

文章编号: 1673-9469 (2017) 02-0095-05

doi:10.3969/j.issn.1673-9469.2017.02.019

## 基于 GAM 的北京 PM<sub>2.5</sub> 浓度变化的影响因素研究

刘珊, 刘迪

(河海大学 地球科学与工程学院, 江苏 南京 211100)

**摘要:** 为辨识与测度不同影响因子对 PM<sub>2.5</sub> 浓度变化的作用机理, 以 2015 年北京 PM<sub>2.5</sub> 浓度时间演变模式为基础, 建立 PM<sub>2.5</sub> 与各大气污染物 (PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub>、CO、O<sub>3</sub>) 及气象因素 (日均温度、风力、风向) 的 GAM 模型, 探索不同因素对 PM<sub>2.5</sub> 浓度变化的影响作用。结果显示: (1) 北京 PM<sub>2.5</sub> 浓度具有夏秋季低、春冬季高的时间分布特点; (2) 2015 年北京 PM<sub>2.5</sub> 浓度变化与 PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub>、CO 大体呈线性正相关, 且正相关程度由强到弱为: CO>PM<sub>10</sub>>SO<sub>2</sub>>NO<sub>2</sub>, 而与 O<sub>3</sub>、温度和风因子的关系更为复杂; (3) GAM 模型的拟合优度 R<sup>2</sup> 为 0.725, 线性回归模型的拟合优度 R<sup>2</sup> 为 0.519, 相比较, GAM 模型对 PM<sub>2.5</sub> 浓度变化的解释度提高了 20.6%。研究表明, GAM 模型对于建立 PM<sub>2.5</sub> 浓度变化与影响因素间综合性复杂关系更灵活、更可靠, 优于线性回归模型。

**关键词:** PM<sub>2.5</sub>; 北京; 影响因素; GAM

**中国分类号:** X513

**文献标识码:** A

## Exploring the factors influencing PM<sub>2.5</sub> concentration change in Beijing based on Generalized Additive Model

LIU Shan, LIU Di

(College of Earth Science and Engineering, Hohai University, Nanjing 211100, China)

**Abstract:** Based on the mechanism of PM<sub>2.5</sub> concentration change in Beijing in 2015, the generalized additive model (GMA) of PM<sub>2.5</sub> and atmospheric pollutants (PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, CO, O<sub>3</sub>), as well as the meteorological factors (daily mean temperature, wind scale, wind direction) are established to explore the impact of different factors on PM<sub>2.5</sub> concentration changes. The results show that: (1) Beijing PM<sub>2.5</sub> concentration has the characteristics of low distribution of summer and autumn, however higher in spring and winter; (2) The PM<sub>2.5</sub> concentration in Beijing is linearly positively correlated with PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, CO. The positive correlation is from strong to weak: CO> PM<sub>10</sub>> SO<sub>2</sub>> NO<sub>2</sub>, and the relationship with O<sub>3</sub>, temperature and wind factor is more complicated. (3) The goodness R<sup>2</sup> of the GAM model is 0.725, and the goodness of the linear regression model is 0.519. Compared with the GAM model, the explanatory degree of the PM<sub>2.5</sub> concentration increased by 20.6%. The results show that the GAM model is more flexible and reliable than the linear regression model in establishing the complex relationship between PM<sub>2.5</sub> concentration and influencing factors.

**Key words:** PM<sub>2.5</sub>; Beijing; influencing factors; GAM

随着 PM<sub>2.5</sub> 污染加剧, 以 PM<sub>2.5</sub> 为主要特征污染物的空气污染问题受到国家政府及广大学者的高度重视。国内外学者对 PM<sub>2.5</sub> 进行了大量的实验和研究, 目前主要包括 PM<sub>2.5</sub> 对人类健康的影响<sup>[1]</sup>, PM<sub>2.5</sub> 污染的扩散与传播<sup>[2]</sup>, PM<sub>2.5</sub> 污染物组成成分及其来源<sup>[3]</sup>, PM<sub>2.5</sub> 的时空分布特征<sup>[4]</sup>等方面的研究。

