

DOI: 10.3969/j.issn.1673-6141.2023.03.006

中国区域PM_{2.5}浓度估算以及影响因素解析

曹媛^{1,2}, 宫明艳³, 沈非^{1,2}, 麻金继^{1,2*}, 杨光^{1,2}, 林锡文^{1,2}

(1 安徽师范大学地理与旅游学院, 安徽 芜湖 241002;

2 资源环境与地理信息工程安徽省工程技术研究中心, 安徽 芜湖 241002;

3 安徽师范大学物理与电子信息学院, 安徽 芜湖 241002)

摘要: 基于2018年中国逐日PM_{2.5}数据, 选用随机森林方法构建了高精度PM_{2.5}浓度估算模型, 并在季节和区域尺度上验证了其时空适用性, 进一步利用特征重要性方法系统阐释了各影响因子对PM_{2.5}浓度变化的重要程度, 最后利用偏依赖技术探究了不同影响因素的交互作用对PM_{2.5}浓度变化产生的综合影响。结果表明: (1) 相比于多元线性回归与极端梯度提升树模型, 利用多源数据构建的随机森林模型精度最高, 可准确模拟出PM_{2.5}的浓度, 且在季节和区域尺度上也有良好的适用性; (2) PM_{2.5}浓度估算模型的特征重要性排序分析发现, 对2018年全国日均PM_{2.5}浓度影响显著的因子主要是时空、大气边界层高度等全局性因素, 表明大气污染防治应把握PM_{2.5}传输机制, 强化区域联防联控; (3) 偏依赖交互效应研究发现温度和相对湿度以及年积日、纬度、温度和大气边界层高度的组合对PM_{2.5}浓度变化会产生显著影响, 说明提升空气质量要从多因子协同治理的角度出发。

关键词: 大气遥感; 随机森林; PM_{2.5}; 时空关联

中图分类号: X513

文献标识码: A

文章编号: 1673-6141(2023)03-245-013

Estimation of PM_{2.5} concentration and analysis of influencing factors in China

CAO Yuan^{1,2}, GONG Mingyan³, SHEN Fei^{1,2}, MA Jinji^{1,2*}, YANG Guang^{1,2}, LIN Xiwen^{1,2}

(1 School of Geography and Tourism, Anhui Normal University, Wuhu 241002, China;

2 Engineering Technology Research Center of Resources Environment and GIS, Anhui Province, Wuhu 241002, China;

3 School of Physics and Electronic Information, Anhui Normal University, Wuhu 241002, China)

Abstract: Based on the daily PM_{2.5} data of China in 2018, a high-precision PM_{2.5} concentration estimation model was constructed using random forest method, and the temporal and spatial applicability of the model was verified at seasonal and regional scales. Further, the importance of each influencing factor to the change of PM_{2.5} concentration was systematically explained using the feature importance method. Finally, the comprehensive influence of the interaction of different influencing factors on PM_{2.5} concentration change was explored using the partial dependence technique. The results show that: (1)

基金项目: 国家自然科学基金项目 (42271372)

作者简介: 曹媛 (1997-), 女, 安徽马鞍山人, 硕士研究生, 主要从事大气环境污染方面的研究。E-mail: yuanCao@ahnu.edu.cn

导师简介: 麻金继 (1975-), 安徽马鞍山人, 博士, 教授, 博士生导师, 主要从事大气环境参量遥感研究。E-mail: jinjima@ahnu.edu.cn

收稿日期: 2021-11-15; 修改日期: 2022-01-09

*通信作者。

Compared with the multiple linear regression model and the extreme gradient ascending tree model, the random forest model based on multi-source data has the highest accuracy, which not only can accurately simulate the $PM_{2.5}$ concentration, but also has good applicability at the seasonal and regional scales. (2) According to the ranking results of model feature importance, the factors that had significant impact on the average daily $PM_{2.5}$ concentration in 2018 were mainly global factors such as space-time and atmospheric boundary layer height, which indicated that the prevention and control of air pollution should follow the $PM_{2.5}$ transmission mechanism, and regional joint prevention and control should be strengthened in air pollution prevention. (3) The partial dependent interaction effect study shows that the combination of temperature, relative humidity, annual cumulative day, latitude, temperature and atmospheric boundary layer height has a significant impact on $PM_{2.5}$ concentration change, indicating that to improve air quality should start from the perspective of multi-factor collaborative governance.

Key words: atmospheric remote sensing; random forest; $PM_{2.5}$; spatio-temporal correlation

0 引言

$PM_{2.5}$ 是空气动力学直径小于或等于 $2.5 \mu m$ 的细微颗粒物,可对人体健康^[1,2]、空气环境^[3-6]、气候变化^[7,8]造成极大负面影响,因此,进行 $PM_{2.5}$ 浓度估算并提高其空间分辨率,可为未来气候变化和人类健康研究提供数据基础,同时科学识别 $PM_{2.5}$ 的关键影响因素对降低 $PM_{2.5}$ 浓度也至关重要。目前国内外学者主要从自然因素、人文因素两个视角展开分析^[9-14]。在研究方法上多使用多元线性回归^[15,16]、主成分回归模型^[17]、地理探测器^[18]、地理加权回归模型^[19,20]、空间计量模型^[21,22]、灰色关联模型^[23,24]等进行 $PM_{2.5}$ 浓度的影响因素研究。但是上述模型多基于自变量与因变量之间是线性关系的假设,难以表征影响因子与 $PM_{2.5}$ 浓度之间复杂的非线性关系。机器学习作为复杂科学领域的新起之秀,可用于 $PM_{2.5}$ 浓度的估算与预测,但缺乏对 $PM_{2.5}$ 影响因素作用机制的解释。目前,学界对 $PM_{2.5}$ 影响因素的研究主要停留在年尺度或月尺度^[25],且研究范围多局限于城市层面^[26],日尺度下全国范围内的 $PM_{2.5}$ 研究较少。

本研究采用随机森林方法,以2018年为例,基于气象、地形、社会经济等辅助数据构建了中国日尺度的 $PM_{2.5}$ 浓度估算模型,并以区域和季节为界限评估其时空适用性。最后利用特征重要性和偏依赖技术多维多视角剖析中国 $PM_{2.5}$ 污染的主要驱动因子及其交互机理,尽可能实现 $PM_{2.5}$ 浓度的准确估算,综合反映自然条件和社会经济状况对 $PM_{2.5}$ 造成的客观影响。

