

安徽省空气质量现状分析及预测研究 ——基于VMD-CNN-LSTM混合模型

王昕怡, 谢欣语, 张洛萌, 王士杰

青岛黄海学院信息科学与工程学院, 山东 青岛

收稿日期: 2025年8月12日; 录用日期: 2025年9月10日; 发布日期: 2025年9月18日

摘 要

在全球城市化与工业化快速发展的进程中, 空气质量已成为衡量城市可持续发展和居民生活品质的关键指标。本研究聚焦安徽省空气质量, 研究发现时间上季度变化呈“U”型, 空间上各地市污染物平均浓度差异大。同时通过Pearson相关性分析可知, 空气中主要污染物与AQI指数紧密相关。最后构建VMD-CNN-LSTM混合模型预测空气质量, 结果显示由于空气质量影响因素复杂、预测时间跨度大带来的不确定因素导致预测存在误差。本研究为安徽及其他区域空气质量管控提供了数据支持与决策参考。

关键词

安徽省空气质量, VMD-CNN-LSTM模型, 相关性分析, AQI指数

An Analysis of the Current Situation and a Predictive Study on Air Quality in Anhui Province

—Based on VMD-CNN-LSTM Hybrid Model

Xinyi Wang, Xinyu Xie, Luomeng Zhang, Shijie Wang

School of Information Science and Engineering, Qingdao Huanghai University, Qingdao Shandong

Received: Aug. 12th, 2025; accepted: Sep. 10th, 2025; published: Sep. 18th, 2025

Abstract

In the process of rapid global urbanization and industrialization, air quality has become a key indicator for measuring urban sustainable development and residents' quality of life. This study focuses

on the air quality in Anhui Province and finds that the quarterly variation shows a “U” shape in terms of time, while there are significant differences in the average concentrations of pollutants among various prefecture-level cities in terms of space. Meanwhile, through Pearson correlation analysis, it is found that the main air pollutants are closely related to the AQI index. Finally, a VMD-CNN-LSTM hybrid model is constructed to predict air quality. The results show that there are errors in the prediction due to complex influencing factors of air quality and uncertain factors caused by the large time span of prediction. This study provides data support and decision-making references for air quality management and control in Anhui and other regions.

Keywords

Air Quality in Anhui Province, VMD-CNN-LSTM Model, Correlation Analysis, AQI Index

Copyright © 2025 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

在全球环境治理的大背景下，各国政府、企业及社会公众正以协同共进的姿态，在政策制定、技术创新、环保实践等多个层面形成合力，共同推动空气质量改善成为普遍共识与持续行动。近年来，安徽省在经济总量跃上新高度，产业创新与开放活力双轮驱动，民生福祉显著改善方面成绩斐然，但随着地区人口与产业高度集聚，城市规模持续扩张，城市运行中的能源消耗、交通出行等活动强度不断加大，空气污染状况却显著。空气质量作为生态核心指标，其改善本是环境优化与可持续发展的关键验证，如今却面临严峻挑战。尽管安徽省发展成就斐然，但面对人口产业集聚带来的空气污染挑战，空气质量的改善已刻不容缓、势在必行。

Hongliang Gu [1]参照 2019~2023 年 *TROPOMI* 数据，揭示了安徽省城乡空气污染差异，并研究发现景观格局与污染物浓度显著相关，研究为城乡差异化治理提供了依据。Długosz-Lisiecka Magdalena [2]应用因子法，分析 2010~2020 年 *B-7* 浓度与气象条件关系，发现其与气象、气溶胶浓度高度相关，证实温湿、日照与 *TSP* 浓度长期关系，并依据 11 年数据预测变化。董清浩[3]分析 2016~2021 年安徽省空气质量，发现 AQI 呈空间正相关且南北振荡，年降水量为主要影响因素，构建的 *SSA-SVR-SVR* 模型对合肥市 AQI 预测效果最优，建议实施区域差异化治理并推广组合预测方法；王航[4]针对传统空气质量预测模型精度低、参数经验化等问题，提出 *VMD-SSA-BiLSTM* 混合模型。通过各项对比筛选，结合 *PSO* 优化与 *SSA* 特征提取，优化模型结构。王学梅[5]分析 2015~2019 年安徽省 PM_{2.5} 时空分布特征，基于主成分分析建立 BP 神经网络和三次指数平滑组合模型，为研究区域 PM_{2.5} 治理提供科学依据。

综上所述，无论是在国内还是国际上，丰富的研究已经丰富了空气质量治理的理论与实践，本文剖析了 AQI 的时空特征，并构建了模型以实现精准预测，从而为区域差异化治理提供了科学依据，推动安徽省走持续发展之路。

2. 安徽省空气质量大数据集的构建

2.1. 研究区域概括

安徽省地处中国华东腹地，横跨长江、淮河两大流域，是长三角一体化战略与中部崛起战略的交汇枢纽，同时也是长江经济带的重要节点与先进制造业新兴增长极，其地理坐标大致为北纬 29°41'~34°38'，

东经 114°54'~119°37'(见图 1)。安徽省地形复杂多样,地势南高北低,水系纵横交织。由于地形复杂与气候多变的叠加效应,该地区局部区域存在空气污染扩散条件受限问题,加之工业化与城市化快速推进,大气污染防治面临结构性挑战。鉴于此,系统研究安徽省空气质量的时空演变规律及其驱动机制,对于科学制定区域协同治理策略、推动绿色低碳转型具有重大现实意义。

