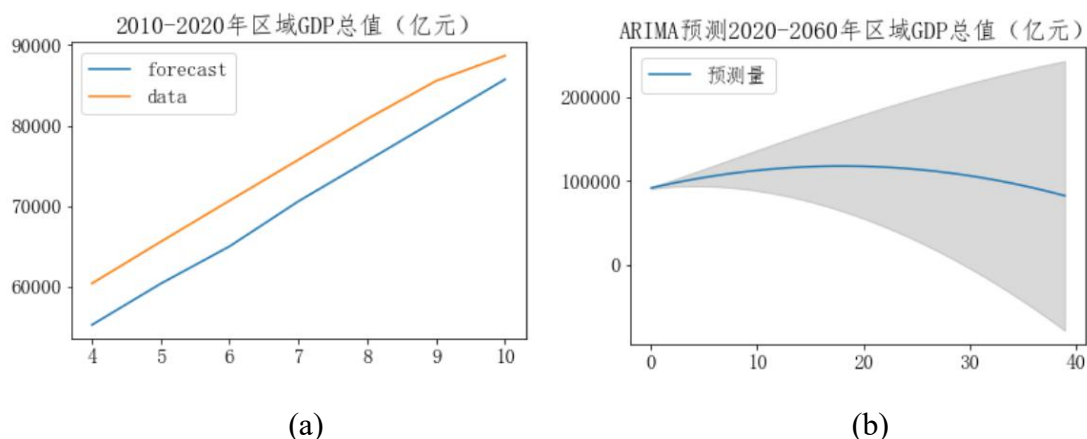


图 4.7 a. 基于 ARIMA 模型的 2010-2020 年区域 GDP 预测值与实际值对比
b. 基于 ARIMA 模型的 2020-2060 年区域 GDP 总值预测值

4.3.3 基于人口和经济变化的能源消费量预测模型

方案一：通过 2010-2020 年的区域人口、GDP 总值的时序曲线，对 2010-2020 年区域能源消耗量的时序曲线进行多元线性回归。之后通过建立的多元线性回归模型预测 2020-2060 年区域能源消耗量的变化情况。

其中模型 $r^2=0.9325$ ， p 值=0.00002，证明模型精度达到要求。

多元线性回归表达式如下：能源消耗量=-43717.48+人口数量*8.41+0.048*GDP 总值；

观察数据可以发现，人口数量和 GDP 总值与能源消耗量之间存在着明显的正相关关系，尤其值得注意的是，人口数量对能源消耗量的影响程度较为显著。在前文中，我们进行了人口数量的未来增长预测，并预见到其增长趋势将逐渐趋于平稳。这一趋势也直接影响了未来的能源消耗量，使其呈现出逐渐平稳的态势。这一趋势预示着我们在可持续发展和资源管理方面需要更多地关注能源效率和替代能源的使用，以实现资源的可持续利用和环境的保护。这也强调了人口增长与经济发展之间的复杂关系，对于能源政策和规划的重要性。

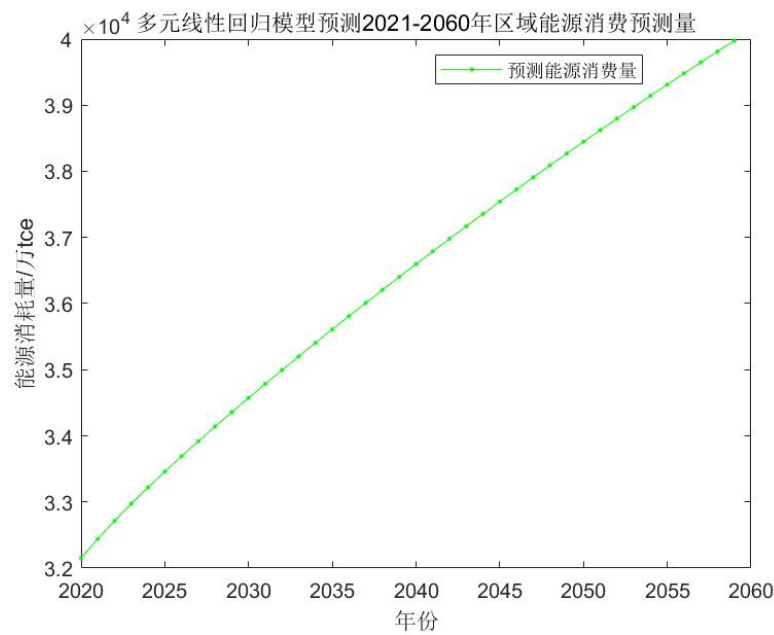


图 4.8 基于多元线性回归模型预测 2021-2060 年区域能源消费预测量

方案二：基于 BP 神经网络以及 LSTM 模型，利用 2010-2050 年区域人口、GDP 总值、人口增长率、GDP 增长率的时序变化曲线，以及前文通过多元线性回归模型得到的 2010-2050 年的能源消费量时序变化曲线，对 2050-2060 年的区域能源消费量进行模拟预测。

其中选用 BP 神经网络训练时，选用 2010-2050 年间 40 年的区域人口、GDP、人口增长率、GDP 增长率作为输入数据，区域能源消费量时间序列数据为输出数据。其中设置前 30 年为训练样本，后 10 年为测试样本。神经网络设置输入层节点 401 个，隐含层节点 10 个，输出层节点 1 个。该模型的平均绝对误差 MAE 为 105.7849，均方误差 MSE 为 13402.5063，均方根误差 RMSE 为 115.7692，拟合精度较好。

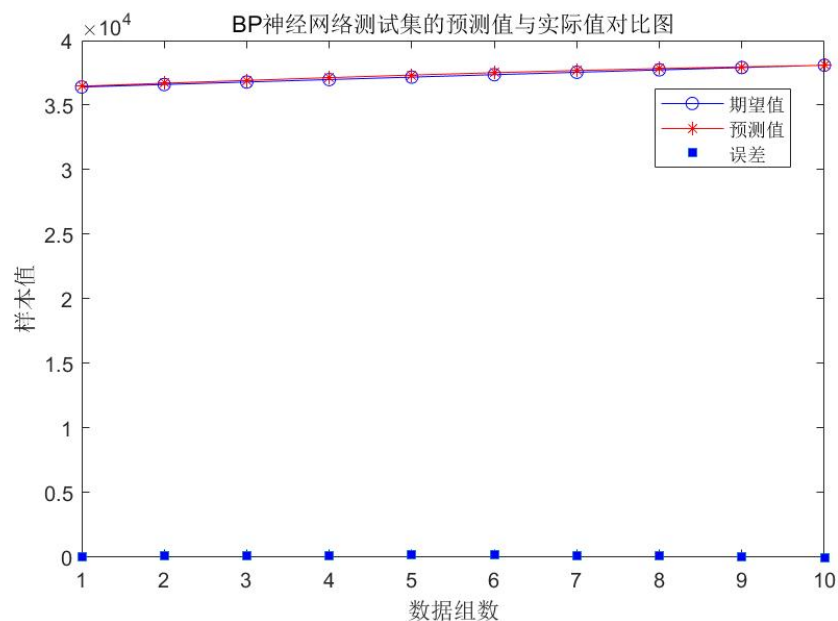


图 4.9 基于 BP 神经网络预测能源消费量测试集的预测值和实际值对比图

之后利用训练得到的 BP 神经网络模型对 2050-2060 年的区域能源消费量进行预测，预测结果如表 4.1 所示，该区域的能源消费量保持平缓增长。

表 4.1: BP 神经网络预测 2050-2060 年区域能耗量结果

年份	区域能耗量（万 tce）
2051	36990.62952
2052	37034.30231
2053	37073.92285
2054	37109.93673
2055	37142.74198
2056	37172.69256
2057	37200.10223
2058	37225.24842
2059	37248.37602
2060	37269.70096

选用 LSTM 模型训练时，输入，输出及测试数据和训练 BP 神经网络相同。模型参数设置上，numhidden_units1=140; numhidden_units2=150; numhidden_units3=100。训练五次，减少随机误差，最后一次训练过程如下图所示（图 4.10）。

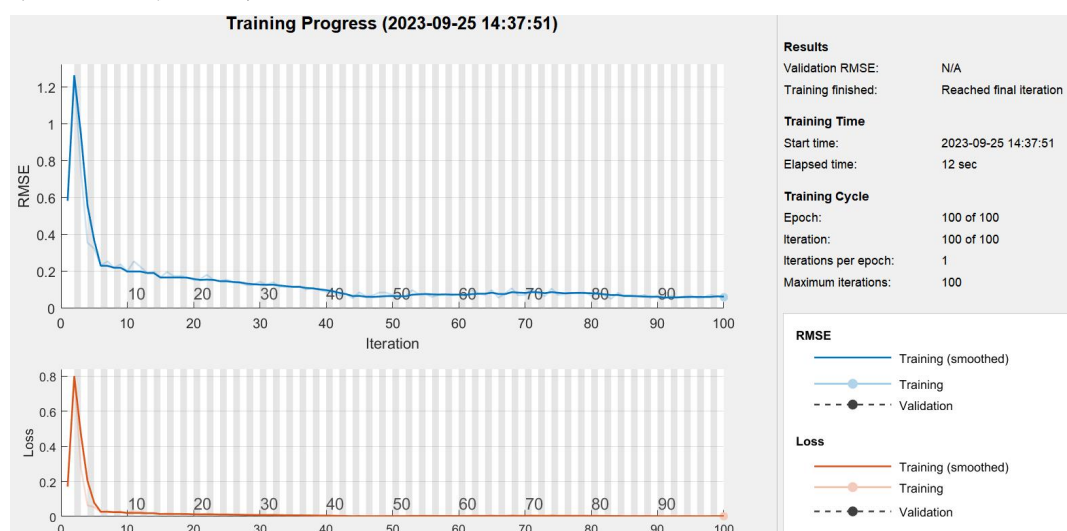


图 4.10 第五次 LSTM 训练过程

LSTM 网络模型的预测绝对平均误差 MAE 为 289.5815，平均绝对误差百分比 MAPE 为 0.0073，均方根误差 RMSE 为 346.58，模型拟合精度总体而言较好，但是与多元线性回归模型和 BP 神经网络相比略输一筹。机器学习（BP 神经网络和 LSTM 模型）算法模拟的能源消费量到后期都呈现增长偏缓态势，这样的预测结果相较于多元线性回归的线性上升态势更加符合实际以及国家发展总体规划目标。

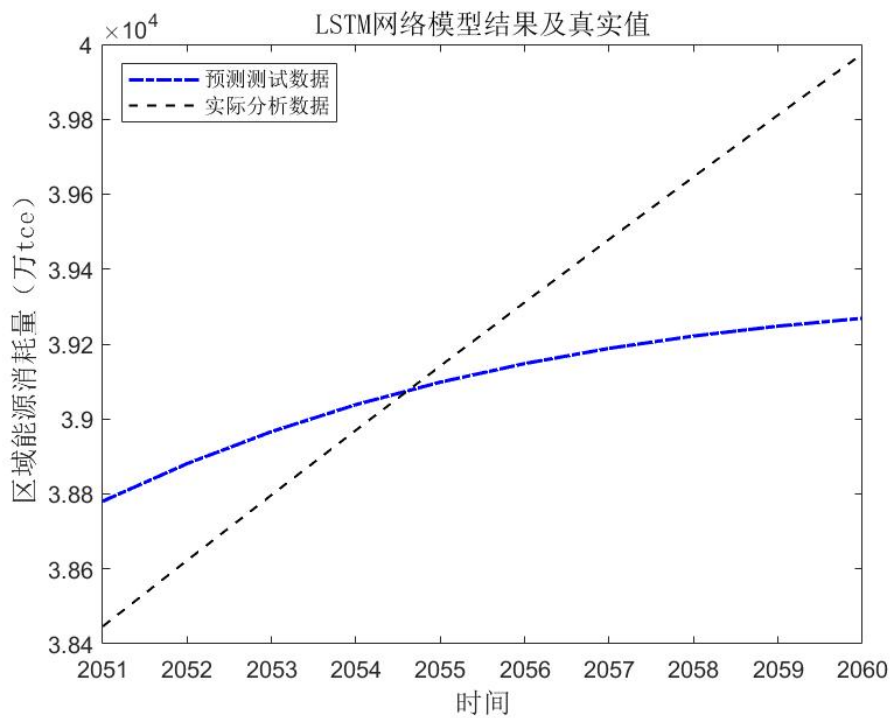


图 4.11 基于 LSTM 网络预测能源消费量测试集的预测值和实际值对比图

表 4.2: LSTM 模型网络预测 2050-2060 年区域能耗量结果

年份	区域能耗量（万 tce）
2051	38664.813
2052	38762.461
2053	38845.941
2054	38916.688
2055	38976.207
2056	39025.965
2057	39067.355
2058	39101.645
2059	39129.953
2060	39153.250

最终，我们对这两种模型的结果进行了详细的分析。在与多元线性回归模型进行比较后，神经网络预测模型展现出了更为精确和全面的能耗量预测。

我们的研究表明，区域能耗量的预测呈现出明显的上升趋势，逐步增加，并在后期阶段逐渐趋于平稳，最终达到饱和状态。这一趋势与能源需求的不断增长以及区域经济发展的特点相一致。通过深入研究这一现象，我们可以更好地理解能源消耗的动态变化，并为未来的能源规划和可持续发展提供有力的支持。

4.3.4 基于多因素的区域碳排放量预测模型

首先，对该模型的输入数据进行预处理，包括统计 2010-2020 年各部门的能源消费量，

2010-2020 年各部门的化石能源消费量与非化石能源以及二次能源（电或热）消费量，计算 2010-2020 年各部门的非化石能源消费占比（包括非化石能源和二次能源）。其中，化石能源消费量主要为煤炭、油品和天然气消费量；非化石能源以及二次能源（电或热）消费量主要为热力、电力和其他能源消费量。下图（图 4.12-4.14）展示了 2010-2020 年各指标因子的时间序列变化情况，可以宏观掌握这 10 年间各指标量的变化情况。注意，能源供应部门的能源消费量为负值表示为该部门进行的非化石能源发电量。

观察这 11 年的各部门能源消费量以及部门内的非化石能源消费占比变化，我们可以得知以下信息：（1）各部门的能源消费量以及能源消费总量总体都呈现缓慢上升趋势。其中，工业部门的能源消费量约占能源消费总量的 60%，对总体的能源消费量变化影响较大。其次是能源供应部门，约占 25%。（2）建筑部门的非化石能源以及二次能源（电或热）的消费量远高于化石能源消费量，而其他部门都是化石能源消费量高于非化石能源以及二次能源（电或热）的消费量。（3）建筑部门的非化石能源消费占比很高，且不断上升，从 2010 年的 77.55% 上升至 2020 年的 97.78%。交通部门的非化石能源消费占比是最低的，这 11 年都不高于 5%。

