

春节期间的PM_{2.5}污染短期暴露健康效应评估* ——以长三角地区25个城市为例

The Evaluation of Health Effect of Short-term Exposure to PM_{2.5} during Spring Festival: A Case Study of 25 Cities in the Yangtze River Delta

戴劭勍 李佳佳 杨维旭 陈方煜 江辉仙 DAI Shaoqiang, LI Jiajia, YANG Weixu, CHEN Fangyu, JIANG Huixian

摘要 春节人口迁徙是中国一年一度在全国范围内的人口大型迁徙活动。基于空间全生命周期流行病学理论框架,集成多源时空数据和机器学习算法实现日尺度的PM_{2.5}高时空分辨率制图,结合腾讯位置大数据评估长三角区域人群PM_{2.5}污染短期暴露健康风险。结果如下:(1) 基于随机森林构建的PM_{2.5}高时空分辨率制图模型空间交叉验证结果R²达到0.8以上,具有良好的精度;(2) 春节人口迁徙行为使得全国范围内人口在短期内大规模流动,导致长三角地区PM_{2.5}短期暴露健康风险增加和减少的迁徙人口分别有6 070万人和6 175万人;(3) 春节人口迁徙行为导致的PM_{2.5}污染短期暴露健康风险危害不容忽视,增加风险均值在0.25—0.39,高值在0.9以上;(4) PM_{2.5}污染短期暴露健康效应具有强烈的时空异质性,两天内的健康风险变化极大(0.84, -0.75);(5) 在长三角地区的部分城市中,大部分迁徙人口的春节迁徙行为导致PM_{2.5}污染短期暴露健康风险增加,如上海和苏州,暴露人口约有1 320万人和600万人。以期为PM_{2.5}污染短期暴露健康效应估算提供实证分析,丰富了空间全生命周期流行病学的应用案例。

Abstract Human mobility during the Spring Festival is a large-scale annual human mobility across China. This paper evaluated the human health risks of short-term exposure to PM_{2.5} in the Yangtze River Delta through a combination of the daily-level high spatiotemporal mapping of PM_{2.5} based on integrated multi-sources spatiotemporal data and machine learning algorithm, and Tencent mobility data based on the theoretical framework of spatial life-course epidemiology. The results showed that: (1) the accuracy of the random forest model for PM_{2.5} spatiotemporal mapping by spatial cross-validation is more than 0.8. (2) Human mobility behavior enables large-scale mobility across the country in a short period during the Spring Festival, which leads to 60.7 million and 61.75 million populations who have increased and decreased the health risk of short-term exposure to PM_{2.5} respectively. (3) We should not ignore the health risks of short-term exposure to PM_{2.5} caused by human mobility behavior during the Spring Festival, because the estimated average increased risks are from 0.25 to 0.39. The highest risk is more than 0.9. (4) This reveals strong spatiotemporal heterogeneity in the health effect of short-term exposure to PM_{2.5}. Even the value of risks could change from 0.84 to -0.75 in two days. (5) Most of the travel populations in the Yangtze River Delta have increased the health risk of short-term exposure to PM_{2.5} caused by human mobility during the Spring Festival, such as Shanghai and Suzhou, with the exposed populations reaching to nearly 13.2 million and 6 million. This paper provides a practice case study to evaluate the health risk of short-term exposure to PM_{2.5} and enriches the application of spatial life-course epidemiology.

关键词 PM_{2.5}短期污染暴露;春节人口迁徙;时空制图;机器学习

Key words short-term exposure to PM_{2.5} pollution; human mobility during the Spring Festival; spatiotemporal mapping; machine learning

文章编号 1673-8985 (2020) 05-0022-08 中图分类号 TU984 文献标志码 A

DOI 10.11982/j.supr.20200504

作者简介

戴劭勍

荷兰特文特大学地理信息与对地观测学院

博士研究生

李佳佳

中国科学院城市环境研究所城市环境与健康重点实验室
硕士研究生

杨维旭

福建省同安第一中学 二级教师,硕士

陈方煜

福建师范大学地理科学学院

江辉仙 (通信作者)

福建师范大学地理科学学院

福建省陆地灾害监测评估工程技术研究中心
副教授,硕士生导师

0 引言

在过去的几十年中,随着人口的快速增长、工业化与城市化进程的快速推进,城市空气污染日益严重并逐渐演变为全球性公共卫生问题。其中以PM_{2.5}为代表的环境细颗粒物

* 基金项目:福建省自然科学基金项目“基于移动终端的大型公共建筑物智能消防疏散系统关键技术研究”(编号2018J01740)资助。

(fine particulate matter) 污染由于其对呼吸系统的损害^[1],与糖尿病^[2]、心血管疾病^[3]等慢性病患病率的高关联性,已经成为严重威胁公众健康的环境风险因子(对全球疾病负担的贡献率位列第6)^[4]。在改革开放40年的快速发展背景下,中国也面临着严峻的PM_{2.5}污染导致的公众健康问题。

PM_{2.5}污染对公众健康的影响研究需要估算人群的PM_{2.5}污染暴露健康效应。PM_{2.5}污染暴露健康效应的研究可以细分为PM_{2.5}污染长期与短期暴露的健康效应。PM_{2.5}污染长期暴露健康效应的研究核心在于估算由于PM_{2.5}污染长期暴露造成的死亡人数以及减少的平均预期寿命。如估算中国2013年因PM_{2.5}暴露导致的死亡人数为107万—137万人^[5-6],大约占到全部死亡人数的30%^[7],同时,PM_{2.5}污染也使中国人平均预期寿命下降了1.25岁^[8]。这类研究通常基于长时间跨度的大型人群队列或者纵向研究数据在宏观尺度上(区域、国家和全球尺度)的分析,由于人群队列研究与纵向研究的人群样本数据通常采用长期居住地或者工作地等主要活动地点作为暴露地点,因此这类研究可以较为准确地估算PM_{2.5}污染长期暴露的健康效应。相反,当前关于PM_{2.5}污染短期暴露的健康效应研究相对较少,PM_{2.5}污染短期暴露已在多项研究中被证实存在严重的疾病负担,如增加死亡率^[9]、急性下呼吸道感染风险^[10]、精神疾病急诊入院率^[11]、心脏系统疾病致死率^[12]及急性冠状动脉事件风险^[13]。此外,PM_{2.5}污染短期暴露给社会带来了严重的经济负担,Wei等^[14]利用超过9 500万人的美国医保记录与PM_{2.5}浓度分析,发现平均每1 μg/m³的PM_{2.5}污染短期暴露会造成每年增加2 050例住院病例以及3 100万美元住院费用。然而,PM_{2.5}污染具有强烈的时空异质性,加上暴露人群所处的不同社会经济环境与复杂的人类移动,造成PM_{2.5}污染短期暴露健康效应存在强烈的时空分异性^[15-17]。由于过去难以精准刻画短期大规模的时空动态人口与精细尺度的PM_{2.5}时空分布,如何准确估算大区域PM_{2.5}污染短期暴露导致的健康效应仍然存