对于污染较严重的北京, 较多的研究工作主要着眼于 PM<sub>2.5</sub> 污染物特征及来源解析<sup>[5]</sup>, 而对于其影响因素的研究较少, 且一般研究中只考虑了其线性相关关系却忽略了不同变量间复杂的非线性关系。大量实验表明对于解释变量之间的非线性关系 GAM 优于线性回归模型, 且对于探索变量之间的关系及

收稿日期: 2017-03-28

基金项目: 国家自然科学基金资助项目 (41201394)

作者简介: 刘珊 (1993-), 女, 江西吉安人, 硕士, 主要从事生态与环境空间分析方面的研究。

预测更可靠、更灵活。因此,本研究以北京为研究区域,基于2015年北京PM<sub>2.5</sub>浓度时间演变模式,通过GAM模型建立PM<sub>2.5</sub>与各大气污染物(PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub>、CO、O<sub>3</sub>)及气象因素(日均温度、风力、风向)之间的非线性关系,探索不同变量对PM<sub>2.5</sub>的影响作用,为北京地区控制PM<sub>2.5</sub>浓度提供有力支持。

## 1 数据与方法

### 1.1 空气质量指标数据

以北京2015年PM<sub>2.5</sub>浓度的时间序列变化为研究对象,PM<sub>2.5</sub>数据为北京35个PM<sub>2.5</sub>监测站点2015年全年日均PM<sub>2.5</sub>值,研究变量中其他空气污染物指标(PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub>、CO、O<sub>3</sub>)也为2015年时间序列数据,均来源于中国环境监测总站。图1为北京35个PM<sub>2.5</sub>监测站点的空间位置分布图。其中27个为原有空气质量监测站点,5个为交通站点(属于污染源监测),3个为区域站点(监控外地传输过来的PM<sub>2.5</sub>污染),综合考虑其他影响因素数据的局限性,北京日均PM<sub>2.5</sub>数据由35个监测站点的日平均值计算而来。

### 1.2 气象数据

PM<sub>2.5</sub>主要来源于燃料燃烧、道路及土壤扬尘等,故本研究考虑的主要气象因素为日均温度、风力和风向3个指标。气象数据来源于中国气象数据网,为北京2015年时间序列数据。图2为2015年北京日均温度时间演变图,深色折线为日均温度,浅色竖线为温度日变化范围。由图2可知,全年日

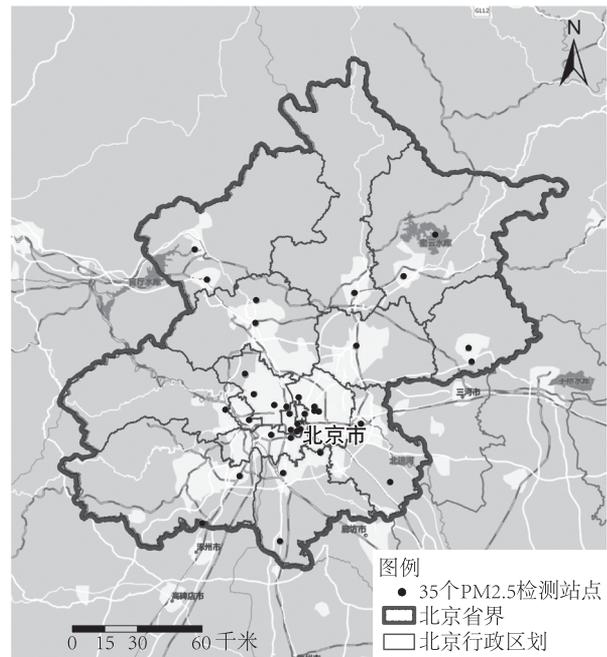


图1 PM<sub>2.5</sub>监测站点空间位置分布图

Fig.1 The distribution of BEIJING's PM<sub>2.5</sub> monitoring site

均温度变化呈正态分布,且全年温度日变化幅度整体差异不大。图3为2015年北京风向图,风力全年69.5%天数为3~4级,主要风向为西南风(35.3%)、东北风(21.6%)、东南风(13.6%)、西北风(12.3%)。

### 1.3 研究方法

本研究主要分析PM<sub>2.5</sub>浓度时间演变与潜在影响因素之间的关系,用线性回归模型描述这种复杂的综合性关系有所欠缺,而GAM模型具有解决因变量与影响因子之间高度非线性和非单调关系的突出能力,其针对不同的变量,采取不同的建模策略,可对部分影响因子进行线性拟合,而对其他因子通