1 数据与方法

1.1 数据来源

所用的基础数据为2018年中国 $PM_{2.5}$ 浓度数据,采用 $PM_{2.5}$ 浓度的24 h滑动均值作为当天的日均值,并剔除缺失值,最终得到2018年的有效天数为353天。该数据采集于中国环境监测总站的全国城市空气质量实时发布平台,数据涉及1456个监测站点,覆盖全国31个省级行政区(香港、澳门、台湾除外)。基于已有研究成果^[27-29]并结合中国 $PM_{2.5}$ 污染现状,遵循科学性、针对性和可操作性原则,从自然和人为因素两个维度遴选出15项指标来定量测度可能影响中国 $PM_{2.5}$ 污染的主要因素。数据详情和来源见表1,其中将径向风(U)和

纬向风 (V) 合成风速 (ws, m/s) 以及风向 (wd, °) 数据, 风向范围为 $-180^{\circ}\sim 180^{\circ}$, 以 0° 表示西风, 逆时针旋转为正, 顺时针旋转为负。在前人研究上, 辅助以经纬度和年积日作为影响因素表征 PM_{2.5} 变化的时空异质性。

表1 影响因素信息表

Table 1 Information table of influencing factors

Data type	Data name	Data code	Units	Data source	
Natural factors	10m u-component of wind	U	$m\cdot s^{-1}$	ERA5 Dataset (https://climate.copernicus.eu)	
	10m v-component of wind	V	$m\cdot s^{-1}$		
	2m temperature	TMP	K		
	Total precipitation	PRE	m		
	Evaporation	EVA	m		
	Relative humidity	RH	%		
	Boundary layer height	BLH	m		
	Elevation	DEM	m		NASA SRTM Digital Elevation Dataset (https://cmr.earthdata.nasa.gov)
	Aspect	ASPECT	(°)		
	Vegetation index	NDVI	—		MOD13A2 v6 Dataset (https://lpdaac.usgs.gov)
Human factors	Population	POP	—	Landscan Dataset (https://landscan.ornl.gov/)	
	Nighttime lights	NTL	DN	NPP VIIRS Dataset (https://eogdata.mines.edu/products/vnl/)	
	Land use	LUC	—	MCD12Q1.006 Dataset (https://lpdaac.usgs.gov)	
Time and space factors	Longitude	LON	(°)	China Environmental Monitoring Station (http://www.cnemc.cn/)	
	Latitude	LAT	(°)		
	Day of year	YOD	day		

1.2 随机森林模型

随机森林 (RF) 是 Breiman^[30] 于 2001 年提出的一种非线性建模工具, 可模拟预测变量和响应变量之间的复杂关系。该模型属于以决策树作为基学习器的一种集成学习算法, 当需要对某个样本进行预测时, 利用 Bootstrap 重采样方法从原始样本中抽取多个样本, 通过构造多个决策树, 统计森林中的每棵树对该样本的预测结果, 进而通过投票法从这些预测结果中选出最后的结果。该模型可以有效地处理大量无需降维的预测变量, 在训练过程中可检测到预测变量之间的相关关系, 对变量间的多元共线性不敏感, 因此无须提前对变量进行筛选和规范化处理^[30], 极其适合多维多因素驱动下的大尺度、高精度 PM_{2.5} 浓度预测与影响因素测度研究。

在模型可解释性方面, 随机森林可以评估建模过程中每个特征的重要性, 产生无偏估计, 从而识别预测变量的重要性并探究预测变量对响应变量之间的偏依赖关系^[31]。基于 Scikit-learn 库的随机森林特征重要性的评判主要通过计算基尼指数 G_i 来决定该特征对节点纯度增加的程度。特征越重要, 对节点纯度增加的效果越好。其计算公式为

$$G_{L,m} = \sum_{k=1}^K \hat{p}_{mk} (1 - \hat{p}_{mk}), \quad (1)$$

$$I_{VM,m} = G_{L,m} - G_{L,l} - G_{L,r}, \quad (2)$$

$$I_{VM,j} = \sum_{m=1}^M I_{VM,m}, \quad (3)$$

$$I_{VM} = \frac{1}{n} \sum_{j=1}^n I_{VM,j}, \quad (4)$$

式中 $G_{L,m}$ 为节点 m 的 Gini 指数, K 为样本集类别数, \hat{p}_{mk} 为节点 m 属于第 K 类的概率, $I_{VM,m}$ 为该自变量在节点 m 的重要性, $G_{L,l}$ 和 $G_{L,r}$ 为由节点 m 分裂的两个新节点的 Gini 指数, $I_{VM,j}$ 为该自变量在第 j 棵树中出现了 M 次的重要性, I_{VM} 为该自变量在随机森林中的归一化重要性得分, n 为随机森林中树的数量。

1.3 变量偏依赖方法

特征重要性只反映了预测变量对响应变量的相对重要性, 无法量化预测变量的变化对响应变量变化趋势的影响。部分依赖图 (PDP) 描述了某一预测变量在排除其他预测变量的情况下, 对响应变量的出现概率在统计结果上的影响, 和数学理论中“偏微分”的概念较为相似, 因此它是一种全局方法, 可显示一个或两个特征对先前拟合模型的预测结果的边际影响, 由此表征特征变量和响应变量之间的交互关系。

本研究主要采用 PDP 来探究特征变量与 $PM_{2.5}$ 的相关关系, 其计算公式为

$$\hat{f}(x_s) = \int \hat{F}(x_s, x_c) dx_c, \quad (5)$$

式中 $\hat{f}(x_s)$ 为变量 x_s 对 $PM_{2.5}$ 的偏依赖效应, $\hat{F}(x)$ 为 $PM_{2.5}$ 浓度估算模型, x_s 为感兴趣的 $PM_{2.5}$ 浓度影响变量, x_c 为剩余的不关注的影响变量, x_s 和 x_c 共同组成预测变量。

1.4 模型精度评估方法

本研究主要利用十折交叉验证技术来测试模型精确度。十折交叉验证技术即将数据集分成十份, 轮流将其中的九份作为训练集, 一份作为测试集进行实验, 并将 10 次输出的结果取平均值作为最终的结果。通过 R^2 和均方根误差 E_{RMS} 指标来对精度进行表征, 其计算公式为

