

基于高频 AQI 数据的关中城市群 空气污染规律探索

胡秋灵*, 杨哲

(陕西师范大学国际商学院, 西安 710119)

摘要 基于高频分时AQI及各污染物浓度数据, 本文使用“AQI小时指数”、首要污染物等进行统计分析并建立VAR模型对关中城市群空气污染的总体情况、日内波动规律以及城市间空气污染的关联规律做了不同层次的挖掘。研究表明: ①关中城市群的空气质量整体较差, 春、冬季空气污染程度明显大于夏季, 空气污染的“季节效应”和“集簇性”明显, 且主要表现为颗粒物污染。②空气质量的日内波动规律在春、冬两季表现为下半年优于上半年, 夏季夜间优于白天。但在不考虑四季AQI小时指数图“相位”差异的情况下, 四季空气污染的日内波动呈现出明显的相似性。③各城市空气污染存在明显的关联规律。城市群内一个城市空气污染的恶化会加剧其他城市的空气污染, 并且对其他城市空气污染的影响峰值会在24小时之内出现, 且该影响会随着空间和时间尺度的增大而逐步衰减。

关键词 高频AQI数据; 关中城市群; AQI小时指数; 首要污染物; VAR模型

中图分类号: F062.2; X830.3

文章编号: 1674-6252(2017)02-0037-06

文献标识码: A

DOI: 10.16868/j.cnki.1674-6252.2017.02.037

Research on Air Pollution Laws in Guanzhong Urban Agglomeration Based on High Frequency AQI Data

HU Qiuling*, YANG Zhe

(International Business School of Shaanxi Normal University, Xi'an 710119)

Abstract: Based on high frequency data of AQI and contaminants, this paper made several layers of researches including the general situation, daily fluctuation rules and correlation laws between cities of air pollution by using statistic analysis with the hour index of AQI and primary pollutants and building the VAR model. The results presented the following three aspects: firstly, the situation of air pollution in Guanzhong urban agglomeration is not optimistic. Air pollution in spring and winter is worse than that in summer, and it is obvious that air pollution which often manifests as pollution of particulates has a seasonal effect and the clustering property. Secondly, air quality in first half day is worse than that in the rest of the day in spring and winter, while in summer, air quality in night is better than that in the day time. However, fluctuations of AQI hour index in four seasons have a noticeable likeness to one another in the situation that with no consideration of the phase position diversity. Thirdly, an impulse acts on one city can cause a plus pollution influence on others, and the peak of this influence will appear in one day generally. What's more, the influence will weaken along with the increase of the distance and time.

Keywords: high frequency AQI data; Guanzhong urban agglomeration; AQI hour index; primary pollutants; VAR model

引言

空气污染危害人类身体健康及经济的可持续发展,《2010年全球疾病负担评估》称目前全球每年因空气污染而过早死亡的人口超过320万,《迈向环境可持续的未

来——中华人民共和国国家环境分析》称每年中国空气污染造成的经济损失,基于支付意愿估算,约占GDP的3.8%。改革开放以来,我国城市建设取得了一定成就,然而,城市化效率总体偏低导致城市空气污染问题日益

资助项目:教育部人文社会科学研究西部项目“基于AQI高频大数据的西部城市群空气污染统计规律挖掘及治污减霾对策设计”(批准号:15XJA910001);陕西省软科学研究计划一般项目“基于AQI高频大数据的关中城市群空气污染统计规律挖掘及治污减霾对策设计”(批准号:

2016KRM096)。

* 责任作者:胡秋灵(1964—),女,副教授,博士,主要研究方向为经济与金融市场计量分析、金融工程、农村金融, E-mail: hu_qi@163.com。

严峻。城市群作为推进城市化的主体形态，其集聚效应无形中造成了更高风险的污染威胁。目前关于城市群空气污染的研究大多集中在京津冀城市群、长三角城市群和珠三角城市群^[1-3]，也有少部分关于关中城市群空气污染的研究^[4,5]。但基于高频分时数据的关中城市群空气污染相关研究我们尚未发现。本文期望通过对关中城市群高频分时 AQI 及各污染物浓度数据的分析，挖掘出关中城市群空气污染的深层次规律，为该区域空气污染治理提供科学依据。