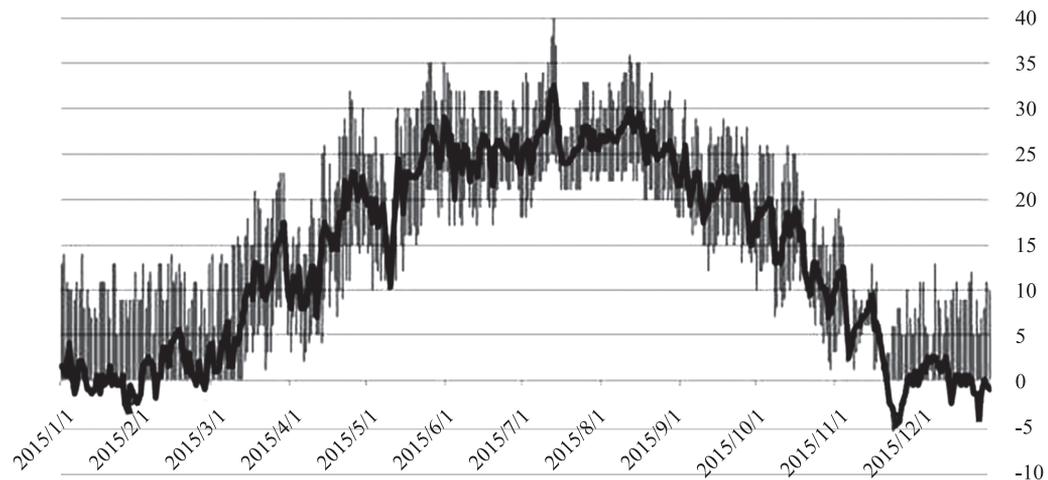


图2 2015年北京日均温度时间演变图

Fig.2 Time series of daily temperature of Beijing in 2015

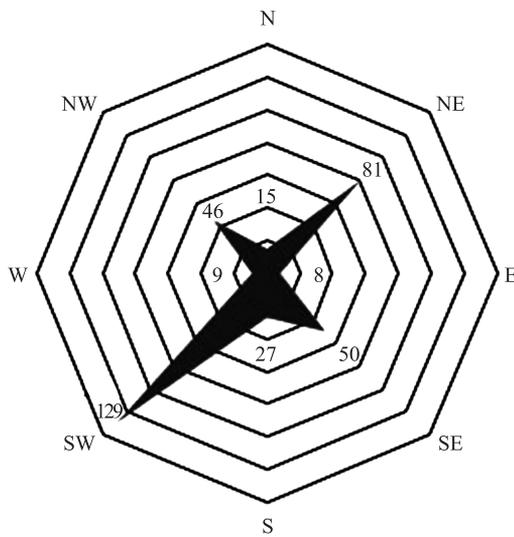


图 3 2015 年北京风向图  
Fig.3 Beijing wind map in 2015

过光滑函数进行拟合，故 GAM 模型在研究变量与影响因子之间的相关性具有高度的灵活性。该模型的公式为

$$y = B_0 + \sum_{i=1} B_i x_i + \sum_{j=1} f_j(x_j) + \sum_{k=1} f_k(x_{1,k}, x_{2,k}) + \varepsilon$$

其中： $y$  为因变量； $x_i$  为连续的线性变量； $x_j$  为连续的非线性变量； $x_{1,k}, x_{2,k}$  为连续的非线性变量对； $B_0, B_i$  为未知系数； $f_j, f_k$  是因变量与非线性变量间的非参数平滑函数。

在对 PM<sub>2.5</sub> 及其影响因素建模之前需要对数据进行探索性分析，对自变量和因变量分别进行 QQplot 验证，分析各数据是否满足正态分布，并对非正态分布的数据进行数据转换，使其满足正态分布特征，并剔除异常值，再对各个变量之间进行相关性和多重共线性分析，选择合理性变量进行建模。图 4 为各空气质量指标及温度因子与 PM<sub>2.5</sub> 相关关系的散点图，从图 4 可以得知，PM<sub>2.5</sub> 与 PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub>、CO 大体呈线性相关关系，而与 O<sub>3</sub> 和温度的关系更为复杂。

