

基于时间序列和多元方法的乌鲁木齐 $\text{PM}_{2.5}$ 浓度分析*

谢心庆, 郑薇, 开璇, 许英

(新疆财经大学 应用数学学院, 新疆 乌鲁木齐 830012)

摘要: 鉴于肺部可吸入颗粒物 $\text{PM}_{2.5}$ 对人体的危害, 利用多元分析及时间序列的方法将 $\text{PM}_{2.5}$ 浓度变化划分为稳定部分(由分子内部的作用引起浓度变化)和不稳定部分(由外部环境因素, 即温度, 风力, 风向及天气等)进行预测. 收集乌鲁木齐 7 个监测站点 2014 年 11 月至 2015 年 3 月(冬季)每天的 $\text{PM}_{2.5}$, PM_{10} , CO , NO_2 , O_3 , O_3 (8 h), SO_2 及天气等相关因素数据, 对 $\text{PM}_{2.5}$ 浓度建立预测模型并进行结果分析. 相关分析表明: 户外 $\text{PM}_{2.5}$ 与 PM_{10} , CO , NO_2 和 SO_2 具有较高的相关性. 对平稳部分利用指数平滑模型预测 $\text{PM}_{2.5}$ 浓度, 得到最好的平滑指数是 0.32. 主成分回归(PCR)模型用于预测 $\text{PM}_{2.5}$ 浓度的不稳定性成分, 得到 R^2 值为 0.803. 最终, 将 2015 年 3 月至 2015 年 5 月的数据利用 5 种性能指标检验模型, 结果表明该模型方法预测效果较好, 有一定的实用价值.

关键词: $\text{PM}_{2.5}$; 大气污染; 指数平滑; PCR

中图分类号: X 513; X 701.2

文献标志码: A

文章编号: 0258-7971(2016)04-0595-07

$\text{PM}_{2.5}$ 是空气质量标准中颗粒物(粒径 $\leq 2.5 \mu\text{m}$) 浓度限值, 主要由水溶性离子、颗粒有机物和微量元素等组成. 近年来, 大量研究发现 $\text{PM}_{2.5}$ 因其粒径较小、相对 PM_{10} 比表面积较大, 因此更容易富集空气中的有机污染物、酸性氧化物、有毒重金属、细菌和病毒. 当 $\text{PM}_{2.5}$ 通过呼吸进入血液时, 就可以产生并导致人体呼吸、内分泌、心血管、神经及免疫等各系统疾病的发生^[1-2]. 我国雾霾现象的出现主要源于二次颗粒物中的 $\text{PM}_{2.5}$, 形成 $\text{PM}_{2.5}$ 的前体物包括二氧化硫(SO_2), 氮氧化物(NO), 挥发性有机化合物(VOCs), 氨气(NH_3)等^[3-4]. 鉴于我国各地雾霾现象的频繁出现及 $\text{PM}_{2.5}$ 对人体的危害, $\text{PM}_{2.5}$ 的研究也是刻不容缓的.

近几年来, 有关 $\text{PM}_{2.5}$ 的相关研究有很多, 在 2000 年 Rinehart 等^[5]在美国加利福尼亚地区建立 20 个观测站, 用以研究 $\text{PM}_{2.5}$ 的空气质量; Appel^[6], Senaratne 等^[7]对 $\text{PM}_{2.5}$ 有关因素进行了分析; Schichtel^[8]对城市 $\text{PM}_{2.5}$ 的危害做了调查研究. 2004 年, 杨复沫等对北京地区 $\text{PM}_{2.5}$ 中的化学元素、水溶性离子和含碳组分进行分析^[9]; 2013 年曲

直对城市大气环境中 $\text{PM}_{2.5}$ 源解析成分谱的建立进行研究^[10]; 张朝能等用克里格插值法、非参数分析法对 $\text{PM}_{2.5}$ 浓度时空特征及气象因素的影响进行分析^[11]. 但已有预测方法上研究甚少, 有仅利用 PCR 方法分析 $\text{PM}_{2.5}$ 浓度, 利用户外 PM_{10} 、 $\text{PM}_{2.5}$ 、臭氧浓度和气象变量来预测室内的污染物 $\text{PM}_{2.5}$ 浓度^[12-15], 预测结果不仅缺少不稳定成分的分析, 而且预测结果仅能达到实际结果的 35%~50%, 预测精度欠佳. 因此, 本文利用 PCR 与时间序列相结合的方法对乌鲁木齐市 $\text{PM}_{2.5}$ 浓度进行分析, 结果表明此法不仅减少预测的不稳定性, 又增加了预测的精度.