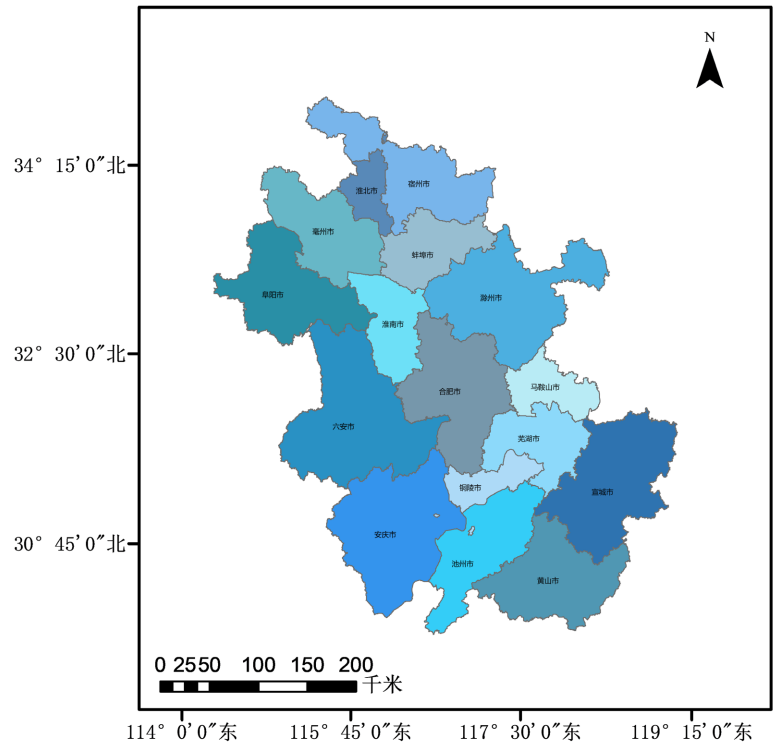


Figure 1. Administrative distribution map in Anhui province
图 1. 安徽省行政分布图

目前,空气质量监测系统通过多站点协同、多参数监测,构建了污染物浓度时空分布数据库。本研究使用的数据集来源于安徽省生态环境厅官网,选取了安徽省合肥市、芜湖市和安庆市等 16 个省级标准环境空气监测站点,采用 2018 年 1 月 1 日 0 时到 2024 年 12 月 31 日 23 时的小时级空气质量监测数据。所收集数据包含了六种主要空气污染物的浓度,以及空气质量指数 AQI 、主要污染物的种类和污染等级的情况。由于监测站点分布广泛且均匀,覆盖区域全面,通过对这些站点收集的数据进行分析,能够准确地反映出市区内部空气污染物的浓度变化情况。

2.2. 数据采集

本文所采用的数据均源自安徽省生态环境厅官方发布渠道,经严格筛选与整理后形成分析基础,具体数据选取范围及内容详见表 1 所示。

空气质量监测网捕获的空气质量数据是依照时间顺序进行采集的,气候变化和地理环境影响导致 AQI 数据存在噪声干扰和非线性等特征。由于直接使用原始历史数据不能较好地进行预测,本研究采用粒子群算法优化的变分模态分解对历史数据进行处理,再结合 $CNN-SLTM$ 构造组合模型,用于城市空气质量预测。为使论文的叙述呈现更具直观性与清晰性,在内容展示上,仅选取合肥市作为城市代表案例,并以 SO_2 和 $PM_{2.5}$ 这两项关键污染物作为重点分析内容。

Table 1. Air quality data indicators in Anhui province
表 1. 安徽省空气质量数据指标

站点	主要污染物	单位	指标释义
安徽省各地市	NO ₂	ug/m ³	用于表示每立方米体积中所含某种物质的质量是多少微克
	SO ₂	ug/m ³	用于表示每立方米体积中所含某种物质的质量是多少微克
	CO	ug/m ³	用于表示每立方米体积中所含某种物质的质量是多少微克
	O ₃	ug/m ³	用于表示每立方米体积中所含某种物质的质量是多少微克
	PM ₁₀	ug/m ³	用于表示每立方米体积中所含某种物质的质量是多少微克
	PM _{2.5}	ug/m ³	用于表示每立方米体积中所含某种物质的质量是多少微克

2.3. 构建空气质量指数大数据集

基于研究收集数据，通过数据筛选、异常数据剔除、丢失数据恢复、异构数据格式转换，对安徽省空气质量数据进行处理和清洗，主要包括以下步骤：

- ① 数据筛选：根据分析的需要，筛选出对研究内容有用的数据。
- ② 数据标准化：对数据进行标准化处理，统一编码。
- ③ 数据去重：去除重复的数据，避免对分析结果的影响。
- ④ 数据异常值处理：对数据进行初步筛选，删除数据中的异常值。
- ⑤ 数据缺失值处理：采用时间序列恢复、历史数据恢复的方法对缺失的数据进行恢复。
- ⑥ 海量数据的归纳：在保证数据安全性和完整性前提下，最大限度地减少海量数据存储空间。

3. 安徽省空气质量现状分析

3.1. 合肥市 AQI 指数变化特征剖析

对安徽省 2018~2022 年的空气污染物数据和空气质量指数进行分析，探究安徽省空气质量波动和变化趋势。其中空气主要污染物有 NO₂、SO₂、CO、O₃、PM10 和 PM2.5。依据年度特征，对安徽省空气质量数据进行分析。

根据表 2，以空气质量指数对应的颜色为背景，绘制 2018~2024 年各市 AQI 指数变化。

Table 2. Air quality index category and corresponding color code table
表 2. 空气质量指数类别及对应颜色对照表

空气质量指数	空气质量指数类别及表示颜色	
0~50	优	绿色
51~100	良	黄色
101~150	轻度污染	橙色
151~200	中度污染	红色
201~300	重度污染	紫色
>300	严重污染	褐红色

在生态环境备受关注的当下，空气质量指数成为衡量城市空气质量的关键指标。而 AQI 的变化能直观反映出当地空气质量动态。安徽省作为华东地区重要省份，其城市 AQI 现状反映着区域生态环境的健康程度。

以安徽省为例，如图 2 所示。

3.2. 安徽省各地市污染物浓度特征剖析

为明确各地市整体污染水平的高低趋势，精准筛选出污染重点防控地市，从而为大气污染防治策略的制定提供科学指引，特针对不同地市间各类污染物平均浓度的差异显著性进行分析，并据此绘制各地市污染物平均浓度面积图，如图 3 所示。

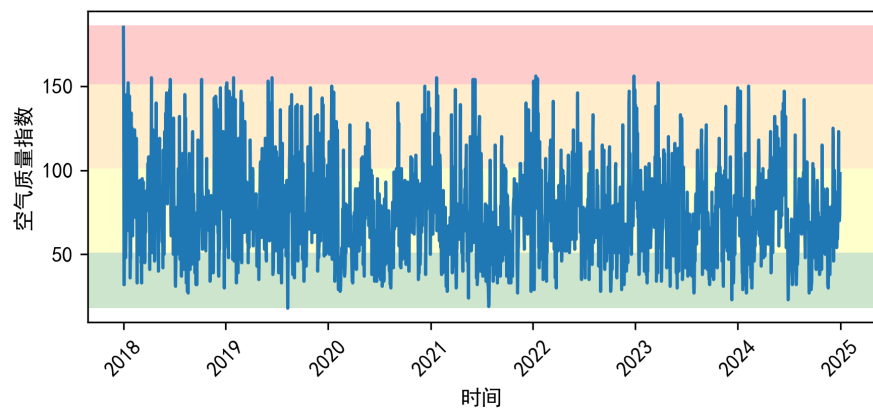


Figure 2. Changes in the AQI Index of Hefei from 2018 to 2024
图 2. 2018~2024 年合肥市 AQI 指数变化