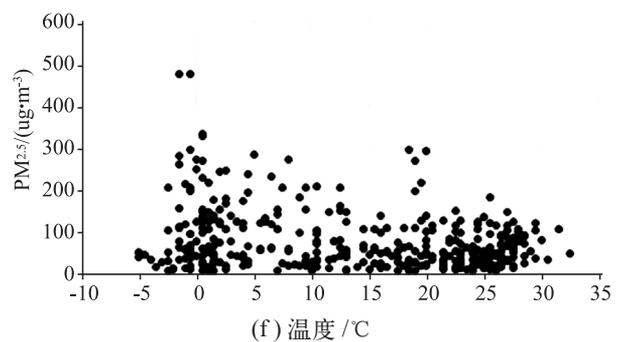
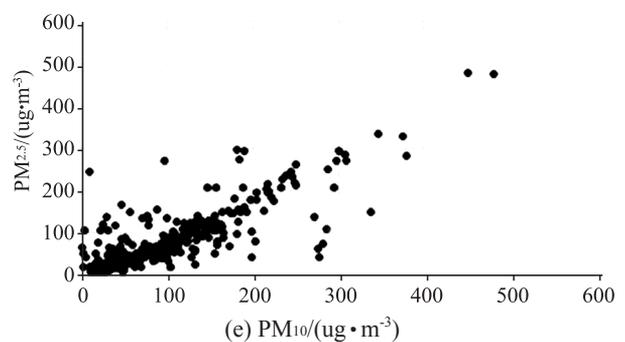
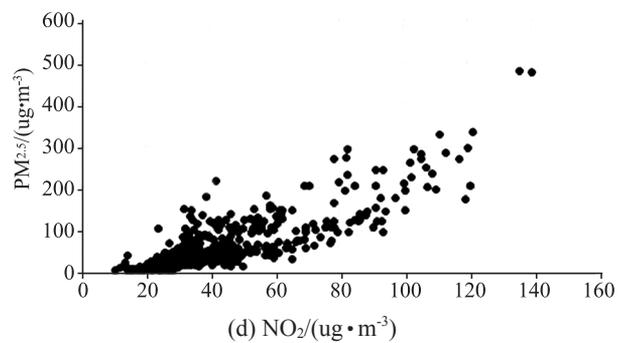
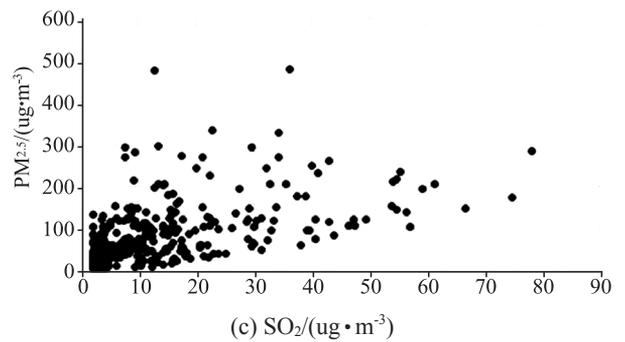
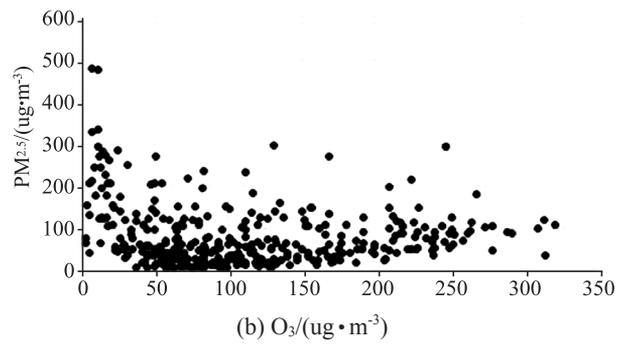
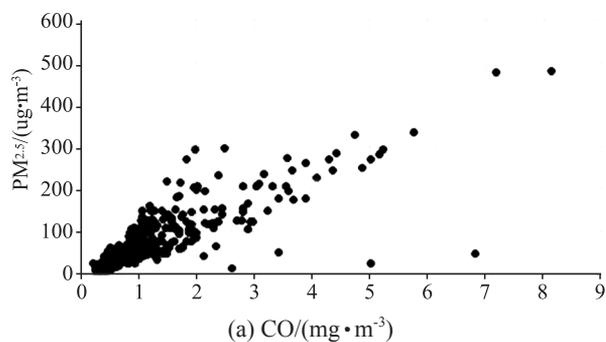


图 4 各空气质量指标及温度因子与 PM<sub>2.5</sub> 相关关系散点图  
Fig.4 The scatter plot between PM<sub>2.5</sub> concentration and air quality index, temperature factor.

