基于机器学习和非参数估计的 $PM_{2.5}$ 风险评估

周 琪[1,2],于 洋[3],刘苗苗[1*],毕 军[1] (1.南京大学环境学院,污染控制与资源化研究国家重点实验室,江苏 南京 210023；2.清华大学环境学院,北京 100084；3.清华大学交叉信息研究院,北京 100084)

摘要：为开展区域风险评估,融合手机信令、气象和地理信息等多源数据,引入随机森林机器学习、非参数估计分位数图示法和非监督学习 K-mean 等方法,构建了区域 $PM_{2.5}$ 风险评估及特征识别评价框架,在南京市区以 0.3km 分辨率网格为基础单元开展了案例研究.结果表明,该技术既可有效模拟 $PM_{2.5}$ 浓度时空分布,十折交叉验证 R^2 达到 0.76,证明了准确度较高,并基于此识别出 4 种主要污染特征;也可有效捕捉短期人口流动导致的风险,在污染浓度不变的情况下短期人口流动会导致风险增加 0.30~0.97 倍.综合 $PM_{2.5}$ 浓度和人口分布,识别出 4 种主要暴露风险模式,其中,研究区域 6.5%的面积为高风险地区,23.0%的面积为低风险地区.“十四五”期间应加快现代科学技术在环境保护领域的应用,实施网格化和差异化的风险控制政策,维护人群健康.

关键词：$PM_{2.5}$；机器学习；非参数估计；暴露风险评估；特征识别

中图分类号：X513,X823 **文献标识码**：A **文章编号**：1000-6923(2022)08-3554-07

Risk assessment of $PM_{2.5}$ pollution based on machine learning and nonparametric estimation. ZHOU Qi[1,2], YU Yang[3], LIU Miao-miao[1*], BI Jun[1] (1.State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing 210023, China；2.School of Environment, Tsinghua University, Beijing 100084, China；3.Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China). *China Environmental Science*, 2022,42(8)：3554~3560

Abstract：A systematic approach of regional $PM_{2.5}$ risk and characterization assessment was developed in this study by integrating random forest model, Quantile-Quantile plot model, and K-mean model, based on multi-source data including mobile phone signals, meteorological data, geographic data, etc. This new approach was further applied in a case study of Nanjing at a 0.3km resolution grid. On the one hand, this new approach effectively simulated the temporal and spatial distribution of the $PM_{2.5}$ concentration with an10-fold cross-validation R^2 of 0.76 and screened out four major pollution characteristics. On the other hand, it effectively captured the short-term population mobility risk. Short-term population mobility increased the $PM_{2.5}$ exposure risk by 0.30~0.97 times, even keeping $PM_{2.5}$ concentration unchanged. After combining $PM_{2.5}$ concentration and population mobility simultaneously, four major $PM_{2.5}$ exposure risk modes were identified. 6.5% of the areas of Nanjing were at high risk, and 23.0% were at low risk. During the 14th Five Year Plan, it is suggested that the government should speed up the application of modern science and technology in environmental protection and implement gridding and differentiated policies on air pollution risk control to promote human health.

Key words：$PM_{2.5}$；machine learning；non-parametric estimation；exposure risk assessment；feather recognition

大范围、高浓度的大气细颗粒物($PM_{2.5}$)污染是城市大气污染最突出的特征之一[1].国内外大量的环境流行病学研究已证明 $PM_{2.5}$ 暴露与呼吸系统疾病、心脑血管疾病等一系列负面健康效应显著相关[2],是我国第四大致死风险因素[6],引起社会的广泛关注.在此背景下,亟需高精度算法精准描绘 $PM_{2.5}$ 暴露风险,为大气污染风险精准防控提供科学依据.根据暴露风险评估的基本原理,这既需要高精度风险受体和 $PM_{2.5}$ 浓度时空分布数据的支撑,也需要高效能特征识别算法的支撑.

作为风险受体,人口分布会对 $PM_{2.5}$ 暴露风险产生影响.目前,考虑高强度人口位置变化的 $PM_{2.5}$ 暴露风险评估研究较少,传统研究多采用以年为单位的长期人口产品开展低频率的 $PM_{2.5}$ 暴露风险评估[3-7].伴随现代通信技术的快速发展,手机信令等大数据的引入使量化短期动态的人口流动成为了可能[8].随着学者对大数据表征人口流动性能力的认识逐步深入,大数据已逐渐被应用于比利时[9]、纽约[10]等发达国家和地区的城市环境风险暴露评估,为在发展中国家应用手机信令数据在 $PM_{2.5}$ 暴露风险评估研究中引入精细化人口流动分布提供了良好示范.另一方面,人口出行伴随的交通工具的使用

收稿日期：2021-12-27

基金项目：国家自然科学基金资助项目(72174084);国家自然科学基金资助项目(71761147002);中央高校基本科研业务费(0211-14380171)

* 责任作者, 副教授, liumm@nju.edu.cn

可能会影响局地 $PM_{2.5}$ 浓度.之前的研究因为难以直接获得人口流动数据,一般采用代理变量近似表征人口的流动性[11],因此难以判断高强度人口流动是否会显著影响污染时空分布特征.因此,本文将融合高精度手机信令数据,实现以上两方面的联合突破.