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (6)$$

$$E_{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (7)$$

式中 y_i 是样本 i 的真实值, \hat{y}_i 是样本 i 的估计值, \bar{y} 为真实值的平均值, n 为样本的数量。

2 $PM_{2.5}$ 浓度估算模型的构建与验证

2.1 随机森林模型构建

$PM_{2.5}$ 成因复杂, 可通过地基和遥感反演进行监测^[32], 其本底排放、传输沉降受气象、地形、社会经济等众多因素的影响。本研究通过结合自然因素和社会经济因素, 并引入时空信息, 立足于 $PM_{2.5}$ 的产生、传输、消

耗与沉降角度,构建时空PM_{2.5}浓度估算模型,实现PM_{2.5}的浓度预测。其中引入时空信息以避免影响因子,尤其是气象变量的空间和时间异质性,从而进一步探究PM_{2.5}浓度的时空差异特征。以2018年为例,在日尺度上以中国作为研究区域,将自然因素(10 m风速、10 m风向、降水量、温度、相对湿度、蒸发量、高程、坡向、植被指数)、人为因素(灯光、人口)、以及时间(年积日)、空间(经纬度)信息作为模型输入,PM_{2.5}地面监测站点日平均浓度数据作为模型输出,根据监测站点空间位置提取输入信息。整个数据集共有51万条数据,按照分布样本的均匀性原则,随机将数据集的80%用于训练,20%用于测试,将这个过程重复10次,并以均方根误差 E_{RMS} 和 R^2 作为精度验证指标。通过不断迭代主要参数,以在测试集上代价函数最小的超参数组合作为最优超参数,最终设置最优参数组合基评估器($n_{\text{estimators}}$)为119,最大深度(max_depth)为39,最大特征个数(max_features)为12。在最优模型上,训练集 R^2 为0.99, $E_{\text{RMS}} = 3.97 \mu\text{g}\cdot\text{m}^{-3}$ [图1(a)];在测试集上 R^2 为0.91, E_{RMS} 为 $10.13 \mu\text{g}\cdot\text{m}^{-3}$ [图1(b)],说明随机森林模型能够有效地建立本研究所使用的影响因子与PM_{2.5}浓度之间的回归关系。以2018年8月20日为例,本研究所构建的模型基本能很好地反映出真实PM_{2.5}浓度[图2]。

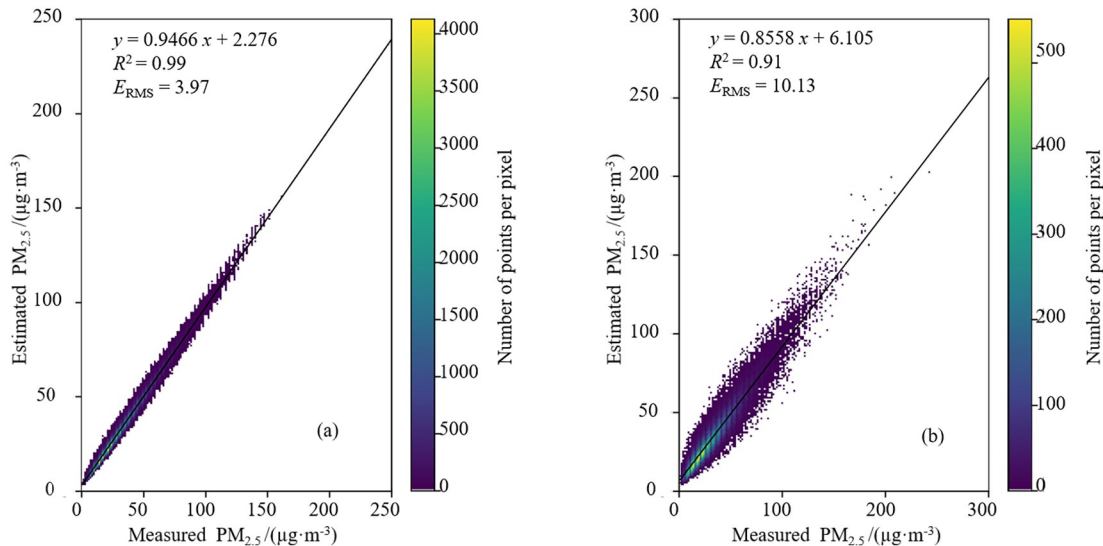


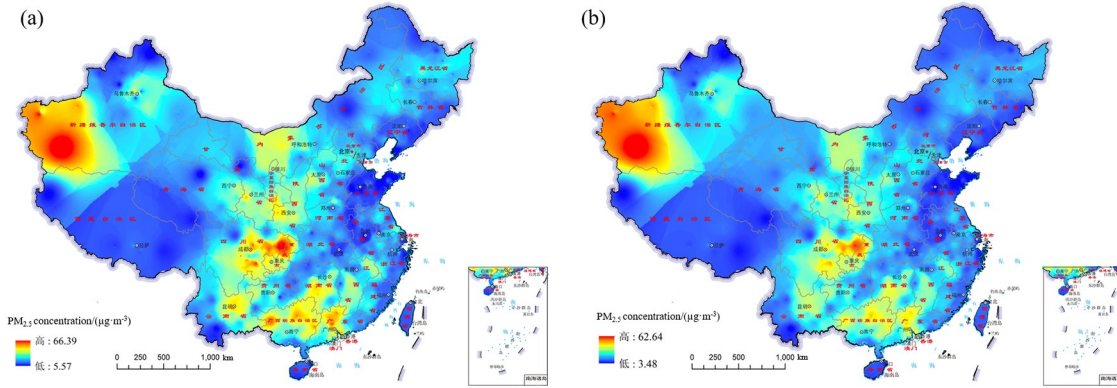
图1 随机森林模型反演精度。(a) 训练集;(b) 测试集

Fig. 1 Random forest model inversion accuracy. (a) Train dataset; (b) test dataset

2.2 模型时空适用性验证

为进一步验证上述模型在时间和空间维度上的表现性能,分别在季节和区域尺度上进行验证。首先根据中国传统的季节划分依次构建春、夏、秋、冬模型进行训练,以探究不同季节随机森林模型的适用性。模型构建过程和上述随机森林模型类似,训练得到的具体参数如表2所示。图3为不同季节模型测试集的反演精度。由图可知,春季模型在测试集上 R^2 为0.85, E_{RMS} 为 $11.42 \mu\text{g}\cdot\text{m}^{-3}$ [图3(a)];夏季模型在测试集上 R^2 为0.78, E_{RMS} 为 $7.15 \mu\text{g}\cdot\text{m}^{-3}$ [图3(b)];秋季模型在测试集上 R^2 为0.91, E_{RMS} 为 $8.29 \mu\text{g}\cdot\text{m}^{-3}$ [图3(c)];冬季模型测试集上 R^2 为0.91, E_{RMS} 为 $13.67 \mu\text{g}\cdot\text{m}^{-3}$ [图3(d)]。季节模型中秋冬模型精度较高,春夏稍次之,说明随机森林模型在秋冬季节的适用性更好。