在较大的研究空白。现有部分研究基于人群时空间行为数据(如个人时空行为轨迹)对微观层面上的污染短期暴露状况展开分析,如依靠居民出行的GPS数据与个人活动日志估算出不同交通方式的空气暴露^[16]与分析不同建成环境对弱势人群暴露效应的影响^[17],但这类研究受限于数据采集的难度,仍然无法在大区域开展。

估算大区域PM_{2.5}污染短期暴露健康效应需要一个全新的跨学科的健康地理研究范式,用以连接PM_{2.5}污染、暴露科学与时空间行为研究。空间全生命周期流行病学(spatial life-course epidemiology)是一门新兴的交叉学科,旨在利用先进的空间信息技术、定位技术与人工智能来研究生物因素、环境因素、行为因素和社会因素对个体健康风险的影响^[18]。空间全生命周期流行病学的核心在于将地理信息系统、遥感、全球定位系统及人工智能技术与多个尺度(从个体尺度到国家尺度)的健康追踪数据结合,对个体全生命周期内的健康风险实现精准估算。这为本研究提供了一个理论框架。当前,先进的地理大数据、对地观测数据与人工智能算法为本文提供了良好的研究基础。基于位置服务的互联网大数据为大区域人群时空间行为的精准测度提供了可靠的数据基础。这类数据已应用至时空间行为的刻画与研究,如人口短期流动模式研究^[19-21],人口迁徙与城市化的关系^[22]以及新型冠状病毒大流行中潜在感染者识别^[23-24]。此外,当前年、季节和月尺度的PM_{2.5}公里制图研究精度较高,在日尺度和小时尺度上精度也有所突破。大部分研究利用时空统计方法(如地理加权回归模型^[25]、时空随机森林^[26]、随机森林回归克里格插值^[27-28]等)、土地利用回归模型^[29]或者机器学习模型(如地理—长短期记忆神经网络^[29]等)融合包括化学传输模式输出(如GEOS-Chem^[30]、WRF-Chem^[31]等)、高时空分辨率遥感反演光学气溶胶厚度(Aerosol Optical Depth即AOD,如MODIS MAIAC^[32],Himawari-8^[33]等)、地面监测站点等多源时空数据实现高时空分辨率PM_{2.5}制图,空间分辨率通常为

0.01°。^[27]或1 km^[26],时间分辨率通常以年^[34]、季节^[35]、月^[36]居多,也有部分日尺度和小时制图结果^[28-30]。

在空间全生命周期流行病学的理论框架下,本文拟利用机器学习等人工智能算法实现PM_{2.5}的大区域高时空分辨率制图,结合腾讯位置大数据,以长三角城市群为例,针对中国一年一度的春节人口大迁徙导致人群PM_{2.5}污染短期暴露环境产生的变化,估算其造成的健康效应变化程度,探究不同城市PM_{2.5}污染短期暴露健康效应变化差异。本文旨在通过实证研究探讨春节人群迁徙行为影响下PM_{2.5}污染短期暴露环境及其健康效应的时空变化,从而丰富空间全生命周期流行病学的实践应用,实现其在空气污染短期暴露健康效应估算领域的研究创新。

1 数据与方法

1.1 数据

1.1.1 腾讯位置大数据

本文使用的腾讯位置大数据来自国家科技基础条件平台—国家地球系统科学数据共享服务平台—长江三角洲科学数据中心(<http://nnu.geodata.cn:8008/>),原始数据来自腾讯位置大数据平台(<https://heat.qq.com/qianxi.php>),通过编写网络挖掘软件获取2017年春节前后各一周(1月23日—2月5日)长三角地区25个主要城市的人口迁徙数据。长三角地区25个主要城市如图1所示。该数据主要包括各个城市人口迁入迁出量排名前10的城市对应的迁入迁出人口数值。

1.1.2 PM_{2.5}站点监测数据与网格产品

本文使用的PM_{2.5}数据包括PM_{2.5}站点监测数据与网格产品。PM_{2.5}站点监测数据由全国1 500多个国家级空气监测站点(<https://quotsoft.net/air/>)获取,包括几种主要的空气污染物(PM_{2.5}、PM₁₀、NO₂、SO₂、O₃、CO和AQI)逐小时浓度和数值。年均PM_{2.5}浓度网格产品(China Regional Estimates,以下简称“CRE Grid产品”,http://fizz.phys.dal.ca/~atmos/martin/?page_id=140#)^[34-36],由加拿大

戴尔豪斯大学 (Dalhousie University) 的大气成分分析小组 (Atmospheric Composition Analysis Group) 基于地理加权回归算法融合卫星观测数据、化学传输模式以及地面监测站点等多源数据得到, 以NetCDF格式存储, 空间分辨率为 $0.01^\circ \times 0.01^\circ$ 。

1.1.3 气象站点监测数据与数字高程模型

本文使用的气象数据来自美国国家气候数据中心。主要气象变量包括:气温、气压、露点温度、风速、降水量。本文使用的数字高程模型 (Digital Elevation Model, DEM) 为美国国家海洋和大气管理局项目GLOBE Topography提供的1 km DEM (<https://www.ngdc.noaa.gov/mgg/topo/DATATILES/elev/all10g.zip>)。

1.2 研究方法

在空间全生命周期流行病学的理论框架下, 本文首先集成了多源时空数据, 其次利用机器学习算法实现全国的1 km尺度PM_{2.5}时空制图。结合腾讯位置大数据估算对应的健康暴露效应, 评估长三角不同城市PM_{2.5}污染短期暴露健康风险的时空差异。

1.2.1 PM_{2.5}时空制图

本文数据涉及的春节迁徙时间段为2017年1月23日—2月5日, 考虑到模型预热等原因, 模型运行时间段设为2017年1月1日—2月14日。数据预处理包括如下操作:首先对该时间段内的PM_{2.5}站点数据进行清洗和聚合处理, 去除站点中的空值, 并将小时尺度监测数据聚合为日尺度, 最终可用于建模的数据共为1 348个站点。其次, 对气象站点数据进行清洗和聚合处理, 去除站点空值。根据气象要素不同, 采用不同方式将小时尺度的气象要素聚合成日尺度, 最终得到408个气象站点。采用薄盘样条插值方法^[39]以X坐标、Y坐标和高程作为协变量进行空间插值得到逐日全国1 km气象要素栅格。本文所有空间数据坐标系均设置为Albers WGS 1984投影坐标系。在这些输入数据的基础上, 本文假定t时刻第*i*个像元的PM_{2.5}浓度符