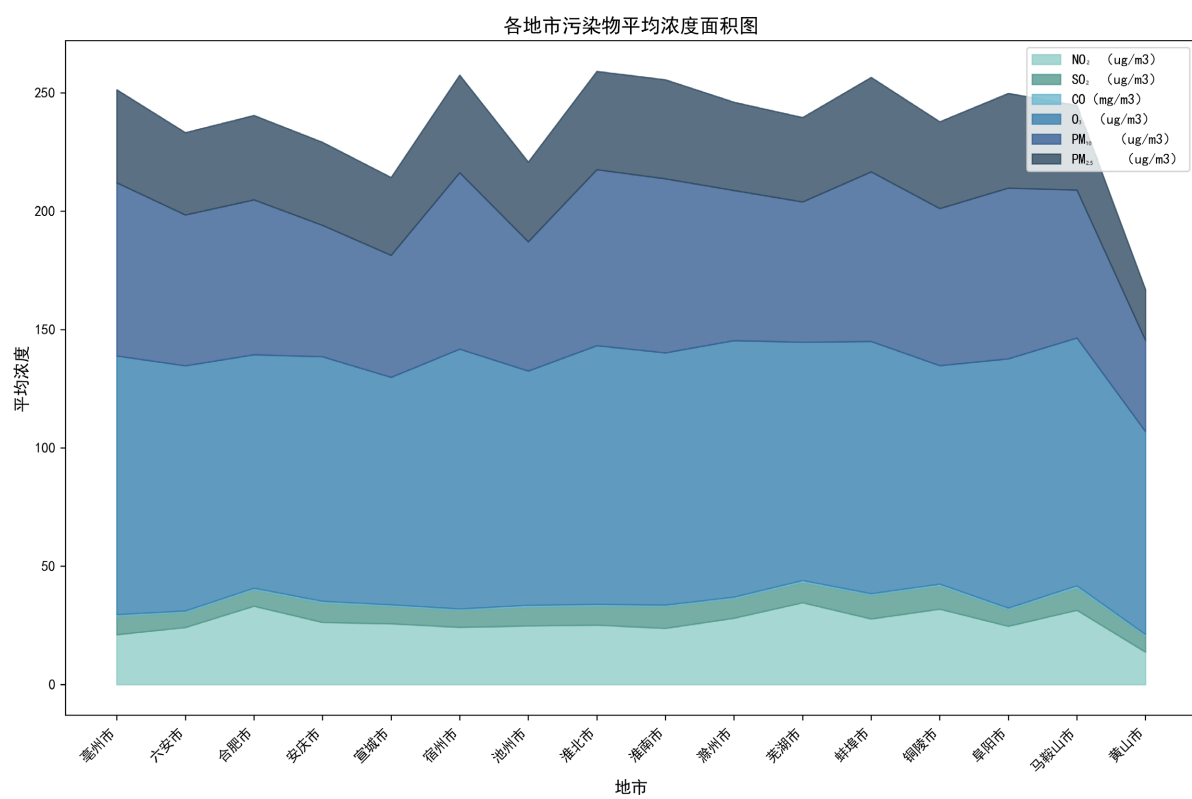


Figure 3. Average concentration area map of pollutants in various cities
图 3. 各地市污染物平均浓度面积图

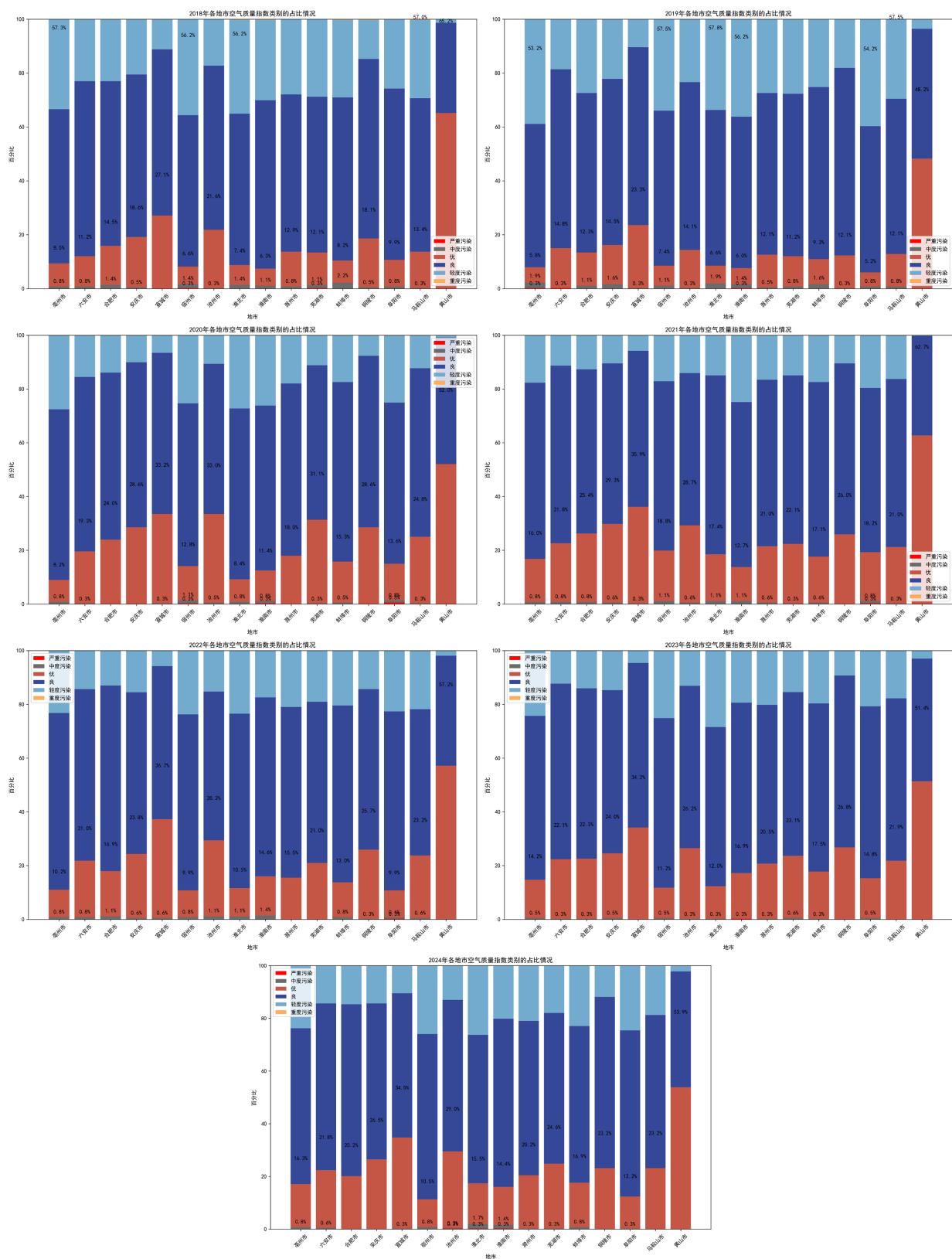


Figure 5. Air quality index stacked bar chart
图 5. 空气质量指数堆积柱状图

3.5. 主要污染物与 AQI 的 Pearson 相关性分析

空气中的主要污染物是计算 AQI 指数的主要特征，从上述分析中可以分析出 AQI 的年度变化特征与空气中各项污染物年度变化的特征密切相关。所以想要对 AQI 指数变化进行分析，必须对空气中各项污染物与 AQI 的相关性关系进行分析，找出对 AQI 指数影响最大的空气污染物，从而可以对空气治理提供更加针对性的方案。