1 文献综述

近年来，国内外研究者已从多个视角对空气污染的特征和原因做了解读，如 Kassomenos 等分析了欧洲三个城市颗粒物污染的来源和季节性特征^[6]，Kimbrough 等分析了美国拉斯维加斯市交通排放的季节变化对当地空气质量的影响^[7]，Zhou 等研究了污染季节包头市可吸入颗粒物的分布特征^[8]，Xu 等研究了宁波市不同程度的污染时期空气中气溶胶成分特征及其来源情况^[9]等。在方法论上，描述性统计分析、相关分析、主成分分析等统计方法已被研究者广泛使用，如 AzidA 等使用主成分分析和人工神经网络方法对空气污染等级进行预测^[10]，Assareh 等对泰国东部地区 1997—2012 年干燥季节的臭氧量做了统计分析^[11]。此外，神经网络、模糊数学、支持向量机和现代计量经济学模型方法等也获得了不同程度的应用，如 Bai 等基于反向传播神经网络模型对大气污染物浓度做了分析和预测^[12]，Kaburlasos 等应用模糊推理对环境臭氧含量进行估计^[13]等。随着高频海量空气质量数据的出现，一些研究者尝试将数据挖掘方法应用于对空气污染物之间关联规律等的研究，如贾瑾使用数据挖掘等方法分析了大气复合污染的时空特征^[14]。这些研究为基于 AQI 高频数据的关中城市群空气污染规律挖掘提供了方法论基础。

2 研究区域与数据来源

2.1 研究区域

本文的研究区域为我国关中城市群。关中城市群是我国已形成的 11 个国家级城市群之一，关中城市群主要包括西安、咸阳、铜川、宝鸡及渭南 5 个城市，其所在的关中平原是经渭河、泾河和洛河冲积而成，是一个由断陷作用形成的槽形地堑，海拔在 400～500m，北部的北山和南部的秦岭构成了天然的屏障^[15]。这种独特的地形特征导致关中城市群构成了一个在环境上相对封闭且城市之间相互影响明显的系统，系统内的污染较难向外扩散，这就导致了在容易出现空气污染的季节系统内污染容易集聚，污染范围广、程度深、时间长。根据近年来在该地区的生活经验，同时参考环保部公布的相关数

据，我们发现关中城市群的空气污染形势十分严峻。

2.2 数据来源及预处理

本文采集了上述 5 个城市 2015 年 1 月 2 日零时到 2015 年 12 月 31 日 23 时的 AQI 值与 SO₂、CO、NO₂、O₃、PM_{2.5} 和 PM₁₀ 6 项污染物的浓度分时数据，其中，AQI（空气质量指数）为定量描述空气质量状况的无量纲指数，我国的 AQI 指数分为 6 个等级，AQI 的数值越大，说明空气质量越差，污染越严重。本文中春季为 3～5 月，夏季为 6～8 月，秋季为 9～11 月，冬季为 1、2 和 12 月^[16]。数据来源于 http://www.pm25.in 网站提供的国家环保部空气质量数据的实时更新。其中，对于少数缺失数据，本文通过试算比较几种常用的插值方法，最终选用误差最小的三次插值法补全。

3 关中城市群空气污染规律统计分析

本部分将从关中城市群空气质量总体情况以及日内波动情况两方面分析关中城市群空气污染的统计规律。

3.1 关中城市群空气质量概况

3.1.1 基于 AQI 时序图的分析

图 1 描述了关中城市群 5 个城市 2015 年全年 8736 个小时的 AQI 分时波动情况。图 1 中，中度污染线对应空气污染达到中度污染时的 AQI 值（151），此时空气质量指数级别达到四级，图 1 中均值为各城市全年 AQI 平均水平。由图 1 可知，5 个城市的 AQI 时序图呈现出相似的波动特征，具体表现在：

① 5 个城市的全年 AQI 均值分布在 87.70～94.83，相差不大，但都高于全国平均水平 79.36^①，反映出关中城市群整体空气质量较差。

② 5 个城市 AQI 值超过“中度污染线”的时段都主要出现在春、冬两季，而夏、秋两季的 AQI 值普遍较低，春、冬两季的空气质量明显劣于夏、秋两季，季节特征明显；空气污染会在某一时间段内集中连续出现，呈现出“集簇性”。

③ 5 个城市的 AQI 值几乎总在相同的时段达到峰值或处于较低水平，一个最明显的例子是，第 7700 小时到第 8700 小时之间，5 个城市都先后出现了 4 次时间上较一致的严重的污染天气。波动上的相似性从一个侧面反映出 5 个城市之间存在空气污染的关联规律。

3.1.2 基于首要污染物的分析

根据环境空气质量指数技术规范（HJ 633—2012）中定义，首要污染物为 AQI 大于 50 时 IAQI 最大的空气污染物。表 1 列出 5 个城市全年共 8736 个小时的首要污染物情况的统计结果，表 2 是对“AQI<50”及各首要污染物出现时所处季节的占比情况的统计。结合两表