然后再分区构建东部、中部和西部模型以探究不同区域随机森林模型的适用性,其中东、中、西部具体



注:此图基于国家自然资源部标准地图服务系统的标准地图[审图号:GS(2019)1827号]绘制,底图无修改

图2 2018年8月20日 $PM_{2.5}$ 浓度原始值与估计值的空间分布。(a)原始值;(b)估计值

Fig. 2 Spatial distribution of original and estimated $PM_{2.5}$ concentration on August 20, 2018. (a) Original value; (b) estimate value

表2 季节模型训练参数表

Table 2 Seasonal model training parameter table

Model	Main parameters		
	n_estimators	max_depth	max_feature
Spring model	119	36	12
Summer model	119	37	12
Autumn model	123	36	13
Winter model	124	33	13

范围参考相关文献^[33]。具体模型构建方法与上述随机森林模型类似,将训练集根据东、中、西区域划分为三部分分别进行训练。具体参数如表3所示。东部模型在测试集上 R^2 为0.91, E_{RMS} 为 $9.41 \mu\text{g}\cdot\text{m}^{-3}$ [图4(a)];中部模型在测试集上 R^2 为0.93, E_{RMS} 为 $9.07 \mu\text{g}\cdot\text{m}^{-3}$ [图4(b)];西部模型在测试集上 R^2 为0.86, E_{RMS} 为 $12.71 \mu\text{g}\cdot\text{m}^{-3}$ [图4(c)]。整体来看,区域模型精度整体表现良好,在测试集上平均 R^2 为0.89, E_{RMS} 平均值为 $10.40 \mu\text{g}\cdot\text{m}^{-3}$ 。其中中部区域模型精度最高,东部地区次之,西部地区相比最差,但整体精度维持在较高水平,说明随机森林模型能较准确地模拟出 $PM_{2.5}$ 浓度。

总体来说,利用随机森林模型构建 $PM_{2.5}$ 浓度估算模型,不论是在时间尺度还是在空间尺度上,都表现出良好的性能,所构建的模型在训练集和测试集上的 R^2 值均大于0.75。其中,东部和中部城市以及秋季和冬季的模型反演精度最好, R^2 值均大于0.9,且均方根误差都在15以下。西部城市和春季拟合精度相对较差的主要原因是西部地区 $PM_{2.5}$ 监测站点稀疏,地形复杂,植被覆盖率低,且沙漠戈壁分布密集,在春季易发生沙尘天气,导致 $PM_{2.5}$ 浓度急剧增加,出现较高的异常值,导致模型精度有些许下降。与之类似,夏季易出现台风等极端天气,对 $PM_{2.5}$ 浓度产生影响,模型精度有所降低,但仍维持在较高水平。各模型具体精度见表4。

2.3 模型精度对比验证

为进一步验证模型的精度,利用多元线性回归和极端梯度提升树方法分别构建 $PM_{2.5}$ 浓度估算模型,最终的模型精度如表5所示。由表可知多元线性回归模型精度最低,无法拟合预测因子和 $PM_{2.5}$ 之间的关系,且在估算过程中会拟合出 $PM_{2.5}$ 的负值,不符合实际情况;极端梯度提升树模型能够在一定程度上拟合预测因子

