

叶瑜, 李秀央, 陈坤, 等. 2011. 大气气溶胶光学厚度与大气污染物及气象因素关系的时间序列研究 [J]. 气候与环境研究, 16 (2): 169-174. Ye Yu, Li Xiuyang, Chen Kun, et al. 2011. Association of aerosol optical depth with air pollutants and meteorological factors: A time series analysis [J]. Climatic and Environmental Research (in Chinese), 16 (2): 169-174.

大气气溶胶光学厚度与大气污染物及气象因素关系的时间序列研究

叶瑜^{1,2} 李秀央¹ 陈坤¹ 刘庆敏³ 项海青⁴

1 浙江大学公共卫生学院流行病与卫生统计学教研室, 杭州 310058

2 宁波市疾病预防控制中心, 宁波 315010

3 杭州市疾病预防控制中心, 杭州 310006

4 杭州市卫生事业发展中心, 杭州 310001

摘要 为探讨大气气溶胶光学厚度 (Aerosol Optical Depth, AOD) 与大气污染物及气象因素的关系, 进而为 AOD 的科学预测提供参考, 应用时间序列研究分析了 2002 年 1 月~2004 年 12 月的杭州市大气 AOD 与污染物及气象因素数据, 并以前 35 个月的数据建立时间序列 ARIMA (Autoregressive Integrated Moving Average) 模型, 用最后 1 个月的资料评估模型拟合效果。结果显示 AOD 与污染物及气象因素存在关联, 经一次差分后, 构建的模型为 ARIMA(0, 0, 2)(0, 0, 0), Ljung-Box Q 检验和残差图显示模型拟合较好。利用该模型计算 2004 年 12 月的 AOD 值, 显示实际值均在模型拟合值的 95% 可信区间内, 符合率达到 100%。时间序列模型表明 AOD 与污染物及气象因素的关系较复杂, 不仅受当日污染物和气象条件的影响, 还受到前几日污染物累积效应的影响。

关键词 气溶胶光学厚度 大气污染物 气象因素 时间序列研究

文章编号 1006-9585 (2011) 02-0169-06 **中图分类号** P468.0 **文献标识码** A

Association of Aerosol Optical Depth with Air Pollutants and Meteorological Factors: A Time Series Analysis

YE Yu^{1,2}, LI Xiuyang¹, CHEN Kun¹, LIU Qingmin³, and XIANG Haiqing⁴

1 *Department of Epidemiology and Health Statistics, Zhejiang University School of Medicine, Hangzhou 310058*

2 *Ningbo Center for Disease Control and Prevention, Ningbo 315010*

3 *Hangzhou Center for Disease Control and Prevention, Hangzhou 310006*

4 *Hangzhou Center for Development of Health Service, Hangzhou 310001*

Abstract To explore the association between Aerosol Optical Depth (AOD) with air pollutants and meteorological factors, and to provide evidence for the prediction of AOD, time series analysis was used. The data of AOD, air pollutants, and meteorological factors were from January 2002 to December 2004 in Hangzhou. The data of the first 35 months was collected to establish ARIMA (Autoregressive Integrated Moving Average) model, and the data of

收稿日期 2009-04-16 收到, 2011-01-30 收到修定稿

资助项目 杭州市科技局重点创新项目 20051323B44, 中央高校基本科研业务费专项资金资助 2010QNA7020

作者简介 叶瑜, 男, 硕士, 主要从事环境流行病学研究。E-mail: yeyu4902@163.com

the last month was used for evaluation. After a time difference, ARIMA(0, 0, 2)(0, 0, 0) model was established, the Ljung-Box Q test and the residual plots showed the model fit well. The true values were all above 95% confidence level. From the model, the AOD was not only affected by the air pollutants and meteorological factors on the same day, but also by those in several days before.