采取 Pearson 相关系数法对空气质量因子和 AQI 之间关系作热力图，可以直观地看清空气质量因子之间的内在关联。公式如下：

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

通过将不同的变量带入 Pearson 公式，其中我们可以得到两个变量之间的相关性系数(见图 6)。根据判别相关性系数的趋向程度得出两个因素的相关程度。与 0 越接近表示相关程度在减弱，反之表示两者相关性在增加。

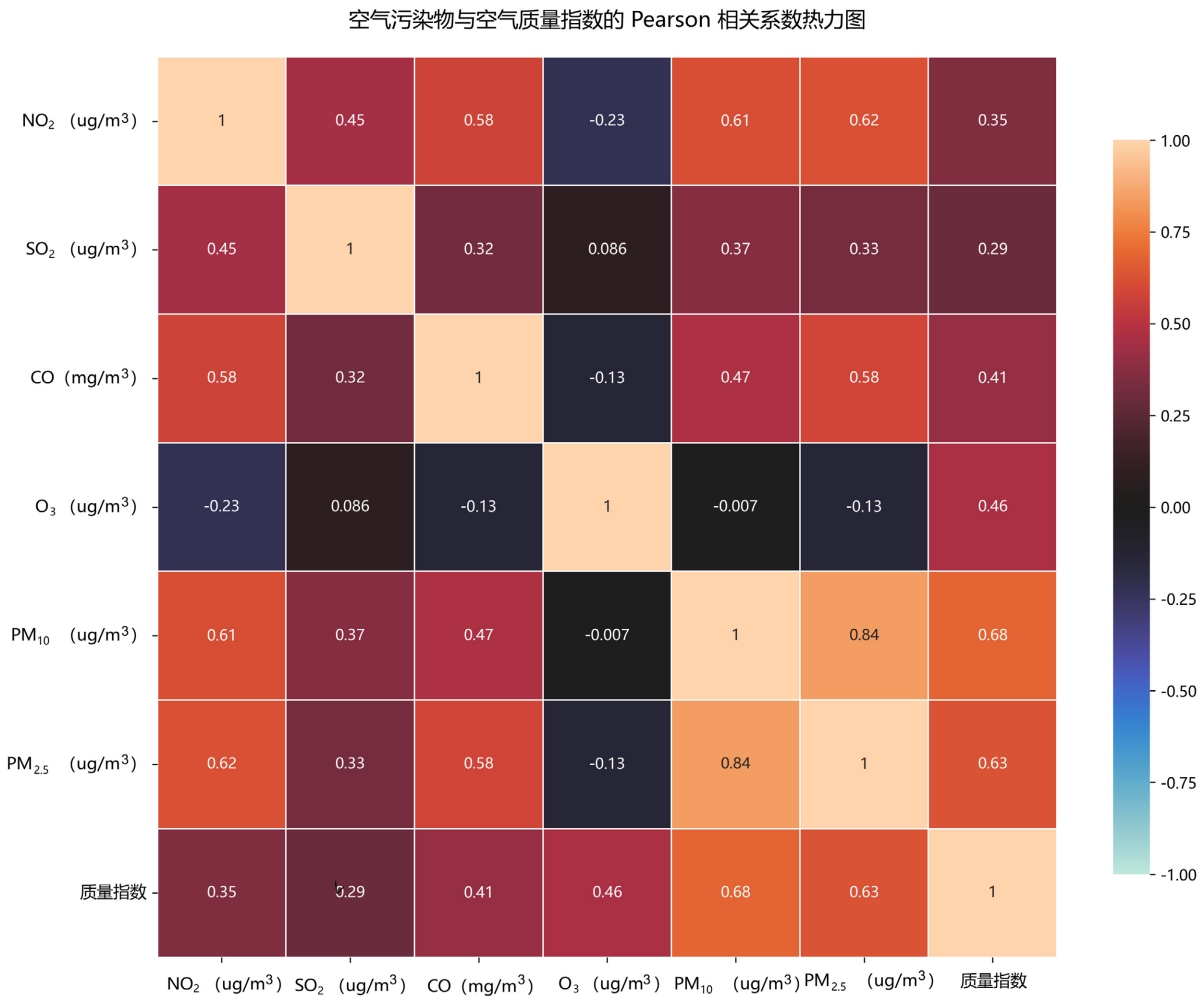


Figure 6. Main pollutant correlation heatmap
图 6. 主要污染物相关性热力图

4. 基于多源数据的安徽省空气质量预测剖析

进行安徽省空气质量预测,旨在通过分析工业、交通、生活等污染源的排放现状及未来变化,考量环保政策落实效果;研究风速、降水等气象条件和沙尘、逆温等特殊气象现象的影响;探究山地、平原等地形地貌对污染物扩散的作用;强化区域传输监测,加强与周边地区协同;利用空气质量监测数据构建模型;关注政策法规与减排措施,评估其对污染源排放的约束和对空气质量的改善成效,从而实现对安徽省空气质量的精准预测,为环境保护和决策制定提供科学依据。

4.1. 粒子群优化 VMD 分解算法

长短时记忆网络(LSTM)是一种特殊的循环神经网络,通过引入记忆单元和三个门控机制(输入门、遗忘门、输出门),能够有效地捕捉和记忆序列数据中的长期依赖关系,用于时间序列的预测。

APSO-VMD 算法实现步骤如图 7 所示:

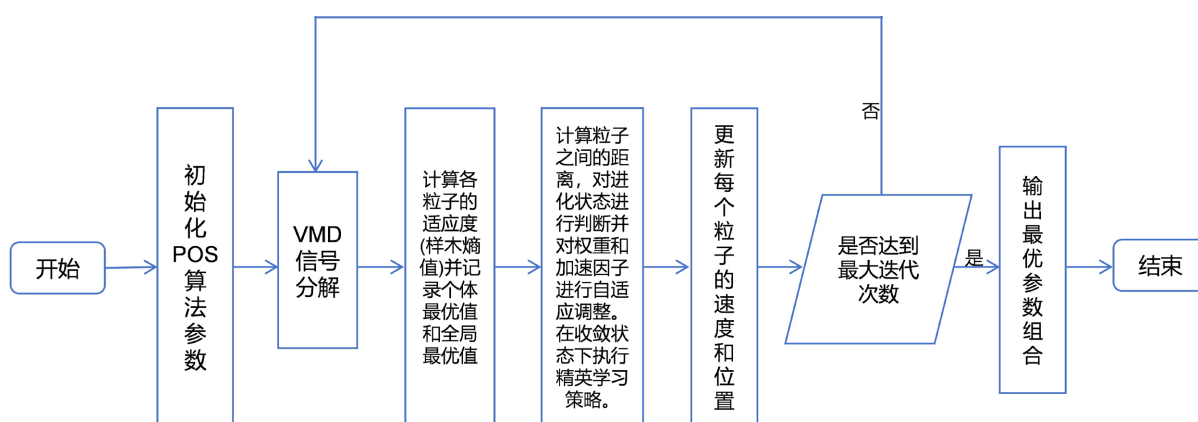


Figure 7. APSO-VMD algorithm flowchart

图 7. APSO-VMD 算法流程图

- ① 初始化: 设置 PSO 算法参数, 例如粒子数量、迭代次数、速度和位置的限制范围等;
- ② 初始化粒子群: 随机生成 N 个粒子, 每个粒子表示一组 VMD 参数 $[k, \alpha]$;
- ③ VMD 分解: 对于每个粒子, 利用对应的 VMD 参数对输入信号进行分解, 得到 K 个 IMF;
- ④ 适应度评价: 根据预设的适应度函数, 计算每个粒子的适应度值。适应度函数的设计应根据具体应用场景进行调整;
- ⑤ 更新粒子速度和位置: 根据 PSO 算法的更新规则, 更新每个粒子的速度和位置;
- ⑥ 更新个体最优和全局最优: 更新每个粒子的个体最优位置和全局最优位置;
- ⑦ 循环以上步骤, 直到满足终止条件;
- ⑧ 输出结果: 输出全局最优解, 即最优 VMD 参数 $[k, \alpha]$, 以及对应的 IMF 分解结果。

