

大数据下新质生产力对空气质量的影响路径

华天源, 刘胜题

上海理工大学管理学院, 上海

收稿日期: 2024年6月15日; 录用日期: 2024年8月19日; 发布日期: 2024年8月26日

摘要

随着工业化和城市化的加速, 空气质量问题已成为经济社会可持续发展的关键, 而新质生产力则是推动经济增长和提升竞争力的重要因素。过去的研究多集中在单一领域。为此, 本研究通过大数据和二次建模方法, 系统分析了空气质量对生产效率和产品质量的影响, 以及新质生产力对空气质量的反馈机制。本文通过固定效应模型和多级指标体系, 实证分析了2012~2022年新质生产力与空气质量的关系, 结果显示新质生产力显著缓解城市空气污染, 且政府预算支出起到正向调节作用, 区域间存在显著差异。本文从政府管理和产业发展角度提出对策建议, 为产业发展和环境保护提供理论支持和政策依据。

关键词

空气质量, 新质生产力, 大数据, 时空演化分析

The Impact Pathways of New Quality Productivity on Air Quality under Big Data

Tianyuan Hua, Shengti Liu

Business School, University of Shanghai for Science & Technology, Shanghai

Received: Jun. 15th, 2024; accepted: Aug. 19th, 2024; published: Aug. 26th, 2024

Abstract

With the acceleration of industrialization and urbanization, air quality issues have become critical to sustainable economic and social development. Simultaneously, new quality productivity is a key driver of economic growth and competitiveness. Previous studies have typically focused on iso-

lated areas. Thus, this study systematically analyzes the impact of air quality on productivity and product quality, as well as the feedback mechanism of new quality productivity on air quality, using big data and secondary modeling methods. We empirically examine the relationship between new quality productivity and air quality from 2012 to 2022 using a fixed-effects model and a multilevel indicator system. The results indicate that new quality productivity significantly mitigates urban air pollution, with government budgetary expenditures playing a positive moderating role, and significant regional differences are observed. This paper proposes policy recommendations from the perspectives of government management and industrial development to provide theoretical support and a policy foundation for both industrial progress and environmental protection.

Keywords

Air Quality, New Quality Productivity, Big Data, Spatio-Temporal Evolution Analysis

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 文献综述

1.1. 空气质量

冉江等基于哨兵五号(Sentinel-5P)卫星遥感数据、地面大气污染国控点监测数据、气象数据、道路交通图层等多源大数据,利用 XGBoost 机器学习模型定量估算研究区的大气颗粒物浓度,综合大气污染要素选择模型综合评判研究区空气质量指数(AQI) [1]。钱坤,张克凡(2023)运用大数据相关理论对宁波空气质量进行分析建模和预测[2]。张平华和贾万祥(2023)利用大数据将时空维度中已知空气质量数据与气象数据动态集成,构建了一种基于深度神经网络技术在时间维度和空间维度相融合的空气质量管理模型[3]。黄钦炎等(2022)研究表明移动物联网的空气质量测大数据分析应用可有效弥补原有固定监测模式的短板,使空气污染治理更加精细化、智能化[4]。张波的研究结果表明最小二乘下的低秩矩阵插补算法对于缺失值的填补效果更好[5]。胡清等(2022)认为环境数据人为造假行为时有发生,并验证了区块链提高环境监测数据质量的可行性[6]。郭炳南,王宇,张浩(2022)借助 283 个城市的面板数据运用 DID 和 PSM-DID 方法评估数字经济为要素的心新质发展对城市空气质量的改善作用,异质性研究表明数字经济对秦岭-淮河一线以北的城市空气质量具有更强的影响效应[7]。陈颖(2021)阐明新时期大气环境监测的要求,探索深入利用大数据的大气环境监测分析技术和发展环境的方法[8]。

1.2. 新质生产力

新质生产力是指在信息化、智能化背景下,以信息技术为基础,通过提高资源利用效率、优化产品和服务质量、改善生产组织方式和管理模式等手段,实现生产要素配置和经济增长方式的变革,从而推动经济社会发展的新动能。简而言之,新质生产力是指利用信息技术和现代管理方法,改变传统生产方式和经济结构,实现经济效益和社会效益双提升的生产力。

韩喜平(2024)认为新质生产力在形成中推动我国经济实现量的合理增长和质的有效提升[9]。张新宁(2024)研究指出以科技创新为核心形成与新质生产力相适应的新型生产关系,要坚持创新驱动发展战略,推动科技革命性突破[10]。米加宁等(2024)提出了推进国家治理体系和治理能力现代化的实现路径。研究认为,算力正在加速成为驱动社会生产力变革的关键力量,引领生产方式从工业化社会向智能化社会转

变, 是塑造未来新发展格局的决定性变量。算力与数据、算法的交互融合, 正在催生数字化、智能化、生态化的新质生产力形态[11]。梁炜和朱承亮(2024)基于颠覆性创新生态系统的视角, 从发展潜力、发展能力、发展实力和发展动力等四个维度, 将新质生产力的理论内涵进一步解析为原创性基础研究的突破、关键技术前沿技术的创新、产业的升级与变革, 以及科学-技术-产业的快速演化[12]。

新质生产力作为国家提出的新兴概念, 更加精准地概括了该地区生产力的衡量体系, 考虑到传统生产对环境的影响作用。但是以往的学者针对空气质量的大量研究尚未涉及到与新质生产力的关系探究, 有关新质生产力的研究也多集中在高质量发展和数字经济维度, 因此探究空气质量的时空演变情况, 及其与新质生产力之间的关系具有重要的现实意义。

2. 相关指标处理与测度

2.1. 数据来源与预处理

2.1.1. 数据来源

数据监测相关数据集来源中国环境监测总站的空气质量状况报告, 该监测数据逐小时记录了中国 373 个城市以及多个站点 2014 年 5 月~2022 年 2 月的空气质量数据, 其中空气质量包括 $PM_{2.5}$, PM_{10} , SO_2 , NO_2 , O_3 , CO_2 等共 1,714,084 条记录直观的展示了各城市的空气质量差异。

