

轨迹数据视角下的分人群 PM_{2.5} 暴露 平等性数据集研发

吴梓昊^{1,2}, 马志锋^{1,2}, 夏吉喆^{1,2*}

1. 深圳大学建筑与城市规划学院、自然资源部大湾区地理环境监测重点实验室、广东省城市空间信息工程重点实验室、深圳市空间信息智能感知与服务重点实验室, 深圳 518060;
2. 亚热带建筑与城市科学全国重点实验室, 深圳 518060

摘要: 作者利用清洗后的大量人群轨迹数据与逐时 1-km 分辨率的 PM_{2.5} 分布数据叠置计算个体的时均 PM_{2.5} 暴露浓度。区别于传统的静态人口分布假设, 该方法能更真实地反映个体在不同时空下的 PM_{2.5} 暴露水平。研究计算了不同性别、年龄、收入及通勤距离群体的时均 PM_{2.5} 暴露浓度, 并在此基础上评估了各城市居民暴露的基尼系数与不同群体的集中指数, 系统识别了各城市 PM_{2.5} 暴露的弱势群体与空间不平等模式, 建立基于轨迹数据的分人群 PM_{2.5} 暴露平等性数据集。该数据集包括研究区 2023 年 1 月 24 日至 2 月 23 日以下数据: (1) 北京、上海、深圳等 6 个城市的逐时 1-km PM_{2.5} 浓度空间分布数据; (2) 基于动态手机轨迹的城市人均 PM_{2.5} 暴露浓度; (3) 各城市不同社会特征 (年龄、收入和性别) 和空间特征 (通勤距离) 群体时均 PM_{2.5} 暴露数据; (4) 各城市居民 PM_{2.5} 暴露基尼系数; (5) 各群体 PM_{2.5} 暴露集中指数 OLS 关联性分析结果; (6) 县级尺度居民时均 PM_{2.5} 暴露量。数据集存储为.shp、.tif 和.xlsx 格式, 由 753 个数据文件组成, 数据量为 10.6 GB (压缩为一个文件, 127 MB)。

关键词: GPS 轨迹数据; PM_{2.5} 浓度反演; PM_{2.5} 暴露评估; 不平等分析

DOI: <https://doi.org/10.3974/geodp.2026.02.08>

CSTR: <https://cstr.escience.org.cn/CSTR:20146.14.2026.02.08>

数据可用性声明:

本文关联实体数据集已在《全球变化数据仓储电子杂志 (中英文)》出版, 可获取:

<https://doi.org/10.3974/geodb.2026.02.08.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2026.02.08.V1>.

1 前言

进入 21 世纪, 伴随着快速的城市化和工业化进程, 中国 PM_{2.5} 污染的问题日益凸显。空气污染, 尤其是 PM_{2.5} 已被证实与各种健康问题有强烈的关系^[1,2]。已有研究^[3]表明长期暴露在较高浓度的 PM_{2.5} 环境中会导致过早死亡。因此, 非常有必要主动防治空气污染。

收稿日期: 2025-12-05; 修订日期: 2026-03-26; 出版日期: 2026-04-25

基金项目: 广西科学技术项目 (GX YD2504850001)

*通讯作者: 夏吉喆, 深圳大学建筑与城市规划学院, xiajizhe@szu.edu.cn

数据引用方式: [1] 吴梓昊, 马志锋, 夏吉喆. 轨迹数据视角下的分人群 PM_{2.5} 暴露平等性数据集研发[J]. 全球变化数据学报, 2026, 10(2): 180–190. <https://doi.org/10.3974/geodp.2026.02.08>. <https://cstr.escience.org.cn/CSTR:20146.14.2026.02.08>.

[2] 马志锋, 吴梓昊, 夏吉喆. 人群轨迹数据视角下中国 6 个城市 PM_{2.5} 暴露平等性 1-km 时均栅格数据集 (2023 年 1–2 月) [J/DB/OL]. 全球变化数据仓储电子杂志, 2026. <https://doi.org/10.3974/geodb.2026.02.08.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2026.02.08.V1>.

为了应对 $\text{PM}_{2.5}$ 污染问题, 中国在 2012 年公布的《环境空气质量标准》(GB3095—2012) 中明确指出将 $\text{PM}_{2.5}$ 浓度纳入污染物扩展标准^[4]。

目前对于 $\text{PM}_{2.5}$ 暴露平等性差异分析大多依赖于人口普查统计数据^[5-7], 其核心假设是个体处于静态居住空间内, 并将区域平均 $\text{PM}_{2.5}$ 浓度作为该区域所有居民的暴露水平。这一方法忽略了个体日常移动行为对空气污染暴露的影响, 难以反映动态、真实的暴露风险分布, 从而在评估环境公平性时产生偏差。 $\text{PM}_{2.5}$ 暴露平等性是指不同社会群体, 如不同收入、年龄、性别、通勤距离等特征群体在 $\text{PM}_{2.5}$ 污染暴露水平上的分布是否公平。若某一群体的平均暴露浓度显著高于其他群体, 则说明存在暴露不平等现象。

本数据集通过将高时空分辨率的 $\text{PM}_{2.5}$ 数据与群体轨迹数据相结合分析, 可以更全面地理解不同社会群体的 $\text{PM}_{2.5}$ 暴露差异以及背后的原因。这种综合分析不仅解释了空气污染对公共健康的影响, 还为制定具有针对性的空气质量管理措施和公共卫生政策提供了科学依据。

2 数据集元数据简介

《人群轨迹数据视角下中国 6 个城市 $\text{PM}_{2.5}$ 暴露平等性 1-km 时均栅格数据集 (2023 年 1-2 月)》^[8] 的名称、作者、地理区域、数据年代、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表 1。

