

DOI: 10.3969/j.issn.1673-6141.2023.03.004

# 基于贝叶斯的大数据模型在大气污染成因分析中的应用

王莉君\*, 周玉, 万丽娟, 程亮亮

(合肥师范学院电子信息与电气工程学院, 安徽 合肥 230601)

**摘要:** 提出了一种基于多维高斯贝叶斯分类算法的复杂系统影响因素的分析方法, 并利用大数据方法建立了不同  $PM_{2.5}$  范围的分类模型, 结合马氏距离开展了影响因素的关键性分析。以合肥市 2013—2018 年间的天气与空气质量数据为基础, 筛选了  $PM_{10}$ 、 $SO_2$ 、 $NO_2$ 、 $CO$ 、 $O_3$  等 8 个大气污染的关键影响因素, 采用散点矩阵对  $PM_{2.5}$  与这些影响因素的相关性进行了分析。利用高斯贝叶斯分类器建立了基于 8 个主要参量的  $PM_{2.5}$  大气污染分析模型, 研究发现,  $PM_{2.5}$  与  $CO$  浓度具有较强的正相关性, 对  $NO_2$  具有选择性, 与  $O_3$  具有负相关性, 而  $CO$  与  $SO_2$  对  $PM_{2.5}$  的产生存在某种竞争机制。

**关键词:** 大数据技术; 高斯贝叶斯模型; 影响因素;  $PM_{2.5}$ ; 相关分析

中图分类号: TP391.4; X511

文献标识码: A

文章编号: 1673-6141(2023)03-227-008

## Application of the Bayesian-based big data model in the analysis of the source of air pollution

WANG Lijun\*, ZHOU Yu, WAN Lijuan, CHENG Liangliang

(School of Electronic Information and Electrical Engineering, Hefei Normal University, Hefei 230601, China)

**Abstract:** A method for analyzing influencing factors of complex systems based on multi-dimensional Gaussian Bayesian classification algorithm is proposed, classification models for diverse range of  $PM_{2.5}$  are established, and then the analysis of the key influencing factors on complex systems is carried out in combination with Mahalanobis distance. Based on the weather and air quality data of Hefei City from 2013 to 2018, 8 main influencing factors for  $PM_{2.5}$ , such as  $PM_{10}$ ,  $SO_2$ ,  $NO_2$ ,  $CO$ ,  $O_3$  and so on, are screened out, and then the correlation between  $PM_{2.5}$  and the influencing factors is analyzed using scatter matrix. The  $PM_{2.5}$  analysis model based on Gaussian Bayesian classifier is established with these data. It is found that  $PM_{2.5}$  has a strong positive correlation with  $CO$  concentration, is selective to  $NO_2$ , and has a negative correlation with  $O_3$ . As for  $CO$  and  $SO_2$ , a certain competitive mechanism between the two factors in the production of  $PM_{2.5}$  is observed.

**基金项目:** 安徽省教育厅省级教学示范课程建设项目 (2020SJXXSFK1991), 安徽省质量工程项目 (2020xsxxkc380), 电子信息仿真设计安徽省重点实验室项目 (2020ZDSYSYB05)

**作者简介:** 王莉君 (1976 - ), 女, 安徽亳州人, 硕士, 讲师, 主要从事信息处理、传感技术、大数据技术研究与应用。E-mail: lijwang@qq.com

**收稿日期:** 2022-02-08; **修改日期:** 2022-04-17

\*通信作者。

**Key words:** big data technology; Gaussian Bayesian model; influencing factors; PM<sub>2.5</sub>; correlation analysis

## 0 引言

近年来随着工业化进程的加快,国内主要城市空气质量先后经历了由好变差、再逐渐向好的过程。依据2012年新版空气质量标准《环境空气质量标准》(GB3095-2012),空气质量主要以PM<sub>10</sub>、PM<sub>2.5</sub>为代表的颗粒物和以NO<sub>2</sub>、SO<sub>2</sub>等为代表的气态污染物的浓度来进行表征<sup>[1,2]</sup>。在所有表征环境污染的指标中,细颗粒物PM<sub>2.5</sub>以其起因复杂、对人体潜在危害大等特点,近年来受到越来越多的关注<sup>[3,4]</sup>。