在 $PM_{2.5}$ 浓度时空分布模拟方面,研究方法包括参数估计模型[12-13]、非参数估计模型[14-18]和模式模拟[19]等.其中,非参数的机器学习算法如人工神经网络[14]、径向基函数神经网络[15]、boosting 方法[16]、反向神经网络[17]、随机森林[18]等在近年来得到了愈发广泛的应用,已成功被发达国家应用于 $PM_{2.5}$ 地面浓度的模拟.

国内外针对污染和风险暴露模式识别已开发了众多方法,比如基于传统统计学的相关系数法[20]、主成分分析法[21]等;基于监督学习法的层次贝叶斯[22]、支持向量机[23]等;基于非监督聚类方法的模糊推理[24]、K-mean 聚类[25]等方法.然而,类似算法在人口和污染高度聚集的发展中国家尚未得到广泛应用.将这些算法应用于污染、人口和暴露的多维度特征识别,可为区域污染格局和暴露风险特征的绘制和差异化风险管理提供丰富的信息.

综上,为了精准评估 $PM_{2.5}$ 暴露风险水平和特征,本文拟采用高精度的、脱敏的手机数据模拟风险受体的逐小时变化,采用随机森林机器学习算法实现 $PM_{2.5}$ 浓度的逐小时模拟,并通过非参估计分位数图示法和无监督机器学习 K-mean 法,对逐小时 $PM_{2.5}$ 浓度和暴露风险进行特征模式识别.

1　数据与方法

1.1　研究区域及数据

选取江苏省南京市作为案例城市开展 $PM_{2.5}$ 风险评估的实证研究.南京市位于长江下游中部地区、江苏省西南部、毗邻安徽省.2018 年南京共设置了 9 个空气质量监测国控站点.选定的研究区域为南京市主城区,地理坐标为 31.97°~32.08°N,118.69°~118.83°E,覆盖了南京市玄武区、秦淮区、鼓楼区、建邺区等(图 1).模拟结果以 0.3km 分辨率的网格展示,时间范围为 2018 年 5 月 28 日~6 月 20 日,时间分辨率为小时.

本文采用的逐小时、基站尺度、脱敏的手机用户数量统计数据来自我国三大运营商之一.截止到 2018 年 6 月 20 日,在南京市 6587km^2 的土地上,已架设了 2 万余个基站单元,负责全市 300 多万用户的日常通讯.$PM_{2.5}$ 地面监测站点数据来自中国环境监测总站[26].此外,本文采用的潜在影响 $PM_{2.5}$ 浓度的因素包括气象条件、污染源数据、基础地理数据等.其中,风速、能见度、温度、露点温度数据来自中国气象局[27];POI(兴趣点)、交通、建筑、水域数据来自 OpenStreetMap[28];土地利用数据来自中国科学院资源环境科学与数据中心[29].

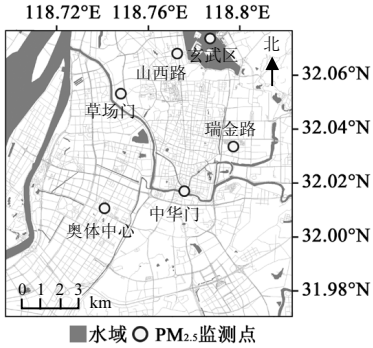

图 1　研究区域及 $PM_{2.5}$ 监测站设置情况

Fig.1　$PM_{2.5}$ monitoring stations in the study area

1.2　风险受体时空分布格局模拟

在不涉及用户隐私的情况下,手机运营公司根据本文需求脱敏提取了研究区域内用户的手机信令数据,在清洗了冗余记录、错误记录、乒乓效应之后,汇总获得了逐小时基站尺度用户统计数.数据集描绘了每个基站单元每小时服务的用户总数.之后,采用基于距离最近原则划分几何平面的泰森多边形算法,将研究区域划分为若干围绕基站的多边形 i,并将对应基站第 h 小时承载的用户数赋值给多边形,记为 $U_{i,h}$,多边形 i 的平均面积为 0.11km^2.考虑到后续随机森林算法对数据形态的要求,参考式(1)将 i 重采样至 0.3km 分辨率的网格 j 中.假设手机用户的分布与南京市人口分布同质,参考式(2)利用区域内总人口对逐小时人口空间分布进行换算,获得逐小时 0.3km 分辨率的人口流动分布格局.

$$U_{j,h} = U_{i,h} \times \frac{S_j}{S_i} \tag{1}$$

$$P_{j,h} = \frac{U_{j,h}}{\sum_1^J U_{j,h}} \times P_{2018} \tag{2}$$

式中: h 为小时代码;i 代表泰森多边形的代码;j 代表

网格的代码;J 为网格总数;S_i 代表泰森多边形的面积;S_j代表网格的面积,0.09km^2;$U_{i,h}$代表泰森多边形 i 中第 h 小时基站所收集的用户数量; $U_{i,h}$为网格 j 中第 h 小时的用户总量;$\sum_1^J U_{j,h}$ 为全部网格第 h 小时的用户总量;P_{2018} 为研究区域内 2018 年的总人口;$P_{j,h}$为估算的网格 j 第 h 小时的总人口.