由于风力、风向因子具有交互作用,根据相关研究,该模型中将风力、风向因子综合为风因子  $w_n$  ( $[$  风力  $\times \cos$  风向  $]$ ),  $w_m$  ( $[$  风力  $\times \sin$  风向  $]$ )。

## 2 结果与分析

### 2.1 PM<sub>2.5</sub> 时间演变模式

根据《环境空气质量指数(AQI)技术规定(试行)》(HJ 633—2012),表1为PM<sub>2.5</sub>浓度及空气质量分级图。图5为2015年北京PM<sub>2.5</sub>浓度时间演变图,图5中竖线表示北京35个监测站点PM<sub>2.5</sub>浓度日变化范围,点表示日均PM<sub>2.5</sub>值,折线表示连续7天PM<sub>2.5</sub>平均值。从图5可知5~9月北京空气质量整体偏好,各地区的PM<sub>2.5</sub>浓度差异较小,日均变化幅度偏小。1~5月和9~12月的空气质量较差,特别是整个冬季空气污染最严重,各地区PM<sub>2.5</sub>日均变化幅度偏大。这可能与冬季独特的气候条件和人类活动有关(如供暖等)。

### 2.2 PM<sub>2.5</sub> 模型

本研究中,综合考虑北京各空气质量指标和气象因素,这些变量从不同方面反映了空气质量,再分别对各个变量进行数据分析、数据处理(正态化、剔除异常值)以及分析各变量与PM<sub>2.5</sub>、各变量间的相关关系,构建其与PM<sub>2.5</sub>之间的关系模型,建立基于GAM的PM<sub>2.5</sub>模型:

$$\log(\text{PM}_{2.5}) = 3.18 + 1.09 \times 10^{-1} \times \text{PM}_{10} + 3.11 \times 10^{-2} \times \text{SO}_2 + 2.16 \times 10^{-3} \times \text{NO}_2 + 1.89 \times 10^{-1} \times \text{CO} + f_1(\log(\text{O}_3)) + f_2(T) + f_3(w_n, w_m) + \varepsilon \quad (2)$$

其中:  $f_1$ 、 $f_2$ 、 $f_3$  分别为PM<sub>2.5</sub>与O<sub>3</sub>、温度和风因子之间的非参数平滑函数。不同变量间的非线性

表1 PM<sub>2.5</sub>浓度及空气质量分级图  
Tab.1 Categories of air quality and corresponding PM<sub>2.5</sub> concentrations

空气质量等级	AQI	PM <sub>2.5</sub>	空气质量
1	0 ~ 50	0 ~ 35	优
2	51 ~ 100	36 ~ 75	良
3	101 ~ 150	76 ~ 115	轻度污染
4	151 ~ 200	116 ~ 150	中度污染
5	201 ~ 300	151 ~ 250	重度污染
6	>300	>250	严重污染

关系在模型中的平滑函数各不相同。图6为GAM模型中O<sub>3</sub>、温度对PM<sub>2.5</sub>浓度变化影响效应图。如图可知:(1)当log(O<sub>3</sub>)<4.4左右时,PM<sub>2.5</sub>浓度呈降低趋势;当log(O<sub>3</sub>)>4.4左右时,PM<sub>2.5</sub>浓度呈增加趋势,且增幅较大。(2)以0℃左右为临界点,随着温度的上升PM<sub>2.5</sub>浓度呈缓慢降低趋势,当温度达到25℃左右,PM<sub>2.5</sub>浓度呈缓慢增加趋势,故推测当温度在0℃左右或25℃左右,PM<sub>2.5</sub>聚集增强。