3 数据研发方法

3.1 数据来源

$\text{PM}_{2.5}$ 分布模拟采用的遥感数据如下。 $\text{PM}_{2.5}$ 地面监测数据来源于中国环境监测总站发布的逐小时实时空气质量监测数据¹; 气溶胶光学厚度 (Aerosol Optical Depth, AOD) 数据采用 MODIS Terra 与 Aqua 卫星的 MCD19A2 产品², 空间分辨率为 1 km, 来源于美国国家航空航天局 (NASA) LAADS DAAC 数据平台; 气象数据提取自欧洲中期天气预报中心 (ECMWF) 发布的 ERA5 再分析数据集³, 包括风速、温度、湿度、降水、边界层高度等逐小时变量; 地形数据采用 SRTM GL1 数字高程模型 (DEM)², 空间分辨率为 30 m, 来源于 NASA; 归一化植被指数 (NDVI) 使用 MODIS MOD13Q1 产品, 空间分辨率为 250 m, 来源于 NASA²; 城市建成区数据 (建筑轮廓与路网) 通过百度地图开放平台⁴获取, 用于表征城市结构与人类活动强度。基于上述多源数据, 构建 Stacking 集成学习模型, 实现逐时 1 km $\text{PM}_{2.5}$ 浓度时空预测, 进而支撑后续人群污染暴露与公平性评估。

本研究所使用的个体轨迹数据由北京摸星科技有限公司提供。这些数据通过多种定位技术, 如 Wi-Fi、蓝牙和 GPS, 定时高频地捕捉并记录用户的行为动态。涵盖了北京、上海、深圳、成都、武汉和西安等 6 个城市的全域空间。数据集中包括了匿名用户 ID、时间

¹ 中国环境监测总站. <https://air.cnemc.cn:18007/>.

² 美国国家航空航天局 (NASA) LAADS DAAC 数据平台. <https://www.earthdata.nasa.gov/>.

³ 欧洲中期天气预报中心 (ECMWF). <https://www.ecmwf.int/>.

⁴ 百度地图开发平台. <https://lbsyun.baidu.com/>.

戳、定位方式和精确的经纬度信息。为了减少居民日常移动模式的波动性对实验误差的影响，本研究选择了 2023 年 1 月 24 日至 2 月 23 日长达一个月的轨迹数据进行分析，为完整反映个体全天 $\text{PM}_{2.5}$ 暴露情况，本研究选取了至少包含 20 个小时以上有效定位记录的个体

表 1 《人群轨迹数据视角下中国 6 个城市 $\text{PM}_{2.5}$ 暴露平等性 1-km 时均栅格数据集 (2023 年 1-2 月)》元数据简表

条 目	描 述
数据集名称	人群轨迹数据视角下中国 6 个城市 $\text{PM}_{2.5}$ 暴露平等性 1-km 时均栅格数据集 (2023 年 1-2 月)
数据集短名	Trajectory_ $\text{PM}_{2.5}$ _Exposure_Equity
作者信息	马志锋, 深圳大学建筑与城市规划学院, 1037341855@qq.com 吴梓昊, 深圳大学建筑与城市规划学院, 2510114024@mails.szu.edu.cn 夏吉喆, 深圳大学建筑与城市规划学院, xiajizhe@szu.edu.cn
地理区域	北京, 上海, 深圳, 成都, 武汉和西安
数据年代	2023.01.24-2023.02.23
数据格式	.tif、.shp、.xlsx
数据量	127 MB (压缩后)
数据集组成	(1) 6 个城市 2023 年 1 月 24 日-2 月 23 日逐时 1-km $\text{PM}_{2.5}$ 浓度空间分布数据; (2) 基于动态手机轨迹的城市人均 $\text{PM}_{2.5}$ 暴露浓度; (3) 各城市不同社会特征 (年龄、收入和性别) 和空间特征 (通勤距离) 群体时均 $\text{PM}_{2.5}$ 暴露数据; (4) 各城市居民 $\text{PM}_{2.5}$ 暴露基尼系数; (5) 各群体 $\text{PM}_{2.5}$ 暴露集中指数与 OLS 关联性分析结果; (6) 县级尺度居民时均 $\text{PM}_{2.5}$ 暴露量
基金项目	国家自然科学基金 (42171400)
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101, 中国科学院地理科学与资源研究所
数据共享政策	(1) “数据”以最便利的方式通过互联网系统免费向全社会开放, 用户免费浏览、免费下载; (2) 最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源; (3) 增值服务用户或以任何形式散发和传播 (包括通过计算机服务器) “数据”的用户需与《全球变化数据学报 (中英文)》编辑部签署书面协议, 获得许可; (4) 摘取“数据”中的部分记录创作新数据的作者需要遵循 10% 引用原则, 即从本数据集中摘取的数据记录少于新数据集总记录量的 10%, 同时需要对摘取的数据记录标注数据来源 ^[9]
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS, GEOSS, PubScholar, CKRSC

轨迹数据。该轨迹数据的空间精度达到 1 m 以内, 时间精度在 1 h 以内。

3.2 算法原理

3.2.1 高时空分辨率 $\text{PM}_{2.5}$ 分布模拟

为精确评估不同人群的 $\text{PM}_{2.5}$ 暴露差异, 需获取高时空分辨率的 $\text{PM}_{2.5}$ 浓度分布数据。本研究提出一种基于 Stacking 集成学习的 $\text{PM}_{2.5}$ 逐时预测模型, 融合多源遥感与城市大数据, 实现 1 km 分辨率、逐小时更新的 $\text{PM}_{2.5}$ 浓度估算。