Key words aerosol optical depth, air pollutants, meteorological factors, time series analysis

1 引言

目前,我国评价空气质量的主要指标为二氧化硫(SO₂)、总悬浮颗粒物(TSP)、PM10(空气动力学当量直径小于等于10 μm的颗粒物)、氮氧化物(NO_x)、二氧化氮(NO₂)、一氧化碳(CO)、臭氧(O₃)等(国家环境保护局和国家技术监督局,1996)。但是一方面由于各地区的主要污染物不同,给综合评价、比较各地的污染水平带来了一定的困难,另一方面由于受风速、地理环境等因素的影响,地面监测点所测得的数据只能准确地反应一定区域范围内的空气污染状况。因此通过某一指标综合地反映更广覆盖区域的大气污染物水平就显得尤为重要。

气溶胶光学厚度(Aerosol Optical Depth, AOD)反映了大气颗粒吸收和散射光的特性,体现了大气气溶胶的光学性质,其通过搭载在卫星上的中分辨率成像光谱仪(MODIS)测得,能够覆盖更广的空间范围。一些研究表明通过卫星测定的AOD能够较准确的反映一定地区范围内的空气质量(Sarigiannis et al., 2002; Alföldy et al., 2007; Paciorek et al., 2008)。本研究利用时间序列模型探讨大气气溶胶与大气污染物及气象因素的关系,一方面为AOD的科学预测提供参考,另一方面为综合评价空气质量提供科学依据。

2 材料与方法

2.1 资料来源

杭州市2002年1月1日~2004年12月29日的气溶胶光学厚度和云量数据来源于欧洲空间局,该数据由搭载在Terra和Aqua卫星上的MODIS测得,这两颗卫星的扫描带每日上、下午2次经过杭州地区,原始数据空间分辨率为250 m,大气气溶胶光学厚度三级产品的空间分辨率为10

km。反演给出的AOD数据产品为参考波长550 nm的气溶胶光学厚度,显示了大气的透明度状况;以0~1表示云量多少(0表示晴天,1表示阴霾天气,随着云量的变化测定值在0~1之间波动)。当云量数值很大时,卫星由于云层过厚而无法获得AOD,从而造成当日AOD数据缺失。取每日AOD测定值、云量测定值的中位数作为当日杭州市的AOD、云量水平。

大气污染物变量包括PM10、SO₂和NO₂,由杭州市环境保护局提供,为各变量每日平均值;气象因素包括温度(T)和相对湿度(RH),该数据为从杭州市气象局获得的每日平均值。

2.2 时间序列分析

时间序列分析是利用现有的观察数据建立能够比较精确地反应时间序列所包含的动态依存关系的数学模型(王振龙和胡永宏,2007),其建模步骤分为4步:(1)序列平稳化和模型识别;(2)模型参数估计;(3)模型的诊断;(4)预测。

当观察数据为平稳序列时,自回归求和移动平均模型[ARIMA(Autoregressive Integrated Moving Average) Model]的表达式为:

$$z_t = \delta_1 z_{t-1} + \delta_2 z_{t-2} + \dots + \delta_p z_{t-p} + a_t - \phi_1 a_{t-1} - \phi_2 a_{t-2} - \dots - \phi_q a_{t-q},$$

其中, z 为时间序列, a 为残差序列, δ 、 ϕ 为自回归和移动平均的系数, p 、 q 为自回归和移动平均的阶数。

当观察数据为非平稳序列时,可通过差分、对数自然转换等方法实现序列的平稳性,此时模型可标记为ARIMA(p, d, q),其中 p 、 d 、 q 分别表示自回归的阶、差分次数、移动平均的阶。如若考虑时间序列的周期性,则模型标记为ARIMA(p, d, q)(P, D, Q)_s,其中 P 、 D 、 Q 、 S 分别表示为季节周期的自回归的阶、差分次数、移动平均的阶和周期。

本研究在对各变量进行平稳化后建立AOD与大气污染物及气象因素的时间序列模型,并通过

Ljung-Box Q 值和残差的自相关系数图 (ACF)、偏相关系数图 (PACF) 诊断模型, 以及利用模型计算 2004 年 12 月的 AOD 值并与实际测定值进行比较以综合评价模型。

