



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 北京航空航天大学

参赛队号 23100060041

1.齐芷璇

队员姓名 2.张嘉仪

3.赵悦

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

题 目：区域碳排放量影响因素分析及双碳目标分析

摘 要：

随着工业化和城市化进程的加快，全球温室气体排放不断增加，导致全球气候变暖、海平面上升、极端天气事件增多等问题日益突出。作为全球最大的排放国之一，中国面临巨大的压力和责任。同时，中国经济快速发展和人口增长导致能源需求不断增加，进而带来碳排放量的持续上升。为了实现可持续发展，中国必须采取行动减少碳排放并转向低碳经济模式。为了应对这一系列问题，中国提出了双碳政策。本文将建立碳排放数学模型，并分析、评价和预测能效提升、产业（产品）升级、能源脱碳和能源消费电气化等重点工程对碳排放的影响。

对于问题一，对碳排放以及人口、经济、能源消费量的现状分析。首先，对于给出的指标构建指标评价体系。以经济、人口、能源消费量和碳排放量作为基本指标，从产业结构构成、能源消费结构等方面进行指标细化，探究各指标之间的关联关系。分析十二五、十三五期间各指标变化趋势及产业结构和能源消费结构构成比例，分析各因素对碳排放量的影响。引入 **Spearman 相关系数**，对碳排放与经济、人口、能源消费量之间进行**相关性分析**，建立**关联模型**。

对于问题二，建立区域碳排放量以及经济、人口、能源消费量的预测模型。首先，基于人口和经济变化建立能源消费量预测模型。对于基于人口的能源消费量预测，分别构建能源消费量与人口的**线性回归模型**进行预测，为了增加预测精度，引入 **LSTM**、**灰色预测模型**再次进行预测。根据预测的结果，建立**加权优化模型**，以精度最小为目标函数，利用 **PSO 算法**求出使优化模型最优的权重。对于基于经济的能源消费量预测，分别构建**线性回归模型**、**灰色预测模型**和 **ARIMA 模型**，实现预测精度的优化。对于区域碳排放量预测模型，构建碳排放量与人口、GDP 和能源消费量的**岭回归模型**进行预测。

对于问题三，设计了**自然、基本、雄心、滞后**四种情景。通过数据分析计算出**自然情景**下到达碳达峰的时间约为 **2040 年**。根据这一特点设计分析自然发展情景、滞后达到碳达峰碳中和情景、按时达到碳达峰碳中和情景、率先碳达峰与碳中和四个情景的拟合模型、计算出 **2025 年、2030 年、2035 年、2050 年和 2060 年**的 **GDP、人口和能源消费量的目标值**。同时根据历史数据计算出 **2010-2020 年非化石能源占比**，根据公式确定了**提高能源利用效率和提高非化石能源消费比重的目标值**（表 6-3）。最终对能效提升、产业（产品）升级、能源脱碳和能源消费电气化等措施进行分析，根据得出的结果规划各种情景下的路径，以实现双碳目标。

关键词： 双碳目标，路径规划，线性回归，关联关系，ARIMA 模型，灰色模型

目录

1. 问题重述	4
2. 模型假设	7
3. 符号说明	8
4. 问题一模型建立与求解.....	9
4.1 建立指标与指标体系.....	9
4.2 区域碳排放量以及经济、人口、能源消费量的现状分析	10
4.2.1 碳排放量状况分析.....	10
4.2.1.1 碳排放总量及变化趋势.....	10
4.2.1.2 单因素 ANOVA 分析碳排放量.....	12
4.2.2 碳排放量影响因素分析.....	15
4.2.3 碳达峰与碳中和面临的挑战.....	17
4.3 相关指标及其关联模型.....	18
4.3.1 相关指标的变化分析.....	18
4.3.1.1 GDP 变化分析.....	18
4.3.1.2 人口变化分析.....	20
4.3.1.3 能源消费量变化分析.....	21
4.3.1.4 各产业与居民生活能源消费量变化分析.....	23
4.3.1.5 各产业 GDP 变化分析.....	26
4.3.2 Spearman 相关性分析.....	30
4.3.3 Spearman 相关性分析结果.....	30
5. 问题二模型建立与求解.....	33
5.1 基于人口和经济变化的能源消费量预测模型.....	33
5.1.1 基于人口的能源消费量预测模型.....	33
5.1.1.1 线性回归预测模型.....	33
5.1.1.2 灰色预测模型.....	37
5.1.1.3 LSTM 预测模型.....	41
5.1.1.4 加权综合预测模型.....	43
5.1.2 基于经济的能源消费量预测模型.....	47
5.1.2.1 线性回归预测.....	47
5.1.2.2 灰色预测模型.....	51
5.1.2.3 ARIMA 预测模型.....	54
5.2 区域碳排放量预测模型.....	61
5.2.1 绘制岭迹图确定模型参数.....	61
5.2.2 岭回归模型的分析结果.....	62
5.2.3 加权预测模型的分析结果.....	63
6. 问题三模型建立与求解.....	64
6.1 情景设计	64
6.2 多情景下碳排放量核算方法.....	65
6.2.1 自然情景下（碳达峰与碳中和）目标与路径预测.....	65
6.2.2 多情景下碳排放量核算.....	67
6.3 双碳目标与路径规划.....	68

6.3.1 GDP、人口和能源消费量目标值.....	68
6.3.2 能源利用效率和提高非化石能源消费比重的目标值.....	69
6.3.3 能效提升、产业（产品）升级、能源脱碳和能源消费电气化的定性分析	71
7. 模型评价	73
7.1 模型的优势	73
7.2 模型的不足	73
7.3 模型的应用前景.....	73
8. 参考文献	75
9. 附录	77

1. 问题重述

1.1 问题背景

工业革命以来，在科技飞速发展的同时，越来越多的人类活动造成了碳排放的飞速增长，对生态系统中原有的碳循环造成了巨大的影响，一定程度上打破了环境中碳排放和碳吸收的平衡，碳排放增长速度远超碳吸收速度，导致环境中二氧化碳浓度急剧上升，引发了温室效应和全球变暖等现象。2020年9月22日，习近平主席在第七十五届联合国大会上提出中国将提高国家自主贡献力度，采取更加有力的政策和措施，二氧化碳排放力争于2030年前达到峰值，努力争取2060年前实现碳中和[1]。

化石能源是推动现代经济增长的重要生产要素，经济生产活动与碳排放活动密切相关。为实现中国民族的伟大复兴，提高国家国际竞争力，推动国家进一步发展少不了对经济的促进和对科技水平的提升。经济增长通常以消耗化石能源为代价，而化石能源消费通常会增加碳排放量，在能源消耗主要为化石能源的情况下，经济增长与能源消费量通常成正相关，能源消费量通常与碳排放量成正相关，而社会全面健康发展既追求经济的增长又追求低碳排放量，在这种情况下，如何处理经济增长与碳排放之间的矛盾关系是确定碳达峰、碳中和路径，解决碳排放过高问题主要难点，必须从提高能源利用效率和提高非化石能源消费比重两个方面入手(图 1-1)。

提高能源利用率（即降低单位 GDP 能耗）的重要途径是实现能效提升和推动产业（产品）升级。开展管理节能、技术节能和结构节能等能效工程，可以在生产过程中降低单位产品与服务的能耗[2]。同时，通过科技创新带来的技术对产业（产品）进行升级，可以增加单位产品与服务的科技附加值，两种方式均可以实现能源利用效率的提升。

提高非化石能源消费比重（即降低单位能耗碳排放）的主要途径是开展能源脱碳工程和能源消费电气化。能源脱碳工程旨在使用新能源发电、火电脱碳与新型电网等方式，提升能源消耗中非化石能源发电的占比；而能源消费电气化则是通过开展以电能代替化石能源的工程，提升非化石能源发电占比。

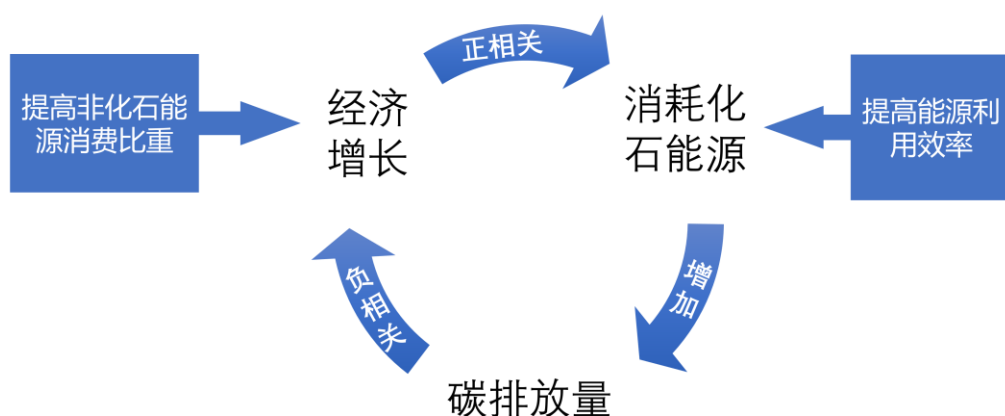


图 1-1 经济增长与碳排放关系图

由此可见，解决经济增长与碳排放的矛盾主要有提高能源利用率和提高非化石能源消费比重两种方式，而提高能源利用率可通过开展节能工程和实现产业升级实现，提高非化石能源消费比重可通过能源脱碳和能源消费电气化实现。

1.2 问题描述

本题目的命题宗旨是根据机理分析建立区域碳排放量数学模型，建立指标体系，并运用模型及题目给定数据分析、评价和预测能效提升、产业（产品）升级、能源脱碳和能源消费电气化等重点工程对碳排放的影响，进而确定双碳（碳达峰与碳中和）目标与路径。

根据上述背景，该题目提供了个附件数据，以及相关参考文献、公式等，拟解决的关键问题如下：

问题一：区域碳排放量以及经济、人口、能源消费量的现状分析

- (1) 建立指标与指标体系
- (2) 分析区域碳排放量以及经济、人口、能源消费量的现状
- (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

问题二：区域碳排放量以及经济、人口、能源消费量的预测模型

- (1) 基于人口和经济变化的能源消费量预测模型
- (2) 区域碳排放量预测模型

问题三：区域双碳（碳达峰与碳中和）目标与路径规划方法

- (1) 情景设计
- (2) 多情景下碳排放量核算方法
- (3) 确定双碳（碳达峰与碳中和）目标与路径

1.3 问题分析

1.3.1 数据集分析

对于题目给出的数据，首先进行数据预处理，将数据进行归类、缺失值处理、可视化处理。分析十二五、十三五期间各指标变化趋势及产业结构和能源消费结构构成比例，分析各因素对碳排放量的影响。

1.3.2 问题一分析

针对问题一，首先对于给出的指标构建指标评价体系。以经济、人口、能源消费量和碳排放量作为一级指标，根据所给数据和调研的内容确定二级指标，完成对各部门的碳排放情况、描述指标之间的相互关系。利用处理后的数据绘制折线图、柱状图、饼图、桑基图，比较分析十二五、十三五的变化趋势，经济增长、人口增长、能源消费结构等方向出发，分析影响因素。引入 Spearman 相关系数，对碳排放与经济、人口、能源消费量之间进行相关性分析，建立关联模型。

1.3.3 问题二分析

首先利用数据处理之后的数据进行处理，建立回归预测模型。对于要求 2、3，分别构建能源消费量与人口以及经济的线性回归模型进行预测，为了增加预测精度，引入 ARIMA、灰色预测模型再次进行预测。根据预测的结果，建立优化模型，以精度最小为目标函数，对三种预测方式进行加权，利用优化模型求解最优的权重。对于区域碳排放量预测模型，构建碳排放量与人口、GDP 和能源消费量的多元线性回归模型进行预测，再利用第一小问的预测模型，提高预测精度。

1.3.4 问题三分析

针对问题四，以 2030 年达到碳达峰，2060 年实现碳中和为确定时间节点，设计自然、基本、雄心、滞后四种不同的情景。分别为，没有人为干预、按时达到碳达峰碳中和、率

先碳达峰与碳中和、滞后达到碳中和情况下的碳减排措施。对于各部门的能源消费量、能源消费品种及其碳排放量的预测方法，选取与问题二相同的预测方法。确定 2025 年、2030 年、2035 年、2050 年和 2060 年的 GDP、人口和能源消费量目标。将目标与提高能源利用效率和提高非化石能源消费比重的目标相关联，确保达到碳达峰和碳中和目标的路径与措施一致。对能效提升、产业（产品）升级、能源脱碳和能源消费电气化等措施进行定性和定量分析，以确定如何实现双碳目标。

本文整体解题流程图如下：

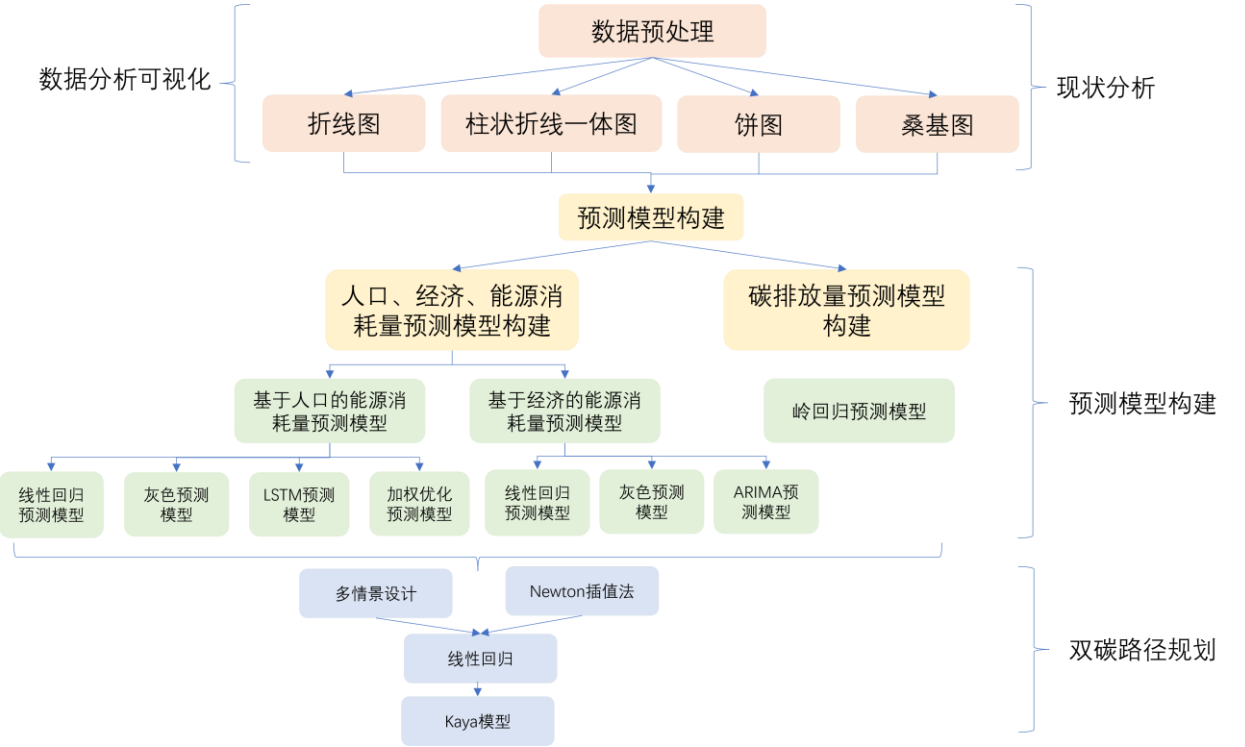


图 1-2 论文整体框架流程图

2. 模型假设

1. 本题目所用数据涉及区域假定位于中国东南沿海，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。因此为了更好的推动该地区的经济发展，亟待解决各产业的能源供给和利用问题。

2. 所分析的数据时间范围：2010-2020 年。其中 2011-2015 年为十二五时期，2016-2020 为十三五时期。

3. 本文将国民经济各行业分为第一产业、第二产业、第三产业及居民生活消费四部分。具体划分如图 2-1。

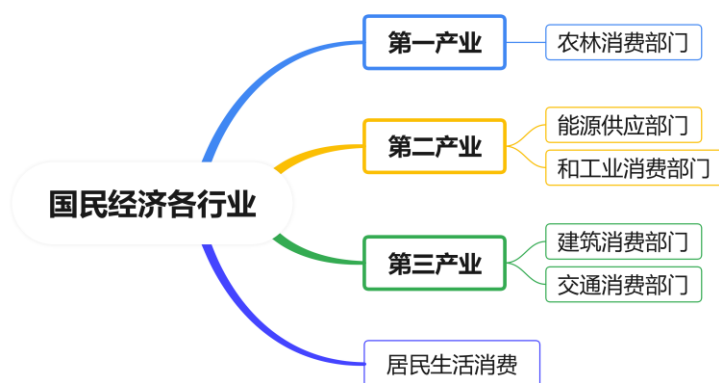


图 2-1 国民经济各行业分类图

4. 经济增长假设。2035 年的 GDP 比基期（2020 年）翻一番；2060 年比基期翻两番；

5. 碳消纳量假设。2060 年生态碳汇的碳消纳量为基期碳排放量的 10%；2060 年工程碳汇或碳交易的碳消纳量为基期碳排放量 10%。

6. 能源消费结构假设。随着积极政策不断推进，能源消费结构会发生变化，非化石能源消费比重逐渐增加。

7. 相关性假设。假设碳排放量与人口、GDP 和能源消费量之间存在线性或非线性的关联关系。

8. 持续趋势假设。假设历史趋势在未来一段时间内会持续，碳排放存在累积效应。

3. 符号说明

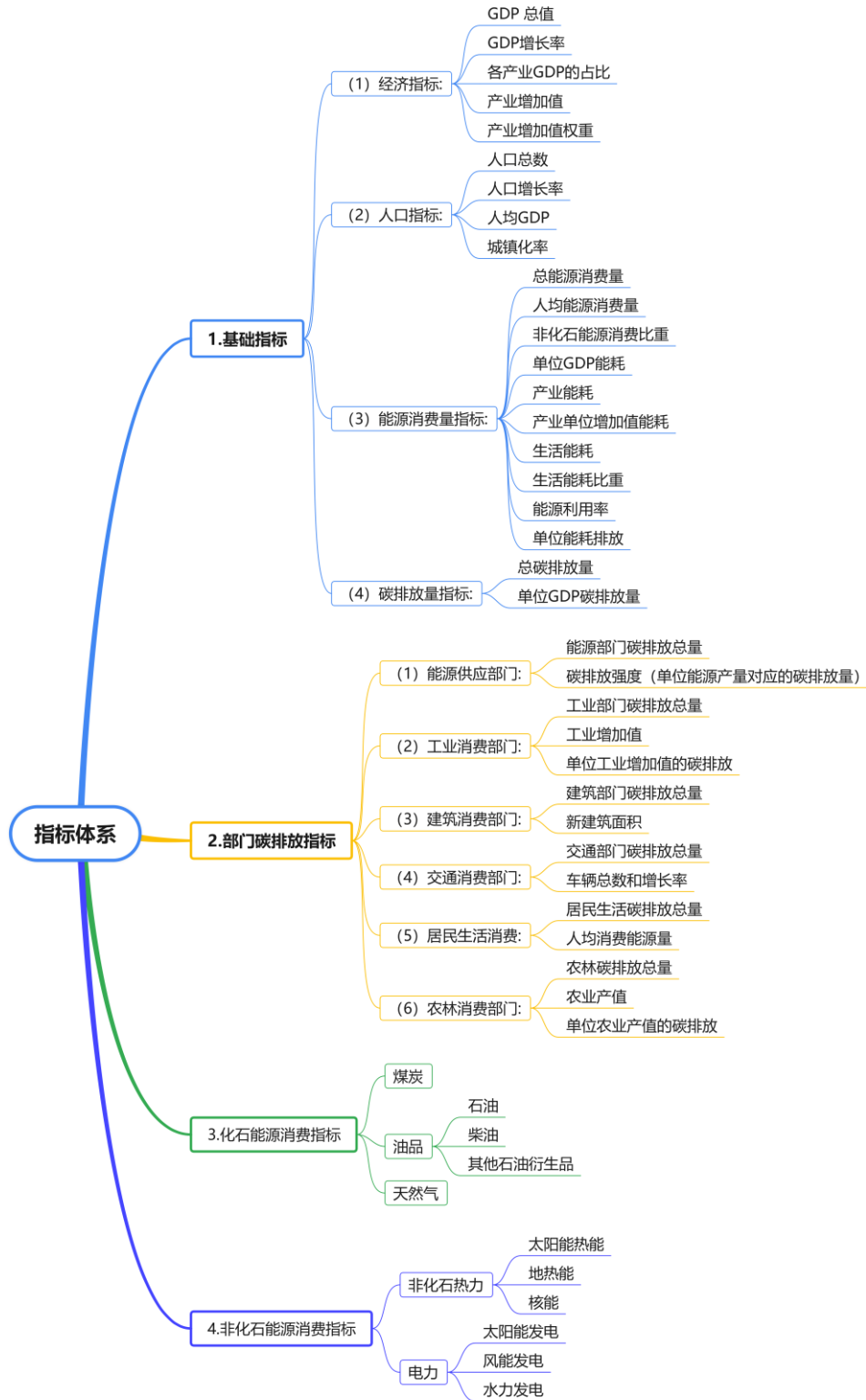
符号	含义
g_i	第 <i>i</i> 产业增加值 ($i=1, 2, 3$)
C	碳排放总量
P	总人口数量
P^t	第 <i>t</i> 年人口数量
G	GDP 总量
E	能源消费总量
g	人均 GDP
e	能源强度 单位 GDP 所消耗的能源
c	能源碳排放强度，单位能源消耗产生的碳排放
h	单位 GDP 产生的碳排放强度
E_i	第 <i>i</i> 产业能耗 ($i=1, 2, 3$)
e_i	第 <i>i</i> 产业单位增加值能耗
E_{life}	生活能耗
n_{life}	生活能耗比重
E^t	第 <i>t</i> 年能耗
a_i	第 <i>i</i> 产业增加值权重
e_p	人均能耗
n_e	能源利用效率

注：其他符号将在文中具体说明

4. 问题一模型建立与求解

4.1 建立指标与指标体系

通过分析问题，提出指标如下：



根据问题研究的需要，首先提出了四个基本指标分别为经济指标、人口指标、能源消费量指标和碳排放量指标。其次，根据数据文件，提出五个部门碳排放指标分别为能源供应部门指标、工业消费部门指标、建筑消费部门指标、交通消费部门指标和农林消费部门指标。同时根据问题要求，区分化石能源消费和非化石能源消费两项指标。这些指标将作为一级指标，具体下分二级指标。

1. 经济指标分为 GDP 总值、GDP 增长率、各产业相对 GDP 总值占比、产业增加值和产业增加值权重；
2. 人口指标分为人口总数、人口增长率、人均 GDP 和城镇化率；
3. 能源消费指标分为总能源消费量、人均能源消费量、非化石能源消费比重、单位 GDP 能耗、产业能耗、产业单位增加值能耗、生活能耗、生活能耗比重、能源利用率和单位能耗排放；
4. 碳排放量指标分为总碳排放和单位 GDP 碳排放量；
5. 能源供应部门分为碳排放总量和碳排放强度（单位能源产量对应的碳排放量）；
6. 工业消费部门分为碳排放总量、工业增加值和单位工业增加值的碳排放；
7. 建筑消费部门分为碳排放总量和新建筑面积；
8. 交通消费部门分为碳排放总量、车辆总数和车辆增长率；
9. 居民生活消费分为碳排放总量和人均消费能源量；
10. 农林消费部门分为碳排放总量、农业产值、单位农业产值的碳排放；
11. 化石能源消费指标分为煤炭、油品、石油、柴油、其他石油衍生品、天然气；
12. 非化石能源消费指标分为非化石热力、太阳能热能、地热能、核能、电力、太阳能发电、风能发电、水力发电。

4.2 区域碳排放量以及经济、人口、能源消费量的现状分析

4.2.1 碳排放量状况分析

4.2.1.1 碳排放总量及变化趋势

根据题目所给出的数据进行分析，从 2011 年到 2020 年（十二五和十三五期间），碳排放总量和碳排放总量的年增长率如表 4-1 所示。2010-2020 年各产业碳排放量变化趋势如图 4-1 所示，从中可以看出碳排放量总体呈波动上升的趋势，且在各产业中第二产业碳排放量最大。

表 4-1 碳排放总量及年增长率统计表

年份	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
碳排放量 (万 tCO2)	65193.3	67502.6	66749.4	64853.3	66074.8	68526.1	70451.6	71502.0	74096.3	72633.3
年增长率 (%)	15.673	3.542	1.116	-2.841	1.884	3.710	2.810	1.491	3.628	-1.974

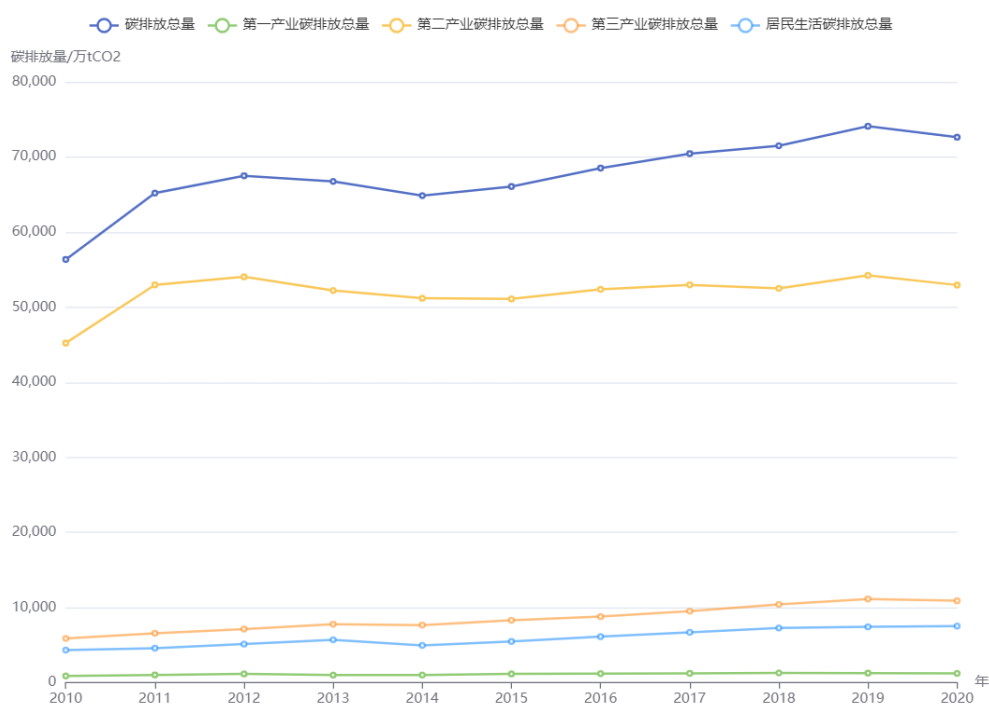


图 4-1 2010-2020 年各产业碳排放量变化趋势折线图

下图 4-2 和图 4-3 分别展示了十二五和十三五期间碳排放量及其增长率的变化趋势。从图 4-2 可以看出，十二五前期碳排放量虽然仍保持上升趋势，但增长率已经大大减少，碳排放增长率变缓；在十二五的后期增长率有小幅上升，但仍保持较小值，碳排放量总体增长不大。从图 4-3 可以看出，十三五期间增长率较为波动但总体在 3% 附近，增长率保持较小值，碳排放总量有较小增长，而在十三五后期，碳排放增长率大幅下降至负值，碳排放总量减少，并有望在未来一段时间内持下降趋势，推测可能是受到了新冠疫情停工停产的影响。