新质生产力指标参考卢江、郭子昂等学者的做法[13], 对新质生产力的评价建立在科技生产力、绿色生产力和数字生产力三个一级指标上。对于科技生产力, 从创新生产力和技术生产力两个方面刻画; 关于绿色生产力, 从资源节约型生产力和环境友好型生产力两个角度进行测度; 对于数字生产力, 则从数字产业生产力和产业数字生产力两个层面进行考量, 针对部分是数据可能存在缺失的问题, 本文使用了插值法进行补充。各一二级指标数据来自于《中国统计年鉴》《中国能源统计年鉴》《中国环境统计年鉴》《中国科技统计年鉴》《中国工业统计年鉴》等。

2.1.2. 数据预处理

针对空气质量监测的原始数据可能存在缺失的情况, 本文利用 Stata 软件对原始数据集进行遍历, 检测是否存在数据缺失情况。同时借助 R 语言函数进行数据可视化展示, 绘制空气质量数据和新质生产力数据的缺失值。相较于缺失值数据, 异常值数据更有可能产生潜在不良影响效果, 因此在进行数据分析前本文借助箱线图检测检查异常值数据。

2.1.3. 异常数据处理

各检测指标数据缺失为完全随机缺失(如机器故障, 停电等), 但是考虑到每个城市空气质量变化具有邻近性, 所以我们考虑采取最近邻法方法填补缺失值。KNN 最近邻法: 欧式距离或相关分析来确定距离具有缺失数据样本最近 k 个样本, 将这 k 个值加权平均来估计该样本的缺失数据。

$$\text{欧式距离: } r = c(x_i, y_i) = \sqrt{\sum_{i=1}^n [c_r(x_i) - c_r(x_j)]^2} \quad (1)$$

在上式中, 我们没有获得目标点的空气质量指数测量值 c_x , 但可以获得其周围若干测量值情况 $c_0, c_1, \dots, c_8$, 这样便可以使用已有数据对目标值 C_x 进行估计:

$$\tilde{C}_x = \frac{\sum_i w_i c_i}{w_i} \quad (2)$$

其中权重 w 与邻点与目标点的距离成反相关, 如:

$$w_i = \frac{1}{d_i} \quad (3)$$

结果：经过这样的调整，完美的调整了数据离群化问题，使得结果更加精准，贴合实际情况。针对个别数据无法实现近邻填补的，本文采用剔除处理，最后得到 987,199 条有效数据。

2.1.4. 数据无量纲化

空气质量二次建模指标和新质生产力指标存在不同的量纲问题，在后续分析中可能导致变量之间的数值差异较大，影响模型的训练和结果解释。无量纲化处理可以消除这种影响，使得不同变量具有可比性，加快模型收敛速度，提高模型的泛化能力。数据无量纲化处理过程如下：

$$x_{standardized} = \frac{x - \bar{x}}{\delta} \text{ (标准化)} \quad (4)$$

$$x_{ij} = \frac{x_{ij} - \min_{1 \leq i \leq n} x_{ij}}{\max_{1 \leq i \leq n} x_{ij} - \min_{1 \leq i \leq n} x_{ij}} \quad (j = 1, 2, \dots, n) \text{ (归一化)} \quad (5)$$

$$x_{ij} = x_{ij} \left(\frac{\max_{1 \leq i \leq n} x_{ij} - \min_{1 \leq i \leq n} x_{ij}}{\max_{1 \leq i \leq n} x_{ij} - \min_{1 \leq i \leq n} x_{ij}} \right) + \min_{1 \leq i \leq n} x_{ij} \quad (j = 1, 2, \dots, n) \text{ (反归一化)} \quad (6)$$

2.2. 指标说明

根据《环境空气质量标准》，常用衡量空气质量的常规大气污染物共有六种，分别为二氧化硫(SO₂)、二氧化氮(NO₂)、粒径小于 10 微米的颗粒物(PM₁₀)、粒径小于 2.5 微米的颗粒物(PM_{2.5})、臭氧(O₃)、一氧化碳(CO)。

颗粒物浓度(PM_{2.5} 和 PM₁₀)：PM_{2.5} 和 PM₁₀ 分别指空气中直径小于等于 2.5 微米和 10 微米的可吸入颗粒物的浓度，通常以微克每立方米(μg/m³)为单位。

臭氧浓度(O₃)：臭氧是一种有毒氧化物，通常以小时平均浓度来衡量，单位为微克每立方米(μg/m³)。自由基循环和 NO_x 循环相互耦合作用，使 NO 不断转化为 NO₂，NO₂ 的光解使 O₃ 逐渐积累，导致污染的产生。总的来说，臭氧污染形成过程受到多种一次污染物因素及光照强度等气象因素的影响[14]。

臭氧指标衡量本文采用最大 8 小时滑动平均法，该指标具体指一个自然日内 8 时至 24 时的所有 8 小时滑动平均浓度中的最大值，其中 8 小时滑动平均值指连续 8 小时平均浓度的算术平均值。其计算公式如下：

$$C_{O_3} = \max_{t=8,9,\dots,24} \left\{ \frac{1}{8} \sum_{i=t-7}^t C_i \right\} \quad (7)$$

其中 C_t 为臭氧在某日 t-1 时至 t 时的平均污染物浓度。

3. 大数据驱动下空气质量二次建模

3.1. 空气质量一次建模 AQI

根据《环境空气质量指数(AQI)技术规定(试行)》，空气质量指数(AQI)可用于判别空气质量等级。

首先需得到各项污染物的空气质量分指数(IAQI)，其计算公式如下：

$$IAQI_p = \frac{IAQI_{Hi} - IAQI_{Lo}}{BP_{Hi} - BP_{Lo}} \cdot (C_p - BP_{Lo}) + BP_{Lo} \quad (8)$$

式中各符号含义如下：IAQI_p 污染物的空气质量分指数，结果进位取整数；C_p 污染物的质量浓度值；BP_{Hi}，BP_{Lo} 与相近的污染物浓度限值的高位值与低位值；IAQI_{Hi}，IAQI_{Lo} 对应的空气质量分指数。

空气质量指数(AQI)取各分指数中的最大值, 即:

$$AQI = \max \{IAQI_1, IAQI_2, IAQI_3, \dots, IAQI_n\} \quad (9)$$

$IAQI_1, IAQI_2, IAQI_3, \dots, IAQI_n$ 为各污染物项目的分指数。在本题中, 对于 AQI 的计算仅涉及为上文表中提供的六种污染物, 因此计算公式如下:

$$AQI = \max \{IAQI_{SO_2}, IAQI_{NO_2}, IAQI_{PM_{10}}, IAQI_{PM_{2.5}}, IAQI_{O_3}, IAQI_{CO}\} \quad (10)$$

空气质量等级范围根据 AQI 数值划分, 当 AQI 小于或等于 50 (即空气质量评价为“优”)时, 称当天无首要污染物; 当 AQI 大于 50 时, IAQI 最大的污染物为首要污染物。若 IAQI 最大的污染物为两项或两项以上时, 并列为首要污染物。IAQI 大于 100 的污染物为超标污染物。