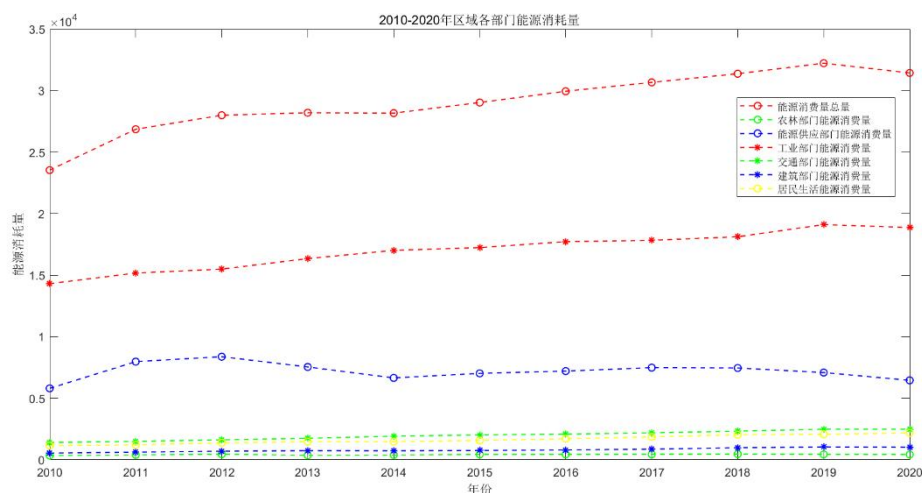


图 4.12 2010-2020 年区域各部门的能源消费量变化时间序列

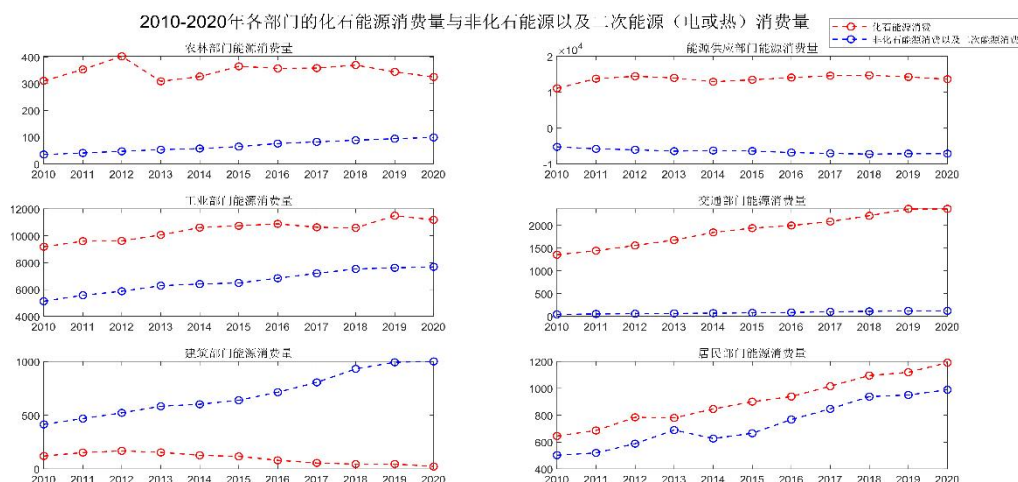


图 4.13 2010-2020 年各部门的化石能源消费量与非化石能源以及二次能源（电或热）消费量

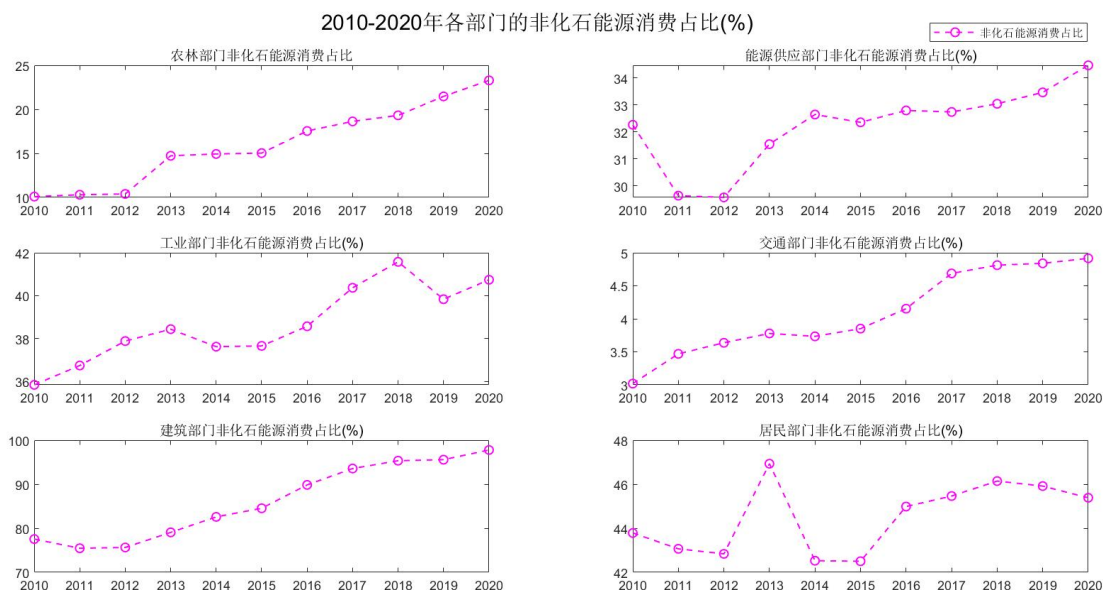


图 4.14 2010-2020 年各部门的非化石能源消费百分比

本实验想要构建 ARIMA-LSTM 或者 ARIMA-BP 神经网络组合模型。所以之后需要先利用 ARIMA 模型对这些自变量指标进行后 40 年的时间序列变化预测，并将其作为后续机器学习模型的输入数据。

接下来以工业部门能源消费量为例，说明 ARIMA 模型的预测过程。ARIMA 模型要求时间序列满足平稳性，首先进行 ADF 检验，根据 t 统计量， p -value 值判断该时间序列是否是平稳的。若不平稳，则通过差分处理使得数据变成平稳时序，并以此确定 ARIMA 模型的 d 值。对 2010-2020 年工业部门能源消费量进行绘图，观察到非平稳时间序列，所以之后对该非平稳时间序列进行 d 阶差分运算，化为平稳时间序列。

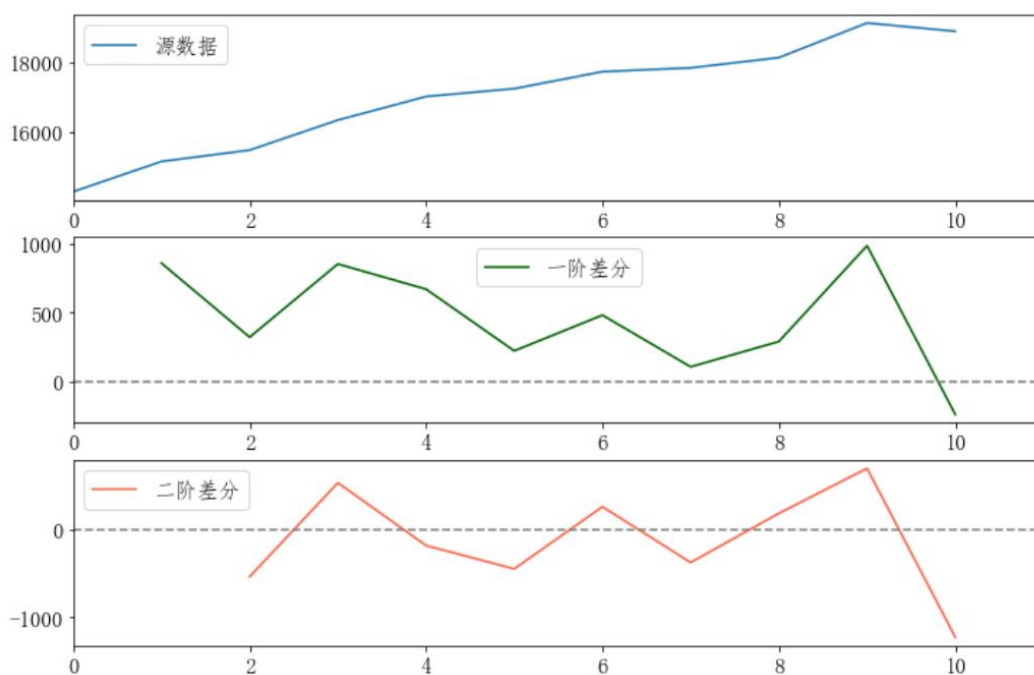


图 4.15 工业部门能源消费量源数据以及一、二阶差分后的数据

表 4.3 原始序列和一阶差分序列的 ADF 检验结果

	原始序列	一阶差分序列
T 统计量	-1.57	-3.88
P_value	0.50	0.002
usedlag	1	0
nobs	9	9
在 99%置信区间下的临界的 ADF 检验的值	-4.473	-4.473
在 95%置信区间下的临界的 ADF 检验的值	-3.289	-3.289
在 90%置信区间下的临界的 ADF 检验的值	-2.772	-2.772

观察到一阶差分序列的 P 值很小，接近 0，且 T 统计量小于 95%置信区间下的临界的 T 值，由此认为该序列为平稳时间序列，将工业部门能源消费量 ARIMA 模型的 d 值定为 1。

之后利用 AIC, BIC 法则确定 ARIMA 模型的 p, q 值。根据 AIC, BIC 最小原则都筛选出 p=1, q=0。由以上得到的 p, d, q 构建 ARIMA (1, 1, 0) 模型，该模型的 AIC 为 150.13，BIC 为 151.04。

表 4.4 ARIMA (1, 1, 0) 模型参数结果

	Coef.	Std.Err.	t	P> t	[0.025	0.975]
const	467.353	67.0247	6.9728	0.0001	335.987	598.7191
ar.L1.D.data	-0.5704	0.3233	-1.7645	0.1157	-1.204	0.0632

	Real	Imaginary	Modulus	Frequency
AR. 1	-1.7532	0	1.7532	0.5

进行残差序列自相关分析，采用 DW 检验方法，其结果靠近 2，为正常；靠近 0，呈现正的自相关关系；靠近 4，则呈现负的自相关关系。该模型的 DW 值为 1.716。最后绘制原数据和预测数据对比图：

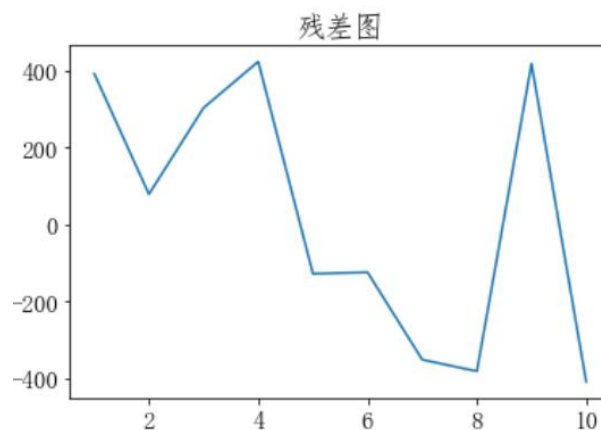
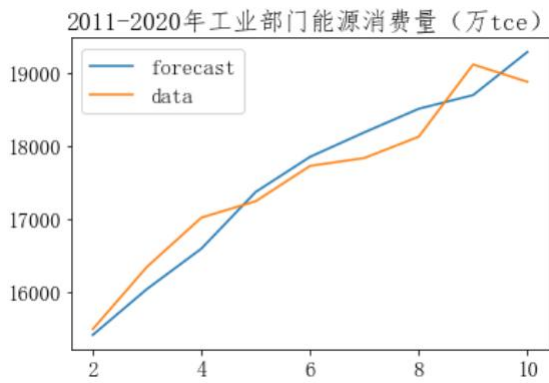
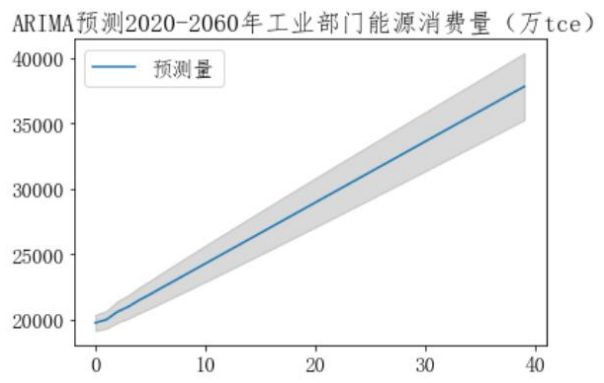


图 4.16 ARIMA 模型预测工业部门能源消耗量的残差结果图



(a)



(b)

图 4.17 a. 基于 ARIMA 模型的 2010-2020 年工业部门能源消耗量预测值与实际值对比
b. 基于 ARIMA 模型的 2020-2060 年工业部门能源消耗量预测结果

接下来利用 ARIMA 模型对其他指标进行预测，具体结果如表 4.5。

表 4.5 指标预测

序号	指标	ARIMA 模型	AIC 值	DW 检验值	2060 年预测值
1	农林部门能源消费量	ARIMA (1, 1, 0)	107.6	2.05	728.55 万 tce
2	能源供应部门能源消费量	ARIMA (5, 1, 0)	162.7	0.94	6122.41 万 tce
3	交通部门能源消费量	ARIMA (1, 1, 0)	110.7	1.74	6957.48 万 tce
5	建筑部门能源消费量	ARIMA (1, 1, 0)	106.8	1.25	2905.86 万 tce
6	居民生活消费量	ARIMA (1, 1, 0)	114.3	1.96	6318.16 万 tce
7	农林消费部门非化石能源占比	ARIMA (1, 1, 0)	38.6	2.36	76.62%
8	能源供应部门非化石能源占比	ARIMA (2, 1, 0)	32.9	1.42	46.52%
9	工业消费部门非化石能源占比	ARIMA (1, 1, 0)	34.4	2.01	60.17%
10	交通消费部门非化石能源占比	ARIMA (1, 1, 0)	-0.8	1.63	12.65%
11	建筑消费部门非化石能源占比	ARIMA (5, 1, 2)	41.1	1.3	100%
12	居民生活消费非化石能源占比	ARIMA (5, 1, 2)	47.1	2.65	56.72%

(1) 农林消费部门

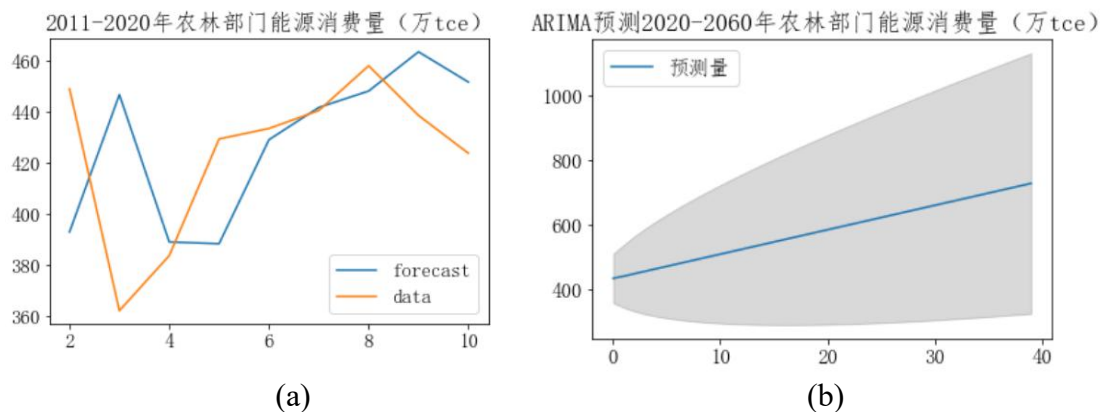


图 4.18 a. 基于 ARIMA 模型的 2010-2020 年农林部门能源消耗量预测值与实际值对比
b. 基于 ARIMA 模型的 2020-2060 年农林部门能源消耗量预测结果