1.3 基于随机森林和非参估计的 $PM_{2.5}$ 浓度特征分析

随机森林算法是基于多个决策树的集成模型,已成功应用于解决各领域的实际问题,包括空气质量[18]、化学信息学[30]、生态学[31]等.本文选择随机森林模型,一是考虑随机森林模型可在没有明确定义函数形状的情况下完成多变量、非参数和非线性分类或回归任务;二是它可输出预测变量重要性排序,有助于识别高强度人口流动是否会影响 $PM_{2.5}$ 浓度的模拟效果.参考 Hu 等[18]的方法,本文构建了 $PM_{2.5}$ 小时浓度的随机森林预测模型.模型中采用的预测变量包括:(1)网格距 $PM_{2.5}$ 地面监测站的距离,m;(2)风速,miles/h;(3)能见度,英里;(4)温度,°C;(5)露点温度,°C;(6)人群流动量,人;(7)POI 数量,个;(8)道路长度,m;(9)交通交叉口数量,个;(10)建筑物面积,m^2;(11)水体面积,m^2;(12)土地利用类型.数据均提前重采样至 0.3km 分辨率.采用超参数调试法对建模中两个重要参数,预测器数目(m_{try})和生长树数目(n_{tree})进行调参.最终,当预测器数目被设置为 10 个,生长树被设置为 1000 个时,模型具有最佳预测精度.采用十折交叉验证法评估模型拟合效果,性能评估指标包括 R^2、均方根误差(RMSE)、平均预测误差(MPE)和相对预测误差(RPE).利用随机森林模型中的 Gini 系数法提取预测变量的重要性排序.

此外,为识别和评估 $PM_{2.5}$ 浓度分布特征,采用分位数图示法逐个检验每小时所有网格 $PM_{2.5}$ 浓度的概率密度分布形式.分位数图示法是一种验证两组数据是否来自同一分布,或者验证一组数据是否来自某个具体分布的方法.分位数图示法的思路是首先将横纵坐标的实际数据排序,分别计算其分位数并标记相同分位数的交点.若两个分布较接近,则交点应分布在 $y=x$ 上.如果 x 轴为已知的固定分布,则可验证数据是否满足这种特定要求.根据预实验结果,本文采用的分布形式包括正态分布、卡方分布、双正态叠加分布和多正态叠加分布.

1.4 基于非监督机器学习的 $PM_{2.5}$ 暴露风险特征识别

为精准评估 $PM_{2.5}$ 暴露风险,从人口和污染两个维度,采用非监督机器学习 K-mean 聚类方法逐小时对人群 $PM_{2.5}$ 暴露风险分类.K-mean 是一种基于欧式距离的聚类算法,认为两个目标之间的距离越近,相似度越大,其中的 K 代表分类的数量.本文采用 Gap Statistic 方法逐个计算每小时最适合的 K 值,并统计出现频次最高的 K 值作为统一的分类数量.然后,基于每小时的分类结果,记录每个网格的类型结果时间序列,并在网格尺度选择出现次数最多的类型作为此网格的主导类型.最后,根据暴露风险的分类和汇总结果,分析人群的 $PM_{2.5}$ 暴露风险特征,并识别高风险地区.

2 结果和讨论

2.1 风险受体流动性评估和特征识别

作为影响人群 $PM_{2.5}$ 暴露风险水平的一大重要因素,在环境 $PM_{2.5}$ 浓度不变的情况下,每时每秒高强度的人口流动会导致风险的快速变化.由图 2 可知,4 个典型时段的平均人口总量在 231.66~349.02 万人之间,相当于同一区域内人口普查常住人口数的 1.30~1.97 倍.在污染浓度不变的情况下短期人口流动会导致风险增加 0.30~0.97 倍.普查数据和人口流动数据之间的绝对差距由两部分构成,一部分是人口普查遗漏的常住人口数量,另一部分则是研究区域承载的短期人口流动数量,比如以休闲娱乐、上班、旅游、出差、打工等为目标的人口群体.因此,在风险评估中引入位置大数据刻画短期高频的人口流动至关重要.

此外,人口的高频流动会导致城市人口分布的时空异质性,从而增加风险的时空异质性.在时间变化方面,研究区域内平均人口从 6:00 的 231.66 万人,增长到 12:00 的 331.92 万人,将导致此时段的风险相提升 43.3%.一天中平均人口数量的 24h 时间变异性为 13.71%,即污染不变时风险的变异系数为 13.71%.在空间变化方面,存在明显的密集区和稀疏区.例如,在图中标注的著名商业中心,新街口商业区,人口密度极大;并且除商业娱乐活动密集外,新街口也是南京重要的交通枢纽,南京地铁一号线和二号线纵横交叉穿过此地区,大量人群在新街口地铁站换乘地

铁.这导致该地区部分网格的小时平均最大人口密度达到了 40.42 万人/km^2,是整个研究区域平均值的几十倍,也即承载着几十倍的 $PM_{2.5}$ 暴露风险.以上发现再次印证了手机信令数据高时空分辨率的优势[8]以及采用高精度手机信令数据模拟高强度、高精度的人口变化情况对于科学量化人群 $PM_{2.5}$ 暴露风险的重要意义[10].

