邓田丰,李梓铭,张少波,等.基于机器学习和时间序列算法融合的北京地区花粉浓度预测技术研究 [J]. 中国环境科学, 2025,45(11):6007-6019.
Deng T F, Li Z M, Zhang S B, et al. Research on pollen concentration prediction technology in Beijing based on the fusion of machine learning and time series algorithms [J]. China Environmental Science, 2025,45(11):6007-6019.

# 基于机器学习和时间序列算法融合的北京地区花粉浓度预测技术研究

邓田丰[1],李梓铭[2*],张少波[1**],孙兆彬[3] (1.成都信息工程大学大气科学学院/高原大气与环境四川省重点实验室/气象灾害预测预警四川省工程研究中心/成都平原城市气象与环境四川省野外科学观测研究站,四川 成都 610225；2.北京市气象台,北京 100089；3.中国气象科学研究院气象影响与风险研究中心,北京 100081)

**摘要**：基于机器学习和时间序列算法,对北京市顺义区 2018~2024 年的花粉浓度进行了时间序列模拟预测研究.研究整合了传统统计模型、机器学习模型和深度学习模型,并引入 36 种气象要素及其变量作为协变量以提升预测精度.结果显示,AutoGluon-TimeSeries(AG-TS)通过模型集成显著降低了过拟合风险,在短期预测(1d)中表现优异,其中 DLinear 模型以 $R$=0.68 和 RMSE=206.5 的综合性能最佳.相比之下,部分模型在预测时有高估或低估现象.研究发现,花粉浓度呈现春季(3~5 月)和秋季(8~10 月)双峰特征,与气象变量的相关性反映了气象条件和植被物候的复杂影响.此外,模型性能随预测时间延长衰减,如 DLinear 在 3d 预测中 $R$ 值下降 18%,RMSE 上升 15%.通过剔除低效模型并优化剩余模型,为花粉过敏预警提供了模型优化的方向和科学支持.

**关键词**：时间序列；预测；花粉浓度；机器学习；模型集成
**中图分类号**：X83 **文献标识码**：A **文章编号**：1000-6923(2025)11-6007-13

**Research on pollen concentration prediction technology in Beijing based on the fusion of machine learning and time series algorithms.** DENG Tian-feng[1], LI Zi-ming[2*], ZHANG Shao-bo[1**], SUN Zhao-bin[3] (1.Chengdu Plain Urban Meteorology and Environment Observation and Research Station of Sichuan Province, Sichuan Provincial Engineering Research Center for Meteorological Disaster Prediction and Early Warning, Plateau Atmosphere and Environment Key Laboratory of Sichuan Province, School of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu 610225, China；2.Beijing Meteorological Station, Beijing 100089, China；3.Center for Meteorological Impact and Risk Research, Chinese Academy of Meteorological Sciences, Beijing 100081, China). *China Environmental Science*, 2025,45(11)：6007~6019

**Abstract**：In this paper, a time-series simulation prediction study of pollen concentration in Shunyi District, Beijing, China, for the period of 2018~2024 was carried out based on machine learning and time-series algorithms. The study integrates traditional statistical models, machine learning models and deep learning models, and introduces 36 meteorological elements and their variables as covariates to improve the prediction accuracy. The results show that AutoGluon-TimeSeries (AG-TS) significantly reduces the risk of overfitting through model integration, and performs well in short-term prediction (1day), with DLinear model having the best overall performance with $R$=0.68 and RMSE=206.5. In contrast, some models overestimated or underestimated the predictions. It was found that the pollen concentration showed double peaks in spring (March~May) and fall (August~October), and the correlation with meteorological variables reflected the complex effects of meteorological conditions and vegetation phenology. In addition, the modeling performance declined with the extension of prediction time, e.g., the $R$-value of DLinear decreased by 18% and the RMSE increased by 15% in the 3-day prediction. By eliminating the inefficient models and optimizing the remaining models, the direction of model optimization and scientific support for pollen allergy early warning were provided.

**Key words**：time series；forecasting；pollen concentration；machine learning；model ensemble

花粉过敏对公共卫生的影响日益显著[1].根据世界卫生组织(WHO)的统计,全球约 30%人口受季节性过敏性鼻炎困扰,花粉暴露是其主要诱因[2-4].在中国,随着城市化快速推进和植被结构变化,花粉浓度在时空分布上呈现明显波动,导致过敏性疾病发病率逐年上升[5-6].以北京市为例,随着花粉浓度增加,过敏性鼻炎和哮喘的风险显著上升,而在相同花粉浓度下,过敏与更高的哮喘风险相关[5].花粉过敏不仅仅会引发鼻塞、结膜炎等短期症状,长期的花粉暴露还可能导致哮喘症状的加重、睡眠障碍以及认知

**收稿日期**：2025-04-09
**基金项目**：中国气象局创新发展专项重点项目(CXFZ2025J074);北京市气象局青年创新团队(CXTD202506)
* 责任作者，高级工程师，lyctze1986@gmail.com; ** 副教授，zsb@cuit.edu.cn

功能的下降[7-8].全球由此产生的直接医疗费用以及间接生产力的损失,每年的总额已经达到了数 10 亿元人民币[9].因此,建立高精度花粉浓度预测体系不仅是环境监测的技术需求,更是公共卫生管理中亟待完成的任务.

围绕花粉预测,国内外学者已探索多种建模方法,包括传统统计方法、机器学习及数值模式等.传统方法主要基于统计模型,如多元回归、时间序列分析、自回归积分滑动平均等,依赖历史花粉数据并结合气象因素进行预测..例如,多元回归可以较好地模拟花粉浓度与气象变量的关系[10],但该方法在处理非线性时存在明显局限性[11].此外,ARIMA 等时序方法能够较好地捕捉时间上的趋势和季节性变化,但它对非平稳时间序列的建模能力有限,且难以适应突发性气象变化[12].因此,尽管传统方法在早期研究中占主导,因其对复杂环境适应能力不足,已逐渐被更先进的机器学习和数值模式方法取代.

近年来,机器学习方法在花粉预测中取得了显著进展.与传统统计方法相比,机器学习方法能够更好地处理高维、非线性和复杂交互的数据,提高预测精度[13].例如,支持向量机(SVM)、随机森林(RF)、极端梯度提升(XGBoost)和神经网络(ANN)等[14-15]方法已被广泛应用于花粉预测研究.其中 RF 方法通过集成多个决策树提高了预测稳定性,并在多项研究中表现出较高的准确度[14].深度学习方法,如卷积神经网络(CNN)和长短时记忆网络(LSTM),在处理时序数据时展现出强大的建模能力.特别是 LSTM 在长期时间依赖性建模方面具有天然优势,使其在花粉浓度预测中表现突出[16].然而,机器学习仍面临数据需求大、可解释性低、计算资源消耗高等挑战,且花粉数据时空异质性较大,模型泛化能力往往较差.

数值模式是另一类常用于花粉预测的方法,主要包括物理-化学过程模型和大气传输模型.数值模式方法通过模拟花粉的释放、传播和沉降过程,结合气象模式(如 WRF-Chem、RegCM-pollen),可以在较大空间尺度上提供高分辨率预测[17-19].相比统计和机器学习方法,数值模式能够更好地描述花粉浓度的物理变化过程,并具备较强的可解释性.然而,该方法对初始条件和输入参数(如花粉释放率、风场、降水等)高度敏感,计算成本高昂,并且在短时间尺度上预测精度不稳定[20].因此,在实际应用中,研究人员常常将数值模式与机器学习方法结合,以优化预测性能.

由于花粉浓度具有的时间相关性和季节性特征,本文基于北京市顺义区连续 7a(2018~2024)的花粉观测数据和气象数据使用时间序列预测模型来预测花粉浓度,时间序列预测模型在处理时变数据时具有天然优势,因为它可以直接学习数据的时序特征,而无需手动提取时间序列特征[21].特别是 Transformer 在近年的发展,使得基于注意力机制的预测模型能够更好地处理长时间跨度的浓度数据[22-23],从而提高预测准确性,这为未来精准花粉预警系统的构建提供了新的思路.