(2) 能源供应部门

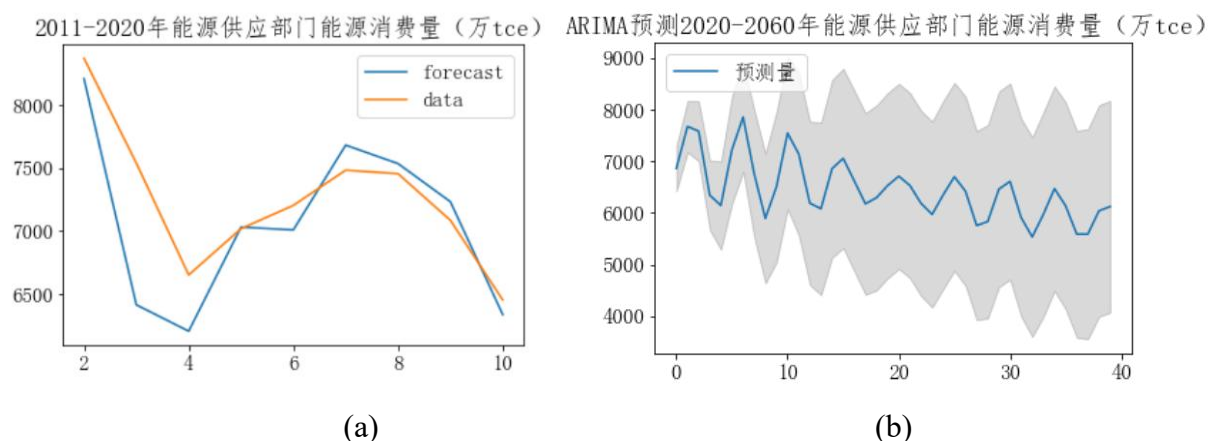


图 4.19 a. 基于 ARIMA 模型的 2010-2020 年能源供应部门能源消耗量预测值与实际值对比
b. 基于 ARIMA 模型的 2020-2060 年能源供应部门能源消耗量预测结果

(3) 交通消费部门

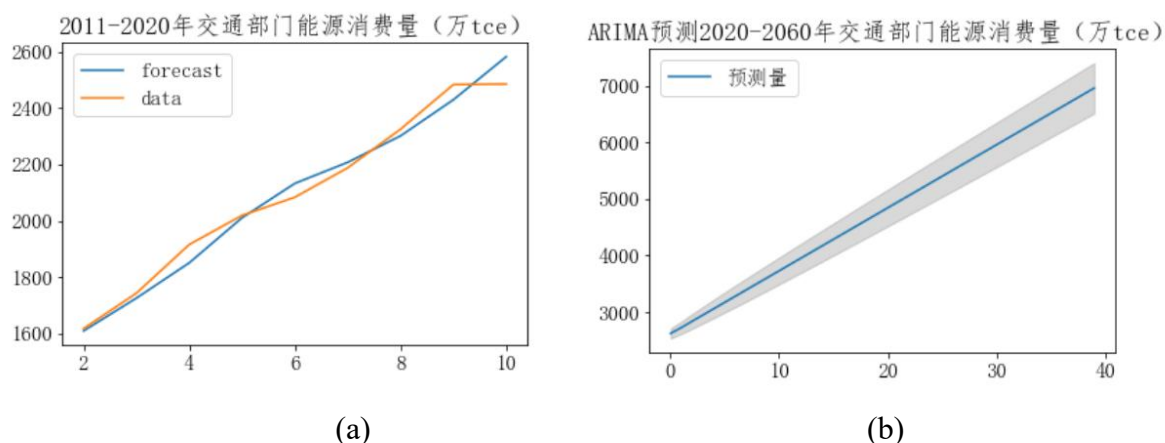


图 4.20 a. 基于 ARIMA 模型的 2010-2020 年交通部门能源消耗量预测值与实际值对比
b. 基于 ARIMA 模型的 2020-2060 年交通部门能源消耗量预测结果

(4) 建筑消费部门

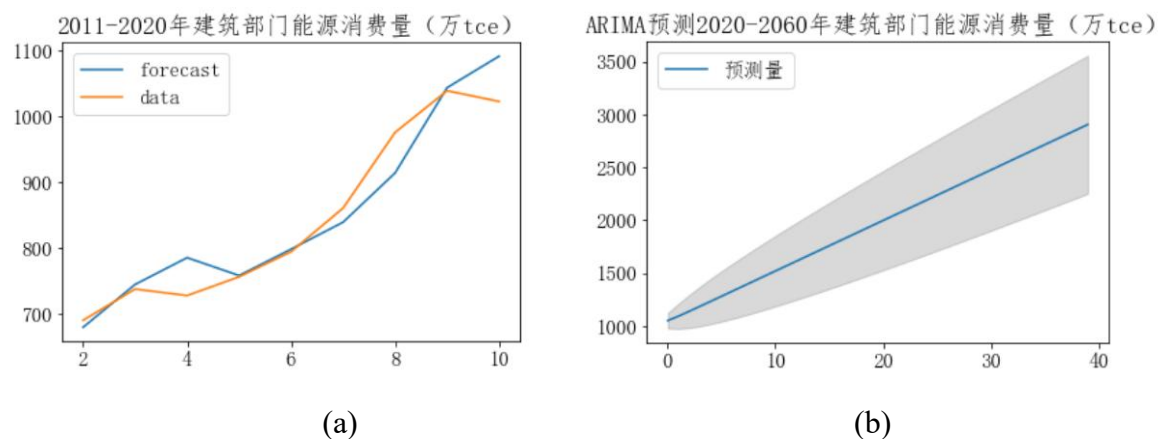


图 4.21 a. 基于 ARIMA 模型的 2010-2020 年建筑部门能源消耗量预测值与实际值对比
b. 基于 ARIMA 模型的 2020-2060 年建筑部门能源消耗量预测结果

(5) 居民生活能源消费量

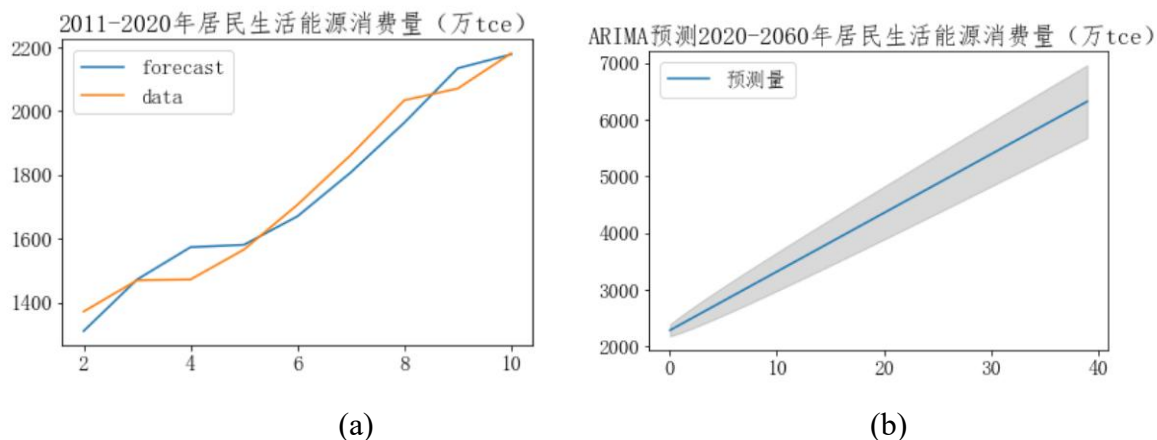


图 4.22 a. 基于 ARIMA 模型的 2010-2020 年居民生活能源消耗量预测值与实际值对比
b. 基于 ARIMA 模型的 2020-2060 年居民生活能源消耗量预测结果

(6) 农林消费部门非化石能源百分比

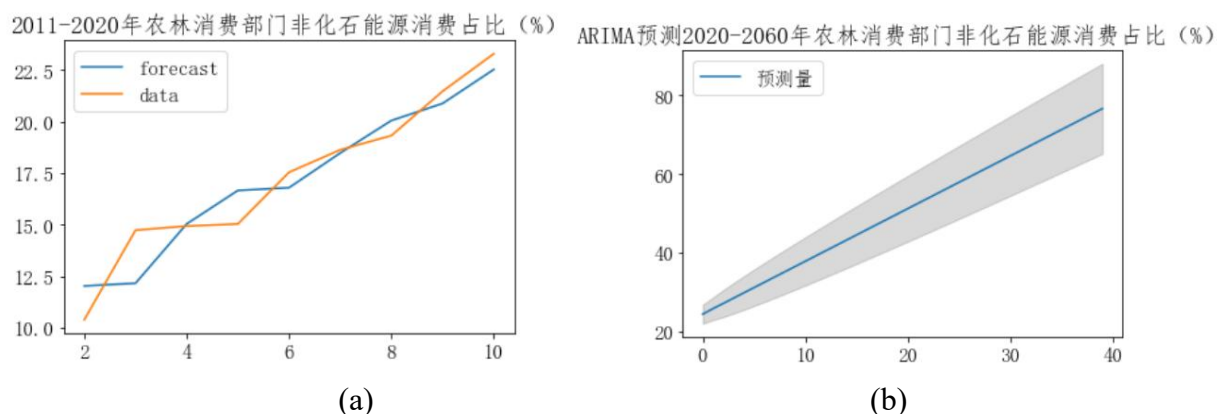


图 4.23 a. 基于 ARIMA 模型的 2010-2020 年农林消费部门非化石能源消费占比预测值与实际值对比
b. 基于 ARIMA 模型的 2020-2060 年农林消费部门非化石能源消费占比预测结果

(7) 能源供应部门

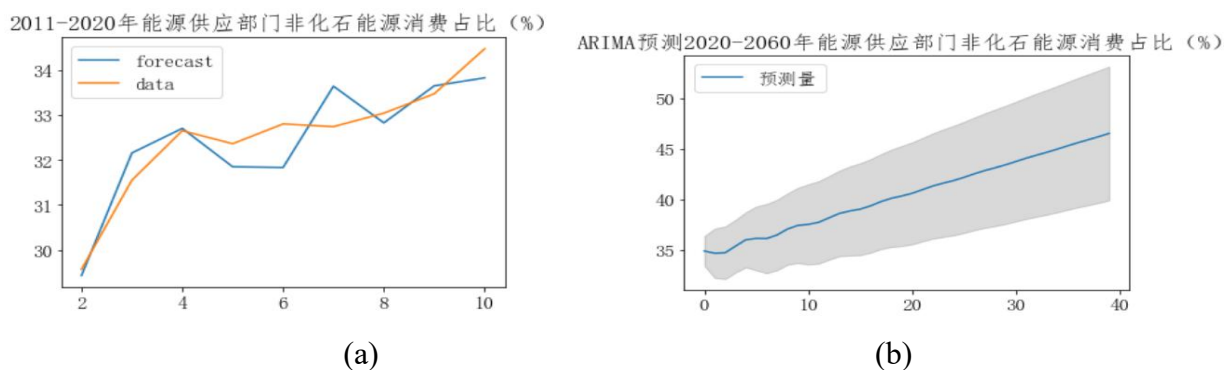
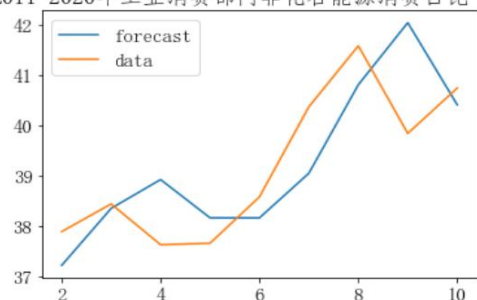


图 4.24 a. 基于 ARIMA 模型的 2010-2020 年能源供应部门非化石能源消费占比预测值与实际值对比
b. 基于 ARIMA 模型的 2020-2060 年能源供应部门非化石能源消费占比预测结果

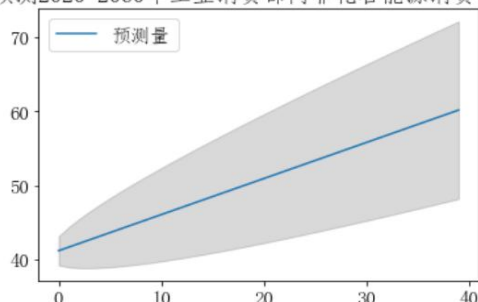
(8) 工业消费部门非化石能源百分比

2011-2020年工业消费部门非化石能源消费占比 (%)



(a)

ARIMA预测2020-2060年工业消费部门非化石能源消费占比 (%)



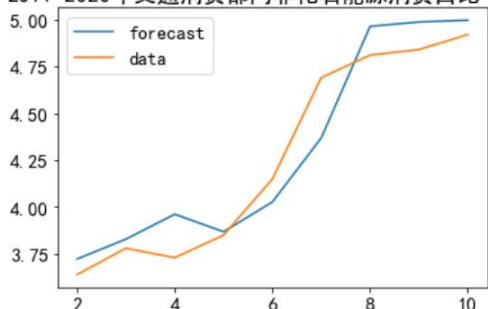
(b)

图 4.25 a. 基于 ARIMA 模型的 2010-2020 年工业消费部门非化石能源消费占比预测值与实际值对比

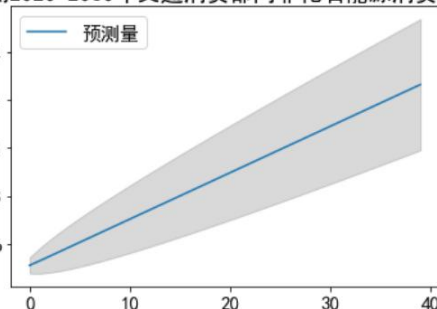
b. 基于 ARIMA 模型的 2020-2060 年工业消费部门非化石能源消费占比预测结果

(9) 交通消费部门非化石能源百分比

2011-2020年交通消费部门非化石能源消费占比 (%) ARIMA预测2020-2060年交通消费部门非化石能源消费占比 (%)



(a)



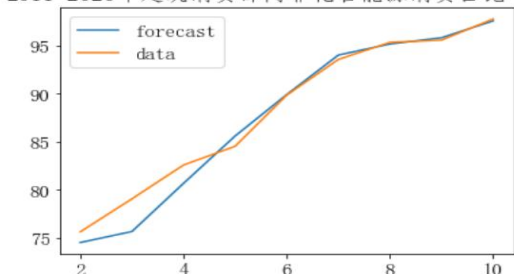
(b)

图 4.26 a. 基于 ARIMA 模型的 2010-2020 年交通消费部门非化石能源消费占比预测值与实际值对比

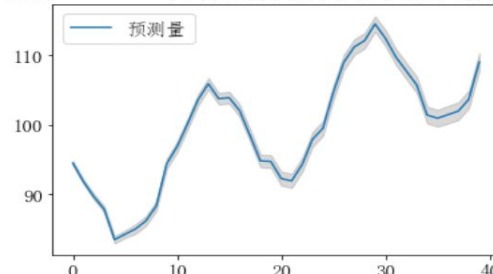
b. 基于 ARIMA 模型的 2020-2060 年交通消费部门非化石能源消费占比预测结果