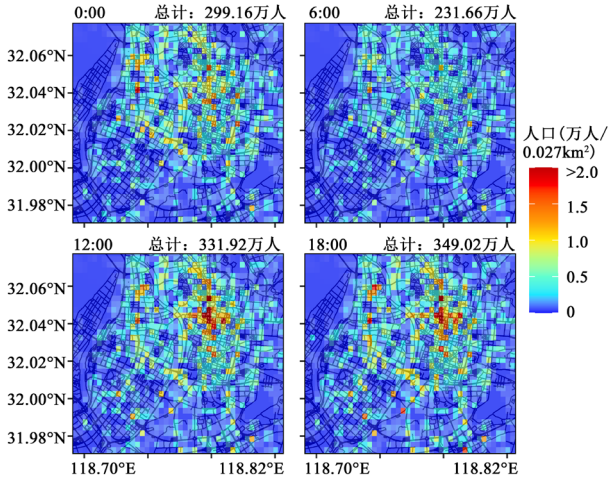

图 2　典型时段的平均人口分布

Fig.2　Average population distribution in typical hours

2.2　$PM_{2.5}$ 污染异质性评估和特征识别

随机森林模型十折交叉验证的结果表明,模型的 R^2 为 0.76,斜率为 1.04,RMSE 为 15.37,MPE 为 10.94,RPE 为 34.93%,模型的准确度和精确度都较高,与之前的研究相当[13,18,32].根据 Gini 系数重要性度量标准,风速(43%)、能见度(25%)、露点温度(10%)、温度(10%)、人口流动(3%)是影响 $PM_{2.5}$ 浓度模拟最重要的 5 个变量,贡献了 Gini 系数下降量 91%的比重.其中,风速、能见度、温度、露点温度为直接影响 $PM_{2.5}$ 扩散条件的气象参数[33].此外,人口流动作为影响 $PM_{2.5}$ 浓度的重要要素,印证了之前的假设,即人口流动伴随的能源消耗可能会影响 $PM_{2.5}$ 的排放[6].

图 3 展示了研究区域典型时段的平均 $PM_{2.5}$ 浓度分布.随机森林模型的引入成功的丰富了 $PM_{2.5}$ 监测的空间覆盖度,更好地捕捉了城市内部 $PM_{2.5}$ 浓度的时空差异.在时间趋势方面,$PM_{2.5}$ 浓度在 6:00~18:00 处于上升状态,21:00 达到最高值 56.15μg/m^3,之后处于下降状态.与南京市地面监测站点的表现相一致,也与其它案例研究一致[34].露点温度与 $PM_{2.5}$ 浓度变化趋势类似.与此相反,风速、温度和边界层高度在 6:00~21:00 处于下降状态,21:00 之后数值逐渐上升.而能见度在白天具有较大的波动,在夜晚一直处于较低值.在空间分布方面,存在明显的清洁区和污染区,比如西部沿江地区的污染物浓度普遍低于东部商业集中区;另一方面,高低污染物地区也伴随时间的变化逐渐迁移、转变.比如在 18:00 的通勤高峰时段,市中心沿道路斑块状的高污染地区呈现向外辐射的趋势.

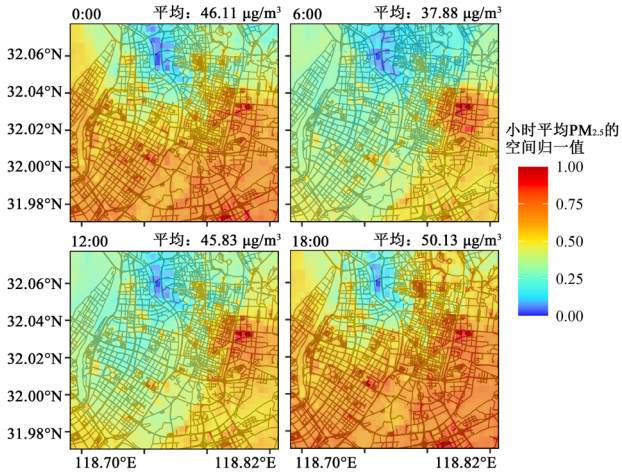

图 3　典型时段的平均 $PM_{2.5}$ 浓度分布(归一值)

