

基于气象因子的 $PM_{2.5}$ 回归预测模型研究

汪宇¹, 彭晓武², 沈劲¹, 嵇萍¹, 邓滢¹, 谢敏¹, 陈多宏^{1*}

(1. 广东省环境监测中心, 国家环境保护区域空气质量监测重点实验室, 广东 广州 510308; 2. 环境保护部华南环境科学研究所, 广东 广州 510655)

摘要:采用 Pearson 相关系数分析了 2013—2016 年 3 大典型城市北京、南京和广州的 $\rho(PM_{2.5})$ 与各气象因子的关系。结果表明, 3 个城市 $\rho(PM_{2.5})$ 与各风速因子最大的相关系数依次为 -0.44 , -0.29 和 -0.37 , 与各气温因子最大的相关系数依次为 -0.44 , -0.33 和 -0.37 , 气压与南京和广州的 $\rho(PM_{2.5})$ 正相关, 气压因子最大的相关系数分别为 0.25 和 0.34 , 湿度与北京 $\rho(PM_{2.5})$ 正相关, 与广州 $\rho(PM_{2.5})$ 负相关, 湿度因子最大的相关系数分别为 0.49 和 -0.36 , 日照时数与北京 $\rho(PM_{2.5})$ 相关系数为 -0.46 , 降水量与南京和广州 $\rho(PM_{2.5})$ 相关系数分别为 -0.20 和 -0.24 ; 采用逐步线性回归方法建立城市次日 $\rho(PM_{2.5})$ 与气象因子的预测模型, 复合相关系数分别为 $0.722\ 8$, $0.770\ 6$ 和 $0.809\ 9$ 。模型预测 3 个城市 2016 年 $PM_{2.5}$ 年均值分别偏高 4 , 5 和 $3\ \mu g/m^3$, 日均值平均相对误差为 $\pm 45.6\%$, $\pm 32.9\%$ 和 $\pm 26.0\%$, 模型对高 $\rho(PM_{2.5})$ 普遍低估。

关键词: 细颗粒物; 气象因子; 相关性; 线性回归; 北京; 南京; 广州

中图分类号: X513

文献标志码: B

文章编号: 1674-6732(2018)04-0008-04

Research on the Regression Model of $PM_{2.5}$ Concentration Based on Meteorological Parameters

WANG Yu¹, PENG Xiao-wu², SHEN Jin¹, JI Ping¹, DENG Ying¹, XIE Min¹, CHEN Duo-hong^{1*}

(1. State Environmental Key Laboratory of Regional Air Quality Monitoring, Guangdong Environmental Monitoring Center, Guangzhou, Guangdong 510308, China; 2. South China Institute of Environmental Sciences, Guangzhou, Guangdong 510655, China)

Abstract: Using Pearson correlation coefficient, relationship between $PM_{2.5}$ concentration [$\rho(PM_{2.5})$] and meteorological factors in three typical cities Beijing, Nanjing and Guangzhou were analyzed during 2013 to 2016. The results showed that the maximum correlation coefficient between $\rho(PM_{2.5})$ and wind speed factor in the three cities was -0.44 , -0.29 and -0.37 in turn, and the maximum correlation coefficient was -0.44 , -0.33 and -0.37 for the temperature factor. Atmospheric pressure was positively correlated with $\rho(PM_{2.5})$ in Nanjing and Guangzhou, and the maximum correlation coefficient of the pressure factor was 0.25 and 0.34 , respectively. Humidity was positively correlated with $\rho(PM_{2.5})$ in Beijing but negatively correlated with $\rho(PM_{2.5})$ in Guangzhou, with the maximum correlation coefficient of humidity factor 0.49 and -0.36 , respectively. The correlation coefficient between sunshine hours and $\rho(PM_{2.5})$ of Beijing was -0.46 . The correlation coefficient between precipitation and $\rho(PM_{2.5})$ in Nanjing and Guangzhou was -0.20 and -0.24 , respectively. The prediction model of next day $\rho(PM_{2.5})$ and meteorological factors was established by the stepwise linear regression method, with the composite correlation coefficients being $0.722\ 8$, $0.770\ 6$ and $0.809\ 9$ respectively. Annual average $\rho(PM_{2.5})$ values in 2016 are overestimated by 4 , 5 and $3\ \mu g/m^3$, while the average relative errors for daily mean are $\pm 45.6\%$, $\pm 32.9\%$ and $\pm 26.0\%$, respectively. The model generally underestimated the high value of $\rho(PM_{2.5})$.