2.3 缺失数据的填充

在开展时间序列分析前, 需对序列的缺失数据进行填充。本研究采用期望值最大似然法 (EM) 进行填充, 并对填充效果进行评价。EM 算法是一种迭代算法, 最初由 Dempster et al. (1977) 提出。本研究利用 2002 ~ 2003 年的 AOD、云量、T、RH、PM10、SO₂ 和 NO₂ 数据以及 2004 年的云量、T、RH、PM10、SO₂ 和 NO₂ 数据进行 EM 填充, 将填充后的数据与实际测得的数据进行比较, 以评价 EM 填充的效果。

所有数据管理和统计分析过程在 SPSS 16.0 FOR WINDOWS 完成。

3 结果与分析

3.1 研究期间 AOD、大气污染物及气象因素的测定情况

2002 年 1 月 1 日 ~ 2004 年 12 月 29 日共 1094 天, 各变量实际测得数据的天数统一简称为实测天数, 实际天数测得的数据简称为实测数据; 数据缺失的天数简称为缺失天数, 缺失天数的数据简称为缺失数据。研究期间各变量测定情况见表 1。

3.2 EM 填充缺失数据情况

2002 ~ 2003 年 AOD 实测数据与 2004 年 AOD 实测数据无显著性差异 (显著性检验结果 $r > 0.05$), 因此认为 AOD 数据来源于同一总体。采用 2002 ~ 2003 年的各变量实测数据和 2004 年的

表 1 2002 ~ 2004 年 AOD、大气污染物及气象因素测定情况

Table 1 Summary of median concentration of AOD, average daily concentrations of air pollutants and meteorological factors in Hangzhou city during 2002 - 2004

项目 (单位)	实测天数/d	均值	最小值	P_{25}	P_{50}	P_{75}	最大值	IQR
AOD	802	0.52	0.042	0.303	0.453	0.658	2.033	0.355
云量	1093	0.31	0.000	0.009	0.063	1.000	1.000	0.991
T (°C)	1056	17.94	-1.400	10.900	18.500	24.975	36.400	14.075
RH (%)	1056	72.19	26.00	61.00	73.00	84.88	99.00	23.88
PM10 浓度 ($\mu\text{g} \cdot \text{m}^{-3}$)	1058	120.46	14.00	80.00	113.00	145.00	476.00	65.00
SO ₂ 浓度 ($\mu\text{g} \cdot \text{m}^{-3}$)	1058	50.80	14.00	36.00	46.00	61.00	132.00	25.00
NO ₂ 浓度 ($\mu\text{g} \cdot \text{m}^{-3}$)	1058	55.76	18.00	42.00	53.00	67.00	135.00	25.00

注: P_{25} 、 P_{50} 、 P_{75} 分别表示各项目中实测数据的数值小于 P_{25} 、 P_{50} 、 P_{75} 的个数恰为实测数据总个数的 25%、50% 和 75%; 四分位间距 IQR (InterQuartile Range) = $P_{75} - P_{25}$ 。

表 2 AOD 与大气污染物及气象因素 ARIMA 模型参数估计

Table 2 Parameter estimation of ARIMA model

变量 (单位)	变量变换	滞后	参数估计值	标准差	r
MA	无	Lag1	0.746	0.030	<0.001
		Lag2	0.206	0.030	<0.001
ΔC	无	Lag0	0.350	0.020	<0.001
		倒数	Lag1	0.151	0.052
ΔR (%)	无	Lag0	0.004	0.001	<0.001
		倒数	Lag1	0.482	0.077
ΔP ($\mu\text{g} \cdot \text{m}^{-3}$)	无	Lag0	0.001	2.159×10^{-4}	<0.001
		倒数	Lag1	0.482	0.077
ΔS ($\mu\text{g} \cdot \text{m}^{-3}$)	无	Lag0	-0.003	0.001	0.004
ΔN ($\mu\text{g} \cdot \text{m}^{-3}$)	无	Lag0	-0.002	0.001	0.005