大气污染的影响因素十分复杂,人类活动向大气排放的各种污染物以及气象条件(如温度、气压、湿度等)都会对大气污染产生影响<sup>[5]</sup>。研究表明,PM<sub>2.5</sub>越高,患呼吸系统病症和心血管病的发病率越高,因此了解PM<sub>2.5</sub>的来源对控制PM<sub>2.5</sub>具有重要作用<sup>[6]</sup>。

随着数据科学技术的发展,采用大数据分析研究方法研究复杂系统的影响因素已成为可能<sup>[7]</sup>。大数据分析技术在复杂系统的模式识别、影响因素分析等方面具有独特优势<sup>[8]</sup>。本文以合肥地区2013-2018年间记录的天气及空气质量数据为基础<sup>[9]</sup>,采用数据可视化技术对PM<sub>2.5</sub>与PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub>等大气污染影响因素的相互关系开展分析<sup>[10]</sup>,筛选了与PM<sub>2.5</sub>相关的主要影响因素并建立了基于多维高斯贝叶斯分类的PM<sub>2.5</sub>预测模型,提出了PM<sub>2.5</sub>污染影响因素分析方法。利用该方法研究了主要因素对PM<sub>2.5</sub>浓度的影响规律,从而对PM<sub>2.5</sub>可能的成因进行探讨。

## 1 多维高斯贝叶斯算法

### 1.1 多维高斯贝叶斯分类

PM<sub>2.5</sub>的影响因素与污染等级可用多维高斯贝叶斯模型来描述。对于某PM<sub>2.5</sub>污染等级(记为C<sub>k</sub>),若有N个影响因素(记为x<sub>1</sub>,x<sub>2</sub>,...,x<sub>N</sub>),这些影响因素的联合分布概率密度记为p(x<sub>1</sub>,x<sub>2</sub>,...,x<sub>N</sub>|C<sub>k</sub>),即<sup>[8,11]</sup>

$$p(x_1, x_2, \dots, x_N | C_k) = \frac{1}{\sqrt{(2\pi)^N |\Sigma_k|}} \exp\left(-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu}_k) \cdot \Sigma_k^{-1} \cdot (\mathbf{X} - \boldsymbol{\mu}_k)^T\right), \quad (1)$$

式中 $\mathbf{X} = (x_1, x_2, \dots, x_N)$ 为由N个影响因素构成的向量, $\boldsymbol{\mu}_k$ 为第k个PM<sub>2.5</sub>污染等级中N个影响因素的均值向量, $\Sigma_k$ 为C<sub>k</sub>污染等级条件下各参数的协方差矩阵, $\Sigma_k^{-1}$ 为其逆阵。利用式(1)即可计算出在特定的污染等级C<sub>k</sub>发生时,N个影响因素 $\mathbf{X} = (x_1, x_2, \dots, x_N)$ 同时发生的联合分布概率密度。将同一组因素代入不同污染等级的概率密度模型中,取值最大的为最可能的污染等级,从而实现由影响因素到污染等级的分析。

### 1.2 基于马氏距离的影响因素关键性评估

式(1)给出了特定分类下N个影响因素的联合分布概率密度,由式(1)可知,对特定的分类,决定p(x<sub>1</sub>,x<sub>2</sub>,...,x<sub>N</sub>|C<sub>k</sub>)大小的主要是指数项 $-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu}_k) \cdot \Sigma_k^{-1} \cdot (\mathbf{X} - \boldsymbol{\mu}_k)^T$ ,即马氏距离。