## 1 资料与方法

### 1.1 花粉数据

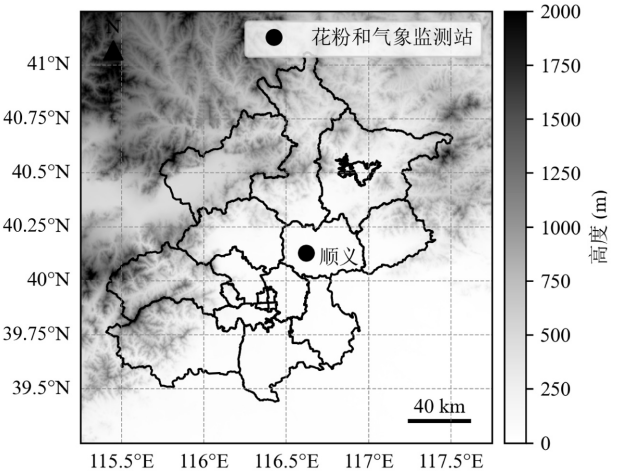



图 1 北京市顺义区花粉监测站点

Fig.1 Pollen monitoring station in Shunyi District, Beijing

背景着色为海拔高度

本研究基于北京市顺义区气传花粉观测站(图 1)2018~2024 年花粉季节(3~10 月)的每日花粉浓度观测数据展开,旨在构建并优化区域花粉浓度预测模型.数据采集期间,花粉由于重力作用自然沉降在采样器表面,通常 24h 采样结束后,使用琼脂碱性复红染色剂对花粉进行染色,并在光学显微镜下进行人工读数.为确保数据质量,本文基于花粉浓度的统计分布特征对原始数据进行了严格的质量控制,通过分析发现,当阈值设定为 $2\times10^3$ 粒/($m^2$·d)时,可涵盖 99%以上的有效观测数据,故剔除了超过 $2\times10^3$ 粒/($m^2$·d)的极值,以消除异常值对模型训练的干扰,从而保障数据的完整性和可靠性.同时对缺测数据

采用线性插值法进行填充,以保持时间序列的连续性.在数据使用方面,2018~2022 年的数据被用于模型的训练与参数优化,而 2023~2024 年的数据则作为独立测试集,用于评估模型的预测性能与泛化能力.图 2 展示了用于模型训练和测试的每日花粉浓度时间序列图,图中清晰呈现了花粉浓度的季节性波动特征,包括春季和秋季的两个显著峰值,这与北京市主要致敏植物的花期高度吻合.

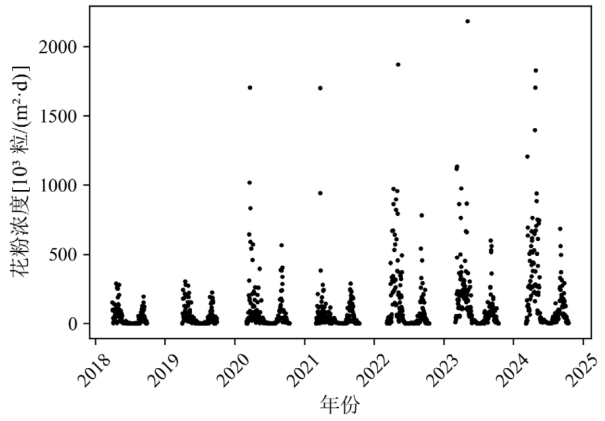



图 2　2018~2024 年北京市顺义区模型训练和预测的实际日总花粉浓度

Fig.2　Actual daily total pollen concentration for model training and prediction in Shunyi District, Beijing from 2018 to 2024

## 1.2　气象数据

气象条件可以通过热力学和动力学作用调控花粉的时空分布,是影响花粉浓度水平的最重要控制因素之一,其变化不仅直接影响植物物候进而决定花粉的释放,还通过风和降水调控花粉的扩散和沉降[24-25].为了全面量化气象条件对花粉浓度的影响,本研究使用的气象数据来自北京市顺义区的国家气象监测站,站号为 54398,坐标为 40.11°N,116.63°E,该数据系统性地整合了包括日平均气压、日平均气温、日降水量、日平均相对湿度、日平均风速、日照时数等在内的 36 种气象要素(表 1),涵盖了大气热力、动力及水分条件的关键指标,能够为花粉浓度的时空分布提供多维度的气候背景信息.此外,研究还引入了年和月日计数等时间特征,以捕捉花粉浓度的季节性波动规律.针对气象数据中可能存在的缺测值,研究采用线性插值方法进行填充,确保数据连续性和完整性.为进一步挖掘气象条件对花粉浓度的潜在影响,研究计算了有效积温(通过累积日平均气温高于发育起点温度的温度值,反映植物生长的热量条件),能够捕捉气象条件对花粉浓度的累积效应和非线性响应,为模型提供更丰富的输入信息.在数据预处理阶段(图 3),气象数据与花粉观测数据通过站点编号和时间戳进行精确对齐,构建了一个多变量时间序列数据集,其中每个时间点的数据包括花粉浓度及其对应的气象条件特征.这种多变量时序数据集的构建不仅能够反映花粉浓度与单一气象因子的关系,还能揭示多气象要素协同作用下的复杂机制,例如高温低湿天气可能促进花粉释放,而强风则可能加速花粉扩散,但降雨又可能通过冲刷作用暂时降低空气中花粉浓度[26].通过将气象数据作为额外的特征输入到模型中,本研究能够更全面地评估气象条件对花粉浓度的影响,并为后续构建高精度预测模型奠定数据基础.

**表 1　气象要素的含义和单位**

Table 1　Meanings and units of meteorological elements

| 物理量名称 | 单位 | 物理量名称 | 单位 |
|---|---|---|---|
| 日平均气压(prs_mean) | hPa | 日极大风速(gust_max) | m/s |
| 日最高气压(prs_max) | hPa | 日平均地表温度(st_mean) | ℃ |
| 日最低气压(prs_min) | hPa | 日最高地表温度(st_max) | ℃ |
| 日平均海平面气压(msl_mean) | hPa | 日最低地表温度(st_min) | ℃ |
| 日平均气温(t2m_mean) | ℃ | 日平均 5cm 地面温度(stl5) | ℃ |
| 日最高气温(t2m_max) | ℃ | 日平均 10cm 地面温度(stl10) | ℃ |
| 日最低气温(t2m_min) | ℃ | 日平均 15cm 地面温度(stl15) | ℃ |
| 日平均相对湿度(rh2m_mean) | % | 日平均 20cm 地面温度(stl20) | ℃ |
| 日最低相对湿度(rh2m_min) | % | 日平均 40cm 地面温度(stl40) | ℃ |
| 日平均水汽压(e_mean) | hPa | 日平均 80cm 地面温度(stl800) | ℃ |
| 日降水量(tp) | mm | 日平均 1.6m 地面温度(stl1600) | ℃ |
| 日夜间降水量(tp_night) | mm | 日平均 3.2m 地面温度(stl3200) | ℃ |
| 日白天降水量(tp_day) | mm | 日平均草面温度(ct_mean) | ℃ |
| 日小型蒸发量(zf_small) | mm | 日最高草面温度(ct_max) | ℃ |
| 日大型蒸发量(zf_big) | mm | 日最低草面温度(ct_min) | ℃ |
| 日平均风速(ws_mean) | m/s | 日平均总云量(tcc) | 1/10 |
| 日平均 10min 风速(ws10_mean) | m/s | 日平均低云量(lcc) | 1/10 |
| 日最大风速(ws_max) | m/s | 日照时数(rzh) | h |

注:所有变量均基于逐小时观测数据计算.

## 1.3　时间序列预测

时间序列预测方法是一种通过对历史数据中的时序依赖性建模,从而推断未来某一时间段内变

量变化的统计学方法[27].其核心假设是数据在时间维度上存在潜在规律,如趋势性、季节性和随机性,且未来状态与历史状态具有关联性[28-29].时间序列预测方法广泛应用于气象、金融、交通等领域[30],其模型大致可分为传统统计模型、机器学习模型和深度学习模型三大类.传统统计模型在捕捉季节性波动和气象影响方面优于传统的历史平均法,能够捕捉花粉数据的季节性和自相关性,结合气象变量后显著提升预测能力,但需要长期高质量数据校准参数,短时间或缺失数据可能导致模型不稳定,其“黑箱”特性使其难以直观解释变量间具体关系[31].机器学习模型可以通过将时间序列转换为基于窗口的回归任务[32],并引入特征工程,还能够捕捉时间序列的自相关性和动态依赖关系[33].深度学习模型,如 LSTM 通过记忆单元和门控机制(输入门、遗忘门、输出门)能有效捕捉花粉数据中的长期依赖关系和非线性动态.还可以通过卷积层过滤噪声,提取关键气象特征,提升模型鲁棒性[34].对单峰季节性和全年分布均表现出高适应性,尤其在预测突发性花粉事件时优于传统方法.但面临计算复杂度高、数据质量依赖和解释性不足等挑战[34].

在本研究中,本文使用 AutoGluon 的 TimeSeries Predictor 工具进行模型训练, AutoGluon-Time Series(以下简写为 AG-TS)是一个开源的自动化机器学习库,专注于概率时间序列预测[35-36].其核心设计理念是易用性和鲁棒性,用户仅需少量代码即可生成高精度的点预测和分位数预测.通过集成多样化的模型(统计模型、机器学习模型、深度学习模型)减少过拟合风险,AG-TS 在有限时间内可以实现高准确性的预测[37].然而,对极端长序列的扩展性、小数据集的适应性以及功能扩展仍是未来改进的重点[38].