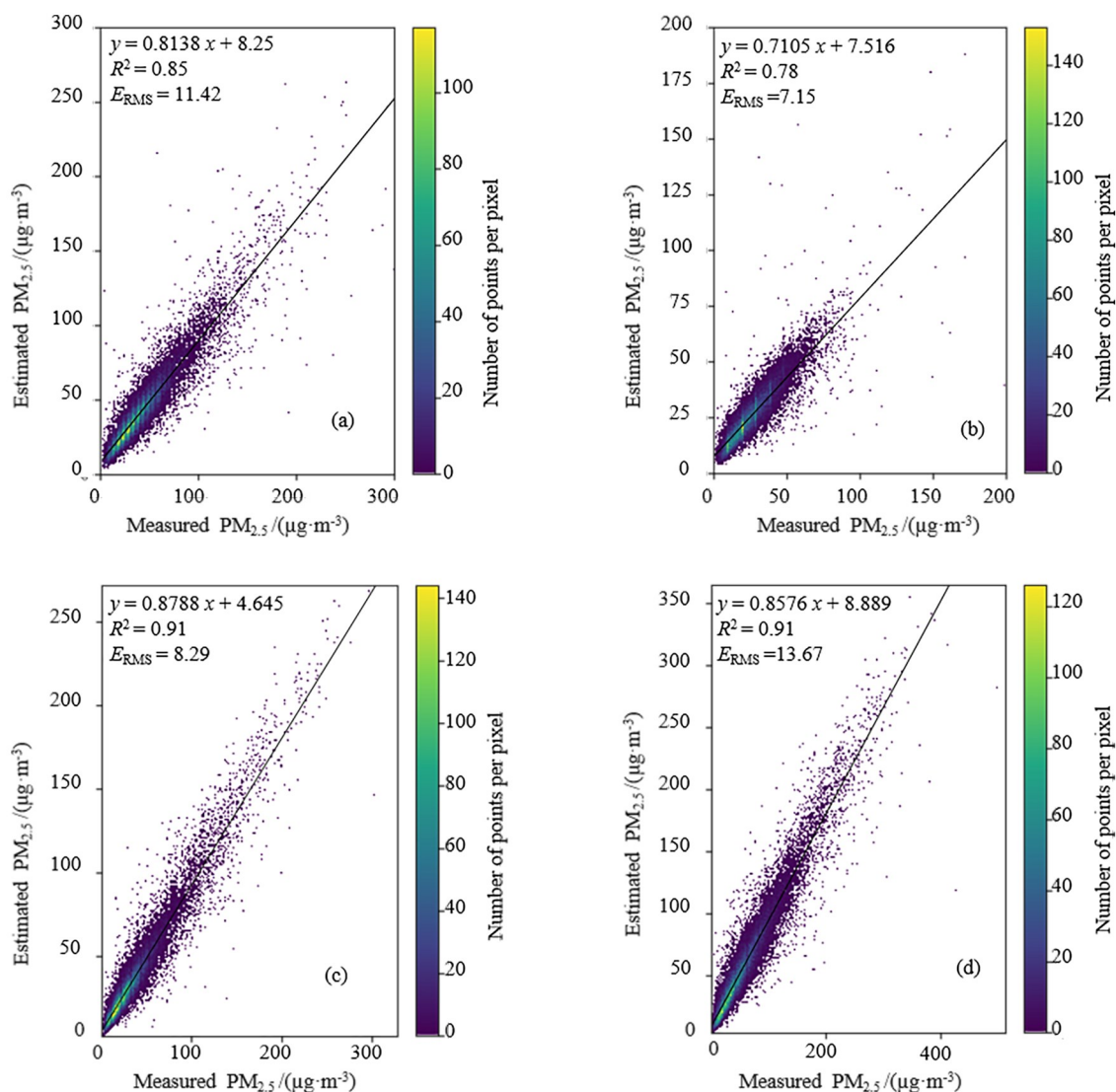


图 3 季节模型测试集反演精度。(a) 春季模型; (b) 夏季模型; (c) 秋季模型; (d) 冬季模型
 Fig. 3 Seasonal model inversion accuracy on test dataset. (a) Spring model; (b) summer model;
 (c) autumn model; (d) winter model

表 3 区域模型训练参数表

Table 3 Area model training parameter table

Model	Main parameters		
	n_estimators	max_depth	max_feature
Eastern model	119	36	12
Central model	119	40	12
Western model	119	30	12

和 PM_{2.5} 之间的关系, 但其精度明显低于随机森林模型, 因此, 优选随机森林方法构建 PM_{2.5} 浓度估算模型。

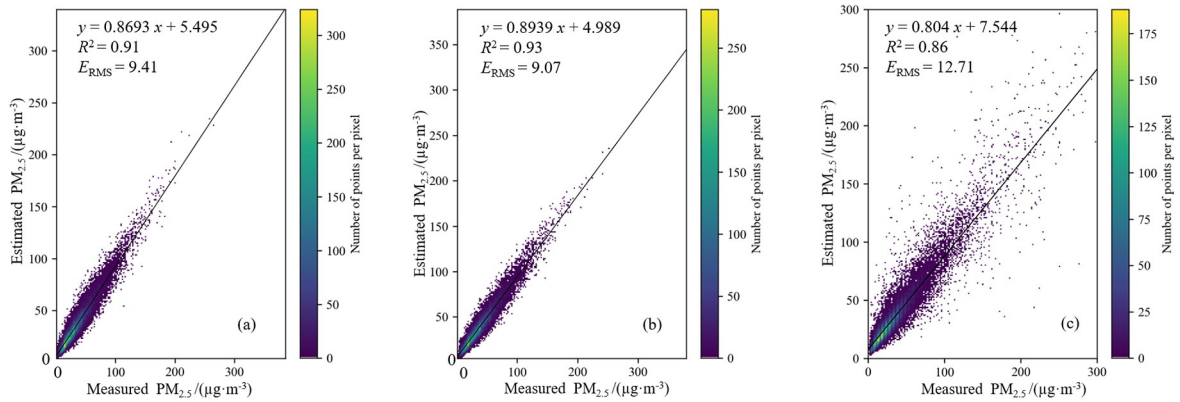


图4 区域模型测试集反演精度。(a) 东部模型; (b) 中部模型; (c) 西部模型

Fig. 4 Area model inversion accuracy on test dataset. (a) Eastern model; (b) central model; (c) western model

表4 模型反演精度表

Table 4 Accuracy table of model inversion

Kind	Model	Train dataset		Test dataset	
		R^2	$E_{RMS}/(\mu\text{g}\cdot\text{m}^{-3})$	R^2	$E_{RMS}/(\mu\text{g}\cdot\text{m}^{-3})$
Overall	China model	0.99	3.97	0.91	10.13
Area	Eastern model	0.99	3.48	0.91	9.41
	Central model	0.99	3.54	0.93	9.07
	Western model	0.98	4.85	0.86	12.71
Season	Spring model	0.98	4.56	0.85	11.42
	Summer model	0.97	2.57	0.78	7.15
	Autumn model	0.99	3.16	0.91	8.29
	Winter model	0.99	5.15	0.91	13.67

表5 模型精度对照表

Table 5 Model precision comparison table

Model	Train dataset		Test dataset	
	R^2	$E_{RMS}/(\mu\text{g}\cdot\text{m}^{-3})$	R^2	$E_{RMS}/(\mu\text{g}\cdot\text{m}^{-3})$
Multiple linear regression	0.24	29.73	0.23	29.72
Extreme gradient boosting	0.78	15.75	0.75	16.47
Random forest	0.99	3.99	0.91	10.13

3 实验结果与讨论

3.1 $\text{PM}_{2.5}$ 影响因素的重要性排序

利用随机森林的特征重要性排序和偏依赖方法来增强模型的可解释性。图5反映了模型特征重要性排序结果。总的来说,年积日、经度、纬度、大气边界层高度、温度是影响 $\text{PM}_{2.5}$ 浓度的主导因子。证明时空因素对 $\text{PM}_{2.5}$ 的浓度分布格局产生主导作用。其次大气边界层作为地表和自由对流层之间的缓冲区,是调节城市地区 $\text{PM}_{2.5}$ 污染最关键的因素之一^[34],其高度会影响大气湍流的运动尺度以及对流活动的演化过程,进而改

变大气环境容量,最终会决定PM_{2.5}排放潜力^[35,36]。温度对PM_{2.5}的影响可分为直接影响和间接影响。直接影响体现在温度不但可以影响PM_{2.5}在大气中的扩散与稀释,还能催化二次污染形成。PM_{2.5}作为一种气溶胶颗粒,其布朗扩散运动与温度高度相关,温度越高,近地面浓度会降低,而高空浓度会显著提升^[37]。间接影响体现在温度决定着人类调节室内气候的意愿强度,超出人类适应范围的温度环境会提升人类通过消耗化石能源来改变环境温度的概率,从而影响PM_{2.5}排放。

在特征重要性排序结果上,土地利用、人口、坡向、灯光和植被覆盖的影响力相对不显著。主要归因于PM_{2.5}的传播过程具有长期性和复杂性,污染源和监测站存在空间错配现象,上述因素所产生的PM_{2.5}污染物会在复杂的大气环境下远距离传播扩散,导致监测站所测得的PM_{2.5}浓度值并非全由本地的污染源产生,从而降低了土地利用和人口等地方性因素的特征重要性。因此,PM_{2.5}传输模拟具有十分重要的研究价值。