1 方法模型

1.1 描述研究区域 乌鲁木齐位于 43.45°N , 87.36°E . 地处北天山西段与东段的结合部分, 东、南、西三面环山, 地势东南高、西北低, 北为准噶尔盆地南缘. 海拔 680~920 m, 市区平均海拔 800 m. 乌鲁木齐属中温带大陆性干旱气候, 春秋两季较

* 收稿日期: 2015-12-17

基金项目: 新疆自治区研究生科研项目基金“乌鲁木齐 $\text{PM}_{2.5}$ 浓度的动态分析研究”(XGJRI2015120).

作者简介: 谢心庆(1988-), 女, 河南人, 硕士生, 主要从事统计学研究. E-mail: celebration.1988@163.com.

通信作者: 许英(1981-), 女, 新疆人, 副教授, 硕士生导师, 主要从事经济图论、统计学方面研究. E-mail: celebration.1988@163.com.

短,冬夏两季较长,昼夜温差大.年平均降水量为 194 mm,最暖的 7、8 月平均气温为 25.7 ℃,最冷的 1 月平均气温为-15.2 ℃.极端气温最高 47.8 ℃,最低-41.5 ℃.乌鲁木齐深处大陆腹地,气候特点是:温差大,寒暑变化剧烈;降水少,且随高度垂直递增;冬季寒冷漫长,四季分配不均,冬季有逆温层出现.乌鲁木齐市因受地理位置及天气的影响,冬季温度低且时间漫长,在长期的室内供暖系统的影响下,造成冬季污染严重,空气污染严重的主要的污染源是污染物的排放^[16].

1.2 样本的地理位置 乌鲁木齐市共有 7 个空气质量监测站,分别是总监测站、收费所监测站、农科院农场监测站、铁路局监测站、31 中学监测站、74 中学监测站和米东区环保局监测站.这 7 个监测站位于乌鲁木齐的不同区域及户外人数最多的时间如表 1.

表 1 各监测站特点
Tab.1 Each monitoring station characteristics

名称	位置	户外活动人数最多时间
监测站	新市区	8:00—11:00 及 18:00—21:00
收费所	天山区	全天
农科院农场	沙依巴克区	全天
铁路局	新市区	全天
31 中学	水磨沟区	9:00—20:00
74 中学	头屯河区	9:00—20:00
米东区环保局	米东区	8:00—11:00 及 18:00—21:00

1.3 样本收集方法 收集了乌鲁木齐 2014 年 11 月—2015 年 3 月每天 9:00—11:00 共 7 个监测站数据,中国环境监测总站公布的乌鲁木齐实时数据及风力、风向、温度和天气等数据.

各监测站收集到的数据包括: AQI, PM_{2.5}, PM₁₀, CO, NO₂, O₃, SO₂.

粒子质量浓度 (PM₁₀ 和 PM_{2.5}) 使用自动监测监控设备检测其浓度 (F701-20). 使用环境空气质量自动监测系统 (EAQM-100) 监测 SO₂ 浓度、CO 浓度、NO₂ 的浓度和 O₃ 浓度.

1.4 统计方法与建模方法

1.4.1 统计方法简要描述 在数据分解之前,为了解决数据量纲不同的问题,在建模前将数据进行标准化处理^[17-18]. 将处理后的数据进行变量间的相关性分析^[19].

1.4.1.1 指数平滑预测^[20] 为解决 PM_{2.5} 浓度稳定部分的预测,使用指数平滑的方法进行数据预测.指数平滑通过计算指数平滑值,配合一定的时间序列预测模型对现象的未来进行预测.其原理是任一期的指数平滑值都是本期实际观察值与前一期指数平滑值的加权平均.指数平滑法的预测方程表示如下:

$$s_t = \alpha y_t + (1 - \alpha) s_{t-1}, \tag{1}$$

式中, y_t 是实际值, s_t 是由指数平滑得到的数值, $\alpha (0 < \alpha < 1)$ 是平滑指数.