(10) 建筑消费部门 (调整预测值超过 100 的设置 100)

2011-2020年建筑消费部门非化石能源消费占比 (%) ARIMA预测2020-2060年建筑消费部门非化石能源消费占比 (%)



(a)



(b)

图 4.27 a. 基于 ARIMA 模型的 2010-2020 年建筑消费部门非化石能源消费占比预测值与实际值对比

b. 基于 ARIMA 模型的 2020-2060 年建筑消费部门非化石能源消费占比预测结果

(11) 居民生活非化石能源百分比

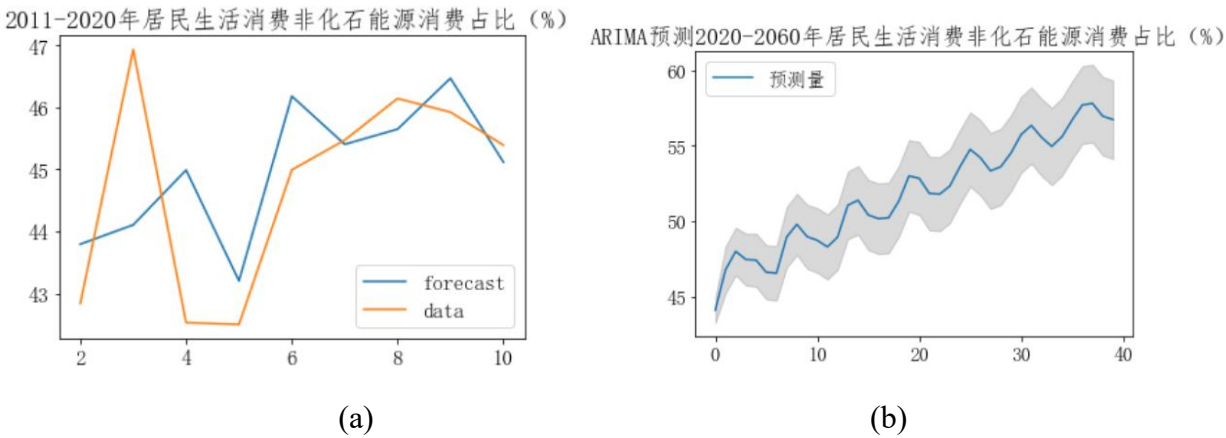


图 4.28 a. 基于 ARIMA 模型的 2010-2020 年居民生活消费非化石能源消费占比预测值与实际值对比
b. 基于 ARIMA 模型的 2020-2060 年居民生活消费非化石能源消费占比预测结果

表 4.6 ARIMA (5, 1, 2) 模拟居民生活消费非化石能源占比模型参数

	Coef.	Std. Err.	t	P> t	[0.025	0.975]
const	0.2975	0.014	21.2953	0.0022	0.2701	0.3248
ar.L1.D.data	-0.7736	0.3391	-2.2817	0.15	-1.4381	-0.1091
ar.L2.D.data	-1.1024	0.2003	-5.5047	0.0315	-1.4949	-0.7099
ar.L3.D.data	-0.8599	0.3511	-2.4493	0.134	-1.548	-0.1718
ar.L4.D.data	-1.0117	0.1844	-5.4873	0.0316	-1.3731	-0.6503
ar.L5.D.data	-0.5774	0.3196	-1.8068	0.2125	-1.2037	0.049
ma.L1.D.data	-1.7282	0.6446	-2.6812	0.1155	-2.9916	-0.4649
ma.L2.D.data	0.9973	0.632	1.578	0.2553	-0.2414	2.2361

	Real	Imaginary	Modulus	Frequency
AR. 1	0.4415	-0.9061	1.008	-0.1778
AR. 2	0.4415	0.9061	1.008	0.1778
AR. 3	-0.5505	-0.8989	1.0541	-0.3375
AR. 4	-0.5505	0.8989	1.0541	0.3375
AR. 5	-1.5342	0	1.5342	-0.5
MA. 1	0.8664	-0.502	1.0013	-0.0836
MA. 2	0.8664	0.502	1.0013	0.0836

通过 ARIMA 模型预测完上述 11 个指标后,即可开始对碳排放量的解释模型进行建模。首先尝试使用多元线性回归模型预测 2021-2060 年的碳排放量时序变化曲线。考虑到包括人口、GDP 在内的共 13 个自变量之间存在多重共线性,采用逐步回归分析,最后选取了第 5 个和第 12 个自变量,选择的是能源供应部门的能源消耗量和交通消费部门的非化石能源消费比重这两个时间序列数据对二氧化碳排放量进行解释。最后建立的碳排放量多元线性回归解释模型的 $R^2=0.949$, p 值=0.000006,证明模型精度达到要求。多元线性回归表达式如下:

碳排放量=24727+2.1027*能源供应部门的能源消耗量+6809.4*交通消费部门的非化石能源消费比重

由此可以看出能源供应部门的能源消耗量和交通部门的非化石能源消费比重都和碳排放量呈现正相关关系。

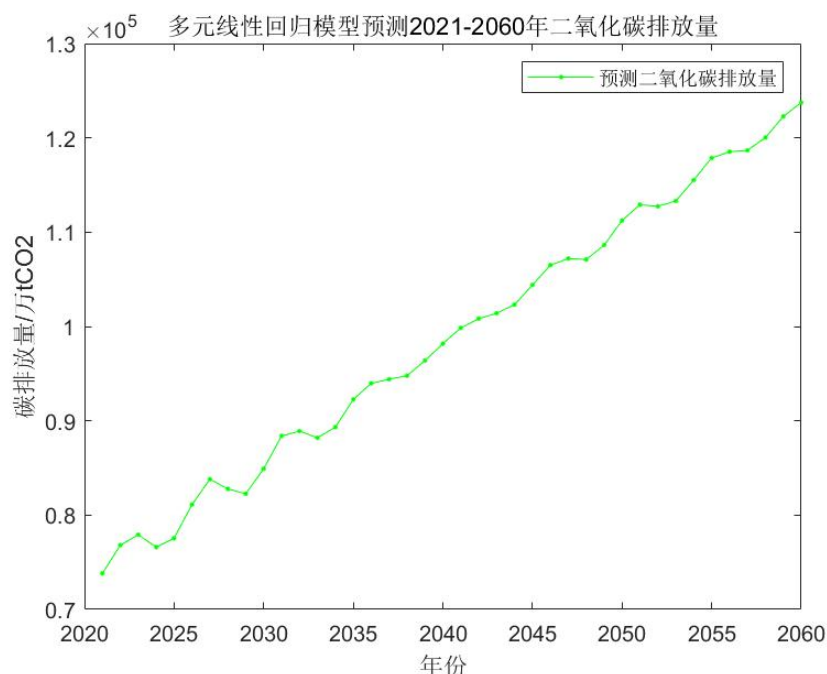


图 4.29 多元线性回归模型预测 2021-2060 年二氧化碳排放量

同时，我们考虑到从机理层面上解释二氧化碳排放量预测模型，并非与上述选取的两个自变量呈线性关系，我们选择了运用机器学习的方法对二氧化碳排放量预测模型进行建模分析。考虑到机器学习方法需要较大样本的输入数据，本实验选择了 2010-2050 年的人口、GDP、各部门的能源消费量以及各部门的非化石能源消费占比这 13 个指标的时间序列数据，以及通过多元线性回归模型得到的 2010-2050 年的二氧化碳排放量时间序列数据作为输入数据训练 BP 神经网络以及 LSTM 模型，最后以期预测 2051-2060 年的二氧化碳排放量。

在构建 BP 神经网络模型中，设置前 30 年为训练样本，后 10 年为测试样本。神经网络设置输入层节点 401 个，隐含层节点 10 个，输出层节点 1 个。该模型的平均绝对误差 MAE 为 1427.7738，均方误差 MSE 为 2882297.0653，均方根误差 RMSE 为 1697.7329。

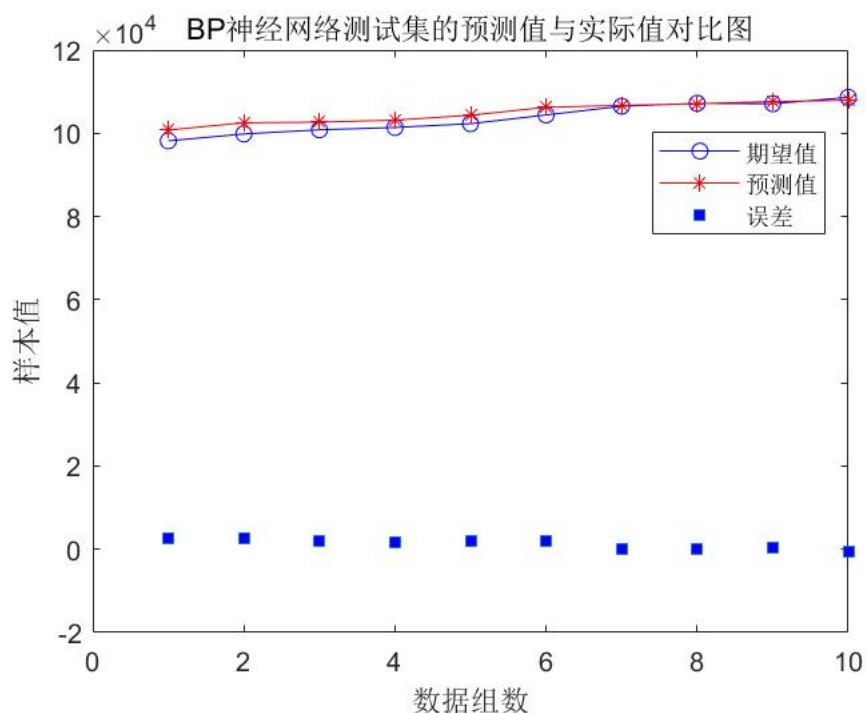


图 4.30 基于 BP 神经网络预测碳排放量测试集的预测值和实际值对比图

表 4.7 BP 神经网络预测 2050-2060 年区域碳排放量结果

年份	区域能耗量（万 tCO ₂ ）
2051	107200.1159
2052	106876.974
2053	107775.1611
2054	108200.1942
2055	107298.0287
2056	106487.826
2057	106634.9483
2058	106986.946
2059	106779.7495
2060	106275.3777

之后，我们运用了 LSTM 模型继续对 2010-2050 年的人口、GDP、各部门能耗量、非石油能源消费占比以及碳排放量进行建模分析，以预测 2051-2060 年的碳排放量时间序列变化。在 LSTM 模型参数设置上，numhidden_units1=160; numhidden_units2=180; numhidden_units3=100。训练 10 次，减少随机误差，最后一次训练过程如下图所示(图 4.31)。

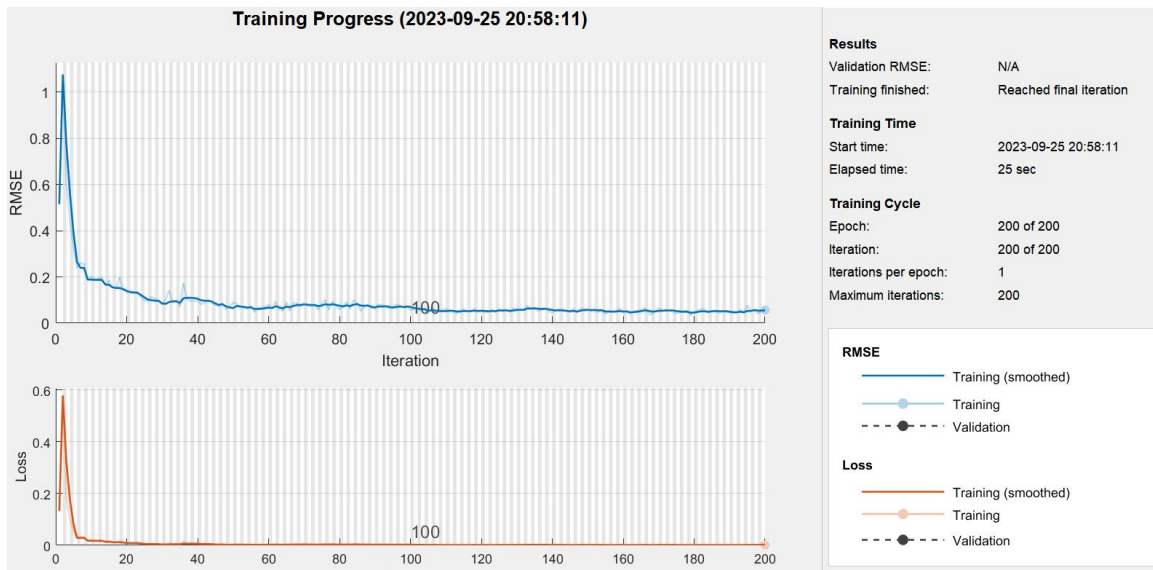


图 4.31 LSTM 训练碳排放量预测模型的过程

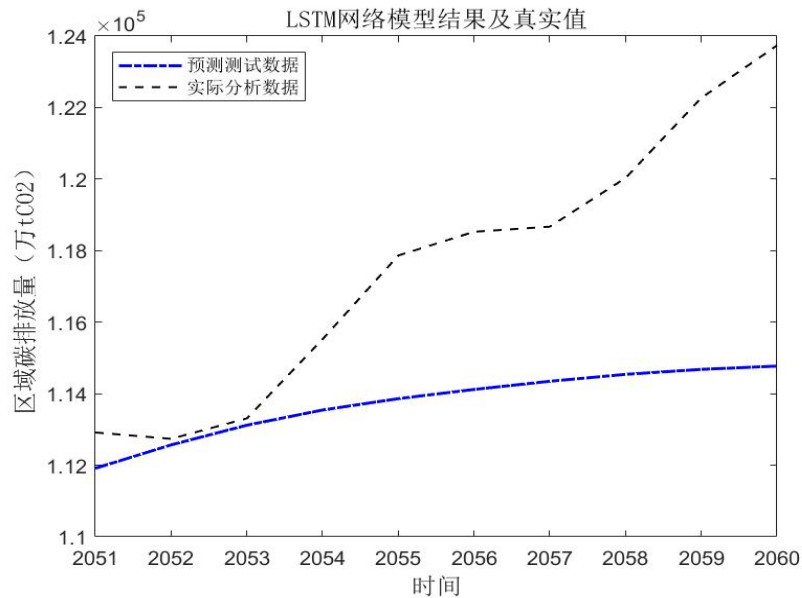


图 4.32 2051-2060 年基于 LSTM 网络的区域碳排放量预测结果

表 4.8 LSTM 预测 2050-2060 年区域碳排放量结果

年份	区域能耗量 (万 tCO ₂)
2051	102532.2308
2052	102771.7289
2053	101877.7307
2054	101716.0669
2055	102740.5691
2056	103620.5664
2057	103153.1398
2058	102290.1322
2059	102496.8005
2060	103636.5742

LSTM 网络模型的预测绝对平均误差 MAE 为 3812.1，平均绝对误差百分比 MAPE 为 0.0317，均方根误差 RMSE 为 4767.4。机器学习（BP 神经网络和 LSTM 模型）算法模拟的碳排放量到后期都呈现增长偏缓的态势，这样的预测结果相较于多元线性回归的线性上升态势更加符合实际情况以及区域双碳目标要求。