根据该模型可知PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub>、CO与PM<sub>2.5</sub>正线性相关,且正相关程度由强到弱为:CO>PM<sub>10</sub>>SO<sub>2</sub>>NO<sub>2</sub>,当PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub>、CO浓度增加时,将促进PM<sub>2.5</sub>浓度呈不同程度地增加。表2为各影响因子对PM<sub>2.5</sub>浓度变化的解释度。

从表2可知,2015年北京PM<sub>2.5</sub>浓度变化主要受CO、PM<sub>10</sub>、风因子的影响。CO对PM<sub>2.5</sub>浓度变化的解释度为23.74%,且呈正相关,北京人口聚集,工业发达,工业生产中矿物燃料燃烧、生活中固体废物焚烧以及汽车废气排放等行为产生了大量CO。PM<sub>10</sub>对PM<sub>2.5</sub>浓度变化的解释度为18.34%,由于建筑施工、道路施工扬尘等增加了PM<sub>10</sub>浓度,

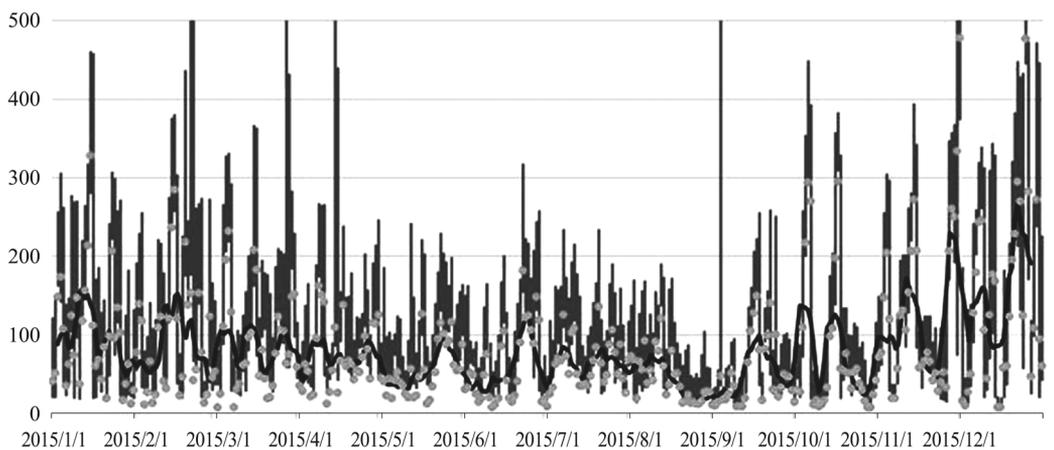


图5 2015年北京PM<sub>2.5</sub>浓度时间演变图

Fig.5 Time series of PM<sub>2.5</sub> concentrations of Beijing in 2015

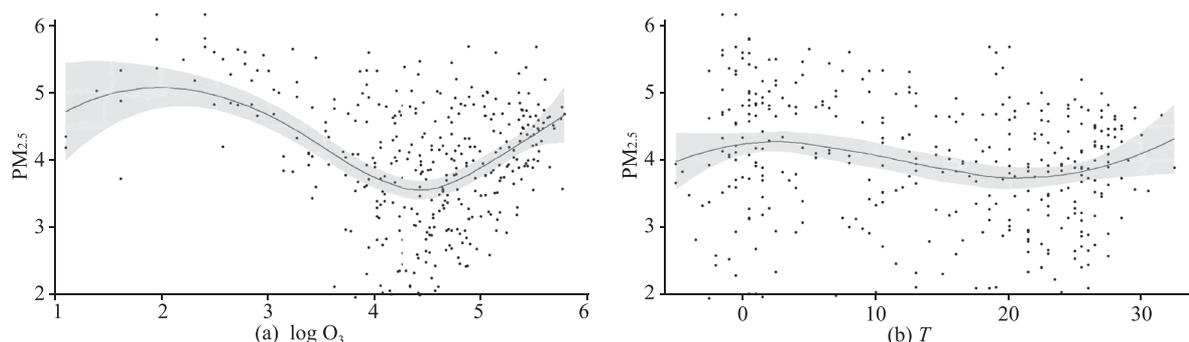
图 6 O<sub>3</sub>、温度对 PM<sub>2.5</sub> 浓度变化影响效应图