预测方程如下:

$$\hat{y}_{t+1} = s_t, \tag{2}$$

这样, $\hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \hat{y}_t, \tag{3}$

即当期的预测值是前一期的实际值与预测值的加权平均.

1.4.1.2 主成分回归模型 主成分回归主要是利用提取主成分的方法建立回归方程. 表达方程如下:

$$y = b_0 + b_1 P_1 + b_2 P_2 + \cdots, b_k P_k + \varepsilon, \tag{4}$$

$$P_i = a_{1i} x_1 + a_{2i} x_2 + \cdots, a_{ni} x_n, \tag{5}$$

这里 b_k 为回归系数, P_k 是提取的主成分, x_i 为变量, a_{ni} 为相对于 x_n 的载荷.

1.4.1.3 模型检验的性能指标 预测的性能指标涉及计算观察值 y 与预测值 $\hat{y}$ 之间的误差.这里采用 5 种性能指标进行模型的检验,分别是规范化的绝对误差 (NAE), 均方根误差 (RMSE), 预测精度 (PA), 可决系数 (R^2) 和性能 IA. 一个好的模型 NAE 和 RMSE 值应接近 0, 而 PA, R^2 和 IA 应该接近 1.

本文进行数据的分析,使用统计软件包括 Eviews, SPSS (Statistical Package for Social Science, version 20).

(1) 规范化的绝对误差 (NAE) 规范化的平均绝对误差是预测值与观测值之间的差异在所有情况下除以观测值的总和 (Gervasi 2008)^[21]. 计算方程如下:

$$NAE = \frac{\sum_{i=1}^N |P_i - O_i|}{\sum_{i=1}^N O_i}, \tag{6}$$

其中 N 是样本的数量, O_i 是实际观测的 PM_{2.5} 浓度值, P_i 是预测得到的 PM_{2.5} 浓度。

(2) 均方根误差 (RMSE) 均方根误差是用来测量数值预测的精度标准^[22-23]. 计算方程如下:

$$\text{RMSE} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (P_i - O_i)^2}, \quad (7)$$

其中 N 是样本的数量, O_i 是实际观测的 PM_{2.5} 浓度值, P_i 是预测得到的 PM_{2.5} 浓度。

(3) 可决系数 (R^2) 可决系数是回归平方和 (ESS) 在总变差 (TSS) 中所占的比重, 可决系数可以作为综合度量回归模型对样本观测值拟合优度的度量指标^[24]. 具体计算公式如下:

$$R^2 = \left(\frac{\sum_{i=1}^N (P_i - \bar{P})(O_i - \bar{O})}{N \cdot S_p \cdot S_o} \right)^2, \quad (8)$$

其中 N 是样本的数量, O_i 是实际观测的 PM_{2.5} 浓度值, P_i 是预测得到的 PM_{2.5} 浓度, $\bar{P}$ 是 P_i 的均值, $\bar{O}$ 为 O_i 的均值, S_p 是 P_i 的标准差, S_o 是 O_i 的标准差。

(4) 性能 IA 性能 IA 是鉴于以上方法的改进 (Gervasi, 2008)^[21]. 具体计算公式如下:

$$\text{IA} = 1 - \left[\frac{\sum_{i=1}^N (\bar{P} - O_i)^2}{\sum_{i=1}^N (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \right], \quad (9)$$

其中 N 是样本的数量, O_i 是实际观测的 PM_{2.5} 浓度值, P_i 是预测得到的 PM_{2.5} 浓度, $\bar{P}$ 是 P_i 的均值, $\bar{O}$ 为 O_i 的均值。

(5) 预测精度 (PA) 预测精度是判断预测值与实际值的接近程度的指标^[21]. 具体计算公式如下:

$$\text{PA} = \frac{\sum_{i=1}^N (P_i - \bar{P})}{\sum_{i=1}^N (O_i - \bar{O})}, \quad (10)$$