4.4 本章小结

通过问题 2 的设问，我们基于 2010-2020 年该区域的人口、GDP、能源消费量、各能源消费，供应部门的能源消费结构以及碳排放量等数据，通过逻辑斯蒂曲线、多项式拟合、ARIMA 模型预测各指标未来 40 年的变化趋势。在探讨经济增长与能源消费量以及能源消费量与碳排放量之间的关联关系时，引入多元线性回归模型、BP 神经网络、LSTM 模型对模型进行解释。

从结果上来看，（1）该区域的人口增长态势将放缓，逐渐趋于平稳。（2）GDP 总值会呈现线性上升的态势，说明该区域经济发展势头良好。（3）关于该区域的能源消费量，本实验利用人口以及经济数据对能源消费量进行回归，并预测未来的发展态势。对比多元线性回归模型和机器学习（BP 神经网络和 LSTM），我们发现在多元线性回归模型中，能源消费量呈线性上升态势；而在机器学习模型当中，能源消费量增长趋于缓慢。基于当下发展环境，我们认为机器学习模型对于能源消费量的回归预测更加合理。伴随着国家关于发展以及节能减排政策的出台，各行各业都在加快产业升级，降低能耗，努力实现经济高质量发展，绿色发展，所以未来社会的能源消耗量也将放缓增长态势。（4）在建立碳排放量预测模型中，我们也对比了多元线性回归模型和机器学习模型对于碳排放量的解释。在建立多元线性回归模型中，考虑到自变量存在的多重共线性，我们通过逐步回归分析，最后选择了能源供应部门的能源消耗量和交通消费部门的非化石能源消费比重两个因子对碳排放量进行回归预测。与多元线性回归模型预测的碳排放量呈线性上升态势不同，采用机器学习方法（BP 神经网络和 LSTM）预测得到的碳排放量增长态势放缓，最后趋于稳定。通过解读双碳政策目标以及了解当下产业形式，我们认为机器学习方法对于碳排放量的回归预测更加合理。从数据分析预测中，我们可以得知各行业的非化石能源消费比重（即降低单位能耗碳排放）都有所上升，这有助于缓解碳排放量的增长态势。

在本章中，我们系统地探讨了区域碳排放量以及经济、人口、能源消费量之间的相互作用关系，也分析了各能源消费，供应部门产业结构调整以及各部门的非化石能源消费比重对于区域碳排放量的影响。最后基于机器学习方法建立了该区域能源消费量以及碳排放量的预测模型，为有关部门宏观把控以及政策制定提供科学的理论依据，助力实现区域双碳目标。

5 问题三的建模与求解

5.1 问题分析

建立区域碳达峰和碳中和的目标与路径规划方法模型是一项综合性任务，需要综合考

考虑多个要素,包括地理信息、人口数据、经济因素等多方面因素。我们将在前两个问题的基础上进一步深入,从机理分析的角度出发,致力于构建区域碳达峰和碳中和的目标与路径规划方法模型。

“碳达峰”为CO₂排放量达到历史峰值,经历平台期后持续下降的过程,被视为排放量由增转降的历史拐点^[3]。“碳中和”通常被解释为人为源的CO₂排放量与人为汇的吸收量相平衡,实现“净零排放”,广义的“碳中和”也可涵盖全部温室气体^[4-5]。从碳达峰到碳中和的排放路径需涵盖爬坡至达峰、经平台期后持续下降至净零排放的完整过程。2020年9月22日,中国提出“CO₂排放力争于2030年前达到峰值,努力争取2060年前实现碳中和”,以上即为“双碳”目标^[6]。

在这一任务中,我们将充分利用地理信息、人口数据、经济因素等多源数据,结合相关科学理论和方法,通过深入研究碳排放的机理和影响因素,提出具体的目标设定和路径规划策略。本题分别建立模型,依据上述三种情景,利用预测模型和附件提供的数据,去构建碳达峰和碳中和模型中需要的指标,自主构建了一个自下而上的中国综合评估模型,模拟碳达峰和碳中和的全过程。

这些策略将为指导区域碳减排工作提供有力支持,推动可持续发展,减缓气候变化的进程,提高生态环境质量,同时促进经济增长和社会福祉的提升。通过科学建模和规划,我们能够更加有效地实现区域碳达峰和碳中和的双碳目标。

5.2 模型建立与求解

5.2.1 情景设计模型构建

本题构建了三种情景,无人为干预下的的自然情景、按时碳达峰与碳中和的基准情景、率先碳达峰与碳中和的进取情景。需要确定每个情景下的核心参数:经济增速、能效提升目标、非化石能源替代目标,并通过实验综合判断参数的合理取值范围,以达到各参数的最佳可行性路径。本题模型以2020为基准年,以2010-2020年碳排放量数据为基准,分别构建线性回归模型、ARIMA时间序列模型,分别构建三种情景,对于每个情景,设定关键参数,如碳达峰和碳中和的时间节点、GDP增长率、人口增长率、能源消费量增长率、能效提升率、非化石能源消费比重等。在此三种情景下对未来碳排放量以及能效提升和非化石能源消费比重进行预测。

1、无人为干预下的的自然情景

该情景构建方法利用向量自相关回归模型,预测在没有任何干预下,碳排放的自然趋势。向量自相关模型(Vector Autoregressive Model, VAR)是一种用于建模多个时间序列变量之间动态关系的统计模型。与单变量自回归模型(AR)不同,VAR模型允许同时考虑多个变量的相互关系。对于一个包含K个变量的VAR(p)模型,公式如下:

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + \varepsilon_t$$

其中: Y_t 是一个K维向量,表示在时间t的观察值, $Y_t = (Y_{1t}, Y_{2t}, \dots, Y_{Kt})$; c 是一个K维向量,表示截距项; A_1, A_2, \dots, A_p 是 $K \times K$ 维的系数矩阵,每个矩阵表示一个滞后期(lag)的影响, p 是滞后阶数; $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ 分别表示在过去时间点的 $t-1, t-2, \dots, t-p$ 的观察值; ε_t 是一个K维向量,表示在时间t的白噪声误差项,通常假设满足正态分布和零均值。

VAR模型的核心思想是,每个变量的当前值可以通过其自身和其他变量的滞后值的线性组合来预测。系数矩阵 A_1, A_2, \dots, A_p 描述了不同滞后期的影响关系。这种模型适用于多变量时间序列分析,用于研究变量之间的动态关系。

根据题目要求,使用 VAR 模型可以建立起能源利用效率与非化石能源消费比重之间的关系,并且利用 2010~2020 年间的数对 2021~2060 年的数据进行预测。预测结果如图 5.1。

对碳排放量预测结果进行差分,可以发现在 20 年以后的 2040 年左右,碳排放总量不再增长,并自此开始稳定负增长,达到碳达峰。而对于碳中和对于非化石能源比重大于 80% 的要求,在该无人干预情境下通过非化石能源比重在 2040 年至 2050 年达到最高 6%左右,之后趋于稳定。因此在该情境下将无法达到碳中和的目标。

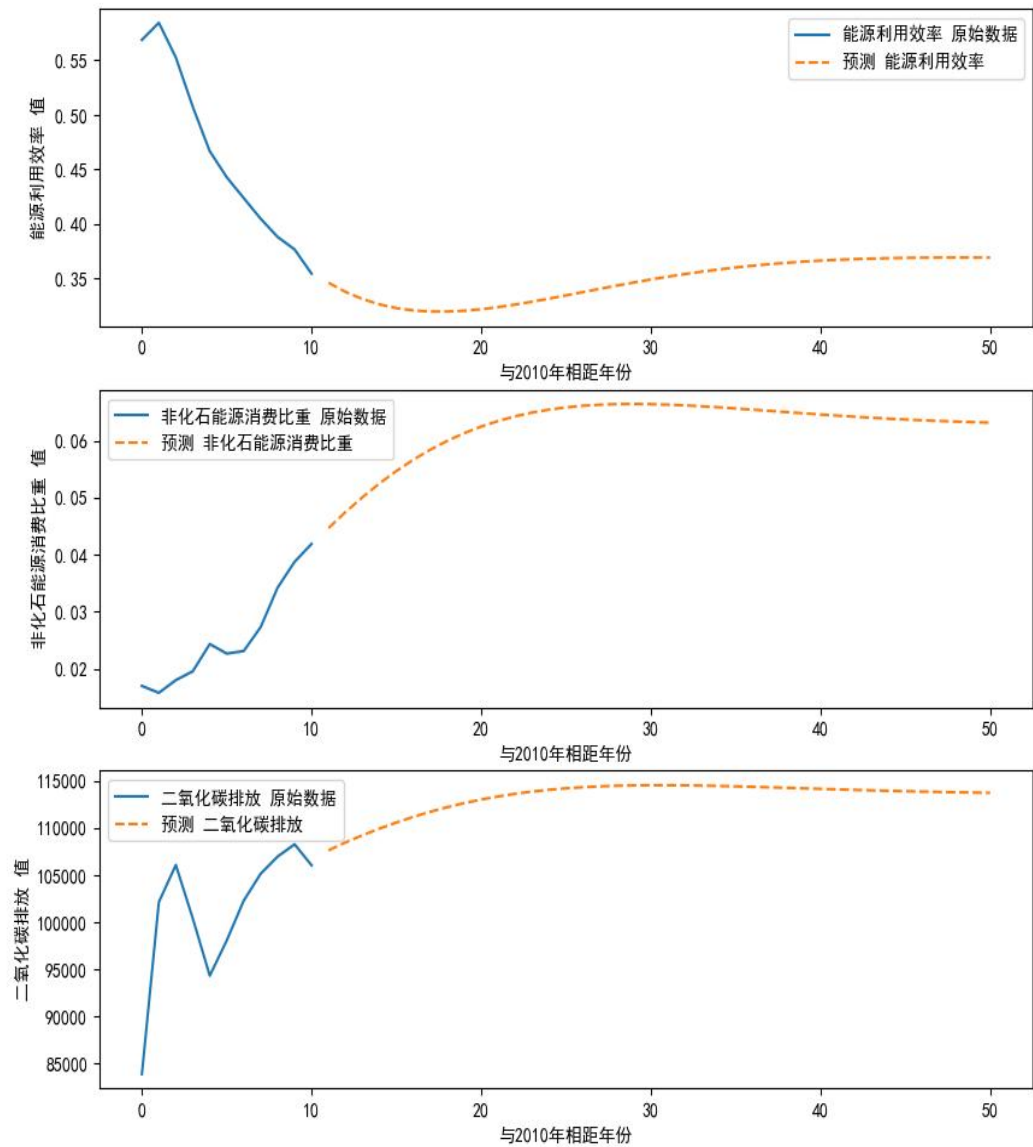


图 5.1 无人干预下的自然情景关联指标 50 年内预测结果

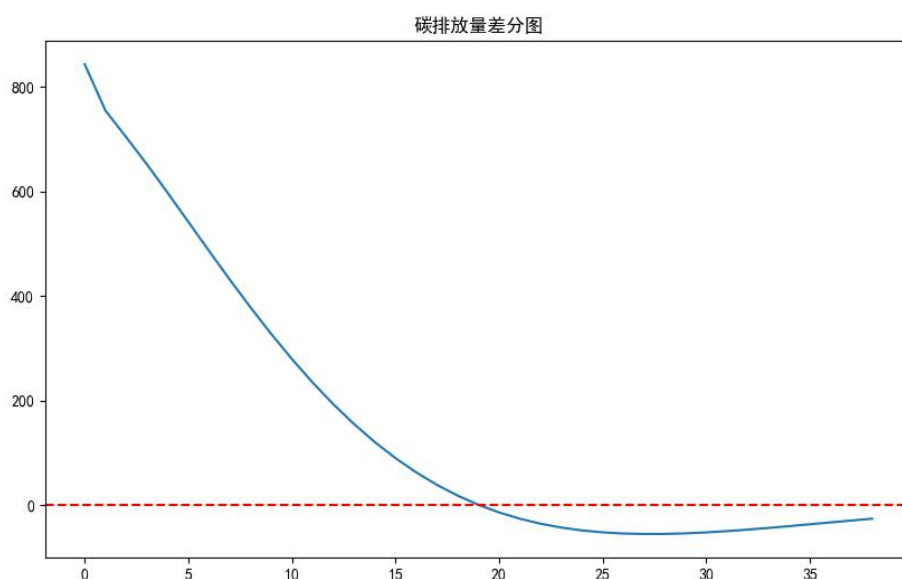


图 5.2 无人干预下的自然情景碳排放量差分图

2、按时碳达峰与碳中和的基准情景与率先碳达峰与碳中和的进取情景

依据政府设定目标，我们将在 2030 年实现碳达峰，也就是将二氧化碳排放量控制在最高点，之后将逐渐减少。到 2060 年，区域实现碳中和，意味着我们的碳排放将不再对全球气候产生负面影响。依据此目标，我们可以基于情景 1 增设一些人为干扰的可能参数。经过文献阅读与分析总结，可以认为以下几个人为干扰参数会显著影响碳排放以及非化石燃料比重：

政策强度：可以描述政府或政府机构在特定领域采取的政策举措的力度、严格程度或影响程度。政策强度与碳排放存在负相关，与非化石燃料比重存在正相关。

技术进步速度：代表在减少碳排放的过程中，新技术的研发和应用速度。与政策强度相关联，主要对降低非化石燃料比重存在积极影响。

能源结构调整速度：是指一个国家或地区在能源消费和生产方面向更清洁、可持续和低碳的方向进行转变的速度。转变包括提升可再生能源的比例、减少煤炭和化石燃料的使用等。主要对降低非化石燃料比重存在积极影响。

分别设定不同情境下三种归一化指标的参数值，结果如表 5.1。

表 5.1 不同情境下三种指标的参数值

	政策强度	技术进步速度	能源结构调整速度
无人干预下的自然情景	0	0	0
按时碳达峰与碳中和的基准情景	0.5	0.5	0.5
率先碳达峰与碳中和的进取情景	1	1	1

假设在按时碳达峰与碳中和的基准情景下，三种参数的大小均为最高值的一半。而在进取情景，三种指标均处于最高水平。

3.情景总结

对于三种不同情景实现碳达峰与碳中和的时间，我们可以进行如表 5.2 的推理与假设。

表 5.2 不同情境下三种情景达到碳达峰与碳中和的时间

	碳达峰	碳中和
无人干预下的自然情景	2040 年	-
按时碳达峰与碳中和的基准情景	2030 年	2060 年
率先碳达峰与碳中和的进取情景	2025 年	2040 年

注释 1：表中“-”代表在当前情境下无法达标。

5.2.2 多情景下碳排放量核算方法

在本问题中，分别根据不同的影响参数对三种场景的碳排放量进行预测。通过构建核算方法，可以帮助决策者和研究人员更好地理解不同因素对碳排放的影响，并为制定政策和规划可持续发展提供有力支持。