AG-TS 采用动态层次化集成策略:首先在扩展模型空间(包含传统统计模型、深度学习模型和预训练时序大模型)中进行多样性采样;然后通过滑动窗口验证评估各模型在近期数据上的表现;最终使用带正则化的加权平均算法生成集成预测,其中权重系数通过最小化验证集的 MASE 指标优化获得.该机制既能充分利用各模型优势,又能通过持续更新的权重调整适应数据分布变化.配置中指定了时间步长为 3d、MASE 为评估指标以及气象数据为协变量.AutoGluon 学习了前面各种类型模型的特点并在最后进行了集成,以便 AutoGluon 自动选择表现最佳的模型.

### 1.4 时序模型优化流程

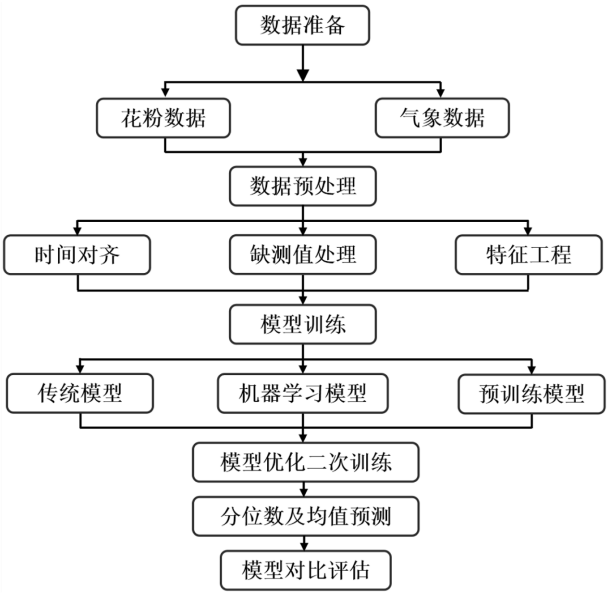



图 3 数据处理和时间序列预测模型技术路线

Fig.3 Technical roadmap for data processing and time series prediction mode

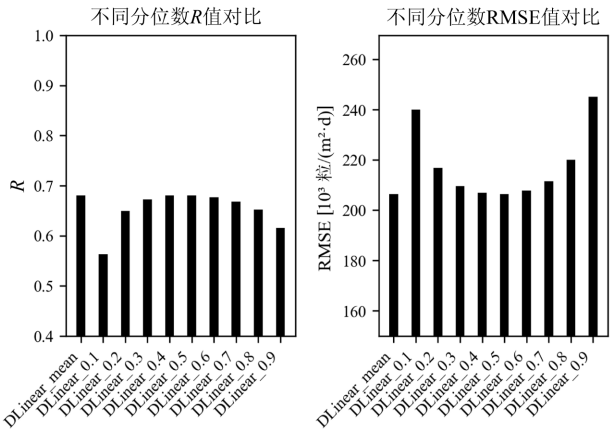



图 4 Dlinear 模型均值预测和不同分位数预测的 $R$ 和 RMSE 值对比

Fig.4 Comparison of $R$ and RMSE values for mean prediction and different quantile predictions of the Dlinear model

使用了 AutoGluon 中全部的时序预测学习器,在实际检验中,发现 AG-TS 的部分学习器表现较差或无法有效拟合数据:例如 DeepAR、PatchTST、Average 和 Croston_2 模型未能捕捉到数据中的日变化趋势,预测结果仅呈现两个低幅峰值; Weighted

Ensemble、Naive 和 Seasonal Naive 模型的预测值与观测值完全一致或滞后,缺乏研究价值;而 Zero 模型则输出全零预测,完全脱离实际数据分布.这些问题的原因可能在于模型本身的算法局限性或假设与数据动态特征不匹配,导致无法有效捕捉实际时序模式.针对这些问题,本文首先剔除了这些明显拟合较差的模型,随后基于剩余有效模型进行了针对性的二次训练.在模型训练完成后,代码进一步处理了测试数据的准备和预测过程.预测部分通过循环遍历测试日期,逐步生成未来 3d 的花粉浓度预测结果,并将不同模型的预测结果合并保存到 CSV 文件中,包含分位数预测(0.1~0.9)及均值预测.以 DLinear 模型为例(图 4),通过对均值预测及不同分位数预测的 $R$ 和 RMSE 进行对比分析,发现均值预测在整体精度和稳定性上表现较好.具体而言,均值预测的 $R$ 值相对高于分位数预测,且 RMSE 值相对较低,表明其能够更准确地捕捉花粉浓度的整体趋势和波动特征.此外,代码还处理了已知协变量的输入,确保模型在预测时可以利用当前的气象信息,从而提升预测精度.根据 $R$ 值大小,研究从 29 个时间序列预测模型中筛选出 $R$ 值大于 0.5 的 10 个模型进行对比分析(图 2).这些模型的对比分析不仅揭示了不同方法在花粉浓度预测中的性能差异,还为模型选择与优化提供了科学依据.

**表 2　筛选后的时间序列预测模型及原理**

Table 2　Selected time series prediction models and their principles

| 模型名称 | 原理描述 | 适用场景 |
|---|---|---|
| Theta[39] | 通过分解线性趋势与残差项进行双指数平滑,适合具有明显趋势但季节性较弱的数据. | 趋势明显、季节性较弱的时间序列数据. |
| Auto ARIM[40] | 基于自回归积分滑动平均模型(ARIMA),通过自动化参数搜索优化传统 ARIMA 的建模流程. | 平稳或可通过差分平稳的时间序列数据,适合自动化建模需求. |
| Chronos Zero Shot(CZS)[41] | 一种预训练模型,通过迁移学习实现跨领域泛化,适合低数据场景. | 数据量较少或需要快速部署的场景. |
| Chronos Fine Tuned(CFT)[42] | 在 Chronos 基础上通过微调适配特定任务,提升预测性能. | 需要结合预训练模型与领域特定数据的场景. |
| Temporal Fusion Transformer(TFT)[43] | 通过特征选择机制动态加权历史信息与外部变量,适合处理多源异构数据. | 多变量输入、复杂非线性关系的时间序列数据. |
| TiDE[44] | 一种基于多层感知机(MLP)的时间序列预测模型,通过结合线性和非线性特征,处理协变量和复杂的依赖关系,提供高效的长时间序列预测. | 具有明显趋势和季节性的时间序列数据. |
| Wave Net[45] | 利用扩张卷积捕捉长期依赖性,适合处理高分辨率时序数据. | 高分辨率、长期依赖性较强的时间序列数据. |
| ARIMA[46] | 经典的自回归积分滑动平均模型,通过线性组合历史值和误差项进行预测. | 平稳或可通过差分平稳的时间序列数据,适合线性关系建模. |
| Simple Feed Forward(SFF)[47] | 一种简单的全连接神经网络模型,通过多层感知机拟合非线性关系. | 非线性关系较强但数据量适中的时间序列数据. |
| DLinear[48] | 通过趋势-季节分解与独立线性层映射,在保持高效计算的同时捕捉长期依赖关系. | 趋势和季节性明显且需要高效计算的时间序列数据. |

## 2　结果与讨论

### 2.1　气象和花粉浓度数据分析

在进行时间序列预测模型预测之前,对气象数据及花粉浓度数据进行了统计分析,以揭示影响花粉浓度变化的关键因素及其时空分布特征.图 5 展示了北京市顺义区 2018~2024 年花粉浓度的频率分布特征,基于 90%分位数分为低浓度组和高浓度组.低浓度组中,花粉浓度集中分布于 0~$50\times10^3$ 粒/($m^2$·d),频率随浓度升高呈显著右偏态衰减特征;而高浓度组则呈现广域分布,花粉浓度集中分布在 $250\times10^3$~$750\times10^3$ 粒/($m^2$·d),其频率亦随浓度递增呈单调递减趋势.值得注意的是,高浓度组在部分天数出现显著跃升现象,表明该区域存在突发性花粉暴露事件,可能对敏感人群产生急性健康风险.整体数据表明,顺义区花粉浓度以低背景水平为主导.

根据相关性分析结果(图 6),气压与草面温度、地面温度、气温及水汽压呈现强负相关($r$<−0.7),可能是地表受热升温导致空气膨胀上升,近地面气压降低,同时温度升高促进蒸发作用,水汽压随之上升,形成

稳定的负反馈循环;气温与草面温度、地面温度、水汽压存在强正相关($r$>0.8),反映太阳辐射是地表能量收支的核心驱动力,地表吸收短波辐射后通过长波辐射和感热通量加热近地空气,同时促进土壤层间热传导,这一过程主导了局地热环境的时空差异;相对湿度与草面温度、降水量及水汽压呈中度正相关($r$>0.3),而与风速负相关($r$<−0.3),可能是因为温度升高导致蒸发,从而增加大气水汽含量,但若未伴随降水或平流补充,可能导致相对湿度下降,另外强风加速近地面空气混合,促进蒸发散热,从而抑制湿度累积;花粉浓度与地面温度、草面温度呈弱正相关($r$>0.1),与降水量、相对湿度、气压呈微弱负相关,这表明适度升温加速植物生长,促进花粉释放,降水直接冲刷大气花粉颗粒,高湿度环境增加花粉吸湿沉降,低压系统常伴随上升气流,可能增强花粉垂直扩散.