依靠标准化 AQI 计算公式, 计算各项监测数据的 IAQI (空气质量分指数)与浓度限值比较, 选取各分指数中最大值作为空气质量指数 AQI。

3.2. 空气质量二次建模 SAQI

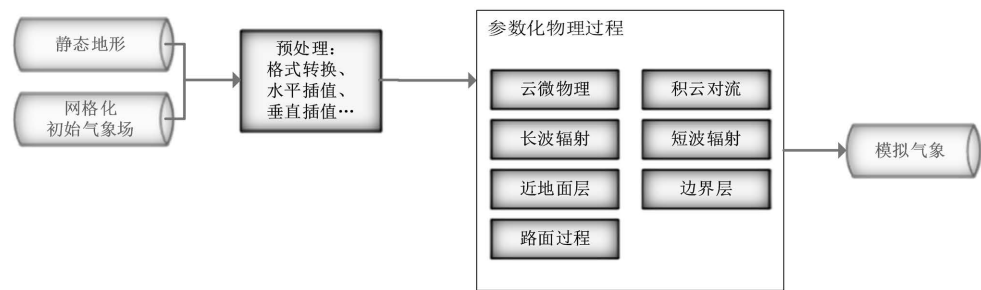


Figure 1. Numerical weather forecasting system, WRF architecture
图 1. 数值天气预报系统 WRF 结构

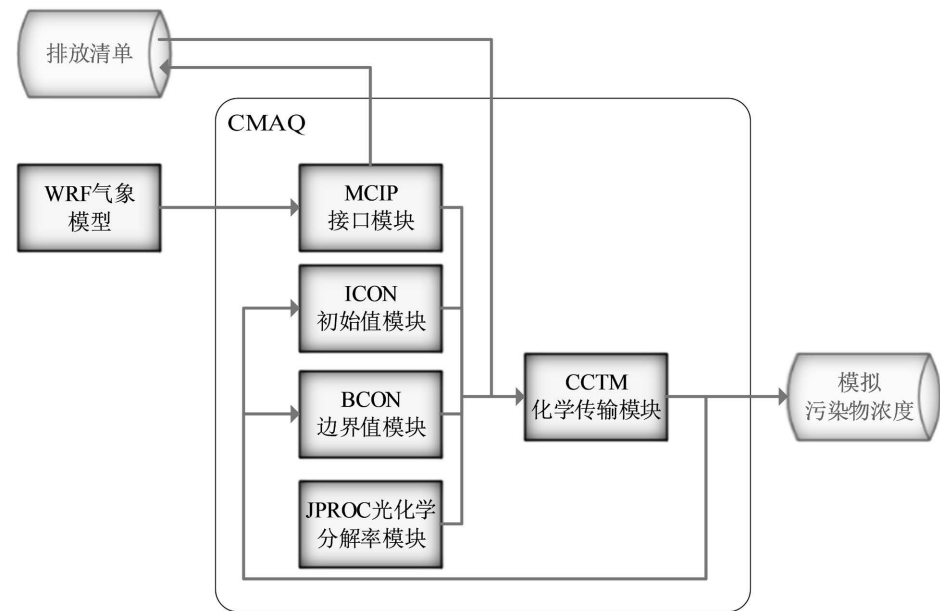


Figure 2. Structure of the air quality prediction and assessment system CMAQ
图 2. 空气质量预测与评估系统 CMAQ 结构

目前常用 WRF-CMAQ 模拟体系(以下简称 WRF-CMAQ 模型)对空气质量进行预报。WRF-CMAQ 模

型主要包括 WRF 和 CMAQ 两部分：WRF 是一种中尺度数值天气预报系统，用于为 CMAQ 提供所需的气象场数据；CMAQ 是一种三维欧拉大气化学与传输模拟系统，其根据来自 WRF 的气象信息及域内的污染排放清单，基于物理和化学反应原理模拟污染物等的变化过程，继而得到具体时间点或时间段的预报结果[15]。WRF 和 CMAQ 的结构如图 1 和图 2 所示。

受制于模拟的气象场以及排放清单的不确定性，以及对包括臭氧在内的污染物生成机理的不完全明晰，WRF-CMAQ 预报模型的结果并不理想。因此，本文提出对空气质量 AQI 二次建模，即在 WRF-CMAQ 模型模拟结果的基础上，建立新的评价体系和更多的数据源进行二次建模以提高空气质量预测的准确性和对新质生产力的影响研究的适用性。

其中，由于实际气象条件对空气质量影响很大(例如湿度降低有利于臭氧的生成)，且污染物浓度实测数据的变化情况对空气质量预报具有一定参考价值，考虑参考空气质量监测点获得的气象与污染物数据进行二次建模，以优化预报模型。二次模型与 WRF-CMAQ 模型关系如图 3 所示。

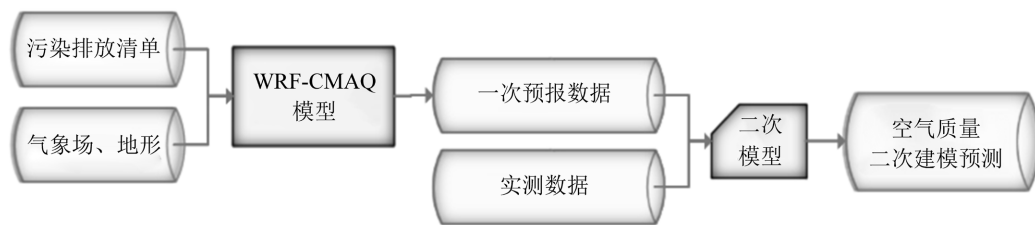


Figure 3. Secondary air quality modelling WRF-CMAQ model
图 3. 二次空气质量建模 WRF-CMAQ 模型

3.3. 空气质量二次建模预测

本文选择遗传算法 + 神经网络算法对部分监测点(本文使用三个监测点)的一次预报数据和实测数据进行二次建模。具体输入变量如表 1 所示。

Table 1. Input variables for secondary modelling of air quality
表 1. 空气质量二次建模输入变量