1.无人干预下的自然情景

对于无认为干预下的自然情景，在情景设计中已经利用向量自相关方法得到 2020~2060 年之间的预测结果，具体碳排放总量变化情况如图 5.3 所示。

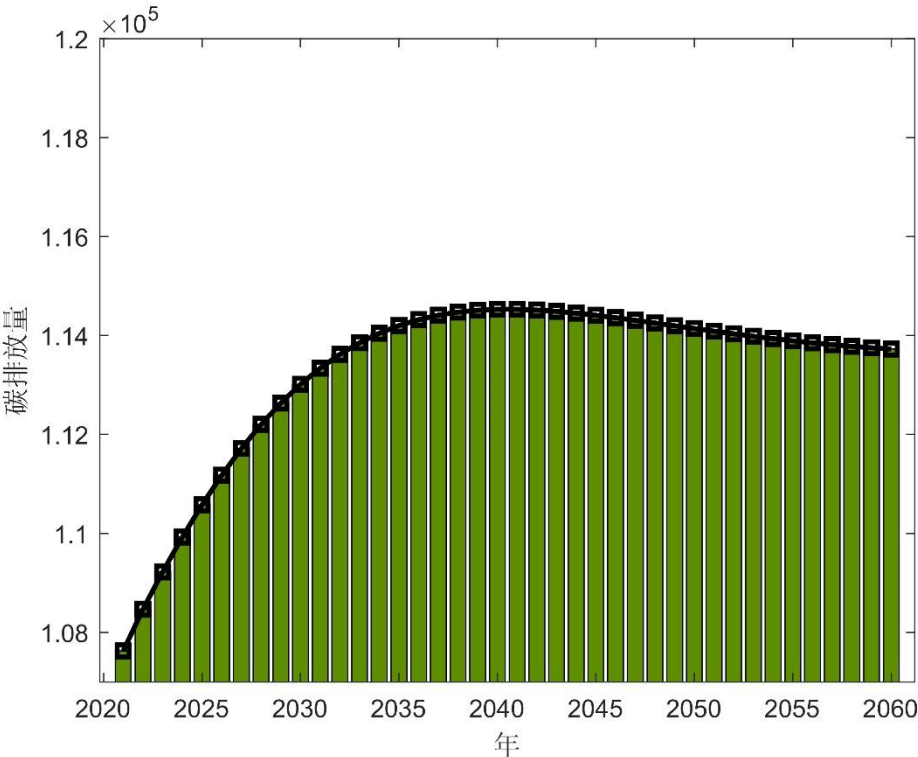


图 5.3 自然情景下 2021~2060 年碳排放预测

根据预测结果，可以得到在 2040 年达到碳达峰的前提下，其它年份的碳排放情况。

2、按时碳达峰与碳中和的基准情景与率先碳达峰与碳中和的进取情景

首先根据题目中假设与情景设计中的假设，可以判断该核算方法需要具有以下三点前提。

前提 1：2035 年的 GDP 比基期（2020 年）翻一番；2060 年比基期翻两番；

前提 2：人口增长率符合人口预测模型。

首先根据附表可以得到基期 2020 年的该区域 GDP，根据前提 1 的假设可以据此得到 2035 年和 2060 年 GDP。利用多项式拟合方法可以据此得到 2020 年~2060 年的每年 GDP 总量，并求得 GDP 每年的变化率，将结果在图 5.4 进行展示。

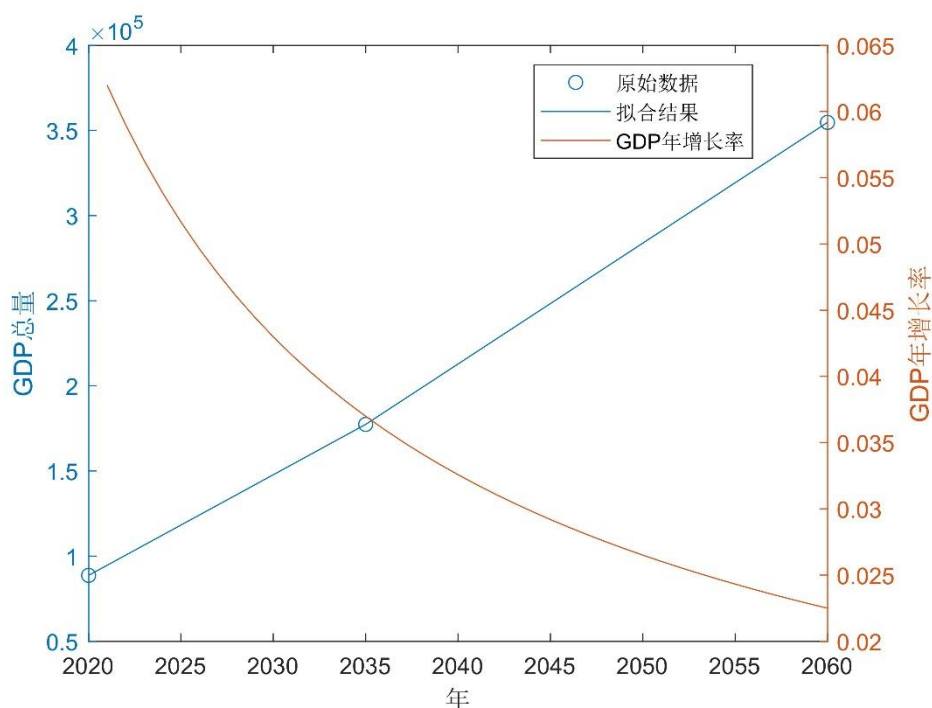


图 5.4 拟合得到的 GDP 总量与年增长率

根据问题二中使用的人口预测模型，可以得到人口在 2021~2060 年间的年变化率结果。将结果在图 5.5 中展示。

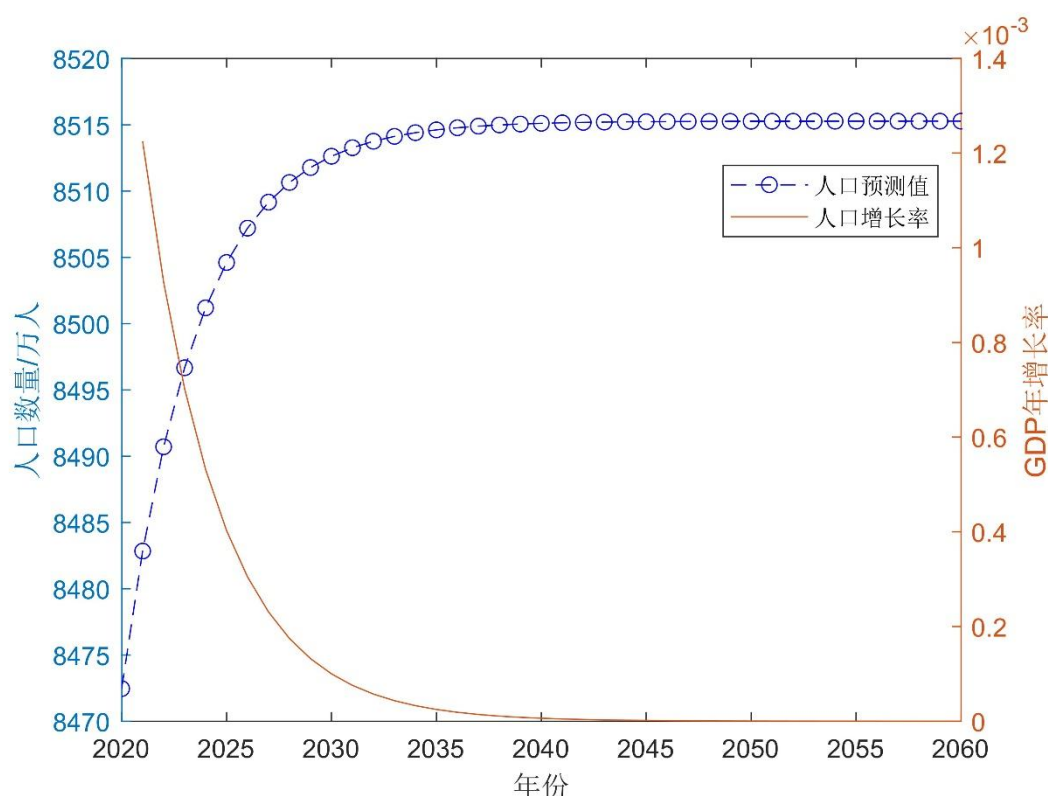


图 5.5 人口预测模型得到的人口与年增长率

Kaya 模型的原始形式为：

$$CO_2 = P * GDP/P * E/GDP * CO_2/E$$

式中， CO_2 是二氧化碳排放量； P 是人口； GDP/P 是人均 GDP； E/GDP 是单位 GDP 能耗； CO_2/E 是单位能耗二氧化碳排放量。采用两边对数并对时间进行一阶偏导即得到：

$$\delta CO_2 = \delta P + \delta GDP/P + \delta E/GDP + \delta CO_2/E$$

式中， δX 是参量 X 对某基准年的相对变化率。

当碳排放达峰时，碳排放增长率为 0，得到碳达峰情况 Kaya 模型：

$$\text{人口增长率} + \text{人均 GDP 增长率} = \text{单位 GDP 能耗下降率} + \text{单位能耗碳排放下降率}$$

即人口和经济增长与碳排放脱钩。因此在已知人口增长率和 GDP 增长率的情况下，碳排放核算问题可以通过以上公式与能效提升变化率和非化石能源占比变化率相关联。同时依据 kaya 模型，可以分别得到基准情景和进取情景下的单位 GDP 能耗增长率与单位能耗碳排放增长率之和。再利用历史数据分别获得在两种情景下的单位 GDP 能耗与单位能耗碳排放。

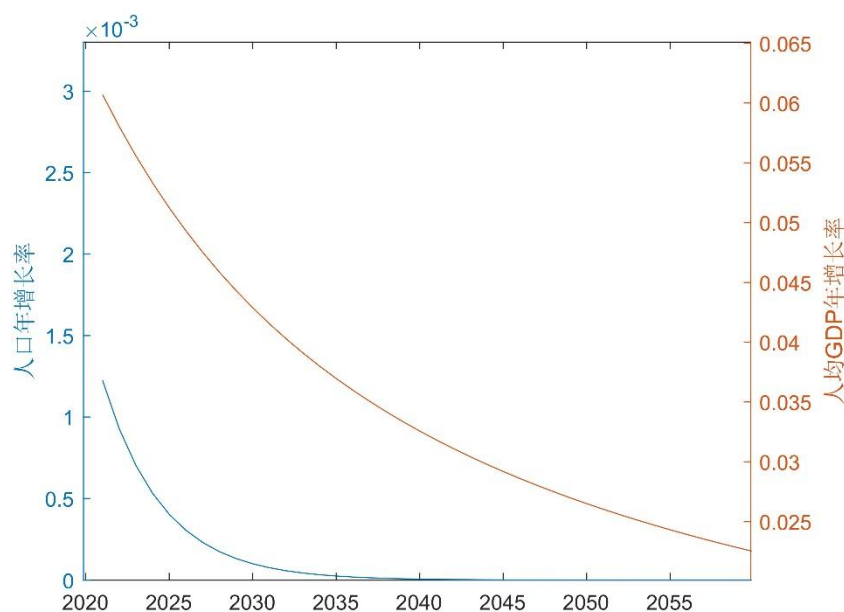


图 5.6 人均 GDP 年增长率和人口增长率

表 5.3 不同情景下人均 GDP 年增长率和人口增长率

达到碳达峰年份	人均 GDP 年增长率	人口增长率
基准情景 2030 年	4.29%	0.01%
进取情景 2025 年	5.12%	0.4%

根据表 5.3 中的数据，且碳达峰时碳排放增长率为 0 的条件，我们可以知道分别在这两种情景下单位 GDP 能耗下降率与单位能耗碳排放下降率之和分别为 4.30%和 5.52%。下面分别计算历史单位 GDP 能耗下降率与单位能耗碳排放下降率从而分配不同指标对于碳达峰情况的贡献值。

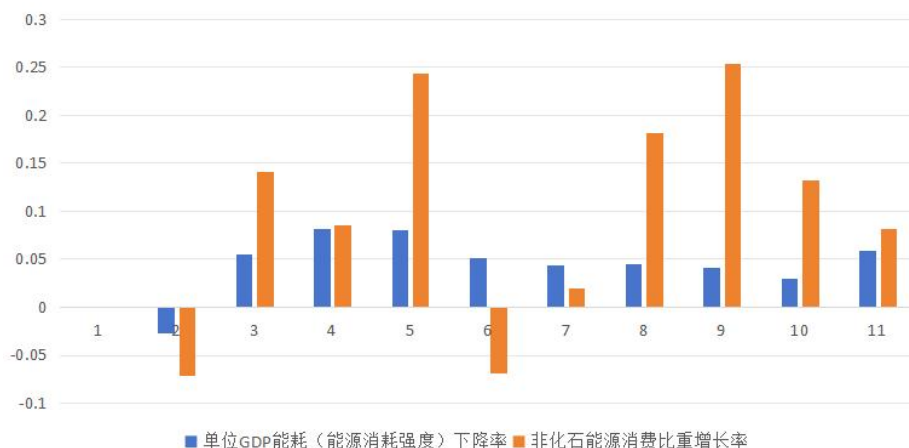


图 5.7 历史单位 GDP 能耗下降率与单位能耗碳排放下降率

通过图 5.7，我们可以看出这两个参数的历史变化情况。在 2016 年之前，除 2010 年外的单位 GDP 能耗下降率一直为正，说明单位 GDP 能耗一直在稳步下降。而非化石能源消费比重在 2011 和 2015 年都存在负增长，说明在十一五期间，我国的能源转型仍然处于探索阶段。直到 2016 年以后，两个参数的变化开始趋于稳定的正值。利用 ARIMA 模型对二

者在 2030 年和 2025 年的比值进行预测（图 5.8），结果分别为 0.45 和 0.49。据此可以分别得到 2030 年和 2025 年碳达峰的单位 GDP 能耗下降率与单位能耗碳排放下降率。结果如表 5.4。

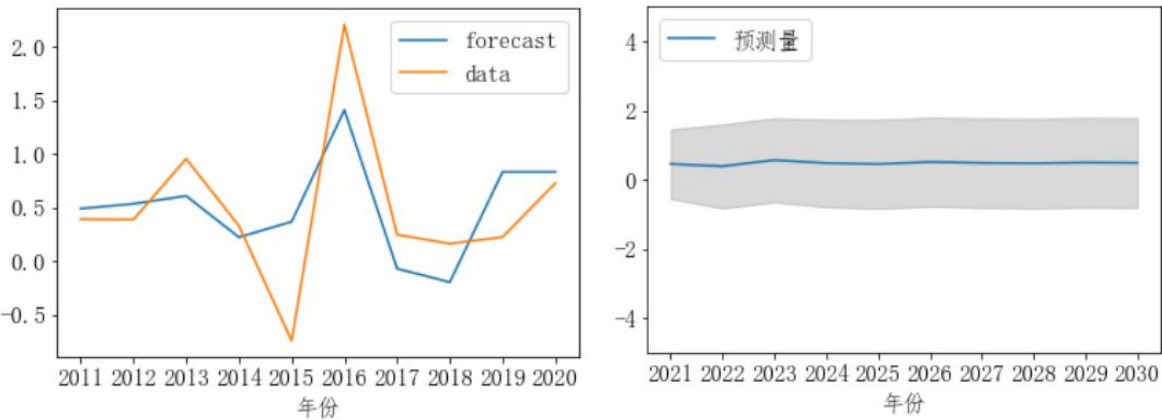


图 5.8 参数比值预测结果

表 5.4 单位 GDP 能耗下降率与单位能耗碳排放下降率

达到碳达峰年份	单位 GDP 能耗下降率	单位能耗碳排放下降率
基准情景 2030 年	2.96%	1.34%
进取情景 2025 年	3.71%	1.81%

根据表 5.6 中下降率与原始数据之间的线性与非线性拟合公式，可以分别得到这两种情境下的单位 GDP 能耗与单位能耗碳排放。结果如表 5.6。

表 5.5 原始数据

年份	单位 GDP 能耗（能源消耗强度）	非化石能源消费比重	单位 GDP 能耗（能源消耗强度）下降率	非化石能源消费比重增长率	比率比值
2010	0.569	0.017	-	-	-
2011	0.585	0.016	-2.76%	-7.08%	0.390
2012	0.553	0.018	5.45%	14.08%	0.387
2013	0.507	0.020	8.19%	8.58%	0.954
2014	0.467	0.024	8.02%	24.38%	0.329
2015	0.443	0.023	5.10%	-6.87%	-0.743
2016	0.424	0.023	4.32%	1.96%	2.204
2017	0.405	0.027	4.47%	18.13%	0.246
2018	0.388	0.034	4.13%	25.34%	0.163
2019	0.377	0.039	2.95%	13.19%	0.224
2020	0.354	0.042	5.89%	8.10%	0.727