其中 N 是样本的数量, O_i 是实际观测的 PM_{2.5} 浓度值, P_i 是预测得到的 PM_{2.5} 浓度, $\bar{P}$ 是 P_i 的均值, $\bar{O}$ 为 O_i 的均值。

1.4.2 建模方法 主要思想是利用时间序列的分解思想将各监测站点 PM_{2.5} 浓度分解成稳定部分 (由分子之间的作用引起浓度变化) 和不稳定部分 (由外部环境因素, 即由于温度和风力等因素影响). 浓度预测的稳定部分可由历史数据预测分析

得到, 不稳定部分由外力的主成分回归得到, 将 2 部分加权建立 PM_{2.5} 浓度的预测模型。

$$\text{PM}_{2.5} = R^2 \cdot \text{Exp}_{\text{预}} + (1 - R^2) \cdot \text{PCR}_{\text{预}}, \quad (11)$$

这里的 $\text{Exp}_{\text{预}}$ 为 PM_{2.5} 浓度稳定部分的预测值 (这部分通过指数平滑得到), $\text{PCR}_{\text{预}}$ 为 PM_{2.5} 浓度不稳定部分的预测值 (这部分由主成分回归得到). R^2 为权重系数。

$$\text{Exp}_{\text{预}t+1} = \alpha \times \text{Exp}_{\text{实}t} + (1 - \alpha) \times \text{Exp}_{\text{预}t}, \quad (12)$$

利用 (12) 式 PM_{2.5} 浓度稳定部分的指数平滑模型, 得到 (1) 式中的 $\text{Exp}_{\text{预}}$

$$\text{Exp}_{\text{实}} = X_{\text{PM}_{2.5}} - \hat{X}_{\text{PM}_{2.5}}, \quad (13)$$

(13) 式中, $\text{Exp}_{\text{实}}$ 是由 7 个监测站每天的实际值减去预测值的均值得到的。

$$\hat{X}_{\text{PM}_{2.5}} = \sum_{i=1}^n a_i x_i, \quad (14)$$

(14) 式为各监测站的回归模型, x_i 为与 PM_{2.5} 强相关的各监测站变量的观测值, a_i 为回归系数, 此式建模得到的 R^2 为 (11) 式中的权重系数。

$$\text{PCR}_{\text{预}} = \hat{X}_{\text{PM}_{2.5}} = b_0 + \sum_{i=1}^k b_i P_i, \quad (15)$$

(15) 式中, $\text{PCR}_{\text{预}}$ 为主成分回归得到的值, 是由 $\hat{X}_{\text{PM}_{2.5}}$ 与中国环境监测总站公布的乌鲁木齐实时各变量数据进行主成分回归得到的, P_i 为提取的主成分, b_i 为回归系数。

$$P_i = \sum_{i=1}^n a_{li} x_i, \quad (16)$$

(16) 式中, a_{li} 为相对于 x_i 的载荷, x_i 为与 PM_{2.5} 强相关的中国环境监测总站公布数值。

2 结果及讨论

2.1 统计描述 利用 SPSS 数据分析软件进行数据分析整理 (见表 2)。

在研究时段所有监测站中, 平均最高的 PM_{2.5} 浓度为 115.86 (±64.62) μg/m³ (农业科学院的农场), 平均最低的 PM_{2.5} 浓度为 53.84 (±37.47) μg/m³ (收费所). 其余变量值见表 2。