本研究构建了一个两层 Stacking 集成模型, 以捕捉 $\text{PM}_{2.5}$ 与各影响因子间的复杂非线性关系, 如图 1 所示。第一层基学习器, 选用 XGBoost^[10]、梯度提升机 (Gradient Boosting Machine, GBM)^[11]和随机森林 (Random Forest, RF)^[12], 三者均在空气质量预测中表现优异, 能有效处理高维异构数据。第二层元学习器, 采用弹性网络 (Elastic Net)^[13], 结合

L1 与 L2 正则化，避免过拟合并提升模型泛化能力。模型训练采用五折交叉验证，针对每个城市独立优化超参数（如学习率、树深度、正则化系数等），以适配区域特征（图 1）。

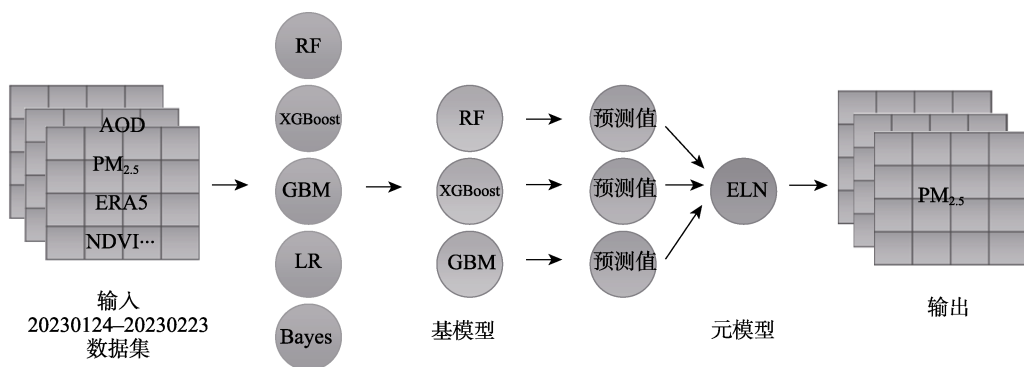


图 1 Stacking 模型架构图

本研究基于相关性筛选相关特征，选取了包括 AOD、NDVI、路网密度等目标特征作为模型训练中的自变量，以监测站点 PM_{2.5} 浓度为因变量。考虑到不同城市在大气环境、地形结构和人类活动方面的差异性，分别训练 6 个城市的独立模型，以增强模型的区域适应性和预测精度。随后将研究区域划分为 1 km×1 km 的规则网格，提取每个网格在对应时段内的特征数据，输入至训练完成的城市独立模型中，输出各网格中心的 PM_{2.5} 预测浓度。

为进一步提升空间连续性，采用反距离权重法对离散预测点进行空间插值，最终生成逐时 1-km PM_{2.5} 浓度空间分布图。

3.2.2 时间加权的 PM_{2.5} 暴露浓度计算

本研究采取时间加权法估算个体的时均 PM_{2.5} 暴露量，该方法综合考虑了每个时间点个体所遭遇的 PM_{2.5} 浓度与其在该点的停留时长，通过累加这些时间点的 PM_{2.5} 浓度乘积，并将总和除以个体全日的记录时长，从而得到个体的时均 PM_{2.5} 暴露浓度。计算原理如图 2 所示。考虑到建筑物对空气污染的阻隔作用^[14]，本研究在对个体 PM_{2.5} 暴露浓度的估算中，根据轨迹数据获取方式来判断个体所处室内还是室外，基于 GPS 信号获取的数据点判定为室外环境，而利用 WiFi 或蓝牙记录的数据点判断为室内环境。同时对每个停留点的 PM_{2.5} 暴露浓度按照室内外暴露修正系数进行调整，确保评估结果的精确性。个体 PM_{2.5} 暴露浓度的时间加权测算公式如下：

$$E_{avg_hour}^{person} = \left(\sum_i^n E_i^{person} * t_i^{person} * \beta_{IO} \right) / T_{day}^{person} \quad (1)$$

式中， $E_{avg_hour}^{person}$ 表示个体的时均 PM_{2.5} 暴露量（ $\mu\text{g}/\text{m}^3$ ）， E_i^{person} 表示个体停留在第 i 个格网暴露的 PM_{2.5} 浓度（ $\mu\text{g}/\text{m}^3$ ）， t_i^{person} 表示个体在第 i 个网格的停留时间长度（h）， β_{IO} 代表室内外 PM_{2.5} 浓度的修正系数，参考已有研究^[14]，本研究假设每类室内外场景的系数等于平均值，取该系数设定为 0.67。 T_{day}^{person} 是个体在一整天中有效记录的总时长（h）。 n 指代观测时间内所停留的格网总数。

为了探究不同群体间的 $\text{PM}_{2.5}$ 暴露差异, 本研究根据个体的特征标签对数据进行了分类聚合, 并根据此计算出每个群体的时均 $\text{PM}_{2.5}$ 暴露浓度。从而为评估和比较各群体的暴露风险提供了一种有效的量化手段, 计算公式如下:

$$E_{\text{avg_hour}}^{\text{group}} = \sum E_{\text{avg_hour}}^{\text{person}} / N^{\text{group}} \tag{2}$$

式中, $E_{\text{avg_hour}}^{\text{group}}$ 表示群体的平均每小时 $\text{PM}_{2.5}$ 暴露浓度 ($\mu\text{g}/\text{m}^3$), N^{group} 表示具备某特征的群体数量。通过计算群体的时均 $\text{PM}_{2.5}$ 暴露量, 能够在相对公平的维度有效描绘出群体层面上的暴露风险, 从而保证后续不同群体 $\text{PM}_{2.5}$ 暴露差异评估的科学性。