^① 所有 AQI 分时数据计算出的平均值。

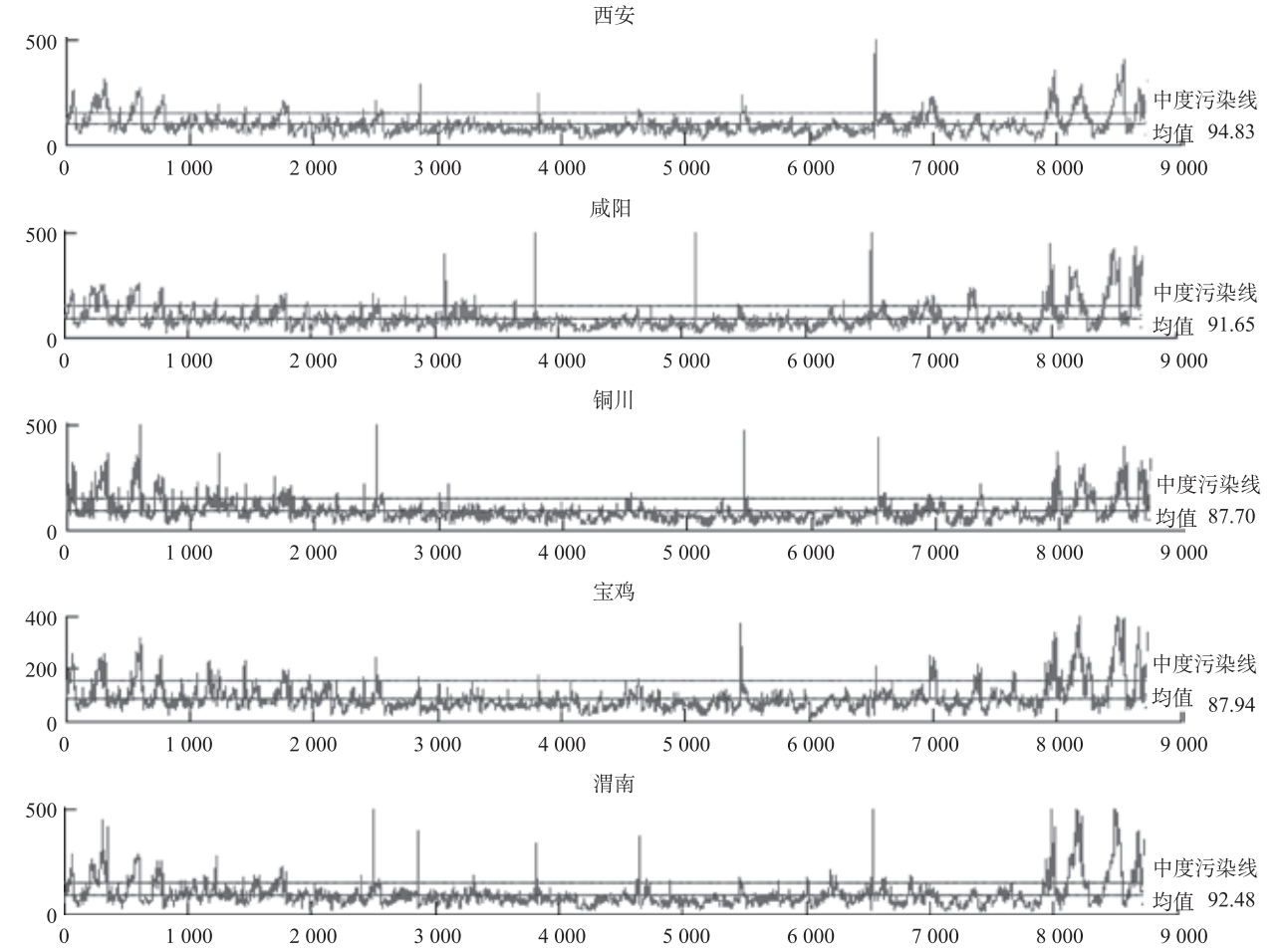


图 1 各城市 AQI 值时序图

可知：

① 5 个城市全年出现次数最多的首要污染物是 PM₁₀，其中，西安全年的首要污染物中，PM₁₀ 占比高达 73.27%，此外 PM₁₀ 作为首要污染物主要出现在春、夏、秋三季；