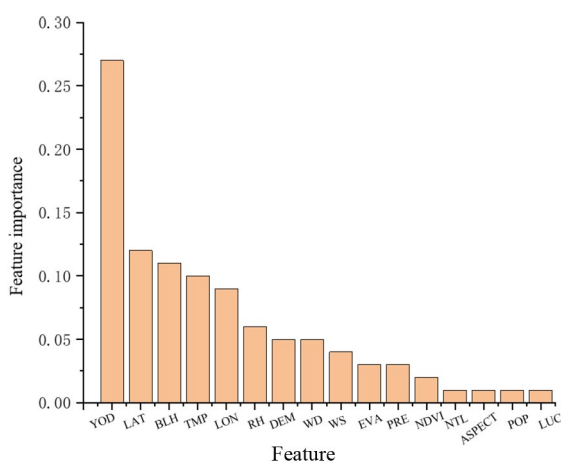


图5 模型特征重要性排序

Fig. 5 Model feature importance ranking

3.2 影响因子的交互效应

PM_{2.5}浓度分布受众多因素的共同影响,各因素对PM_{2.5}的影响并非孤立存在,其双向交互作用可能会对PM_{2.5}浓度产生强烈的增强或抵消效应。为了厘清这种交互效应,在上述重要性分析的基础上,筛选出对PM_{2.5}浓度分布产生显著影响的因素,进而通过将重要因素组合输出偏依赖图以量化因子间交互作用对PM_{2.5}的影响。影响因素两两交互作用对PM_{2.5}浓度影响较为显著的组合如图6所示。

年积日和大气边界层高度的交互作用如图6(a)所示,当处于春季以及冬季时,随着大气边界层高度的降低,PM_{2.5}呈现上升的趋势,尤其大气边界层高度在400 m以下时,易产生高浓度的PM_{2.5};当处于夏季时,随着大气边界层高度的变化,PM_{2.5}浓度未发生显著变化,即在夏季,大气边界层高度对PM_{2.5}的影响较小;纬度与大气边界层高度的交互作用如图6(b)所示,PM_{2.5}浓度与纬度呈现正相关,与大气边界层高度呈现负相关的关系,即随着纬度的升高,大气边界层高度的降低,PM_{2.5}呈现上升的趋势;温度与大气边界层高度的交互作用如图6(c)所示,当气温处于零下,大气边界层高度较低时,PM_{2.5}浓度呈现高值,尤其当气温处于零下5℃,大气边界层高度位于200 m以下时,PM_{2.5}浓度值出现高峰;温度与相对湿度的交互作用如图6(d)所示,总体来看,随着温度升高,PM_{2.5}浓度呈上升的趋势,但当温度在10~25℃区间时,由于气温适宜,人类借助化石燃料调节气温的意愿较小,导致PM_{2.5}浓度值出现低谷,而湿度对PM_{2.5}浓度的影响呈现"倒U型"变

化趋势,相对湿度在50%~60%之间时,PM_{2.5}浓度达到高峰。这从侧面说明在冬季,北方高相对湿度的天气条件易于PM_{2.5}的产生,从而增加了污染天气出现的概率。

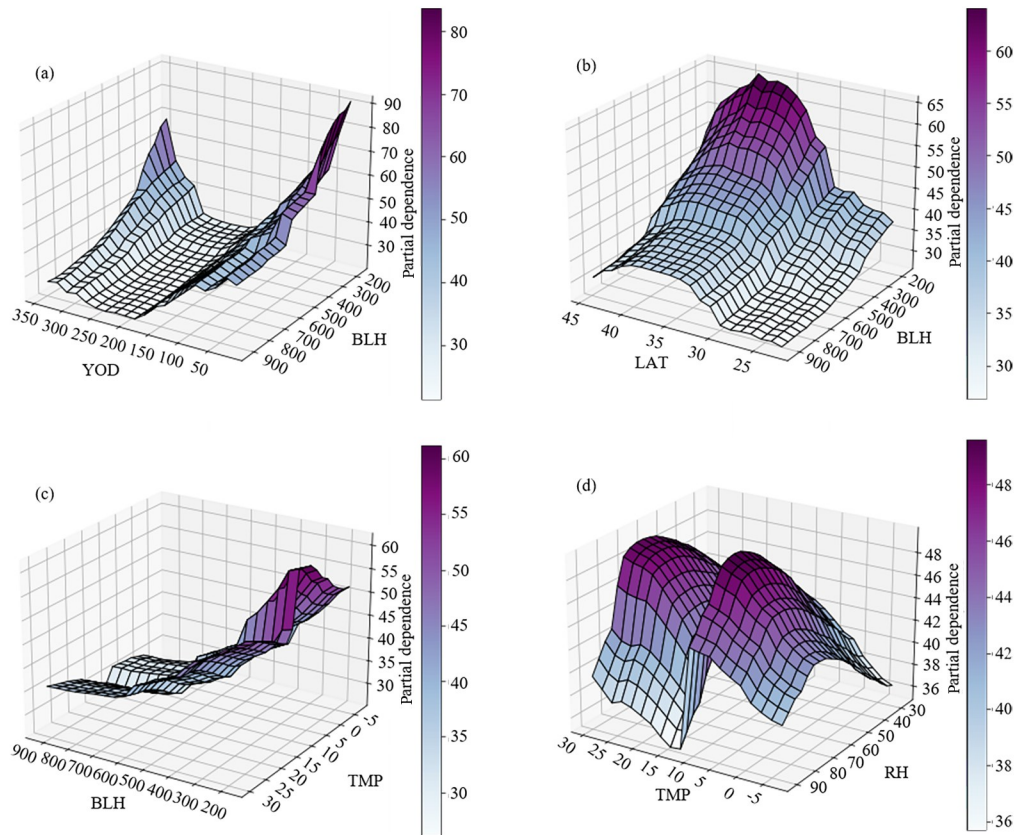


图6 影响因子对PM_{2.5}日均浓度变化的三维空间效应图。(a) 年积日与大气边界层高度;(b) 纬度与大气边界层高度;(c) 大气边界层高度与温度;(d) 温度与相对湿度

Fig. 6 Three-dimensional spatial effect diagram of influence factors on changes of PM_{2.5} daily concentration. (a) AOD and BLH; (b) LAT and BLH; (c) BLH and TMP; (d) TMP and RH