Fig.6 The effects of univariate smoothing functions for PM<sub>2.5</sub> concentration changes in the GAM: O<sub>3</sub> (left), temperature (right)

进而影响了 PM<sub>2.5</sub> 的浓度变化。风因子对 PM<sub>2.5</sub> 浓度变化的解释度为 12.89%，在一定的风力和风速条件下，PM<sub>2.5</sub> 浓度会有不同程度的转移扩散。其他影响因子的解释度相对较低，SO<sub>2</sub>、NO<sub>2</sub>、O<sub>3</sub> 在外界环境影响下产生化学反应进行二次转化转化为 PM<sub>2.5</sub>，北京独特的气候变化也潜在地影响了 PM<sub>2.5</sub> 的聚集扩散，其影响作用相对较小。

该模型的拟合优度  $R^2$  为 0.725，其残差符合正态分布，该模型对 PM<sub>2.5</sub> 浓度变化的解释度为 72.5%，为了评估该模型，本研究根据线性回归模型建立了 PM<sub>2.5</sub> 及各影响因素之间的线性关系，其拟合优度  $R^2$  为 0.519，GAM 与线性回归模型相比，对 PM<sub>2.5</sub> 浓度变化的解释度提高了 20.6%，GAM 对于建立变量间综合性复杂关系更灵活、更可靠。

### 3 结论

GAM 模型对 PM<sub>2.5</sub> 浓度变化的解释度优于线性回归模型，研究结果表明，2015 年北京 PM<sub>2.5</sub> 浓度变化主要受 CO、PM<sub>10</sub>、风因子等因素的影响，与

表 2 各影响因子对 PM<sub>2.5</sub> 浓度变化的解释度  
Tab.2 The influence of each factor on the change of PM<sub>2.5</sub> concentration

影响因子	解释度 /%
PM <sub>10</sub>	18.34
SO <sub>2</sub>	4.96
NO <sub>2</sub>	3.74
CO	23.74
O <sub>3</sub>	5.48
温度	3.35
风因子	12.89
模型综合	72.5

矿物燃料燃烧、钢铁冶炼、汽车废气排放以及建筑施工等人类行为密切相关，且受不同的气候因素的影响。

本次研究是基于时间序列数据进行的，缺少对 PM<sub>2.5</sub> 空间差异的研究；本次研究的影响因素只考虑了空气质量指标因素和气象因素，而 PM<sub>2.5</sub> 浓度变化的影响因素却更为复杂，潜在的可能因素还有，地形因素（高程）、人为因素（人口密度、城市化、工业生产、产业结构等）等；模型具有局限性，由于部分影响因子的数据特征不适用于该模型，导致与 PM<sub>2.5</sub> 有密切关系的影响因子无法纳入模型中进行研究，需要对模型加以改进和优化。

### 参考文献：

- [1]胡子梅, 王军, 陶征楷, 等. 上海市 PM<sub>2.5</sub> 重金属污染水平与健康风险评价 [J]. 环境科学学报, 2013 (12): 3399-3406.
- [2]薛文博, 付飞, 王金南, 等. 中国 PM<sub>2.5</sub> 跨区域传输特征数值模拟研究 [J]. 中国环境科学, 2014 (6): 1361-1368.
- [3]张静, 朱建强, 朱礼波. 兰州市大气 PM<sub>2.5</sub> 的化学组分特征及其来源研究 [J]. 环境工程, 2016 (5): 86-90.
- [4]王占山, 李云婷, 陈添, 等. 2013 年北京市 PM<sub>2.5</sub> 的时空分布 [J]. 地理学报, 2015 (1): 110-120.
- [5]李璇, 聂滕, 齐珺, 等. 2013 年 1 月北京市 PM<sub>2.5</sub> 区域来源解析 [J]. 环境科学, 2015 (4): 1148-1153.
- [6]贺祥, 林振山, 刘会玉, 等. 基于灰色关联模型对江苏省 PM<sub>2.5</sub> 浓度影响因素的分析 [J]. 地理学报, 2016 (7): 1119-1129.
- [7]贺祥, 林振山. 基于 GAM 模型分析影响因素交互作用对 PM<sub>2.5</sub> 浓度变化的影响 [J]. 环境科学, 2017 (1): 22-32.

(责任编辑 王利君)