图 4-2 十二五期间碳排放量变化趋势及增长率变化

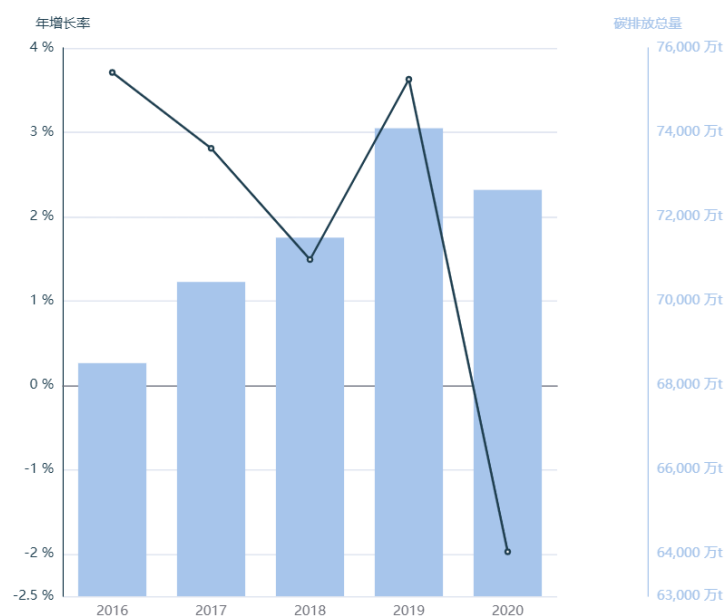


图 4-3 十三五期间碳排放量变化趋势及增长率变化

下图 4-4 展示了 2011-2020 年期间（即十二五和十三五期间）总体碳排放量及其增长率的变化趋势。从图 4-4 可得，十二五期间碳排放量较少，而十三五期间碳排放量总体较多，十二五到十三五期间，碳排放量总体呈增长状态；从十二五到十三五期间，碳排放增长呈波动状态，但从初始较高的 15.67% 下降至 3% 附近，总体上碳排放量呈现延缓增长的趋势。

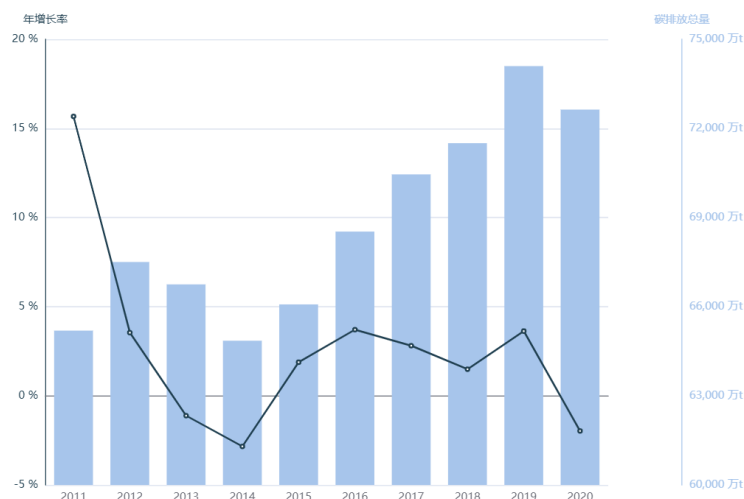


图 4-4 2011 年-2020 年碳排放量变化趋势及增长率变化

4.2.1.2 单因素 ANOVA 分析碳排放量

基于上文对 2011 年-2020 年碳排放总量变化趋势的描述与分析，进一步地，对十二五和十三五期间的碳排放量进行单因素方差分析（Analysis of Variance，简称 ANOVA），来检验十二五期间的碳排放量与十三五期间的碳排放量是否存在显著差异[3]。

(1) 对 2011 年-2020 年碳排放量进行正态性检验
由于样本量较小，所以采用 Shapiro-Wilk 检验法，得到其输出结果如下表 4-2 所示：

表 4-2 Shapiro-Wilk 检验结果表

变量名	样本量	中位数	平均值	标准差	偏度	峰度	S-W 检验
碳排放量 (万 tCO ₂)	10	68014.369	68758.276	3244.33	0.413	-1.252	0.935(0.495)

由表 4-2 可以看出，Shapiro-Wilk 检验[4]下的显著性 P 值为 0.495 大于 0.05，水平上不呈现显著性，不能拒绝原假设（原假设为数据符合正态分布），因此数据满足正态分布。

(2) 对 2011 年-2020 年碳排放量进行方差齐次性检验

此处选择进行利用 MATLAB 的 varstestn 函数进行检验，得到输出结果如下表 4-3 和图 4-5 所示：

表 4-3 方差齐次性检验结果表

组	计数	均值	标准差	Bartlett 统计量	自由度	p 值
1	5	66074.7	1091.19	1.46201	1	0.22661
2	5	71441.9	2118.47			
池化	10	68758.3	1685.02			

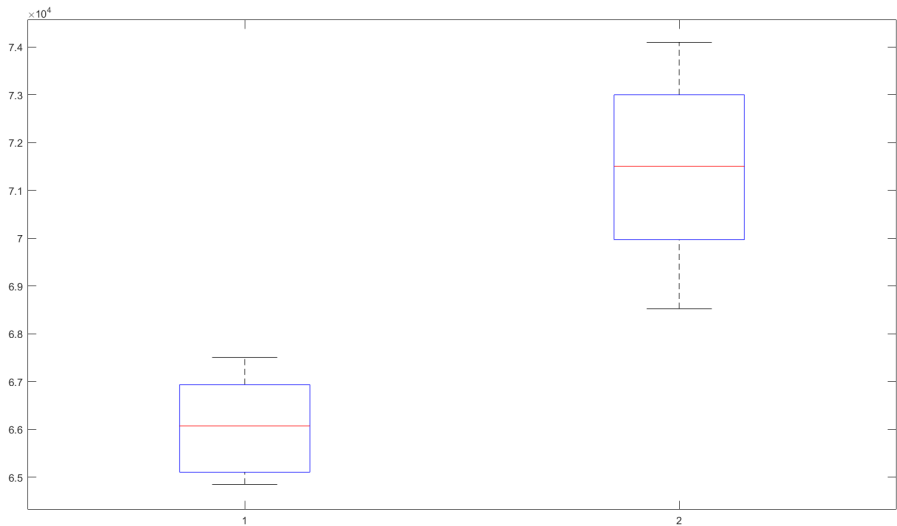


图 4-5 2011 年-2020 年碳排放量方差齐次性检验结果图

可以看出，检验的 p 值=0.22661>0.05,说明在显著性水平 0.05 下接受原假设，认为十二五和十三五两个阶段的碳排放量服从方差相同的正态分布，满足方差分析的基本假定。

(3) 方差分析

在前两步的基础上，对 2011 年-2020 年碳排放量进行方差分析得到结果如下表 4-4 和图 4-6 所示：

表 4-4 方差分析结果表

差异来源	误差平方和 (*10 ⁷)	自由度	均方差 (*10 ⁷)	F 值	p 值 (Prob>F)
组间	7.20167	1	7.20167		
组内	2.27144	8	0.28393	25.36	0.001
合计	9.47311	9			

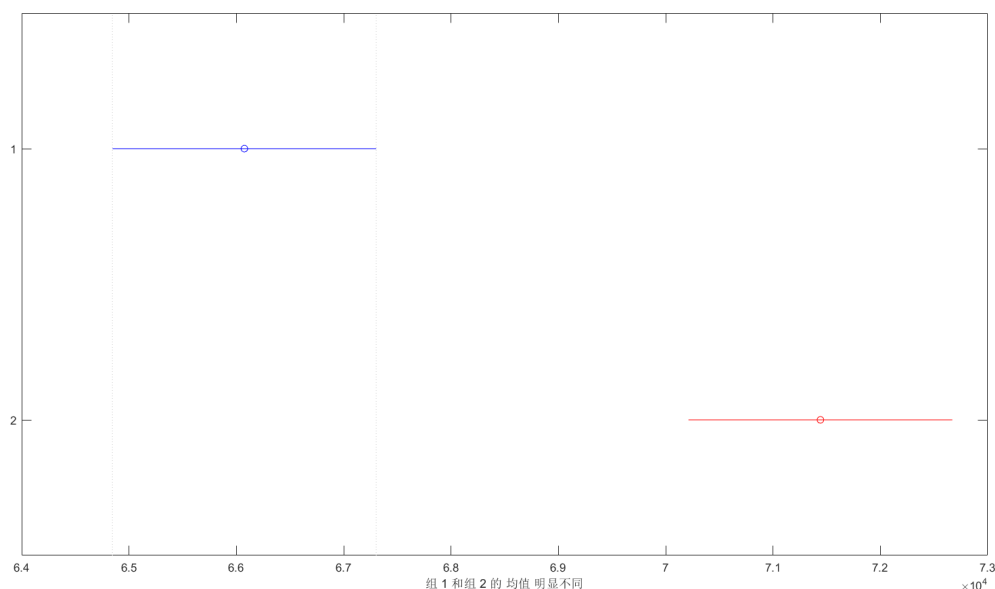


图 4-6 2011 年-2020 年碳排放量方差分析结果图

上表中的 p 值为 0.001 小于 0.05，代表着十二五和十三五两个阶段的碳排放量之间存在显著差异。同时，在图 4-6 所展示的交互式图形上用一圆圈标出了两个阶段的碳排放量均值，用一条之间段标出了每个阶段的碳排放量均值的置信区间。图中可见两条线段不相交，即没有重叠的部分，则说明这两个阶段的碳排放量均值之间的差异是显著的。在图上任意选中一个阶段，选中的阶段用蓝色显示，则与选中的阶段差异显著的另一阶段相应地呈现红色高亮。

4.2.2 碳排放量影响因素分析

十二五和十三五期间各部门占地区生产总值的比例分别如图 4-7、图 4-8 所示。可以看出，在十二五和十三五期间第二产业中的工业消费部门生产总值占总生产值比例最大，而第一产业农业消费部门占比最小。同时，从十二五到十三五期间工业消费部门的占比有所减少，而建筑消费部门占比有所增加，说明在此期间国家更注重基础设施建设，而减少了一定对工业的关注。

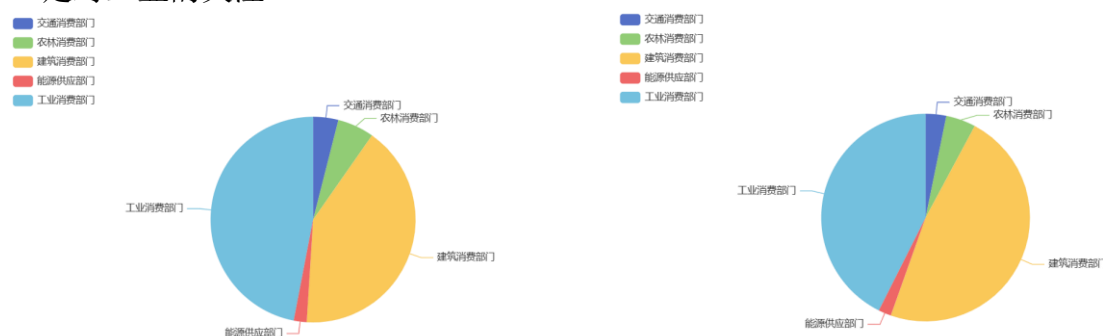


图 4-7 十二五期间各部门占总产值比例图 图 4-8 十三五期间各部门占总产值比例图

十二五和十三五期间各部门能源消耗占比如图 4-9、图 4-10 所示。可以得出，十二五和十三五期间工业消耗部门均占能源消耗比例最大，而农业消费部门则占能源消耗比例最小。同时，从十二五到十三五期间居民消费部门和交通消费部门的能源消耗占比略有增加，由此可看出此期间人民生活水平有所提高且交通更加发达。

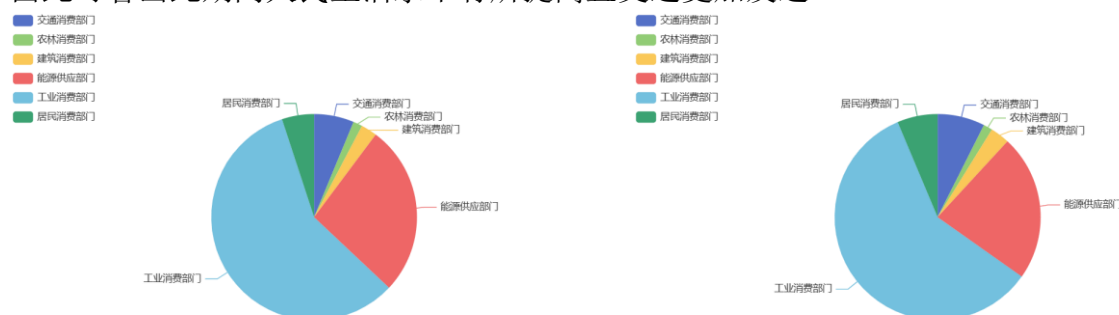


图 4-9 十二五期间各部门能源消耗占比图 图 4-10 十三五期间各部门能源消耗占比图

十二五和十三五期间各能源消费品种结构如图 4-11、图 4-12 所示。可得，十二五和十三五期间煤炭消费均为能源消费的主要形式。从十二五到十三五期间，煤炭消费量有所减少，而新能源电力和天然气消费有所增加，说明国家正在积极减少能源结构中煤炭的占比，提高非化石能源的比重（天然气作为化石能源中较为清洁的能源也在提高使用）。同时，在此期间外地调入电有所增加，由于本题区域位于东南沿海地区，可推测此增长是由于西电东送政策的推行。

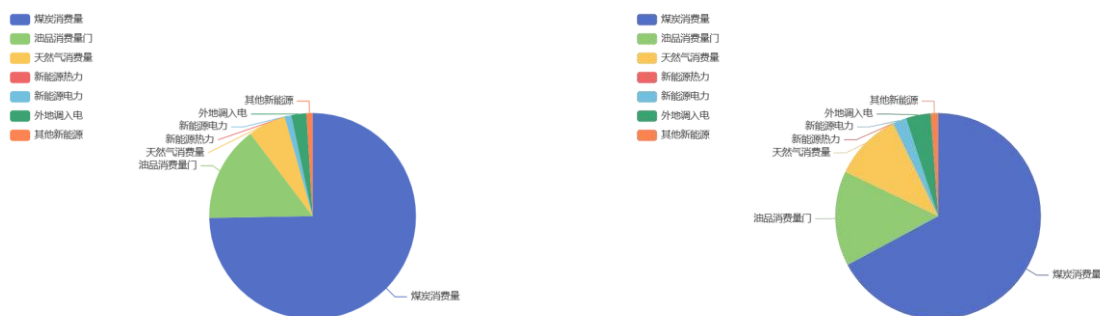


图 4-11 十二五期间能源消费品种结构占比图 图 4-12 十三五期间能源消费品种结构占比图

为更直观观察各部门消耗能源种类，本文绘制了十二五期间和十三五期间能源消费品种桑基图如图 4-13、图 4-14 所示。可以得出，十二五和十三五期间煤炭能源消费仍是能源消耗的主体，能源供应部门和工业消费部门都主要由煤炭进行能源供应，能源供应部门通过煤炭等能源产生的能量会供给给电力和热力，提供这两种能源。十二五和十三五期间虽然能源消耗种类占比总体一致，但从数据中可以看出，在此期间中，煤炭在大多部门的能源消费量均有所降低，虽然由于工业发展对煤炭依赖性较强，工业消费部门对煤炭消费量略有上升，但还是做到了尽可能在不影响发展的同时减少煤炭的使用。而油品、天然气、热力、电力和其他资源在各部门中消费量均有所增加，可以得出在此期间国家对较为清洁能源的提倡推动了这些资源的利用，同时提高了能源消耗中非化石能源的比重，这对双碳政策的实施有着重大的意义[5]。

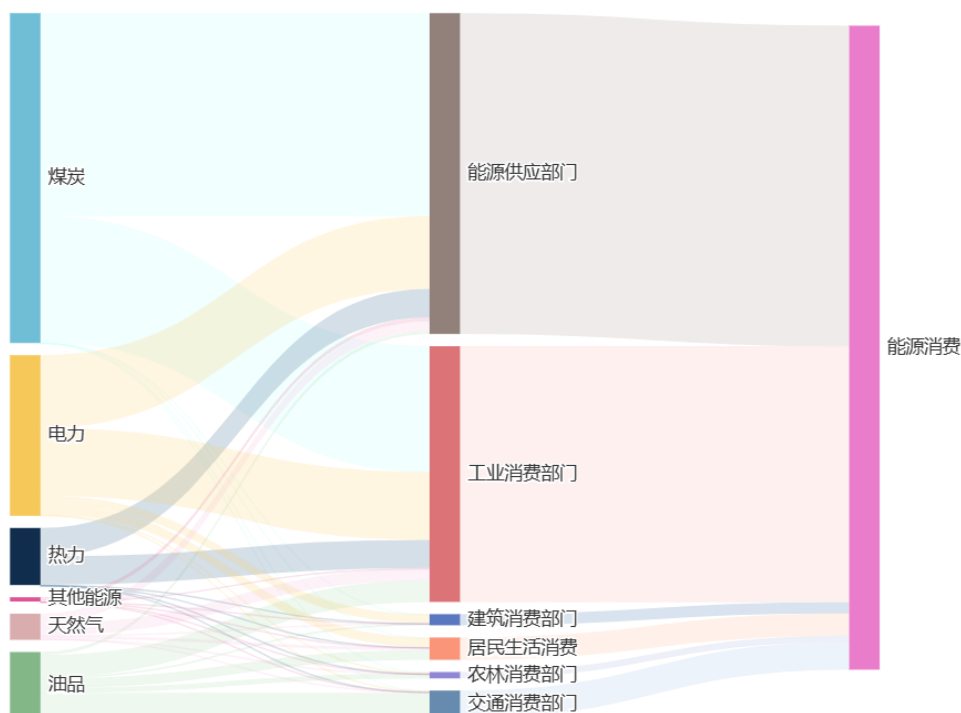


图 4-13 十二五产业部门能源消费品种结构桑基图

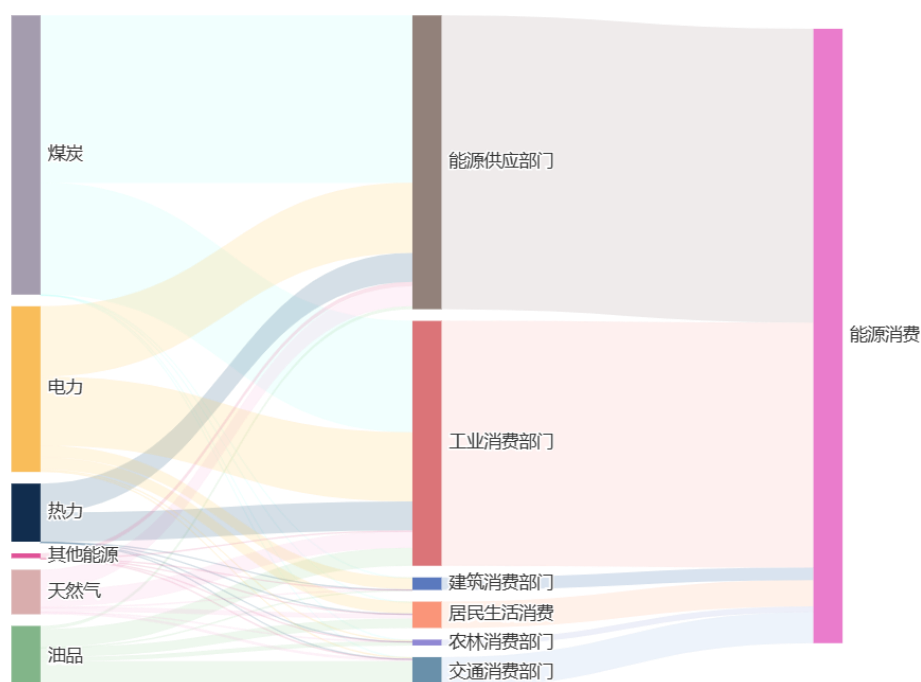


图 4-14 十三五产业部门能源消费品种结构桑基图

综上所述，十二五和十三五期间，工业消费部门的能源消耗和生产总值占比均占据主导地位，说明工业发展产生的经济效益和需要的能源投入都很高，由于我国要实现经济发展和碳排放控制的两手抓，故优化产业结构升级和提高非化石能源占比是重要解决方法。同时，由于技术和基础设施等方面的限制，目前能源消费中煤炭能源仍占据主导地位，但在双碳政策推行下，煤炭使用量在逐渐减少，而其他较为清洁的新型能源的消耗量占比在逐渐增加。

4.2.3 碳达峰与碳中和面临的挑战

碳排放量主要受人口、经济、能源消费量等因素影响，同时也可能受一些环境因素的影响，对影响因素的分析可以为该区域双碳路径规划中差异化的路径选择提供依据[6]。

人口与碳排放量的关系和人口与人均碳排放量的关系如图 4-15、图 4-16 所示。从图中可以看出人口与碳排放量总体上成正相关，人口与人均排放量无明显相关关系，但人口较多时总体上人均碳排放量较大。随着城镇化推进，城市常住人口数增多，居民相关活动产生的碳排放量也会增多，故人口与碳排放量成正相关。

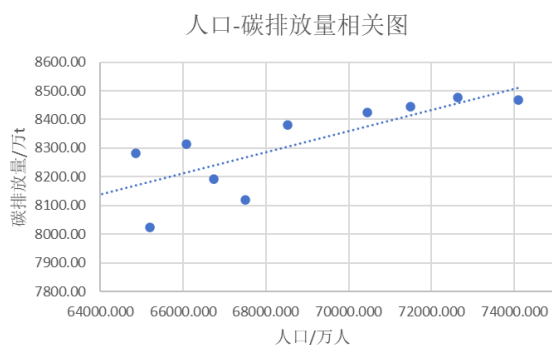


图 4-15 人口-碳排放量拟合

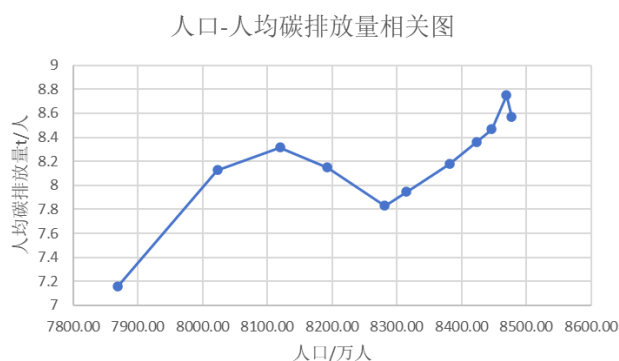


图 4-16 人口-人均碳排放量拟合图

经济因素的影响可由 GDP、产值构成等经济指标与碳排放量的关系，图 4-17、图 4-18 给出了 GDP-碳排放量的关系和该区域产值构成。从图中可以看出 GDP 和碳排放量总体呈正相关，产业构成中第二、三产业占总 GDP 产值较大，而这两个产业包含工业消费部门和建筑消费部门等消耗能源主要为煤炭的部门，因此这两个产业会较大地影响碳排放，如何在保持经济发展趋势的同时降低碳排放量是双碳政策实施的重要挑战之一。

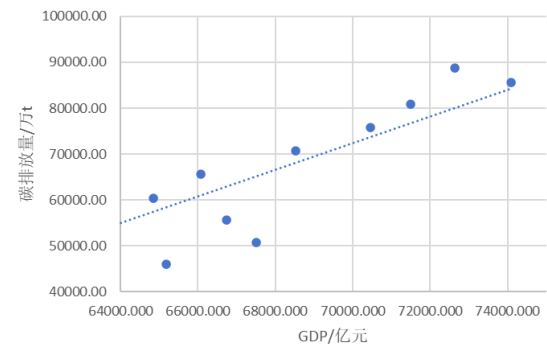


图 4-17 GDP-碳排放量拟合图

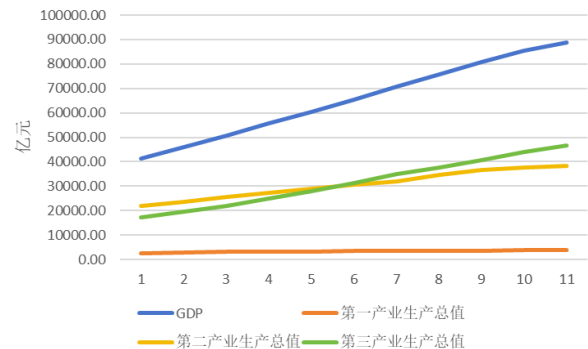


图 4-18 各产业 GDP 折线图

能源消费量如 4.2 中分析，煤炭仍在能源消费结构中占据主导地位，而煤炭消费会导致碳排放增加，故如何减少煤炭在能源消费中的占比，提高非化石能源的占比，是实现双碳政策的重要问题。

除此之外，随着国家政策、科技发展、国际合作关系、天气变化等因素的改变，碳排放量也会受到一定的影响[7]。

4.3 相关指标及其关联模型

4.3.1 相关指标的变化分析

4.3.1.1 GDP 变化分析

表 4-5 为十二五（2011~2015）和十三五（2016~2020）期间 GDP 及其环比、同比数据变化（2020 年 GDP 基准为 41383.87 亿元）：

表 4-5 GDP 及其环比、同比数据表（GDP 单位：亿元）

		第一年	第二年	第三年	第四年	第五年
十二五	GDP	45952.65	50660.2	55580.11	60359.43	65552
	环比（*100%）	0.1104	0.1024	0.0971	0.0860	0.0860
十三五	GDP	70665.71	75752.20	80827.71	85556.13	88683.21
	环比（*100%）	0.0780	0.0720	0.0670	0.0585	0.0366
	同比（*100%）	0.53783	0.4953	0.4543	0.4174	0.3529

反映在图图像上如下所示：

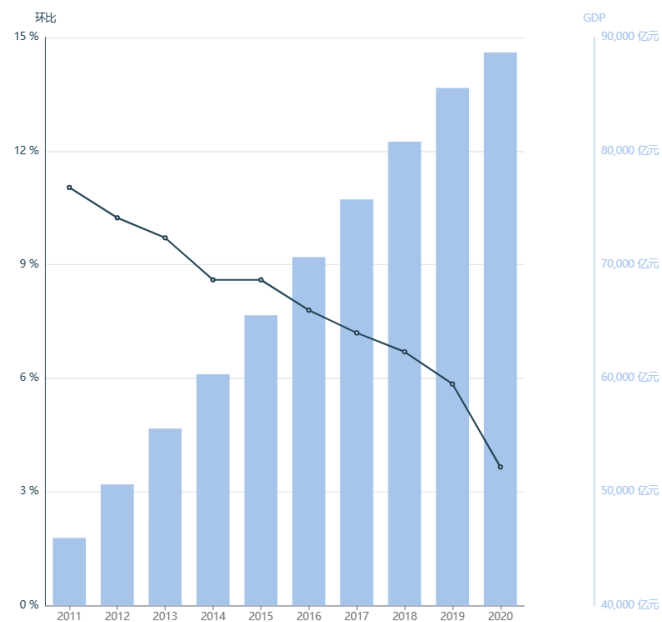


图 4-19 2011~2020 年 GDP 及其环比数据变化

由表 4-5 和图 4-19 可以看出：在十二五（2011~2015）和十三五（2016~2020）期间该区域 GDP 逐年攀升，但环比数据不断下降，表示该区域 GDP 持续增加但增速不断放缓。且 2019 年到 2020 年环比数据下降剧烈，分析由于新冠肺炎疫情等因素影响，导致 GDP 增速大幅下跌。



图 4-20 十三五期间 GDP 同比数据变化

由图 4-20 可以看出，与十二五阶段相比，十三五阶段的 GDP 同比数据虽然不断下降，但是一直处于 0 以上。说明在两个阶段的同一时期，十三五的 GDP 都高于十二五的 GDP。但由于整体上 GDP 增速放缓，所以十三五期间的同比数据也不断下降。

4.3.1.2 人口变化分析

表 4-6 为十二五（2011~2015）和十三五（2016~2020）期间人口及其环比、同比数据变化（2020 年人口基准为 7869.34 万人）：

表 4-6 人口及其环比、同比数据表（人口单位：万人）

		第一年	第二年	第三年	第四年	第五年
十二五	人口	8022.99	8119.81	8192.44	8281.09	8315.11
	环比（*100%）	0.0195	0.0121	0.0089	0.0108	0.0041
十三五	人口	8381.47	8423.50	8446.19	8469.09	8477.26
	环比（*100%）	0.0080	0.0050	0.0027	0.0027	0.0010
	同比（*100%）	0.0447	0.0374	0.0310	0.0227	0.0195