模型	输出结果	输入结果
(1)	SO ₂ 实测浓度预测	气象因子、SO ₂ 预测时点前 24 小时实测浓度 SO ₂ 预测时点一次预告浓度
(2)	NO ₂ 实测浓度预测	气象因子、O ₃ 预测时点前 24 小时实测浓度 NO ₂ 预测时点一次预告浓度
(3)	CO 实测浓度预测	气象因子、CO 预测时点一次预告浓度
(4)	O ₃ 实测浓度预测	气象因子、CO、NO ₂ 预测时点前 24 小时实测浓度 O ₃ 预测时点一次预告浓度
(5)	PM _{2.5} 实测浓度预测	气象因子、SO ₂ 、NO ₂ 、CO 预测时点前 24 小时实测浓度 PM _{2.5} 预测时点一次预告浓度
(6)	PM ₁₀ 实测浓度预测	气象因子、SO ₂ 、NO ₂ 、CO 预测时点前 24 小时实测浓度 PM ₁₀ 预测时点一次预告浓度

3.3.1. 运用 LM-神经网络对预测模型的构建

LM (Levenberg-Marquardt)神经网络算法是对梯度下降算法的一个优化，但是 BP 神经网络具有收敛速度过慢、容易陷入局部最优和降低泛化能力等问题。基于此，为解决上述问题对模型可能造成的拟合误差，我们采用“遗传算法 + 神经网络”结合高斯 - 牛顿算法的特点解决 BP 神经网络的缺点。基于此构建了遗传算法(GA)和 LM 前馈神经网络模型。

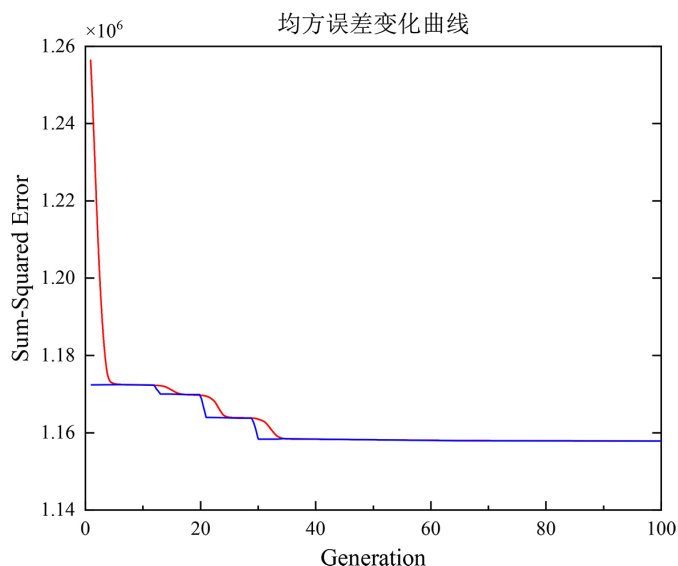
我们首先确定神经网络的基本拓扑结构，并对神经网络权值和阈值进行编码训练得到初试种群，比

较测试集测试误差。然后根据误差大小计算合适的适应度进行交叉变异产生新种群与终止条件对比, 若不满足则返回初试种群进行重新训练, 以此得到神经网络最佳权值和阈值, 用于二次建模。

3.3.2. 确定初始隐藏层节点数

本问题选取的 LM 前馈神经网络是一个输入层节点数为 2, 为工艺参数所包含的两组数据值; 包含一个隐藏层, 隐藏层的节点数为 10; 输出层的节点数为 3, 为结构变量所包含的三组数据值。

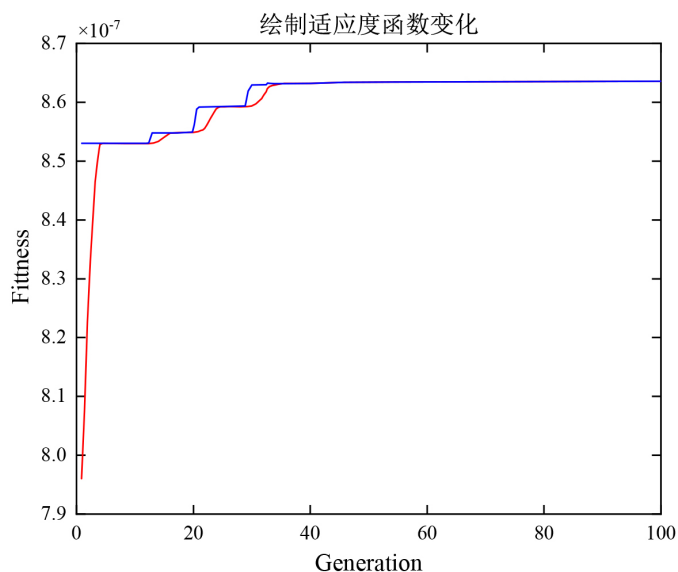
3.3.3. 初试模型求解与遗传算法参数修正



注: 横坐标表示迭代次数, 纵坐标表示均方误差值。

Figure 4. Mean square error variation curve

图 4. 均方误差变化曲线



注: 横坐标表示迭代次数, 纵坐标表示适应度函数值。

Figure 5. Change in fitness function

图 5. 适应度函数变化

通过 MATLAB 软件编程和代码实现, 可以看出当遗传算法迭代次数接近 38 次的时候, 适应度函数达到最优, 此时适应度函数的均方误差值为 $1.16\text{E}-6$ 。详细见图 4 和图 5。

Table 2. Table of some parameters of “Genetic algorithm + LM feed-forward neural network”

表 2. “遗传算法 + LM 前馈神经网络” 部分参数表

参数	参数说明
2000	Maximum number of epochs to train
0	Performance goal
6	Maximum validation failures
$1.00\text{E}-07$	Minimum performance gradient
0.001	Initial μ
0.1	μ decrease factor
10	μ increase factor
$1.00\text{E}+10$	Maximum μ
25	Epochs between displays (NaN for no displays)
FALSE	Generate command-line output
TRUE	Show training GUI

注: 表 2 说明参数修正后神经网络的学习率初始赋值 $\mu = 0.001$, 衰减率为 0.1, 增长率是 10。

3.3.4. 二次建模训练对比

通过对比下图初试模型网络性能图 6 和遗传算法(GA)修正参数后网络性能图 7。很容易发现, 虽然两张图都在 20 次内学习的到了最优的结果停止了对参数的调整, 但是经过遗传算法优化后的模型拟合误差为 0.038 是原始模型拟合性误差 5.034 的 100 多倍。说明遗传算法对模型拟合的性能有大幅提升。