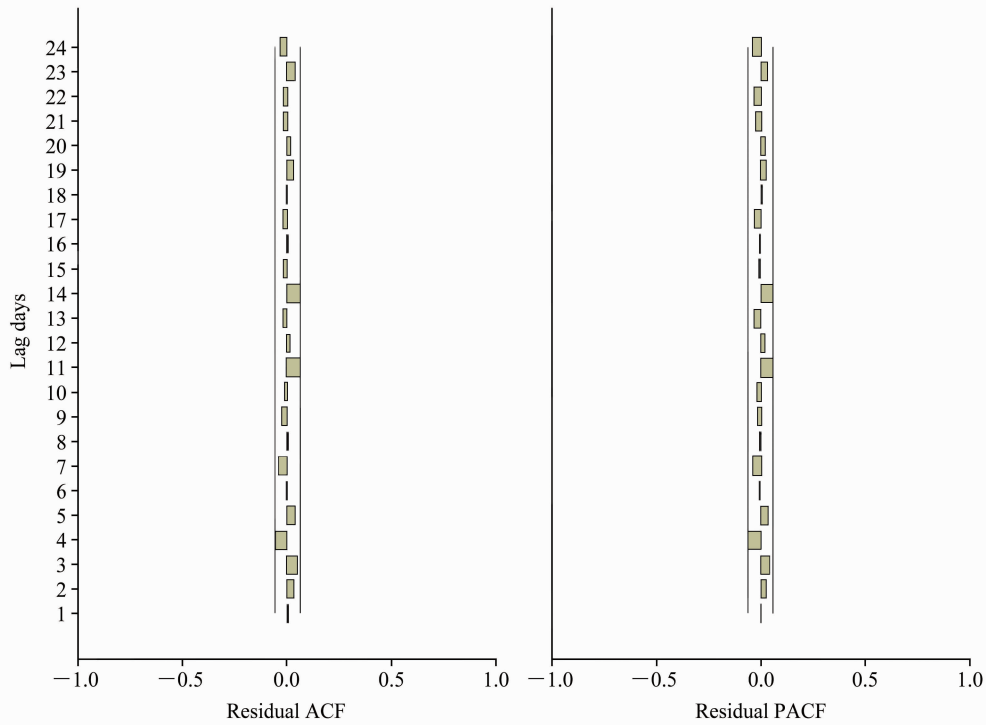


图1 ΔA 残差 ACF、PACF 图
Fig.1 Residual plots of ΔA

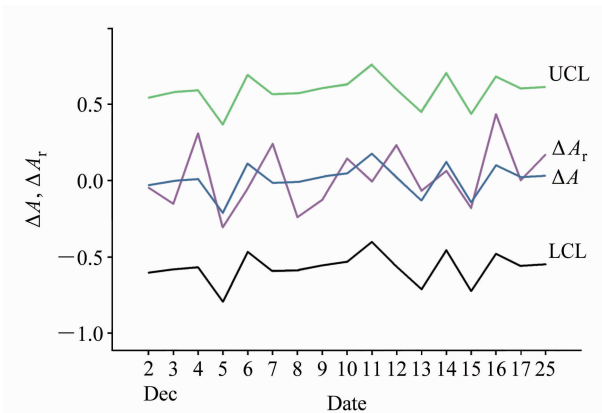


图2 ARIMA 模型预测 (UCL 为 95% CI 上限, LCL 为 95% CI 下限)
Fig.2 Prediction of ARIMA model (UCL means 95% upper CI, LCL means 95% lower CI)

大气污染物和气象因素实测数据进行 AOD 缺失数据 EM 法填充, 将 2004 年填充后结果与 2004 年的实测数据进行比较, 两者无显著性差异 ($r > 0.05$)。说明 EM 法填充缺失数据效果较好, 填充后的数据在统计学意义上与实际较符合。

3.3 时间序列分析结果

对 AOD、云量、 T 、RH、PM10、SO₂ 和

NO₂ 进行一次差分, 以达到序列的平稳性。结果见表 2, 其中 Lag0 表示当日, 即 t 日; Lag1 表示 $t-1$ 日; Lag2 表示 $t-2$ 日; A_t ($C_t, T_t, R_t, P_t, S_t, N_t$) 表示 t 日的 AOD (云量, T , RH, PM10, SO₂, NO₂) 测定值; ΔA ($\Delta C, \Delta T, \Delta R, \Delta P, \Delta S, \Delta N$) 表示 t 日与 $t-1$ 日的 AOD (云量, T , RH, PM10, SO₂, NO₂) 测定值之差; MA 为滑动平均模型的随机波动。