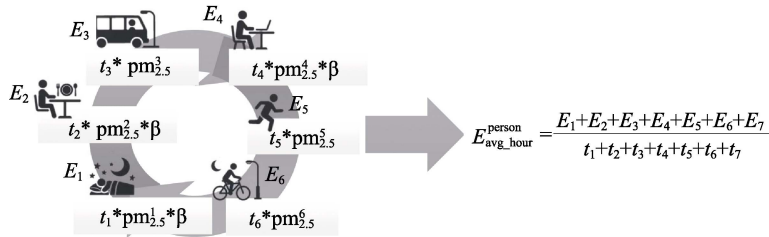


图 2 时均 $\text{PM}_{2.5}$ 暴露计算示意图

3.2.3 群体轨迹驱动的 $\text{PM}_{2.5}$ 暴露平等性评估

(1) 基尼系数计算

基尼系数 (Gini Index, GI), 作为衡量不平等的经典统计工具, 被广泛应用于收入分配、教育水平和健康指标等多个分布领域的不平等研究^[15]。其优势在于能够通过一个规范化的值 (范围从 0 到 1) 来综合地反映一个分布中的不平等程度。同时, 基尼系数对于收入或暴露水平的中间区域的变化异常敏感, 因此, 本研究采用基尼系数来量化和比较不同城市或群体之间的暴露不平等, 基尼系数计算公式如下:

$$GI = \frac{\sum_{i=1}^n \sum_{j=1}^n \frac{|y_i - y_j|}{2n \sum_{i=1}^n y_i}}{\tag{3}}$$

式中, $|y_i - y_j|$ 是个体间 $\text{PM}_{2.5}$ 暴露值的差异 ($\mu\text{g}/\text{m}^3$), 表示群体内部不同个体间差异的绝对值。 $2n \sum_{i=1}^n y_i$ 则是所有个体 $\text{PM}_{2.5}$ 暴露值的两倍平均数。其中, i 和 j 分别表示群体中不同个体的编号, n 为该群体内的样本总数。它归一化了总差异, 使得基尼系数不受样本大小的影响。

(2) 集中指数计算

由于计算机制的限制, 基尼系数仅适用于一维数据, 无法解析影响不平等的多个维度间的复杂关系, 例如, 它无法直接揭示收入水平与 $\text{PM}_{2.5}$ 暴露不平等之间的相互作用。因此, 本研究采用集中指数^[16] (Concentration Index, CI) 来探讨社会 and 空间特征与 $\text{PM}_{2.5}$ 暴露不平等之间的关系, 集中指数计算公式如下:

$$CI = \frac{2}{\bar{y}} \sum_{i=1}^n (y_i - \bar{y}) \left(x_i - \frac{1}{2} \right) \tag{4}$$

式中， $\bar{y}$ 表示 PM_{2.5} 暴露水平的平均值 ($\mu\text{g}/\text{m}^3$)， x_i 表示某项社会或空间特征的值， y_i 代表关联个体的 PM_{2.5} 暴露浓度 ($\mu\text{g}/\text{m}^3$)。其中， i 表示按照社会或空间特征值从小到大排序后的个体位次， n 为样本总数。集中指数 (CI) 的计算结果范围在-1 至 1 之间，以收入与 PM_{2.5} 暴露浓度的关联性为例，CI 值趋近于 0 意味着收入不同的群体间 PM_{2.5} 的暴露差异较小，当 CI 值小于 0 时，表明低收入群体承受较高的暴露量，反之，CI 值大于 0 时代表高收入人群承受较高暴露量。

4 数据结果与验证

4.1 数据集组成

数据集共包括以下 5 方面数据：(1) 各城市逐时 1-km PM_{2.5} 浓度空间分布数据；(2) 基于动态手机轨迹的城市人均 PM_{2.5} 暴露浓度数据；(3) 各城市不同社会 and 空间特征群体时均 PM_{2.5} 暴露数据；(4) 各城市居民 PM_{2.5} 暴露基尼系数；(5) 各城市不同社会 and 空间特征群体 PM_{2.5} 暴露集中指数和 OLS 关联性分析结果；(6) 各城市县级尺度居民时均 PM_{2.5} 暴露量数据。

4.2 数据结果分析

4.2.1 各城市 PM_{2.5} 暴露平等性评估结果

时均 PM_{2.5} 暴露均值为评估群体暴露水平的直观指标，如表 2 所示，各城市中西安的暴露均值最高，为 69.22 $\mu\text{g}/\text{m}^3$ ；相对的，深圳的暴露均值最低，仅 19.96 $\mu\text{g}/\text{m}^3$ 。其他城市如成都、武汉、上海和北京的暴露均值分别为 56.36 $\mu\text{g}/\text{m}^3$ 、50.46 $\mu\text{g}/\text{m}^3$ 、43.74 $\mu\text{g}/\text{m}^3$ 和 38.44 $\mu\text{g}/\text{m}^3$ 。这些数据与各城市 PM_{2.5} 年均监测值及其排名基本一致。

基尼系数展示了各城市的 PM_{2.5} 暴露不平等对比情况。北京以 0.49 的基尼系数排在最高，意味着北京城市内部，居民间的 PM_{2.5} 暴露水平差异最大。相对而言，深圳以 0.10 的基尼系数展示了其 PM_{2.5} 暴露水平较为均匀，城市内部差异较小。

表 2 城市人均 PM_{2.5} 暴露浓度和暴露基尼系数对比表

城市	北京	上海	深圳	武汉	成都	西安
人均 PM _{2.5} 暴露浓度 ($\mu\text{g}/\text{m}^3$)	38.44	43.74	19.96	50.46	56.36	69.22
基尼系数	0.49	0.24	0.10	0.18	0.20	0.17