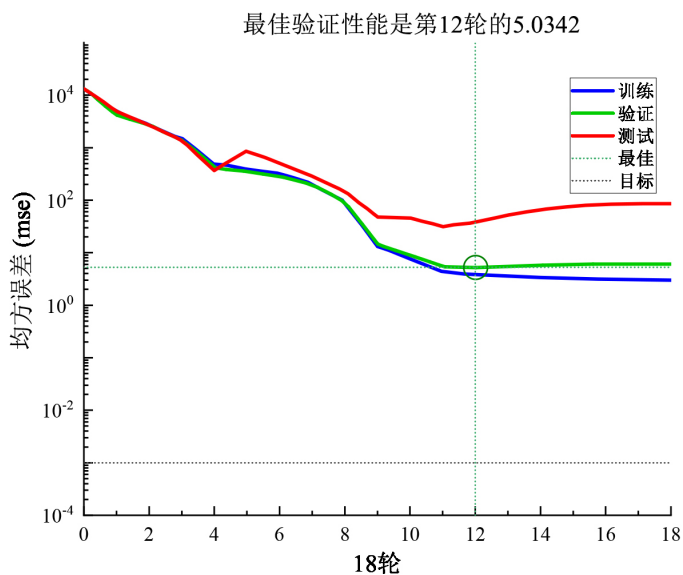
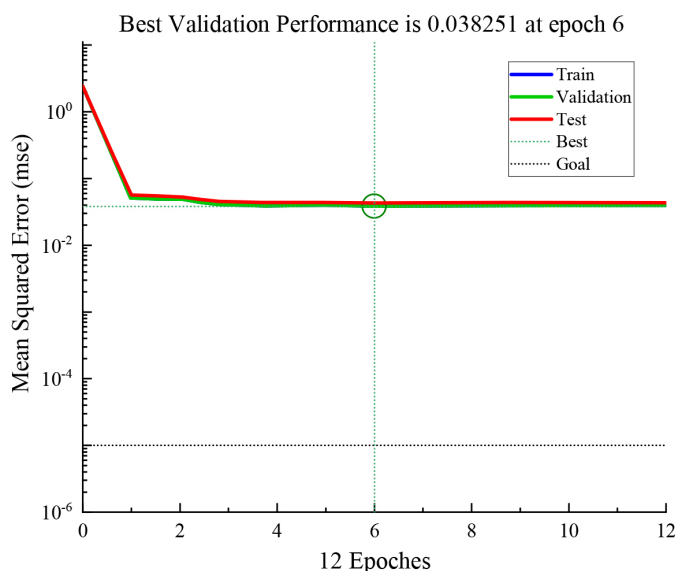


Figure 6. Initial parametric model performance effects

图 6. 初始参数模型性能效果



注：横坐标表示学习轮次，纵坐标表示均方误差。

Figure 7. Effect of model performance after correction of parameters

图 7. 修正参数后模型性能效果

3.3.5. 二次建模预测结果

通过 Matlab 对集合数据的运行, 得到了四个集合的拟合情况, 其中 $R_T = 0.82$, $R_{Va} = 0.82$, $R_{Test} = 0.80$, $R_{All} = 0.82$ 。因此本文采用修正参数后的训练模型作为二次建模模型。

4. 空气质量与新质生产力实证分析

4.1. 实证模型设定

为新质生产力对检验空气污染的影响, 设定如下面板固定效应模型:

$$SAQI_{it} = \beta_0 + \beta_1 NQF_{it} + \beta_2 Contr_{it} + \mu_i + \eta_t + \varepsilon_{it} \quad (11)$$

上式子中, 下标 i 、 t 分别代表城市和年份; $SAQI_{it}$ 是第 i 个农产品区域公用品牌第 t 年的空气污染指数, 是本文的核心被解释变量; NQF_{it} 为新质生产力, 是本文的核心解释变量; $Contr_{it}$ 代表控制变量, 城市规模(Size)、居民能源消耗指数(Energy)、产业结构(Stru)、氮氧化物排放(NO)、城市人口增长率(Rate)和规模以上工业数(Company); β 代表常数项; μ_i 为不可观测的个体因素; η_t 为不可观测的时间因素, ε_{it} 为误差项。

此外, 为检验政府一般预算支出在新质生产力与城市空气污染之间的调节作用, 设定如下面板回归模型:

$$SAQI_{it} = \beta_0 + \beta_1 NQF_{it} + \beta_2 Exp_{it} + \beta_3 Medor_{it} + \beta_4 Contr_{it} + \mu_i + \eta_t + \varepsilon_{it} \quad (12)$$

其中, 政府一般预算支出 Exp_{it} 为调节变量, $Medor_{it}$ 代表新质生产力与政府一般预算支出的交乘项, $Contr_{it}$ 为调节效应的关键解释变量。

4.2. 变量选取, 样本选择

4.2.1. 变量选取

将本文的核心解释变量为城市新质生产力、被解释变量为城市空气污染。此外, 基于数据可得性,

还选取了与空气污染相关的控制变量, 如城市规模(*Size*)、居民能源消耗指数(*Energy*)、产业结构(*Stru*)、氮氧化物排放(*NO*)、城市人口增长率(*Rate*)和规模以上工业数(*Company*)。变量汇总如表 3 所示:

Table 3. Summary of variables for empirical analysis

表 3. 实证分析变量汇总

变量类型	变量名称	符号	说明
被解释变量	空气污染指数	SAQI	通过本文的二次建模得到
解释变量	新质生产力	NQF	通过熵权法 TOSIS 法及其测度指标体系计算
调节变量	政府预算支出	Medor	政府一般预算支出
控制变量	城市规模	Size	地级市户籍人全市口
	居民能源消耗指数	Energy	Ln (1 + 人工或天然气居民家庭用量)
	产业结构	Stru	第二产业占地区生产总值的比重
	氮氧化物排放	NO	全市工业氮氧化物排放量
	城市人口增长率	Rate	全市人口自然增长率
	规模以上工业数	Company	全市规模以上工业企业数