表 5.6 单位 GDP 能耗与单位能耗碳排放

达到碳达峰年份	单位 GDP 能耗	单位能耗碳排放
基准情景 2030 年	0.454	0.0227
进取情景 2025 年	0.450	0.0220

5.2.3 双碳目标与路径确定

双碳目标与路径的确定在全球应对气候变化中扮演关键角色。这些目标涵盖了减排和碳吸收的双重使命，对于减缓气候变化、实现可持续发展至关重要。确定路径与目标不仅有助于减少温室气体排放，降低全球气温上升速度，还推动了绿色技术创新、国际合作和可持续经济增长。同时，双碳路径的确定需要综合考虑社会、经济和生态系统的可持续性，为全球社区提供了战略性指导，以实现更清洁、更健康、更繁荣的未来。

本问题首先要求对实现双碳目标的对碳排放产生影响的其他指标在关键时间节点的目标值进行确定，这些指标分别为 GDP 总量、人口、能源消费量、能源提升效率和非化石能源消费比重。基于前文的关联分析，可以确定 GDP、人口与能源消费量这三种指标与双碳政策的实施并不具有相关性，因此可以直接延续使用问题二中的 GDP、人口与能源消费量预测模型，首先对该三种指标的目标值进行确定。

GDP：在问题二中，通过对比指数增长、ARIMA 以及多项式拟合三种不同模型的拟合优度以及实用性，最终选定使用多项式拟合的方法来建立 GDP 的增长模型；

人口：在问题二中，使用逻辑斯蒂模型对 2010-2020 年区域人口总量进行回归分析，建立模型，以预测 2020-2060 年区域人口总量变化时间序列；

能源消耗量：在问题二中，通过对比多元回归与 BP 神经网络、LSTM 时序模型的拟合结果，最终选择建立 BP 神经网络模型进行预测。

各时间节点的目标值如表 5.7。从 2025-2060 年，GDP 在稳步增长，人口增长趋势归于平缓，与此同时能源消耗量增多，单位 GDP 能耗也在下降。

除了该三种指标外，还需要分别确定提高能源利用效率和提高非化石能源消费比重的目标值。该两种指标是与实现双碳政策息息相关的两个参数，因此将国务院颁布的政策数据作为辅助数据进行预测。

能源利用效率可以利用单位 GDP 能耗进行表示。根据国务院与 2021 年发布的碳达峰方案，预计到 2025 年非化石能源消费比重需要达到 20%左右，单位 GDP 能耗需要比 2020 年下降 13.5%，非化石能源消费比重达到 20%，为实现碳达峰奠定坚实基础。到 2030 年，非化石能源消费比重要达到 25%左右，2060 年非化石能源消费比重要达到碳中和要求中的 80%。根据这些已知目标数据与历史数据我们可以分别构建单位 GDP 能耗与非化石能源消费比重的变化模型。

通过对数据点分布情况的观察与分析，对于非化石能源比重的路径确定我们使用多项式曲线拟合；对于单位 GDP 碳排放我们采用指数拟合。最终得到的结果如图 5.9-10。综合以上分析将这五个指标的数据分别整理，结果如表 5.7。

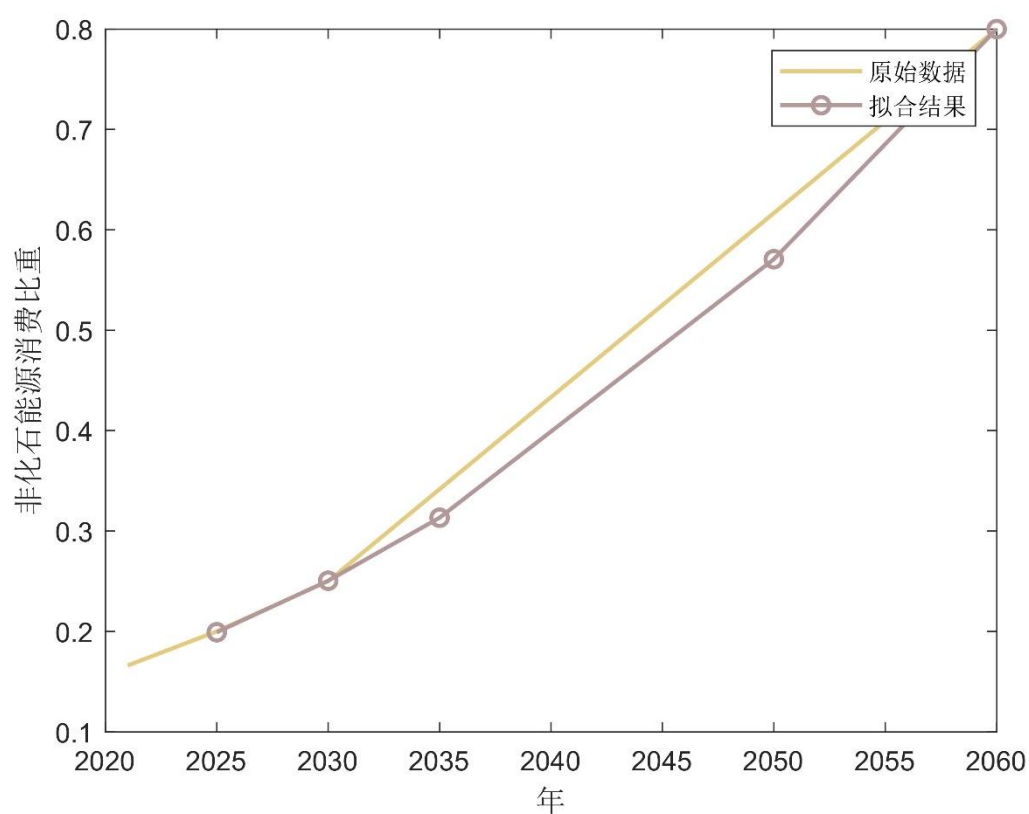


图 5.9 非化石能源消费比重拟合结果

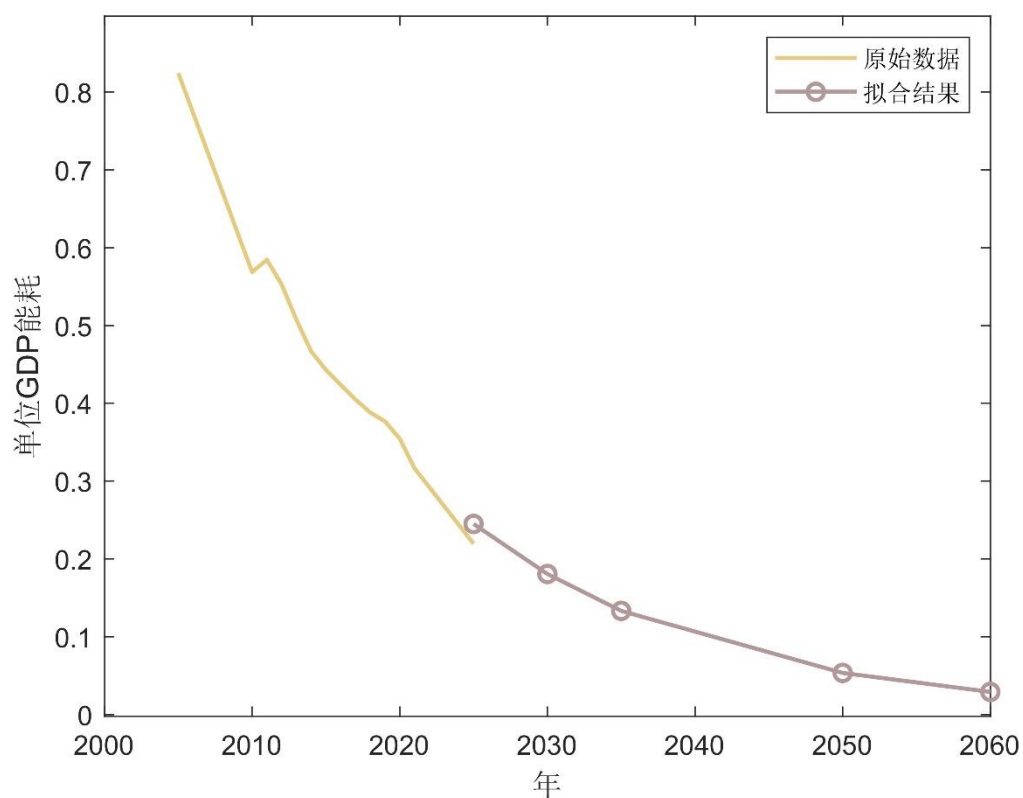


图 5.10 单位 GDP 能耗拟合结果

表 5.7 各时间节点的不同指标目标值

年份（年）	2025	2030	2035	2050	2060
GDP（亿元）	112789.8	135092.1	156568.6	216044	251565.7
人口（万）	8504.604	8512.617	8514.607	8515.254	8515.264
能源消耗量（万 tce）	33219.35	34358.02	35406.37	38268.69	38703.4
单位 GDP 能耗 （万 tCO ₂ /亿元）	0.2449	0.1806	0.1332	0.0534	0.0290
非化石能源消费比重	0.1992	0.2504	0.3132	0.5707	0.8000

能效提升可以通过单位 GDP 碳排放进行表示。能效提升越快，单位 GDP 碳排放下降速度越快。能效提升是实现双碳目标的重中之重。企业必须从降低产品残次率、提升企业效率、提升产品质量、降低单位能耗才能实现综合能效的提升。根据国际能源署分析，要将全球平均气温上升控制在 2 摄氏度以内，到 2050 年前，全球需要削减与能源有关的二氧化碳排放量，降低幅度为 40%至 70%。如果全球的温室气体排放从当前的 330 亿吨降至 2050 年左右的 100 亿吨，那么在此过程中，各项措施对全球减少二氧化碳排放的贡献可分为如下几部分：能源节约和提高能源效率贡献了 37%，可再生能源发展贡献了 32%，燃料替代贡献了 8%，核电发展贡献了 3%，二氧化碳捕获利用与封存（CCUS）贡献了 9%，而其余 12%的贡献来自其他技术。通过这些数据，我们可以节能增效是减少碳排放的第一手段。

产业升级是指通过引入新的技术、改进生产方法、提高产品质量、增加附加值等手段，使原有产业结构向更高端、更高附加值、更具竞争力的方向转变的过程，通常伴随着经济结构的优化和转型，旨在提高国家或地区的产业竞争力、创新能力和可持续性，同时减少资源浪费和环境污染，因此部分转型过程会对 GDP 和非化石能源消费比重等参数产生积极影响。但在某些情况下，产业升级可能导致短期内的碳排放增加。例如，新的生产线或设备的建设可能需要大量能源，导致短期内的碳排放上升。然而，这种情况通常是暂时的，随着新技术的投入使用和优化，碳排放往往会逐渐减少。总的来说，产业升级对碳排放的影响取决于多个因素的综合作用。在追求可持续发展的过程中，产业升级通常应当朝着减少碳排放的方向努力，但需要在减排和经济增长之间寻找平衡，以确保经济的可持续性。政府、企业和社会各界需要共同合作，制定和执行可持续发展战略，以最大程度地减少碳排放并应对气候变化。

能源脱碳的方式包括可再生能源投资、碳定价政策和碳捕获技术等，不同的方式会对碳排放量造成直接影响。在 2017 年刘强等人[7]对中国深度脱碳路径的研究中，提出了电力、工业、建筑、交通四个重点行业领域到 2050 年的最大碳减排潜力，并提出了这些行业和领域的深度脱碳路径。通过该脱碳路径，我国一次能源需要会在 2040 年达到峰值 62.5 亿 tce，然后不断下降到 2050 年的 40.1 亿 tce。

提高非化石能源消费比重的必要条件是提高非化石能源发电占比，需要由新能源发电、脱碳火电和包含储能在内的新型电网等能源生产部门完成，因此能源消费电气化也是实现双碳目标的重要方法之一。中电联党委书记、常务副理事长杨昆曾经指出“电气化水平是现代文明进步的重要标志，是实现‘双碳’目标的必然选择。”我国电气化发展虽然在人均用电量和电能占终端能源消费比重已经位居世界前列，但在非化石能源发电量、人均年生活用电量、单位发电量二氧化碳排放强度等方面仍然具有很大提升空间。

5.3 结果定性定量分析

最终确定双碳目标和路径需要政府、企业、研究机构和公众的合作，以制定全面、可行的计划，实现减排和碳吸收的目标，应对气候变化并推动可持续发展。这是一项复杂的任务，需要不断的监测和更新，以确保目标的实现。基于这些观察，我们得出以下定性和定量的结论和建议：

1、相较于主要发达经济体，中国的碳达峰和碳中和目标表现出了雄心勃勃的气候承诺，但也意味着中国需要付出更为艰巨的努力。回顾主要发达经济体的碳达峰特征，我们可以发现，它们通常经历了一个相对较长的爬坡期和平台期，这种现象在排放一直上升直至达到峰值的过程中相对常见。此外，从碳达峰到碳中和，它们通常需要 36 到 77 年的时间来完成这一转型。相对而言，中国在碳达峰过程中也将经历一段平台期，但与主要发达经济体相比，中国在碳达峰后到碳中和之间的时间约束更为紧迫，需要在大约 30 年的时间内快速实现深度的碳减排。

从公平的角度看，主要发达经济体在碳达峰时往往具有较高的历史累积碳排放量、人均碳排放量和人均国内生产总值（GDP）。此外，这些经济体在碳达峰时通常已经实现了经济增长与碳排放的脱钩，并成功实现了以更清洁的能源结构，以减少油气等高碳能源的使用。相比之下，中国需要在人均碳排放量、人均 GDP 以及能源消费结构等方面面临更大的挑战，以达到碳排放峰值。这意味着中国必须在绿色转型方面付出更大的努力，找到在应对气候变化的同时实现经济可持续发展的途径，并更加迅速地实现低碳化的生产和生活方式。这个过程将需要全社会的协同合作和创新，以应对气候挑战并为未来的可持续发展创造条件。