合如下关系:

$$PM25_{i,t} = f(PM25_{i,a}, H_i, Temp_{i,t}, DP_{i,t}, Press_{i,t}, Precip_{i,t}, Wind_{i,t}, doy) \quad (1)$$

其中, $PM25_{i,t}$ 代表t时刻第*i*个像元的预测PM_{2.5}浓度 (单位: $\mu g/m^3$), $f(\dots)$ 代表拟合的机器学习时空映射模型, 本文选用随机森林^[40]与极端梯度提升树算法^[41]拟合模型。 $PM25_{i,a}$ 代表第*i*个像元上CRE Grid产品2017年的PM_{2.5}浓度 (单位: $\mu g/m^3$), H_i 代表第*i*个像元上DEM的高程值, $Temp_{i,t}$ 代表t时刻第*i*个像元的气温 (单位: $^\circ C$), $DP_{i,t}$ 代表t时刻第*i*个像元的露点温度 (单位: $^\circ C$), $Press_{i,t}$ 代表t时刻第*i*个像元的气压 (单位: Hpa), $Precip_{i,t}$ 代表t时刻第*i*个像元的降雨量 (单位: mm), $Wind_{i,t}$ 代表t时刻第*i*个像元的风速 (单位: m/s), doy 代表t时刻是一年中的第几天。

为了比较不同模型的精度, 本文采用五折空间交叉验证^[42]。该验证方法根据站点数据的空间自相关性将所有站点数据划分为多个网格, 这些网格每次被抽取一部分标记为训练数据, 其余的为验证数据, 每次基于训练数据拟合机器学习模型, 用验证数据验证精度, 重复5次完成空间交叉验证与模型训练验证过程, 最终计算出不同模型的预测精度, 确定较优机器学习模型作为最后制图模型。本文选用均方根误差 (Root Mean Square Error, RMSE)、归一化均方根误差 (normalized Root Mean Square Error, nRMSE), 以及预测值与验证值回归方程的R²作为精度评估的指标。

$$RMSE = \sqrt{\frac{\sum_{j=1}^n (PM25_{k,t} - PM25_{j,t})^2}{n}} \quad (2)$$

$$nRMSE = \frac{RMSE}{PM25_{k,t}} \quad (3)$$

其中, RMSE和nRMSE分别为均方根误差与归一化均方根误差, $PM25_{k,t}$ 和 $PM25_{j,t}$ 分别为t时刻第*k*个像元上预测的PM_{2.5}浓度和第*j*个监测站点的PM_{2.5}浓度 (单位: $\mu g/m^3$), 其

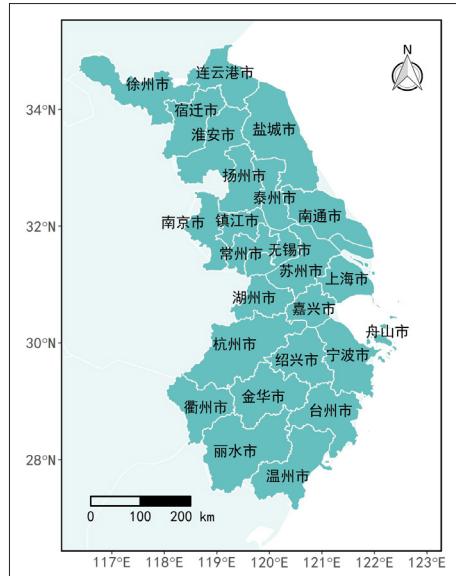


图1 研究区概况

Fig.1 Study area

资料来源:笔者自绘。

中*k*和*j*是一一对应的。*n*为用于验证的监测站点数量。 $(PM25_{k,t})$ 是t时刻所有监测站点的平均PM_{2.5}浓度值 (单位: $\mu g/m^3$)。

选定最优算法后将所有站点数据作为输入拟合模型, 最后将研究时间段内的模型自变量结果输入模型中, 得到全国日尺度PM_{2.5}的1 km制图结果。

1.2.2 PM_{2.5}短期暴露健康效应估算

本文根据一项以272个中国城市为研究单元的PM_{2.5}暴露与日死亡率研究^[43-44]所提供的PM_{2.5}暴露关联死亡率估算PM_{2.5}短期暴露导致的健康效应。选用人口死亡归因分数 (Attribute Fraction, AF) 作为短期暴露健康效应的健康风险指标。该指标随空气污染暴露程度变化而变化, 用于估算因春运迁徙造成人群空气污染暴露环境变化导致的特定原因死亡率。该指标包括总死亡、心血管疾病、高血压疾病、冠心病、中风、呼吸道疾病和慢性阻塞性肺疾病的特定原因死亡率。

$$AF = 1 - e^{-\beta \Delta C} \quad (4)$$

式中, β 是暴露响应函数的因果系数, 取值见表1, 代表每增加 $10 \mu g/m^3$ 关联的死亡

率。 ΔC 是由于春运迁徙导致的人群PM_{2.5}污染短期暴露环境变化程度,定义为目的地城市空气污染浓度与出发地城市空气污染浓度之差的绝对值。根据这个差值的正负性区分健康风险的增减。

2 研究结果

2.1 长三角区域人口春节迁徙行为

通过25个城市在春节期间的迁入迁出人口数据对比,发现不同城市净迁徙人口在时间尺度上的变化趋势存在一些相似的模式(净迁徙人口为每个城市每日迁入人口与迁出人口的差)(见图2)。如果净迁徙人口为负值,说明迁出人口多于迁入人口,反之则说明迁出人口少于迁入人口。如果净迁徙人口接近0值,说明迁出迁入人口数量大致相同,差异不显著。25个城市根据净迁徙人口可以分为3类:第一类包括上海、苏州、南京、杭州和宁波5个长三角地区经济较为发达的核心城市,在春节前以迁出为主,春节后以迁入为主;第二类包括宿迁、徐州和盐城,这几个城市劳动力输出较多,经济发展水平较低,春节前以迁入为主,春节后以迁出为主;第三类为其余17个城市,净迁徙人口整体接近于0,春节前后迁入迁出变化较小。

从迁徙城市的分布来看,2017年春节期间与长三角地区25个城市具有迁入迁出联系的城市共有102个,其中包括长三角地区25个城市的区域内迁徙和77个非长三角地区城市的跨区域迁徙。在77个非长三角地区城市中,人口迁入长三角的有70个,人口迁出长三角的有56个,既有迁入又有迁出的有49个。其中,迁入迁出均有的49个城市中,与长三角联系最多的分别是重庆、北京、贵阳、深圳、广州和成都(以上6个城市与长三角25个城市在春节前后的联系数量均超过100)。