Key words: $PM_{2.5}$; Meteorological factor; Correlation; Linear regression; Beijing; Nanjing; Guangzhou

大气中的 $\rho(PM_{2.5})$ 与排放源、气象条件等因素有关。研究表明, 不利气象条件是造成 $PM_{2.5}$ 污染事件发生的重要原因, 由于污染物背景浓度、排放源特征、气候条件、地形等差异, 影响不同地区的

收稿日期: 2018-01-10; 修订日期: 2018-03-19

基金项目: 国家科技支撑计划基金资助项目(2014BAC21B04); 国家自然科学基金面上资助项目(21477045)

作者简介: 汪宇(1991—), 男, 硕士, 工程师, 从事空气质量预报工作。

* 通信作者: 陈多宏 E-mail: chenduohong@139.com

$\rho(PM_{2.5})$ 的主要气象因子有所差异^[1-3]。排放源强在短期内的变化量级不大,且精细化的实时排放源数据无法全面有效地获取,因此对 $\rho(PM_{2.5})$ 构建回归模型时,缺少能直接表征污染源排放变化的相关因子,主要考虑气象等因子的影响^[4-6]。识别主要影响气象因子有助于加深对各地大气污染形势的认识,辅助空气质量预报^[7],预判大气污染过程,采取大气污染防治措施。

现选取了北京、南京和广州3个城市,比较分析影响这3个城市的主要气象因子,并采用逐步回归方法建立 $PM_{2.5}$ 与气象参数等相关性模型,对次日 $\rho(PM_{2.5})$ 进行预测,并运用多种方式对预报结果进行检验。

1 研究方法

1.1 数据来源

2013—2016年 $\rho(PM_{2.5})$ 数据来自北京、南京和广州市环保局网站上每日公布空气质量日报。气象数据来自国家气象科学数据共享服务平台(<http://data.cma.cn/>)。气象参数 $X_1—X_{13}$ 分别表示:平均风速、最大风速、极大风速、平均本站气压、日最低本站气压、日最高本站气压、平均气温、日最低气温、日最高气温、平均相对湿度、最小相对湿度、日照时数、20:00—次日20:00降水量。

1.2 分析方法

为了便于比较不同量级的数据之间的相关性,需对 $\rho(PM_{2.5})$ 和气象数据进行离差标准化处理。采用Pearson相关系数法^[8]计算各气象因子与 $\rho(PM_{2.5})$ 的相关性,按气象要素类别筛选对 $\rho(PM_{2.5})$ 影响较大的气象因子。气象条件对 $\rho(PM_{2.5})$ 的影响包括:①水平风的扩散稀释作用;②垂直方向的扩散稀释作用;③气温影响气溶胶化学反应;④水汽影响气溶胶化学反应;⑤降雨的湿清除作用;等。基于初步筛选后的气象因子和前一日 $\rho(PM_{2.5})$ 因子,采用多元线性回归的方法^[8],保留影响显著的气象因子,构建3个城市气象驱动因子的 $\rho(PM_{2.5})$ 预报模型。

2 气象因子相关性分析

表1为气象参数与城市 $PM_{2.5}$ 日均值的相关性。由表1可见, $X_1—X_3$ 与各城市 $\rho(PM_{2.5})$ 呈现较强负相关。风速越大,扩散条件越好, $\rho(PM_{2.5})$ 越低。风速因子影响最大的是北京,其次是广州。

$X_4—X_6$ 与 $\rho(PM_{2.5})$ 正相关。气压升高可能与垂直扩散条件变差有关。气压因子对北京的影响非常小,对广州的影响大于对南京的影响。 $X_7—X_9$ 与 $\rho(PM_{2.5})$ 负相关。温度升高往往伴随大气湍流活动加强,利于 $PM_{2.5}$ 扩散。整体而言,气温因子对南京和广州的影响高于北京。 X_{10} 和 X_{11} 与北京 $\rho(PM_{2.5})$ 呈现正相关,而与广州 $\rho(PM_{2.5})$ 呈现负相关,与南京相关性较弱。

表1 3个城市 $PM_{2.5}$ 日均值与各参数相关系数

自变量		北京	南京	广州
风速	X_1	-0.37	-0.21	-0.21
	X_2	-0.41	-0.22	-0.34
	X_3	-0.44	-0.29	-0.37
气压	X_4	0.05	0.25	0.33
	X_5	0.08	0.25	0.33
	X_6	0.06	0.24	0.34
气温	X_7	-0.18	-0.30	-0.30
	X_8	-0.14	-0.33	-0.37
	X_9	-0.44	-0.25	-0.17
湿度	X_{10}	0.49	-0.07	-0.22
	X_{11}	0.46	-0.15	-0.36
日照	X_{12}	-0.46	-0.03	0.10
降水	X_{13}	-0.06	-0.20	-0.24