4.2.2. 样本选择

本文的空气污染相关数据来源于科技部重点研发计划项目《地理大数据挖掘与时空模式发现》, 该项目在全国主要城市设定观测点, 从 2014 年开始持续进行观测, 本文选取了 2014 至 2022 年 385 个地级市的空气污染数据。关于新质生产力计算的原始数据、控制变量相关数据均来源于历年《中国城市统计年鉴》, 将空气污染数据与新质生产力数据通过“城市名称、年份”使用 Stata 进行合并, 得到研究所需的非平衡面板数据。并将数据缺失的样本进行了插补和剔除处理, 将研究数据进行上述处理后, 最终得到 385 个地级市从 2014 至 2022 年一共 2357 个样本, 基本满足研究所需。

4.3. 检验与分析

4.3.1. 描述性统计

相关核心变量的均值、标准差、最小值、最大值的统计结果如表 4 所示。从中可以看出全是工业氮氧化物排放量、城市人口和规模以上工业企业数的最大值与最小值差别较大, 可能由于经济发展状况导致。

Table 4. Descriptive statistics for variables

表 4. 变量描述性统计

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max
NQF	2357	0.0216	0.0415	0.00106	0.443
SAQI	2357	68.84	21.50	24.60	159.9
Size	2357	464.1	329.1	20	3416
Energy	2357	8.055	1.621	0.693	12.39
Stru	2357	44.18	10.28	10.68	79.36
NO	2357	26,322	31,681	2	474,805
Rate	2357	5.845	5.634	-16.64	38.80
Company	2357	776.2	440.8	1	1538

4.3.2. 相关性分析

通过相关性分析，初步探明各个变量间的相关性，变量间的相关系数如表 5 所示。

Table 5. Table of correlation coefficients

表 5. 相关系数表

	SAQI	NQF	Size	Energy	Stru	NO	Rate	Company
SAQI	1							
NQF	0.015	1						
Size	0.212***	0.475***	1					
Energy	0.309***	0.440***	0.394***	1				
Stru	0.287***	-0.120***	-0.124***	0.052**	1			
NO	0.258***	0.190***	0.259***	0.163***	0.221***	1		
Rate	0.008	-0.003	0.112***	-0.106***	0.115***	0.077***	1	
Company	-0.078***	0.095***	-0.040*	0.034*	0.064***	0.068***	-0.0340	1

注：***、**、*分别表示在 1%、5%、10%的水平上显著。

4.3.3. 多重共线性检验

为排除多重共线性对回归结果的影响，将与空气污染相关的变量纳入回归模型进行 VIF 检验，结果如表所示，由表 6 可知，各个变量的 VIF 值都小于 10，处于可接受的范围，故认为不存多重共线性问题。

Table 6. Results of multicollinearity tests

表 6. 多重共线性检验结果

Variable	VIF	1/VIF
Size	1.500	0.665
NQF	1.480	0.674
Energy	1.380	0.725
NO	1.170	0.856
Stru	1.150	0.872
Rate	1.070	0.937
Company	1.030	0.973
Mean	VIF	1.250

4.3.4. 豪斯曼检验

为确定使用随机效应还是固定效应模型，故需要 Hausman 检验。其原假设与备择假设分别为：

H_0 = 模型中不同个体的截距项与所有解释变量不相关。

H_1 = 模型中不同个体的截距项与所有解释变量相关。

借助 Stata17 对研究数据进行检验豪斯曼，检验结果如表 7 所示，新质生产力作为解释变量时的卡方值为 876.53，P 值都是 0.000，在 1%水平下显著拒绝原假设，证明使用固定效应模型更优。

Table 7. Results of the Hausman test
表 7. Hausman 检验结果

Hausman Test	Chi-sq 值	自由度	P 值
NQF	876.53	5	0.000

4.4. 基准回归结果

新质生产力的回归结果如表 8 所示，采用逐步回归法，在使用地区和年份固定效应后，依次将控制变量纳入回归模型，由表可知，将控制变量纳入回归模型，并固定年份和地区效应后，新质生产力的回归系数为-1.214，在 1%水平显著，这表明新质生产力显著缓解城市空气污染的程度，图 8 反映出新质生产力与城市空气污染的拟合关系，拟合曲线的效率为负，再次证实新质生产力负向影响城市空气污染。

Table 8. Results of capital mismatch regressions
表 8. 资本错配回归结果

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	SAQI	SAQI	SAQI	SAQI	SAQI	SAQI	SAQI
<i>NQF</i>	-1.270*** (-3.70)	-1.040*** (-3.00)	-1.022*** (-2.95)	-1.037*** (-2.97)	-1.204*** (-3.43)	-1.216*** (-3.47)	-1.214*** (-3.46)
<i>Size</i>		-0.000*** (-3.85)	-0.000*** (-3.83)	-0.000*** (-3.85)	-0.000*** (-3.58)	-0.000*** (-3.72)	-0.000*** (-3.72)
<i>Energy</i>			-0.005** (-2.05)	-0.005** (-2.05)	-0.005** (-2.12)	-0.005** (-2.08)	-0.005** (-2.08)
<i>Stru</i>				0.000 (0.41)	0.000 (0.55)	0.000 (0.42)	0.000 (0.41)
<i>NO</i>					0.000*** (3.55)	0.000*** (3.57)	0.000*** (3.56)
<i>Rate</i>						0.002*** (2.77)	0.001*** (2.77)
<i>Company</i>							0.000 (0.33)
<i>_cons</i>	0.830*** (88.27)	0.981*** (24.34)	1.016*** (23.19)	1.008*** (20.86)	0.996*** (20.62)	0.994*** (20.62)	0.993*** (20.51)
<i>i.FE</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>i.YEAR</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	2357	2357	2357	2357	2357	2357	2357
<i>r2_w</i>	0.534	0.537	0.538	0.538	0.541	0.542	0.542