4.2.2 各社会特征人群 PM_{2.5} 暴露平等性评估结果

图 3 展示了 6 个城市内部不同收入、年龄、性别群体的时均 PM_{2.5} 平均暴露量的汇总统计。年龄段与性别标签源自轨迹数据供应商，涵盖了 22 岁至 85 岁用户，以 5 岁为一个分组间隔。收入水平采用个体所在地区的房价作为替代变量，通过三阶层模型划分为 9 个级别^[17]。

如表 3 所示，在收入维度上，北京、深圳、武汉和成都的 CI 值均大于零，表明在这些

城市较高的 $\text{PM}_{2.5}$ 暴露集中在高收入人群中。相反，上海和西安的 CI 值小于零，表明在这两个城市低收入人群承受了相对较大的 $\text{PM}_{2.5}$ 暴露。从不同特征群体与 $\text{PM}_{2.5}$ 暴露关联性分析结果可以看出，在年龄维度上，北京、武汉、成都和西安的居民 $\text{PM}_{2.5}$ 暴露量随年龄增长而减少。相反，在上海，年龄的增长与 $\text{PM}_{2.5}$ 暴露量的增加呈正相关；在性别维度上，北京和上海的男性群体相较于女性经历更高水平的 $\text{PM}_{2.5}$ 暴露，而在成都，女性的暴露水平高于男性。

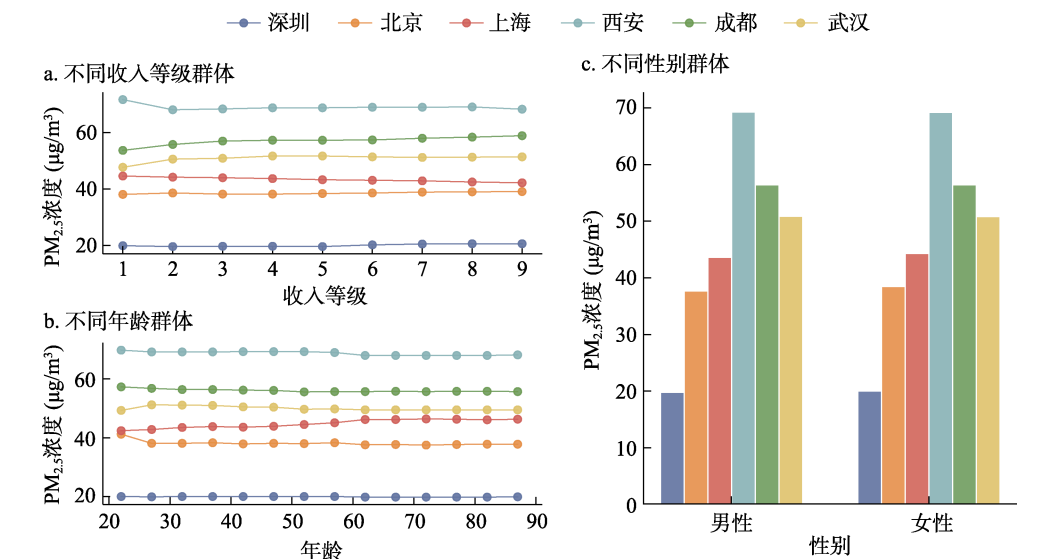


图 3 各城市不同社会特征群体时均 $\text{PM}_{2.5}$ 暴露量统计图

表 3 各城市不同社会特征群体 $\text{PM}_{2.5}$ 暴露集中指数和暴露关联性 OLS 分析

城市	北京	上海	深圳	武汉	成都	西安
收入 (CI)	0.0077	-0.0115	0.0039	0.0123	0.0190	-0.0050
年龄 (CI)	-0.0066	0.0090	-0.0015	-0.0032	-0.0047	-0.0034
性别 (CI)	0.0027	-0.0339	-0.0078	-0.0284	-0.0087	-0.0093
收入 (OLS)	0.315*	-0.701*	0.237*	4.8***	4.8***	-2.17**
年龄 (OLS)	-1.44***	1.29***	0.077	-0.403*	-0.432**	-0.713**
性别 (OLS)	21.6**	11.51**	3.6*	0.232	-4.66**	-2.57*

*表示 $P<0.1$ ，边缘显著性；**表示 $P<0.05$ ，显著；***表示 $P<0.01$ ，高度显著。

4.2.3 各空间特征人群 $\text{PM}_{2.5}$ 暴露平等性评估结果

为了探讨不同空间特征群体的 $\text{PM}_{2.5}$ 暴露差异情况。本研究采用三阶层模型对居民的通勤距离进行划分，以探究日常通勤距离与 $\text{PM}_{2.5}$ 暴露量的关系。由表 4 可知，北京和上海的中长距离通勤者承受了更多的暴露，值得注意的是当通勤距离超过特定阈值时，暴露水平有所降低 (图 4)；而在武汉短途通勤者会遭遇更多的暴露；在深圳、西安和成都， $\text{PM}_{2.5}$ 暴露集中在中长距离通勤者，但是结果并非显著。

表 4 各城市不同空间特征群体 PM_{2.5} 暴露集中指数和暴露关联性 OLS 分析

城市	北京	上海	深圳	武汉	成都	西安
通勤距离 (CI)	0.0308	0.0749	0.0128	-0.0580	0.0099	0.0157
通勤距离 (OLS)	2.2***	0.8***	0.2	-1.9***	0.1	0.1