令 $y = -(\mathbf{X} - \boldsymbol{\mu}_k) \cdot \Sigma_k^{-1} \cdot (\mathbf{X} - \boldsymbol{\mu}_k)^T$ ,可以通过求y对X中任一因素x<sub>i</sub>的偏导数,从而获取该因素x<sub>i</sub>对y的影响规律<sup>[8]</sup>:

$$\frac{\partial y}{\partial x_i} = -\frac{\partial(\mathbf{X} - \boldsymbol{\mu}_k)}{\partial x_i} \cdot \Sigma_k^{-1} \cdot (\mathbf{X} - \boldsymbol{\mu}_k)^T - (\mathbf{X} - \boldsymbol{\mu}_k) \cdot \Sigma_k^{-1} \cdot \frac{\partial(\mathbf{X} - \boldsymbol{\mu}_k)}{\partial x_i}^T. \quad (2)$$

对于特定的污染等级,偏导数 $\frac{\partial y}{\partial x_i}$ 的绝对值越大,说明参数 $x_i$ 对影响因素向量 $\mathbf{X}$ 隶属于该污染等级的影响最大。若 $x_i$ 是原始影响参量时,则说明该参量对特定的分类的表征能力影响越强,即越关键。

## 2 数据准备

为构建 $\text{PM}_{2.5}$ 污染等级分类模型,从天气后报网站(<http://www.tianqihoubao.com/weather/top/hefei.html>)下载了2013—2018年间的合肥市天气和空气质量数据用于建模与分析。天气数据包括日照、风向、风力等文本型数据,空气质量数据包括空气质量指数(AQI)、城市空气质量指数排名(#AQI)、 $\text{PM}_{2.5}$ 、 $\text{PM}_{10}$ 、 $\text{SO}_2$ 、 $\text{NO}_2$ 、 $\text{CO}$ 、 $\text{O}_3$ 、日最高温度(Max temp.)、日最低温度(Min temp.)等数值型数据。为便于分析,对全部数据采用以下方法进行预处理:

- 1) 考虑与大气污染直接相关的一次数据,不考虑AQI、#AQI等评估污染程度的二次数据;
- 2) 暂不考虑天气、风向、风力等文本型数据;
- 3) 采用数字化方法以自然日对日期进行编号;
- 4) 考虑样本数据的完整性,对数据不完整的样本进行剔除;
- 5) 参考《环境空气质量指数(AQI)技术规定(试行)》,依据 $\text{PM}_{2.5}$ 的数值将空气污染状况分为优、良、轻度污染、中度污染、重度污染、严重污染共6级<sup>[12,13]</sup>。

依据上述方法处理后,得到2013—2018年间1873个日历日的天气和空气质量数据,如表1所示。

表1 2013—2018年间合肥市天气及空气质量情况数据收集表

Table 1 Data collection table for weather and air quality of Hefei City from 2013 to 2018

$\text{PM}_{10}$ concentration/ ( $\mu\text{g}\cdot\text{m}^{-3}$ )	$\text{SO}_2$ concentration/ ( $\mu\text{g}\cdot\text{m}^{-3}$ )	$\text{NO}_2$ concentration/ ( $\mu\text{g}\cdot\text{m}^{-3}$ )	$\text{CO}$ concentration/ ( $\mu\text{g}\cdot\text{m}^{-3}$ )	$\text{O}_3$ concentration/ ( $\mu\text{g}\cdot\text{m}^{-3}$ )	Max temp. / $^{\circ}\text{C}$	Min temp. / $^{\circ}\text{C}$	Date number	$\text{PM}_{2.5}$ level
200	37	44	1.35	42	24	11	298	3
247	38	55	1.58	46	24	15	299	3
88	16	27	0.87	46	21	15	300	2
55	15	35	1.08	25	18	15	301	2
...	...	...	...	...	...	...	...	...