② 5 个城市全年出现次数次多的首要污染物为 PM_{2.5}，主要出现在春、秋、冬三季；

③ 5 个城市 O₃ 污染主要出现在夏季，西安的 O₃ 污染在 5 个城市中最为严重；

④ 5 个城市“AQI<50”的情况主要出现在春、夏、秋三季，从一个侧面反映出冬季空气质量在一年中相对最差。

表 1 各城市首要污染物占比

项目	西安	咸阳	铜川	宝鸡	渭南
O ₃	1.15%	0.41%	0.27%	0.44%	0.46%
PM _{2.5}	25.59%	39.81%	42.66%	40.73%	41.79%
PM ₁₀	73.27%	59.76%	57.07%	58.83%	57.73%

注：在首要污染物统计中，CO 没有出现，SO₂ 和 NO₂ 极少出现，且分别在咸阳和渭南两地出现一次，占比 0.01%，因此，以上三种污染物均未在表中体现。

表 2 “AQI<50”及各首要污染物季节分布情况

项目	春季	夏季	秋季	冬季
AQI<50	24.26%	39.14%	33.78%	2.82%
O ₃	20.00%	78.57%	1.43%	0.00%
PM _{2.5}	33.41%	13.27%	26.67%	26.65%
PM ₁₀	36.61%	28.61%	21.92%	12.86%

4.2 关中城市群空气污染日内波动规律

4.2.1 AQI 小时指数的构建

季节指数用来反映某季度的变量水平与总平均值之间的比较稳定的关系，绘制季节指数图可以帮助我们更清晰地总结月度变迁对待研究变量的影响，本文类似地定义 AQI 小时指数，用来反映一个季节内某一时点的 AQI 水平与季节平均水平之间的关系。其构建过程如下：第一步，计算某个季节 AQI 各时点平均值。AQI 序列以 24 小时为一个周期，设一个季节共有 n 个周期，则：

$$\bar{x}_k = \frac{\sum_{i=1}^n x_{ik}}{n}, \quad k=1,2,\cdots,24$$

① AQI 全国平均水平是根据收集到的全国 190 个城市 2015 年全年

第二步，计算某个季节 AQI 的总平均值，

$$\overline{x} = \frac{\sum_{i=1}^n \sum_{k=1}^{24} x_{ik}}{24n}$$

第三步，用各时点平均值除以总平均值得到各时点的小时指数 H_k ($k=1,2,\cdots,24$)，即

$$H_k = \frac{\overline{x_k}}{\overline{x}}, k=1,2,\cdots,24$$

H_k 大于 1，说明该时点的 AQI 值常常会高于该季节总平均值；反之，则说明该时点的 AQI 值常常会低于该季节总平均值。通过对 AQI 小时指数图的观察，可以总结出某个季节一天之中空气污染的波动规律。

4.2.2 关中城市群空气污染日内波动规律分析

该部分以西安市为例进行规律分析，首先分别计算出西安市每个季节的 24 个 AQI 小时指数值，然后通过观察四季 AQI 小时指数图分析西安市不同季节空气污染的日内变化规律。西安市四季 AQI 小时指数变化情况见图 2。观察图 2 我们发现：

- ①春、冬两季的小时指数波动状态大体一致，夏季明显异于春、冬两季，而秋季则介于两者之间。
- ②春、冬两季的小时指数从 22 时开始到次日 13 时基本都处于大于 1 的状态，并且在凌晨 2 时到 3 时之间达到一个小的峰值，在 11 时左右达到一个较大的峰值，此时的空气质量全天最差，在 14 时到 21 时之间小时指数小于 1，最小值出现在 17 时左右，此时的空气质量全天最优；夏季一天中，从 8 时到 22 时，小时指数基本都大于 1，两个峰值分别出现在 11 时和 19 时左右，22 时到次日 7 时之间小时指数小于 1，最小值出现在 1 时左右。总结以上规律可知，一天内的空气质量变化情况为：春、冬两季的下半日（14 时到 23 时）相对优于上半日（0 时到 13 时），夏季夜间相对优于白天。
- ③春、夏、秋、冬四季的小时指数曲线在形状上具有相似性，只是在“相位”上有所不同，或者说，春、冬季的小时指数图可以经过左右平移相似地得到秋季和