*表示 P<0.1, 边缘显著性; **表示 P<0.05, 显著; ***表示 P<0.01, 高度显著。

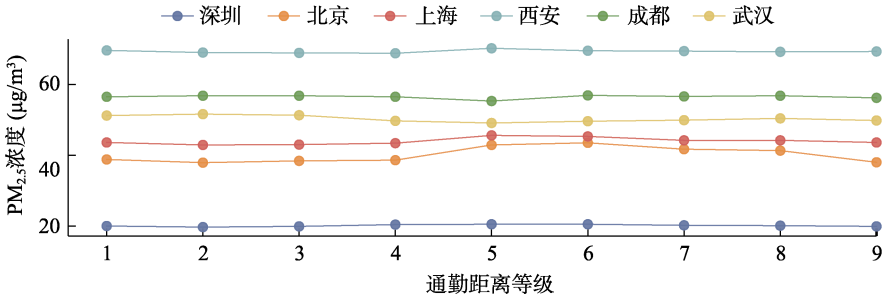


图 4 各城市不同通勤距离群体时均 PM_{2.5} 暴露量统计图

图 5 描绘了 6 个城市县级层面的居民时均 PM_{2.5} 暴露量的空间分布,从空间分布上看,各城市 PM_{2.5} 暴露的分布特征有明显异质性。在北京高 PM_{2.5} 暴露群体主要分布于城市东南及东北部;在上海环市中心居民的 PM_{2.5} 暴露风险更高,个别区县居民的 PM_{2.5} 暴露量比市中心多出 26%;在成都与武汉,市中心的居民遭遇了更高水平的 PM_{2.5} 暴露。空间自相关分析 (Moran's I) 显示,成都与武汉的居民 PM_{2.5} 暴露量呈现出显著的空间正相关性,且集聚程度高于其他城市,表明其暴露分布具有明显的中心集聚模式。此外,成都的居民 PM_{2.5} 暴露最高值位于城市南部和东南部;西安的 PM_{2.5} 暴露分布模式与北京和上海类似,主要集中在城市中心环线以外的郊区;深圳的 PM_{2.5} 暴露分布呈现出多中心分布模式。

4.3 数据结果验证

本研究所采用的逐时 PM_{2.5} 分布数据是由 Stacking 集成模型的预测获得,模型在所有城市中均呈现出较高的 R² 值 (0.660 至 0.853 之间) 和较低的 RMSE 值,性能明显优于传统的单一模型,且在不同浓度区间上均展现出优秀性能。与实测日均值对比,模型预测的误差率维持在 10%–20% 范围内,与半年均值相比较,模型预测的误差率仅在 0.9%–10% 之间,具备良好的时空外推能力与稳定性,证明了其可靠性和精确度。

图 6 展示了 6 个城市 PM_{2.5} 浓度的实测与预测值散点分布及拟合结果。Stacking 模型在各浓度区间均表现稳健,预测值与实测值基本沿 45° 对角线分布,显示出良好的一致性。局部存在一定误差,如北京在 40–150 μg/m³ 范围内部分预测值偏高,深圳、武汉、成都的高浓度区间亦呈现类似趋势。该现象可能与训练数据多源于城市中心、PM_{2.5} 浓度普遍较高,导致对边缘站点的高估有关。整体而言,模型在不同城市中表现出较高的准确性与稳定性。