注：***、**、*分别表示在 1%、5%、10%的水平上显著；括号内为 t 值，以下各表同。

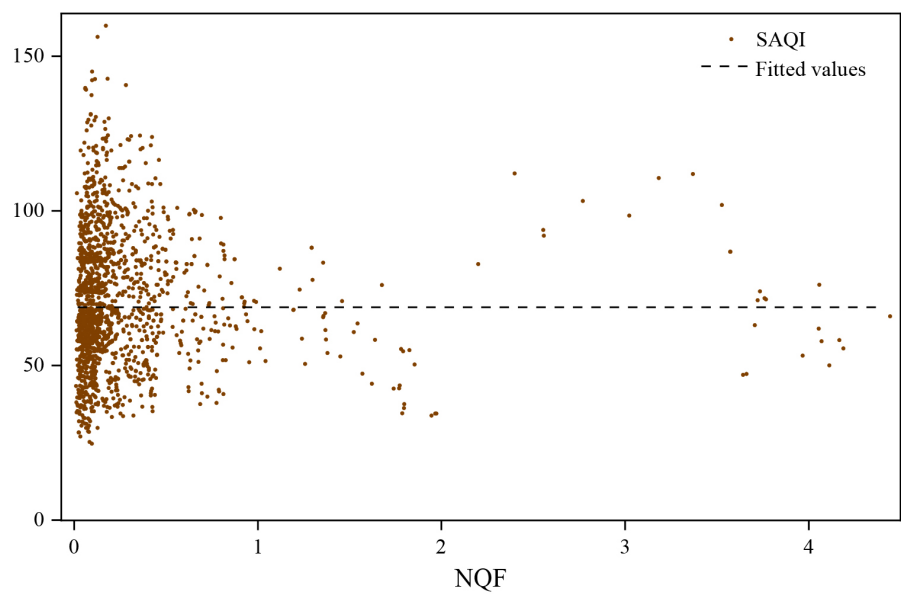


Figure 8. Fitted curves of new quality productivity and urban air pollution
图 8. 新质生产力与城市空气污染拟合曲线

4.5. 调节效应检验

政府预算支出在新质生产力与城市空气污染之间的调节作用检验结果如表 9 所示。表第(1)列是基准回归结果，表第(2)列是将调节变量纳入基准回归模型的结果，可以看到政府预算支出(*Exp*)的回归系数为 0.056，且在 1%水平显著。表第(3)列是将调节变量，调节变量与核心解释变量的交乘项同时纳入基准回归模型的结果，由表可知，新质生产力与政府预算支出的交乘项(*Medor*)回归系数为-0.138，且在 1%水平显著，而新质生产力(*NQF*)回归系数在 1%水平显著为负，这表明政府预算支出在在新质生产力与城市空气污染之间起到正向调节作用。图 9 展示了政府预算支出的调节效应，可以看出，相比于低水平的政府预算支出，在高水平的政府预算支出时的斜率更大，再次证实政府预算支出正向调节作用。

Table 9. Results of the moderating effect of government budget expenditures
表 9. 政府预算支出的调节效应结果

	(1)	(2)	(3)
	SAQI	SAQI	SAQI
<i>NQF</i>	-0.121*** (-3.46)	-0.073** (-2.01)	-0.110*** (-2.76)
<i>Exp</i>		0.056*** (4.61)	0.096*** (4.46)
<i>Moder</i>			-0.138** (-2.26)
<i>Size</i>	-0.335*** (-3.72)	-0.169* (-1.75)	-0.104 (-1.03)
<i>Energy</i>	-0.484** (-2.08)	-0.498** (-2.16)	-0.468** (-2.02)

续表

<i>Stru</i>	0.192 (0.41)	0.272 (0.58)	0.357 (0.76)
<i>NO</i>	0.032*** (3.56)	0.025*** (2.72)	0.024*** (2.63)
<i>Rate</i>	0.150*** (2.77)	0.156*** (2.90)	0.155*** (2.88)
<i>Company</i>	0.020 (0.33)	0.021 (0.34)	0.023 (0.38)
<i>_cons</i>	0.099*** (20.51)	0.045*** (3.51)	0.064*** (4.18)
<i>i.FE</i>	Yes	Yes	Yes
<i>i.YEAR</i>	Yes	Yes	Yes
<i>N</i>	2357	2357	2357
<i>r2_w</i>	0.542	0.547	0.548

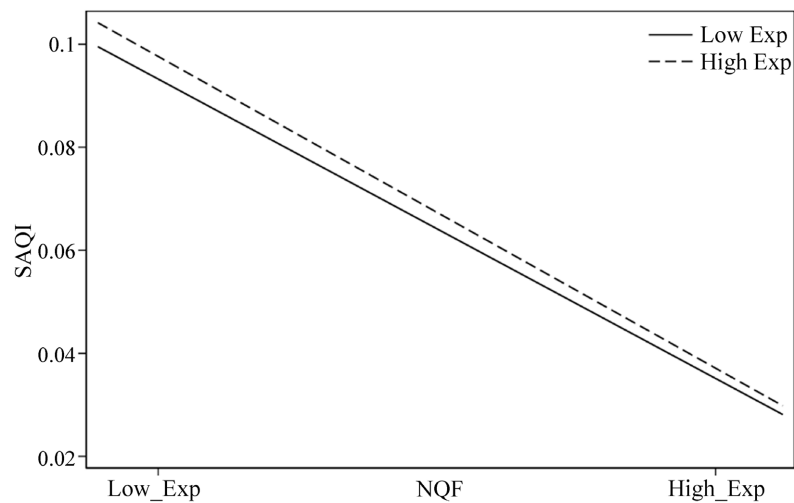


Figure 9. Diagram of the moderating effect of government budget expenditures
图 9. 政府预算支出的调节效应图

4.6. 稳健性检验

稳健性检验结果如表 10 所示, 由表可知, 在排除异常值影响和替换被解释变量重新回归后, 新质生产力的回归系数和显著性与基准回归的基本一致, 说明基准回归结果稳健。

Table 10. Robustness test results
表 10. 稳健性检验结果

	(1)	(2)	(3)	(4)
	SO ₂	PM _{2.5}	NO ₂	SAQI
<i>NQF</i>	-1.602* (-1.66)	-0.927*** (-3.00)	-0.258* (-1.67)	-1.003** (-2.20)

续表

<i>Size</i>	−0.000	−0.000***	−0.000***	−0.000***
	(−0.26)	(−4.24)	(−8.73)	(−3.66)
<i>Energy</i>	0.015**	−0.004**	−0.000	−0.005*
	(2.33)	(−2.01)	(−0.04)	(−1.95)
<i>Stru</i>	0.002*	0.000	0.000	0.000
	(1.72)	(0.52)	(0.99)	(0.62)
<i>NO</i>	−0.000***	0.000***	0.000***	0.000***
	(−5.21)	(3.44)	(3.27)	(3.58)
<i>Rate</i>	−0.000	0.001***	0.000	0.001**
	(−0.33)	(3.11)	(1.46)	(2.04)
<i>Company</i>	0.000*	0.000	0.000	0.000
	(1.85)	(0.01)	(0.10)	(0.40)
<i>_cons</i>	−2.474***	0.715***	0.476***	0.975***
	(−18.53)	(16.72)	(22.43)	(20.37)
<i>i.FE</i>	Yes	Yes	Yes	Yes
<i>i.YEAR</i>	Yes	Yes	Yes	Yes
<i>N</i>	2357	2357	2357	2357
<i>r2_w</i>	0.773	0.565	0.461	0.554