3个城市不同相对湿度区间中 $\rho(PM_{2.5})$ 、 X_3 和 X_{13} 见表2。由表2可见,与低相对湿度的因子相比,高相对湿度的因子,北京 X_3 明显下降,扩散能力下降,南京和广州的 X_3 变化不大。 X_{13} 与北京的 $PM_{2.5}$ 日均值负相关性较高,与南京和广州的相关性较低。 X_{13} 对北京的 $PM_{2.5}$ 日均值影响较小,而与南京和广州的相关性较高,可能与北京降水多以小雨为主,湿清除能力不足有关,而在南京和广州,降水量增多加强了湿清除作用。

表2 3个城市不同相对湿度区间中 $\rho(PM_{2.5})$ 、 X_3 和 X_{13}

相对湿度因子		$\geq 80\%$	79% ~ 60%	59% ~ 40%	39% ~ 20%	$< 20\%$
$\rho(PM_{2.5})/(\mu g \cdot m^{-3})$	北京	131	108	75	43	20
	南京	43	59	63	71	72
	广州	34	36	44	65	60
$X_3/(0.1 m \cdot s^{-1})$	北京	61	65	83	100	123
	南京	87	82	81	82	91
	广州	83	91	81	78	81
$X_{13}/(0.1 mm \cdot d^{-1})$	北京	90	13	4	0	0
	南京	199	72	12	3	2
	广州	248	102	21	1	0

考虑同一类的因子两两之间具有较强的正相关性,仅选取相关性最高的 X_3 代表3个城市水平扩散条件因子。南京选取 X_5 , 广州选取 X_6 作为表征垂直扩散条件因子。北京选取 X_9 , 南京和广州选取 X_8 , 作为表征气温的因子。北京选取 X_{10} , 南京和广州选取 X_{11} 作为表征湿度的因子。北京选取 X_{12} , 南京和广州选取 X_{13} 作为表征晴雨的因子。

3 构建回归模型

根据上一步选取的气象因子,加入 X_{14} 因子[前一日 $\rho(\text{PM}_{2.5})$],使用2013—2015年观测数据,采用逐步回归方法,根据回归参数检验,剔除不显著的回归因子,得到各城市多元线性回归预报方程。

3.1 北京市

基于气象因子 X_3 、 X_9 、 X_{10} 、 X_{12} 和 X_{14} , 得到北京 $Y(\text{PM}_{2.5}$ 日均值) 的回归方程如下:

$$Y = 0.147\ 1 - 0\ X_3 - 0.258\ 2\ X_9 + 0.087\ 6\ X_{10} - 0.069\ 9\ X_{12} + 0.472\ 1\ X_{14} \quad (1)$$

因 X_3 与 X_9 相关性近似达到1,同时引入 X_3 和 X_9 进入回归方程,造成其中一个系数为零。保留 X_3 , 剔除 X_9 , 重新建立回归方程如下:

$$Y = 0.144\ 4 - 0.255\ 6\ X_3 + 0.087\ 6\ X_{10} - 0.070\ 0\ X_{12} + 0.472\ 1\ X_{14} \quad (2)$$

3.2 南京市

基于气象因子 X_3 、 X_5 、 X_8 、 X_{11} 、 X_{13} 和 X_{14} , 得到南京 Y 的回归方程如下:

$$Y = 0.260\ 7 - 0.212\ X_3 - 0.106\ 8\ X_5 - 0.127\ 4\ X_8 - 0.020\ 2\ X_{11} - 0.168\ 3\ X_{13} + 0.660\ 7\ X_{14} \quad (3)$$

X_3 、 X_5 、 X_8 、 X_{13} 、 X_{14} 回归因子的 t 检验 P 值均 <

0.01。 X_{11} 未通过 P 值为0.05的 t 检验,剔除后,重新建立回归方程如下:

$$Y = 0.253\ 7 - 0.209\ 4\ X_3 - 0.108\ 0\ X_7 - 0.132\ 8\ X_8 - 0.193\ 5\ X_{13} + 0.660\ 8\ X_{14} \quad (4)$$

3.3 广州市

基于气象因子 X_3 、 X_6 、 X_8 、 X_{11} 、 X_{13} 和 X_{14} , 得到广州 Y 的回归方程如下:

$$Y = 0.362\ 5 - 0.257\ X_3 - 0.085\ 0\ X_6 - 0.167\ 7\ X_8 - 0.077\ 1\ X_{11} - 0.010\ 4\ X_{13} + 0.633\ X_{14} \quad (5)$$