从图 7 可以看出,北京的花粉季节主要集中在每年的 3~10 月,这与植物生长周期和气候条件密切相关.3~5 月是春季花粉高峰期,尤其是 4 月,花粉浓度达到全年峰值,主要来源于桦树、松树等早春开花植物,这些植物的花粉释放量较大,对过敏症患者的影响尤为显著[49-50].夏季(6~7 月)花粉浓度相对较低,但仍有一定波动.8~10 月是秋季花粉高峰期,主要花粉类型来自禾本科植物、艾蒿和莎草等秋季开花植物,这一时期的花粉浓度虽不及春季峰值,但仍对过敏人群构成显著威胁[51].此外,花粉浓度的年际变化和季节性分布还受到环境因素和人为活动的共同影响.例如,城市化进程中的绿地规划可能改变局部植被类型,进而影响花粉的种类和浓度[52];空气污染(如 $PM_{2.5}$)可能通过与花粉颗粒的相互作用加剧过敏反应[53-54];气候变化可能导致花粉季节延长和花粉生成量增加[55-56].这些因素的复杂交互作用使得花粉浓度的预测更具挑战性,但也为多变量时间序列模型的构建提供了丰富的数据基础.

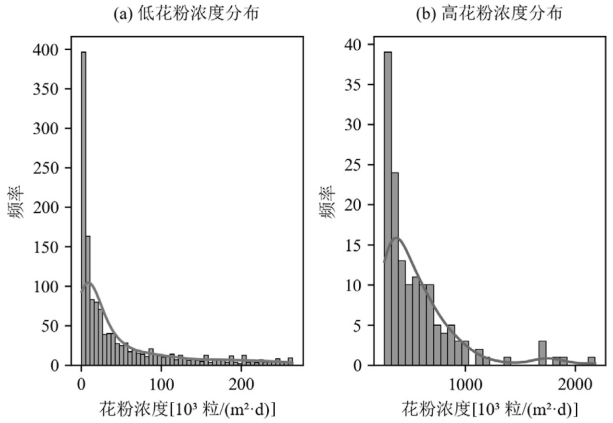
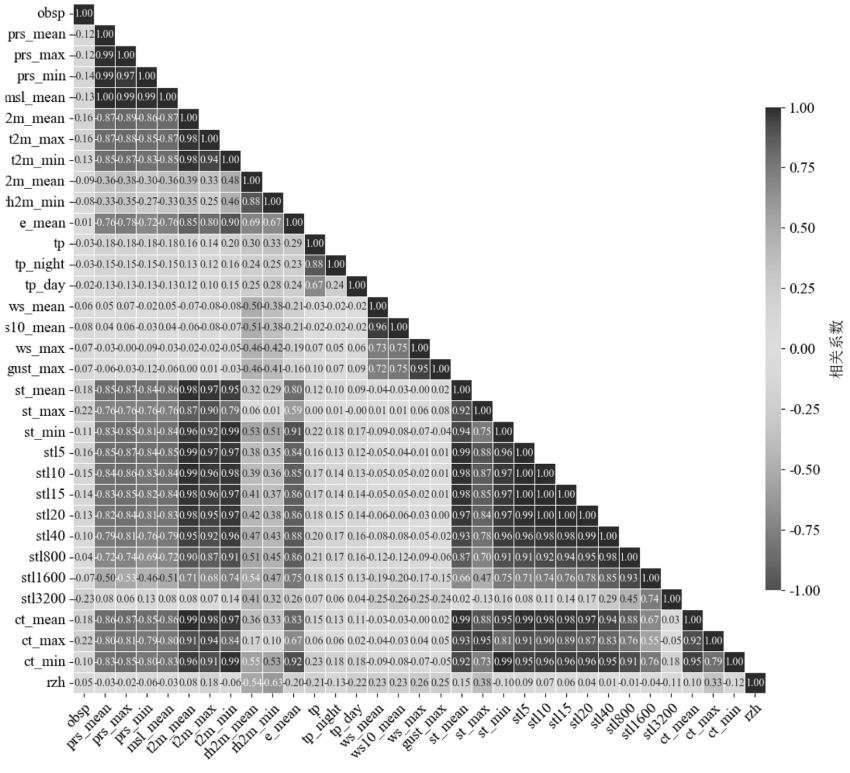



图 5 2018~2024 年期间北京市顺义区的花粉浓度频率

Fig.5 Pollen Concentration Frequency in Shunyi District, Beijing from 2018 to 2024



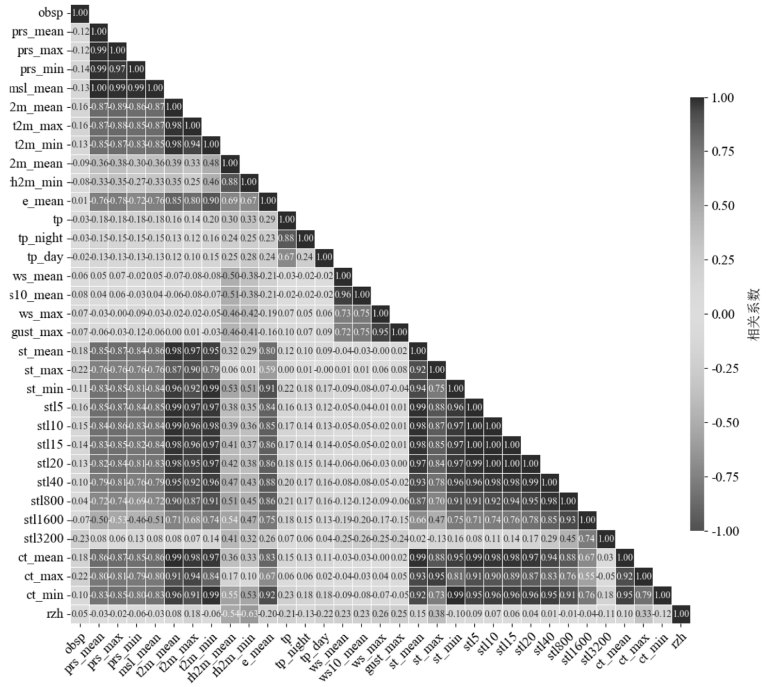



图 6　花粉浓度与气象要素的相关系数

Fig.6　Correlation coefficients between pollen concentration and meteorological factors

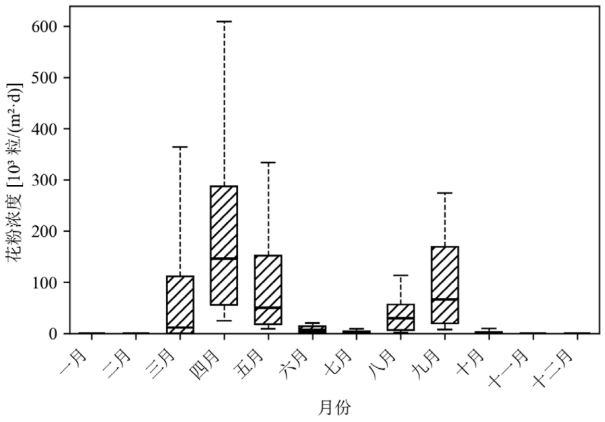



图 7　2018~2024 年期间北京市顺义区的月花粉浓度分布

Fig.7　Distribution of monthly pollen concentration in Shunyi District, Beijing from 2018 to 2024

## 2.2　时间序列预测模型对比

对比了不同时间序列模型预测花粉浓度的表现(图 8),发现多数模型在 1d 预测期内能大致跟踪浓度变化趋势,比如 DLinear 模型($R$=0.68, RMSE=206.5×10$^3$ 粒/(m$^2$·d))整体效果较好,说明它更适合处理花粉数据的线性变化规律.其他模型存在明显问题:其中 TFT 模型(图 8(f))整体预测值偏低,可能是模型对数据波动“学习不足”,导致预测时过于保守;Theta 模型(图 8(g))在春季出现系统性高估,尤其当实际浓度较低时预测值显著偏高,这可能与其基于历史均值的趋势外推机制有关,无法适应春季物候突变的低浓度波动;WaveNet 模型(图 8(h))和 ARIMA 模型(图 8(j))的预测问题在于前者春季出现过多峰值,后者 3 月份峰值过于高估,可能是 WaveNet 模型的扩张卷积结构对局部波动敏感,ARIMA 模型的线性假设与非线性花粉动态冲突.

为了对花粉过敏患者提供有价值的早期预警,本文集中研究了时间序列预测模型在不同时间跨度内的表现,重点分析了 1~3d 的预测结果.通过对多种模型预测结果的分析,我们明确了预测时间对模拟结果稳定性的影响.由图 9 可见,各模型在提前 1d 和提前 3d 的预测中,花粉浓度的拟合线形状较为一致.然而,这种一致性也凸显了多次预测与提前期为 1d 时的共同问题.值得注意的是,多数模型在多次时间推移预测中表现出显著差异,这反映了预测的时间滞后性,以及预测性能相应的降低和改进.通过图 10 的分析更说明,各个模型在预测第 1d 的 $R$ 值和 RMSE 值均显著优于第 2~3d 的预测结果,表明短期预测 1d 的精度和稳定性明显高于第 2~3d.然而,随着预测时间延长至第 2~3d,$R$ 值逐渐下降,RMSE 值则显著上升,表明模型的预测性能随时间的推移而衰减.总体而言,预测 1d 的效果较好,尽管存在时间滞后性,短期预测依然是相对稳定和有效的.