Fig.3　Normalized average $PM_{2.5}$ concentration distribution in typical hours

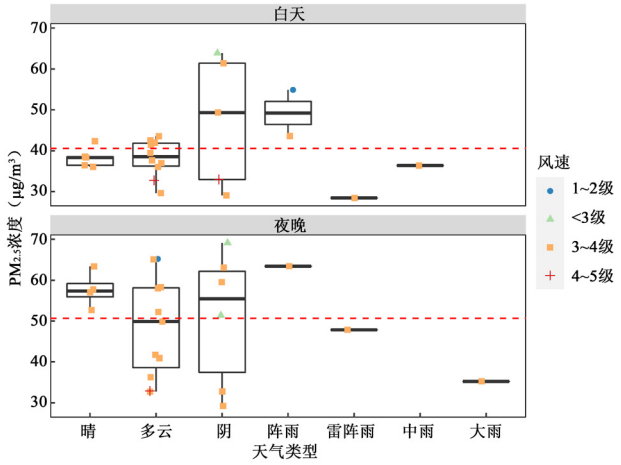

图 4　白天夜晚各天气类型下 $PM_{2.5}$ 浓度的抖动图和箱线图

Fig.4　The jitter and box plot of $PM_{2.5}$ concentration under different weather types during the day and night

虚线分别为白天和夜晚 $PM_{2.5}$ 浓度的平均值

此外,天气类型也是影响 $PM_{2.5}$ 浓度的重要因素.图 4 展示了白天和夜晚各天气类型下 12h 平均 $PM_{2.5}$ 浓度的抖动图和箱线图.夜晚的平均浓度高于白天.

阵雨天气除外的下雨天气,$PM_{2.5}$ 浓度偏低.晴天时 $PM_{2.5}$ 浓度分布较为集中,多云和阴天时 $PM_{2.5}$ 浓度分布较为离散,说明了多云和阴天天气对 $PM_{2.5}$ 浓度的影响较为复杂.此外,风速对 $PM_{2.5}$ 浓度影响较为显著,当风速为 1~2 级时(蓝色散点),$PM_{2.5}$ 浓度均处于各天气类型组中的高值,当风速为 4~5 级时(红色散点),$PM_{2.5}$ 浓度均处于各天气类型组中的低值.总体而言,天气类型、风速、边界层高度、能见度、温度和湿度等多种因素会对 $PM_{2.5}$ 浓度变化产生较大影响.

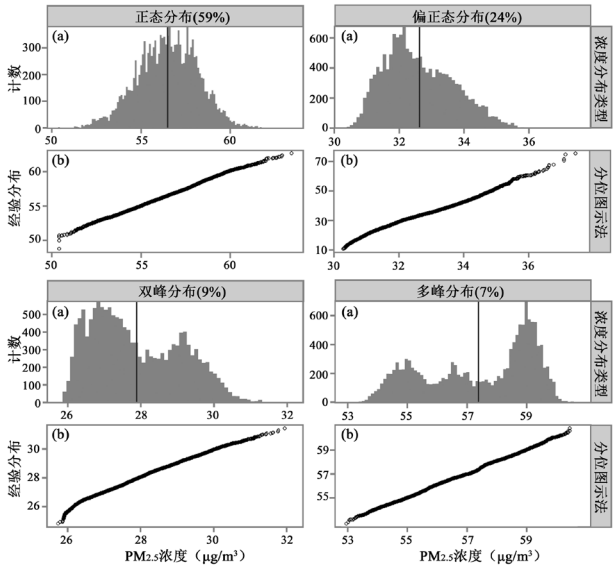

图 5 $PM_{2.5}$ 浓度分布直方图及非参数分位数图示法拟合的 $PM_{2.5}$ 浓度分布散点图

Fig.5 Histogram plot of $PM_{2.5}$ concentration and Quantile-Quantile plot of $PM_{2.5}$ concentration distribution

对逐小时所有网格内 $PM_{2.5}$ 浓度数值的分布逐个进行识别后发现,$PM_{2.5}$ 浓度分布类型可划归为 4 种主要类型:正态分布、偏正态分布、双峰分布和多峰分布.图 5a 中的直方图展示了 $PM_{2.5}$ 浓度分布 4 种分类的典型案例.基于上述定性的分类,使用分位数图示法,分别对逐时污染浓度概率密度分布进行模拟和验证,案例结果如图 5b.分位数图示法基本分布在 $y=x$ 的直线上,说明探索的经验分布函数对于样本分布有较好的表征效果.统计结果表明,82%的时间满足正态或者偏正态分布,即 $PM_{2.5}$ 浓度处于中等水平的地区较多,而处于高、低浓度段的地区较少.16%的时间概率密度函数出现了多个峰值,其中 73%出现在晚上(18:00~第 2 天 6:00).南京市应着重关注夜晚复杂的空气污染状况,提前做好应急措施,降低突发大气污染事件对人群健康的影响.

2.3 高精度 $PM_{2.5}$ 暴露风险评估和特征识别

图 6 展示了 $PM_{2.5}$ 暴露风险特征的 K-mean 聚类典型案例,可划归为 4 种分类方法:(1)人口流动主导模式:两类地区具有相似 $PM_{2.5}$ 浓度,但人口流动数量有显著差异,高人口流动地区具有较高风险;(2)污染主导模式:两类地区具有相似人口流动情况,但 $PM_{2.5}$ 浓度有显著差异,高污染地区具有较高风险;(3)高污染高流动模式(低污染低流动模式):高(低)污染地区和高(低)人口流动地区重合,两者叠加具有较高(低)风险;(4)高污染低流动模式(低污染高流动模式):高(低)污染地区和低(高)人口流动地区重合,风险水平需结合实际数据计算.