## 3 影响因素及相关性分析

### 3.1 $\text{PM}_{2.5}$ 与影响因素的关系分析

$\text{PM}_{2.5}$ 与各影响因素的相关性采用散点矩阵进行分析<sup>[14]</sup>,利用开源算法计算了合肥市2013—2018年间的 $\text{PM}_{2.5}$ 与主要影响因素( $\text{PM}_{10}$ 、 $\text{SO}_2$ 、 $\text{NO}_2$ 、 $\text{CO}$ 、 $\text{O}_3$ 、Max temp.、Min temp.及日期)的相互关系散点矩阵如图1所示。

由图可知,单一影响因素与其他影响因素之间的相互关系如下:

- 1) 日期:不同日期的温度分布呈“^”形状,越靠近年中 $\text{O}_3$ 浓度越大。日期对 $\text{PM}_{10}$ 、 $\text{SO}_2$ 、 $\text{NO}_2$ 以及 $\text{CO}$ 的影响表现为夏季时这些污染物浓度相对较小且分布越集中,冬季时这些污染物浓度相对较高且越分散,这一发

现与陈相如<sup>[15]</sup>的研究结果相似。

2) 温度: 日最高、最低温度呈现显著的线性关系。除 $O_3$ 外一般情况下温度越高污染物浓度越小, 且分布越集中, 温度越低污染物浓度越分散。而温度对 $O_3$ 的影响呈现相反的特征, 即温度越高 $O_3$ 的含量越高。

3)  $O_3$ :  $PM_{2.5}$ 严重污染时通常 $O_3$ 含量较低, 而当 $O_3$ 浓度较高时,  $CO$ 、 $NO_2$ 、 $SO_2$ 以及 $PM_{10}$ 浓度通常较小, 但 $O_3$ 对这些污染物的作用是直接还是间接的还有待进一步分析。

4)  $CO$ :  $CO$ 对 $PM_{10}$ 、 $SO_2$ 、 $NO_2$ 的影响呈现一定的正相关性, 即 $CO$ 浓度越高这些污染物的含量越高。 $PM_{2.5}$ 浓度较高时通常 $CO$ 含量较高, 可见 $CO$ 可能与 $PM_{2.5}$ 的发生具有某种直接关系。