反映在图像上如下所示：

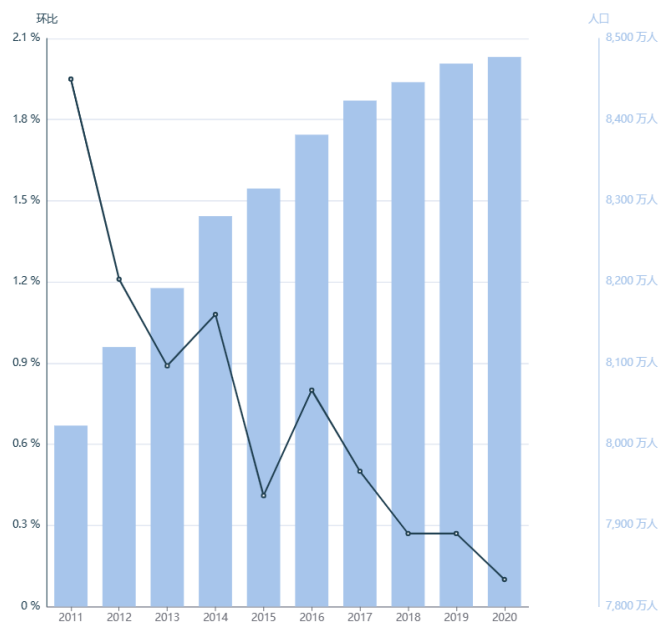


图 4-21 2011~2020 年人口及其环比数据变化

由表 4-6 和图 4-21 可以看出：在十二五（2011~2015）和十三五（2016~2020）期间该区域人口不断增加，与此同时其环比数据也不断波动，但整体呈下降趋势且始终为正数，这表示该区域人口持续增加但增速朝着下降的方向变化。



图 4-22 十三五期间人口同比数据变化

由图 4-22 可以看出，与十二五阶段相比，十三五阶段的人口同比数据不断下降，但一直处于 0 以上。说明在两个阶段的同一时期，十三五的人口都高于十二五的人口。但由于整体上人口增速放缓，所以十三五期间的同比数据也不断下降。

4.3.1.3 能源消费量变化分析

表 4-7 为十二五（2011~2015）和十三五（2016~2020）期间能源消费量及其环比、同比数据变化（2020 年能源消费量基准为 23539.3144312995 万 tce）：

表 4-7 能源消费量及其环比、同比数据表（能源消费量单位：万 tce）

		第一年	第二年	第三年	第四年	第五年
十二五	能源消费量	26860.03	27999.22	28203.10	28170.51	29033.61
	环比（*100%）	0.1411	0.0424	0.0073	-0.0012	0.0306
十三五	能源消费量	29947.98	30669.89	31373.13	32227.51	31438.00
	环比（*100%）	0.0315	0.0241	0.0229	0.0272	-0.0245
	同比（*100%）	0.11	0.10	0.11	0.14	0.08

反映在图像上如下所示：

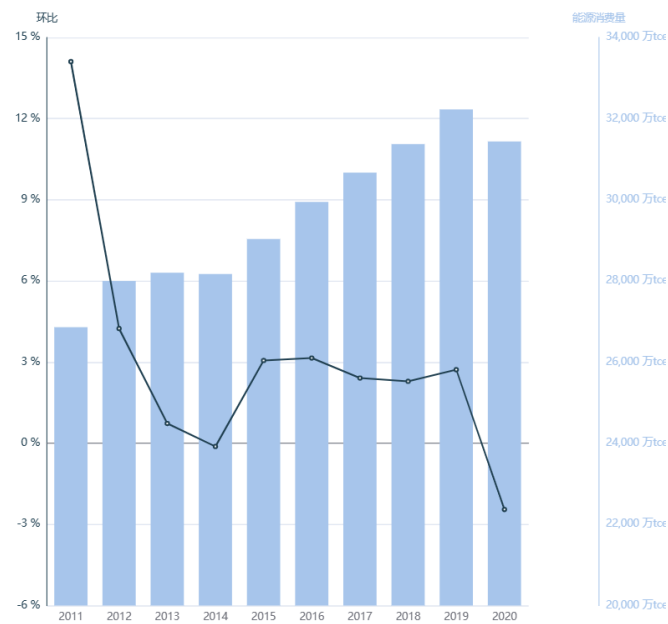


图 4-23 2011~2020 年能源消费量及其环比数据变化

由表 4-7 和图 4-23 可以看出：在十二五（2011~2015）和十三五（2016~2020）期间该区域能源消费量除了在 2014 年与 2020 年出现下降外，其余年份都不断增加。与此同时，其环比数据在十二五初期经历了大幅下降后发生了小范围的波动，随后在十三五后期再次开始大幅度下降，整体上呈快速下降趋势，且开始出现负数。这表示该区域能源消费量虽然有所增加但增速快速下降，近年来已经开始出现能源消费量逐年减少的趋势。



图 4-24 十三五期间能源消费量同比数据变化

由图 4-24 可以看出，与十二五阶段相比，十三五阶段的能源消费量同比数据呈现先上升后下降的趋势，且一直处于 0 以上。说明在两个阶段的同一时期，十三五的 GDP 都高于十二五的 GDP。但由于其环比数据经历过小范围的波动，所以其同比数据也出现了波动。

4.3.1.4 各产业与居民生活能源消费量变化分析

表 4-8 为十二五（2011~2015）和十三五（2016~2020）期间各产业和居民生活能源消费量及其环比、同比数据变化（2020 年第一、二、三产业和居民生活能源消费量基准分别为 345.36、20113.07、1932.84、1148.05 万 tce）：

表 4-8 各产业和居民生活能源消费量及其环比、同比数据表（能源消费量单位：万 tce）

第一产业能源消费量						
		第一年	第二年	第三年	第四年	第五年
十二五	能源消费量	393.87	448.95	362.18	383.72	429.37
	环比（*100%）	0.1405	0.1398	-0.1932	0.0595	0.1190
十三五	能源消费量	433.46	440.49	457.99	438.58	423.78
	环比（*100%）	0.0095	0.0162	0.0397	-0.0424	-0.0337
	同比（*100%）	0.1005	-0.0189	0.2646	0.1429	-0.0130
第二产业能源消费量						
		第一年	第二年	第三年	第四年	第五年
十二五	能源消费量	23145.47	23869.22	23889.14	23670.13	24261.96
	环比（*100%）	0.1508	0.0313	0.0008	-0.0092	0.0250
十三五	能源消费量	24929.67	25317.56	25580.66	26196.12	25325.99
	环比（*100%）	0.0275	0.0156	0.0104	0.0241	-0.0332
	同比（*100%）	0.0771	0.0607	0.0708	0.1067	0.0439
第三产业能源消费量						
		第一年	第二年	第三年	第四年	第五年
十二五	能源消费量	2115.56	2308.96	2481.88	2644.20	2776.11
	环比（*100%）	0.0945	0.0914	0.0749	0.0654	0.0499
十三五	能源消费量	2878.53	3049.20	3300.79	3522.38	3507.62
	环比（*100%）	0.0369	0.0593	0.0825	0.0671	-0.0042
	同比（*100%）	0.3606	0.3206	0.3300	0.3321	0.2635
居民生活能源消费量						
		第一年	第二年	第三年	第四年	第五年
十二五	能源消费量	1205.13	1372.09	1469.91	1472.45	1566.16
	环比（*100%）	0.0497	0.1385	0.0713	0.0017	0.0636
十三五	能源消费量	1706.31	1862.64	2033.69	2070.43	2180.60
	环比（*100%）	0.0895	0.0916	0.0918	0.0181	0.0532
	同比（*100%）	0.4159	0.3575	0.3835	0.4061	0.3923

反映在图像上如下所示：

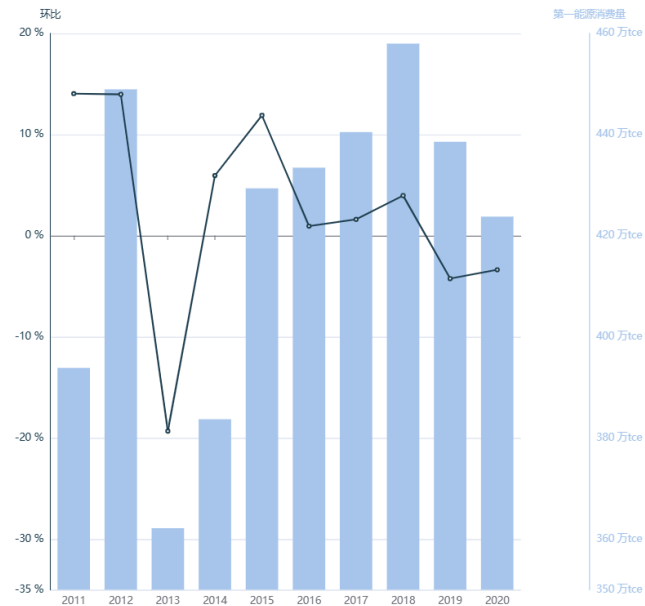


图 4-25 2011~2020 年第一产业能源消费量及其环比数据变化

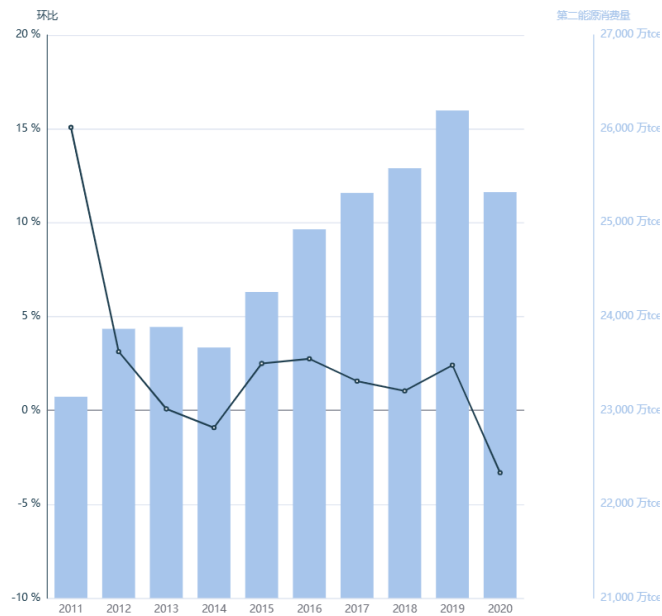


图 4-26 2011~2020 年第二产业能源消费量及其环比数据变化

由表 4-8、图 4-25、图 4-26、图 4-27、图 4-28 可以看出：在十二五（2011~2015）和十三五（2016~2020）期间该区域第一、二、三产业和居民生活能源消费量虽然在个别年份出现了一定的下降，但其整体上都呈现上升趋势。与此同时，第一、二、三产业能源消费量的环比数据整体上都呈现下降趋势，尤其是第二、三产业的能源消费量环比数据在十三五后期都出现了大幅度的下降，且都降到了负数。第一产业的能源消费量环比数据在十三五后期虽然没有大幅度下降，但是到后面也下降到了负数。而居民生活能源消费量的环比数据则一直处于较大幅度的波动之中，并无明显趋势，且一直保持在 0 以上。



图 4-27 2011~2020 年第三产业能源消费量及其环比数据变化

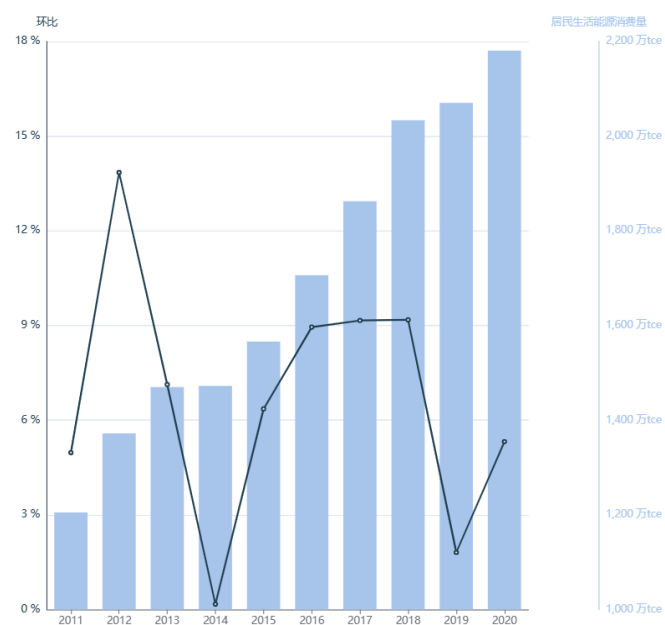


图 4-28 2011~2020 年居民生活能源消费量及其环比数据变化

这表示该区域第一、二、三产业的能源消费量虽然在十二五初期不断增加，但随着时间的推移，其增速不断降低，到十三五后期更是降到 0 以下，也即第一、二、三产业的能源消费量开始逐年减少。而居民生活的能源消费量虽然增速不一，但却逐年攀升，并未出现减少的趋势。

值得注意的是第一产业的能源消费量及其环比数据在 2013 年出现突变，分析其可能是由于某些特殊原因造成的异常值。

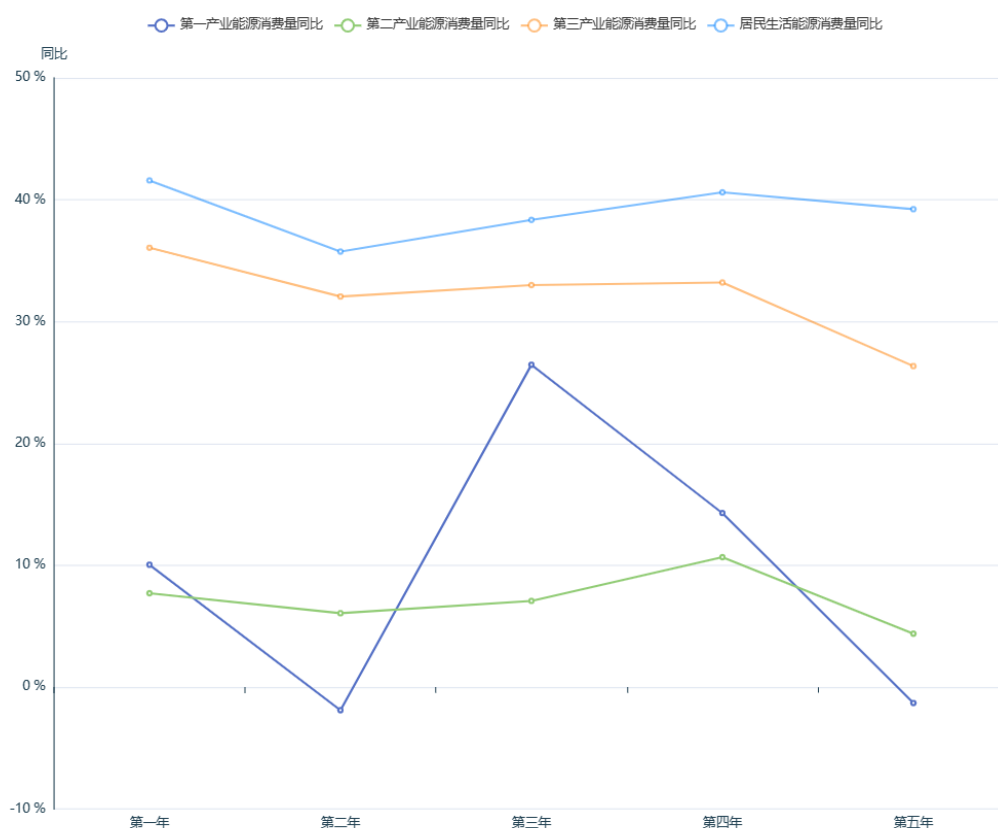


图 4-29 十三五期间各产业与居民生活能源消费量同比数据变化

由图 4-29 可以看出，与十二五阶段相比，十三五阶段的能源消费量同比数据除了第一产业外都不断下降，且一直处于 0 以上。说明在两个阶段的同一时期，十三五的能源消费量除了第一产业外都高于十二五的能源消费量。而十三五阶段第一产业的能源消费量在第二、第五年则低于十二五的能源消费量。与此同时，在各产业之间，居民生活的能源消费量同比曲线整体处于最高位，其次是第三产业和第二产业，说明居民生活的能源消费量在两个阶段同比增长得最多。

4.3.1.5 各产业 GDP 变化分析

表 4-9 为十二五（2011~2015）和十三五（2016~2020）期间各产业 GDP 及其环比、同比数据变化（2020 年第一、二、三产业 GDP 基准分别为 2409.24、21853.60、17121.03 亿元）。

由图 4-30、图 4-31、图 4-32、表 4-9 可以看出：在十二五（2011~2015）和十三五（2016~2020）期间该区域第一、二、三产业的 GDP 都呈现上升趋势。与此同时，第一、二、三产业的 GDP 环比数据整体上都呈现下降趋势，尤其是第二产业的 GDP 环比数据在十三五后期出现了大幅度的下降。但除了第一产业的 GDP 环比数据在 2017 年左右出现过负数以外，其余各产业的 GDP 环比数据目前都处于 0 以上。

这表示该区域第一、二、三产业的 GDP 在不断增加的同时，其增速不断降低。尤其是 2019 到 2020 年期间，受新冠肺炎疫情等因素影响，代表农业生产的第一产业的 GDP 环比上升，相应的 GDP 也上升；而代表工业和服务业的第二、三产业 GDP 环比则大幅下降但仍为正，相应的 GDP 虽增加但增速大幅下降。

表 4-9 各产业 GDP 及其环比、同比数据表（GDP 单位：亿元）

第一产业 GDP						
		第一年	第二年	第三年	第四年	第五年
十二五	GDP	2736.86	3057.82	3228.54	3358.61	3636.08
	环比（*100%）	0.1360	0.1173	0.0558	0.0403	0.0826
十三五	GDP	3690.61	3568.54	3591.61	3726.61	3916.81
	环比（*100%）	0.0150	-0.0331	0.0065	0.0376	0.0510
	同比（*100%）	0.3485	0.1670	0.1125	0.1096	0.0772
第二产业 GDP						
		第一年	第二年	第三年	第四年	第五年
十二五	GDP	23739.96	25612.91	27298.13	28907.54	30700.41
	环比（*100%）	0.0863	0.0789	0.0658	0.0590	0.0620
十三五	GDP	32013.02	34514.33	36533.74	37730.14	38183.23
	环比（*100%）	0.0428	0.0781	0.0585	0.0327	0.0120
	同比（*100%）	0.3485	0.3475	0.3383	0.3052	0.2437
第三产业 GDP						
		第一年	第二年	第三年	第四年	第五年
十二五	GDP	19475.83	21989.47	25053.44	28093.28	31215.51
	环比（*100%）	0.1375	0.1291	0.1393	0.1213	0.1111
十三五	GDP	34962.07	37669.33	40702.36	44099.38	46583.18
	环比（*100%）	0.1200	0.0774	0.0805	0.0835	0.0563
	同比（*100%）	0.7952	0.7131	0.6246	0.5697	0.4923

反映在图像上如下所示：

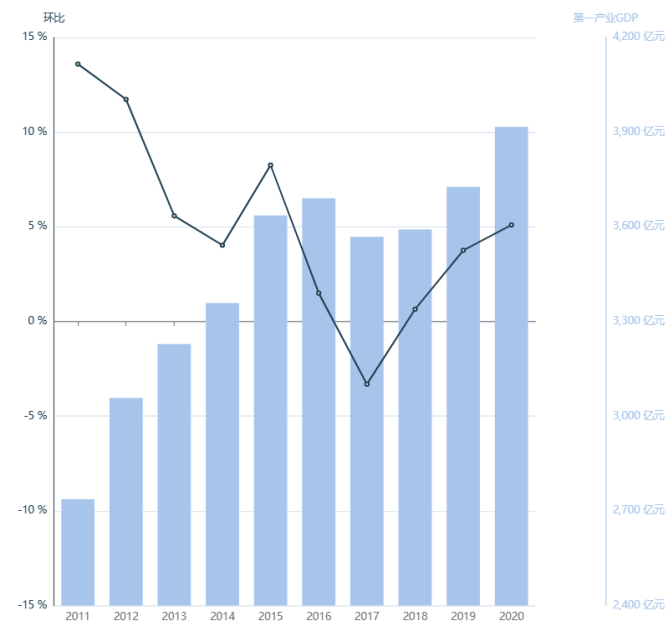


图 4-30 2011~2020 年第一产业 GDP 及其环比数据变化

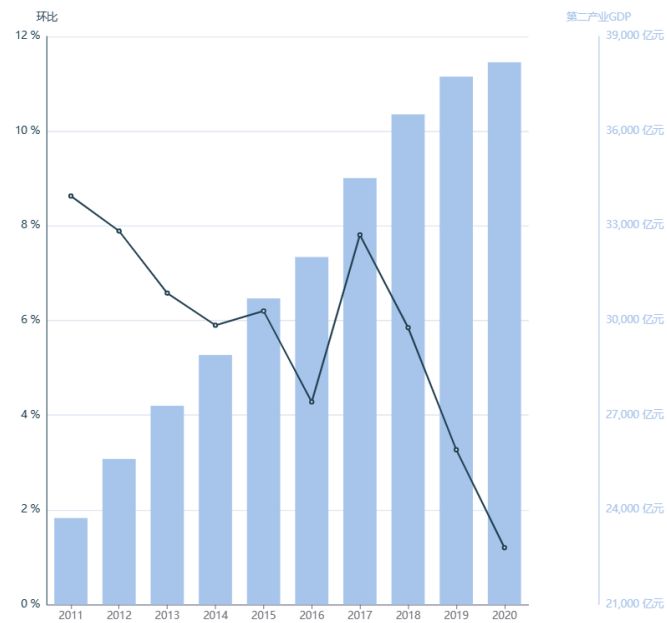


图 4-31 2011~2020 年第二产业 GDP 及其环比数据变化

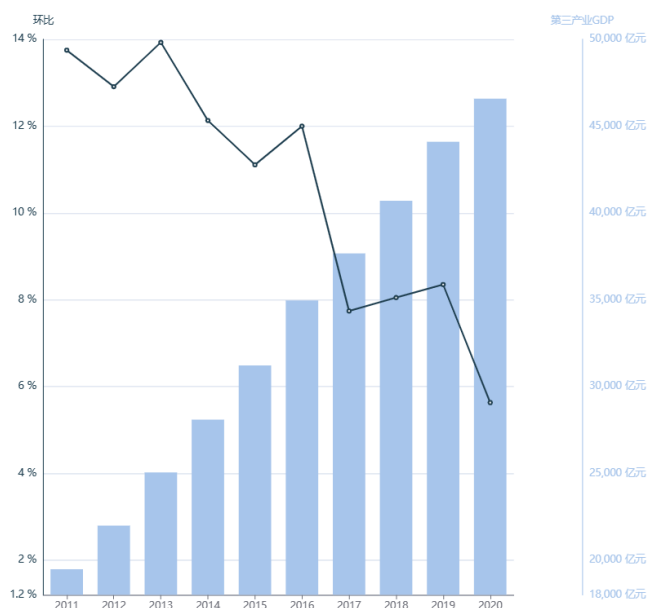


图 4-32 2011~2020 年第三产业 GDP 及其环比数据变化

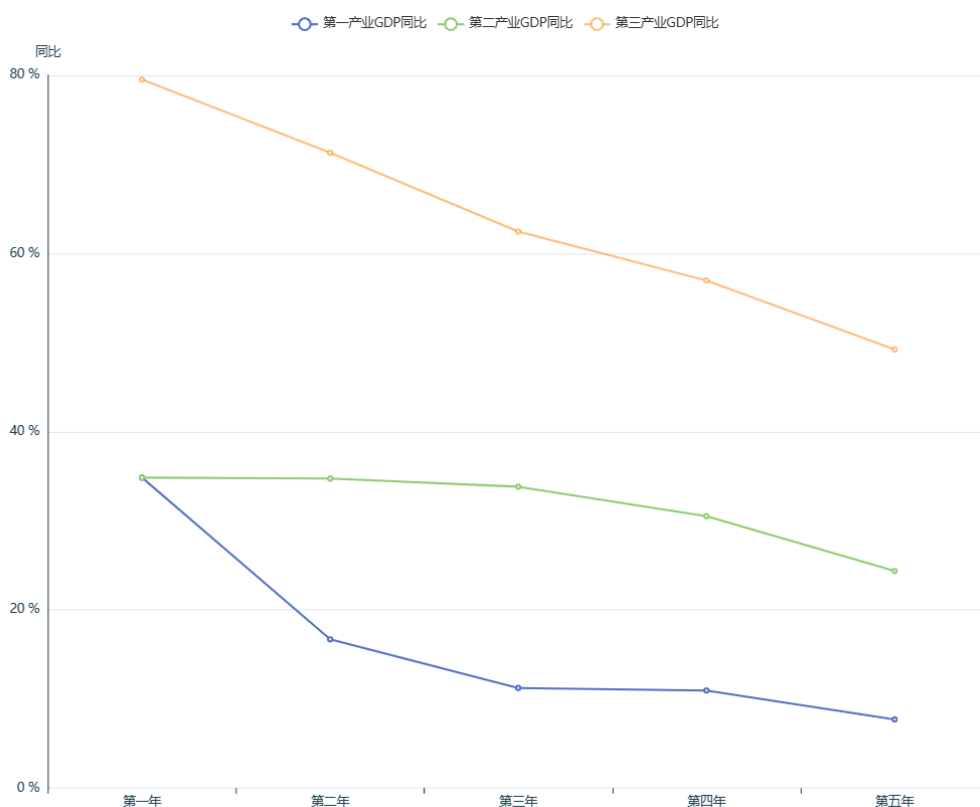


图 4-31 十三五期间各产业 GDP 同比数据变化

由图 4-31 可以看出，与十二五阶段相比，十三五阶段的各产业 GDP 同比数据虽然不断下降，但是一直处于 0 以上。说明在两个阶段的同一时期，十三五的各产业 GDP 都高于十二五的各产业 GDP。但由于整体上各产业的 GDP 增速放缓，所以十三五期间的同比数据也不断下降。与此同时，在各产业之间，第三产业的 GDP 同比曲线整体处于最高位，其次是第二产业和第一产业，说明第三产业的 GDP 在两个阶段同比增长得最多。

4.3.2 Spearman 相关性分析

相关分析是指对两个或多个具备相关性的变量元素进行分析，从而衡量因素之间的相关密切程度，只有存在一定的联系或者概率的元素才可进行相关性分析。相关性分析常用皮尔逊相关系数（Pearson correlation）和斯皮尔曼相关系数（Spearman correlation）来衡量，皮尔逊相关系数通常用于衡量两个连续性随机变量间的相关系数，适用于分析连续、正态分布、线性关系的数据，而斯皮尔曼相关系数适用于处理反映观测对象等级、顺序关系的数据定序数据[8]。由于本问题中数据不符合正态分布的要求，故采用 Spearman 方法进行相关性分析。Spearman 相关系数是秩相关系数（Coefficient of Rank Correlation），又称等级相关系数，反映的是两个随机变量的变化趋势方向和强度之间的关联，是将两个随机变量的样本值按数据的大小顺序排列位次，以各要素样本值的位次代替实际数据而求得的一种统计量[9]。Spearman 相关系数计算方法如公式（4.1）所示，其中 d_i 表示第 i 个数据对的位次值之差， n 为总的观测样本数。

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{2.1}$$

4.3.3 Spearman 相关性分析结果

Spearman[10]相关性分析后，所得的关于 GDP、人口和能源消费量的相关系数表如表 4-10 所示，相关系数通常为[0,1]之前的数，且相关系数越接近 1 表示相关性越大，越接近 0 则表示相关性越小。通过 Spearman 相关性分析获得的相关系数热力图如图 4-32 所示。从图 4-32 可以看出人口、GDP、能源消费量与碳排放量之间相关系数均在[0.8,1]之间，可得出这人口、GDP、能源消费量对碳排放量有较大影响，其中能源消费量与碳排放量的相关系数为 0.918，此时 Spearman 相关系数最大，由此可见能源消费量与碳排放量之间关系最密切，对碳排放量影响最大。

表 4-10 人口、GDP、能源消费量与碳排放量相关系数表

	碳排放量	GDP	人口	能源消费量
碳排放量	1	0.882	0.882	0.918
GDP	0.882	1	1	0.982
人口	0.882	1	1	0.982
能源消费量	0.918	0.982	0.982	1