各变量相关关系值见表 3. 由表 3 可知, PM_{2.5} 分别与 PM₁₀、CO、NO₂ 和 SO₂ 均有较高的相关关系。

2.2 指数平滑预测 对由实际数据剔除不稳定部分后, 得到的稳定部分 (即随机成分) 进行指数平

表 2 描述统计
Tab.2 Describes the statistical

									μg/m ³
地点	属性	AQI	PM _{2.5}	PM ₁₀	CO	NO ₂	O ₃	O ₃ (8 h)	SO ₂
监测站	min	11	1	4	0.41	4	1 000	1 000	2
	max	457	327	557	7.07	151	57 000	61 000	86
	mean	131.71	87.17	140.29	2.45	57.55	8 830	9 460	28.46
	std	101.28	69.55	123.48	1.38	32.16	10 530	11 030	20.72
收费所	min	23	4	15	0.27	3	1 000	1 000	1
	max	415	204	515	6.71	137	56 000	60 000	92
	mean	93.76	53.84	121.98	2.34	71.14	7 610	8 330	16.04
	std	59.29	37.47	90.34	1.44	30.44	11 310	10 720	15.15
农科院农场	min	33	21	33	0.15	2	1 000	1 000	2
	max	489	306	589	7.07	122	33 000	36 000	119
	mean	163.31	115.86	206.57	2.73	57.32	8 740	8 860	30.9
	std	87.63	64.62	117.28	1.45	29.83	6 490	6 630	22
铁路局	min	7	2	7	0.02	2	1 000	1 000	1
	max	476	459	576	8.5	146	46 000	54 000	145
	mean	138.53	101.88	153.2	2.68	68.57	6 290	8 430	35.73
	std	97.13	86.62	120.46	1.78	38.03	7 360	9 770	30.62
31 中学	min	30	1	18	0.22	6	1 000	1 000	1
	max	448	291	548	3.915	132	89 000	94 000	112
	mean	105.21	61.65	126.54	1.79	67.32	11 370	13 910	28.31
	std	70.55	51.99	91.34	0.79	27.88	17 450	17 970	22.38
74 中学	min	12	5	6	0.44	2	1 000	1 000	2
	max	500	535	694	7.89	114	79 000	82 000	80
	mean	142.77	89.88	174.19	2.01	51.04	10 720	13 960	22.4
	std	108.84	105.49	137.33	1.24	25.86	12 530	15 520	16.87
米东区 环保局	min	10	5	6	0.38	8	1 000	1 000	1
	max	454	316	554	7.29	92	21 000	24 000	170
	mean	123.14	81.5	163.3	2.45	45.05	4 930	5 350	23.02
	std	77.55	57.9	109.76	1.44	20.59	4 630	4 650	26.31

表中数据为平均每小时的浓度值(除 O₃(8 h)外)

滑的预测.

2.2.1 随机成分的稳定性检验 首先,利用 SPSS 软件将 PM_{2.5}数据利用 7 个监测站(1048 组)的数据进行拟合(见表 4),经检验 2 次拟合的 R^2 最大.其次,用实际值减去拟合值($\hat{y}$)得到拟合残差(随机因

素),利用 eviews 软件检验残差的平稳性(见表 5),P 值接近于 0,从而得到随机成分是平稳过程.

2.2.2 建立指数平滑模型 利用 eviews 软件进行指数平滑,得到最合适的 α 值为 0.32.最后得到随机部分的预测方程如下:

表 3 Pearson 相关系数
Tab.3 Pearson correlation coefficients

相关系数	PM _{2.5}	PM ₁₀	CO	NO ₂	O ₃	O ₃ (8 h)	SO ₂
PM _{2.5}	1	0.863	0.823	0.711	-0.272	-0.278	0.631
PM ₁₀	—	1	0.764	0.620	-0.304	-0.290	0.538
CO	—	—	1	0.823	-0.435	-0.437	0.541
NO ₂	—	—	—	1	-0.357	-0.298	0.490
O ₃	—	—	—	—	1	0.843	-0.269
O ₃ (8 h)	—	—	—	—	—	1	-0.311
SO ₂	—	—	—	—	—	—	1

表 4 几种回归结果比较
Tab.4 Several regression results

类型	一次线性	二次线性	指数	Logistic
R ²	0.008 8	0.255 1	0.020 5	0.020 498

表 5 随机成分的稳定性检验
Tab.5 The stability test of the random component

稳定性检验	T 检验	<i>P</i> 值	
Dickey-Fuller 检验统计量	-5.433 058	0.0000	
检验的 标准度	1%	-3.476 805	—
	-5%	-2.881 83	—
	-10%	-2.577 668	—

$$s_t = 0.32y_t + 0.68s_{t-1}, \tag{17}$$

2.3 主成分回归 由 2.2 经 SPSS 得到的拟合值 $\hat{y}$ 作为引起 PM_{2.5}变化的不稳定成分,利用 SPSS 将 PM₁₀、CO、O₃、NO₂、SO₂ 成分及天气等变量作为自变量 $\hat{y}$ 对进行主成分回归.经 SPSS 提取主成分检验,系统共选取了 4 个主成分,4 个主成分可以解释全部变量的 85.887%.图 1 为提取的主成分包涵成分,表 6 为旋转主成分载荷矩阵和特征值及方差贡献.