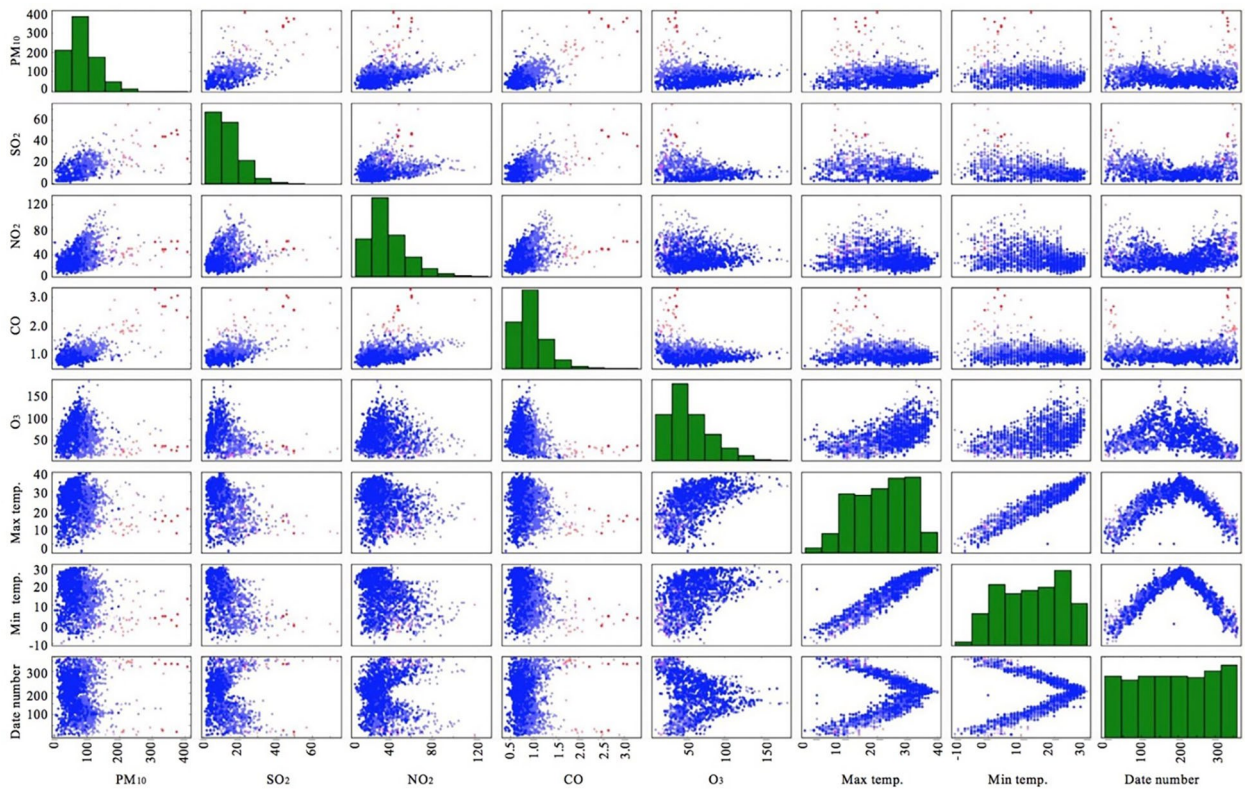


图1  $PM_{2.5}$ 的8个主要影响因素及其相互关系散布矩阵

Fig. 1 8 main influence factors to  $PM_{2.5}$  and their interrelation scatter matrix

5)  $NO_2$ :  $NO_2$ 浓度较高时 $CO$ 、 $SO_2$ 、 $PM_{10}$ 等污染物的浓度通常也较高, 而 $NO_2$ 浓度在 $40\sim 60\ \mu\text{g}/\text{m}^3$ 附近时 $PM_{2.5}$ 一般较高时, 这种现象在其他文献中亦有报道<sup>[16]</sup>。

6)  $SO_2$ : 在 $SO_2$ 、 $PM_{10}$ 、 $NO_2$ 、 $CO$ 的相互关系中,  $SO_2$ 对其他污染物的影响与 $CO$ 对其他污染物的影响相似, 发生 $PM_{2.5}$ 严重污染时通常 $SO_2$ 含量较高。

7)  $PM_{10}$ :  $PM_{10}$ 浓度与 $SO_2$ 、 $NO_2$ 以及 $CO$ 的浓度具有显著的相关性, 一般 $PM_{10}$ 越大这些污染物的含量越高,  $PM_{2.5}$ 污染也较为严重。

样本中全部1873天的大气污染分布情况如图2所示。在1873组数据中, 需选择一定数量的数据用于训练和建模, 其余数据作为检验。本研究采用随机抽样方法, 选择70%的天气数据用于训练, 30%的数据用于检验, 利用 $PM_{10}$ 、 $SO_2$ 等8个参量, 采用多维高斯贝叶斯模型对 $PM_{2.5}$ 进行分析。