表 4-11 第一、二、三产业能源消费量与碳排放量相关系数表

	碳排放量	第一产业能源消费量	第二产业能源消费量	第三产业能源消费量	居民生活能源消费量
碳排放量	1	0.682	0.945	0.891	0.882
第一产业	0.682	1	0.655	0.545	0.518
第二产业	0.945	0.655	1	0.964	0.945
第三产业	0.891	0.545	0.964	1	0.991
居民生活能源消费量	0.882	0.518	0.945	0.991	1

表 4-12 第一、二、三产业生产总值与碳排放量相关系数表

	碳排放量	第一产业生产总值	第二产业生产总值	第三产业生产总值
碳排放量	1	0.773	0.882	0.882
第一产业	0.773	1	0.927	0.927
第二产业	0.882	0.927	1	1
第三产业	0.882	0.927	1	1

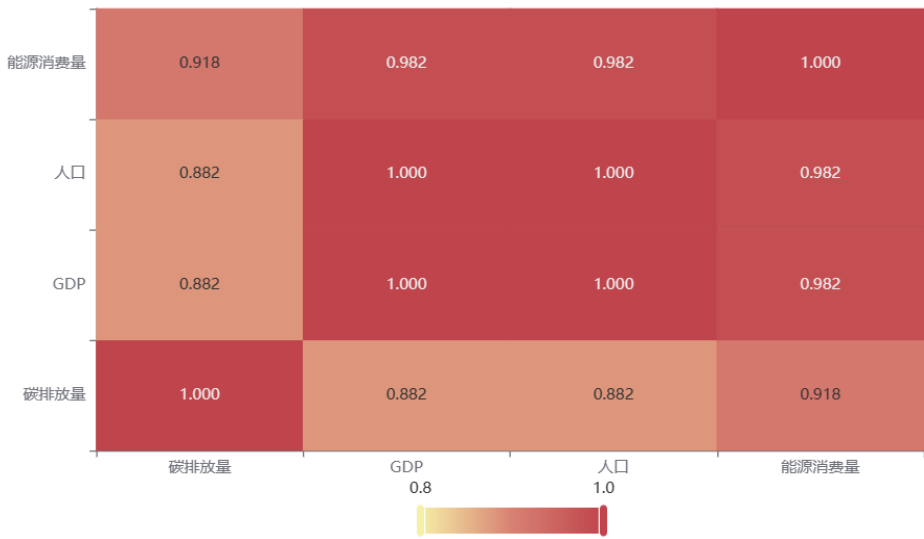


图 4-32 人口、GDP、能源消费量与碳排放量相关系数热力图

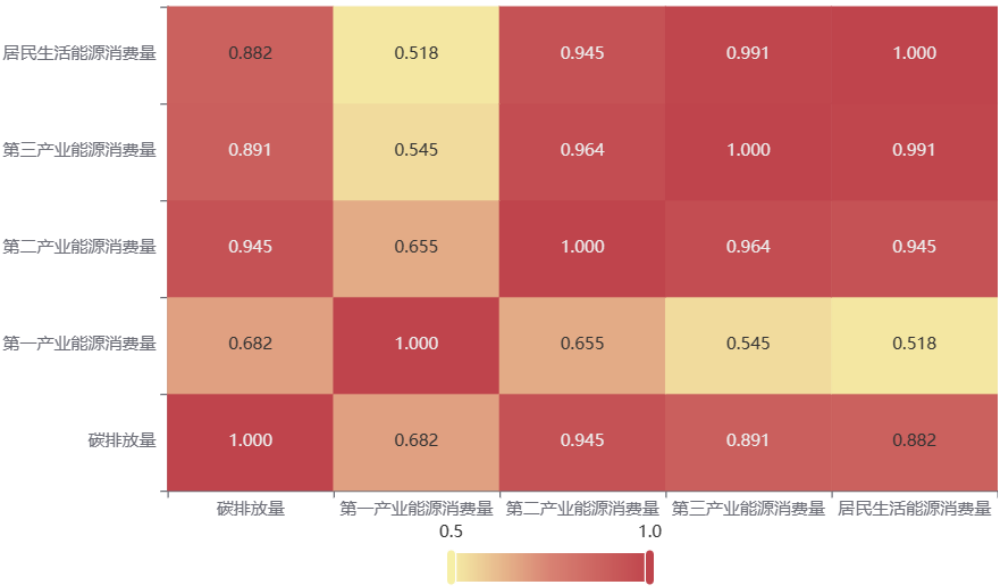


图 4-33 第一、二、三产业能源消费量与碳排放量相关系数热力图

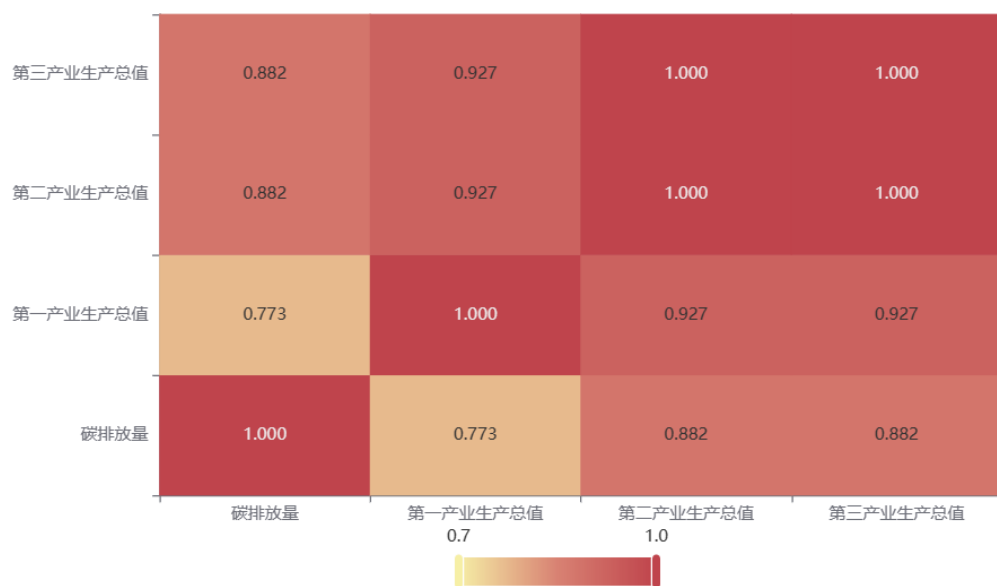


图 4-34 第一、二、三产业生产总值与碳排放量相关系数热力图

表 4-11 通过 Spearman 相关性分析同样可得到碳排放量与第一、二、三产业能源消费量的相关系数和碳排放量与第一、二、三产业生产总值的相关系数。第一、二、三产业能源消费量与碳排放量相关系数表如表 4-11 所示，第一、二、三产业能源消费量与碳排放量相关系数热力图如图 4-33 所示，第一、二、三产业生产总值与碳排放量相关系数表如表 4-12 所示，第一、二、三产业生产总值与碳排放量相关系数热力图如图 4-34 所示。

从图 4-33 可以得出，第一、二、三产业能源消费量与碳排放量相关系数均在 0.5 以上，均与碳排放量有较大关系。其中，第一产业农林消费部门的能源消费量与碳排放量的相关系数为 0.682，在其中最低，说明能源消费量中第一产业对碳排放量影响最小；第二产业工业消费部门的能源消费量与碳排放量的相关系数为 0.945，在其中最高，说明能源消费量中第二产业对碳排放量影响最大。从图 4-34 可以得出，第一、二、三产业生产总值与碳排放量相关系数均在 0.7 以上，均与碳排放量有较大关系。其中，第一产业农林消费部门的生产总值与碳排放量的相关系数为 0.773，在其中最低，说明能源消费量中第一产业对碳排放量影响最小；第二产业能源供应部门和工业消费部门以及第三产业交通消费部门和建筑消费部门的生产总值与碳排放量的相关系数均为 0.882，在其中最高，说明生产总值中第二产业和第三产业对碳排放量影响最大。

综上所述，能源消费量对碳排放量影响最大，而第二产业的能源消耗和生产总值对与碳排放量都有较强的联系。

5. 问题二模型建立与求解

5.1 基于人口和经济变化的能源消费量预测模型

5.1.1 基于人口的能源消费量预测模型

5.1.1.1 线性回归预测模型

线性回归预测模型是根据已知的变量(自变量)来预测某个连续的数值变量(因变量)。一元线性回归模型也被称为简单线性回归模型,是指模型中只含有一个自变量和一个因变量,本问使用的就是一元线性回归模型[11]。一元线性回归模型数学公式如下:

$$y = a + bx + \varepsilon \quad (3.1)$$

其中 a 为模型的截距项, b 为模型的斜率项, ε 为模型的误差项。

绘制理想的拟合线,需使误差项 ε 达到最小。由于误差项是 y 与 $a+bx$ 的差,结果可能为正值或负值,因此误差项 ε 达到最小的问题需转换为误差平方和最小的问题,采用最小二乘法的思路。误差平方和的公式可以表示为:

$$J(a,b) = \sum_{i=1}^n \varepsilon^2 = \sum_{i=1}^n (y_i - [a + bx_i])^2 \quad (3.2)$$

回归系数 a 和 b 的求解公式如下:

$$a = \bar{y} - b\bar{x} \quad (3.3)$$

$$b = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2} \quad (3.4)$$

其中 $\bar{x}$, $\bar{y}$ 分别为自变量和因变量的平均值。

本问中建立基于人口的能源消费量预测模型,总人口数量为 P , 能源消费总量为 E , 设第 t 年人口数量为 P^t , 第 t 年能源消费量为 E^t , 则线性回归模型如下所示:

$$E^t = a + bP^t \quad (3.5)$$

将收集到数据导入进行线性回归得到表 5-1。分析得出, F 检验的显著性 P 值为 $0.000***$, 水平上呈现显著性, 拒绝回归系数为 0 的原假设, 因此模型基本满足要求。 R^2 代表曲线回归的拟合程度, 越接近 1 效果越好。 VIF 值代表多重共线性严重程度, 用于检验模型是否呈现共线性, 即解释变量间存在高度相关的关系 (VIF 应小于 10 或者 5, 严格为 5) 若 VIF 出现 $\inf$, 则说明 VIF 值无穷大, 此时考虑检查共线性, 或者使用岭回归。 F 检验是为了判断是否存在显著的线性关系, R^2 是为了判断回归直线与此线性模型拟合的优劣。在线性回归中主要关注 F 检验是否通过, 而在某些情况下 R^2 大小和模型解释度没有必然关系。

表 5-1 人口-能源消费量回归分析表

线性回归分析结果 n=11 因变量：能源消费量									
	非标准化系数		标准化系数		t	P	VIF	R ²	调整 R ²
	B	标准误差	Beta						
常数	-70760.134	8593.918	----	-8.234	0.000***	----		0.937	0.931
人口	12.064	1.039	0.968	11.616	0.000***	1			
									F=134.936
									P=0.000***

注：***、**、*分别代表 1%、5%、10%的显著性水平

求解得出线性回归模型为：

$$E^t = -70760.134 + 12.064P^t \quad (3.6)$$

为检验线性回归模型与真实数据的拟合程度，绘制了人口-能源消费量回归拟合效果图如图 5-1 所示。



图 5-1 人口-能源消费量回归拟合效果图

建立人口-能源消费量线性回归方程后，为预测十四五至二十一五期间的能源消费量，需先预测此期间的常驻人口数量，因此需建立时间和人口之间的线性回归方程，如下所示：

$$P^t = a + bt \quad (3.7)$$

表 5-2 年份-人口回归分析表

线性回归分析结果 n=11 因变量：人口									
	非标准化系数		标准化系数		t	P	VIF	R ²	调整 R ²
	B	标准误差	Beta						
常数	-108334.379	11712.369	-----	-9.25	0.000***	-----		0.917	
年份	57.869	5.813	0.957	9.956	0.000***	1		0.908	F=99.12 P=0.000***

注：***、**、*分别代表 1%、5%、10%的显著性水平

其中 P^t 表示第 t 年的常住人口数量， t 表示年份。将数据导入 MATLAB 中进行线性回归分析得到年份和人口线性回归分析表如表 5-2 所示。由线性回归结果可以得出时间和人口之间的线性回归方程为：

$$P^t = -108334.379 + 57.869t \quad (3.8)$$

F 检验的显著性 P 值为 0.000***，水平上呈现显著性，拒绝回归系数为 0 的原假设，因此模型基本满足要求。VIF 不为无穷，说明可以使用线性回归。为检验线性回归模型与真实数据的拟合程度，绘制了年份-人口回归拟合效果图如图 5-2 所示，从图中可以看出预测值与拟合值相差较小，拟合较为合理。



图 5-2 年份-人口回归拟合效果图

表 5-3 人口-能源消费量回归预测结果表

年份	常住人口数量	能源消费量
2021	8619.788545	33230.92481
2022	8677.658	33929.07487
2023	8735.527455	34627.22493
2024	8793.396909	35325.37499
2025	8851.266364	36023.52504
2026	8909.135818	36721.6751
2027	8967.005273	37419.82516
2028	9024.874727	38117.97522
2029	9082.744182	38816.12527
2030	9140.613636	39514.27533
2031	9198.483091	40212.42539
2032	9256.352545	40910.57545
2033	9314.222	41608.72551
2034	9372.091455	42306.87556
2035	9429.960909	43005.02562
2036	9487.830364	43703.17568
2037	9545.699818	44401.32574
2038	9603.569273	45099.47579
2039	9661.438727	45797.62585
2040	9719.308182	46495.77591
2041	9777.177636	47193.92597
2042	9835.047091	47892.07603
2043	9892.916545	48590.22608
2044	9950.786	49288.37614
2045	10008.65545	49986.5262
2046	10066.52491	50684.67626
2047	10124.39436	51382.82631
2048	10182.26382	52080.97637
2049	10240.13327	52779.12643
2050	10298.00273	53477.27649
2051	10355.87218	54175.42655
2052	10413.74164	54873.5766
2053	10471.61109	55571.72666
2054	10529.48055	56269.87672
2055	10587.35	56968.02678
2056	10645.21945	57666.17683
2057	10703.08891	58364.32689
2058	10760.95836	59062.47695
2059	10818.82782	59760.62701
2060	10876.69727	60458.77707

5.1.1.2 灰色预测模型

灰色预测模型是一种基于灰色系统理论的预测方法，它可以对含有不确定因素的系统进行预测，它通过对待预测对象的历史数据进行分析 and 建模，来预测未来的趋势和发展规律[12]。与传统的统计方法和时间序列方法相比，灰色预测模型更适用于短期、非线性、小样本的预测问题。灰色预测模型的核心思想是将待预测对象分为发展趋势和随机变动两个部分，利用已知的发展趋势建立数学模型，然后根据该模型对未知的发展趋势进行预测。具体而言，灰色预测模型通常包括 GM(1,1)模型和 GM(2,1)模型两种常见形式。在 GM(1,1)模型中，通过对原始数据进行累加生成级比数列，并利用级比数列构建微分方程，从而得到一阶线性常微分方程模型。通过求解该微分方程，可以得到对未来发展趋势的预测结果。

而在 GM(2,1)模型中，首先对原始数据进行累减生成白化数列，然后通过指数平滑累减生成新的级比数列，最后通过建立二阶线性常微分方程模型，得到对未来发展趋势的预测结果[13]。灰色预测模型在实际应用中通常需要通过参数优选和模型检验来提高预测精度。此外，由于灰色预测模型具有较好的适应性和泛化能力，可以处理样本不足、数据缺失等问题，因此在各个领域，如经济学、环境科学、管理学等都得到了广泛的应用。本问使用 GM(1,1)模型对区域能源消费量进行预测。

(1) 数据的检验与处理

灰色预测需要将数据的级比进行校验，只有通过级比校验才适合使用灰色预测，其余数据则选择平移变换或选择平滑指数预测、回归预测等其他方法。设参考数据列 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ ，依次累加生成序列（1-AGO）：

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)) = (x^{(0)}(1), x^{(0)}(1) + x^{(0)}(2), \dots, x^{(0)}(1) + \dots + x^{(0)}(n)) \quad (3.9)$$

式中： $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, n$ ， $x^{(1)}$ 的均值生成序列 $z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n))$ ，

式中 $z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1), k = 2, 3, \dots, n$ ，建立灰微分方程：

$$x^{(0)}(k) + az^{(1)}(k) = b, k = 2, 3, \dots, n \quad (3.10)$$

相应的白化微分方程为

$$\frac{dx^{(1)}}{dt} + ax^{(1)}(t) = b \quad (3.11)$$

设参考数据为 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ ，计算序列的级比

$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}, k = 2, 3, \dots, n$ ，如果所有级比 $\lambda(k)$ 都落在可容覆盖 $\theta = (e^{-\frac{2}{n+1}}, e^{\frac{2}{n+2}})$ 内，则序

列 $x^{(0)}$ 可以作为模型 GM(1,1) 数据进行灰色预测。表 5-4 展示了序列值和级比值。若所有的

级比值都位于区间 $(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+2}})$ （即 $(0.846, 1.181)$ ）内，说明数据适合模型构建。

表 5-4 人口-能源消费量灰色预测级比检验结果表

索引项	原始值	级比值
7869.34	23539.314	—
8022.99	26860.026	0.876
8119.81	27999.218	0.959
8192.44	28203.104	0.993
8281.09	28170.506	1.001
8315.11	29033.608	0.97
8381.47	29947.977	0.969
8423.5	30669.886	0.976
8446.19	31373.127	0.978
8469.09	32227.505	0.973
8477.26	31437.998	1.025

(2) 建立模型

按照白化微分方程建立 GM(1,1)模型，则可以得到预测值

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{\hat{b}}{\hat{a}})e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}}, k = 0, 1, \dots, n-1, \dots \quad (3.12)$$

而且

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), k = 0, 1, \dots, n-1, \dots \quad (3.13)$$

表格 5-5 展示了发展系数、灰色作用量、后验差比值。由发展系数和灰色作用量可以构建灰色预测模型。其中发展系数表示数列的发展规律和趋势，灰色作用量反映数列的变化关系。后验差比值可以验证灰色预测的精度，后验差比值越小，则说明灰色预测精度越高。一般后验差比值 C 值小于 0.35 则模型精度高，C 值小于 0.5 说明模型精度合格，C 值小于 0.65 说明模型精度基本合格，如果 C 值大于 0.65，则说明模型精度不合格。此处后验差比值为 0.028，模型精度高[14]。

表 5-5 展示了灰色预测模型的拟合结果表。相对误差值越小越好，一般情况下小于 20% 即说明拟合良好。从表中可以看出模型平均相对误差为 1.076%，意味着模型拟合效果良好。

为了直观展示灰色预测模型拟合效果，图 5-3 给出了真实值和拟合值之间的关系，从图中可以看出，两条曲线趋势相近，拟合效果较好。图 5-4 给出了根据拟合曲线趋势，在十三五到二十一五之间能源消费量随人口的增长曲线。

表 5-5 人口-能源消费量灰色预测模型参数表

发展系数 a	灰色作用量 b	后验差比 C 值
-0.019	26352.943	0.028

表 5-6 人口-能源消费量灰色预测模型拟合结果表

索引项	原始值	预测值	残差	相对误差 (%)
7869.34	23539.314	23539.314	0	0
8022.99	26860.026	27072.345	-212.319	0.79
8119.81	27999.218	27603.412	395.806	1.414
8192.44	28203.104	28144.897	58.207	0.206
8281.09	28170.506	28697.004	-526.499	1.869
8315.11	29033.608	29259.942	-226.334	0.78
8381.47	29947.977	29833.923	114.054	0.381
8423.5	30669.886	30419.163	250.724	0.817
8446.19	31373.127	31015.883	357.243	1.139
8469.09	32227.505	31624.309	603.196	1.872
8477.26	31437.998	32244.671	-806.673	2.566

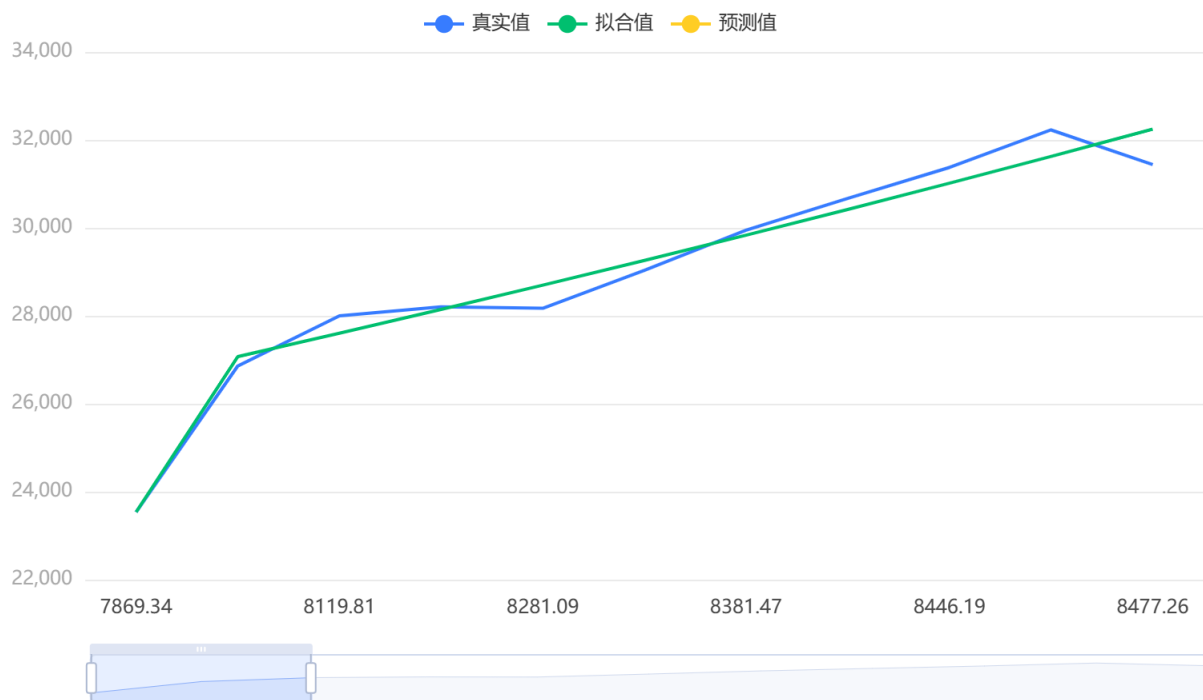


图 5-3 人口-能源消费量灰色预测模型拟合图

表 5-7 灰色预测模型的预测结果表

预测阶数	预测值
1	32877.201
2	33522.14
3	34179.731
4	34850.221
5	35533.863
6	36230.917
7	36941.644
8	37666.314
9	38405.199
10	39158.578
11	39926.736
12	40709.963
13	41508.554
14	42322.811
15	43153.04
16	43999.556
17	44862.678
18	45742.731
19	46640.047
20	47554.967
21	48487.833
22	49439
23	50408.825
24	51397.674
25	52405.922
26	53433.948
27	54482.14
28	55550.894
29	56640.614
30	57751.71
31	58884.602
32	60039.718
33	61217.493
34	62418.372
35	63642.808
36	64891.263
37	66164.209
38	67462.126
39	68785.503
40	70134.841



图 5-4 人口-能源消费量灰色预测模型预测图

5.1.1.3 LSTM 预测模型

LSTM (Long Short-Term Memory, 长短期记忆网络) 模型是一种特殊的循环神经网络 (RNN)，它在处理序列数据时具有出色的记忆能力。相比于传统的 RNN，LSTM 模型能够更好地解决长期依赖问题，对于处理时间序列、自然语言处理、机器翻译等任务非常有效[15]。LSTM 模型通过引入了一组称为“门” (gates) 的结构来控制信息的流动和记忆的更新。主要包括输入门、遗忘门和输出门。这些门的作用是根据输入数据的重要性对信息进行筛选、存储和输出。在 LSTM 模型中，每个时刻的隐藏状态包含了当前时刻的输入、前一时刻的隐藏状态以及前一时刻的记忆状态。通过门的控制和记忆单元的更新，LSTM 模型可以选择性地忽略或保留过去的信息，从而更好地捕捉序列数据中的长期依赖关系。

LSTM 模型通过反向传播算法进行训练，通常使用梯度下降来调整模型参数以最小化预测误差。训练过程中，LSTM 模型学习到的权重可以帮助模型自动提取输入数据中的关键特征，并将其应用于未来的预测或分类任务中。由于 LSTM 模型在处理长期依赖问题上的优越性能，它在许多领域取得了显著的应用。例如，可以将 LSTM 模型用于时间序列预测、语言模型构建、情感分析、机器翻译等任务。此外，LSTM 模型还可以与其他神经网络结构相结合，构建更加复杂和强大的深度学习模型，如 LSTM-CNN 模型和 LSTM-Attention 模型等[16]。

(1) 模型数据预处理

为了消除不同场景由于不同因素造成的影响，采用最大/最小归一化方法对样本数据进行归一化处理，公式如下：

$$x = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (3.14)$$

式中：X 为原始样本数据； $\min(X)$ 为原始样本数据的最小值； $\max(X)$ 样本数据的最大值；x 为归一化处理后的样本数据。

(2) 模型结构

LSTM 预测模型需要选择网络输入与输出变量。具体地，选用三个核心指标的数据作为样本数据，确定网络输入变量的个数。模型示意图如下图所示。LSTM 重复包含了 4 个交互的层，它主要包含细胞状态和门两个要素。其中细胞状态是重复单元顶端传送的向量。细胞状态只需要进行一些初级的线性变换，因此保持不变很容易。门是一个可以让信息选择性通过的结构，LSTM 通过精心设计的各种门来去除或新增信息到细胞状态。LSTM 有三个门来保护和控制细胞状态：遗忘门、输入门和输出门。遗忘门的确定从细胞状态中丢弃哪些信息，此处由 **sigmoid** 函数层来完成，**sigmoid** 函数控制每个部分有多少量可以通过。输入门确定什么样的新信息要输入细胞状态中，它的第一部分是 **sigmoid** 层，用于确定要更新哪个值；第二部分是 **tanh** 层，产生新的待用值。旧细胞状态通过遗忘门，待用值通过输入门，两者结合的结果更新细胞状态。输出门确定输出什么值，输出门中的 **sigmoid** 函数决定要输出的部分，与 **tanh** 将细胞状态变换后相乘，输出需要的信息。这三个门都使用 **sigmoid** 函数作为选择工具，**tanh** 函数作为变换工具，这两个函数结合起来实现三个门的功能。遗忘门、输入门和输出门的公式如下：

遗忘门：

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (3.15)$$

输入门：

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (3.16)$$

$$\tilde{C}_t = \tanh(W_C \times [h_{t-1}, x_t] + b_C) \quad (3.17)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (3.18)$$