从迁徙城市的分布来看,2017年春节期间与长三角地区25个城市具有迁入迁出联系的城市共有102个,其中包括长三角地区25个城市的区域内迁徙和77个非长三角地区城市的跨区域迁徙。在77个非长三角地区城市中,人口迁入长三角的有70个,人口迁出长三角的有56个,既有迁入又有迁出的有49个。其中,迁入迁出均有的49个城市中,与长三角联系最多的分别是重庆、北京、贵阳、深圳、广州和成都(以上6个城市与长三角25个城市在春节前后的联系数量均超过100)。

nRMSE值,预测散点图分布在1:1线两侧的均匀程度,随机森林的预测精度要优于极端梯度提升树。

根据空间交叉验证结果,本文选用随机森林作为制图模型进行制图(见图4)。由于长三角区域城市迁徙涉及城市包含非长三角地区,故PM_{2.5}制图的范围以全国为准。全国区域春节期间PM_{2.5}污染浓度值约为100 μg/m³,污染较为严重的区域主要是京津冀区域、河南省、陕西省和山西省南部、湖北省和湖南省交界处以及四川省和重庆市的交界处。从时空分布来看,2017年1月23日整体空气污染较轻,随时间变化全国整体污染程度加重,京津冀区域有明显的高污染热点出现;1月27日整体区域污染程度降低;1月28日污染程度再次加重,并且有两个较为明显的污染热点区域(京津冀区域,湖北省和湖南省交界处);1月29日—2月1日,全国的空气污染程度较轻;2月2日起,全国整体污染程度有所加重,污染热点区域仍然为京津冀地区。

2.3 PM_{2.5}污染短期暴露健康风险

为了区分AF增减,将目的地与出发地的PM_{2.5}浓度差值的正负性与AF相乘作为最后参考的AF值。当AF大于0时,说明目的地PM_{2.5}浓度大于出发地,PM_{2.5}污染短期暴露的健康风险增加;当AF小于0时,说明PM_{2.5}污染短期暴露的健

表1 PM_{2.5}暴露响应函数的因果系数与95%置信区间
Tab.1 Cause-specific coefficients (β) and 95% confidence intervals of concentration-response functions for daily PM_{2.5}

死亡原因	β
总死亡	0.22(0.15,0.28)
心血管疾病	0.27(0.18,0.36)
高血压疾病	0.39(0.13,0.65)
冠心病	0.30(0.19,0.40)
中风	0.23(0.13,0.34)
呼吸道疾病	0.29(0.17,0.42)
慢性阻塞性肺疾病	0.38(0.23,0.53)

资料来源:笔者自制。

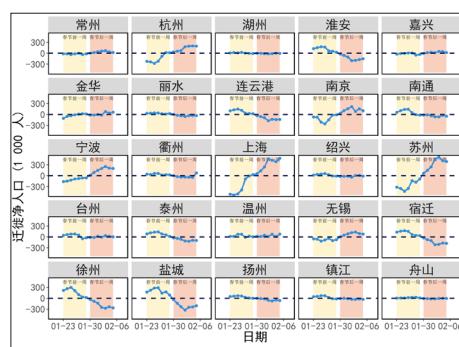


图2 长三角地区25个城市春节期间净迁徙人口

Fig.2 Net migration populations of 25 cities in the Yangtze River Delta region during the Spring Festival

资料来源:笔者自绘。

2.2 PM_{2.5}时空制图

空间交叉验证结果如图3所示,a为随机森林,b为极端梯度提升树,绿色线为1:1线,蓝色线为观测值与预测值拟合的线性方程。两种算法拟合的R²均超过0.75, RMSE值均在45 μg/m³左右,nRMSE均在0.6左右。整体的散点分布有较明显的线性趋势,散点图回归直线与1:1的直线较为接近。根据R²、RMSE和

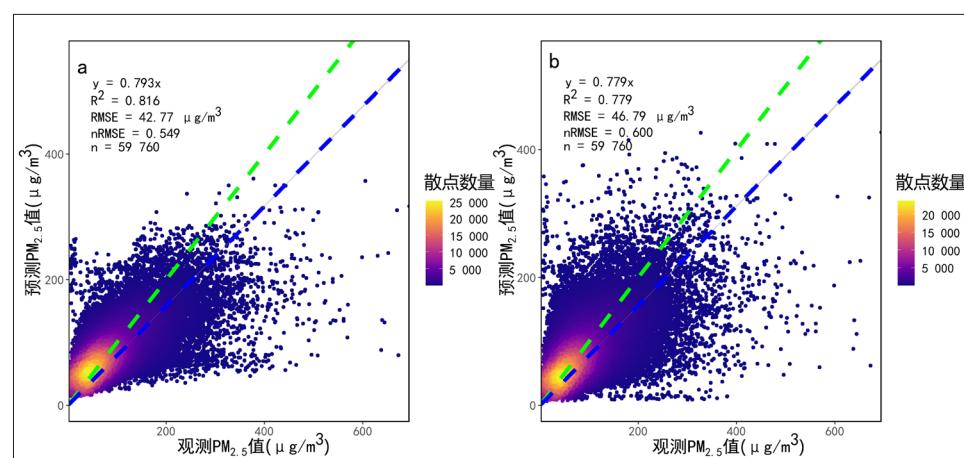
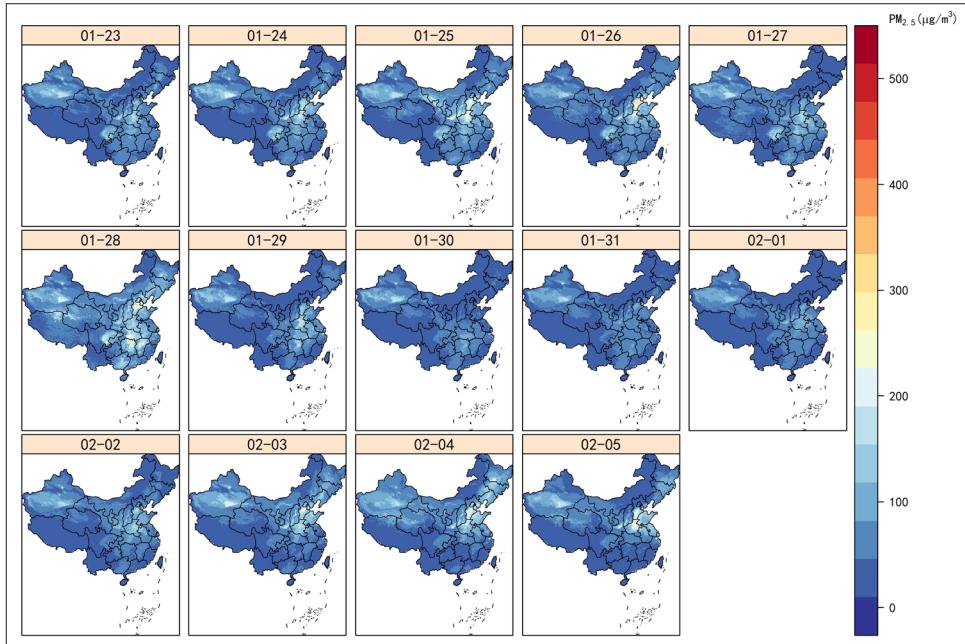