风险受体主导模式出现频次达 89 次,占所有类型的 15%,污染主导模式占据 5%,高污染高流动(低污染低流动)模式占据 35%,高污染低流动(低污染高流动)模式占据 44%.单因素主导类型出现的频次最少,总计占据 20%.大部分情况下,暴露风险受到受体分布和污染分布的双重影响.

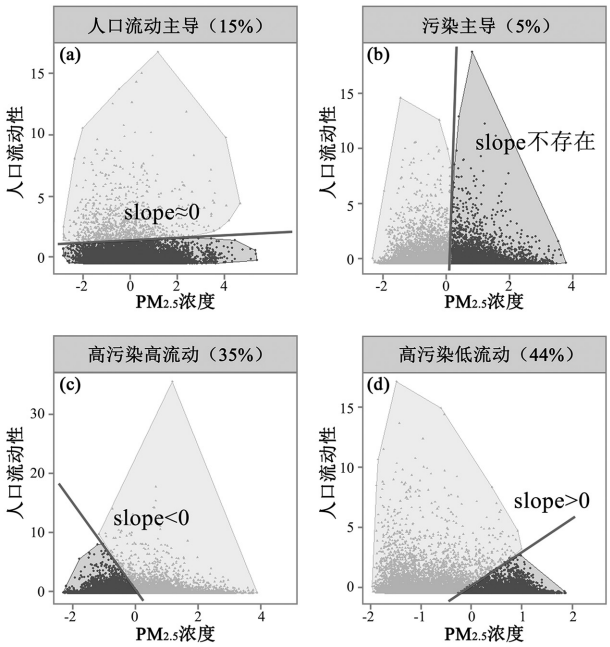

图 6 $PM_{2.5}$ 暴露聚类模式归纳

Fig.6 Cluster types of $PM_{2.5}$ exposure

每个网格的主导暴露风险分类结果见图 7a.经过分类汇总和最大概率筛选后,网格受两种分类方法主导,分别为高污染高流动模式(低污染低流动模

式)和高污染低流动模式(低污染高流动模式).说明在城市复杂的环境下,人口暴露风险特征受到人口流动和 $PM_{2.5}$ 污染浓度的共同影响.其中,研究区域 6.5%的面积更高概率出现高污染高流动模式,45.4%的面积呈现高污染低流动模式,25.1%的面积呈现低污染高流动模式,23.0%呈现低污染低流动模式.高风险主要集中在南京市的市中心,如新街口、夫子庙、大行宫等地区;低风险地图主要集中在西北和东北的长江、玄武湖地区.

此外,4 个典型时段的主导暴露风险分类结果见图 7b,存在较明显的时空异质性.其中,6:00 几乎不存在高污染高流动模式,不论是受体还是 $PM_{2.5}$ 浓度都处于一天中的较低值.0:00 和 18:00 暴露风险类型和总体的分布较为相似,高污染高流动和低污染低流动的空间覆盖范围均较小.12:00 时高污染高流动模式和低污染低流动模式的覆盖面积都存在较为明显的扩大,从典型时段的人口分布和 $PM_{2.5}$ 浓度分布也可观察到两者高值和低值分别的高度重复.因此,需要针对特定时段的风险分布特征动态调整 $PM_{2.5}$ 风险重点管控的区域.

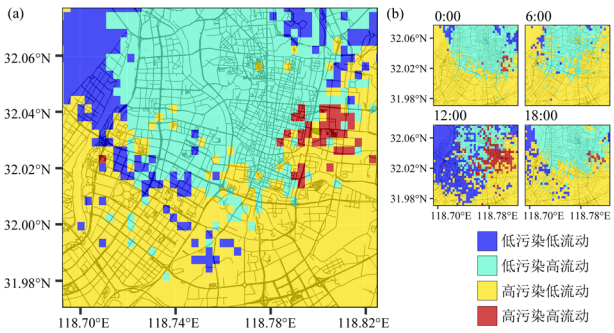

图 7　总体和典型时段 $PM_{2.5}$ 暴露聚类地图

Fig.7　Cluster map of $PM_{2.5}$ exposure in general and typical hours

3　结论与政策建议

3.1　结论

3.1.1　创新性地采用手机信令数据提取风险受体,即人口流动的时空分布特征,相比传统静态人口模拟的方法,提高了其分布的时空分辨率.人口流动性风险的引入为高异质性人群暴露风险特征的识别提供坚实的数据基础.

3.1.2　引入人口流动作为表征交通能源消耗的替代指标纳入随机森林 $PM_{2.5}$ 浓度预测模型中,发现人口流动为重要特征之一,可提高模型的模拟效果.此外,随机森林算法可以较好的实现对 $PM_{2.5}$ 时空分布的模拟,模拟结果的 R^2 达到 0.76.