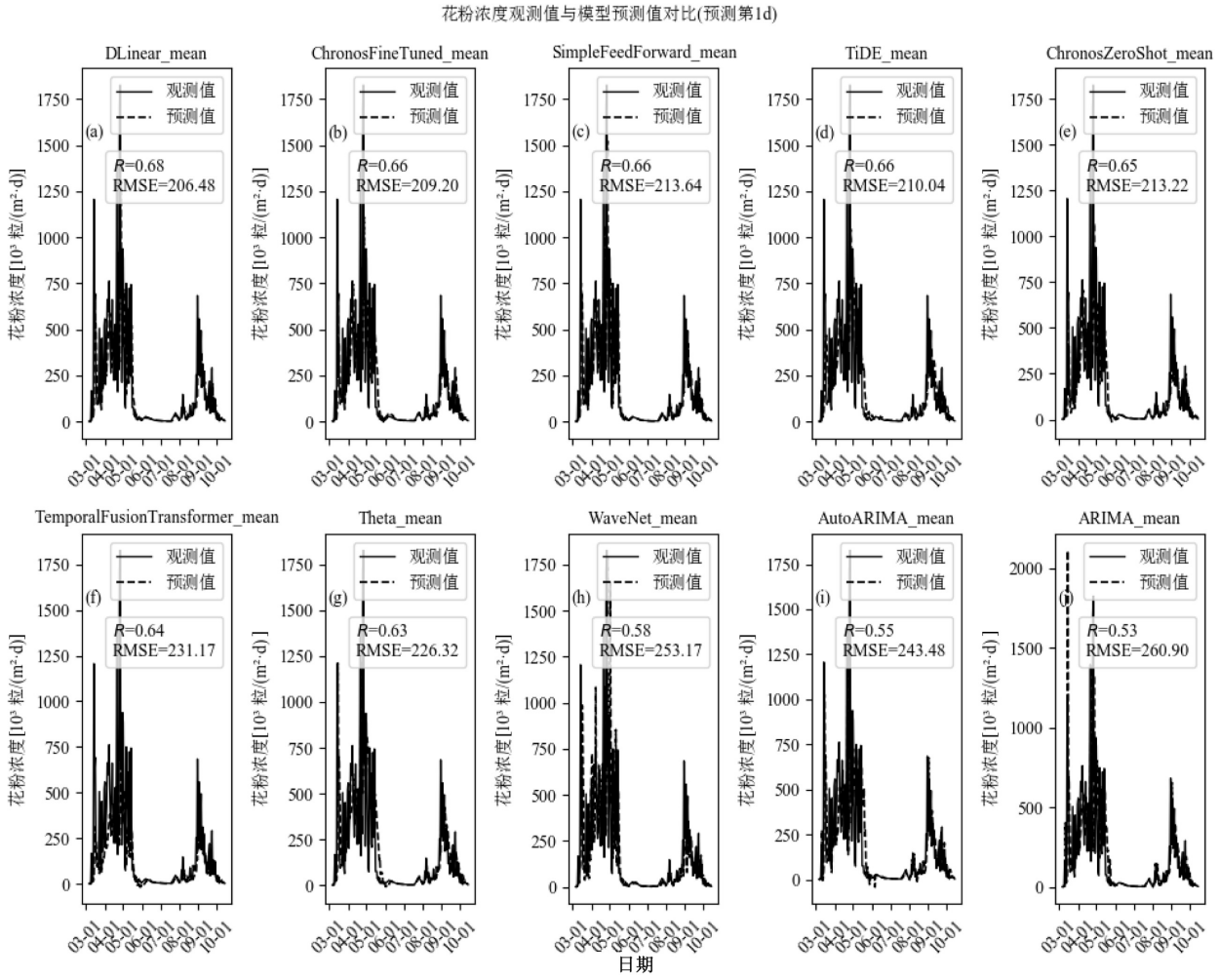



图 8 10 种时间序列预测模型第 1d 的花粉浓度预测与观测对比

Fig.8 Comparison of pollen concentration prediction and observation on the first day of 10time series prediction models

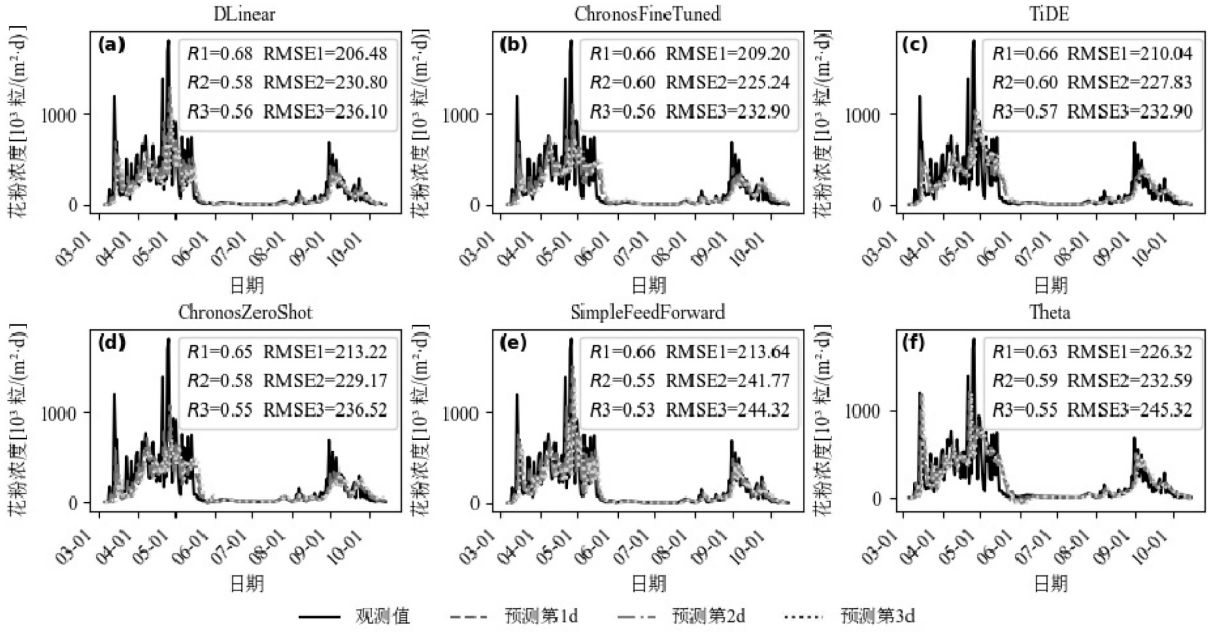



图 9 6 种时间序列预测模型预测 1~3d 的花粉浓度与观测对比

Fig.9 Comparison of 6time series prediction models predicting pollen concentration for 1~3 days with observations

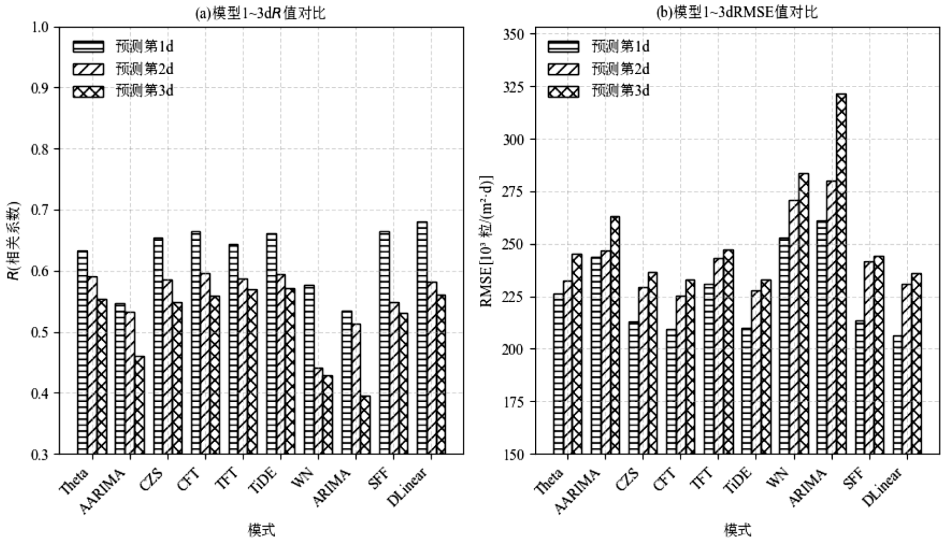



图 10　10 种时间序列预测模型预测 1~3d 的 *R* 和 RMSE 对比

Fig.10　Comparison of *R* and RMSE predictions for 1~3 days using 10 time series prediction model

## 2.3　讨论

本研究基于北京市顺义区连续 7a(2018~2024 年)的花粉浓度观测数据与气象数据,系统构建并优化了多变量时间序列预测模型,旨在提升区域花粉浓度预警的精度与可靠性.通过整合传统统计模型、机器学习模型及深度学习模型的优势,结合 AG-TS 框架的自动化集成能力,本研究揭示了时间序列预测在花粉浓度动态建模中的核心价值,并深入探讨了气象-物候交互作用对花粉时空分布的影响机制.