X_3 、 X_6 、 X_8 、 X_{11} 、 X_{14} 5个回归因子的 t 检验 P 值均 <0.01。 X_{13} 未通过 P 值为0.05的 t 检验,剔除后,重新建立回归方程如下:

$$Y = 0.361\ 8 - 0.259\ 2\ X_3 - 0.083\ 0\ X_6 - 0.166\ 4\ X_8 - 0.079\ 1\ X_{11} + 0.633\ 0\ X_{14} \quad (6)$$

4 回归模型的检验

分别使用回归方程(2)(4)和(6),预测2016年北京、南京和广州的 $\text{PM}_{2.5}$ 日均值,将实测值与预测值对比,得到结果见图1(a)(b)(c)。回归模型预测的3城市 $\rho(\text{PM}_{2.5})$ 与实测值较为吻合,北京、南京和广州的 $\text{PM}_{2.5}$ 预测的年均值分别偏高4,5,和3 $\mu\text{g}/\text{m}^3$,日均值平均误差分别为 ± 33 , ± 16 和 $\pm 9\ \mu\text{g}/\text{m}^3$,日均值平均相对误差为 $\pm 45.6\%$, $\pm 32.9\%$ 和 $\pm 26.0\%$ 。在实测值较低时,北京预测值略偏高。回归模型预测的南京和广州 $\rho(\text{PM}_{2.5})$ 与实测值较为接近,实测值较低时,预测值没有明显高估。当北京、南京、广州实测 $\rho(\text{PM}_{2.5})$ 分别 > 200,150 和 100 $\mu\text{g}/\text{m}^3$ 时,预测值全部偏低。

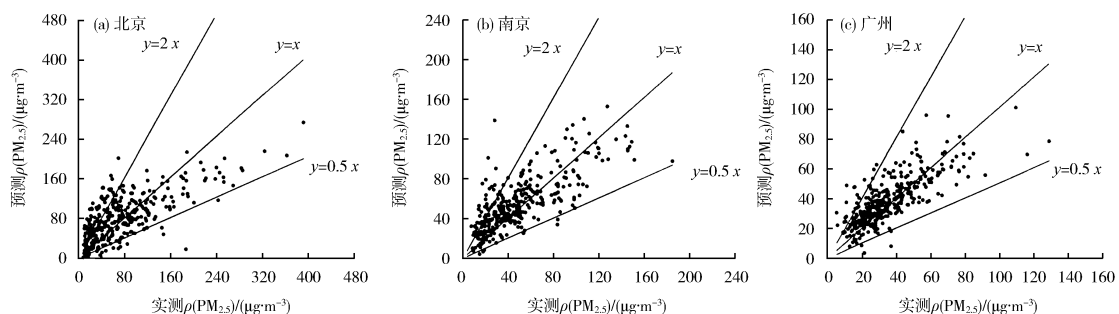


图1 实测与回归模型预测 $\rho(\text{PM}_{2.5})$ 散点图

3个城市回归方程检验参数见表3。由表3可见,北京、南京和广州各因子复合相关系数分别为0.722 8,0.770 6和0.809 9,各城市的回归因子均

通过了 P 值为0.01的 t 检验,说明 $\rho(\text{PM}_{2.5})$ 与各因子高度相关。根据调整后的 R^2 值,所选取的回归因子分别可以解释北京、南京和广州 $\text{PM}_{2.5}$ 日均

值变化的 52.07%、59.20% 和 65.60%。3 个城市回归方程的标准误差分别为 0.101 7、0.083 4 和 0.098 4,说明回归方程拟合程度比较好。3 个城市的 F 检验值均 <0.01 ,说明方程回归效果显著。

表 3 3 个城市回归方程检验参数

参数	北京	南京	广州
复合相关系数	0.722 8	0.770 6	0.809 9
回归因子 t 检验 P 值	<0.01	<0.01	<0.01
调整后 R^2	0.520 7	0.592 0	0.656 0
标准误差	0.101 7	0.083 4	0.098 4
显著性 F 值	<0.01	<0.01	<0.01

5 结论

研究了气象因子与北京、南京和广州 $PM_{2.5}$ 日均值的相关性。风速与 3 个城市 $\rho(PM_{2.5})$ 呈负相关,影响程度大小分别为:北京 $>$ 广州 $>$ 南京。气压与南京和广州的 $\rho(PM_{2.5})$ 正相关,对北京影响较小,对广州影响大于南京。气温与 3 个城市 $\rho(PM_{2.5})$ 负相关,对南京和广州的影响程度大于北京。湿度与北京 $\rho(PM_{2.5})$ 正相关,而与广州 $\rho(PM_{2.5})$ 负相关,与南京相关性较小。日照时数与北京 $\rho(PM_{2.5})$ 正相关,降水量与南京和广州 $\rho(PM_{2.5})$ 负相关。