利用多维高斯贝叶斯模型对8个因素共同作用下的 $PM_{2.5}$ 浓度情况进行分析, 得到562个检验样本的



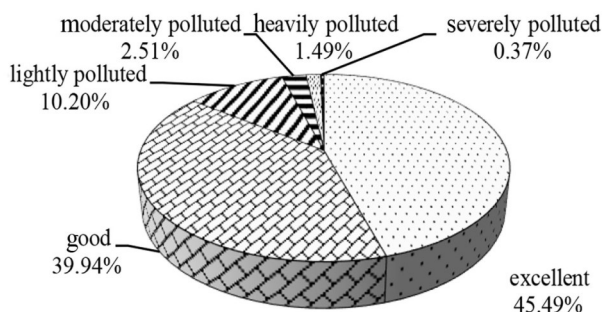


图 2 2013—2018 年间合肥市 PM<sub>2.5</sub> 大气污染天数分布情况

Fig. 2 Distribution of PM<sub>2.5</sub> air pollution in Hefei from 2013 to 2018

PM<sub>2.5</sub>浓度预测数据。分析表明,模型对从优到严重污染共6种由PM<sub>2.5</sub>表征的污染等级进行预测的正确识别率为92.70%。不同PM<sub>2.5</sub>污染等级的预测结果和混淆矩阵分别如表2和表3所示,其中PM<sub>2.5</sub>等级(0~4)分别代表空气质量优、良、轻度污染、中度污染和重度污染,由于待分析的样本中无严重污染的样本,故表2和表3中未出现严重污染(PM<sub>2.5</sub>等级5级)的数据。表2表明,模型对PM<sub>2.5</sub>为优与良的分析结果准确率较高,随着污染等级增加,识别率有所下降。重度污染时模型的识别率为50%,其原因可能是由于样本数量不足所致(全部1873个样本中,重度污染天数仅为28天)。此外,从表3分析结果的混淆矩阵可以发现,相邻的污染等级容易发生误识别。

表 2 合肥市 PM<sub>2.5</sub> 污染模型分析结果

Table 2 Analysis results of PM<sub>2.5</sub> pollution model in Hefei City

PM <sub>2.5</sub> pollution level	Number of days	Correctly recognized days	Recognition rate/%
0	376	363	96.54
1	141	125	88.65
2	38	29	76.32
3	5	3	60.00
4	2	1	50.00

表 3 模型分析结果混淆矩阵

Table 3 Confusion matrix for model analysis results

PM <sub>2.5</sub> pollution level	0	1	2	3	4
0	0.97	0	0.16	0	0
1	0.03	0.89	0.03	0.40	0
2	0	0.09	0.76	0.00	0.50
3	0	0.02	0.03	0.60	0
4	0	0	0.02	0	0.50

### 3.2 影响因素的关键性分析

利用第1.2节提出的方法,对PM<sub>2.5</sub>大气污染的主要影响因素进行分析,结果如表4所示。由表可知,对PM<sub>2.5</sub>严重污染影响最大的为PM<sub>10</sub>、CO和O<sub>3</sub>,其中PM<sub>10</sub>和CO的偏导数数值为正,说明其浓度越高形成PM<sub>2.5</sub>严重污染的可能性越大,O<sub>3</sub>的偏导数数值为负,说明其浓度越高越不易发生严重污染。对重度及中度污染,PM<sub>10</sub>和CO的影响逐渐下降,但仍占主导地位,而SO<sub>2</sub>的影响逐渐增加,O<sub>3</sub>的影响仍为负值,NO<sub>2</sub>的影响为负值,说明NO<sub>2</sub>较低时发生PM<sub>2.5</sub>中度及重度污染的可能性高。对轻度污染,SO<sub>2</sub>与PM<sub>10</sub>的影响较大。

从表4还可以看出,PM<sub>2.5</sub>对SO<sub>2</sub>、NO<sub>2</sub>以及CO这类污染物具有某种选择性。对于严重及重度污染,CO浓度高时发生PM<sub>2.5</sub>污染的可能性大,但这种影响随着PM<sub>2.5</sub>浓度的降低逐渐下降,说明CO可能对PM<sub>2.5</sub>升高具有某种促进作用。而对于SO<sub>2</sub>而言,随着PM<sub>2.5</sub>浓度的升高,SO<sub>2</sub>的影响逐渐减弱。可见,SO<sub>2</sub>与CO对PM<sub>2.5</sub>的产生可能具有某种竞争机制。而NO<sub>2</sub>对PM<sub>2.5</sub>的影响比较复杂,对于轻度、中度及重度污染而言,较低的NO<sub>2</sub>含量有利于PM<sub>2.5</sub>的生成,而对严重污染,较高的NO<sub>2</sub>浓度促进了PM<sub>2.5</sub>的生成。

表4 PM<sub>2.5</sub>大气污染的主要影响因素的马氏距离偏导数

Table 4 Main influence factors to PM<sub>2.5</sub> pollution and their partial derivatives of Mahalanobis distance