花粉过敏已成为全球性公共卫生问题,其发病率与城市化进程、气候变化及植被结构变化密切相关[57-59].通过分析花粉浓度以及气象数据揭示了花粉浓度的分布特征,低浓度占主导,而高浓度虽频率较低,但对过敏人群影响显著[60].这一分布与北京植被物候特征密切相关,春季(3~5 月)的桦树、松树花粉与秋季(8~10 月)的禾本科、艾蒿花粉分别形成年度双峰[4].相关性分析进一步表明,地表温度与气温是驱动花粉浓度的核心气象因子,其正相关性源于热量积累对植物花期与花粉成熟的促进作用[5].然而,降水量与相对湿度的负相关揭示了气象因子的双重作用:短期降雨虽可冲刷花粉,但长期高湿度可能抑制释放[61].值得注意的是,草面温度的正相关暗示地下热力条件通过调控植物根系活动间接影响花粉生成,这一发现拓展了传统气象驱动的单一视角[62].此外,城市化导致的绿地类型变化(如人工植被替代原生植物)可能改变花粉种类组成[6],而 $PM_{2.5}$ 与花粉的协同效应[63]进一步复杂化了健康风险评估.这些发现为多变量时序模型的构建提供了关键依据,例如通过引入 NDVI 卫星数据[64],可增强模型对空间异质性的捕捉能力.

与以往研究相比,本研究在花粉浓度时间序列预测方法上展现出多维度创新优势.首先,通过 AG-TS 框架,实现了传统统计模型(如 ARIMA、Theta)、深度学习模型(如 TFT、WaveNet)及预训练模型(如 Chronos)的自动化集成与优化,突破了传统单一模型依赖的局限性,有效结合不同算法的优势,显著提升了模型泛化性与预测精度.其次,研究系统性整合了 36 种气象要素及衍生特征作为协变量输入,充分量化气象条件对花粉浓度的动态影响,克服了传统方法因忽略多源数据交互而导致的预测偏差.此外,基于 *R* 值和 RMSE 指标,通过剔除低效模型,如 DeepAR、Zero 模型等和二次训练优化模型性能,从 29 个候选模型中筛选出 10 个高精度模型($R$>0.5),并解决了传统研究中模型冗余与过拟合问题.最后,结合均值预测与分位数预测(0.1~0.9),既保障了整体趋势的准确性,又量化了预测不确定性,为公共卫生预警提供了更全面的决策支持.

时间序列预测模型在本研究中展现出对花粉浓度动态变化的高效捕捉能力,尤其是在短期预测(1d)中表现突出.DLinear 模型以 $R$=0.68 和 RMSE=206.48 的综合性能在捕捉线性趋势与季节性特征

方面表现最佳,其通过趋势−季节分解与独立线性层映射,在保持计算效率的同时有效建模长期依赖性[65-66].相比之下,TFT 模型虽在多变量输入和非线性关系建模中具有优势[67],但预测值普遍偏低,表明其对数据波动的学习能力不足;Theta 模型因依赖历史均值外推[39],可能在春季低浓度时出现系统性高估;WaveNet 模型的扩张卷积结构虽能捕捉长期依赖,但对局部噪声敏感[68],可能导致春季预测峰值过多.

此外,模型性能随预测时间延长显著衰减:1d 预测的 $R$ 值普遍高于 2~3d 预测,例如 DLinear 模型在 3d 预测中 $R$ 值下降 18%,RMSE 值上升 15%.这一现象表明,模型性能随预测时间延长而逐渐衰减存在多种原因.气象条件是影响花粉扩散的关键因素之一,随着预测时间范围的扩大,气象要素的不确定性会逐渐累积.短期预测可以依赖相对准确的气象数据,但中长期预测中,气象模型的误差会不断叠加,导致花粉浓度预测的偏差增大.例如,风向的微小变化可能在几天后显著改变花粉的传输路径[69],而模型难以完全捕捉这种非线性效应.此外,花粉扩散本身涉及复杂的物理过程,包括颗粒的悬浮、湍流混合、干湿沉降等[70].花粉的粒径和密度等特性使其在大气中的滞留时间和传播距离存在较大变数[71],时间跨度越长,这些因素的交互作用越复杂.短期预测中,局部源主导的花粉浓度可能相对稳定,而中长期预测需考虑跨区域传输和多次沉降−再悬浮过程,进一步增加了建模难度.从模型角度看,现有方法在捕捉长期依赖关系和复杂非线性动态时存在局限性.虽然线性模型能较好地拟合局部趋势,但对花粉浓度随气象条件变化的滞后效应或突发天气事件的响应能力不足.同时,季节性变化和年际波动可能使模型在长期预测中难以平衡历史规律与实时动态.这些因素共同导致预测误差随时间延长而放大,表现为 $R$ 值下降和 RMSE 上升.未来研究可通过融合更精细的气象耦合机制、引入花粉源动态变化模块或改进模型对长期时空关联的建模能力来缓解这一问题.

这些问题反映出三个共性原因:一是模型对季节转换期的气象突变响应不足[72];二是训练数据中春秋季的高浓度样本较少,模型缺乏对极端值的预测能力;三是花粉释放与植物生长阶段、天气因素存在复杂关联[73],现有模型可能只捕捉了部分简单规律,却忽略了更复杂的生物−环境交互作用[74].因此,短期预测(1d)因其较高的稳定性和实用性,可作为当前花粉预警系统的核心输出,而中长期预测需通过引入动态气象预报数据或耦合物理模型以提升可靠性.

在国际上也有类似的研究,Cordero 等[75]采用了 LightGBM(轻量级梯度提升机)和 ANN(人工神经网络)两种机器学习方法对西班牙马德里社区 10 个站点的花粉季峰值日期以及每日花粉浓度进行了预测,数据集覆盖了从 1993~2018 年.与本研究相比,Cordero 等研究的时空范围要更广,从结果上来看,其峰值预测与每日浓度预测的效果要优于本文,$R^2$ 值可以达到 0.6.Csépe 等[76]使用 CL(计算智能)方法对法国里昂和匈牙利塞格德两个城市进行了每日花粉浓度的预测,该方法包括多层感知器(MLP)、树算法、MLPRegressor,数据集覆盖了从 1997~2006 年.与本研究相比,该研究对每日花粉浓度预测了 1~7d,并且效果要优于本研究,预测 1d 的 $R$ 值可以达到 0.9,同时也出现了模式的预测随时间衰减的现象.但本研究的优势在于使用 AG−TS 框架集成 29 种模型,显著提升泛化能力,引入 36 种气象要素及其衍生变量,以及对 1~3d 的预测进行了比较分析,通过剔除低效模型和二次训练优化流程,为花粉预测提供了可复用的技术路线.另外本研究集成的模型中不包含上述等人使用的模型,基于他们使用模型的预测效果,在后续的研究中可以加入这些模型以提升模型的准确性.

相较于传统统计方法,多变量时间序列模型通过整合气象要素、物候特征及历史浓度数据的动态耦合关系,能够实现未来 72h 的高分辨率预测.未来可以整合卫星遥感数据、社交媒体健康报告及医院就诊记录,提升数据的时空分辨率与覆盖范围;结合数值模式(如 WRF−Chem)的物理机制与机器学习的非线性拟合能力,开发混合模型,以增强对极端事件的预测鲁棒性[77-79];基于实时数据流与边缘计算技术[80],构建自适应更新模型[81],实现分钟级花粉浓度预警;通过迁移学习或联邦机器学习技术[82-84],将北京市模型适配至其他气候区,验证模型的普适性与可迁移性.此外,加强跨学科合作将是推动花粉预警系统从研究向实际应用转化的关键.

## 3　结语

本研究通过综合应用传统统计模型、机器学习模型及深度学习模型,并结合 AG-TS 框架的自动化集成能力,对花粉浓度进行了预测.研究表明,北京花粉季节呈双峰模式,春季(3~5 月)和秋季(8~10 月)为高峰期,峰值出现在 4 月,与北京市主要致敏植物的花期高度吻合.花粉浓度呈右偏分布,低浓度(0~50×$10^3$ 粒/($m^2$·d))占主导,高浓度(250×$10^3$~750×$10^3$ 粒/($m^2$·d))频率低但可引发急性健康风险.气象因素对花粉浓度的时空分布起到了重要的调控作用,尤其是温度、降水量、风速等气象要素对花粉浓度的影响较为显著.花粉浓度与地面温度、草面温度呈弱正相关,与降水量、相对湿度和气压呈微弱负相关.在对比的时间序列预测模型中,DLinear 模型在 1d 预测期内表现最优($R$=0.68,RMSE=206.5×$10^3$ 粒/($m^2$·d)),但模型预测性能随预测时间延长而下降,第 1d 预测精度显著高于第 2~3d.本研究通过 AG-TS 框架整合多种模型和大量气象协变量,显著提高了花粉浓度预测的精度和可靠性,为花粉过敏预警提供了技术支持,这对于改善公共卫生状况、降低花粉过敏对人群健康的影响具有重要意义.