利用 SPSS 进行主成分回归(PCR),回归的 $R^2 = 0.803$,则回归模型对不稳定成分 PM_{2.5}浓度的解释程度到达 80.3%.膨胀因子 VIF 为 1.6(接近于 2),可以认为回归结果不存在多重共线性.图 2 为拟合回归的散点图.最后得到的回归方程为:

$$\text{PM}_{2.5} = 0.48 + 0.73\text{PC1} + 0.096\text{PC2} - 0.62\text{PC3} - 0.21\text{PC4}. \tag{18}$$

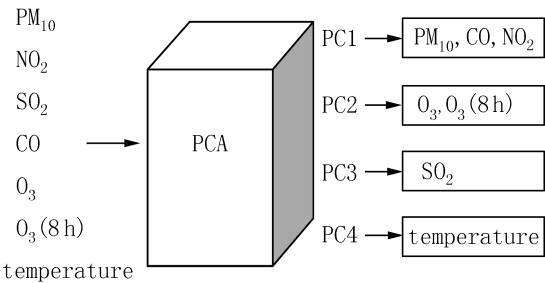


图 1 主成分的提取
Fig.1 Principal component extraction

表 6 旋转主成分载荷矩阵
Tab.6 Rotating principal component load matrix

载荷矩阵	PC1	PC2	PC3	PC4
PM ₁₀	0.872	—	—	—
CO	0.892	—	—	—
NO ₂	0.839	—	—	—
O ₃	—	0.921	—	—
O ₃ (8 h)	—	0.901	—	—
SO ₂	—	—	-0.662	—
temperature	—	—	—	0.967
特征值	3.842	1.464	0.879	0.685
贡献度	32.616	23.472	16.910	12.889
贡献度累计	32.616	56.088	72.998	85.887

2.4 性能指标的检验结果 将 2015 年 3 月 9 日至 2015 年 5 月 27 日(共 87 组)每天 9:00—11:00 共 7 个监测站数据,中国环境监测总站公布的乌鲁木齐实时数据和温度数据代入(17)和(18)中进行预测,将实际值与预测值利用式(6)~(10)的计算

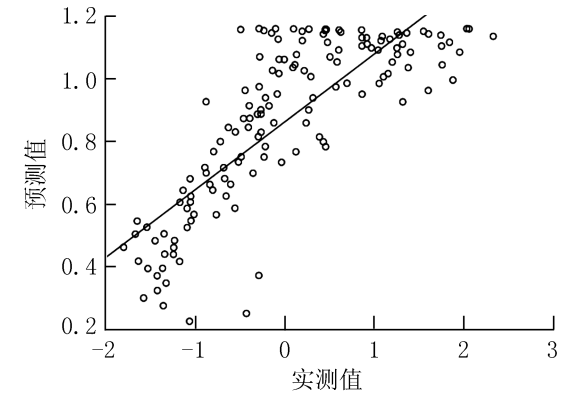


图 2 不稳定成分的拟合散点图

Fig.2 Unstable components fitting scatter plot

公式进行比较,结果见表 7.

从表 7 的结果可看出,NAE、RMSE 的值相对较接近于 0,PA、 R^2 、IA 的值近相较于于 1,这说明这种预测模型及方法具有一定实用性,但考虑到除温度外其他环境因素的影响,模型尚且需要改进.