4 结 论

1) 综合对PM_{2.5}浓度产生影响的自然因素和社会经济因素,辅以时空信息,构建随机森林模型模拟2018年中国区域的PM_{2.5}浓度空间分布,结果表明,不论是分区域还是分季节,均可以达到较好的模拟效果。其中西部地区因站点稀疏,植被覆盖率低,地形起伏大等原因导致精度有所降低,夏季台风以及春季沙尘等极端天气的频发导致这两个季节的模型精度有些许下降,不过仍维持在较高水平。

2) 从相对模型特征重要性排序结果来说,在2018年日尺度上对PM_{2.5}浓度影响靠前的因子主要是时空、大气边界层高度等全局性因素。这表明,从全国与区域联防联控视角来看,治理PM_{2.5}污染不仅需控制污染源排放,更重要的是从全局出发,把握PM_{2.5}扩散与传播的空间规律,构建大气污染防治一体化协作机制。

3) 偏依赖交互效应表明,大气边界层高度分别和年积日、纬度与温度的组合以及温度和相对湿度的组合对PM_{2.5}浓度变化产生显著影响,在该协同作用下有助于PM_{2.5}浓度的生成,说明改善空气环境应从多因子协同治理的角度出发。

参考文献:

- [1] Zhao B, Zheng H T, Wang S X, *et al.* Change in household fuels dominates the decrease in PM_{2.5} exposure and premature mortality in China in 2005–2015 [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115(49): 12401-12406.
- [2] Lu M, Lin B L, Inoue K, *et al.* PM_{2.5}-related health impacts of utilizing ammonia-hydrogen energy in Kanto Region, Japan [J]. *Frontiers of Environmental Science & Engineering*, 2018, 12(2): 13.
- [3] Cai W, Li K, Liao H, *et al.* Weather conditions conducive to Beijing severe haze more frequent under climate change [J]. *Nature Climate Change*, 2017, 7(4): 257-262.
- [4] Duan J X, Zhai W X, Cheng C Q, *et al.* Socio-economic factors influencing the spatial distribution of PM_{2.5} concentrations in China: An exploratory analysis [J]. *Environmental Science*, 2018, 39(5): 2498-2504.
段杰雄, 翟卫欣, 程承旗, 等. 中国PM_{2.5}污染空间分布的社会经济影响因素分析 [J]. 环境科学, 2018, 39(5): 2498-2504.
- [5] Yin X M, Zhu B, Xiong Y J, *et al.* Objective analysis on circulation types and its links to haze days over Beijing during 2007–2016 [J]. *China Environmental Science*, 2020, 40(1): 123-134.
尹晓梅, 朱彬, 熊亚军, 等. 2007–2016年北京天气分型与霾日的关联 [J]. 中国环境科学, 2020, 40(1): 123-134.
- [6] Zhao H J, Che H Z, Zhang X Y, *et al.* Characteristics of visibility and particulate matter (PM_{2.5}) in an urban area of Northeast China [J]. *Atmospheric Pollution Research*, 2013, 4(4): 427-434.
- [7] Ma X Y, Shi G Y, Guo Y F, *et al.* Radiative forcing by greenhouse gases and sulfate aerosol [J]. *Acta Meteorologica Sinica*, 2005, 63(1): 41-48.
马晓燕, 石广玉, 郭裕福, 等. 温室气体和硫酸盐气溶胶的辐射强迫作用 [J]. 气象学报, 2005, 63(1): 41-48.
- [8] Wang Y S, Zhang J K, Wang L L, *et al.* Researching significance, status and expectation of haze in Beijing-Tianjin-Hebei region [J]. *Advances in Earth Science*, 2014, 29(3): 388-396.
王跃思, 张军科, 王莉莉, 等. 京津冀区域大气霾污染研究意义、现状及展望 [J]. 地球科学进展, 2014, 29(3): 388-396.
- [9] Li X, Gao Z Q, Li Y B, *et al.* Meteorological conditions for severe foggy haze episodes over North China in 2016-2017 winter [J]. *Atmospheric Environment*, 2019, 199: 284-298.
- [10] Zhao X L, Zhou W Q, Han L J, *et al.* Spatiotemporal variation in PM_{2.5} concentrations and their relationship with socioeconomic factors in China's major cities [J]. *Environment International*, 2019, 133: 105145.
- [11] Aguilera I, Eeftens M, Meier R, *et al.* Land use regression models for crustal and traffic-related PM_{2.5} constituents in four areas of the SAPALDIA study [J]. *Environmental Research*, 2015, 140: 377-384.
- [12] He X, Lin Z S. Interactive effects of the influencing factors on the changes of PM_{2.5} concentration based on GAM model [J]. *Environmental Science*, 2017, 38(1): 22-32.
贺祥, 林振山. 基于GAM模型分析影响因素交互作用对PM_{2.5}浓度变化的影响 [J]. 环境科学, 2017, 38(1): 22-32.
- [13] Zhang L K, Lu S F, Jiao K L, *et al.* Pollution characteristics of particulate matter in urban districts of Baotou and their relationships with meteorological conditions [J]. *Journal of Atmospheric and Environmental Optics*, 2017, 12(6): 401-410.
张连科, 鲁尚发, 焦坤灵, 等. 包头城区冬春大气颗粒物污染特征及其与气象条件关系 [J]. 大气与环境光学学报, 2017, 12(6): 401-410.
- [14] Zheng K L, Huang Y, Yao X Y, *et al.* Correlation between PM_{2.5}, NO₂ and tourism activities, weather factors in Zhangjiajie City [J]. *Journal of Atmospheric and Environmental Optics*, 2020, 15(5): 347-356.
郑凯莉, 黄毅, 姚小云, 胡晓琰. 张家界市PM_{2.5}、NO₂与旅游活动及天气因素的相关性分析 [J]. 大气与环境光学学报, 2020, 15(5): 347-356.
- [15] Zhang S P, Han L J, Zhou W Q, *et al.* Relationships between fine particulate matter (PM_{2.5}) and meteorological factors in winter at typical Chinese cities [J]. *Acta Ecologica Sinica*, 2016, 36(24): 7897-7907.