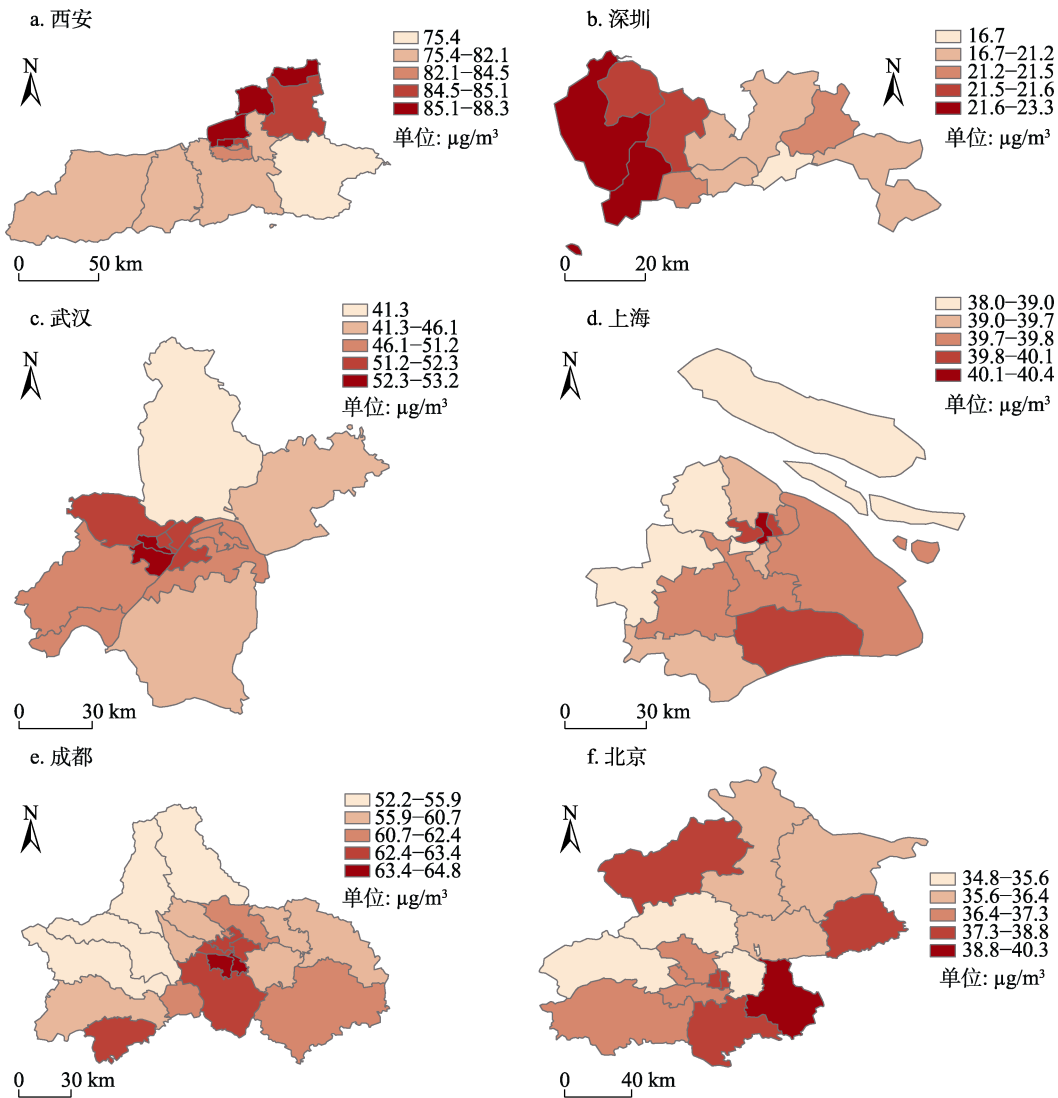


图 5 各城市区县级居民时均 $PM_{2.5}$ 暴露量空间分布图

5 讨论和总结

本数据集通过叠加精细化的数据轨迹与高精度 $PM_{2.5}$ 分布，采用时间加权方法精确计算了个体及群体的 $PM_{2.5}$ 暴露浓度，进而评估了不同城市的整体暴露差异情况，深入探讨了城市内部不同特征群体的 $PM_{2.5}$ 暴露平等性问题。

数据结果表明，时间加权的 $PM_{2.5}$ 暴露浓度测算值与各城市的年均 $PM_{2.5}$ 监测值基本一致，表明了本研究暴露浓度预测方法的可靠性与稳定性。此外，研究区域内的城市均面临一定程度的 $PM_{2.5}$ 暴露不公平问题。基于基尼系数计算结果，北京地区的不平等程度最为严重，而深圳的不平等程度较低，表明不同地区的 $PM_{2.5}$ 暴露差异存在异质性。为解释差异产生的原因，本研究从不同社会和空间特征群体的暴露差异角度切入，结果表明，居民