图3 不同机器学习模型空间交叉验证结果比较

Fig.3 Comparison between different machine learning models via spatial cross validation

资料来源:笔者自绘。

图4 日尺度PM_{2.5}公里制图结果Fig.4 Daily level PM_{2.5} mapping at 1-km resolution

资料来源：笔者自绘。

表2 PM_{2.5}暴露导致的AF平均值及置信区间Tab.2 The average value and 95% confidence intervals of AF causing by PM_{2.5} exposure

死亡原因	$\overline{AF}(+)(n=2858)$	$\overline{AF}(-)(n=2914)$
总死亡	0.28(0.21, 0.33)	0.28(0.21, 0.33)
心血管疾病	0.32(0.24, 0.38)	0.33(0.25, 0.39)
高血压疾病	0.40(0.19, 0.53)	0.41(0.19, 0.54)
冠心病	0.34(0.24, 0.41)	0.35(0.26, 0.42)
中风	0.29(0.19, 0.37)	0.29(0.19, 0.38)
呼吸道疾病	0.34(0.23, 0.42)	0.34(0.24, 0.43)
慢性阻塞性肺疾病	0.40(0.29, 0.48)	0.40(0.29, 0.49)

资料来源：笔者自制。

康风险减少，其中AF绝对值越接近于1，说明健康风险越大。根据春节人口迁徙的5 772对城市迁徙数据进行PM_{2.5}污染暴露差异计算，得到对应暴露的健康风险。共有2 858对城市迁徙行为，6 070万迁徙人口的PM_{2.5}污染短期暴露健康风险增加；2 914对城市迁徙行为，6 175万迁徙人口的PM_{2.5}污染短期暴露健康风险减少；长三角地区城市春节迁徙人群有接近一半因迁徙导致PM_{2.5}暴露环境变化增加了健康风险，同时剩余的超过一半的迁徙人群健康风险减少。详细的各类特定死亡原因的AF平均值与置信区间见表2，可以看到增加和减少的人群的总死亡率、中风、

呼吸道疾病与慢性阻塞性肺疾病的AF平均值相同，而风险增加人群的心血管疾病、高血压疾病与冠心病的AF平均值略低于风险减少人群（均相差0.1）。

对长三角城市区域内25个主要城市春节期间的迁徙人口按照健康风险变化（增加/减少）与出行属性（出发地/目的地）进行统计与可视化（见图5）。25个城市按照人口迁徙—健康风险变化模式可以分为3大类。第一类城市“风险增加—出发地”与“风险减少—目的地”迁徙人口明显大于“风险增加—目的地”与“风险减少—出发地”。这类城市共有7个（无锡、温州、

台州、苏州、上海、宁波、南京），其中以上海和苏州为典型城市，这两个城市“风险增加—出发地”与“风险减少—目的地”迁徙人口均位列第一、二位，分别为1 300万人与600万人左右，是其他模式迁徙人口的3倍。这类城市大部分迁徙人口在春节期间迁徙目的地的PM_{2.5}污染程度比该城市大，春节迁徙行为导致PM_{2.5}污染短期暴露健康风险增加。第二类城市“风险增加—出发地”与“风险减少—目的地”迁徙人口明显小于“风险增加—目的地”与“风险减少—出发地”。这类城市共有12个（镇江、扬州、盐城、徐州、宿迁、泰州、衢州、连云港、金华、淮安、杭州、常州），其中以杭州和徐州为典型城市。杭州为“风险增加—目的地”与“风险减少—出发地”迁徙人口位列第一的城市，迁徙人口分别达到了470万与432万人，接近其他模式迁徙人口的2倍。这类城市大部分迁徙人口在春节期间迁徙目的地的PM_{2.5}污染程度比该城市小，春节迁徙行为导致PM_{2.5}污染短期暴露健康风险减少。第三类城市在4种人口迁徙—健康风险变化模式中的迁徙人口差距较小。这类城市共有6个（舟山、绍兴、南通、丽水、嘉兴、湖州），其中以绍兴为典型



图5 长三角地区25个城市春节迁徙的人口迁徙—健康风险变化模式

Fig.5 The changing mode of human mobility-health risk in 25 cities of the Yangtze River Delta during the Spring Festival

资料来源：笔者自绘。

城市,绍兴在4种模式下的迁徙人口均为100万人左右。这类城市一半的迁徙人口在春节期间迁徙目的地的PM_{2.5}污染程度比该城市大,春节迁徙行为导致PM_{2.5}污染短期暴露健康风险增加,其余迁徙人口则相反。

为了进一步研究更具体的健康风险变化,根据数据完整性(即每对迁徙城市在研究时间段均发生迁徙行为)筛选出2 408对城市迁徙数据,经过处理得到172对城市14天的迁徙数据以及对应的健康风险。其中以长三角城市为出发地的城市迁徙对共31对,上海、宁波和温州是联系最多的3个城市,上海有5个,宁波和温州各3个;以长三角城市为目的地的城市迁徙对共141对,杭州、南京和苏州是联系最多的3个城市,分别有12个、11个和10个,上海涉及9个位居第4。本文

主要分析联系最多的这6个城市(上海、宁波、温州、杭州、南京和苏州),健康风险指标选用总死亡率进行表征。对AF的变化值进行可视化,结果如图6所示,其中a和b分别代表长三角地区城市作为出发地和目的地的时间变化矩阵,c和d分别代表长三角地区城市作为出发地和目的地的矩阵(origin-destination, OD矩阵)。

时间变化矩阵显示以长三角地区城市为出发地的迁徙行为造成的PM_{2.5}污染短期暴露风险大部分都是增加的,同时有大部分迁徙城市在2月5日的PM_{2.5}污染短期暴露风险是减少的,其原因可能是2月5日已经处于春节复工后,长三角区域经济生产活动大幅增加,PM_{2.5}污染加重。而以长三角地区城市为目的地的迁徙行为造成的PM_{2.5}污染短期暴露健康风险大部分都是减少