**参考文献：**


[1] Lake I R, Jones N R, Agnew M, et al. Climate change and future pollen allergy in Europe [J]. Environmental Health Perspectives, 2017, 125(3):385-391.

[2] Liu Y, Lu C, Li Y, et al. Outdoor air pollution and indoor window condensation associated with childhood symptoms of allergic rhinitis to pollen [J]. International Journal of Environmental Research and Public Health, 2022,19(13):8071.

[3] Öçal R, Bayar Muluk N, Mullol J. Epidemiology of allergic rhinitis [J]. All Around the Nose: Basic Science, Diseases and Surgical Management, 2020:297-301.

[4] Wang X Y, Ma T T, Wang X Y, et al. Prevalence of pollen-induced allergic rhinitis with high pollen exposure in grasslands of northern China [J]. Allergy, 2018,73(6):1232-1243.

[5] Zhao Y, Sun Z, Xiang L, et al. Effects of pollen concentration on allergic rhinitis in children: A retrospective study from beijing, a chinese megacity [J]. Environmental Research, 2023,229:115903.

[6] Zong H, Yin R, Wu X, et al. Comparison of two methods for assessing the pollen allergy risk of urban parks in chengdu city, china [J]. Urban Forestry & Urban Greening, 2023,86:128014.

[7] Bürgler A, Glick S, Hartmann K, et al. Rationale and design of a panel study investigating six health effects of airborne pollen: The EPOCHAL study [J]. Frontiers in Public Health, 2021,9:689248.

[8] Chico-Fernández J, Ayuga-Téllez E. Study of the interrelations of the concentrations of six tree pollen types and six atmospheric pollutants with rhinitis and allergic conjunctivitis in the community of Madrid [J]. Applied Sciences, 2024,14(7):2965.

[9] Zhou Y, Dai J, Liu H, et al. Tourist risk assessment of pollen allergy in tourism attractions: A case study in the summer palace, beijing, china [J]. Frontiers in Public Health, 2022,10:1030066.

[10] Janati A, Bouziane H, del Mar Trigo M, et al. Poaceae pollen in the atmosphere of tetouan (NW morocco): Effect of meteorological parameters and forecast of daily pollen concentration [J]. Aerobiologia, 2017,33:517-528.

[11] Forkuor G, Hounkpatin O K, Welp G, et al. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: A comparison of machine learning and multiple linear regression models [J]. PLOS One, 2017,12(1):e0170478.

[12] Schaffer A L, Dobbins T A, Pearson S A. Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: A guide for evaluating large-scale health interventions [J]. BMC Medical Research Methodology, 2021,21:1-12.

[13] Wang Q, Nguyen T T, Huang J Z, et al. An efficient random forests algorithm for high dimensional data classification [J]. Advances in Data Analysis and Classification, 2018,12:953-972.

[14] Ruan W, Li Z, Sun Z, et al. Enhancing pollen prediction in beijing, a chinese megacity: Leveraging ensemble learning models for greater accuracy [J]. Aerosol and Air Quality Research, 2024,24(11):240123.

[15] Shokouhi B V, de Hoogh K, Gehrig R, et al. Spatiotemporal modelling of airborne birch and grass pollen concentration across switzerland: A comparison of statistical, machine learning and ensemble methods [J]. Environmental Research, 2024,263:119999.

[16] Picornell A, Hurtado S, Antequera-Gómez M L, et al. A deep learning LSTM-based approach for forecasting annual pollen curves: Olea and Urticaceae pollen types as a case study [J]. Computers in Biology and Medicine, 2024,168:107706.

[17] Liu L, Solmon F, Vautard R, et al. Estimates of common ragweed pollen emission and dispersion over europe using RegCM-pollen model [J]. Biogeosciences Discussions, 2015,12(21).

[18] Tagliaferro S, Adani M, Pepe N, et al. The impact of the spatial resolution of vegetation cover on the prediction of airborne pollen concentrations over northern Italy [J]. Agricultural and Forest Meteorology, 2024,355:110153.

[19] Werner M, Guzikowski J, Kryza M, et al. Extension of WRF-chem for birch pollen modelling—a case study for Poland [J]. International Journal of Biometeorology, 2021,65(4):513-526.

[20] Zhu X, Ma X, Zhang Z, et al. Floating in the air: forecasting allergenic pollen concentration for managing urban public health [J]. International Journal of Digital Earth, 2024,17(1):2306894.

[21] Zhang J, Zheng Y, Qi D, et al. DNN-based prediction model for spatio-temporal data [C]//Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2016:1-4.

[22] Liu Y, Gong C, Yang L, et al. DSTP-RNN: a dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction [J]. Expert Systems with

Applications, 2020,143:113082.

[23] Wang H W, Li X B, Wang D, et al. Regional prediction of ground-level ozone using a hybrid sequence-to-sequence deep learning approach [J]. Journal of Cleaner Production, 2020,253:119841.

[24] Baklanov A, Smith Korsholm U, Nuterman R, et al. Enviro-HIRLAM online integrated meteorology–chemistry modelling system: Strategy, methodology, developments and applications (v7. 2) [J]. Geoscientific Model Development, 2017,10(8):2971-2999.

[25] Pacher B, Lalic B, Eitzinger J, et al. Agricultural meteorology and climatology [M]. Firenze: Firenze University Press, 2018.

[26] Neeson F. Saclay: Pollen and fungal spore monitoring [D]. Dublin City University, 2024.

[27] Parmezan A R S, Souza V M, Batista G E. Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model [J]. Information Sciences, 2019,484:302-337.

[28] Le Guen V, Thome N. Shape and time distortion loss for training deep time series forecasting models [J]. Advances in Neural Information Processing Systems, 2019,32.

[29] Mirzavand M, Ghazavi R. A stochastic modelling technique for groundwater level forecasting in an arid environment using time series methods [J]. Water Resources Management, 2015,29:1315-1328.

[30] Lim B, Zohren S. Time-series forecasting with deep learning: A survey [J]. Philosophical Transactions of the Royal Society A, 2021, 379(2194):20200209.

[31] Gessman D J. Pollen forecasting in sarasota, florida [M]. Florida: University of South Florida, 2017.

[32] Yin Q, Sun C, Cao M, et al. Sliding window GBDT for electricity demand forecasting [C]//International Conference on Intelligent Computing. Springer, 2023:817-827.

[33] Kontopoulou V I, Panagopoulos A D, Kakkos I, et al. A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks [J]. Future Internet, 2023,15(8):255.

[34] Picornell A, Hurtado S, Antequera-Gómez M L, et al. A deep learning LSTM-based approach for forecasting annual pollen curves: Olea and urticaceae pollen types as a case study [J]. Computers in Biology and Medicine, 2024,168:107706.

[35] Erickson N, Mueller J, Shirkov A, et al. Autogluon-tabular: Robust and accurate automl for structured data [J]. Arxiv Preprint Arxiv, 2003,06505,2020.

[36] Shi X, Mueller J, Erickson N, et al. Multimodal automl on structured tables with text fields [C]//8th ICML workshop on automated machine learning (AutoML). 2021.

[37] Westergaard G, Erden U, Mateo O A, et al. Time series forecasting utilizing automated machine learning (AutoML): A comparative analysis study on diverse datasets [J]. Information-an International Interdisciplinary Journal, 2024,15(1):39.

[38] Shchur O, Turkmen A C, Erickson N, et al. AutoGluon–TimeSeries: AutoML for probabilistic time series forecasting [C]//International Conference on Automated Machine Learning. PMLR, 2023:9-1.

[39] Spiliotis E, Assimakopoulos V, Nikolopoulos K. Forecasting with a hybrid method utilizing data smoothing, a variation of the theta method and shrinkage of seasonal factors [J]. International Journal of Production Economics, 2019,209:92-102.

[40] Ospina R, Gondim J A, Leiva V, et al. An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: Methodology and case study in Brazil [J]. Mathematics, 2023,11(14): 3069.

[41] Chang M, Fokoue A, Uceda-Sosa R, et al. CHRONOS: A schema-based event understanding and prediction system: 21 [C]//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 38. 2024:22871-22877.

[42] Arango S P, Mercado P, Kapoor S, et al. ChronosX: Adapting pretrained time series models with exogenous variables [J]. Arxiv Preprint Arxiv: 2503.12107,2025.