表 7 性能指标结果分析

Tab.7 Performance indicators results analysis

NAE	RMSE	PA	R^2	IA
0.33	0.27	0.68	0.61	0.71

3 结 论

本文利用多元分析及时间序列的方法将 $PM_{2.5}$ 浓度变化划分为 2 部分进行预测.利用 ADF 检验得到 $PM_{2.5}$ 浓度的稳定部分是一个平稳的过程.对稳定部分利用指数平滑模型预测 $PM_{2.5}$ 浓度,得到最合适的平滑指数是 0.32.利用主成分回归(PCR)模型预测 $PM_{2.5}$ 浓度的不稳定性成分,得到预测结果 R^2 值为 0.803,说明不稳定成分的预测值可以解释不稳定成分的 80.3%.两部分的加权值作为最终预测结果.最后,利用 2 个月的数据进行 5 种性能指标检验模型,结果表明该模型方法预测效果较好,具有很好的应用前景.

参考文献:

[1] 覃辉艳.大气污染颗粒物 $PM_{2.5}$ 诱导人支气管上皮细胞凋亡及其机制探讨[D].南宁:广西医科大学,2012.
TAN H Y.Air pollution particles $PM_{2.5}$ human bronchial epithelial cells induced apoptosis and the mechanism

study[D].Nanning:Guangxi Medical University,2012.

[2] MORRIS R D,NAUMOVA E N,MUNASINGHE R L. Ambient air pollution and hospitalization for congestive heart failure among elderly people in seven large US cities[J]. American Journal of Public Health, 1995, 85 (10):1 361-1 365.

[3] WANG Y,ZHUANG G S,ZHANG X Y,et al.The ion chemistry,seasonal cycle,and sources of $PM_{2.5}$ and TSP aerosol in shanghai[J].Atmospheric Environment,2006, 40(16):2 935-2 952.

[4] EVANS J,DONKELAAR A V,MARTIN R V,et al.Estimates of Global Mortality Attributable to Particulate Air Pollution Using Satellite Imagery[J].Environmental Research,2013,120(1):33-42.

[5] RINEHART L R,FUJITA E M,CHOW J C,et al.Spatial distribution of $PM_{2.5}$ associated organic compounds in central California[J]. Atmospheric Environment, 2006, 40(2):290-303.

[6] APPEL B R,TOKIWA J H,HSU J,et al.Visibility as related to atmospheric aerosol constituents[J].Atmospheric Environment, 1985,19(9):1 525-1 534.

[7] SENARATNE I,SHOOTER D.Elemental composition in source identification of brown haze in Auckland, New Zealand[J]. Atmospheric Environment, 2004, 38 (19): 3 049-3 059.

[8] SCHICHTEL B A,HUSAR R B,FALKE S R,et al.Haze trends over the United States, 1980—1995[J].Atmospheric Environment,2001,35(30):5205-5210.

[9] 杨复沫,贺克斌,马永亮,等.北京 $PM_{2.5}$ 化学物种的质量平衡特征[J].环境化学,2004,5(3):327-311.
YANG F M,HE K B,MA Y L,et al.Beijing $PM_{2.5}$ mass balance characteristics of chemical species[J].Journal of Environmental Chemistry,2004,5 (3):327-311.

[10] 曲直.城市大气环境中 $PM_{2.5}$ 源解析成分谱的建立[D].长春:吉林大学,2013.
QU Z.Source of $PM_{2.5}$ analytical composition in the urban atmospheric environment for the establishment of the spectrum[D].Changchun:Jilin University,2013.

[11] 张朝能,王梦华,胡振丹,等.昆明市 $PM_{2.5}$ 浓度时空变化特征及其与气象条件的关系[J].云南大学学报:自然科学版,2016,38(1):90-98.
ZHANG C N,WANG M H,HU Z D,et al.Kunming $PM_{2.5}$ concentrations of temporal variation characteristics and relationship with meteorological conditions[J]. Journal of Yunnan University:Natural Science Edition, 2016,38(1):90-98.