1、建议中国采用全面综合的政策工具来支持碳中和行动的有效实施。尽管有研究和评论指出，目前发达经济体发布的战略和政策可能还不足以确保实现碳中和目标^[7]，但主要发达经济体已经制定了包括顶层设计、地方实施路线图以及企业和金融机构行动方案等多层次、全方位的政策来推动碳中和目标的实现。目前，中国已经颁布了双碳的“1+N”政策体系，作为一个全面综合的政策工具。《关于完整准确全面贯彻新发展理念做好碳达峰碳中和工作的意见》^[8]和《2030 年前碳达峰行动方案》^[9]明确了中国实现达峰的总体目标，并提出了相关的措施和实施路径。这个政策体系也是中国继续推动生态文明建设和绿色转型的重要政策基础，可以在未来进一步细化和完善。

2、建议推动能源结构调整，发展可再生能源，并确保不同资源禀赋的地区都能公平过渡向绿色能源。在全球范围内，发达经济体的经验表明，未来几十年，天然气将成为低碳能源的支柱和重要的过渡性替代能源。然而，在中国，仅依赖天然气无法实现对高比例煤炭的替代^[10]。因此，优化能源结构和大规模发展可再生能源是中国提前实现碳达峰的主要策略。这种能源结构的转型可能导致传统能源行业工人失业问题。为减轻这些地区的社会经济影响，建议参考发达国家的经验，动员财政、银行、社会基金等资源，为受影响最严重的地区提供有针对性的支持，包括改造矿区、帮助煤炭工人再就业，以及兴建经济适用房等措施。

为了实现中国碳达峰、碳中和目标，应认真落实“四个革命、一个合作”能源战略要求，采取更加有力的政策和措施，加快推进新技术普及、新业态创新、低碳技术部署。进一步完善经济激励政策和部门协同管理机制，依靠市场竞争促进可再生能源发展。^[11]

各发达经济体的碳中和路径均以能源转型为重点，采用了多元化的政策工具，并且注重低碳和负碳技术的革新。根据发达经济体的政策展望，在实现碳中和时，均难以将绝对排放量降为零，都需要通过碳移除手段进行抵消。通过对比分析，发现中国的碳达峰和碳中和目标是具有雄心的气候承诺，相较其他发达经济体需要付出更大努力。建议运用全面

综合的政策工具支撑碳中和目标的有效落实，加快中国的气候立法，在兼顾公正转型的同时推动能源结构调整，注重可再生能源和能效方面的新技术开发应用。^[12]

9 模型评价与推广

这个模型的数据收集范围较广，且数据涉及范围有限，所以数据存在一定的误差、不确定性，我们本着科学的原则进行了最优化的处理，以提高其准确性。本文中的模型也具有一定的参考价值。

本文中模型的优点：

- 1、灵活的使用回归的方式是问题简化；
- 2、忽略次要因素的影响，使问题进一步得简化；
- 3、模型具有一定广泛性，可以用于不同地区、不同时间的模型求解；
- 4、模型对于问题的多个因素进行了分析求解，考虑多个指标使得模型预测 结果更加精确，模型更加适用。

模型的缺点：

- 1、主要使用多元线性回归方法对不同指标间关系进行建模，限制了模型的精确度；
- 2、研究数据有限，主要因素较少，不能更准确的反映研究对象与各个变量之间的关系拟合以及各个因素之间的关系；
- 3、收集的数据受限制较少，不确定成分多，存在一定偏差。

本模型所使用的方法适用与受多因素控制的并具有一定变化规律的变量，可以推广到同类问题的研究。

10 参考文献

- [1]唐杰等, 基于 Kaya 模型的碳排放达峰实证研究[J].深圳社会科学, 第 5 卷第 3 期, 2022 年 5 月, 第 50-59 页.
- [2]L.Zhang,J.Huang,A.H.AKhan.An ARIMA-based method for short-term wind speed forecasting[J]. Solar Energy, 2016(139):195-204
- [3]胡鞍钢.中国实现 2030 年前碳达峰目标及主要途径[J]. 北京工业大学学报: 社会科学版, 2021, 21 (3): 1-15.
- [4]巢清尘.“碳达峰和碳中和”的科学内涵及我国的政策措施[J]. 环境与可持续发展, 2021, 46 (2): 14-19.
- [5]邓旭,谢俊,滕飞.何谓“碳中和”?[J]. 气候变化研究进展, 2021, 17 (1): 107-113.
- [6]习近平.在第七十五届联合国大会一般性辩论上的讲话[N]. 人民日报, 2020-09-23 (1).
- [7]高翔.应对气候变化不能“光说不练” [N]. 光明日报, 2021-11-03 (8).
- [8]国务院.中共中央国务院关于完整准确全面贯彻新发展理念做好碳达峰碳中和工作的意见[EB/OL]. 2021 [2022-03-17].
- [9]国务院.国务院关于印发 2030 年前碳达峰行动方案的通知[EB/OL]. 2021 [2022-03-17].
- [10]杜祥琬.能源革命: 为了可持续发展的未来[J]. 中国人口·资源与环境, 2014, 24 (7): 1-4.

- [11]余碧莹,赵光普,安润颖,陈景明,谭锦潇,李晓易.碳中和目标下中国碳排放路径研究[J].北京理工大学学报,2021.
- [12]樊星,李路,秦圆圆,高翔.主要发达经济体从碳达峰到碳中和的路径及启示[J].气候变化研究进展, 2023, 19(1): 102-115.

11 附录

11.1 问题一部分代码

matlab 代码计算各指标并多元线性回归建立关联模型，计算指标同比增长率

```
x_fit=x';
[b,bint,r,rint,stats]=regress(x_fit(:,6),[ones(11,1) x_fit(:,1:5)],0.05) % 95%的置信区间,
Y=x_fit(:,6);
Y_NiHe = b(1) + b(2) .* x(1,:) + b(3) .* x(2,:) + b(4) .* x(3,:) +b(5) .* x(4,:)+b(6) .* x(5,:);
x_fit_2=x';
[b_2,bint,r,rint,stats_2]=regress(x_fit(:,7),[ones(11,1) x_fit(:,1:6)],0.05); % 95%的置信区间,
Y_2=x_fit(:,7);
Y_NiHe_2 = b_2(1) + b_2(2) .* x(1,:) + b_2(3) .* x(2,:) + b_2(4) .* x(3,:) +b_2(5) .*
x(4,:)+b_2(6) .* x(5,:)+b_2(7) .* x(6,:);
function yoy_growth = calculateYoYGrowth(x)
yoy_growth=[];
for i=1:(length(x)-1)
% 计算同比增长率
yoy_growth =[yoy_growth ((x(:,i+1) - x(:,i)) ./ x(:,i)) * 100] ;
end
end
%% 相关性
data=readtable('F:\shumo\nenghao.xlsx','Sheet','Sheet','Range','D2:N5');
% data =transpose(data);
data1 = table2array(data);
% h = heatmap(xvalues,yvalues, rho, 'FontSize',10, 'FontName','宋体');
h = heatmap(xvalues,yvalues, data1);
h.Title = '碳排放量与人口、经济、能源的相关性分析';
h.XLabel = '年份';
h.YLabel = '能源、人口、经济与碳排放量数据';
colormap summer
%% arima
year = 2010:2020;
energy_consumption = readtable('F:\shumo\nenghao.xlsx','Sheet','Sheet','Range','D2:N2');
energy_consumption= table2array(energy_consumption);
% model
AR_Order =2;
MA_Order =0;
AR = arima(AR_Order, 0, MA_Order);
if size(energy_consumption, 2) > 1
    energy_consumption = energy_consumption'; % 如果是行向量，转换为列向量
end
```

```

enyc= estimate(AR,energy_consumption);
step = 40;
en = forecast(enyc,step,'Y0',energy_consumption);
if size(carbonEmissions, 2) > 1
    carbonEmissions = carbonEmissions'; % 如果是行向量，转换为列向量
end
tpyc= estimate(AR,carbonEmissions);
tp = forecast(tpyc,step,'Y0',carbonEmissions);
if size(nonFossilEnergy, 2) > 1
    nonFossilEnergy = nonFossilEnergy'; % 如果是行向量，转换为列向量
end
nonyc= estimate(AR,nonFossilEnergy);
non = forecast(nonyc,step,'Y0',nonFossilEnergy);
year1 = 2021:2060;
AR_Order =2;
MA_Order =0;
AR = arima(AR_Order, 0, MA_Order);
if size(datanew,2) > 1
    datanew = datanew'; % 如果是行向量，转换为列向量
end
enyc= estimate(AR,datanew);
step = 10;
result= forecast(enyc,step,'Y0',datanew);
figure;
plot(result,'b-', 'LineWidth', 2);

```

11.2 问题二代码

1、人口预测模型部分代码

```

%输出
pnum=zeros(n,1);
for i=0:1:n-1
    pnum(i+1)=xm/(1+(xm/x(1)-1)*exp(-r0*i));
end
year1=2010:2020;
%计算均方根误差
rmse = sqrt(mean((pnum'-x).^2))
fnum=zeros(n+40,1);
for i=0:1:n+39
    fnum(i+1)=xm/(1+(xm/x(1)-1)*exp(-r0*i));
end

```

2、GDP 增长模型部分代码

```

z=log(y);p=polyfit(x,z,1);x1=2010:2060;x2=2010:2020;

```

```

z1=polyval(p,x1);y1=exp(z1);
z2=polyval(p,x2);y2=exp(z2); %计算均方根误差
rmse = sqrt(mean((y2-y).^2));
plot(x1,y1,'b.-',x,y,'r*');
xlabel('年份')ylabel('GDP/亿元')legend('预测 GDP')
title('指数增长模型预测 2010-2060 年区域 GDP 总值')

```

3、多元线性回归模型部分代码（包括能耗量和碳排放量的预测）

```

%% 预测
Xnew=data1(12:end,1:2);Znew=b1(1)+b1(2)*Xnew(:,1)+b1(3)*Xnew(:,2)figure('color','w');
plot([2020:2060],Znew, 'g.-')legend('预测能源消费量')xlabel('年份')ylabel('能源消耗量/万 tce')
title('多元线性回归模型预测 2021-2060 年区域能源消费量')

```

4、BP 神经网络

```

%% 第五步 网络参数配置（训练次数，学习速率，训练目标最小误差等）
net.trainParam.epochs=1000; % 训练次数，这里设置为 1000 次
net.trainParam.lr=0.01; % 学习速率，这里设置为 0.01
net.trainParam.goal=0.00001; % 训练目标最小误差，这里设置为 0.00001
%% 第六步 BP 神经网络训练
net=train(net,inputn,outputn);%开始训练，其中 inputn,outputn 分别为输入输出样本
%% 第七步 测试样本归一化
inputn_test=mapminmax('apply',input_test,inputps); % 对样本数据进行归一化
%% 第八步 BP 神经网络预测
an=sim(net,inputn_test); %用训练好的模型进行仿真
%% 第九步 预测结果反归一化与误差计算
test_simu=mapminmax('reverse',an,outputps); %把仿真得到的数据还原为原始的数
量级
error=test_simu-output_test; %预测值和真实值的误差
%%第十步 真实值与预测值误差比较
figure('units','normalized','position',[0.119 0.2 0.38 0.5] , 'color','w')
title('BP 神经网络测试集的预测值与实际值对比图')
[c,l]=size(output_test);MAE1=sum(abs(error))/l;MSE1=error*error'/l;RMSE1=MSE1^(1/2);
predict_y = zeros(10,1); % 初始化 predict_y
pre_test=mapminmax('apply',new_X(:,:),'inputps');% 对预测数据进行归一化
for i = 1: 10
    result = sim(net, pre_test(:,i));
    predict_y(i) = result;
end

```

5、LSTM 部分代码

```

error_test=LSTMoutput_test'-output_test';
pererror_test=error_test./output_test';error=error_test';pererror=pererror_test';
avererror=sum(abs(error))/(ntest);averpererror=sum(abs(pererror))/(ntest);
RMSE = sqrt(mean((error).^2));

```

6、ARIMA（部分 python 代码）

```
df={"time":[2010,2011,2012,2013,2014,2015,2016,2017,2018,2019,2020],
    "data":[ 56360.052,65193.342,67502.613,66749.376,64853.276,66074.810,68526.125,70451.55
    7,71502.003,74096.331,72633.324]}
data = pd.DataFrame(df)
data["diff_1"] = data["data"].diff(1)  #一阶差分
data["diff_2"] = data["data"].diff(1)  #二阶差分
def diff(df,col):
    font = {"size":15,
            "family":"fangsong"}
    matplotlib.rc("font",**font)
    matplotlib.rcParams['axes.unicode_minus']=False

    df["diff_1"] = df[col].diff(1)  #一阶差分
    df["diff_2"] = df["diff_1"].diff(1)  #二阶差分

diff(data,"data")

#未差分平稳性检测（ADF 检验、单位根检验）
from statsmodels.tsa.stattools import adfuller as ADF
print(u'原始序列的 ADF 检验结果为：', ADF(data["data"],1))
#返回值依次为 adf、pvalue、usedlag、nobs、critical values、icbest、regresults、resstore p<0.05
时表示稳定
#一阶差分平稳性检测（ADF 检验、单位根检验）
from statsmodels.tsa.stattools import adfuller as ADF
print(u'一阶差分序列的 ADF 检验结果为：', ADF(data["diff_1"][1:],1))
#返回值依次为 adf、pvalue、usedlag、nobs、critical values、icbest、regresults、resstore p<0.05
时表示稳定

#模型构建
arima513 = ARIMA(data["data"],order=(1,1,0)).fit()
arima513.summary2()
```

11.3 问题三代码

python 代码向量自相关：

```
df = pd.DataFrame(data)
# 拟合 VAR 模型
model = VAR(df)
model_fit = model.fit()
# 定义预测期数
n_forecast = 40
# 进行预测
forecast = model_fit.forecast(model_fit.endog, steps=n_forecast)
```

```

# 打印预测结果
print("预测结果: ")
print(forecast)
# 绘制每个变量的预测结果
variables = df.columns
n_variables = len(variables)
fig, axes = plt.subplots(n_variables, 1, figsize=(10, 5 * n_variables))
for i, var in enumerate(variables):
    axes[i].plot(df[var], label=f'{var} 原始数据')
    axes[i].plot(range(len(df), len(df) + n_forecast), forecast[:, i], linestyle='--', label=f'预测 {var}')
    axes[i].set_xlabel('与 2010 年相距年份')
    axes[i].set_ylabel(f'{var} 值')
    # axes[i].set_title(f'{var} 预测结果')
    axes[i].legend()

```

matlab 代码指数与多项式拟合

```

y1=[0.8245 0.568804088 0.584515274 0.552686687 0.507431602 0.466712588 0.442909569
0.423797878 0.404871223 0.388148147 0.376682582 0.354497722 0.316497722 0.219497722];
x1=[2005 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2025];
% 使用 polyfit 进行指数拟合
coefficients = polyfit(x1, log(y1), 1);
x1_=[2025 2030 2035 2050 2060];
% 生成多项式模型
model = exp(polyval(coefficients, x1));
predictx=exp(polyval(coefficients,x1_));
y2=[0.166 0.2 0.25 0.8];
x2=[2021 2025 2030 2060];
% 使用 polyfit 进行拟合
coefficients = polyfit(x2, y2, 2);
x1_=[2025 2030 2035 2050 2060];
% 生成多项式模型
model = polyval(coefficients, x1);
predictx=polyval(coefficients,x1_);

```