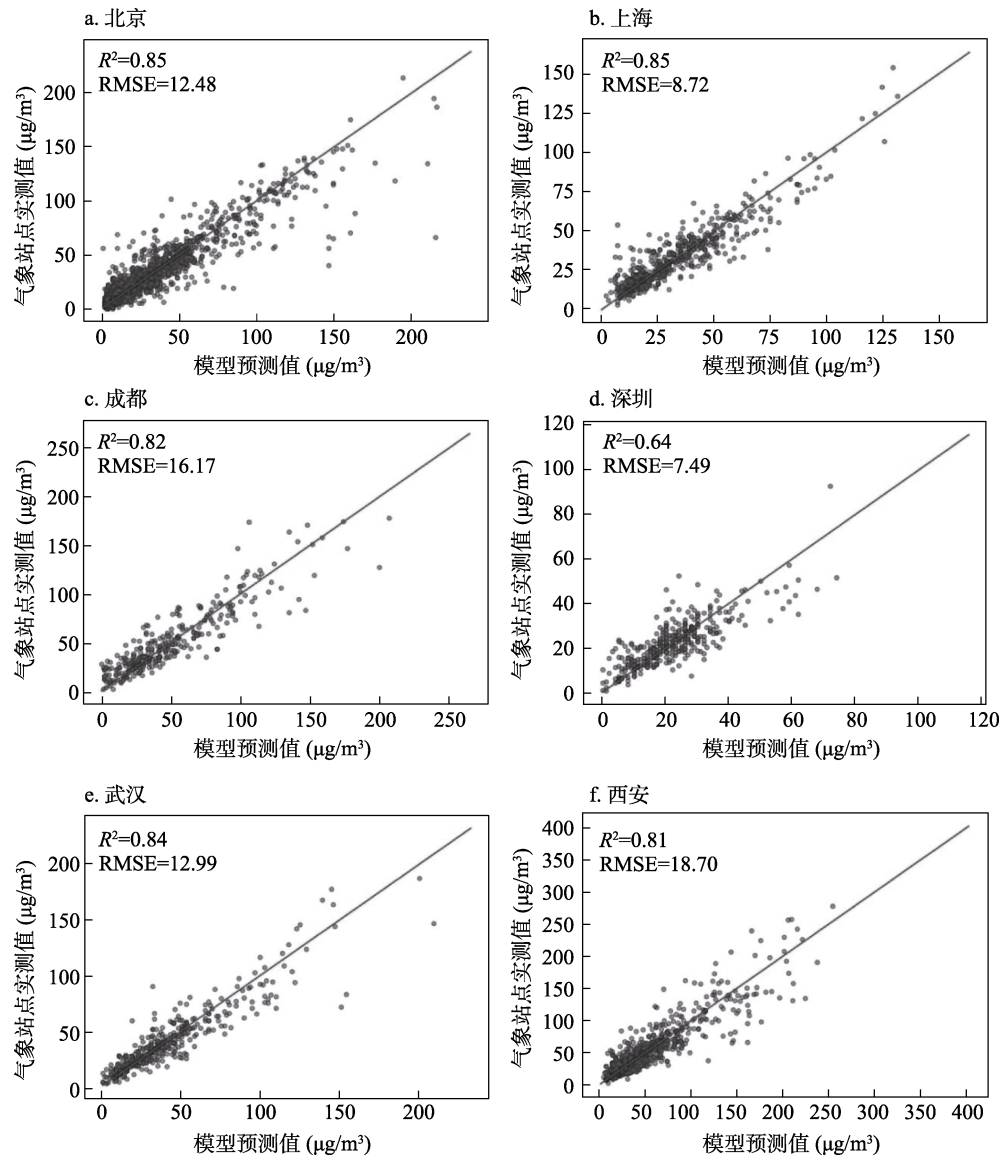


图 6 各城市 Stacking 模型拟合结果图

的 $\text{PM}_{2.5}$ 暴露水平受到其社会和空间特征的显著影响，而这些影响在不同城市间有差异。这一现象可能与城市的城市结构与产业布局有关，北京的城市结构呈环状单中心，居住在郊区的居民工作机会相对较少，导致他们拥有更长的通勤距离，从而遭受更高的暴露量；在上海和西安重污染产业位于郊区，导致郊区的 $\text{PM}_{2.5}$ 浓度偏高，间接促成了郊区居民承受较高的污染暴露水平；武汉市中心的 $\text{PM}_{2.5}$ 污染浓度较高，因此市中心的高收入人群会遭遇更高的暴露量；成都的居民 $\text{PM}_{2.5}$ 暴露量最高值地区位于城市南部和东南部，这与成都的多中心城市布局有关；而深圳的 $\text{PM}_{2.5}$ 暴露分布与城市的多中心结构高度一致，显示了多中心分布模式，使得不同区域的居民暴露量受收入或通勤距离的影响较小。

这些结果揭示了城市空间布局和产业布局对不同人群环境暴露风险的复合影响，也为

精准识别城市中的 $\text{PM}_{2.5}$ 暴露弱势群体提供了关键依据,为城市 $\text{PM}_{2.5}$ 污染治理提供了精确的数据支持,有助于环境措施的精细化,对于实现环境正义具有重要意义。本数据集仅覆盖了中国的6个城市,可能未能完全反映中国 $\text{PM}_{2.5}$ 暴露不平等的真实情况,但从整体来看,本数据集还是具有一定的代表性。后续的研究需要引入更多的影响因子以进一步挖掘暴露不平等现象背后更深层的原因。

作者分工: 夏吉喆对数据集的开发做了总体设计;吴梓昊采集和处理了数据并撰写了数据论文;吴梓昊和马志锋设计了模型和算法并做了数据验证。

利益冲突声明: 本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献

- [1] Chen, H., Kwong, J. C., Copes, R., *et al.* Exposure to ambient air pollution and the incidence of dementia: a population-based cohort study [J]. *Environment International*, 2017, 108: 271–277.
- [2] Guo, Y., Zeng, H., Zheng, R., *et al.* The association between lung cancer incidence and ambient air pollution in China: a spatiotemporal analysis [J]. *Environmental Research*, 2016, 144: 60–65.
- [3] Maji, K. J., Dikshit, A. K., Arora, M., *et al.* Estimating premature mortality attributable to $\text{PM}_{2.5}$ exposure and benefit of air pollution control policies in China for 2020 [J]. *Science of the Total Environment*, 2018, 612: 683–693.
- [4] 环境保护部, 国家质量监督检验检疫总局. 环境空气质量标准: GB 3095—2012 [S]. 第3版. 北京: 中国环境科学出版社, 2012: 3.
- [5] Collins, T. W., Grineski, S. E., Shaker, Y., *et al.* Communities of color are disproportionately exposed to long-term and short-term $\text{PM}_{2.5}$ in metropolitan America [J]. *Environmental Research*, 2022, 214: 114038.
- [6] Ouyang, W., Gao, B., Cheng, H., *et al.* Exposure inequality assessment for $\text{PM}_{2.5}$ and the potential association with environmental health in Beijing [J]. *Science of the Total Environment*, 2018, 635: 769–778.
- [7] Bravo, M. A., Anthopolos, R., Bell, M. L., *et al.* Racial isolation and exposure to airborne particulate matter and ozone in understudied US populations: environmental justice applications of downscaled numerical model output [J]. *Environment International*, 2016, 92: 247–255.
- [8] 马志锋, 吴梓昊, 夏吉喆. 人群轨迹数据视角下中国6个城市 $\text{PM}_{2.5}$ 暴露平等性 1-km 时均栅格数据集 (2023年1–2月) [J/DB/OL]. 全球变化数据仓储电子杂志, 2026. <https://doi.org/10.3974/geodb.2026.02.08.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2026.02.08.V1>.
- [9] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策 [OL]. (2017年更新). <https://doi.org/10.3974/dp.policy.2014.05>.
- [10] Chen, Z. Y., Zhang, T. H., Zhang, R., *et al.* Extreme gradient boosting model to estimate $\text{PM}_{2.5}$ concentrations with missing-filled satellite data in China [J]. *Atmospheric Environment*, 2019, 202: 180–189.
- [11] Danesh Yazdi, M., Kuang, Z., Dimakopoulou, K., *et al.* Predicting fine particulate matter ($\text{PM}_{2.5}$) in the Greater London Area: an ensemble approach using machine learning methods [J]. *Remote Sensing*, 2020, 12(6): 914.
- [12] Sun, J., Gong, J., Zhou, J. Estimating hourly $\text{PM}_{2.5}$ concentrations in Beijing with satellite aerosol optical depth and a random forest approach [J]. *Science of the Total Environment*, 2021, 762: 144502.
- [13] Zou, H., Hastie, T. Regularization and variable selection via the elastic net [J]. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 2005, 67(2): 301–320.
- [14] 邵智娟. 室内 $\text{PM}_{2.5}$ 污染及对人群健康的影响研究 [D]. 南京: 南京大学, 2019.
- [15] Jbaily, A., Zhou, X., Liu, J., *et al.* Air pollution exposure disparities across US population and income groups [J]. *Nature*, 2022, 601(7892): 228–233.
- [16] Josa, I., Aguado, A. Measuring unidimensional inequality: practical framework for the choice of an appropriate measure [J]. *Social Indicators Research*, 2020, 149(2): 541–570.
- [17] Saunders, P. Social class and stratification [M]. London: Routledge, 2006.