[12] FERRO A R,KOPPERUD R J,HILDEMANN L M,et

- al.Source strengths for indoor human activities that re-suspend particulate matter [J]. Environ Sci Technol, 2004, 38(6): 1759-1764.
- [13] FROMME H, TWARDELLA D, DIETRICH S, et al. Particulate matter in the indoor air of classrooms—exploratory results from Munich and surrounding area [J]. Atmospheric Environment, 2007, 41(4): 854-866.
- [14] GOLD D R, ALLEN G, DAMOKOSH A, et al. Comparison of outdoor and classroom ozone exposures for school children in Mexico City [J]. Air Waste Manage Assoc, 1996, 46(4): 335-342.
- [15] UL-SAUFIE A Z, YAHYA A S, RAMLI N A. Improving multiple linear regression model using principal component analysis for predicting PM_{10} concentration in Seberang Prai, Pulau Pinang [J]. Int J Environ Sci, 2010, 2(2): 403-408.
- [16] 王春华, 吕爱华, 余晓丽, 等. 乌鲁木齐大气污染现状及影响因素分析 [J]. 新疆农业大学学报, 2010, 33(4): 349-353.
- WANG C H, LV A H, YU X L, et al. Urumqi air pollution status quo and influence factors analysis [J]. Journal of Xinjiang Agricultural University, 2010, 33(4): 349-353.
- [17] ABDUL-WAHAB S A, BAKHEIT C S, AL-ALAWI S M. Principal component and multiple regression analysis in modelling of ground-level ozone and factors its concentrations [J]. Environ Model & Software, 2005, 20(10): 1263-1271.
- [18] ÖZBAY B, KESKIN G A, AYBERK S, et al. Multivariate methods for ground-level ozone modeling [J]. Atmos Res, 2011, 102(1/2): 57-65.
- [19] ÖZBAY B. Modeling the effects of meteorological factors on SO_2 and PM_{10} concentrations with statistical approaches [J]. Clean Soil Air Water, 2012, 40(6): 571-577.
- [20] 韩忠庚. 数学建模方法及其应用 [M]. 2版. 北京: 高等教育出版社, 2009.
- HAN Z G. Mathematical modeling method and its application [M]. 2nd Edition. Beijing: Higher Education Press, 2009.
- [21] GERVASI O, MURGANTE B, LAGANA A, et al. Computational Science and its Applications its ICCSA 2008 [M]. Milan: Springer Press, 2008.
- [22] ALSHITAWI M, AWBI H B, MAHYUDDIN N. Particulate matter mass concentration (PM_{10}) under different ventilation methods in classrooms [J]. Int J Vent, 2009, 8(2): 93-108.
- [23] KOCAK M, KUBILAY N, TUĞRUL S, et al. Atmospheric nutrient inputs to the northern levantine basin from a long-term observation: sources and comparison with riverine inputs [J]. Biogeosciences Discussions, 2010, 7: 5081-5117.
- [24] 何书元. 应用时间序列分析 [M]. 2版. 北京: 北京大学出版社, 2014.
- HE S Y. Application of time series analysis [M]. 2nd Edition. Beijing: Peking University Press, 2014.

An analysis of $\text{PM}_{2.5}$ concentration based on time sequence and multivariate methods in Urumqi City

XIE Xin-qing, ZHENG Wei, KAI Xuan, XU Ying

(Institute of Applied Mathematics, Xinjiang University of Finance and Economics Urumqi 830012, China)

Abstract: In this study, the concentrations of PM_{10} , CO , NO_2 , O_3 , O_3 (8 h) and SO_2 concentrations have been employed to predict the concentration of $\text{PM}_{2.5}$ using multivariate statistical and time series analysis methods. The data have been collected from seven monitoring stations in Urumqi (Xinjiang province, China) from November 2014 to March 2015 (winter). The bivariate correlation analysis shows that the outdoor $\text{PM}_{2.5}$ concentration is highly correlated with that of PM_{10} , CO , NO_2 and SO_2 . Further, using exponential smoothing model to predict the stability of $\text{PM}_{2.5}$ concentration, the best smoothing index is 0.32. Principal component regression (PCR) has been used for modeling a prediction the instability of $\text{PM}_{2.5}$ concentration and values is 0.803 for explaining the instability of $\text{PM}_{2.5}$ concentration. The performance indicators of models indicate that the prediction for $\text{PM}_{2.5}$ has an effective prediction.

Key words: $\text{PM}_{2.5}$; air pollution; exponential smoothing; PCR