的。这说明长三角地区春节期间整体空气污染程度比其他区域轻。此外,PM_{2.5}污染程度具有强烈的时空异质性,会使得部分城市迁徙行为前后两天健康风险差异巨大。如上海到北京的迁徙行为的PM_{2.5}污染短期暴露健康风险在两天内(2月4日、2月5日)从0.84变化为-0.75。以长三角地区为出发地和目的地的迁徙行为造成的PM_{2.5}短期暴露健康风险增加值最高分别可达到0.96和0.75,减少值最高分别可达到0.78和0.96,均值分别为0.25与0.39,对公众而言这是不容忽视的健康风险。

3 讨论与结论

3.1 讨论

3.1.1 PM_{2.5}时空制图

本文基于随机森林和极端梯度提升树构建了日尺度全国PM_{2.5}公里制图模型。其中采用空间交叉验证方法比较精度,与既有研究存在较大差别。大部分PM_{2.5}时空制图研究精度评估采用的交叉验证方法不考虑空间自相关性,如10折交叉验证,在多个日尺度PM_{2.5}制图中被采用^{[37]7, [45]56, [46]11989, [47]133}。但是PM_{2.5}站点分布具有显著的空间自相关性,普通的交叉验证可能会导致模型过拟合,在区域制图时可能会导致极大的不确定性^[48]。这种情况下采用空间交叉验证方法可以有效地评估模型精度,避免由于训练数据的空间自相关性带来的模型过拟合问题。从理论上来说,空间交叉验证的精度是低于普通交叉验证的。因此,采用空间交叉验证避免模型过拟合可能是本研究模型精度中RMSE和nRMSE数值不如其他同类型研究的部分原因。由于验证方式不同,本文结果与其他日尺度PM_{2.5}制图的精度对比仅能作为制图精度评估的参考,不能从根本上代表模型的优劣与制图的准确度。与现有研究相比,无论是R²还是RMSE^{[37]7, [45]56, [46]11989, [47]133},本文的PM_{2.5}时空制图结果都具有足够的可信度。

3.1.2 PM_{2.5}污染短期暴露的健康效应

从本文研究结果来看,春节迁徙行为造成长三角地区PM_{2.5}污染短期暴露健康风险增加和减少比例几乎持平,但健康风险不容忽视。长三角地区有一半的城市人群因春节迁徙导致的

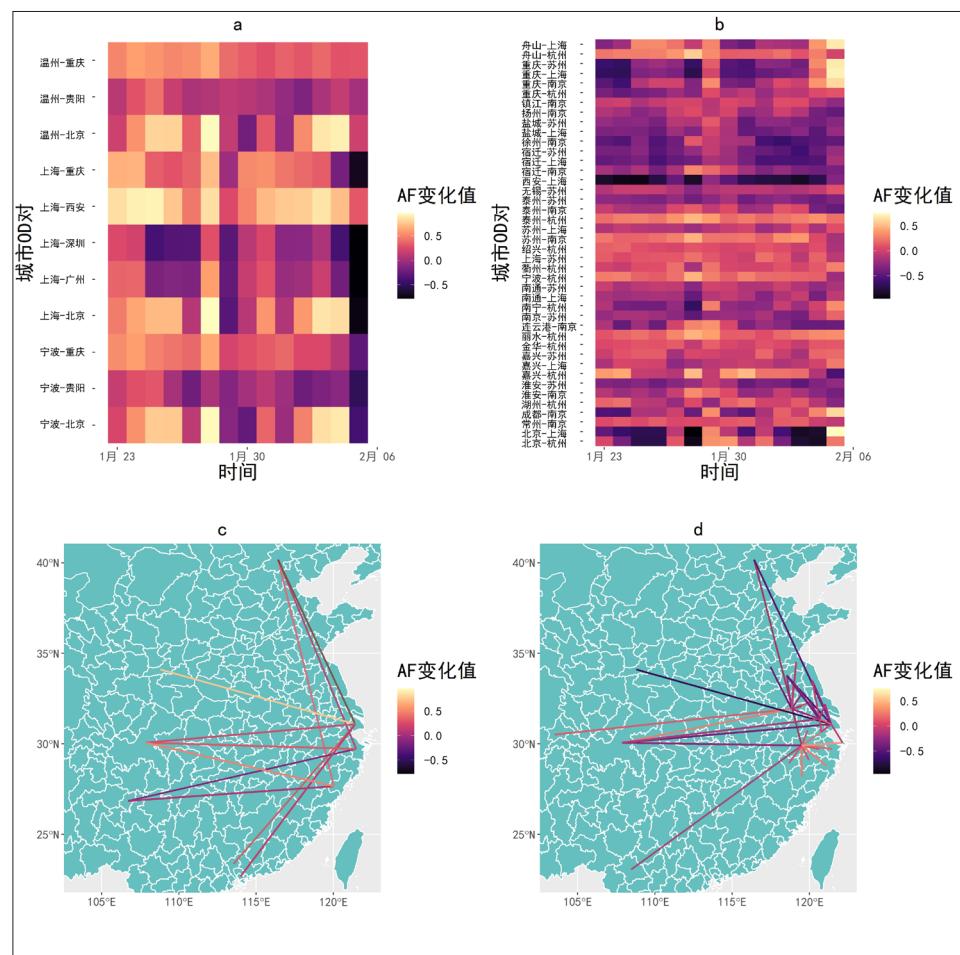


图6 AF时间变化矩阵与OD矩阵

Fig.6 The time variation and OD matrices of AF

资料来源:笔者自绘。

PM_{2.5}污染短期暴露健康风险是减少的,另一半则是增加的。(本文采用集成多源时空数据的高时空分辨率PM_{2.5}制图评估短期暴露的健康效应与Kloog等^[49]采用的基于卫星的新型暴露评估模型有相似之处。)尽管Shi等^[50]的研究从城市化与移民角度分析城乡人口的PM_{2.5}暴露差异不大,但本文表明PM_{2.5}污染短期暴露健康效应是随时间与空间变化的,主要原因是PM_{2.5}具有强烈的时空异质性^{[15]575}。本文在空间全生命周期流行病学的理论框架下,分析具有中国特色的春节迁徙行为造成的PM_{2.5}污染短期暴露健康效应,在应用层面丰富了宏观区域PM_{2.5}污染短期暴露健康效应的研究。同时,集成多源时空数据的高时空分辨率制图实现PM_{2.5}环境暴露的精准测度是对空间全生命周期流行病学10大热点问题中的精准测度环境暴露指标的一次实证研究探索^[51],也丰富了空间全生命周期流行病学在空气污染暴露领域的实际应用。