4.2. CNN-LSTM 混合模型原理

CNN-LSTM 混合架构借助时空特征耦合提升建模能力, 利用 CNN 的空间滤波器提取局部时序模式, 通过 LSTM 的门控机制建模全局动态演化规律, 进而提高模型预测性能。在应用该架构进行质量指数预测时, 先运用 VMD 分解原始质量指数数据序列, 得到平滑稳定的 IMF 子序列, 缓解原始数据的非线性问题; 接着将子序列与时间等特征合并构建特征矩阵并归一化; 然后把处理好的特征矩阵输入 CNN-LSTM 模型训练, 输出子序列预测值; 最终叠加重构预测子序列, 得到质量指数的最终预测结果。如图 8 所示:

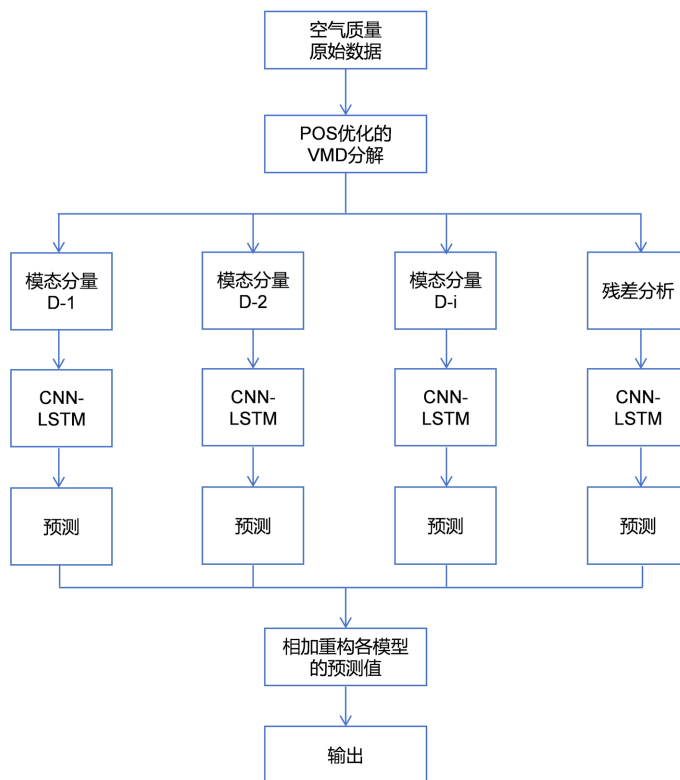


Figure 8. CNN-LSTM algorithm flowchart

图 8. CNN-LSTM 算法流程图

4.3. 预测安徽省未来 30 天的空气质量指数——以合肥市为例

拆分复杂数据、分离特征成分，给 *VMD-CNN-LSTM* 模型提供优质数据，提升安徽省空气质量预测精度与可靠性，基于 *PSO* 优化 *VMD* 分解原始质量指数数据序列。

图 9 中呈现了经 *PSO-VMD* 算法分解原始质量指数数据序列所得的 5 个 *IMF* 子序列。其中，*IMF1* 波动平缓，振幅与频率低，反映质量指数中稳定、变化慢的成分；*IMF2-IMF5* 波动复杂，*IMF3-IMF5* 更剧烈，体现原始数据中快速变化的高频成分。这是因为 *PSO-VMD* 算法借助粒子群优化找最优参数，能依信号局部特征自适应分解数据，且安徽省空气质量受工业排放、气象等多因素影响，这些因素在时间和频率上差异大，致使原始数据波动复杂，算法便将其按不同尺度和频率特征分解开来。

分析合肥市空气质量相关数据的信号特征，合肥市 2018~2024 年空气质量数据波动幅度较大，在不同时间点差异显著，数值在 25~175 左右的区间大幅波动，说明合肥市质量状况在这段时间内变化剧烈，不同时段质量水平差异明显。合肥市空气质量数据波动频繁且无固定规律，属于非平稳信号。

为助力安徽省空气质量的精准监测与高效管理，本研究构建 *CNN-LSTM* 预测模型，深入剖析每日实际质量指数与预测质量指数的波动特征，以及二者相对偏差的变化情况，如图 10，旨在精准探寻质量指数的波动规律，明晰预测偏差根源。

图 11 中展示了安徽省质量检测相关的每日实际质量指数波动、预测每日质量指数波动以及相对偏差变化。质量指数波动频繁，主要是因安徽省内地理环境多样，工业分布、植被覆盖、降水等环境因素复杂，且受季节性气象条件如冬季取暖燃煤排放、夏季降水等影响。预测值与实际值前期偏差波动大后期渐稳，原因在于预测模型数据可能覆盖不全，未充分考虑突发因素，模型参数及算法存在局限，不过随着时间推移通过优化等手段使偏差逐渐稳定。

通过对比实际与预测值的偏差,为空气质量防控和治理提供科学决策依据,使用 *PSO-VMD* 对数据进行分解,将各子序列与时间等特征合并构建特征矩阵,经归一化处理后输入 *CNN-LSTM* 模型训练,输出子序列预测值,从而实现对空气质量的精准分析与预测。

图 12 中展示了不同 *IMF* 分量的质量指数波动曲线,包括实际、预测及向后预测 30 天的情况。实际与预测曲线有一定相似性,说明 *CNN-LSTM* 模型在一定程度上能捕捉质量指数波动特征。但各曲线间也存在明显偏差,尤其向后预测 30 天的曲线差异更大。