输出门：

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (3.19)$$

$$h_t = o_t \times \tanh(C_t) \quad (3.20)$$

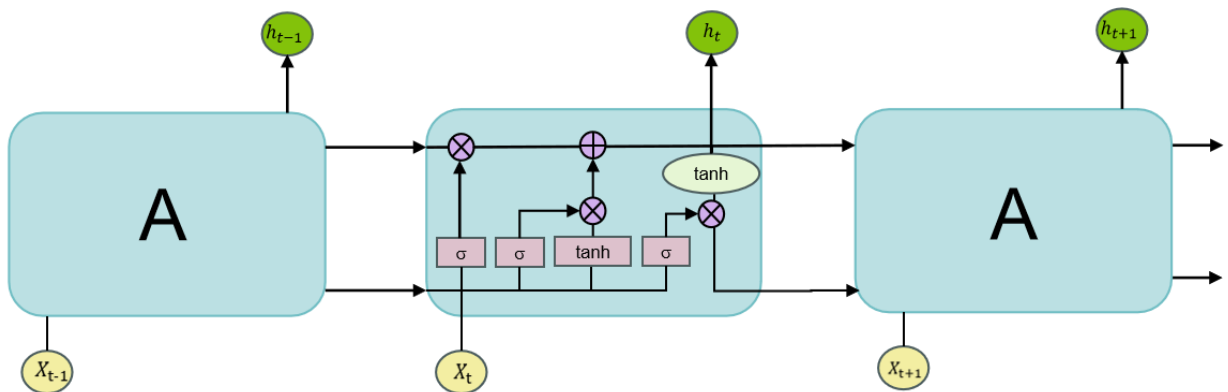


图 5-5 LSTM 预测模型示意图

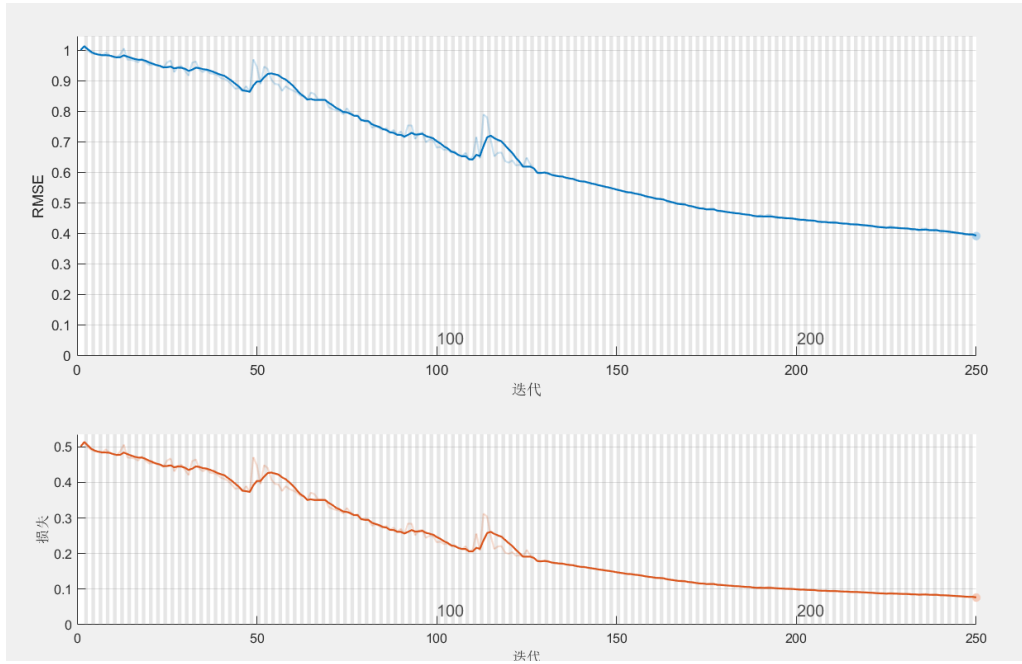


图 5-6 LSTM 预测模型训练图

5.1.1.4 加权综合预测模型

上文分别通过线性回归模型、灰色预测模型和 LSTM 预测模型对人口与能源消耗量的关系进行了建模，并预测了十三五到二十一五期间人口和能源消费量的变化。最后，本节通过对以上三种预测模型的结果进行加权优化分析，以提高模型精度[17]。

设 $\hat{y}_i (i=1, 2, \dots, m)$ 是一个问题的 m 种预测方法得到的预测结果， y'_{ij} 是第 $i (i=1, 2, \dots, m)$ 种预测方法对第 j 个原始数据 $y_j (j=1, 2, \dots, n)$ (n 为原始数据数量) 的模拟值。由此可以写出， m 种方法的加权预测值为：

$$J = \alpha_1 \hat{y}_1 + \alpha_2 \hat{y}_2 + \dots + \alpha_m \hat{y}_m \quad (3.21)$$

其中 $\sum_{i=1}^m \alpha_i = 1, \alpha_i \geq 0, (i=1, 2, \dots, m)$

设 $e_{ij} = \hat{y}_{ij} - y_j$ 表示第 i 种预测方法的预测值与第 j 个历史数据的模拟值之间的误差，记

$$E_j = \begin{bmatrix} e_{1j}^2 & e_{1j}e_{2j} & \cdots & e_{1j}e_{mj} \\ e_{1j}e_{2j} & e_{2j}^2 & \cdots & e_{2j}e_{mj} \\ \cdots & \cdots & \cdots & \cdots \\ e_{1j}e_{mj} & e_{2j}e_{mj} & \cdots & e_{mj}^2 \end{bmatrix} \quad (3.22)$$

$$\begin{aligned}
L &= \sum_{j=1}^n (\hat{J}_j - y_j)^2 = \sum_{j=1}^n (\alpha_1 \hat{y}_1 + \alpha_2 \hat{y}_2 + \cdots + \alpha_m \hat{y}_m - y_j)^2 \\
&= \sum_{j=1}^n (\alpha_1 e_{1j} + \alpha_2 e_{2j} + \cdots + \alpha_m e_{mj} - y_j)^2 = \sum_{j=1}^n x^T E_j x
\end{aligned} \tag{3.23}$$

规划模型需要进行大量运算，求解过程耗时较长，所以我们决定使用多目标粒子群优化算法来求解多元规划模型。该算法可以在以误差最小为目标函数和权重和为 1 的约束条件下，更高效地求得最优解，避免耗费过多时间和计算资源。

PSO(Particle Swarm optimization, PSO)算法又称粒子群算法，是智能优化算法的一种，如表 5-8 所示。它源于对鸟群捕食的行为研究，鸟群的任务是找到鸟巢，鸟群在寻找的过程中不断彼此传递信息来判断自己寻找到的最优解是不是整个鸟群的最优解，最终得到全局最优的位置，整个鸟群都聚集在鸟巢附近，即找到了最优解[18]。PSO 算法是一种基于迭代的优化算法，系统初始化设定为一组随机解，通过迭代搜寻找到最优值。让粒子在解空间跟随位置最优的粒子进行搜索。与遗传算法相比较，PSO 没有许多需要调整的参数，这使它更加简单容易操作。PSO 具有收敛速度快、对种群数量变化不敏感且具有良好的可扩展性的优点。

表 5-8 PSO 算法伪代码

Algorithm PSO

Step1: for step in max_steps

Step2: for i in population

Step3: initialize v_i^k, x_i^k

Step4: calculate $fitness_i^k$

Step5: if $fitness_i^k < \min \{ fitness_i^k \}, t = 0, 1, \dots, k-1$

Step6: $P_{best_i} = fitness_i^k$

Step7: if $fitness_i^k < \min \{ fitness_i^k \}, t = 0, 1, \dots, k-1, i = 0, 1, \dots, population$

Step8: $G_{best_i} = fitness_i^k$

Step9: $v_i^{k+1} = \omega v_i^k + c_1 r_1 (P_{best_i} - x_i^k) + c_2 r_2 (G_{best_i} - x_i^k)$

Step10: $x_i^{k+1} = x_i^k + v_i^{k+1}$

Step11: step = step + 1

Step12: if step > max_steps or $fitness_i^k = 0$

Step13: end

PSO 算法求解优化问题的过程中，每个优化问题的解都可以被看作搜索空间中的一只鸟，即空间中的“粒子”。粒子有速度和位置两个属性，速度代表移动的快慢，位置则代表移动的方向。PSO 算法会选择一个适应度函数来衡量它们当前位置的好坏，每个粒子都有一个适应值，增大或减小适应度的方向搜索，通过粒子间相互比较位置好坏粒子们也会会追随当前的最优粒子在解空间中搜索。PSO 算法初始化为一组随机解，通过迭代找到最优解，在每一次迭代过程中粒子通过跟随两个极值来进行自我更新。PSO 初始化为一组随机，然后通过迭代找到最优解。在每一次迭代中，粒子通过跟踪 P_{best} 和 G_{best} 这两个极值来更新自己。 P_{best} 是粒子本身所找到的最优解，它是个体极值；将个体极值与整个粒子群里的其他粒子进行共享，找到最优的个体极值作为整个粒子群的当前全局最优解 G_{best} ，它是全局极值。粒子群中的所有粒子根据自己的当前个体极值 P_{best} 和全局极值 G_{best} 进行自我更新。

PSO 算法优化求解得到的是一组粒子在设定好的空间内的位置，在本文中，每个粒子的位置代表一个检测任务的开始执行时间，通过求解粒子群的最优位置来求取检测任务的最佳开始时间。PSO 算法的速度和位置更新标准公式如下式所示：

$$v_i^{k+1} = v_i^k + c_1 r_1 (P_{best_i} - x_i^k) + c_2 r_2 (G_{best} - x_i^k) \quad (3.24)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (3.25)$$

在上述公式中， $i=1,2,\dots,N$ ，其中 N 是此种群中粒子的总数。 v_i^k 是 k 时刻第 i 个粒子的速度， x_i^k 是 k 时刻第 i 个粒子的位置。 c_1 和 c_2 是学习因子，用于衡量粒子是更倾向于向自身的最佳位置 P_{best} 移动还是更倾向于向整个种群的最佳位置 G_{best} 移动，通常取 $c_1 = c_2 = 2$ 。 r_1 和 r_2 是两个(0,1)之间的独立随机数。通常会对粒子速度 v 和位置 x 进行限制，速度 v 的最大值为 V_{max} ，最小值为 V_{min} ，位置 x 的最大值为 X_{max} ，最小值为 X_{min} 。如果 v 大于 V_{max} ，则 $v = V_{max}$ ；如果 v 小于 V_{min} ，则 $v = V_{min}$ 。位置 x 同理。

通过对标准 PSO 算法公式加入惯性因子 ω ，可以提高 PSO 算法性能，在面对不同的搜索问题时，可以通过调整 ω 来调整全局和局部寻优能力[19]。加入惯性因子 ω 的 PSO 算法速度和位置更新公式如下式所示。

$$v_i^{k+1} = \omega v_i^k + c_1 r_1 (P_{best_i} - x_i^k) + c_2 r_2 (G_{best} - x_i^k) \quad (3.26)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (3.27)$$

其中 ω 为惯性因子，当惯性因子 ω 较大时，粒子对探索空间搜索能力增强，而局部细调能力较弱；当惯性因子 ω 较小时，算法的局部搜索能力较强，而搜索新空间能力减弱。

本文中 PSO 算法的适应度函数根据误差最小的目标函数来进行设置。在算法执行过程中，首先进行随机初始化粒子群，包括设定种群大小、初始位置和初始速度。接下来将每

个粒子所在的位置带入适应度函数，计算每个粒子的适应度值。对于粒子个体，算法需要将该粒子的适应度值与其经历过的最优位置 P_{best} 的适应度值进行比较，如果此刻位置适应度更优，则将其作为当前时刻该粒子的最好位置，即个体最优解；对于种群而言，将粒子适应度值和全局所经历最优位置 G_{best} 的适应度值进行比较，如果此粒子适应度较优，则将当前粒子的位置作为全局最优位置，并更新 G_{best} 的索引号。更新 P_{best} 和 G_{best} 后，再通过公式更新每个粒子的速度和位置。每次迭代后，判断此时是否到达了最大迭代次数或是否已找到全局最优解，若满足如上终止条件，则算法结束并输出全局最优解；若不满足条件则继续计算粒子适应度，重复更新 P_{best} 和 G_{best} 以及粒子的速度和位置进行迭代，指导达到终止条件。PSO 算法流程图如图 5-7 所示。

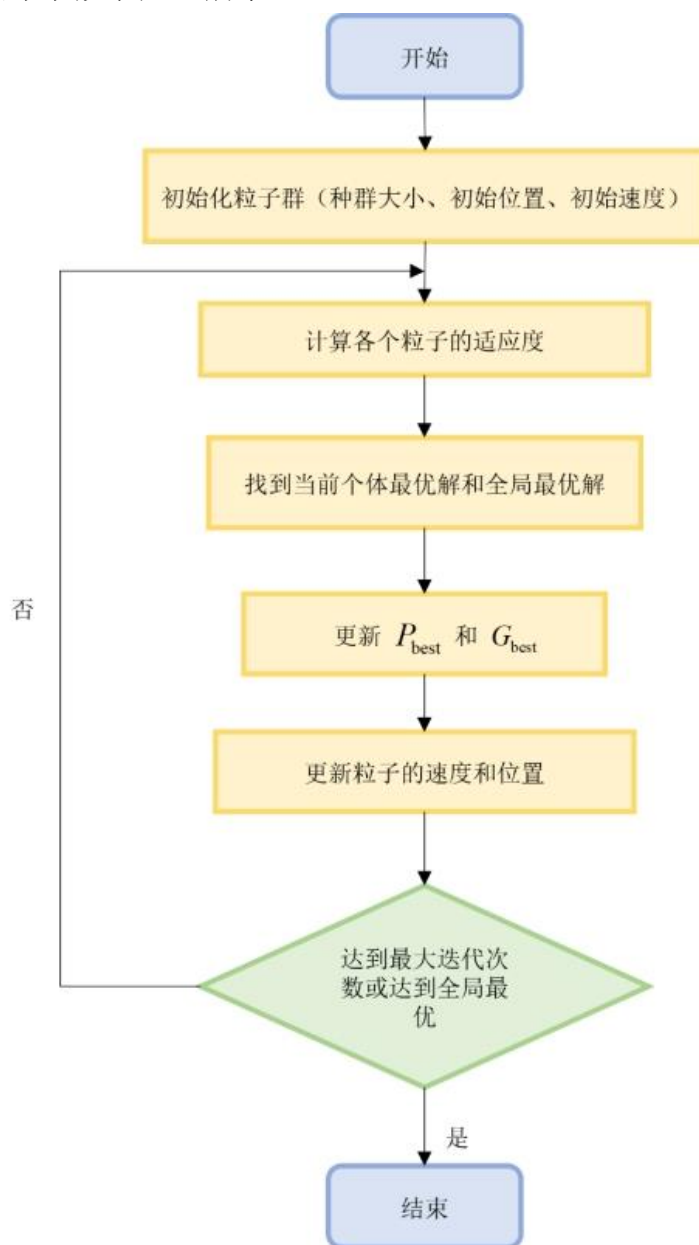


图 5-7 PSO 算法流程图

表 5-9 加权优化预测结果

年份	2021	2022	2023	2024	2025
人口	8591.621	8643.532	8695.756	8748.296	8801.153
人口-能源消费量	32889.18174	33515.43605	34145.46638	34779.30894	35416.97579
年份	2056	2057	2058	2059	2060
人口	7951.225784	7830.246394	7694.467128	7546.078094	7385.96663
人口-能源消费量	25163.45386	23703.9585	22065.91743	20275.75213	18344.16742

5.1.2 基于经济的能源消费量预测模型

此问与 5.1.1 节问题类型相同，故可建立相同的模型。5.1.2 节中分别使用线性回归模型、灰色预测模型和 ARIMA 预测模型对经济与能源消耗量进行预测。从拟合结果来看，三种模型均拟合较好，可以用于进行预测。

5.1.2.1 线性回归预测

本问中建立基于经济的能源消费量预测模型，生产总值为 GDP，能源消费总量为 E，设第 t 年生产总值为 GDP^t ，第 t 年能源消费量为 E^t ，则线性回归模型如下所示：

$$E^t = a + bGDP^t \quad (3.28)$$

将收集到数据导入进行线性回归得到表 5-10。F 检验的结果分析可以得到，显著性 P 值为 0.000***，水平上呈现显著性，拒绝回归系数为 0 的原假设，因此模型基本满足要求。对于变量共线性表现，VIF 全部小于 10，因此模型没有多重共线性问题，模型构建良好。求解得出线性回归模型为：

$$E^t = 19513.993 + 0.145GDP^t \quad (3.29)$$

为检验线性回归模型与真实数据的拟合程度，绘制了经济-能源消费量回归拟合效果图如图 5-8 所示。

建立人口-能源消费量线性回归方程后，为预测十四五至二十一五期间的能源消费量，需先预测此期间的常驻人口数量，因此需建立时间和人口之间的线性回归方程，如下所示：

$$GDP^t = a + bt \quad (3.30)$$

其中 P^t 表示第 t 年的常驻人口数量，t 表示年份。将数据导入 MATLAB 中进行线性回归分析得到年份和人口线性回归分析表如表 5-11 和表 5-12 所示。表 5-12 预测了十四五至二十一五期间的能源消费量。

由线性回归结果可以得出时间和人口之间的线性回归方程为：

$$GDP^t = -9754168.464 + 4873.306t \tag{3.31}$$

F 检验的结果分析可以得到，显著性 P 值为 0.000***，水平上呈现显著性，拒绝回归系数为 0 的原假设，因此模型基本满足要求。

对于变量共线性表现，VIF 全部小于 10，因此模型没有多重共线性问题，模型构建良好。为检验线性回归模型与真实数据的拟合程度，绘制了年份-经济回归拟合效果图如图 5-9 所示，从图 5-9 中可以看出预测值与拟合值相差较小，拟合较为合理。

表 5-10 经济-能源消费量线性回归预测表

线性回归分析结果 n=11 因变量：能源消费量									
	非标准化系数		标准化系数	t	P	VIF	R ²	调整 R ²	F
	B	标准误差	Beta						
常数	19513.993	1171.039	-----	16.664	0.000***	-----	0.886	0.873	F=69.864 P=0.000***
GDP	0.145	0.017	0.941	8.358	0.000***	1			

注：***、**、*分别代表 1%、5%、10%的显著性水平

表 5-11 年份-经济回归分析表

线性回归分析结果 n=11 因变量：GDP									
	非标准化系数		标准化系数	t	P	VIF	R ²	调整 R ²	F
	B	标准误差	Beta						
常数	-9754168.464	106334.504	-----	-91.731	0.000***	---	0.999	0.999	F=8528.057 P=0.000***
年份	4873.306	52.771	0.999	92.347	0.000***	1			

注：***、**、*分别代表 1%、5%、10%的显著性水平

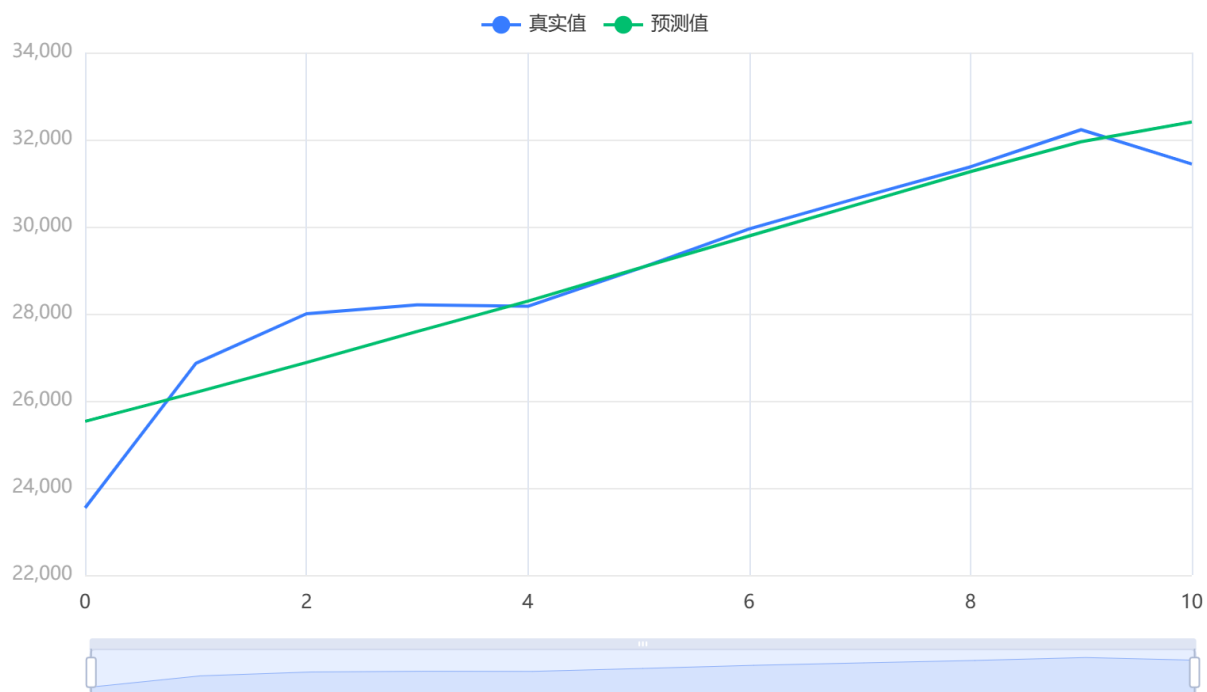


图 5-8 人口-能源消费量回归拟合效果图

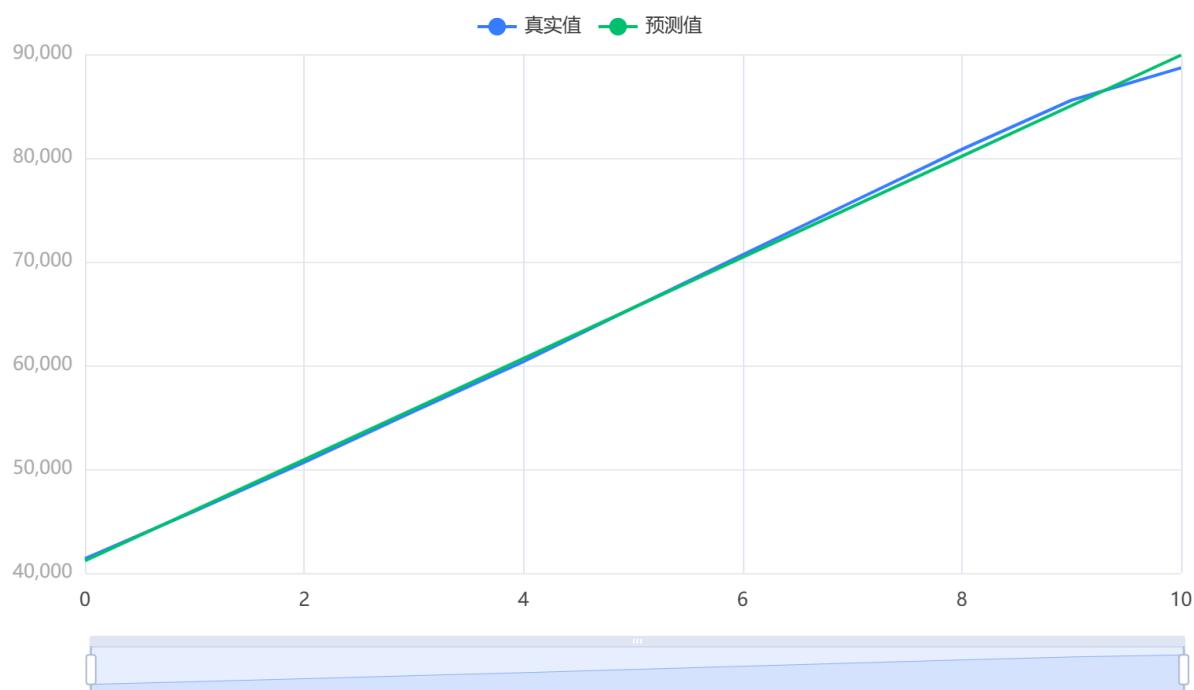


图 5-9 年份-经济回归拟合效果图

表 5-12 人口-能源消费量回归预测结果表

年份	GDP	能源消费量
2021	94782.85648	33257.50719
2022	99656.16243	33964.13655
2023	104529.4684	34670.76591
2024	109402.7743	35377.39528
2025	114276.0803	36084.02464
2026	119149.3862	36790.654
2027	124022.6922	37497.28336
2028	128895.9981	38203.91273
2029	133769.3041	38910.54209
2030	138642.61	39617.17145
2031	143515.916	40323.80081
2032	148389.2219	41030.43018
2033	153262.5279	41737.05954
2034	158135.8338	42443.6889
2035	163009.1398	43150.31826
2036	167882.4457	43856.94763
2037	172755.7516	44563.57699
2038	177629.0576	45270.20635
2039	182502.3635	45976.83571
2040	187375.6695	46683.46508
2041	192248.9754	47390.09444
2042	197122.2814	48096.7238
2043	201995.5873	48803.35316
2044	206868.8933	49509.98253
2045	211742.1992	50216.61189
2046	216615.5052	50923.24125
2047	221488.8111	51629.87061
2048	226362.1171	52336.49998
2049	231235.423	53043.12934
2050	236108.729	53749.7587
2051	240982.0349	54456.38806
2052	245855.3409	55163.01743
2053	250728.6468	55869.64679
2054	255601.9528	56576.27615
2055	260475.2587	57282.90551
2056	265348.5647	57989.53488
2057	270221.8706	58696.16424
2058	275095.1766	59402.7936
2059	279968.4825	60109.42296
2060	284841.7884	60816.05232

5.1.2.2 灰色预测模型

数据处理方式如 5.1.2 小节，如 5-13 展示了序列值和级比值。若所有的级比值都位于区间 $(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+2}})$ （即（0.846，1.181））内，说明数据适合模型构建。

表 5-13 经济-能源消费量灰色预测级比检验结果表

索引项	原始值	级比值
41383.87	23539.314	—
45952.65	26860.026	0.876
50660.2	27999.218	0.959
55580.11	28203.104	0.993
60359.43	28170.506	1.001
65552	29033.608	0.97
70665.70683	29947.977	0.969
75752.20149	30669.886	0.976
80827.71193	31373.127	0.978
85556.13387	32227.505	0.973
88683.21463	31437.998	1.025

按照白化微分方程建立 GM(1,1)模型，则可以得到预测值

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{\hat{b}}{\hat{a}})e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}}, k = 0, 1, \dots, n-1, \dots \quad (3.32)$$

而且

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), k = 0, 1, \dots, n-1, \dots \quad (3.33)$$

下表展示了发展系数、灰色作用量、后验差比值。由发展系数和灰色作用量可以构建灰色预测模型。其中发展系数表示数列的发展规律和趋势，灰色作用量反映数列的变化关系。后验差比值可以验证灰色预测的精度，后验差比值越小，则说明灰色预测精度越高。一般后验差比值 C 值小于 0.35 则模型精度高，C 值小于 0.5 说明模型精度合格，C 值小于 0.65 说明模型精度基本合格，如果 C 值大于 0.65，则说明模型精度不合格。此处后验差比值为 0.028，模型精度高。

表 5-14 经济-能源消费量灰色预测模型参数表

发展系数 a	灰色作用量 b	后验差比 C 值
-0.019	26352.943	0.028

表 5-14 展示了灰色预测模型的拟合结果表。相对误差值越小越好，一般情况下小于 20%即说明拟合良好。从表 5-15 中可以看出模型平均相对误差为 1.076%，意味着模型拟合效果良好。

表 5-15 人口-能源消费量灰色预测模型拟合结果表

索引项	原始值	预测值	残差	相对误差（%）
41383.87	23539.314	23539.314	0	0
45952.65	26860.026	27072.345	-212.319	0.79
50660.2	27999.218	27603.412	395.806	1.414
55580.11	28203.104	28144.897	58.207	0.206
60359.43	28170.506	28697.004	-526.499	1.869
65552	29033.608	29259.942	-226.334	0.78
70665.70683	29947.977	29833.923	114.054	0.381
75752.20149	30669.886	30419.163	250.724	0.817
80827.71193	31373.127	31015.883	357.243	1.139
85556.13387	32227.505	31624.309	603.196	1.872
88683.21463	31437.998	32244.671	-806.673	2.566

为了直观展示灰色预测模型拟合效果，下图给出了真实值和拟合值之间的关系，从图中可以看出，两条曲线趋势相近，拟合效果较好。图 5-10 给出了根据拟合曲线趋势，在十三五到二十一五之间能源消费量随人口的增长曲线。



图 5-10 经济-能源消费量灰色预测模型拟合图

表 5-16 灰色预测模型的预测结果表

预测阶数	预测值
1	32877.201
2	33522.14
3	34179.731
4	34850.221
5	35533.863
6	36230.917
7	36941.644
8	37666.314
9	38405.199
10	39158.578
11	39926.736
12	40709.963
13	41508.554
14	42322.811
15	43153.04
16	43999.556
17	44862.678
18	45742.731
19	46640.047
20	47554.967
21	48487.833
22	49439
23	50408.825
24	51397.674
25	52405.922
26	53433.948
27	54482.14
28	55550.894
29	56640.614
30	57751.71
31	58884.602
32	60039.718
33	61217.493
34	62418.372
35	63642.808
36	64891.263
37	66164.209
38	67462.126
39	68785.503
40	70134.841



图 5-11 经济-能源消费量灰色预测模型预测图

5.1.2.3 ARIMA 预测模型

ARIMA (Autoregressive Integrated Moving Average, 差分整合移动平均自回归模型) 又称整合移动平均自回归模型, 是一种广泛应用于时间序列分析和预测的经典统计模型 [20]。它可以通过对时间序列数据进行拟合和预测, 识别和描述时间序列数据中的趋势、季节性和随机性成分, 从而实现对未来数据走势的预测和分析。

ARIMA 模型包含三个部分: 自回归 (AR)、积分 (I) 和移动平均 (MA)。其中, 自回归部分表示当前值与历史值的相关性和依赖性; 移动平均部分表示当前值与历史误差值的相关性和依赖性; 积分部分则表示对时间序列数据的差分处理, 使其满足平稳性要求。