根据主要气象影响因子构建的回归模型都能较好地预测 $\rho(PM_{2.5})$,但是对 $\rho(PM_{2.5})$ 高值存在低估。一方面,由于线性回归模型对样本的依赖,而高 $\rho(PM_{2.5})$ 样本数量不多,故对高浓度的模拟

效果表现不佳;另一方面,高 $\rho(PM_{2.5})$ 污染过程中, $\rho(PM_{2.5})$ 增长迅速,呈现非线性增长特性,故线性模型的模拟会普遍低估 $\rho(PM_{2.5})$ 的高值。总体而言,应用实测的 $\rho(PM_{2.5})$ 数据和气象因子数据样本,构建 $PM_{2.5}$ 线性模型,通过引入天气形势预报的气象因子,来预测 $\rho(PM_{2.5})$,仍然是一种可用且高效的方法。后续研究可以考虑采用分位数回归的方法^[9],使气象因子回归模型能够更好地捕捉到高 $\rho(PM_{2.5})$ 。

[参考文献]

- [1] 赵晨曦,王云琦,王玉杰,等. 北京地区冬春 $PM_{2.5}$ 和 PM_{10} 污染水平时空分布及其与气象条件的关系[J]. 环境科学, 2014,35(2):418-427.
- [2] 姜迪,李聪. 南京市 $PM_{2.5}$ 扩散与气象条件的关系[J]. 环境监测管理与技术,2016,28(1):36-40.
- [3] 朱倩茹,刘永红,徐伟嘉,等. 广州 $PM_{2.5}$ 污染特征及影响因素分析[J]. 中国环境监测,2013,29(2):15-21.
- [4] 徐伟嘉,李红霞,黄慎,等. 珠海市空气质量预测模型研究与分析[J]. 环境科学与技术,2015,38(11):142-147,158.
- [5] 刘闯,王帅,林宏,等. 沈阳市冬季环境空气质量统计预报模型建立及应用[J]. 中国环境监测,2014,30(4):10-15.
- [6] 许杨,王凯,贾桥莲,等. 2007—2008 年武汉市空气质量预报及检验[J]. 气象与环境学报,2012,28(2):81-84.
- [7] 叶斯琪,陈多宏,谢敏,等. 珠三角区域空气质量预报方法及预报效果评估[J]. 环境监控与预警,2016,8(3):10-13.
- [8] 魏凤英. 现代气候统计诊断与预测技术[M]. 2 版. 北京:气象出版社, 2007: 37-40, 106-117.
- [9] 陈建宝,丁军军. 分位数回归技术综述[J]. 统计与信息论坛, 2008(3):89-96.

· 简讯 ·

《自然》气候科学报告称 —— 碳捕获封存是缓解气候变化有效工具

科技日报北京 6 月 18 日电 英国《自然·通讯》杂志近日发表的一项气候科学研究报告称,碳捕获地质封存将是一种有效的缓解气候变化工具。全新开发的存储安全计算器表明:在管理得当的情况下,98% 的注入地下的二氧化碳(CO_2)在地下保留 10000 年以上的概率为 50%;在管理欠佳的情况下,78% 的注入地下的 CO_2 可以保留在地下。

碳捕获和封存涉及将工业生产过程中的化石燃料释放的 CO_2 注入地下。尽管该技术有望减少全球 CO_2 排放,并有助于实现《巴黎协定》目标——把全球平均气温升幅控制在工业化前水平以上低于 $2^\circ C$ 之内,并努力将气温升幅限制在工业化前水平以上 $1.5^\circ C$ 之内。但一直有人担心,注入地下的 CO_2 会在未来泄漏到大气中。

此次,英国阿伯丁大学研究人员朱安·阿卡尔德及其同事,开发了一个计算程序——存储安全计算器(Storage Security Calculator),以评估地质 CO_2 封存对于减缓气候变化的有效性。该计算器量化了 2020 年至 2050 年注入地下的 CO_2 的固定量,以及 10000 年内向大气中泄漏的 CO_2 总量。研究人员计算了两种场景,一种是注入地密度适中,并使用当前最佳实践对场地进行管理,另一种是场地管理不善的情况。他们发现,如果场地得到适当管理,年度泄漏率低于 0.01%,许多人认为这是有效减缓气候变化的可接受限度。对于管理不当的场地,虽然在头 100 年里泄漏率超过了这个限值,但在 1000 年后,它会降低到可接受的水平。

摘自 www. hbt. jiangsu. gov. cn 2018-06-25