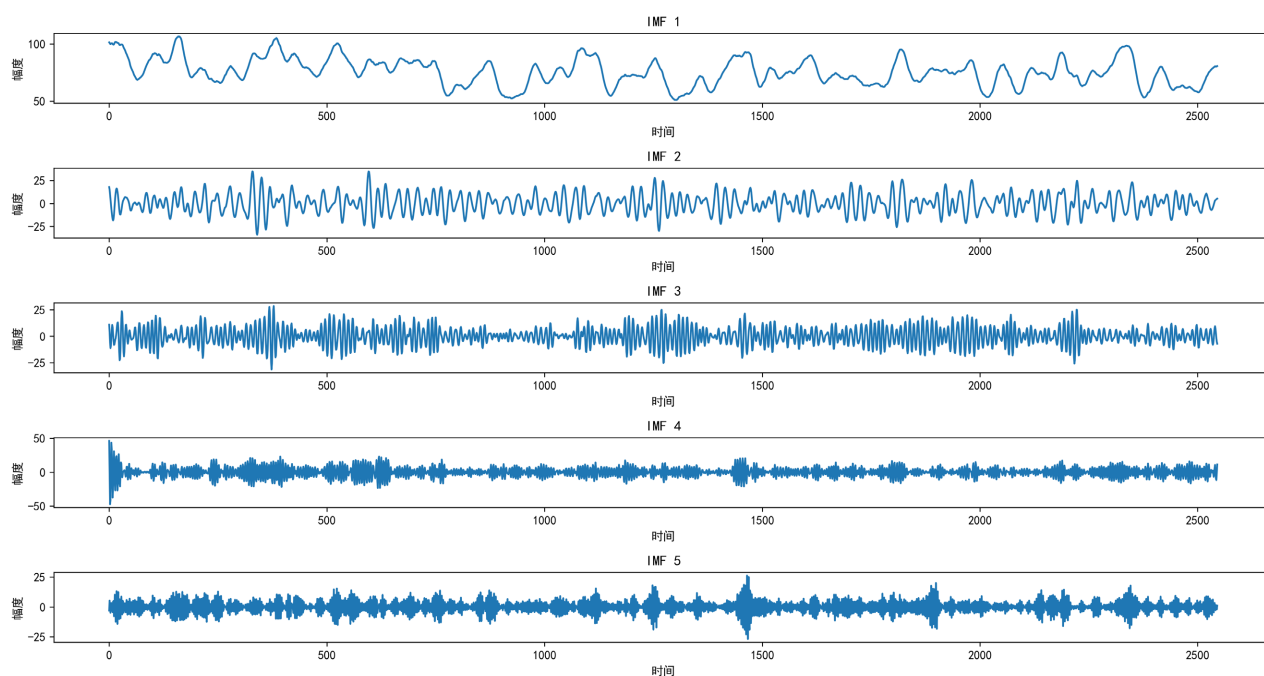


Figure 9. Time-frequency domain waveform of *IMF* components after *PSO-VMD* decomposition

图 9. 基于 *PSO-VMD* 分解后 *IMF* 分量的时频域波形图

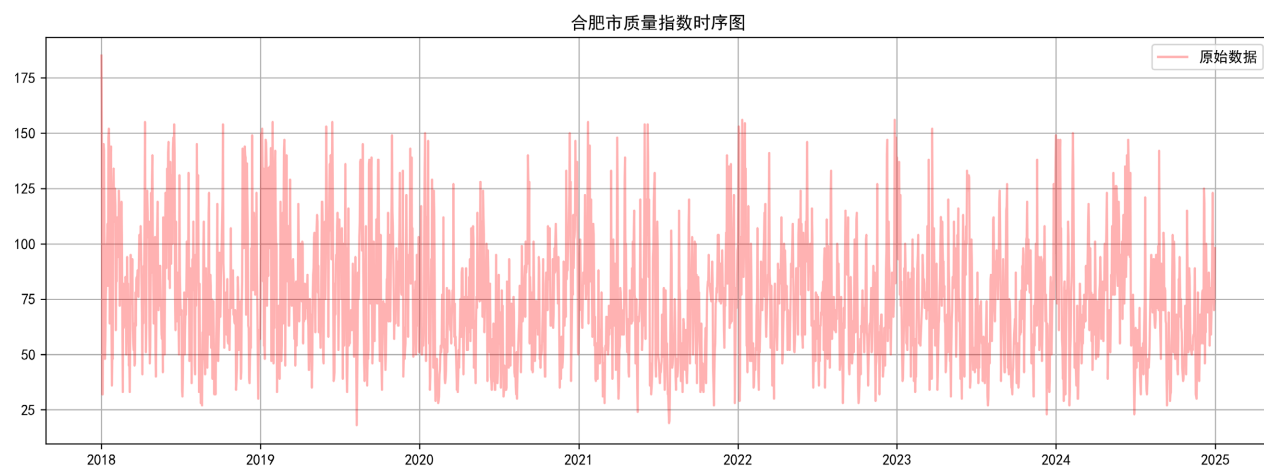


Figure 10. Original data time series chart

图 10. 原始数据时序图

原因在于,一方面空气质量受气象、人为排放等多种复杂因素影响,模型难以完全涵盖这些动态变化;另一方面,预测时间跨度增加,不确定因素增多,模型对长期趋势的把握能力有限。此外,不同 *IMF*

分量特征复杂程度有别，模型对复杂波动特征的学习不够精准，导致预测存在偏差。

构建 *VMD-CNN-LSTM* 模型并叠加重构预测子序列以输出质量指数预测结果，是为精准预测安徽省空气质量指数，为管控决策提供科学依据。

图 13 中呈现了 *VMD-CNN-LSTM* 模型的测试集实际质量指数波动曲线、测试集预测质量指数波动曲线及向后预测 30 天质量指数波动曲线。红色预测曲线与青色实际曲线多数时候波动趋势相近，说明模型能捕捉部分规律，但也有偏离，存在误差；蓝色向后预测 30 天曲线末端与前两者差异大。这是因为 *VMD-CNN-LSTM* 模型虽结合多种优势能学习数据规律，可空气质量影响因素复杂，模型难以完全覆盖。且预测时间跨度增至 30 天，不确定因素增多，模型应对能力有限，导致预测误差增大。预测结果如表 3 所示：