ARIMA 模型的核心思想是: 通过分析时间序列数据的自相关性和偏自相关性函数, 确定 ARIMA 模型中 p 、 d 、 q 三个参数的取值, 进而建立 ARIMA 模型, 对时间序列数据进行预测。其中, p 表示自回归项数, d 表示差分次数, q 表示移动平均项数 [21]。

在实际应用中, ARIMA 模型可以与其他技术结合使用, 如指数平滑法、回归分析、灰色模型等。同时, 由于 ARIMA 模型对数据平稳性的要求较高, 因此在使用 ARIMA 模型进行预测之前, 需要对原始数据进行平稳性检验和差分处理。

总的来说, ARIMA 模型在处理时间序列数据中的趋势、季节性和周期性等问题上具有较好的应用效果。它广泛应用于经济学、气象学、金融学、社会学等领域, 成为时间序列分析和预测的重要工具之一。

采用 ARIMA 模型预测时序数据, 必须是稳定的, 如果不稳定的数据, 是无法捕捉到规律的。比如股票数据用 ARIMA 无法预测的原因就是股票数据是非稳定的, 常常受政策和新闻的影响而波动, 可以使用 ADF 检验, 该检验用于稳定性检验, 使用差分分析对数据进行稳定性处理。ADF 检验表如下所示。该模型要求序列必须是平稳的时间序列数据。通过分析 t 值, 分析其是否可以显著地拒绝序列不平稳的原假设。若呈现显著性 ($P < 0.05$),

则说明拒绝原假设，该序列为一个平稳的时间序列，反之则说明该序列为一个不平稳的时间序列。临界值 1%、5%、10%不同程度拒绝原假设的统计值和 ADF Test result 的比较，ADF Test result 同时小于 1%、5%、10%即说明非常好地拒绝该假设。差分阶数本质上就是下一个数值，减去上一个数值，主要是消除一些波动使数据趋于平稳，非平稳序列可通过差分变换转化为平稳序列。AIC 值是衡量统计模型拟合优良性的一种标准，数值越小越好。临界值是对应于一个给定的显著性水平的固定值。该序列检验的结果显示，基于变量能源消费量在差分为 0 阶时，显著性 P 值为 0.007***，水平上呈现显著性，拒绝原假设，该序列为平稳的时间序列。在差分为 1 阶时，显著性 P 值为 0.511，水平上不要呈现显著性，不能拒绝原假设，该序列为不平稳的时间序列。在差分为 2 阶时，显著性 P 值为 0.146，水平上不要呈现显著性，不能拒绝原假设，该序列为不平稳的时间序列。

表 5-17 ARIMA 预测模型 ADF 检验表

ADF 检验表							
变量	差分阶数	t	P	AIC	临界值		
					1%	5%	10%
能源消费量	0	-3.522	0.007***	112.052	-4.332	-3.233	-2.749
	1	-1.546	0.511	97.206	-5.354	-3.646	-2.901
	2	-2.386	0.146	100.397	-4.665	-3.367	-2.803

注：***、**、*分别代表 1%、5%、10%的显著性水平

图 5-12 展示了原始数据 0 阶差分后的时序图。

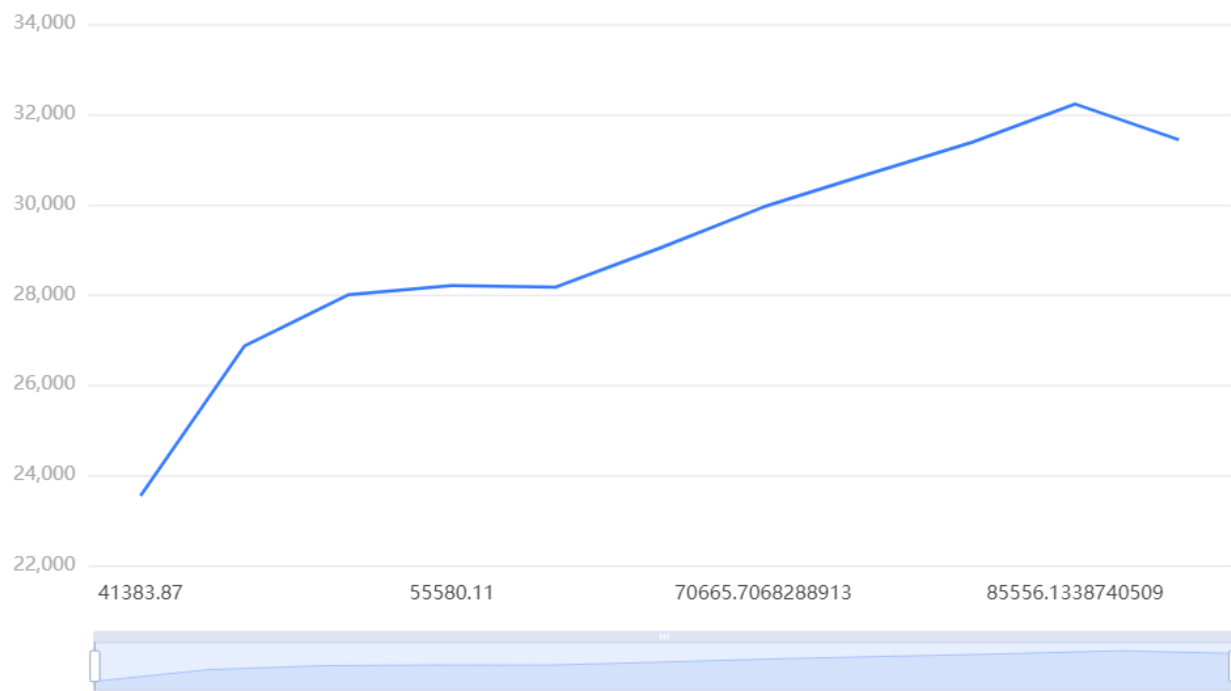


图 5-12 ARIMA 预测模型 0 阶差分时序图

图 5-13 展示了自相关图(ACF)，包括系数，置信上限和置信下限。其中横轴代表延迟数目，纵轴代表自相关系数。自相关(ACF)图在 q 阶进行截尾，偏自相关(PACF)图拖尾，ARMA 模型可简化为 MA(q)模型。倘若自相关与偏自相关图均拖尾，可结合 PACF、ACF 图中最显著的阶数(最小值)作为 p、q 值。倘若自相关与偏自相关图均截尾，可以选择更换更高的差分，或则不适合建立 ARMA 模型[22]。截尾(Truncation)是指在 ARIMA 模型中只保留残差项的部分信息，截去残差项的高频成分，使其更加平稳。例如，在 ARIMA(p, d, q)模型中，可以截去高于 q 阶的残差项，即只保留 ARMA(q)模型中的残差项。截尾是在置信区间内，ACF 或 PACF 在某阶后就恒等于零(或在 0 附近随机波动)。这样可以减少模型的复杂度，提高预测的准确性。但同时，也可能会丢失一些重要的信息。拖尾(Tail Truncation)是指在 ARIMA 模型中只保留残差项的后几项信息，以捕捉残差项的长期依赖性。拖尾是在置信区间内，ACF 或 PACF 始终有非零取值，不呈现在某阶后就恒等于零(或在 0 附近随机波动)。例如，在 ARIMA(p, d, q)模型中，可以保留 ARMA(q-1)模型中的残差项，即将 ARIMA(p, d, q)模型中的残差项作为 ARMA(q-1)模型中的自回归项使用。这样可以捕捉时间序列的长期依赖性和周期性，并提高预测的准确性。但同时，也可能会增加模型的复杂度，导致过拟合的问题。

图 5-14 展示了偏自相关图(PACF)，包括系数，置信上限和置信下限。偏自相关(PACF)图在 p 阶进行截尾，自相关(ACF)图拖尾，ARMA 模型可简化为 AR(P)模型。

图 5-15 展示了模型的残差自相关图(ACF)，包括系数，置信上限和置信下限。其中横轴代表延迟数目，纵轴代表自相关系数。相关系数均在虚线内，自回归模型(AR)残差为白噪声序列，时间序列要求模型残差为白噪声序列。

图 5-16 展示了模型的残差偏自相关图(PACF)，包括系数，置信上限和置信下限。其中存在滞后阶数的偏自相关系数超过置信区间，可以认为这个滞后阶数是显著的。

表 5-18 展示 ARIMA 预测模型相关参数系统基于 AIC 信息准则自动寻找最优参数，建立了一个 ARIMA 模型（1，1，0）来预测能源消费量。经过 Q 统计量的分析发现，Q6 的结果在水平上没有显著性，因此无法拒绝模型残差为白噪声序列的假设。同时，模型的拟合度 R² 为 0.783，表明模型的预测效果较好，基本满足要求。

表 5-18 ARIMA 预测模型参数表

ARIMA 模型（1,1,0）检验表		
项	符号	值
	Df Residuals	8
样本数量	N	11
Q 统计量	Q6 (P 值)	0.248(0.618)
信息准则	AIC	172.143
	BIC	173.05
拟合优度	R ²	0.783

注：***、**、*分别代表 1%、5%、10%的显著性水平



图 5-13 ARIMA 预测模型最终差分数数据自相关图 (ACF)

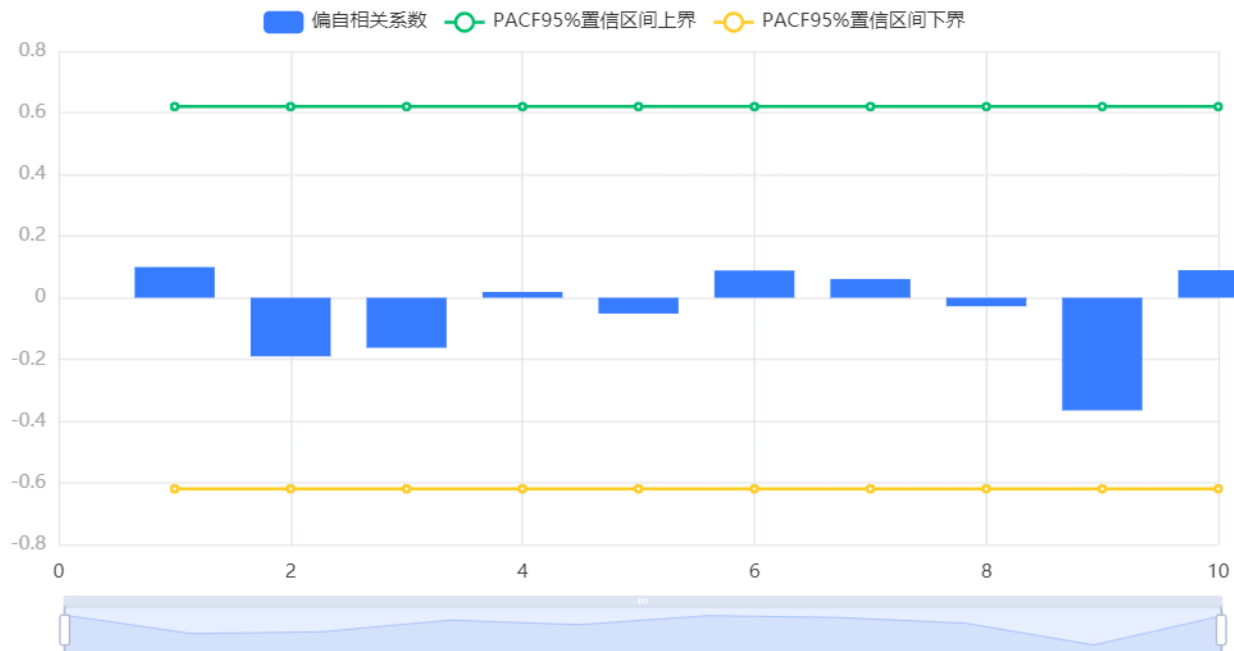


图 5-14 ARIMA 预测模型最终差分数数据偏自相关图 (PACF)

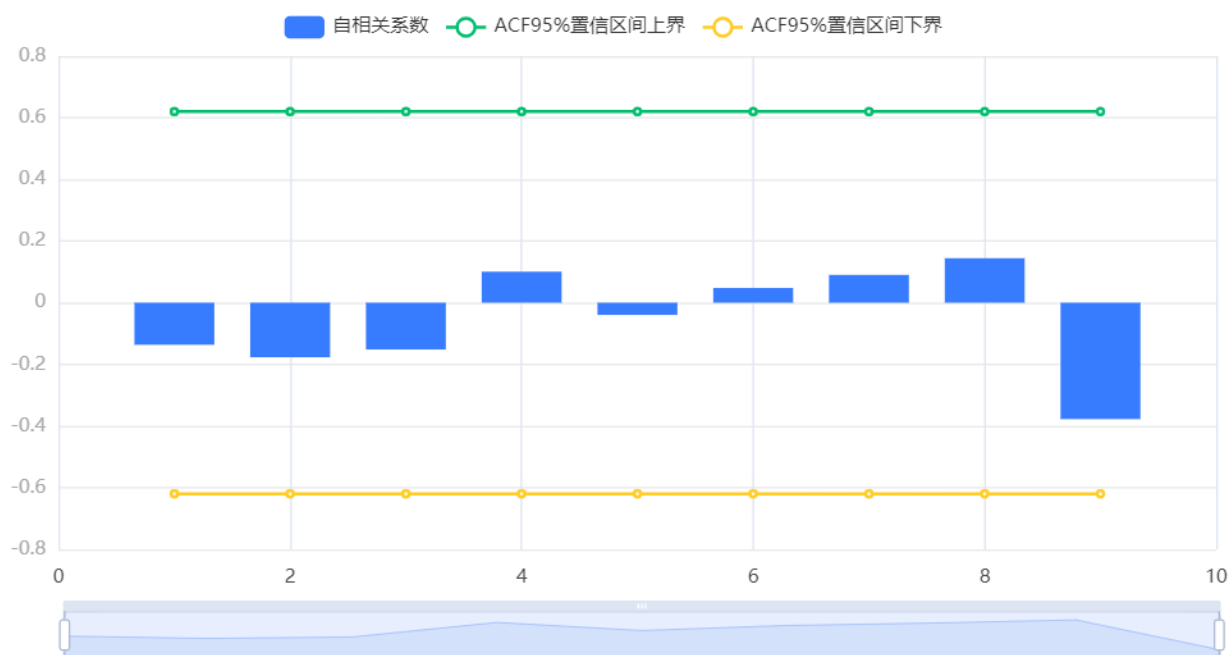


图 5-15 ARIMA 预测模型残差自相关图 (ACF)

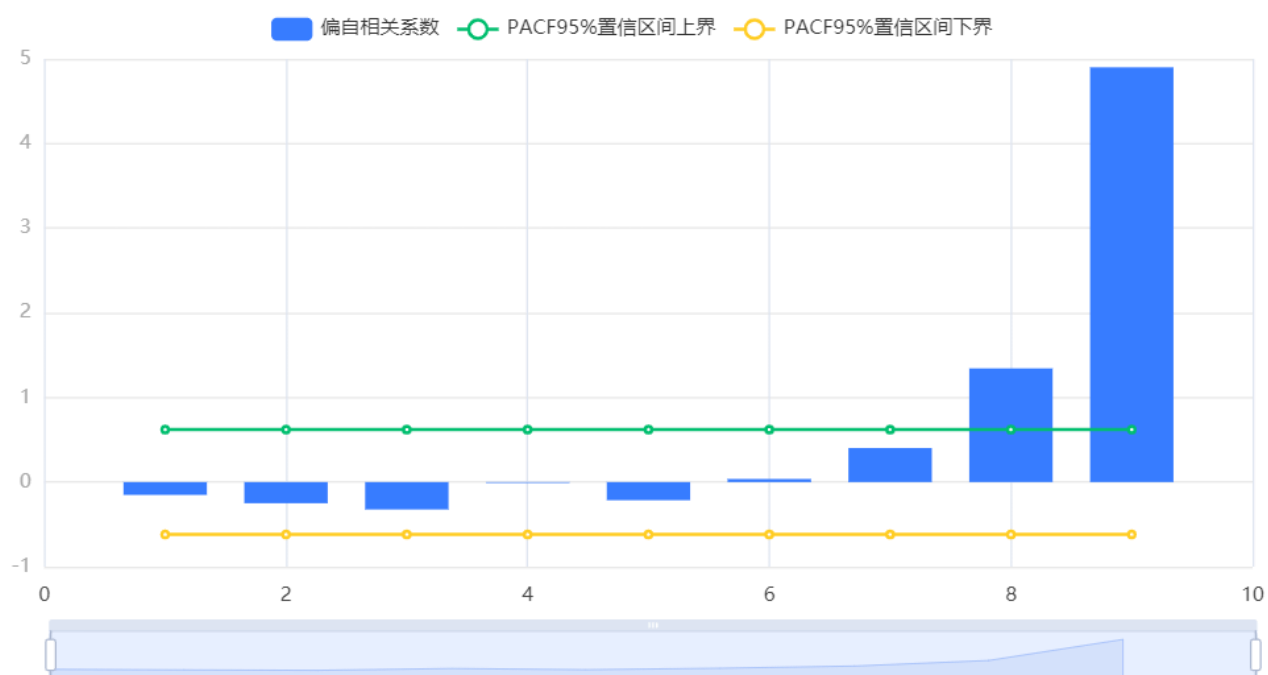


图 5-16 ARIMA 预测模型残差偏自相关图 (PACF)

表 5-19 展示本次模型参数结果，包括模型的系数、标准差，T 检验结果等，用于分析模型公式。基于变量能源消费量，系统基于 AIC 信息准则自动寻找最优参数，模型结果为 ARIMA 模型 (1, 1, 0) 检验表且基于 0 差分数据，模型公式如下：

$$E(t) = 854.33 + 0.439 \times E(t-1) \quad (3.34)$$

表 5-19 ARIMA 预测模型检验表

模型参数表						
	系数	标准差	t	$P> t $	0.025	0.975
常数	854.33	525.306	1.626	0.104	-175.251	1883.911
ar.L1.D. 能源消费量	0.439	0.578	0.76	0.447	-0.694	1.572

注：***、**、*分别代表 1%、5%、10%的显著性水平

为了直观展示 ARIMA 预测模型拟合效果,图 5-17 给出了真实值和拟合值之间的关系,从图 5-17 中可以看出,两条曲线趋势相近,拟合效果较好。图 5-18 给出了根据拟合曲线趋势,在十三五到二十一五之间能源消费量随人口的增长曲线。



图 5-17 经济-能源消费量 ARIMA 预测模型拟合图



图 5-18 经济-能源消费量 ARIMA 预测模型预测图

表 5-20 ARIMA 预测模型的预测结果表

预测值	
阶数（时间）	预测结果
1	31570.153637687847
2	32107.216869405114
3	32822.16468179049
4	33615.26108739891
5	34442.68988201936
6	35285.20164725592
7	36134.3396919083
8	36986.38880627063
9	37839.71681732964
10	38693.60667572442
11	39547.74336596225
12	40401.9884948361
13	41256.28126317833
14	42110.59496058266
15	42964.91785258247
16	43819.2447839695
17	44673.57348994783
18	45527.90297554299
19	46382.2328036409
20	47236.56278220778
21	48090.89282687897
22	48945.22290059123
23	49799.55298706187
24	50653.883079137544
25	51508.213173675635
26	52362.543269295515
27	53216.87336539065
28	54071.20346169458
29	54925.53355809024
30	55779.86365452619
31	56634.193750979844
32	57488.523847441276
33	58342.85394390613
34	59197.18404037248
35	60051.51413683949
36	60905.844233306794
37	61760.17432977422
38	62614.50442624171
39	63468.834522709214
40	64323.164619176736

5.2 区域碳排放量预测模型

由于影响碳排放量的各指标之间存在着高度相关性，为了避免传统的线性回归模型在处理这样的问题时会出现系数估计不稳定，导致过拟合的情况，本文采用岭回归模型来对碳排放量进行预测。

岭回归模型（Ridge Regression），也被称为 Tikhonov 正则化，是线性回归模型的一个变形。其核心思想是在损失函数中添加一个正则项 $\lambda \|\beta\|^2$ 来稳定系数估计，解决多重共线性的问题，此时模型的目标如下公式所示：

$$\min_{\beta} (\|Y - X\beta\|^2 + \lambda \|\beta\|^2) \quad (3.35)$$

其中， λ 是正则化参数，为一个非负值，它的选择是通过交叉验证来确定的，从而得到最小的预测误差。当 $\lambda = 0$ 时，岭回归模型就退化为普通的最小二乘法线性回归模型。当 λ 增大时，正则项的影响增强，对估计系数的约束也更强。

5.2.1 绘制岭迹图确定模型参数

岭回归模型参数 λ 的选择原则是各个自变量的标准化回归系数趋于稳定时的最小 λ 值。一般情况下， λ 值越小，偏差越小。

本文以人口、GDP、能源消费总量、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、能源供应部门、非化石能源消费比重、非化石能源发电比重为岭回归的自变量，碳排放量为岭回归的因变量，得到如下图 5-19、5-20 所示的各自变量岭迹图：

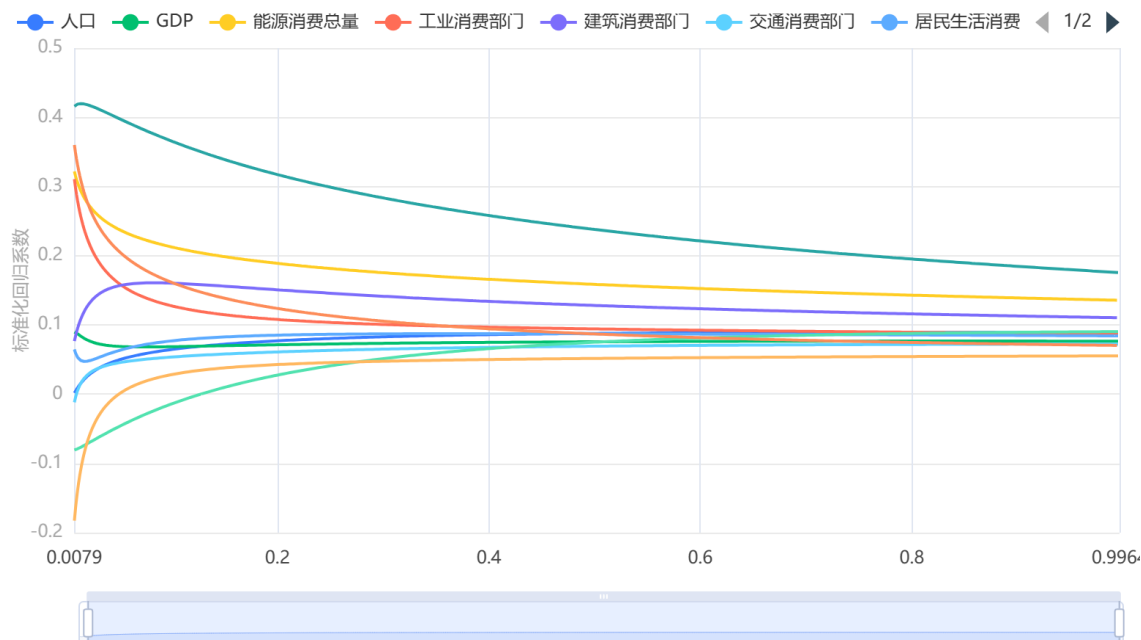


图 5-19 各自变量岭迹图（局部）



图 5-20 各自变量岭迹图（整体）

由图 5-19、5-20 可以看出，本次岭回归模型的各个自变量的标准化系数随着横坐标 λ 值的不断增大趋于稳定。因此，根据上文中提到的 λ 值的选取原则，将其确定为 0.147。

5.2.2 岭回归模型的分析结果

下表 5-21 为本次岭回归模型分析结果的集中展示：

表 5-21 岭回归模型分析结果表

因变量	自变量	非标准化系数	标准化系数	R^2	调整 R^2	F
碳排放量	常数	8515.373	—	0.983	1.167	-5.351
	人口	1.736	0.072			
	GDP	0.021	0.07			
	能源消费量	0.387	0.199			
	工业消费部门	0.363	0.115			
	建筑消费部门	4.632	0.156			
	交通消费部门	0.736	0.058			
	居民生活消费	1.128	0.082			
	农林消费部门	1.123	0.009			
	能源供应部门	2.308	0.34			
	非化石能源消费比重	9721.063	0.038			
	非化石能源发电比重	89650.488	0.14			

由上表 5-21 可知本次岭回归模型的拟合优度 R^2 为 0.983, 即说明模型表现为较为优秀。结合表 5-21 中的各自变量的回归系数, 最终得到碳排放量的预测模型公式如下:

$$\begin{aligned} \text{碳排放量 } C = & 8515.373 + 1.736 \times \text{人口 } P + 0.021 \times \text{GDP} + 0.387 \times \text{能源消费总量 } E \\ & + 0.363 \times E_{\text{工业消费部门}} + 4.632 \times E_{\text{建筑消费部门}} + 0.736 \times E_{\text{交通消费部门}} \\ & + 1.128 \times E_{\text{居民生活消费}} + 1.123 \times E_{\text{农林消费部门}} + 2.308 \times E_{\text{能源供应部门}} \\ & + 9721.063 \times \text{非化石能源消费比重 } NFF_{\text{消费比重}} \\ & + 89650.488 \times \text{非化石能源发电比重 } NFF_{\text{发电比重}} \end{aligned} \quad (3.36)$$

利用上面的公式, 对碳排放量数据进行拟合, 最终得到本次岭回归模型的原始数据图、模型拟合值如下图 5-21 所示:

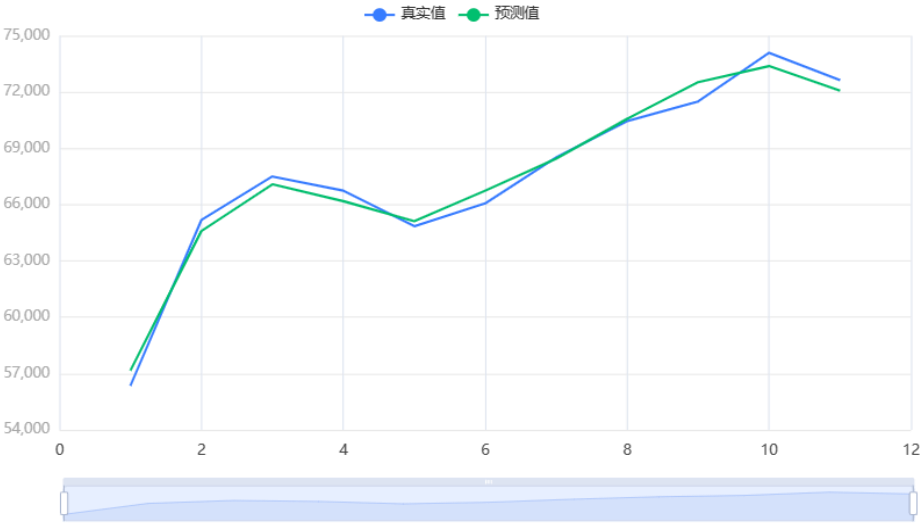


图 5-20 岭回归模型的原始数据及拟合数据图

5.2.3 加权预测模型的分析结果

基于 5.2.2 部分由岭回归模型得到的碳排放量关联公式, 利用 5.1.1.4 部分使用的加权综合预测模型, 最终得到碳排放量的预测结果如下表 5-22 所示 (由于数据较多, 下表仅展示了 2021~2025 年以及 2056~2060 年份的碳排放量预测结果):

表 5-22 加权综合模型预测结果表

年份	碳排放量	年份	碳排放量
2021	78473.42	2056	67078.84
2022	76147.76	2057	66974.95
2023	77751.05	2058	66731.63
2024	74155.61	2059	66415.67
2025	75245.46	2060	65589.38

6. 问题三模型建立与求解

6.1 情景设计

根据问题要求，我们设计了四种情景如下：



图 6-1 区域双碳目标四种情景

- 1.自然情景：碳排放和经济增长不进行特别政策干预，未来发展按照当前趋势进行。
- 2.基准情景：这个情景按照国家碳达峰和碳中和的时间节点（2030 年碳达峰，2060 年碳中和）设定。这个情景下，制定政策和措施来确保在规定的时间内达到这些目标，模拟达到目标的路径。
- 3.雄心情景：设定更积极的碳达峰和碳中和目标，采取更积极的政策干预。提前碳达峰碳中和时间。
- 4.滞后情景：按时完成碳达峰目标后未及时采取降碳减排措施，导致碳中和目标之后完成。