PM <sub>2.5</sub> pollution level	PM <sub>10</sub>	SO <sub>2</sub>	NO <sub>2</sub>	CO	O <sub>3</sub>	Max temp.	Min temp.	Date number
5 (Severely polluted)	0.516	0.067	0.171	0.481	-0.991	0.096	0.270	-0.115
4 (Heavily polluted)	0.234	0.173	-0.302	0.371	-1.417	-0.026	0.211	-0.050
3 (Moderated polluted)	0.262	0.137	-0.247	0.375	-1.332	0.014	0.257	-0.068
2 (Lightly polluted)	0.437	1.085	-0.124	0.097	0.157	-0.256	-0.580	0.234

## 4 结 论

以PM<sub>2.5</sub>大气污染为对象,采用大数据分析开展研究,得到以下结论:

1) 提出一种多维高斯贝叶斯模型结合马氏距离估算的大数据分析方法用于大气污染分析,利用数据可视化技术筛选了PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub>、CO、O<sub>3</sub>等8个对PM<sub>2.5</sub>具有关键影响的参量;

2) 研究表明,CO、NO<sub>2</sub>、SO<sub>2</sub>与PM<sub>2.5</sub>浓度具有较强的相关性,控制这些污染物的生成对降低PM<sub>2.5</sub>具有明显作用。O<sub>3</sub>的浓度对PM<sub>2.5</sub>污染具有负相关性;

3) 建立了基于多维高斯贝叶斯的PM<sub>2.5</sub>分类模型,模型对PM<sub>2.5</sub>污染等级的预测正确率达到92.70%。但由于重度污染、严重污染样本数量偏少,模型识别率偏低;

4) 分析表明,PM<sub>2.5</sub>对CO、NO<sub>2</sub>、SO<sub>2</sub>具有较强的选择性,其中以NO<sub>2</sub>最为显著,NO<sub>2</sub>浓度在40~60 μg/m<sup>3</sup>时PM<sub>2.5</sub>通常较高。CO与SO<sub>2</sub>对PM<sub>2.5</sub>的影响可能具有某种竞争机制。