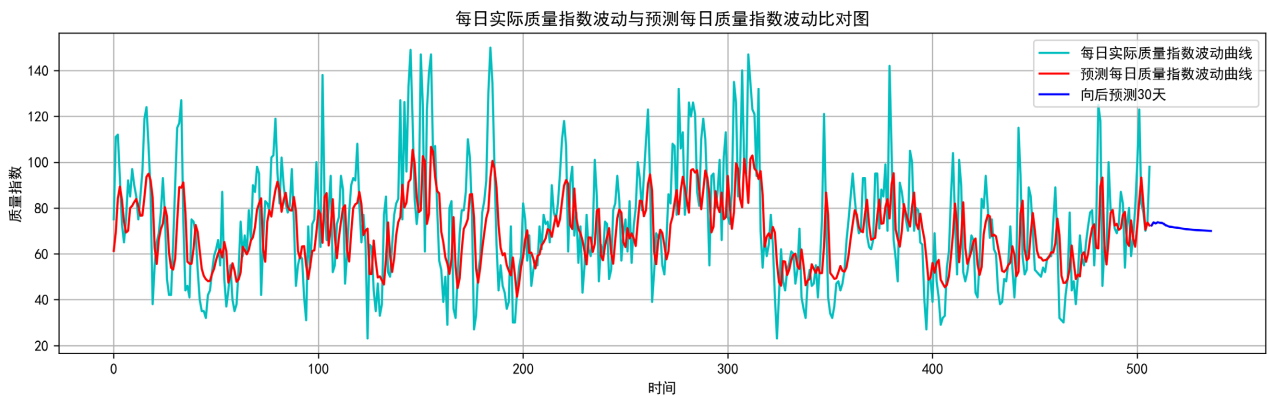
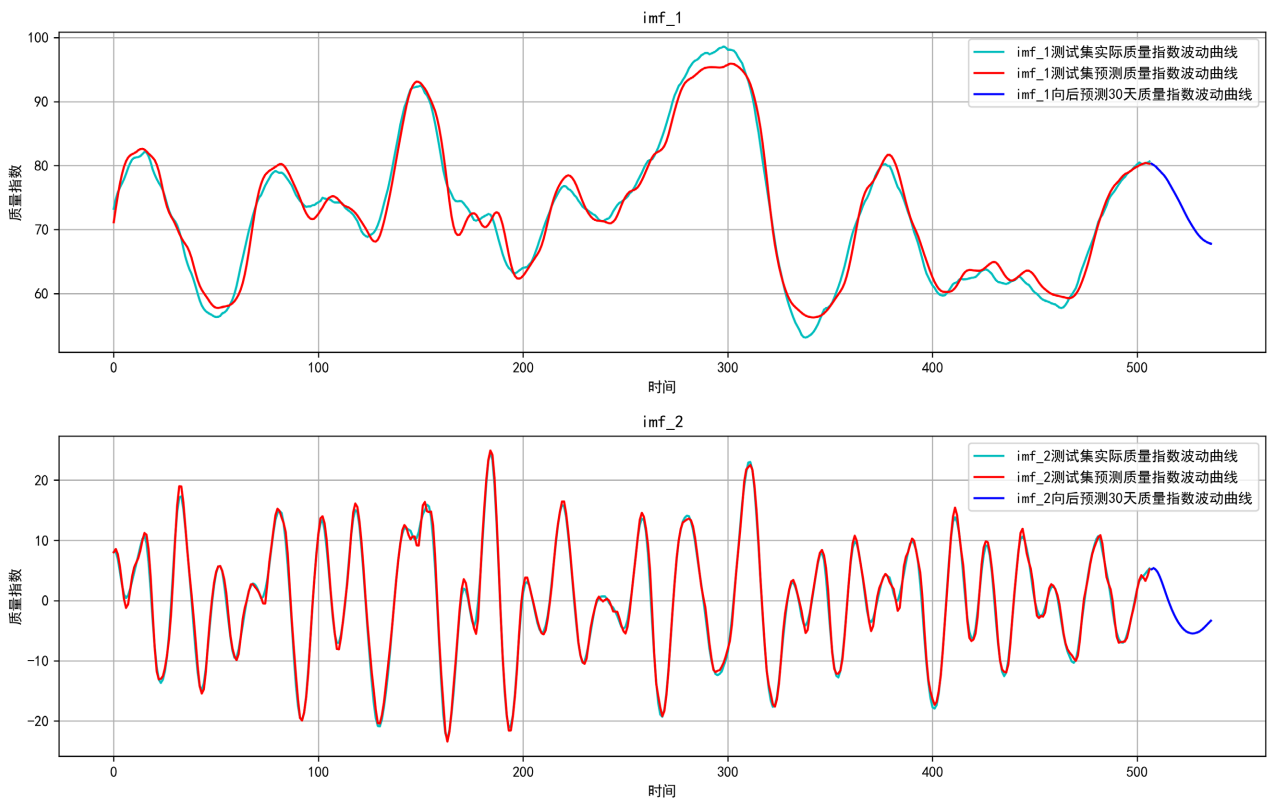


Figure 11. Daily actual quality index fluctuation vs. Predicted daily quality index fluctuation comparison chart