6.2 多情景下碳排放量核算方法

首先明确题目中所给出的基本假设：

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

假设 2：2060 年生态碳汇的碳消纳量为基期碳排放量的 10%；

假设 3：2060 年工程碳汇或碳交易的碳消纳量为基期碳排放量 10%。

根据假设可以利用插值法推算出 2030 年和 2060 年 GDP 数值。

6.2.1 自然情景下（碳达峰与碳中和）目标与路径预测

Newton 插值法是多项式插值的一种方法，其基于差商的概念。由于该方法可以逐点扩展，所以适合应用在诸如本题中需要多次插值的场合。

Newton 插值法的关键是差商的概念。对于一组点 $(x_0, y_0)(x_1, y_1) \dots (x_n, y_n)$ ，定义差商为：

零阶差商：

$$f[x_i] = y_i \quad (4.1)$$

一阶差商：

$$f[x_i, x_{i+1}] = \frac{f[x_{i+1}] - f[x_i]}{x_{i+1} - x_i} \quad (4.2)$$

二阶差商：

$$f[x_i, x_{i+1}, x_{i+2}] = \frac{f[x_{i+1}, x_{i+2}] - f[x_i, x_{i+1}]}{x_{i+2} - x_i} \quad (4.3)$$

N 阶差商：

$$f[x_0, x_1, \dots, x_n] \quad (4.4)$$

基于上述差商定义，n 阶 Newton 插值多项式为：

$$P_n(x) = f[x_0] + (x - x_0)f[x_0, x_1] + (x - x_0)(x - x_1)f[x_0, x_1, x_2] + \dots \quad (4.5)$$

结合假设 1，利用 MATLAB 编写 Newton 插值代码，来对 2020~2060 年的 GDP 数据进行预测。

首先得到 2 阶 Newton 插值多项式如下：

$$\begin{aligned} P_2 = & (1625137101350983 \times x) / 274877906944 \\ & + (4160350979458515 \times (x - 2020) \times (x - 2035)) \times 140737488355328 \\ & - 820694236182246415 / 68719476736 \end{aligned} \quad (4.6)$$

最终得到 2020-2060 年的 GDP 曲线如下图 6-2 所示。

根据 5.1.2 部分的经济与能源消费量关系式

$$E^t = 19513.993 + 0.145GDP^t \quad (4.7)$$

已知 2020~2060 年份的 GDP 数据，可以利用该式得到 2020~2060 年份的能源消费量数据，如图 6-3 所示。

基于上述能源消费量数据，再通过 5.1.1 部分的人口与能源消费量关系式 $E^t = -70760.134 + 12.064P^t$ 可以得到 2020~2060 年份的人口数据，如图 6-4 所示。

根据人均 GDP 定义：

$$g = \frac{G}{P} \quad (4.8)$$

得到 2020-2060 年的人均 GDP 数据如图 6-5 所示。

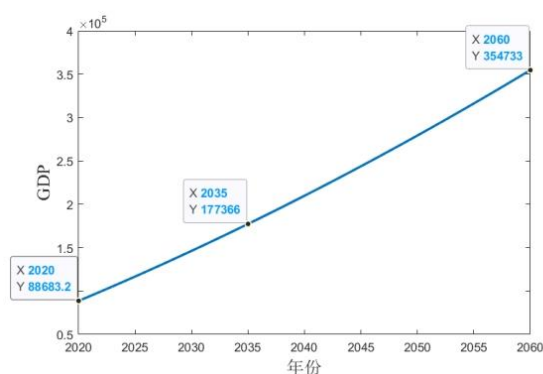


图 6-2 2020-2060 年的 GDP 曲线

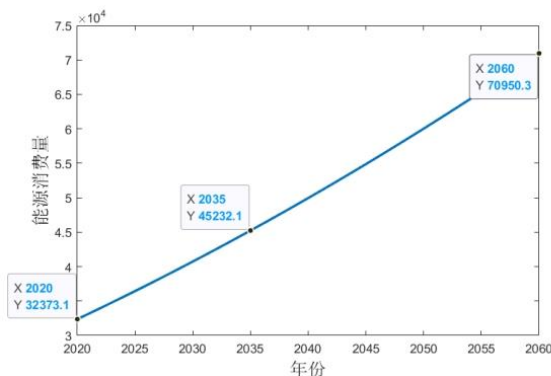


图 6-3 2020-2060 年的能源消费量曲线

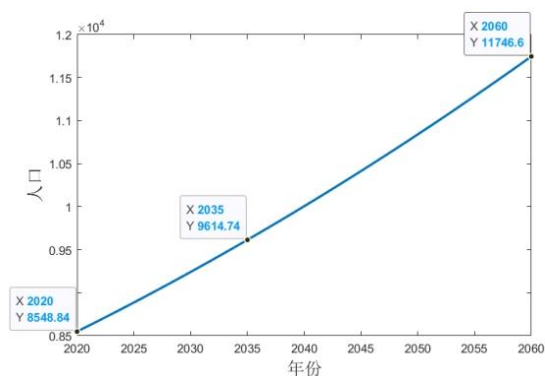


图 6-4 2020-2060 年的人口曲线

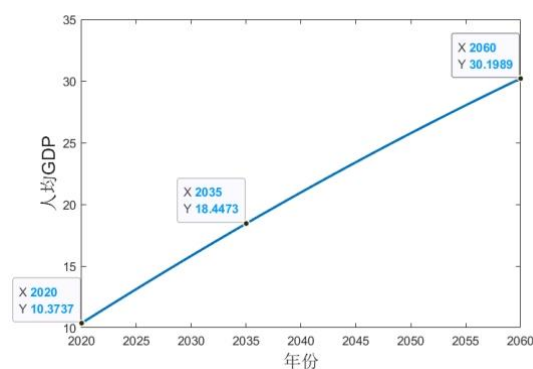


图 6-5 2020-2060 年的人均 GDP 曲线

根据单位 GDP 能耗定义：

$$e = \frac{E}{G} \quad (4.9)$$

得到 2020~2060 年的单位 GDP 能耗数据如下图 6-6 所示：

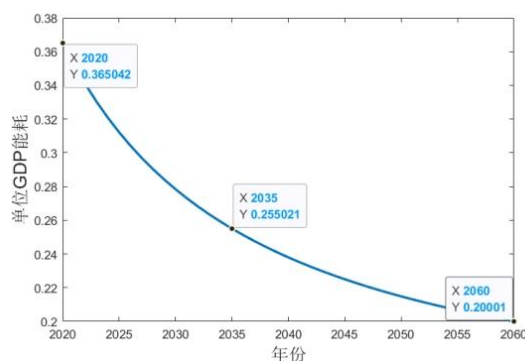


图 6-6 2020-2060 年的单位 GDP 能耗曲线

通过题目给出的数据可以通过计算得到 2010-2020 年的单位能耗二氧化碳排放量的数据，公式为：

$$c = \frac{C}{E} \tag{4.10}$$

进而通过线性回归预测出 2020-2060 年份的单位能耗二氧化碳排放量数据，如图 6-7 所示：

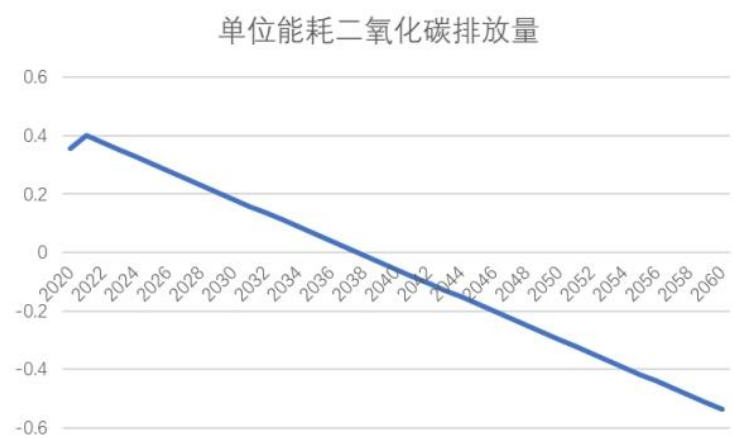


图 6-6 2020-2060 年的单能耗二氧化碳排放量曲线

可以看到大约在 2040 年左右单位能耗二氧化碳排放量接近为 0，因此可以认为自然情境下碳达峰大约在 2040 年可以实现。

6.2.2 多情景下碳排放量核算

根据情景设计理念，四个情景在能效提升和非化石能源比重两个指标特点如表 6-1 所示。同时根据 6. 2. 1 节结论设计四个情景预期碳达峰碳中和时间。

表 6-1 四个情景对应能源利用效率和提高非化石能源消费比重指标特点

情景	能效提升	非化石能源比重	预期碳达峰时间	预期碳中和时间
自然	较少	较慢	2040	2060
基准	适中	逐步	2030	2050
雄心	强烈	较快	2025	2045
滞后	较慢	很慢	2045	2070

因此可以根据特点设计不同的曲线参数，得到四个场景碳排放量核算结果如图 6-7 所示。

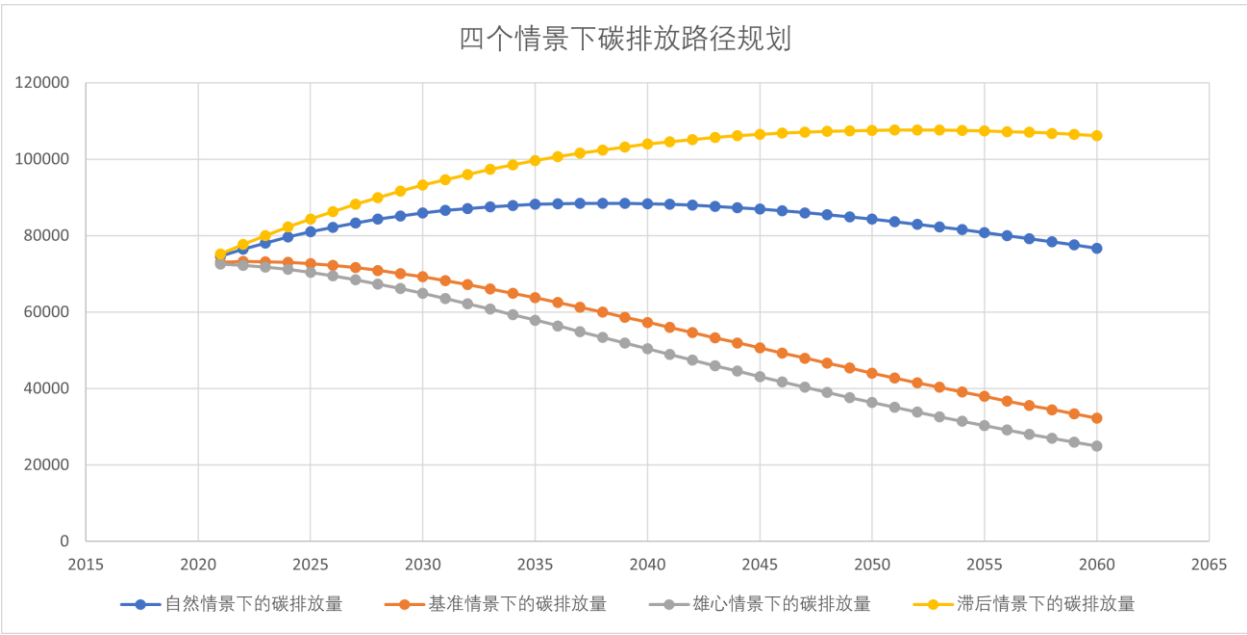


图 6-7 多情景下碳排放量核算结果

6.3 双碳目标与路径规划

6.3.1 GDP、人口和能源消费量目标值

由于 GDP 及人口目标不作为碳排放的因变量，而是影响变量。因此实际值不受情景设计影响，直接利用前面算出的模型预测目标值即可。但不同情景的人口及 GDP 增长率不同，因而能源消费量与碳排放目标相关，因此区分四个不同场景下的能源消费目标。计算结果如表 6-2。

表 6-2 四个情景对应 GDP、人口和能源消费量目标值

年份	GDP 目标	人口目标	自然能源消 费目标	基准能源消 费目标	雄心能源消 费目标	滞后能源消 费目标
2025	116766.23	8696.138	38771.8491	38575.8341	39364.68	38380.61
2030	146327.3	8914.544	45510.3715	45051.3701	46912.74	44596.54
2035	177366.42	9132.95	51670.4442	50890.724	54077.03	50122
2050	279352.11	9788.166	66877.5021	64874.3339	73252.31	62929.22
2060	354732.84	10224.98	74507.6743	71547.0116	84124.54	68701.17

6.3.2 能源利用效率和提高非化石能源消费比重的目标值

通过利用十二五、十三五时期数据，可以计算出 2010-2020 年非化石能源占比如图 6-X 所示。

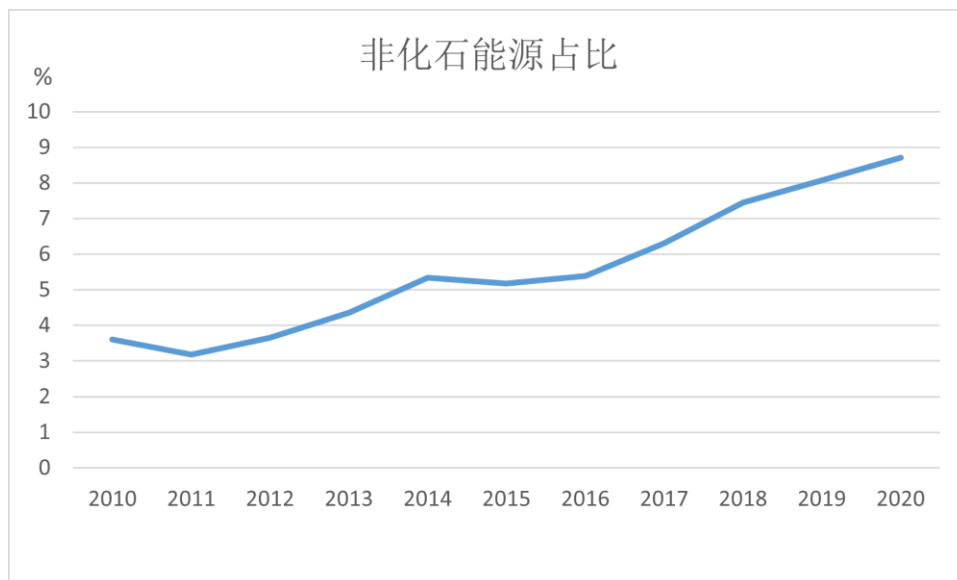


图 6-7 2010-2020 年的非化石能源占比曲线

因此可以计算出自然情景非化石能源消费比重 $NFF_{\text{比重}}$ 以及自然情景非化石能源消费比重增长率 $NFF_{\text{比重增长率}}$ 。有如下递推公式：

$$\begin{aligned} NFF_{\text{比重增长率}}(i) &= (1 - NFF_{\text{占比增加因子}}(i) \times (1 - NFF_{\text{占比}}(i))) \\ NFF_{\text{占比}}(i+1) &= NFF_{\text{占比增长率}}(i) \times NFF_{\text{占比}}(i) \end{aligned} \quad (4.11)$$

同时根据能源利用效率计算公式：

$$n_e = \frac{1}{gdp} = \frac{G}{E} \quad (4.12)$$

因此可以分别计算出四个场景非化石能源消费比重的目标值、非化石能源消费比重增长率的目标值、能源消费量的目标值和能源利用效率的目标值如表 6-3 所示。

表 6-3 四个情景对应提高能源利用效率和提高非化石能源消费比重的目标值

年份	非化石能源消费 比重目标值	非化石能源消费比重 增长率的目标值	能源消费量的目标值	提高能源效率的目 标值
自然情景				
2025	0.17480381	1.106611871	38771.84911	0.332046776
2030	0.254088002	1.063729195	45510.37149	0.311017647
2035	0.325754633	1.044103689	51670.44423	0.291320331
2050	0.502023204	1.020661943	66877.50207	0.239402171
2060	0.593116702	1.014198968	74507.67435	0.21003884
基准情景				
2025	0.255640045	1.138074532	38575.83409	0.33036808
2030	0.393069215	1.068760544	45051.3701	0.307880834
2035	0.505125209	1.042558415	50890.724	0.286924233
2050	0.731735116	1.015512579	64874.33388	0.232231407
2060	0.82164875	1.009126926	71547.01158	0.201692664
雄心情景				
2025	0.29360948	1.144984202	39364.67983	0.33712385
2030	0.453408481	1.067746657	46912.74006	0.320601427
2035	0.577057902	1.040122925	54077.02911	0.304888767
2050	0.804054635	1.012992783	73252.30607	0.262222131
2060	0.882680272	1.007044704	84124.53658	0.237148995
滞后情景				
2025	0.13183393	1.071258152	38380.61266	0.32869618
2030	0.174382705	1.050225314	44596.53931	0.304772523
2035	0.214846168	1.038328949	50121.9966	0.282590113
2050	0.324721887	1.021456344	62929.22429	0.225268476
2060	0.389290579	1.01610136	68701.16803	0.193670166

6.3.3 能效提升、产业（产品）升级、能源脱碳和能源消费电气化的定性分析

(1) 能效提升

定性分析方面能效提升可以减少能源消耗、促进清洁能源转型和刺激技术创新和产业升级。能效提升意味着在生产和使用过程中更有效地利用能源，减少能源浪费。通过采用节能技术、改进生产工艺和优化设备等方式，可以降低碳排放。这将有助于减少温室气体的排放，缓解气候变化的影响。此外，能效提升可以降低能源消耗，减少对传统高碳能源的依赖。促进清洁能源的应用，如太阳能和风能，将减少化石燃料的使用，从而降低碳排放。能效提升需要引入先进技术和设备，推动产业结构升级和转型。这将鼓励企业进行技术创新，开发更高效的产品和生产方式，进一步减少碳排放。

定量分析方面，通过衡量单位产出所需的碳排放量来评估能效提升的效果。常用的指标包括碳排放强度（单位产值或单位能源消耗的碳排放量）和碳排放系数（单位产出所需的碳排放量）。通过与基准数据进行比较，可以定量评估能效提升对碳排放量的影响。对能源利用效率进行改善，通过评估单位能源消耗下的产出增加，可以定量衡量能效提升的效果。通常使用能源利用效率指标，如能源转换效率、能源利用率和能源消耗单位产出指标等，来评估能效改进的程度。可以对碳排放削减量计算，通过比较能效提升前后的碳排放量，可以计算出由能效提升带来的碳排放削减量。这可以通过实际监测数据、能源消耗情况和碳排放因子等进行定量分析。

(2) 产业（产品）升级

定性分析方面，能源脱碳意味着减少或消除对高碳化石燃料的依赖，转向更清洁的能源来源，如可再生能源（太阳能、风能、水能等）和核能。这将有助于降低碳排放，减缓气候变化的影响。能源脱碳推动技术创新和能效提升，优化能源利用方式，减少能源浪费。通过改进生产工艺、采用节能设备、推广能源管理系统等措施，能够在单位能源消耗下获得更多的产出，降低碳排放。同时推动能源结构转型。能源脱碳鼓励减少化石燃料在能源结构中的比重，增加可再生能源的占比。通过增加可再生能源的装机容量、建设更多的清洁能源发电厂和优化能源供应系统，可以减少碳排放并实现能源结构的转型。

定量分析方面，可以对碳排放削减量计算，通过对能源脱碳前后的碳排放量进行比较，可以定量计算由能源脱碳带来的碳排放削减量。这需要考虑能源转换效率、可再生能源装机容量、能源消耗情况和碳排放因子等因素，以实际数据进行准确计算。可以进行能源替代效果评估，能源脱碳将化石燃料替代为更清洁的能源来源。通过分析替代效果，比如可再生能源替代传统能源所减少的碳排放量，可以定量评估能源脱碳对碳排放的影响。对碳排放强度指标分析，通过衡量能源脱碳后单位产出所需的碳排放量，如碳排放强度和碳排放系数，可以定量评估能源脱碳的效果。与基准数据进行比较，可以进一步衡量能源脱碳对碳排放量的减少程度。

(3) 能源脱碳

定性分析方面，可以进行技术创新和能效提升、清洁能源使用、循环经济和可持续发展。产品升级和产业升级通常伴随着技术创新和能效提升。通过引入更高效的生产技术、设备和工艺，可以减少能源消耗和物质浪费，从而降低碳排放。这样的升级有助于改进产品的制造过程，提高资源利用效率，减少环境污染。升级过程中，可以推动清洁能源的应用，如太阳能、风能、水能等。替代传统的高碳能源，减少化石燃料的使用，是减少碳排放的重要途径之一。产品和产业升级可以促进清洁能源技术的采用，实现能源结构的转型。产品升级和产业升级可以推动循环经济和可持续发展理念的实施。通过优化材料选择、产品设计和废弃物处理等方面，减少资源消耗和环境负荷。采用循环利用和再生利用的方式，降低碳排放，减少对自然资源的依赖。

定量分析方面，可以对碳强度指标、资源利用效率改善、碳排放削减量进行计算。碳强度指标是通过衡量单位产出所需的碳排放量来评估产品升级和产业升级的效果。常用的指标包括碳排放强度（单位产值或单位能源消耗的碳排放量）和碳排放系数（单位产出所需的碳排放量）。通过与基准数据进行比较，可以定量评估升级对碳排放量的影响。资源利用效率改善是通过评估单位资源消耗下的产出增加，可以定量衡量产品升级和产业升级的效果。这可以通过资源利用效率指标，如精益生产、资源利用率和废品率等，来评估升级改进的程度。碳排放削减量计算是通过比较升级前后的碳排放量，可以计算出由产品升级和产业升级带来的碳排放削减量。这可以通过实际监测数据、能源消耗情况和碳排放因子等进行定量分析。

(4) 能源消费电气化

定性分析方面，可以从能源替代效果、能效提升、结构优化方面来分析。电气化推动传统能源的替代，如可再生能源等。这有助于减少化石燃料的使用，降低对环境的压力，减少碳排放。电气化可以促进能效的提升，通过改善设备和系统的设计、制造和使用，减少能源消耗和物质浪费，从而减少碳排放。例如，智能家居系统可以准确感知用户需求，控制家电的使用时间和能耗，实现节能减排。电气化改变能源产业结构，通过推动清洁能源和高效设备的应用，促进能源供给侧结构的优化，从而减少碳排放。例如，发展太阳能和风能等清洁能源，可以大大提高能源结构的清洁度。

定量分析方面，可以对能源消耗量、碳强度指标和碳排放削减量进行计算。能源消耗量计算方面，电气化可以减少传统能源的消耗，降低单位能源消耗下的碳排放量。通过能源消耗量的分析，可以定量评估电气化对碳排放的影响。碳强度指标分析方面，通过衡量电气化后的单位产出所需的碳排放量，如碳排放强度和碳排放系数，可以定量评估电气化对碳排放量的影响。与基准数据进行比较，可以进一步衡量电气化对碳排放量的减少程度。碳排放削减量计算方面，通过比较电气化前后的碳排放量，可以计算出由电气化带来的碳排放削减量。这需要考虑到电气化过程中的能源消耗情况、碳排放因子等因素，以实际数据进行准确计算。

7. 模型评价

7.1 模型的优势

(1) 数据处理完善全面：在解决问题之前，我们首先对数据进行了细致全面的处理和分析，先对数据进行预处理，进而对数据进行分类整理，提取出需要分析的关键数据，进而对数据进行可视化处理，通过绘制直方图、折线图、饼图、桑基图等来对数据进行展示和分析，使数据更加清晰，有助于后续建模和提取关键指标。

(2) 问题分析细致：对问题背景进行细致分析，从经济、人口、能源消耗量以及产业结构等多个方面对碳排放量进行分析，分析并预测了能效提升、产业（产品）升级、能源脱碳和能源消费电气化等重点工程对碳排放的影响。

(3) 问题假设全面：设置多个假设前提对问题进行约束，避免因模型太复杂而导致处理困难。

(4) 预测模型种类多样：对于基于人口的能源消费量预测，本文使用线性回归预测、灰色预测、LSTM 预测和加权优化预测四种预测模型，对十三五到二十一五期间的人口和能源消费量进行了预测，并将多种模型进行对比分析。在其中利用加权平均的方法综合考虑各模型的结果，有助于减小模型预测的误差。对于基于经济的能源消费量预测，本文使用线性回归预测、灰色预测和 ARIMA 预测三种预测模型，对十三五到二十一五期间的经济和能源消费量进行了预测，多种模型的构建可以更全面的对问题进行分析，指导实际应用中模型的选择。

(5) 情景设计全面：本文设置了自然、基准、雄心、滞后四种情景，从碳达峰到达时间的各种可能性，对情景分析较为全面。

(6) 路径规划：本文提出了一种综合性的路径规划，旨在解决碳减排问题并实现碳达峰和碳中和目标。为了达成这一目标，我们探讨了政策、技术和产业结构等多个方面的措施。

总而言之，本文综合考虑了各种方面因素，并对给定数据进行了全面的处理和分析，采用了多种方法模型，系统地解决了双碳政策预测和路径规划的问题，是一个非常有意义和参考价值的建模。

7.2 模型的不足

(1) 没有考虑特殊情况：本文对情景进行了假设，忽略了突发因素对碳排放以及经济、人口、能源消耗量等因素的影响，在实际应用中可能会导致预测不准确。

(2) 情景设计复杂：情景设计是一个综合考虑政策、技术和发展路径等多个因素的复杂任务。当设计情景时，我们需要确保所构建的情景具有合理性、可行性，并能够为政策制定提供有价值的建议。为确保情景设计的准确性和有效性，我们需要依据可靠的数据、可信的模型和科学的方法来进行分析和预测。同时，还应该充分考虑利益相关者的观点和反馈，使设计的情景更具参考价值。

(3) 个别有待细化：在路径规划和政策建议领域，更多的关注应该放在细节和实施问题上。具体而言，需要对政策和技术措施的实施方式、监测方法以及评估指标等方面展开更加详细的讨论。这样做可以确保政策和技术措施能够顺利实施，并对其效果进行全面的监测和评估。

7.3 模型的应用前景

(1) 预测和管理碳排放：通过建立碳排放预测模型，可以对碳排放进行实时监测和管理。这可以帮助政府、组织和企业及时发现和解决碳排放问题，并制定相应的政策和措施。

(2) 政策制定与实施：区域双碳目标的实现需要制定和实施一系列政策和措施。路径规划方法可以帮助政策制定者评估不同政策的实施效果，预测碳排放的变化趋势，并为政策制定提供科学依据，确保政策的可行性和有效性。

(3) 推动企业可持续发展：碳排放预测模型可以帮助企业深入了解其碳排放数量和趋势，并为企业提供准确的排放量数据。这能够推动企业朝着更加清洁、低碳的方向发展，实现可持续发展。

(4) 能源转型与规划：能源领域是碳排放的主要来源之一，实现区域双碳目标需要进行全面的能源转型。路径规划方法可以帮助确定能源转型的路径和时间表，评估不同能源结构的碳排放影响，优化能源规划，并提供针对性的能源政策建议。

(5) 城市规划与建设：城市是碳排放集中的区域，实现区域双碳目标需要在城市规划与建设中考考虑碳排放因素。路径规划方法可以帮助城市规划者评估不同城市发展方案的碳排放影响，优化城市布局和交通规划，推动低碳建筑和可持续交通等绿色技术的应用。

(6) 产业转型与升级：实现区域双碳目标需要进行产业结构的调整和转型升级。路径规划方法可以帮助产业决策者评估不同产业政策的碳排放效果，发掘新兴低碳产业的发展机会，促进绿色技术与创新的应用，并提供产业转型的路径和策略建议。

8. 参考文献

- [1] 唐杰,崔文岳,温照杰等.基于 Kaya 模型的碳排放达峰实证研究[J].深圳社会科学,2022,5(03):50-59.
- [2] 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(03):1-15.
- [3] 杨小勇.方差分析法浅析——单因素的方差分析[J].实验科学与技术,2013,11(01):41-43.
- [4] 唐诵六.数据分布类型检验及其在土壤学中的应用 II.Shapiro—Wilk W 检验法及其计算机程序[J].土壤,1984(03):112-115.
- [5] 许士春,习蓉,何正霞.中国能源消耗碳排放的影响因素分析及政策启示[J].资源科学,2012,34(01):2-12.
- [6] 张友国.碳达峰、碳中和工作面临的形势与开局思路[J].行政管理改革,2021(03):77-85.DOI:10.14150/j.cnki.1674-7453.2021.03.005.
- [7] 徐翔.碳达峰和碳中和——物流业面临深刻变革[J].中国储运,2021(08):30-31.DOI:10.16301/j.cnki.cn12-1204/f.2021.08.005.
- [8] 徐维超.相关系数研究综述[J].广东工业大学学报,2012,29(03):12-17.
- [9] 谢文华. Spearman 相关系数的变量筛选方法[D].北京工业大学,2015.
- [10] 贾科,杨哲,魏超等.基于斯皮尔曼等级相关系数的新能源送出线路纵联保护[J].电力系统自动化,2020,44(15):103-111.
- [11] 王大荣,张忠占.线性回归模型中变量选择方法综述[J].数理统计与管理,2010,29(04):615-627.DOI:10.13860/j.cnki.sljt.2010.04.003.
- [12] 谢乃明,刘思峰.离散 GM(1,1)模型与灰色预测模型建模机理[J].系统工程理论与实践,2005(01):93-99.
- [13] 杨华龙,刘金霞,郑斌.灰色预测 GM(1,1)模型的改进及应用[J].数学的实践与认识,2011,41(23):39-46.
- [14] 张大海,江世芳,史开泉.灰色预测公式的理论缺陷及改进[J].系统工程理论与实践,2002(08):140-142.
- [15] 王鑫,吴际,刘超等.基于 LSTM 循环神经网络的故障时间序列预测[J].北京航空航天大学学报,2018,44(04):772-784.DOI:10.13700/j.bh.1001-5965.2017.0285.
- [16] 宋刚,张云峰,包芳勋等.基于粒子群优化 LSTM 的股票预测模型[J].北京航空航天大学学报,2019,45(12):2533-2542.DOI:10.13700/j.bh.1001-5965.2019.0388.
- [17] 刘海波,苏炳凯,项静恬.最优加权组合预测方法在气候预测中的应用研究[J].气象,1999(11):20-24.
- [18] Kennedy J., Eberhart R. Particle swarm optimization[C]. Proceedings of ICNN'95-international conference on neural networks, 1995: 1942-1948.
- [19] Shi Y., Eberhart R. C. Empirical study of particle swarm optimization[C]. Proceedings Of the 1999 Congress on Evolutionary Computation-CEC99, 1999: 1945-1950.
- [20] 韩超,宋苏,王成红.基于 ARIMA 模型的短时交通流实时自适应预测[J].系统仿真学报,2004(07):1530-1532+1535.
- [21] 薛可,李增智,刘浏等.基于 ARIMA 模型的网络流量预测[J].微电子学与计算机,2004(07):84-87.DOI:10.19304/j.cnki.issn1000-7180.2004.07.022.
- [22] 孙红果,邓华.样本自相关系数与偏自相关系数的研究[J].蚌埠学院学

报,2016,5(01):35-39.DOI:10.13900/j.cnki.jbc.2016.01.008.