3.2 结论

本文基于空间全生命周期流行病学理论框架,以2017年春节迁徙期间长三角地区25个城市为例,集成多源时空数据实现高时空分辨率PM_{2.5}制图,结合腾讯位置大数据估算PM_{2.5}污染短期暴露的健康效应。从本文研究结果来看,春节迁徙行为导致长三角地区PM_{2.5}污染短期暴露健康风险增加和减少人数分别为6 070万人和6 175万人。健康风险增加值最高值在0.78与0.96之间,均值在0.25与0.39之间。PM_{2.5}污染短期暴露健康效应的时空异质性更强烈,前一天与后一天的暴露健康风险变化值可以达到1.5以上。此外,长三角地区的7个城市大部分迁徙人口由于春节迁徙行为导致PM_{2.5}污染短期暴露健康风险增加,包括无锡、温州、台州、苏州、上海、宁波、南京,暴露人口从143万人到1 320万人不等。此外,6个城市的半迁徙人口由于春节迁徙行为导致PM_{2.5}污染短期暴露健康风险增加,包括舟山、绍兴、南通、丽水、嘉兴、湖州,暴露人口从30万人到155万人不等。对于这些城市的迁徙人口应当加强对迁徙人群PM_{2.5}污染暴露相关疾病的的关注。本文仍然存在一些不足,如腾讯人口迁徙数据获取不

够完整,仅有每个城市迁入迁出前10城市的数据,PM_{2.5}时空制图的精度还可以进一步提升等。这些都有待于进一步分析与探究。■

(致谢:感谢国家科技基础条件平台—国家地球系统科学数据共享服务平台—长江三角洲科学数据中心(<http://nnu.geodata.cn:8008/>)提供数据支撑。)

参考文献 References

- XING Y F, XU Y H, SHI M H, et al. The impact of PM_{2.5} on the human respiratory system[J]. Journal of Thoracic Disease, 2016, 8(1): 69-74.
- BOWE B, XIE Y, LI T, et al. The 2016 global and national burden of diabetes mellitus attributable to PM_{2.5} air pollution[J]. The Lancet Planetary Health, 2018, 2(7): 301-312.
- DELFINO R J, SIOUTAS C, MALIK S. Potential role of ultrafine particles in associations between airborne particle mass and cardiovascular health[J]. Environmental Health Perspectives, 2005, 113(8): 934-946.
- FOROUZANFAR M H, AFSHIN A, ALEXANDER L T, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015[J]. The Lancet, 2016, 388(10053): 1659-1724.
- LIU J, HAN Y, TANG X, et al. Estimating adult mortality attributable to PM_{2.5} exposure in China with assimilated PM_{2.5} concentrations based on a ground monitoring network[J]. Science of the Total Environment, 2016, 568: 1253-1262.
- TIAN L, ZENG Q, DONG W, et al. Addressing the source contribution of PM_{2.5} on mortality: an evaluation study of its impacts on excess mortality in China[J]. Environmental Research Letters, 2017, 12(10): 104016.
- SONG C, HE J, WU L, et al. Health burden attributable to ambient PM_{2.5} in China[J]. Environmental Pollution, 2017, 223: 575-586.
- APTE J S, BRAUER M, COHEN A J, et al. Ambient PM_{2.5} reduces global and regional life expectancy[J]. Environmental Science & Technology Letters, 2018, 5(9): 546-551.
- LU F, XU D, CHENG Y, et al. Systematic review and meta-analysis of the adverse health effects of ambient PM_{2.5} and PM₁₀ pollution in the Chinese population[J]. Environmental Research, 2015, 136: 196-204.
- HORNE B D, JOY E A, HOFMANN M G, et al. Short-term elevation of fine particulate matter air pollution and acute lower respiratory infection[J]. American Journal of Respiratory and Critical Care Medicine, 2018, 198(6): 759-766.
- LEE S, LEE W, KIM D, et al. Short-term PM_{2.5} exposure and emergency hospital admissions for mental disease[J]. Environmental Research, 2019, 171: 313-320.
- MAT T, GUAITA R, PICHUELE M, et al. Short-term effect of fine particulate matter (PM_{2.5}) on daily mortality due to diseases of the circulatory system in Madrid (Spain)[J]. Science of the Total Environment, 2010, 408(23): 5750-5757.
- WILLIAMS A M, PHANEUF D J, BARRETT M A, et al. Short-term impact of PM_{2.5} on contemporaneous asthma medication use: behavior and the value of pollution reductions[J]. Proceedings of the National Academy of Sciences, 2019, 116(12): 5246-5253.
- WEI Y, WANG Y, DI Q, et al. Short term exposure to fine particulate matter and hospital admission risks and costs in the Medicare population: time stratified, case crossover study[J]. BMJ, 2019, 367: l6258.
- COLMER J, HARDMAN I, SHIMSHACK J, et al. Disparities in PM_{2.5} air pollution in the United States[J]. Science, 2020, 369(6503): 575-578.
- 郭文伯,张艳,柴彦威.城市居民出行的空气污染暴露测度及其影响机制——北京市郊区社区的案例分析[J].地理研究, 2015, 34 (7) : 1310-1318.
- GUO Wenbo, ZHANG Yan, CHAI Yanwei. Measurement of residents' daily travel air pollution exposure and its mechanism: a case study of suburban communities in Beijing[J]. Geographical Research, 2015, 34(7): 1310-1318.
- MA X, LI X, KWAN M P, et al. Who could not avoid exposure to high levels of residence-based pollution by daily mobility? Evidence of air pollution exposure from the perspective of the Neighborhood Effect Averaging Problem (NEAP)[J]. International Journal of Environmental Research and Public Health, 2020, 17(4): 1223.
- JIA P. Spatial life-course epidemiology[J]. The Lancet Planetary Health, 2019, 3(2): 57-59.
- 潘竟,赖建.中国城市间人口流动空间格局的网络分析——以国庆—中秋长假和腾讯迁徙数据为例[J].地理研究, 2019, 38 (7) : 1678-1693.
- PAN Jing, LAI Jian. A network analysis of the spatial pattern of population flow among cities in China: taking the National Day - Mid-autumn Festival and Tencent's migration data as examples[J]. Geographical Research, 2019, 38(7): 1678-1693.
- 周素红,张琳,林荣平.地理环境暴露与公众健康研究进展[J].科技导报, 2020, 38 (7) : 43-52.
- ZHOU Suhong, ZHANG Lin, LIN Rongping. Progress and prospect of the research on geographical