4.7. 异质性分析

为了探究新质生产力对城市空气污染的影响在不同地区是否存在差异，本文按照国家统计局的划分标准，将研究样本划分为东部地区、中部地区和西部地区，进行分组回归，结果如表 11 所示。由表 11 可知，相比于东部地区，新质生产力对城市空气污染的影响在中西部地区更显著，这说明新质生产力对城市空气污染的影响具有区域异质性。

Table 11. Results of the regional heterogeneity test

表 11. 地区异质性检验结果

	(1)	(2)	(3)
	东部地区	中部地区	西部地区
<i>NQF</i>	−0.044	−0.508***	−0.233*
	(−1.24)	(−7.11)	(−1.81)
<i>Size</i>	−0.000***	0.000*	−0.000
	(−3.42)	(1.78)	(−1.15)
<i>Energy</i>	−0.000	−0.001**	−0.000
	(−0.11)	(−2.32)	(−0.94)
<i>Stru</i>	0.000***	−0.000	0.000
	(5.42)	(−1.33)	(0.16)

续表

NO	0.000***	0.000***	0.000
	(4.80)	(2.79)	(0.59)
Rate	0.000***	-0.000	0.000
	(3.39)	(-1.12)	(0.29)
Company	-0.000*	-0.000	-0.000
	(-1.71)	(-0.06)	(-0.11)
_cons	0.083***	0.092***	0.080***
	(11.50)	(10.13)	(10.37)
i.FE	Yes	Yes	Yes
i.YEAR	Yes	Yes	Yes
N	971	850	536
r2_w	0.689	0.681	0.481

5. 结论与对策建议

5.1. 结论

本文首先对所有收集到的原始数据进行处理，目的在于检测和修正异常数据，保证数据和模型的可靠性。针对空气质量的预测问题，本文利用标准化空气质量进行一次建模，然后利用遗传算法和神经网络进行空气质量预测的二次建模，保证了预测的精度。分析结果显示：(1) 新质生产力与城市空气污染的拟合曲线的效率为负，即新质生产力显著缓解城市空气污染；(2) 相比于低水平的政府预算支出，在高水平的政府预算支出时的斜率更大，即政府财政预算支出在新质生产力与城市空气污染之间起到正向调节效应；(3) 相比于东部地区，新质生产力对城市空气污染的影响在中西部地区更显著，新质生产力对城市空气污染的影响具有显著的区域异质性。

5.2. 对策建议

第一，政府应加强环境管理和监测体系建设，完善环境监测网络覆盖，提高监测数据的时空分辨率和准确性，在中西部地区建立更多的监测站点，加强监测数据的采集和分析，及时发现和解决环境问题。

第二，产业界应积极响应政府政策，加大对环保技术研发和应用的投入，推动产业向高端、绿色、智能方向发展；加强与政府、科研机构和社会组织的合作，共同开展环保项目，提高生产过程中的资源利用率和能源利用效率，减少污染物排放。

第三，政府应加大对新质生产力发展的支持力度，通过增加环保产业的投资，鼓励企业推行绿色生产，推动产业升级，减少污染物排放，实现经济效益和环境效益的双赢。

参考文献

[1] 冉江, 黄荷筠, 邹镭坤. 基于多源大数据的长三角城市群 AQI 遥感估算[J]. 复旦学报(自然科学版), 2023, 62(6): 786-795.

[2] 钱坤, 张克凡. 大数据融合在空气质量预测领域的应用——以宁波市为例[J]. 中国管理信息化, 2023, 26(11): 178-182.

[3] 张平华, 贾万祥. 基于大数据技术的雾霾天气预测模型研究[J]. 安徽职业技术学院学报, 2023, 22(1): 7-12, 20.

[4] 黄钦炎, 林庆丰, 黄欣龙. 基于移动物联的空气质量监测大数据分析应用[J]. 价值工程, 2022, 41(21): 146-148.

- [5] 张波, 宋国君. 大规模空气质量监测数据缺失处理方法实证研究[J]. 中国环境科学, 2022, 42(5): 2078-2087.
- [6] 胡清, 吕广丰, 高菁阳, 付冬雪, 苗波. 区块链技术在环境监测中的应用研究[J]. 中国环境管理, 2022, 14(3): 21-29.
- [7] 郭炳南, 王宇, 张浩. 数字经济发展改善了城市空气质量吗?——基于国家级大数据综合试验区的准自然实验[J]. 广东财经大学学报, 2022, 37(1): 58-74.
- [8] 陈颖. 基于大数据技术的大气污染监测系统研究[J]. 无线互联科技, 2021, 18(18): 46-47.
- [9] 韩喜平, 马丽娟. 发展新质生产力与推动高质量发展[J]. 思想理论教育, 2024(4): 4-11.
- [10] 张新宁. 科技创新是发展新质生产力的核心要素论析[J]. 思想理论教育, 2024(4): 20-26.
- [11] 米加宁, 李大宇, 董昌其. 算力驱动的新质生产力: 本质特征、基础逻辑与国家治理现代化[J]. 公共管理学报, 2024, 21(2): 1-14.
- [12] 梁炜, 朱承亮. 颠覆性创新生态系统视角下新质生产力的逻辑内涵及监测框架[J]. 西北大学学报(哲学社会科学版), 2024, 54(3): 38-47.
- [13] 卢江, 郭子昂, 王煜萍. 新质生产力发展水平、区域差异与提升路径[J]. 重庆大学学报(社会科学版), 2024, 30(3): 1-17.
- [14] 谷精华, 范文玉, 刘健, 等. CuO-ZnO-SiO₂ 制备及其催化臭氧氧化降解硝基苯废水[J]. 石油化工, 2020, 49(5): 434-440.
- [15] 郭林浩. 基于 WRF/CMAQ 模式的区域大气环境承载力研究——以内蒙古自治区呼、包、鄂、巴四市为例[D]: [硕士学位论文]. 呼和浩特: 内蒙古大学, 2023.