[43] Wang S, Lin Y, Jia Y, et al. Unveiling the multi-dimensional spatio-temporal fusion transformer (MDSTFT): A revolutionary deep learning framework for enhanced multi-variate time series forecasting [J]. IEEE Access, 2024,12:115895-115904.

[44] Das A, Kong W, Leach A, et al. Long-term forecasting with tide: Time-series dense encoder [J]. Arxiv Preprint Arxiv: 2304.08424, 2023.

[45] Chen Y, Wu Z, Shen Z, et al. Long-time prediction of multi-indicator water quality based on the improved WaveNet under time series decomposition and phase space reconstruction [J]. Measurement Science and Technology, 2024,36(1):15802.

[46] Shumway R H, Stoffer D S, Shumway R H, et al. ARIMA models [J]. Time Series Analysis and Its Applications: With R Examples, 2017:75-163.

[47] González-Arjona D, López-Pérez G, González A G. Non-linear QSAR modeling by using multilayer perceptron feedforward neural networks trained by back-propagation [J]. Τάλαντα. Proceedings of the Dutch Archaeological and Historical Society, 2002,56(1):79-90.

[48] Chaoqun Su. DLinear makes efficient long-term predictions [C]//Proceedings of ACM Conference (Baidu KDD CUP 2022). Baidu KDD CUP 2022:4.

[49] Qi C, Jiang J, Ye C, et al. Prediction of annual peak concentration dates for major airborne allergenic pollens in Beijing [J]. Acta Ecologica Sinica, 2023,43(7):2650-2662.

[50] 张 婧,蒋雨婷,高 露,等.2022~2023 年太原市春季气传花粉分布情况与过敏性鼻炎患者 sIgE 检测阳性的相关性研究 [J]. 中华预防医学杂志, 2024,58(6):823-829.
Zhang J, Jiang Y, Gao L, et al. Correlation between the distribution of airborne pollen in spring 2022~2023 in Taiyuan and positivity of sIgE tests in patients with allergic rhinitis[J]. Chinese Journal of Preventive Medicine, 2024,58(6):823–829.

[51] Sun A, Zhang H, Li X, et al. Study on the characteristics of airborne pollen types and concentration variations in northern Beijing [J]. Scientia Geographica Sinica, 2023,43(4):737-744.

[52] Calatayud V, Cariñanos P. Mapping pollen allergenicity from urban trees in valencia: A tool for green infrastructure planning [J]. Environmental Research, 2024,252:118823.

[53] Candeias J, Zimmermann E J, Bisig C, et al. The priming effect of diesel exhaust on native pollen exposure at the air-liquid interface [J]. Environmental Research, 2022,211:112968.

[54] Cariñanos P, Adinolfi C, Diaz de la Guardia C, et al. Characterization

of allergen emission sources in urban areas [J]. Journal of Environmental Quality, 2016,45(1):244−252.

[55] Barnes C S. Impact of climate change on pollen and respiratory disease [J]. Current Allergy and Asthma Reports, 2018,18:1−11.

[56] Mousavi F, Oteros J, Shahali Y, et al. Impacts of climate change on allergenic pollen production: A systematic review and meta−analysis [J]. Agricultural and Forest Meteorology, 2024,349:109948.

[57] d'Amato G, Chong−Neto H J, Monge Ortega O P, et al. The effects of climate change on respiratory allergy and asthma induced by pollen and mold allergens [J]. Allergy, 2020,75(9):2219−2228.

[58] Demain J G. Climate change and the impact on respiratory and allergic disease: 2018 [J]. Current Allergy and Asthma Reports, 2018,18:1−5.

[59] Ravindra K, Goyal A, Mor S. Pollen allergy: Developing multi−sectorial strategies for its prevention and control in lower and middle−income countries [J]. International Journal of Hygiene and Environmental Health, 2022,242:113951.

[60] Sun Z, Yuxin Z, Xingqin A, et al. Effects of airborne pollen on allergic rhinitis and asthma across different age groups in Beijing, China [J]. Science of the Total Environment, 2024,912:169215.

[61] Lang Yongyao, Yang Yun, Liu Qing, et al. An analysis of meteorological and environmental factors linked to thunderstorm asthma in China [J]. Journal of Asthma, 2024:1−8.

[62] Li J, An X, Sun Z, et al. Construction and application of a pollen emissions model based on phenology and random forests [J]. Egusphere, 2024,2024:1−41.

[63] Capone P, Lancia A, D'Ovidio M C. Interaction between air pollutants and pollen grains: Effects on public and occupational health [J]. Atmosphere, 2023,14(10):1544.

[64] Zhong J, Xiao R, Wang P, et al. Identifying influence factors and thresholds of the next day's pollen concentration in different seasons using interpretable machine learning [J]. Science of the Total Environment, 2024,935:173430.

[65] Cheng S T, Lyu Y J, Lin Y H. Linear model and gradient feature elimination algorithm based on seasonal decomposition for time series forecasting [J]. Mathematics (2227−7390), 2025,13(5):883.

[66] Khaldi B, Sellami M A, Hamzi O S. Time series forecasting using linear models: A comprehensive investigation and empirical analysis [D]. Ouargla: Kasdi Merbah University, 2024.

[67] Llaquet Vélez S. Implementation of the temporal fusion transformer for demand forecasting [D]. Barcelona: Universitat Politècnica de Catalunya, 2025.

[68] Zhou R, Zhang Y, Wang Q, et al. A hybrid self−adaptive DWT−WaveNet−LSTM deep learning architecture for karst spring forecasting [J]. Journal of Hydrology, 2024,634:131128.

[69] Rojo J, Rapp A, Lara B, et al. Effect of land uses and wind direction on the contribution of local sources to airborne pollen [J]. Science of the Total Environment, 2015,538:672−682.

[70] Mandal M, Popek R, Przybysz A, et al. Breathing fresh air in the city: Implementing avenue trees as a sustainable solution to reduce particulate pollution in urban agglomerations [J]. Plants, 2023,12(7): 1545.

[71] Xu Q, Zhang S, Gaillard M jose, et al. Studies of modern pollen assemblages for pollen dispersal−deposition−preservation process understanding and for pollen−based reconstructions of past vegetation, climate, and human impact: A review based on case studies in China [J]. Quaternary Science Reviews, 2016,149:151−166.

[72] Lops Y, Choi Y, Eslami E, et al. Real−time 7−day forecast of pollen counts using a deep convolutional neural network [J]. Neural Computing and Applications, 2020,32(15):11827−11836.

[73] Oh J W. Pollen and climate [M]//Pollen Allergy in a Changing World: A Guide to Scientific Understanding and Clinical Practice. Springer, 2023:137−154.

[74] Zewdie G K, Lary D J, Levetin E, et al. Applying deep neural networks and ensemble machine learning methods to forecast airborne ambrosia pollen [J]. International Journal of Environmental Research and Public Health, 2019,16(11):1992.

[75] Cordero J M, Rojo J, Gutiérrez−Bustillo A M, et al. Predicting the olea pollen concentration with a machine learning algorithm ensemble [J]. International Journal of Biometeorology, 2021,65(4):541−554.

[76] Csépe Z, Makra L, Voukantsis D, et al. Predicting daily ragweed pollen concentrations using neural networks and tree algorithms over lyon (france) and szeged (hungary) [J]. Acta Climatologica Et Chorologica, 2014,47:17−32.

[77] Han T, Hu X, Zhang J, et al. Rebuilding high−quality near−surface ozone data based on the combination of WRF−chem model with a machine learning method to better estimate its impact on crop yields in the beijing−tianjin−hebei region from 2014 to 2019 [J]. Environmental Pollution, 2023,336:122334.

[78] Lv L, Wei P, Li J, et al. Application of machine learning algorithms to improve numerical simulation prediction of $PM_{2.5}$ and chemical components [J]. Atmospheric Pollution Research, 2021,12(11):101211.

[79] Ma X, Liu H, Peng Z. Improving WRF−chem $PM_{2.5}$ predictions by combining data assimilation and deep−learning−based bias correction [J]. Environment International, 2025,195:109199.

[80] Li Q Y, Others. A novel real−time monitoring, notification, analytics system, and personal thermal sensations model for indoor air quality and energy efficiency in commercial buildings [J]. 2021: xvii, 138.

[81] Wang X, Han Y, Leung V C, et al. Convergence of edge computing and deep learning: A comprehensive survey [J]. IEEE Communications Surveys and Tutorials, 2020,22(2):869−904.

[82] Das B, Adel A, Jan T, et al. Water quality management using federated deep learning in developing southeastern asian country [J]. Water Resources Management, 2024:1−17.

[83] Somu N, Sriram A, Kowli A, et al. A hybrid deep transfer learning strategy for thermal comfort prediction in buildings [J]. Building and Environment, 2021,204:108133.

[84] Wang L, Geng X, Ma X, et al. Cross−city transfer learning for deep spatio−temporal prediction [J]. Arxiv Preprint Arxiv:1802.00386, 2018.

**作者简介：**邓田丰(1998−),男,天津人,硕士,主要从事气象健康方向研究.1102254630@qq.com.