3.1.3　采用非参数学习和非监督学习的方法,对南京市 $PM_{2.5}$ 污染和人群 $PM_{2.5}$ 暴露风险进行了科学的分析和刻画.成功识别出南京市区内的 4 种主要污染特征和 4 种主要风险模式,发现复杂污染特征主要出现在晚上,高风险主要出现在新街口、夫子庙、大行宫等地区.总体看来,研究区域 6.5%的面积为高风险,23.0%的面积为低风险,而剩余 70.5%面积的风险需要具体问题具体分析.

3.2　政策启示

引入了手机信令大数据、气象数据、土地数据等多源数据对人群 $PM_{2.5}$ 暴露风险开展了精细化模拟,为多源大数据融合支撑环境研究提供了示范.引入大数据机器学习算法实现了城市环境风险的评估和特征精准化刻画,为精准网格化风险管理提供了助力和手段.“十四五”期间我国大气污染防控可融合移动通讯等现代科学技术,对城市环境风险开展高精度、高频率实时监测,为政府的环境风险调控、应急管理和风险交流提供数据支撑.

“十四五”期间我国大气污染防控目标需要根据污染和暴露风险特征优化设计.研究发现,人口流动、污染和暴露风险特征具有时空异质性,需要在识别出的重点污染时段和高风险地区制定差异化的 $PM_{2.5}$ 污染控制措施.研究区 $PM_{2.5}$ 暴露风险管控的优先次序为高污染高流动地区>高污染低流动和低污染高流动地区>低污染低流动地区.其中,建议将高污染高流动地区列入空气污染重点管控地区,引入网格化微型监测站等新型监测单元,构建现代环境空气质量的检测和预警预报平台,对空气质量变化情况进行及时预警,从而实现区域污染的网格化检测、准确监控和精准执法.

参考文献:

[1] Ma Z, Hu X, Huang L, et al. Estimating ground-level $PM_{2.5}$ in China using satellite remote sensing [J]. Environmental Science & Technology, 2014,48(13):7436-7444.

[2] 朱翠云,何　清,赵竹君,等.乌鲁木齐市区与南郊山区颗粒物污染特征对比分析 [J]. 中国环境科学, 2022:1-17.
Zhu C, He Q, Zhao Z, et al. Comparative analysis of particulate pollution characteristics between Urumqi urban area and southern mountainous area [J]. China Environmental Science, 2022:1-17.

[3] 阮芳芳,曾贤刚,段存儒.基于城市面板数据的 $PM_{2.5}$ 对公共健康的影

响 [J]. 中国环境科学, 2020,40(12):5451-5458.
Ruan F, Zeng X, Duan C. Influence of $PM_{2.5}$ pollution on public health based on urban panel data [J]. China Environmental Science, 2020, 40(12):5451-5458.

[4] Liu M, Huang Y, Ma Z, et al. Spatial and temporal trends in the mortality burden of air pollution in China: 2004~2012 [J]. Environmental International, 2017,98:75-81.

[5] He C Y, Han L J, Zhang R Q. More than 500 million Chinese urban residents (14% of the global urban population) are imperiled by fine particulate hazard [J]. Environmental Pollution, 2016,218:558-562.

[6] Shen H, Tao S, Chen Y, et al. Urbanization-induced population migration has reduced ambient $PM_{2.5}$ concentrations in China [J]. Science Advances, 2017,3(7): e1700300.

[7] Silva R A, West J J, Lamarque J F, et al. Future global mortality from changes in air pollution attributable to climate change [J]. Nature Climate Change, 2017,7(9):647.

[8] Deville P, Linard C, Martin S, et al. Dynamic population mapping using mobile phone data [J]. PNAS, 2014,111(45):15888-15893.

[9] Dewulf B, Neutens T, Lefebvre W, et al. Dynamic assessment of exposure to air pollution using mobile phone data [J]. International Journal of Health Geographics, 2016,15:14.

[10] Nyhan M, Grauwin S, Britter R, et al. "Exposure Track" The impact of mobile-device-based mobility patterns on quantifying population exposure to air pollution [J]. Environmental Science & Technology, 2016,50(17):9671-9681.

[11] Wang F, Ren J, Liu J, et al. Spatial correlation network and population mobility effect of regional haze pollution: Empirical evidence from Pearl River Delta urban agglomeration in China [J]. Environment, Development and Sustainability, 2021,23(11):15881-15896.

[12] Kloog I, Chudnovsky A A, Just A C, et al. A new hybrid spatio-temporal model for estimating daily multi-year $PM_{2.5}$ concentrations across northeastern USA using high resolution aerosol optical depth data [J]. Atmospheric Environment, 2014,95:581-590.

[13] Ma Z, Hu X, Sayer A M, et al. Satellite-based spatiotemporal trends in $PM_{2.5}$ concentrations: China, 2004~2013 [J]. Environmental Health Perspectives, 2016,124(2):184-192.

[14] Gupta P, Christopher S A. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: A neural network approach [J]. Journal of Geophysical Research-Atmospheres, 2009,114(D20):205.