9. 附录

折线图代码：（java）

```
option = {
  title: {
  },
  tooltip: {
    trigger: 'axis'
  },
  legend: {
    data: ['碳排放总量', '第一产业碳排放总量', '第二产业碳排放总量', '第三产业碳排放总量', '居民生活碳排放总量']
  },
  grid: {
    left: '3%',
    right: '4%',
    bottom: '3%',
    containLabel: true
  },
  toolbox: {
    feature: {
      saveAsImage: {}
    }
  },
  xAxis: {
    name: '年',
    type: 'category',
    boundaryGap: false,
    data: ['2010', '2011', '2012', '2013', '2014', '2015', '2016', '2017', '2018', '2019', '2020']
  },
  yAxis: {
    name: '碳排放量/万tCO2',
    type: 'value'
  },
  series: [
    {
      name: '碳排放总量',
      type: 'line',
      data:
[56360.052, 65193.342, 67502.613, 66749.376, 64853.276, 66074.810, 68526.125, 70451.557, 71502.003, 74096.331, 72633.324]
    }
  ]
}
```

```

    },
    {
        name: '第一产业碳排放总量',
        type: 'line',
        data:
[896.07, 1031.176, 1165.275, 1007.478, 1020.853, 1162.512, 1211.034, 1245.019, 1295.4
87, 1278.384, 1238.759
]
    },
    {
        name: '第二产业碳排放总量',
        type: 'line',
        data:
[45225.697, 52975.787, 54048.282, 52229.084, 51187.98, 51101.873, 52382.224, 52975.8
49, 52506.881, 54235.439, 52954.049
]
    },
    {
        name: '第三产业碳排放总量',
        type: 'line',
        color: 'rgba(255, 127, 0, 0.5)',
        data:
[5898.284, 6584.556, 7147.774, 7791.472, 7676.077, 8314.531, 8801.016, 9524.015, 1042
2.113, 11150.957, 10906.028
]
    },
    {
        name: '居民生活碳排放总量',
        type: 'line',
        color: 'rgba(0, 128, 255, 0.5)',
        data:
[4340.001, 4601.822, 5141.282, 5721.341, 4968.367, 5495.894, 6131.851, 6706.674, 7277
.522, 7431.551, 7534.489
]
    }
]
};

```

折线柱状一体图代码：（java）

```

const colors = ['a7c5eb', '#204051'];
option = {
    color: colors,
    tooltip: {
        trigger: 'axis',
        axisPointer: {

```

```

        type: 'cross'
    }
},
grid: {
    right: '20%'
},
toolbox: {
    feature: {
        dataView: { show: true, readOnly: false },
        restore: { show: true },
        saveAsImage: { show: true }
    }
},
legend: {
    data: ['Precipitation', 'Temperature']
},
xAxis: [
    {
        type: 'category',
        axisTick: {
            alignWithLabel: true
        },
        // prettier-ignore
        data: ['2011', '2012', '2013', '2014', '2015', '2016', '2017', '2018',
'2019', '2020']
    }
],
yAxis: [
    {
        min: '60000',
        type: 'value',
        name: '碳排放总量',
        position: 'right',
        alignTicks: true,
        offset: 80,
        axisLine: {
            show: true,
            lineStyle: {
                color: colors[0]
            }
        },
        axisLabel: {
            formatter: '{value} 万t'
        }
    }
]

```

```

    },
    {
      type: 'value',
      name: '年增长率',
      position: 'left',
      alignTicks: true,
      axisLine: {
        show: true,
        lineStyle: {
          color: colors[1]
        }
      },
      axisLabel: {
        formatter: '{value} %'
      }
    }
  ],
  series: [
    {
      name: '碳排放量',
      type: 'bar',
      yAxisIndex: 0,
      data: [
65193.342, 67502.613, 66749.376, 64853.276, 66074.810, 68526.125, 70451.557, 71502.0
03, 74096.331, 72633.324
      ]
    },
    {
      type: 'line',
      yAxisIndex: 1,
      data:
[15.6729636, 0.035421886*100, -0.011158645*100, -0.02841*100, 0.018835*100, 0.0370
99*100, 0.028098*100, 0.01491*100, 0.036283*100, -0.01974*100]
    }
  ]
};

```

饼图代码: (java)

```

option = {
  title: {

    left: 'center'
  },
  tooltip: {

```

```

    trigger: 'item'
  },
  legend: {
    orient: 'vertical',
    left: 'left'
  },
  series: [
    {
      name: 'Access From',
      type: 'pie',
      radius: '50%',
      data: [ { value: 19244.60279, name: '交通消费部门' },
        { value: 5387.293921, name: '农林消费部门' },
        { value: 18269.80783, name: '建筑消费部门' },
        { value: 25928.70668, name: '居民生活消费' },
        { value: 261543.0061, name: '工业消费部门' },
      ],
      emphasis: {
        itemStyle: {
          shadowBlur: 10,
          shadowOffsetX: 0,
          shadowColor: 'rgba(0, 0, 0, 0.5)'
        }
      }
    }
  ]
};

```

桑基图代码：（java）

```

option = {
  series: {
    type: 'sankey',
    layout: 'none',
    emphasis: {
      focus: 'adjacency'
    },
    data: [
      {
        name: '煤炭'
      },
      {
        name: '油品'
      },
      {
        name: '天然气',

```

```

        itemStyle: {
            color: '#d9adad'
        },
    },
    {
        name: '热力',
        itemStyle: {
            color: '#112d4e'
        },
    },
    {
        name: '电力'
    },
    {
        name: '其他能源',
        itemStyle: {
            color: '#e05297'
        },
    },
    {
        name: '农林消费部门',
        itemStyle: {
            color: '#9088d4'
        },
    },
    {
        name: '能源供应部门',
        itemStyle: {
            color: '#92817a'
        },
    },
    {
        name: '工业消费部门'
    },
    {
        name: '交通消费部门'
    },
    {
        name: '建筑消费部门'
    },
    {
        name: '居民生活消费',
        itemStyle: {
            color: '#fa9579'
        },
    },

```

```

    },
    {
      name: '能源消费'
    }
  ],
  links: [
    {
      source: '煤炭',
      target: '农林消费部门',
      value: 37.5693228,
      lineStyle: {color: '#b9fffc'}
    },
    {
      source: '煤炭',
      target: '能源供应部门',
      value: 12897.23761,
      lineStyle: {color: '#b9fffc'}
    },
    {
      source: '煤炭',
      target: '工业消费部门',
      value: 8005.640609,
      lineStyle: {color: '#b9fffc'}
    },
    {
      source: '煤炭',
      target: '交通消费部门',
      value: 2.1471858,
      lineStyle: {color: '#b9fffc'}
    },
    {
      source: '煤炭',
      target: '建筑消费部门',
      value: 9.6622426,
      lineStyle: {color: '#b9fffc'}
    },
    {
      source: '煤炭',
      target: '居民生活消费',
      value: 10.32686137,
      lineStyle: {color: '#b9fffc'}
    },
  ],

```

```

{
  source: '油品',
  target: '农林消费部门',
  value: 313.5703037,
  lineStyle: {color: '#a8dda8'}
},
{
  source: '油品',
  target: '能源供应部门',
  value: 191.3316511,
  lineStyle: {color: '#a8dda8'}
},
{
  source: '油品',
  target: '工业消费部门',
  value: 1381.334339,
  lineStyle: {color: '#a8dda8'}
},
{
  source: '油品',
  target: '交通消费部门',
  value: 1606.762112,
  lineStyle: {color: '#a8dda8'}
},
{
  source: '油品',
  target: '建筑消费部门',
  value: 107.2094765,
  lineStyle: {color: '#a8dda8'}
},
{
  source: '油品',
  target: '居民生活消费',
  value: 607.9521612,
  lineStyle: {color: '#a8dda8'}
},
{
  source: '天然气',
  target: '农林消费部门',
  value: 0,
  lineStyle: {color: '#fbbedf'}
},
{
  source: '天然气',

```

```

    target: '能源供应部门',
    value: 624.9108112,
    lineStyle: {color: '#fbbedf'}
  },
  {
    source: '天然气',
    target: '工业消费部门',
    value: 742.25036,
    lineStyle: {color: '#fbbedf'}
  },
  {
    source: '天然气',
    target: '交通消费部门',
    value: 84.2552064,
    lineStyle: {color: '#fbbedf'}
  },
  {
    source: '天然气',
    target: '建筑消费部门',
    value: 26.9192,
    lineStyle: {color: '#fbbedf'}
  },
  {
    source: '天然气',
    target: '居民生活消费',
    value: 181.1726,
    lineStyle: {color: '#fbbedf'}
  },
{
  source: '热力',
  target: '农林消费部门',
  value: 0,
  lineStyle: {color: '#2d6187'}
},
{
  source: '热力',
  target: '能源供应部门',
  value: 1813.071192, //负
  lineStyle: {color: '#2d6187'}
},
{
  source: '热力',
  target: '工业消费部门',
  value: 1790.813162,

```

```

    lineStyle:{color: '#2d6187'}
  },
  {
    source: '热力',
    target: '交通消费部门',
    value: 1.409898896,
    lineStyle:{color: '#2d6187'}
  },
  {
    source: '热力',
    target: '建筑消费部门',
    value: 5.416854337,
    lineStyle:{color: '#2d6187'}
  },
  {
    source: '热力',
    target: '居民生活消费',
    value: 15.43127624,
    lineStyle:{color: '#2d6187'}
  },
  {
    source: '电力',
    target: '农林消费部门',
    value: 52.47829488,
    lineStyle:{color: '#ffd369'}
  },
  {
    source: '电力',
    target: '能源供应部门',
    value: 4625.832649, //负的
    lineStyle:{color: '#ffd369'}
  },
  {
    source: '电力',
    target: '工业消费部门',
    value: 4303.030914,
    lineStyle:{color: '#ffd369'}
  },
  {
    source: '电力',
    target: '交通消费部门',
    value: 63.77034578,
    lineStyle:{color: '#ffd369'}
  },

```

```

    {
      source: '电力',
      target: '建筑消费部门',
      value: 557.7914276,
      lineStyle: {color: '#ffd369'}
    },
    {
      source: '电力',
      target: '居民生活消费',
      value: 602.2640172,
      lineStyle: {color: '#ffd369'}
    },
  {
    source: '其他能源',
    target: '农林消费部门',
    value: 0,
    lineStyle: {color: '#e6739f'}
  },
  {
    source: '其他能源',
    target: '能源供应部门',
    value: 237.612,
    lineStyle: {color: '#e6739f'}
  },
  {
    source: '其他能源',
    target: '工业消费部门',
    value: 31.926,
    lineStyle: {color: '#e6739f'}
  },
  {
    source: '其他能源',
    target: '交通消费部门',
    value: 0,
    lineStyle: {color: '#e6739f'}
  },
  {
    source: '其他能源',
    target: '建筑消费部门',
    value: 0,
    lineStyle: {color: '#e6739f'}
  },
  {
    source: '其他能源',

```

```

    target: '居民生活消费',
    value: 0,
    lineStyle: {color: '#e6739f'}
  },
  {
    source: '农林消费部门',
    target: '能源消费',
    value: 403.6179214,
    lineStyle: {color: '#a6b1e1'}
  },
  {
    source: '能源供应部门',
    target: '能源消费',
    value: 20389.99591,
    lineStyle: {color: '#ad9d9d'}
  },
  {
    source: '工业消费部门',
    target: '能源消费',
    value: 16254.99538,
    lineStyle: {color: '#fbb4ae'}
  },

  {
    source: '交通消费部门',
    target: '能源消费',
    value: 1758.344749,
    lineStyle: {color: '#a7c5eb'}
  },

  {
    source: '建筑消费部门',
    target: '能源消费',
    value: 706.999201,
    lineStyle: {color: '#3f72af'}
  },
  {
    source: '居民生活消费',
    target: '能源消费',
    value: 1417.146916,
    lineStyle: {color: '#ffbb91'}
  }
]

```

```
}  
};
```

ANOVA分析代码 (Python)

```
import numpy  
import pandas  
from spsspro.algorithm import statistical_model_analysis  
#生成案例数据  
data = pandas.DataFrame({  
    "A": numpy.random.random(size=20)  
})  
#时间序列分析 (ARIMA)  
result = statistical_model_analysis.arima_analysis(data=data, p=0, d=0, q=0,  
forecast_num=10)  
print(result)
```

相关系数热力图代码: (Python)

```
import pandas  
  
from spsspro.algorithm import descriptive_analysis  
#生成案例数据  
data = pandas.DataFrame({  
    "A": [1, 2, 3],  
    "B": [2, 3, 4]  
})  
#相关性分析  
result = descriptive_analysis.correlation_analysis(data)  
print(result)
```

线性回归预测代码: (Python)

```
import numpy  
import pandas  
from spsspro.algorithm import statistical_model_analysis  
#生成案例数据  
data_x1 = pandas.DataFrame({  
    "A": numpy.random.random(size=100),  
    "B": numpy.random.random(size=100)  
})  
data_x2 = pandas.DataFrame({"C": numpy.random.choice(["1", "2", "3"],  
size=100)})  
data_y = pandas.Series(data=numpy.random.choice([1, 2], size=100), name="Y")  
#线性回归  
result = statistical_model_analysis.linear_regression(data_y=data_y,  
data_x1=data_x1, data_x2=data_x2)  
print(result)
```

灰色预测模型代码：（Python）

```
import pandas
from spsspro.algorithm import no_parameter_test
#生成案例数据
data = pandas.Series([1, 2, 3, 4, 5], name="A")
index = pandas.Series(["1", "2", "3", "4", "%"], name="B")
#灰色预测
print(no_parameter_test.grey_forecasting_analysis(data, index))
```

LSTM 预测模型代码：（Python）

```
import keras
from keras.models import Sequential
from keras.layers import LSTM, Dense

# 定义模型
model = Sequential([
    LSTM(50, return_sequences=True, input_shape=(n_steps, input_dim)), #
    # n_steps是时间步长, input_dim是输入维度
    LSTM(50),
    Dense(1) # 回归问题, 输出维度为1
])

# 编译模型
model.compile(optimizer="adam", loss="mse")

# 训练模型
model.fit(X_train, y_train, epochs=50, batch_size=32, validation_data=(X_val, y_val))

def now():
    return datetime.now().strftime("%m_%d_%H_%M_%s")

def parse_result_tf(tf_data):
    """
    parse the result of model output in tensorflow
    :param tf_data: the output of tensorflow
    :return: data in DataFrame format
    """
    return pd.DataFrame.from_dict({"ds": tf_data["times"].reshape(-1), "y":
    tf_data["mean"].reshape(-1)})

def generate_time_series(
    start_date=datetime(2006, 1, 1),
    cnt=4018, delta=timedelta(days=1), timestamp=False
```

```

):
    """
    generate a time series/index
    :param start_date: start date
    :param cnt: date count. If =cnt are specified, delta must not be; one is
required
    :param delta: time delta, default is one day.
    :param timestamp: output timestamp or format string
    :return: list of time string or timestamp
    """

    def per_delta():
        curr = start_date
        while curr < end_date:
            yield curr
            curr += delta

    end_date = start_date + delta * cnt

    time_series = []
    if timestamp:
        for t in per_delta():
            time_series.append(t.timestamp())
    else:
        for t in per_delta():
            time_series.append(t)
        # print(t.strftime("%Y-%m-%d"))
    return time_series

def LSTM_predict_tf(train_data, evaluation_data, forecast_cnt=365, freq="D",
model_dir=""):
    """
    predict time series with LSTM model in tensorflow
    :param train_data: data use to train the model
    :param evaluation_data: data use to evaluate the model
    :param forecast_cnt: how many point needed to be predicted
    :param freq: the interval between time index
    :param model_dir: directory of pre-trained model(checkpoint, params)
    :return:
    """

    model_directory = "./model/LSTM_%s" % now()
    params = {
        "batch_size": 3,

```

```

        "window_size": 4,
        # The number of units in the model's LSTMCell.
        "num_units": 128,
        # The dimensionality of the time series (one for univariate, more than
one for multivariate)
        "num_features": 1,
        # how many steps we train the model
        "global_steps": 3000
    }
    # if there is a pre-trained model, use parameters from it
    if model_dir:
        model_directory = model_dir
        params = read_model_param(model_dir + "/params.txt")

    # create time index for model training(use int)
    time_int = range(len(train_data) + len(evaluation_data))

    data_train = {
        tf.contrib.timeseries.TrainEvalFeatures.TIMES:
time_int[:len(train_data)],
        tf.contrib.timeseries.TrainEvalFeatures.VALUES: train_data["y"],
    }

    data_eval = {
        tf.contrib.timeseries.TrainEvalFeatures.TIMES:
time_int[len(train_data):],
        tf.contrib.timeseries.TrainEvalFeatures.VALUES: evaluation_data["y"],
    }

    reader_train = NumpyReader(data_train)
    reader_eval = NumpyReader(data_eval)

    """
    define in
tensorflow/contrib/timeseries/python/timeseries/input_pipeline.py
    """
    train_input_fn = tf.contrib.timeseries.RandomWindowInputFn(
        reader_train, batch_size=params["batch_size"],
window_size=params["window_size"])

    """
    define in tensorflow/contrib/timeseries/python/timeseries/estimators.py
    """
    estimator_lstm = ts_estimators.TimeSeriesRegressor(

```

```

        model=_LSTMModel(num_features=params["num_features"],
num_units=params["num_units"]),
        optimizer=tf.train.AdamOptimizer(learning_rate=0.01),
        model_dir=model_directory
    )

    if not model_dir:
        """
        website:
https://www.tensorflow.org/api\_docs/python/tf/estimator/Estimator#train
        """

        estimator_lstm.train(input_fn=train_input_fn,
steps=params["global_steps"])

    evaluation_input_fn =
tf.contrib.timeseries.WholeDatasetInputFn(reader_eval)
    evaluation = estimator_lstm.evaluate(input_fn=evaluation_input_fn, steps=1)
    # Predict starting after the evaluation
    (predictions,) = tuple(estimator_lstm.predict(
        input_fn=tf.contrib.timeseries.predict_continuation_input_fn(
            evaluation, steps=forecast_cnt)))

    save_model_param(model_directory, params)
    if "loss" in evaluation.keys():
        print("loss:%.5f" % evaluation["loss"])
        f = open(model_directory + "/%s" % evaluation["loss"], "w")
        f.close()
        model_log(
            evaluation["loss"],
            average_loss=-1 if "average_loss" not in evaluation.keys() else
evaluation["average_loss"],
            content=model_dir
        )

    evaluation = parse_result_tf(evaluation)
    predictions = parse_result_tf(predictions)
    first_date = evaluation_data["ds"].tolist()[0]
    evaluation["ds"] = generate_time_series(first_date, cnt=len(evaluation),
delta=delta_dict[freq])
    latest_date = evaluation_data["ds"].tolist()[-1]
    predictions["ds"] = generate_time_series(latest_date, cnt=len(predictions),
delta=delta_dict[freq])

    return evaluation, predictions

```

加权优化预测模型代码：（MATLAB）

```
% 假设您有三种预测结果：regressionResult, neuralNetworkResult,
greyPredictionResult
% 这些结果可以是向量或矩阵，具体取决于您的数据和模型

% 定义目标函数，这里我们使用均方误差作为目标函数
objectiveFunction = @(weights) mean((weights(1) * regressionResult + weights(2)
* neuralNetworkResult + weights(3) * greyPredictionResult).^2);

% 定义约束条件，权重非负且权重之和为1
lb = [0, 0, 0]; % 最小权重
ub = [1, 1, 1]; % 最大权重
Aeq = [1, 1, 1]; % 权重之和等于1
beq = 1; % 权重之和为1

% 使用fmincon函数来优化权重
initialGuess = [1/3, 1/3, 1/3]; % 初始权重猜测
options = optimoptions('fmincon', 'Display', 'iter'); % 可选的优化选项
optimalWeights = fmincon(objectiveFunction, initialGuess, [], [], Aeq, beq, lb,
ub, [], options);

% 使用最优权重进行加权平均预测
weightedPrediction = optimalWeights(1) * regressionResult + optimalWeights(2) *
neuralNetworkResult + optimalWeights(3) * greyPredictionResult;

% 输出最优权重和加权平均预测结果
disp('最优权重: ');
disp(optimalWeights);
disp('加权平均预测结果: ');
disp(weightedPrediction);
```

PSO算法代码：（MATLAB）

```
% 主程序 PSO
clear
close all
clc

SearchAgents_no = 30 ; % 种群规模
dim = 10 ; % 粒子维度
Max_iter = 1000 ; % 迭代次数
ub = 5 ;
lb = -5 ;
c1 = 2 ; % 学习因子1
c2 = 2 ; % 学习因子2
```

```

w = 0.8 ; % 惯性权重
vmax = 3 ; % 最大飞行速度
pos = lb + rand(SearchAgents_no, dim).*(ub-lb) ; % 初始化粒子群的位置
v = - vmax +2*vmax* rand(SearchAgents_no, dim) ; % 初始化粒子群的速度
% 初始化每个历史最优粒子
pBest = pos ;
pbestfit = zeros(SearchAgents_no,1);
for i = 1:SearchAgents_no
pbestfit(i) = sum(pos(i,:).^2) ;
end
%初始化全局历史最优粒子
[gBestfit, index] = min(pbestfit) ;
gBest = pos(index, :) ;
Convergence_curve = zeros(Max_iter,1);

for t=1:Max_iter
    for i=1:SearchAgents_no
        % 更新个体的位置和速度
        v(i,:) =
w*v(i,:)+c1*rand*(pBest(i,:)-pos(i,:))+c2*rand*(gBest-pos(i,:)) ;
        pos(i,:) = pos(i,:)+v(i,:) ;
        % 边界处理
        v(i,:) = min(v(i,:), vmax);
        v(i,:) = max(v(i,:), -vmax);
        pos(i,:) =min(pos(i,:), ub);
        pos(i,:) =max(pos(i,:), lb);
        % 更新个体最优
        f1 = sum(pos(i,:).^2);
        if f1<pbestfit(i)
            pBest(i,:) = pos(i,:) ;
            pbestfit(i) = f1;
        end
        % 更新全局最优
        if pbestfit(i) < gBestfit
            gBest = pBest(i,:) ;
            gBestfit = pbestfit(i) ;
        end
    end
    % 每代最优解对应的目标函数值
    Convergence_curve(t) = gBestfit;
    disp(['Iteration = ' num2str(t) ', Evaluations = ' num2str(gBestfit)]);
end

figure('unit','normalize','Position',[0.3,0.35,0.4,0.35],'color',[1 1

```

```

1], 'toolbar', 'none')
subplot(1, 2, 1);
x = -5:0.1:5; y=x;
L=length(x);
f=zeros(L, L);
for i=1:L
    for j=1:L
        f(i, j) += e(i)e(j);
    end
end
surfc(x, y, f, 'LineStyle', 'none');
xlabel('x_1');
ylabel('x_2');
zlabel('F')
title('Objective space')

subplot(1, 2, 2);
semilogy(Convergence_curve, 'Color', 'r', 'linewidth', 1.5)
title('Convergence_curve')
xlabel('Iteration');
ylabel('Best score obtained so far');

axis tight
grid on
box on
legend('PSO')
display(['The best solution obtained by PSO is : ', num2str(gBest)]);
display(['The best optimal value of the objective function found by PSO is : ',
num2str(gBestfit)]);

```

ARIMA 预测模型代码：（Python）

```

import numpy
import pandas
from spsspro.algorithm import statistical_model_analysis
#生成案例数据
data = pandas.DataFrame({
    "A": numpy.random.random(size=20)
})
#时间序列分析(ARIMA)
result = statistical_model_analysis.arima_analysis(data=data, p=0, d=0, q=0,
forecast_num=10)
print(result)

```

Newton插值法代码：（MATLAB）

```

clc;clear all;clf;close all;
xi=[2020 2035 2060];

```

```

y0=88683.215;
yi=[y0 2*y0 4*y0];
x=sym('x');
p= Newton_fun(x,xi,yi)
a=2020:1:2060;
p=subs(p,x,a);
plot(a,p)

function p= Newton_fun(x,xi,yi)
n=length(xi);
f=zeros(n,n);

% 对差商表第一列赋值
for k=1:n
    f(k)=yi(k);
end
% 求差商表
for i=2:n
    for k=i:n
        f(k,i)=(f(k,i-1)-f(k-1,i-1))/(xi(k)-xi(k+1-i));
    end
end
disp('差商表如下: ');
disp(f);

%求插值多项式
p=0;
for k=2:n
    t=1;
    for j=1:k-1
        t=t*(x-xi(j));
        disp(t)
    end
    p=f(k,k)*t+p;
    disp(p)
end
p=f(1,1)+p;

end

```

岭回归代码：（Python）

```

import numpy
import pandas
from spsspro.algorithm import statistical_model_analysis
#生成案例数据

```

```
data_x1 = pandas.DataFrame({
    "A": numpy.random.random(size=100),
    "B": numpy.random.random(size=100)
})
data_y = pandas.Series(data=numpy.random.random(size=100), name="Y")
#岭回归, 输入参数详细可以光标放置函数括号内按shift+tab查看, 输出结果参考spsspro模板分析报告
result = statistical_model_analysis.ridge_regression(y=data_y, x1=data_x1)
print(result)
```