- environment exposure and public health[J]. *Science & Technology Review*, 2020, 38(7): 43-52.
- [21] 王楠, 王会, 杜云, 等. 青藏高原人口流入流出时空模式研究[J]. *地理学报*, 2020, 75 (7) : 1418-1431.
- WANG Nan, WANG Hui, DU Yun, et al. Research on the temporal and spatial patterns of population inflow and outflow in the Qinghai-Tibet Plateau[J]. *Acta Geographica Sinica*, 2020, 75(7): 1418-1431.
- [22] 赵梓. 基于大数据的中国人口迁徙空间格局及其对城镇化影响研究[D]. 长春: 吉林大学, 2018.
- ZHAO Zi. Research on the spatial pattern of Chinese population migration and its impact on urbanization based on big data[D]. Changchun: Jilin University, 2018.
- [23] 刘张, 千家乐, 杜云艳, 等. 基于多源时空大数据的区际迁徙人群多层次空间分布估算模型——以COVID-19疫情期间自武汉迁出人群为例[J]. *地球信息科学学报*, 2020, 22 (2) : 147-160.
- LIU Zhang, QIAN Jiale, DU Yunyan, et al. Multi-level spatial distribution estimation model of the inter-regional migrant population using multi-source spatio-temporal big data: a case study of migrants from Wuhan during the spread of COVID-19[J]. *Journal of Geo-information Science*, 2020, 22(2): 147-160.
- [24] YANG Z, ZENG Z, WANG K, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions[J]. *Journal of Thoracic Disease*, 2020, 12(3): 165-174.
- [25] SONG W, JIA H, HUANG J, et al. A satellite-based geographically weighted regression model for regional PM_{2.5} estimation over the Pearl River Delta region in China[J]. *Remote Sensing of Environment*, 2014, 154: 1-7.
- [26] WEI J, HUANG W, LI Z, et al. Estimating 1-km-resolution PM_{2.5} concentrations across China using the space-time random forest approach[J]. *Remote Sensing of Environment*, 2019, 231: 111221.
- [27] LIU Y, CAO G, ZHAO N, et al. Improve ground-level PM_{2.5} concentration mapping using a random forests-based geostatistical approach[J]. *Environmental Pollution*, 2018, 235: 272-282.
- [28] STAFOGGIA M, BELLANDER T, BUCCI S, et al. Estimation of daily PM₁₀ and PM_{2.5} concentrations in Italy, 2013-2015, using a spatiotemporal land-use random-forest model[J]. *Environment International*, 2019, 124: 170-179.
- [29] MA J, DING Y, CHENG J C P, et al. A temporal-spatial interpolation and extrapolation method based on geographic Long Short-Term Memory neural network for PM_{2.5}[J]. *Journal of Cleaner Production*, 2019, 237: 117729.
- [30] LEE H M, PARK R J, HENZE D K, et al. PM_{2.5} source attribution for Seoul in May from 2009 to 2013 using GEOS-Chem and its adjoint model[J]. *Environmental Pollution*, 2017, 221: 377-384.
- [31] SAIDE P E, CARMICHAEL G R, SPAK S N, et al. Forecasting urban PM₁₀ and PM_{2.5} pollution episodes in very stable nocturnal conditions and complex terrain using WRF-Chem CO tracer model[J]. *Atmospheric Environment*, 2011, 45(16): 2769-2780.
- [32] HE Q, GU Y, ZHANG M. Spatiotemporal trends of PM_{2.5} concentrations in central China from 2003 to 2018 based on MAIAC-derived high-resolution data[J]. *Environment International*, 2020, 137: 105536.
- [33] CHEN J, YIN J, ZANG L, et al. Stacking machine learning model for estimating hourly PM_{2.5} in China based on Himawari 8 aerosol optical depth data[J]. *Science of the Total Environment*, 2019, 697: 134021.
- [34] VAN DONKELAAR A, MARTIN R V, LI C, et al. Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors[J]. *Environmental Science & Technology*, 2019, 53(5): 2595-2611.
- [35] LI R, MA T, XU Q, et al. Using MAIAC AOD to verify the PM_{2.5} spatial patterns of a land use regression model[J]. *Environmental Pollution*, 2018, 243: 501-509.
- [36] HUANG C S, LIN T H, HUNG H, et al. Incorporating satellite-derived data with annual and monthly land use regression models for estimating spatial distribution of air pollution[J]. *Environmental Modelling & Software*, 2019, 114: 181-187.
- [37] 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.
- [38] VAN DONKELAAR A, MARTIN R V, BRAUER M, et al. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter[J]. *Environmental Health Perspectives*, 2015, 123(2): 135-143.
- [39] GREEN P J, SILVERMAN B W. Nonparametric regression and generalized linear models: a roughness penalty approach[M]. London: CRC Press, 1993.
- [40] BREMAN L. Random forest[J]. *Machine Learning*, 2001, 45(1): 5-32.
- [41] CHEN T, GUESTRIN C. Xgboost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery, 2016: 785-794.
- [42] VALAVI R, ELITH J, LAHOZ-MONFORT J J, et al. blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models[J]. *Methods in Ecology and Evolution*, 2019, 10(2): 225-232.
- [43] CHEN K, WANG M, HUANG C, et al. Air pollution reduction and mortality benefit during the COVID-19 outbreak in China[J]. *The Lancet Planetary Health*, 2020, 4(6): 210-212.
- [44] CHEN R, YIN P, MENG X, et al. Fine particulate air pollution and daily mortality: a nationwide analysis in 272 Chinese cities[J]. *American Journal of Respiratory and Critical Care Medicine*, 2017, 196(1): 73-81.
- [45] CHEN G, LI S, KNIBBS L D, et al. A machine learning method to estimate PM_{2.5} concentrations across China with remote sensing, meteorological and land use information[J]. *Science of the Total Environment*, 2018, 636: 52-60.
- [46] LI T, SHEN H, YUAN Q, et al. Estimating ground-level PM_{2.5} by fusing satellite and station observations: a geo-intelligent deep learning approach[J]. *Geophysical Research Letters*, 2017, 44(23): 11985-11993.
- [47] ZHAN Y, LUO Y, DENG X, et al. Spatiotemporal prediction of continuous daily PM_{2.5} concentrations across China using a spatially explicit machine learning algorithm[J]. *Atmospheric Environment*, 2017, 155: 129-139.
- [48] MEYER H, REUDENBACH C, WÖLLAUER S, et al. Importance of spatial predictor variable selection in machine learning applications - moving from data reproduction to spatial prediction[J]. *Ecological Modelling*, 2019, 411: 108815.
- [49] KLOOG I, RIDGWAY B, KOUTRAKIS P, et al. Long- and short-term exposure to PM_{2.5} and mortality: using novel exposure models[J]. *Epidemiology*, 2013, 24(4): 555-561.
- [50] SHI G, LU X, DENG Y, et al. Air pollutant emissions induced by population migration in China[J]. *Environmental Science & Technology*, 2020, 54(10): 6308-6318.
- [51] JIA P, LAKERVELD J, WU J, et al. Top 10 research priorities in spatial life-course epidemiology[J]. *Environmental Health Perspectives*, 2019, 127(7): 074501.