[15] 梁 泽,王玥瑶,岳远紊,等.耦合遗传算法与 RBF 神经网络的 $PM_{2.5}$ 浓度预测模型 [J]. 中国环境科学, 2020,40(2):523-529.
Liang Z, Wang Y, Yue Y, et al. A coupling model of genetic algorithm and RBF neural network for the prediction of $PM_{2.5}$ concentration [J]. China Environmental Science, 2020,40(2):523-529.

[16] Reid C E, Jerrett M, Petersen M L, et al. Spatiotemporal prediction of fine particulate matter during the 2008 Northern California wildfires using machine learning [J]. Environmental Science & Technology, 2015,49(6):3887-3896.

[17] Di Q, Koutrakis P, Schwartz J. A hybrid prediction model for $PM_{2.5}$mass and components using a chemical transport model and land use regression [J]. Atmospheric Environment, 2016,131:390-399.

[18] Hu X F, Belle J H, Meng X, et al. Estimating $PM_{2.5}$ concentrations in the conterminous United States using the random forest approach [J]. Environmental Science & Technology, 2017,51(12):6936-6944.

[19] Li R, Mei X, Wei L, et al. Study on the contribution of transport to $PM_{2.5}$ in typical regions of China using the regional air quality model RAMS-CMAQ [J]. Atmospheric Environment, 2019,214:116856.

[20] 刘文军,郑国义,田 学,等.西安市 $PM_{2.5}$ 相关因素多元回归分析模型 [J]. 经济数学, 2015,32(1):85-88.
Liu W, Zheng G, Tian X. $PM_{2.5}$ factors associated with multivariate regression analysis model in Xi'an [J]. Journal of Quantitative Economics, 2015,32(1):85-88.

[21] 卢月明,王 亮,仇阿根,等.一种基于主成分分析的协同克里金插值方法 [J]. 测绘通报, 2017(11):51, 57-63.
Lu M, Wang L, Qiu A. A CoKriging interpolation method based on principal component analysis [J]. Bulletin of Surveying and Mapping, 2017(11):51, 57-63.

[22] Vaidyanathan A, Dimmick W F, Kegler S R, et al. Statistical air quality predictions for public health surveillance: evaluation and generation of county level metrics of $PM_{2.5}$ for the environmental public health tracking network [J]. International Journal of Health Geographics, 2013,12:12.

[23] Pozdnoukhov A, Kanevski M. Monitoring network optimisation for spatial data classification using support vector machines [J]. International Journal of Environment and Pollution, 2006,28(3/4): 465-484.

[24] Kaburlasos V G, Athanasiadis I N, Mitkas P A. Fuzzy lattice reasoning (FLR) classifier and its application for ambient ozone estimation [J]. International Journal of Approximate Reasoning, 2007,45(1):152-188.

[25] Sfetsos A, Vlachogiannis D. A new approach to discovering the causal relationship between meteorological patterns and PM_{10} exceedances [J]. Atmospheric Research, 2010,98(2-4):500-511.

[26] 中国环境监测总站.全国城市空气质量实时发布平台.
China National Environmental Monitoring Centre. National air quality real-time release platform.

[27] 中国气象局.中国地面气象站逐小时观测资料. http://data.cma.cn/.
China Meteorological Administration. Hourly observation data from stationary meteorological sites. http://data.cma.cn/.

[28] Geofabrik. Open street map. https://www.openstreetmap.org/.

[29] 中国科学院地理科学与资源研究所.资源环境科学与数据中心. https://www.resdc.cn/.
Institute of Geographic Sciences and Natural Resources Research. Resource and environment science and data center. https://www.resdc.cn/.

[30] Svetnik V, Liaw A, Tong C, et al. Random forest: A classification and regression tool for compound classification and QSAR modeling [J]. Journal of Chemical Information and Computer Sciences, 2003,43(6): 1947-1958.

[31] Cutler D R, Edwards T C, Beard K H, et al. Random forests for classification in ecology [J]. Ecology, 2007,88(11):2783-2792.

[32] Shao Y, Ma Z, Wang J, et al. Estimating daily ground-level $PM_{2.5}$ in China with random-forest-based spatiotemporal kriging [J]. Science of the Total Environment, 2020,740:139761.

[33] 李名升,任晓霞,于 洋,等.中国大陆城市 $PM_{2.5}$ 污染时空分布规律 [J]. 中国环境科学, 2016,36(3):641-650.
Li M, Ren X, Yu Y. Spatiotemporal pattern of ground-level fine particulate matter ($PM_{2.5}$) pollution in mainland China [J]. China Environmental Science, 2016,36(3):641-650.

[34] 许 珊,邹 滨,胡晨霞.面向场景的城市 $PM_{2.5}$ 浓度空间分布精细模拟 [J]. 中国环境科学, 2019,39(11):4570-4579.
Xu S, Zou B, Hu C. Urban scene-oriented simulation of the spatial distribution of $PM_{2.5}$concentration in an intra-urban area at fine scale [J]. China Environmental Science, 2019,39(11):4570-4579.

作者简介: 周 琪(1994-),女,江苏徐州人,博士后,主要从事环境健康和环境管理研究.发表论文 9 篇.