根据表 2 可得时间序列模型为 ARIMA(0, 0, 2)(0, 0, 0), 模型如下 (其中 a_t 为滑动平均模型 t 日的随机波动):

$$A_t = A_{t-1} - 0.746a_{t-1} - 0.206a_{t-2} + 0.350 \times (C_t - C_{t-1}) + 0.151 / (C_{t-1} - C_{t-2}) + 0.004 \times (R_t - R_{t-1}) + 0.001 \times (P_t - P_{t-1}) + 0.482 / (P_{t-1} - P_{t-2}) - 0.003 \times (S_t - S_{t-1}) - 0.002 \times (N_t - N_{t-1}).$$

3.4 模型评价

Ljung-Box Q 检验 ($r > 0.05$, 表明模型适合) 和残差 ACF、PACF 图 (图 1, 残差表现为围绕零均值随机波动) 均显示模型拟合较好。同时, 利用该模型计算 2004 年 12 月的 ΔA 值, 并与 2004 年

12 月实测数据的 ΔA 值 [用 ΔA_t 表示] 比较, 实际值均在模型拟合值的 95% 可信区间 (Confidence Interval, CI) 内。

4 结语与讨论

时间序列研究是利用变量随着时间发展的规律建立数学模型, 而不仅仅考虑各变量间的因果关系。采用时间序列研究的方法, 与时间变化相关的一些变量不再成为潜在的混杂因素, 因此被广泛应用于环境、医学等领域的研究, 用于预测模型的建立、大气污染与人群急性健康效应研究等方面。同时本研究还采用了 EM 算法填充缺失数据。EM 算法在处理数据缺失时, 考虑了数据的分布, 是以模型为基础的方法 (周艺彪等, 2004)。在数据缺失较多时, EM 算法是比较合适的选择。

由于目前反映大气污染水平的指标较多, 且大多通过地面监测点测得, 污染物水平又受到温度、湿度、气压等气象因素的影响, 因此在综合评价、比较及真实反映各地区空气质量上存在一定的局限性。因此本研究采用 AOD 作为反映大气污染状况的综合指标, 其不仅扩大了监测的区域范围, 也整体反映了颗粒物及气态污染物的综合水平。一些研究也证实了卫星测得的 AOD 数据能够较为真实地反映空气质量 (Sarigiannis et al., 2002; Alföldy et al., 2007; Paciorek et al., 2008), 且与通过太阳直接辐照度推算的大气气溶胶光学厚度值 (宋磊和吕达仁, 2006) 来的更加直接和客观。本研究利用 2002 年 1 月~2004 年 11 月的资料建立 ARIMA 模型, 以探讨 AOD 水平与污染物及气象资料的关系, 从而为 AOD 预测模型的建立提供依据和参考。模型拟合的评价指标显示模型拟合较好, 2004 年 12 月的 AOD 实测数据均在模型拟合值的 95% CI 内, 符合率达到 100%。同时本研究对 EM 算法的填充效果进行了验证, 结果也显示采用 EM 算法填充数据是比较合适的。

ARIMA 模型显示 AOD 与云量、RH、PM10、SO₂ 和 NO₂ 存在一定的关联, 部分上也解释了 Alföldy et al. (2007) 得出 PM10 与 AOD 线性关系较弱的原因, 即不仅是由于气溶胶成分过