图 11. 每日实际质量指数波动与预测每日质量指数波动对比图



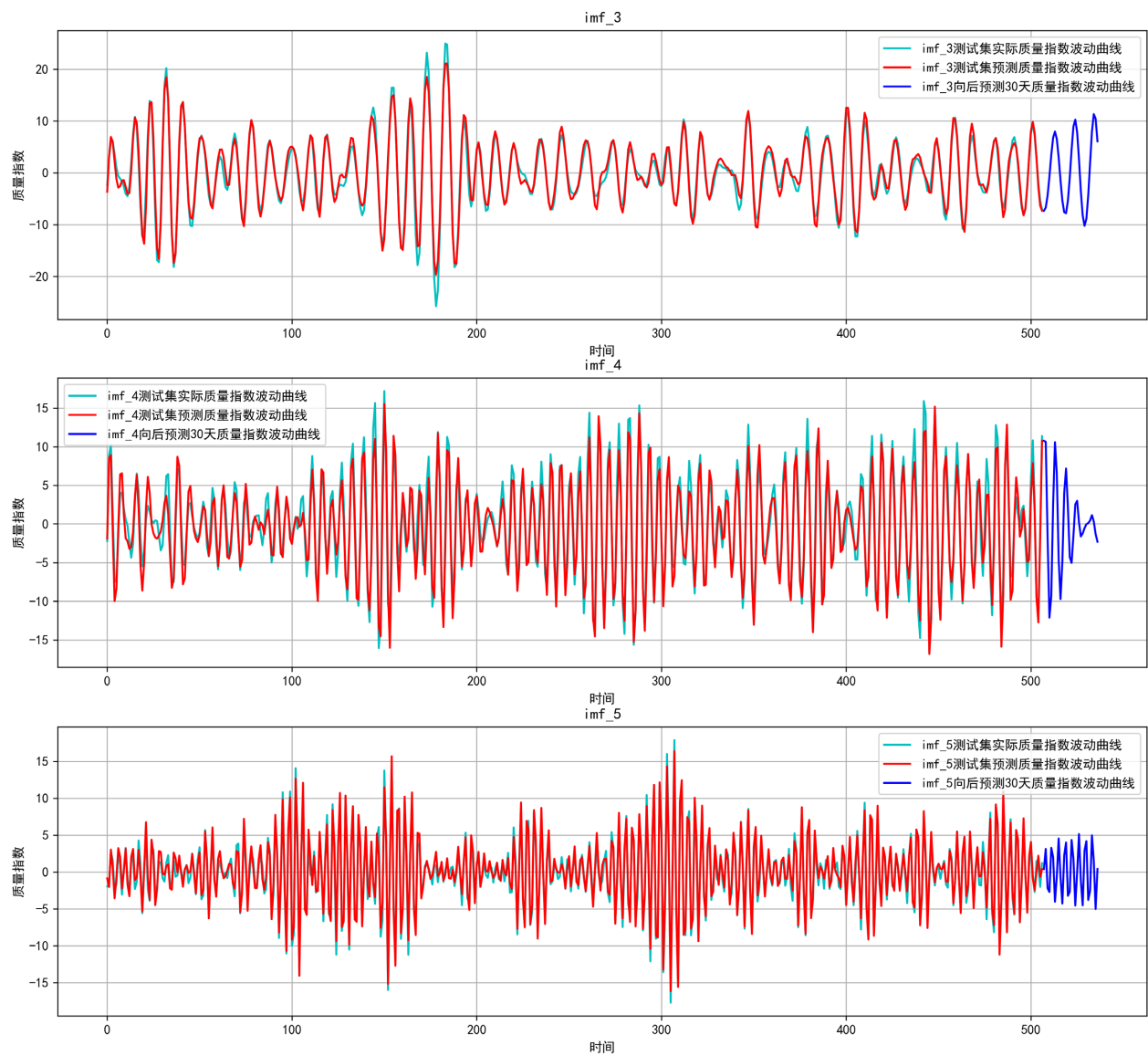


Figure 12. Predicted post-subsequence volatility curve

图 12. 预测后子序列的波动曲线图

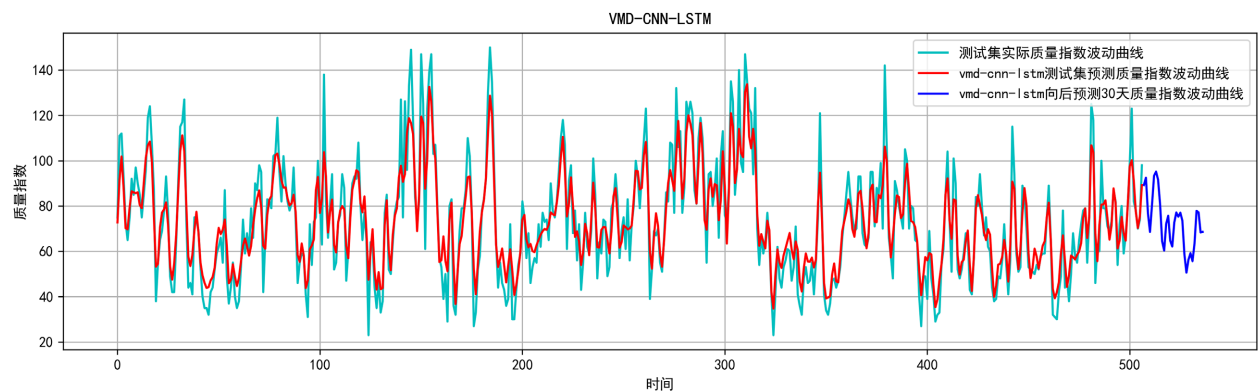


Figure 13. Comparison chart of air quality index fluctuations in Hefei city

图 13. 合肥市空气质量指数波动对比图

Table 3. Prediction results table in Hefei city
表 3. 合肥市预测结果表

合肥市	
未来三十天	空气质量指数预测值
1	90.23
2	93.12
3	76.97
4	70.56
5	82.30
6	92.10
7	94.53
8	94.53
9	83.82
10	65.36
11	62.83
12	76.30
13	77.58
14	67.05
15	68.28
16	77.00
17	74.65
18	73.70
19	81.61
20	76.89
21	58.65
22	53.89
23	59.94
24	55.76
25	53.18
26	67.97
27	78.61
28	69.32
29	63.79
30	69.32

本研究以合肥市为例，致力于预测安徽省未来 30 天的空气质量指数。分析合肥市 2018~2024 年空气质量数据发现，其波动幅度大、无固定规律，属于非平稳信号。接着，剖析每日实际与预测质量指数的波动特征及偏差变化。研究发现，空气质量指数波动受地理环境、工业分布、气象条件等多种复杂因素

影响, 预测值与实际值前期偏差波动大, 后期因模型优化渐趋稳定。不同 *IMF* 分量预测曲线显示, 模型虽能捕捉部分波动特征, 但受环境因素复杂多变、预测时间跨度增加以及模型对复杂波动特征学习不精准等因素影响, 存在明显偏差。最终, 输出合肥市未来 30 天的空气质量指数预测值, 为安徽省空气质量管控决策提供了科学依据, 但模型仍有优化空间, 后续研究可进一步完善, 以提升空气质量预测的准确性和可靠性。

5. 结论

时间维度上, 季度变化呈现“U”型态势, 夏季受光化学反应驱动, 臭氧等污染物浓度攀升, 拉高整体污染水平; 年初年末则因气象扩散条件优越及季节性生产活动更迭, 污染物浓度相对趋低。年度视角下, 各地市空气质量等级占比在不同年份间波动明显。空间层面, 各地市污染物平均浓度差距悬殊, 部分地市污染负荷较重, 是污染防治的关键区域。以合肥市为例, 2018~2024 年 *AQI* 指数波动频繁且幅度大, 在 25~175 区间大幅震荡, 空气质量动态变化剧烈。借助 *Pearson* 相关性分析, 明确空气中主要污染物与 *AQI* 指数关联紧密。其中, *PM*10、*PM*2.5 等污染物与 *AQI* 指数相关性尤为突出, 对 *AQI* 指数的起伏变化影响显著。构建的 *VMD-CNN-LSTM* 模型具备一定捕捉空气质量指数波动特征的能力, 测试集内预测曲线与实际曲线在多数时段波动趋势契合, 但仍存在偏差。当向后预测 30 天, 曲线与测试集曲线差异显著, 这源于预测时间跨度延展致使不确定因素增多, 模型对长期趋势的把控力不足, 且不同 *IMF* 分量特征的复杂程度差异也干扰了预测精度。

总体而言, 本研究全面剖析了安徽省空气质量的现状、污染物关联及预测模型表现。后续可基于这些结论, 进一步优化预测模型, 融入更多影响因子, 提升模型精度与可靠性。同时, 为政府及相关部门制定科学的空气质量管控策略、优化资源配置提供有力的数据支撑与决策参考, 助力安徽省空气质量持续改善。

参考文献

- [1] Gu, H. and Zhang, W. (2025) Quantitative Assessment of Urban-Rural Spatiotemporal Heterogeneity in Air Pollutants Using GEE Multi-Source Data across the Anhui Province, China. *Atmospheric Pollution Research*, **16**, Article 102464. <https://doi.org/10.1016/j.apr.2025.102464>
- [2] Długosz-Lisiecka, M. and Nowański, S. (2023) Factor Regression as a Tool to Predict Be-7 Concentration in the Air. *Atmospheric Environment*, **297**, Article 119612. <https://doi.org/10.1016/j.atmosenv.2023.119612>
- [3] 董清浩. 安徽省空气质量空间统计分析及基于 SSA 的组合预测[D]: [硕士学位论文]. 合肥: 安徽建筑大学, 2024.
- [4] 王航. 基于粒子群优化的 VMD-SSA-BiLSTM 模型的合肥市空气质量预测研究[D]: [硕士学位论文]. 合肥: 合肥大学, 2024.
- [5] 王学梅. 安徽省 *PM*2.5 时空变化及其气象影响因素分析[D]: [硕士学位论文]. 合肥: 安徽农业大学, 2020.