- 张淑平, 韩立建, 周伟奇, 等. 冬季PM_{2.5}的气象影响因素解析 [J]. 生态学报, 2016, 36(24): 7897-7907.
- [16] Jing R H, Ma J J, Wang C. Methods of PM_{2.5} inversion based on multi-source data [J]. *Journal of Atmospheric and Environmental Optics*, 2015, 10(1): 51-62.
景瑞环, 麻金继, 汪超. 基于多源数据的PM_{2.5}反演方法 [J]. 大气与环境光学学报, 2015, 10(1): 51-62.
- [17] Zhang H, Dong X G, Li Q. Principal component regression analysis of influencing factors of PM_{2.5} concentration [J]. *Journal of Changchun University of Technology*, 2017, 38(2): 105-110.
张红, 董小刚, 李群. PM_{2.5}浓度影响因素的主成分回归分析 [J]. 长春工业大学学报, 2017, 38(2): 105-110.
- [18] Zhou L, Zhou C H, Yang F, et al. Spatio-temporal evolution and the influencing factors of PM_{2.5} in China between 2000 and 2011 [J]. *Acta Geographica Sinica*, 2017, 72(11): 2079-2092.
周亮, 周成虎, 杨帆, 等. 2000—2011年中国PM_{2.5}时空演化特征及驱动因素解析 [J]. 地理学报, 2017, 72(11): 2079-2092.
- [19] Wang S J, Liu X P, Yang X, et al. Spatial variations of PM_{2.5} in Chinese Cities for the joint impacts of human activities and natural conditions: A global and local regression perspective [J]. *Journal of Cleaner Production*, 2018, 203: 143-152.
- [20] Tu M, Liu Z, He C, et al. The relationships between urban landscape patterns and fine particulate pollution in China: A multiscale investigation using a geographically weighted regression model [J]. *Journal of Cleaner Production*, 2019, 237: 117744.
- [21] Peng L S, Sun H, Nie F F. The evolution of temporal and spatial pattern and influencing factors of the air pollution in China [J]. *Journal of Environmental Economics*, 2017, 2(1): 42-56.
彭丽思, 孙涵, 聂飞飞. 中国大气污染时空格局演变及影响因素研究 [J]. 环境经济研究, 2017, 2(1): 42-56.
- [22] Liu H M, Fang C L, Huang J J, et al. The spatial-temporal characteristics and influencing factors of air pollution in Beijing-Tianjin-Hebei urban agglomeration [J]. *Acta Geographica Sinica*, 2018, 73(1): 177-191.
刘海猛, 方创琳, 黄解军, 等. 京津冀城市群大气污染的时空特征与影响因素解析 [J]. 地理学报, 2018, 73(1): 177-191.
- [23] Han J, Li Y Z, Chen X C, et al. Analysis of the influencing factors of PM_{2.5} concentration in the urban areas of China's towns based on grey correlation model [J]. *Environmental Protection Science*, 2018, 44(3): 69-73.
韩婧, 李元征, 陈新闻, 等. 基于灰色关联模型的中国城镇PM_{2.5}浓度影响因素分析 [J]. 环境保护科学, 2018, 44(3): 69-73.
- [24] He X, Lin Z S, Liu H Y, et al. Analysis of the driving factors of PM_{2.5} in Jiangsu Province based on grey correlation model [J]. *Acta Geographica Sinica*, 2016, 71(7): 1119-1129.
贺祥, 林振山, 刘会玉, 等. 基于灰色关联模型对江苏省PM_{2.5}浓度影响因素的分析 [J]. 地理学报, 2016, 71(7): 1119-1129.
- [25] Nan Y, Zhang Q Q, Zhang B H. Influencing factors of long-term variations on gridded PM_{2.5} of typical regions in China based on GAM model [J]. *Environmental Science*, 2020, 41(2): 499-509.
南洋, 张倩倩, 张碧辉. 基于GAM模型分析中国典型区域网格化PM_{2.5}长期变化影响因素 [J]. 环境科学, 2020, 41(2): 499-509.
- [26] Liu Z H, Huang J W, Kong D Y. The spatial-temporal characteristics and influencing factors of PM_{2.5} in Wuhan metropolitan area [J]. *Environmental Protection Science*, 2019, 45(3): 51-59.
刘子豪, 黄建武, 孔德亚. 武汉城市圈PM_{2.5}的时空特征及其影响因素解析 [J]. 环境保护科学, 2019, 45(3): 51-59.
- [27] Wei J, Li Z Q, Lyapustin A, et al. Reconstructing 1-km-resolution high-quality PM_{2.5} data records from 2000 to 2018 in China: Spatiotemporal variations and policy implications [J]. *Remote Sensing of Environment*, 2021, 252: 112136.
- [28] Wei J, Huang W, Li Z Q, 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.
- [29] Zheng Y, Zhang Q, Liu Y, et al. Estimating ground-level PM_{2.5} concentrations over three megalopolises in China using

- satellite-derived aerosol optical depth measurements [J]. *Atmospheric Environment*, 2016, 124: 232-242.
- [30] Breiman L. Random forests [J]. *Machine learning*, 2001, 45(1): 5-32.
- [31] Guidotti R, Monreale A, Ruggieri S, *et al.* A survey of methods for explaining black box models [J]. *ACM Computing Surveys*, 2019, 51(5): 1-42.
- [32] Wang C, Liu Q M, Ying N, *et al.* Air quality evaluation on an urban scale based on MODIS satellite images [J]. *Atmospheric Research*, 2013, 132: 22-34.
- [33] Zhao X Y, Wang W J, Wan W Y. Regional inequalities of residents' health level in China: 2003—2013 [J]. *Acta Geographica Sinica*, 2017, 72(4): 685-698.
赵雪雁, 王伟军, 万文玉. 中国居民健康水平的区域差异: 2003-2013 [J]. *地理学报*, 2017, 72(4): 685-698.
- [34] Miao Y C, Liu S H, Guo J P, *et al.* Unraveling the relationships between boundary layer height and PM_{2.5} pollution in China based on four-year radiosonde measurements [J]. *Environmental Pollution*, 2018, 243: 1186-1195.
- [35] Zhang J, Liu D Y, Qian Y Y, *et al.* Boundary layer features of the successional fog and haze episode in Jiangsu area [J]. *Journal of Arid Meteorology*, 2018, 36(3): 483-491.
张静, 刘端阳, 钱映月, 等. 一次持续性雾霾天气的边界层结构特征 [J]. *干旱气象*, 2018, 36(3): 483-491.
- [36] Zhang Y H, Li S Y. Climatological characteristics of planetary boundary layer height over Japan [J]. *International Journal of Climatology*, 2019, 39(10): 4015-4028.
- [37] Li W G, Duan F K, Zhao Q, *et al.* Investigating the effect of sources and meteorological conditions on wintertime haze formation in Northeast China: A case study in Harbin [J]. *Science of the Total Environment*, 2021, 801: 149631.