于复杂, 同时 AOD 的值还受到前两日 PM10 值的影响。此外, 模型显示云量大小对 AOD 也有一定的影响。众所周知, 气溶胶对云凝结核的形成有重要的作用, 气溶胶微粒会改变云层, 增加大气中液滴浓度, 从而增加云量, 但是云量对 AOD 的影响还尚不明确。因此, 本研究结果可在统计学上为云量对 AOD 影响的研究提供一定的线索。从模型中也可以看出, AOD 的值与 RH、SO₂ 和 NO₂ 也不仅仅是单一的线性关系。这表明当日 AOD 不仅受当日大气污染物和气象因素的影响, 同时也受到前几日各变量的累积效应影响。同时, 本研究也存在局限性, 首先 AOD 的缺失数据所占百分比比较高, 其测定受云量、云层厚度等因素影响较大; 其次卫星每日测定的 AOD 值, 其经度、纬度各不相同, 将数据取中位数合并, 会损失相当一部分的信息; 再次, 由于受到模型本身的限制, 无法推断影响 AOD 水平变化的关键因子; 最后, 影响 AOD 值的污染物及气象因素较多, 除上述纳入的变量外, 包括硫酸盐离子、硝酸盐离子、PM2.5、CO、气压等 (Molnár and Mészáros, 2001; Sunder Raman et al., 2008), 而本次模型纳入的变量相对较少。

通过上述讨论, 可得到大气气溶胶与大气污染物及气象因素的关系如下:

(1) 时间序列研究模型显示 AOD 不仅受当日污染物和气象因素的影响, 还受到前几日累积效应的影响。从中也可以推断, 当日大气污染水平受到当日污染排放和多日污染物累积的影响;

(2) AOD 与大气污染物和气象因素的关系不是简单的线性关系, 其与各变量的关系较复杂;

(3) 利用时间序列分析方法建立的 ARIMA 模型, 其模型的拟合度较好。

参考文献 (References)

Alföldy B, Osán J, Tóth Z, et al. 2007. Aerosol optical depth, aerosol composition, and air pollution during summer and winter conditions in Budapest [J]. *Science of the Total Environment*, 383 (1-3): 141-163.

Dempster A P, Laird N M, Rubin D B. 1977. Maximum likelihood from incomplete data via the EM algorithm [J]. *Journal of the Royal Statistical Society*, 39B (1): 1-38.

国家环境保护局, 国家技术监督局. 1996. 环境空气质量标准

- (GB3095-1996) [S]. 北京: 中国环境科学出版社, 5pp. State Bureau of Environmental Protection, State Bureau of Technical Supervision. 1996. Ambient Air Quality Standard (GB3095-1996) [S] (in Chinese). Beijing: China Environmental Science Press, 5pp.
- Molnár A, Mészáros E. 2001. On the relation between the size and chemical composition of aerosol particles and their optical properties [J]. *Atmos. Environ.*, 35: 5053-5058.
- Paciorek C J, Liu Y, Moreno-Macias H, et al. 2008. Spatiotemporal associations between GOES aerosol optical depth retrievals and ground-level PM_{2.5} [J]. *Environ. Sci. Technol.*, 42 (15): 5800-5806.
- Sarigiannis D A, Soulakellis N, Schäfer K, et al. 2002. ICAROS: An integrated computational environment for the assimilation of environmental data and models for urban and regional air quality [J]. *Water, Air, and Soil Pollution: Focus*, 2 (5-6): 641-654.
- 宋磊, 吕达仁. 2006. 上海地区大气气溶胶光学特性的初步研究 [J]. *气候与环境研究*, 11 (2): 203-208.
- Song Lei, Lü Daren. 2006. Investigating of atmospheric optical characteristics over Shanghai region [J]. *Climatic and Environmental Research* (in Chinese), 11 (2): 203-208.
- Sunder Raman R, Hopke P K, Holsen T M. 2008. Characterization of fine aerosol and its inorganic components at two rural locations in New York State [J]. *Environmental Monitoring and Assessment*, 144 (1-3): 351-366.
- 王振龙, 胡永宏. 2007. 应用时间序列分析 [M]. 北京: 科学技术出版社, 250pp.
- Wang Zhenlong, Hu Yonghong. 2007. The Application of Time Series Analysis [M] (in Chinese). Beijing: Scientific and Technical Press, 250pp.
- 周艺彪, 姜庆五, 赵根明. 2004. 不完全数据处理方法——多重填充 [J]. *中华预防医学杂志*, 38 (6): 424-426.
- Zhou Yibiao, Jiang Qingwu, Zhao Genming. 2004. The processing of incomplete data—Multiple imputation [J]. *Chinese Journal of Preventive Medicine* (in Chinese), 38 (6): 424-426.