**参考文献:**

- [1] General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Ministry of Environmental Protection of the People's Republic of China. Ambient air quality standard: GB 3095—2012 [S]. Beijing: China Environmental Science Press, 2012.  
国家质量监督检验检疫总局, 环境保护部. 环境空气质量标准: GB 3095—2012 [S]. 北京: 中国环境科学出版社, 2012.
- [2] Luo Y P, Chen T, Ding H, *et al.* Characteristics analysis of PM<sub>2.5</sub> and NO<sub>2</sub> pollution at Beijing traffic stations during the Spring Festival from 2014 to 2018 [J]. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, 2020, 59(1): 50-63.  
罗银萍, 陈同, 丁卉, 等. 2014—2018年春节期间北京交通站PM<sub>2.5</sub>及NO<sub>2</sub>污染特征分析 [J]. 中山大学学报 (自然科学版), 2020, 59(1): 50-63.
- [3] Zhou Y M, Zhao X Y. Correlation analysis between PM<sub>2.5</sub> concentration and meteorological factors in Beijing area [J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2017, 53(1): 111-124.  
周一敏, 赵昕奕. 北京地区PM<sub>2.5</sub>浓度与气象要素的相关分析 [J]. 北京大学学报 (自然科学版), 2017, 53(1): 111-124.
- [4] Huang L K, Xu Y Y, Wang G Z, *et al.* Study on pollution characteristics and toxic effects of fine particulate matter in Harbin, China [J]. *China Environmental Science*, 2019, 39(12): 5326-5332.  
黄丽坤, 徐媛媛, 王广智, 等. 哈尔滨市大气细颗粒物污染特征及其毒性效应 [J]. 中国环境科学, 2019, 39(12): 5326-5332.
- [5] Liu K K, Zhang H, Liu G J. Elemental composition characteristics of PM<sub>2.5</sub> and PM<sub>10</sub>, and heavy pollution analysis in Hefei [J]. *Environmental Science*, 2019, 40(8): 3415-3420.  
刘可可, 张红, 刘桂建. 合肥市PM<sub>2.5</sub>和PM<sub>10</sub>中元素组成特征及重污染成因分析 [J]. 环境科学, 2019, 40(8): 3415-3420.
- [6] Zhang J Y, Tian L, Li S M, *et al.* Relationship between ambient PM<sub>2.5</sub> exposure and outpatient number for respiratory diseases in Chaoyang District, Beijing, China: A time series analysis [J]. *Journal of Environmental Hygiene*, 2020, 10(4): 367-371.  
张金艳, 田霖, 李书明, 等. 北京市朝阳区大气PM<sub>2.5</sub>暴露与医院呼吸系统疾病门诊量关系的时间序列分析 [J]. 环境卫生学杂志, 2020, 10(4): 367-371.
- [7] Li W, Hu H, Xu F C, *et al.* Studies on big data analysis technology used in environmental sciences [J]. *Environmental Monitoring in China*, 2015, 31(3): 118-122.  
李蔚, 胡昊, 徐富春, 等. 大数据解析技术在大气环境监测中的应用研究 [J]. 中国环境监测, 2015, 31(3): 118-122.
- [8] Zhu J X, Chen X D, Lü B L, *et al.* Smart failure/fault diagnosis and influence analysis for mechanical equipment with multivariate Gaussian Bayesian method [J]. *Journal of Mechanical Engineering*, 2020, 56(4): 35-41.  
朱建新, 陈学东, 吕宝林, 等. 基于多维高斯贝叶斯的机械设备失效/故障智能诊断及参数影响分析 [J]. 机械工程学报, 2020, 56(4): 35-41.
- [9] Past-weather website. Historical weather query of Hefei City [OL]. <http://www.tianqihoubao.com/weather/top/hefei.html>.  
天气后报网站. 合肥历史天气查询 [OL]. <http://www.tianqihoubao.com/weather/top/hefei.html>.
- [10] McKinney W. *Python for Data Analysis* [M]. Sebastopol, CA: O'Reilly, 2012.
- [11] Zhang J, Yuan M, Liu Y Y. The Bayes classifier based on the normal mixture model and its application [J]. *Acta Mathematicae Applicatae Sinica*, 2020, 43(4): 742-755.  
张婧, 袁敏, 刘妍岩. 基于正态混合模型的贝叶斯分类方法及其应用 [J]. 应用数学学报, 2020, 43(4): 742-755.
- [12] Ministry of Environmental Protection of the People's Republic of China. Technical regulation on ambient air quality index (on trial): HJ 633—2012 [S]. Beijing: China Environmental Science Press, 2012.  
环境保护部. 环境空气质量指数 (AQI) 技术规定: HJ 633—2012 [S]. 北京: 中国环境科学出版社, 2012.
- [13] Han F R, Lu X, Wu T Y, *et al.* Characteristics of air pollution in Jinhua City and its relationship with meteorological

- conditions [J]. *Journal of Zhejiang Meteorology*, 2021, 42(4): 13-18.
- 韩芙蓉, 鹿翔, 吴天贻, 等. 金华市大气污染特征及与气象条件的关系分析 [J]. *浙江气象*, 2021, 42(4): 13-18.
- [14] Liu X H. *High-Dimensional Data Analysis Based on Scatter Plot Classification* [D]. Tianjin: Tianjin University, 2019.
- 刘训瀚. 基于散点图模式探索的高维数据分析 [D]. 天津: 天津大学, 2019.
- [15] Chen X R. Research on characteristics of air pollutant emission changes in eco-city based on big data analysis [J]. *Environmental Science and Management*, 2022, 47(1): 60-63.
- 陈相如. 基于大数据分析的生态城市大气污染物排放变化特征研究 [J]. *环境科学与管理*, 2022, 47(1): 60-63.
- [16] Xu D, Zhang Q, Fan G Q, *et al.* Characteristics of particles pollution based on lidar network in Hangzhou and surrounding areas [J]. *Journal of Atmospheric and Environmental Optics*, 2019, 14(3): 171-178.
- 徐达, 张全, 范广强, 等. 基于激光雷达观测网的杭州及周边地区颗粒物污染特征研究 [J]. *大气与环境光学学报*, 2019, 14(3): 171-178.