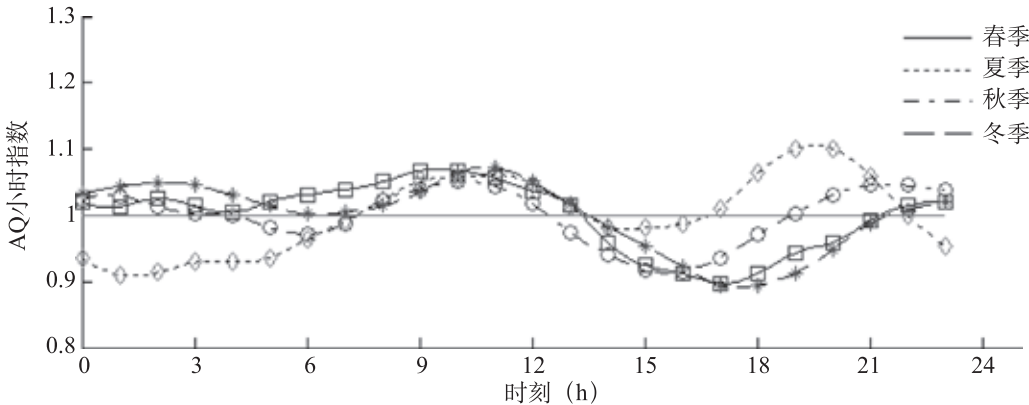


图2 西安市四季 AQI 小时指数图

夏季的小时指数图，即四个季节有较一致的日内空气污染相对变化，只是一天内空气质量的优或劣所处时点不同。

通过对其他 4 个城市四季 AQI 小时指数的分别考察，同样能得到类似上述的空气污染日内变化规律。

5 基于 VAR 模型的关中城市群空气污染关联规律分析

该部分使用向量自回归（VAR）模型研究关中城市群内 5 个城市空气污染的关联规律。因为在做首要污染物分析时，我们发现 5 个城市分时首要污染物全年占比居于前两位的分别是 PM_{10} 和 $PM_{2.5}$ （具体的首要污染物占比统计结果见表 1），同时考虑到 $PM_{2.5}$ 对人体健康较强的危害性，该部分我们选择 $PM_{2.5}$ 作为城市间空气污染关联规律研究的分析对象。

5.1 VAR 模型简介和指标选取

VAR 模型的数学表达式如下：

$$y_t = \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + u_t, t=1,2,\cdots,T$$

其中， y_t 是 5 维内生变量列向量； p 是滞后阶数； T 是样本个数。5×5 维矩阵 $\Phi_1, \cdots, \Phi_p$ 是待估计的系数矩阵， u_t 是 5 维扰动列向量，也被称为脉冲值（impulse）。根据前文说明，该部分我们选用西安、咸阳、铜川、宝鸡和渭南 5 个城市的 2015 年全年分时 $PM_{2.5}$ 时间序列数据进行建模分析。

5.2 VAR 模型的建立

首先，检验数据序列的平稳性。进行 VAR 动态回归模型拟合时，各参与建模的数据序列必须满足平稳性要求。本文采用 ADF 检验法对上述各数据序列的平稳性进行检验（详见表 3），从表 3 可以看出，5 个城市的 $PM_{2.5}$ 序列在 99% 的置信水平下都是平稳的，可以建立 VAR 模型。

然后，确定模型的滞后期。采用施瓦兹（SC）准则确定的模型的滞后期为 3。

表3 ADF 检验结果				
变量	ADF 统计量	临界值（1%）	伴随概率 P	结论
西安（xa）	-8.133 3	-3.430 9	0.0000	平稳
咸阳（xy）	-6.731 2	-3.430 9	0.0000	平稳
铜川（tc）	-8.525 0	-3.430 9	0.0000	平稳
宝鸡（bj）	-9.791 4	-3.430 9	0.0000	平稳
渭南（wn）	-8.120 2	-3.430 9	0.0000	平稳

实证结果显示，VAR 模型中 5 个函数的拟合优度都在 0.967 以上，由图 3 可知，模型的特征方程的特征根

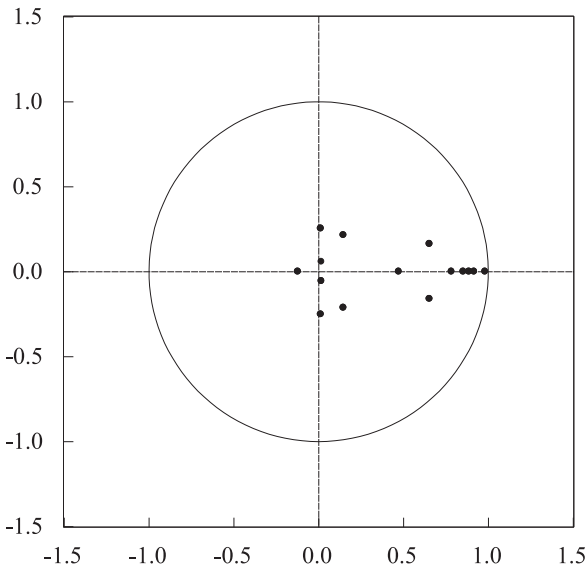


图3 VAR 模型平稳性检验

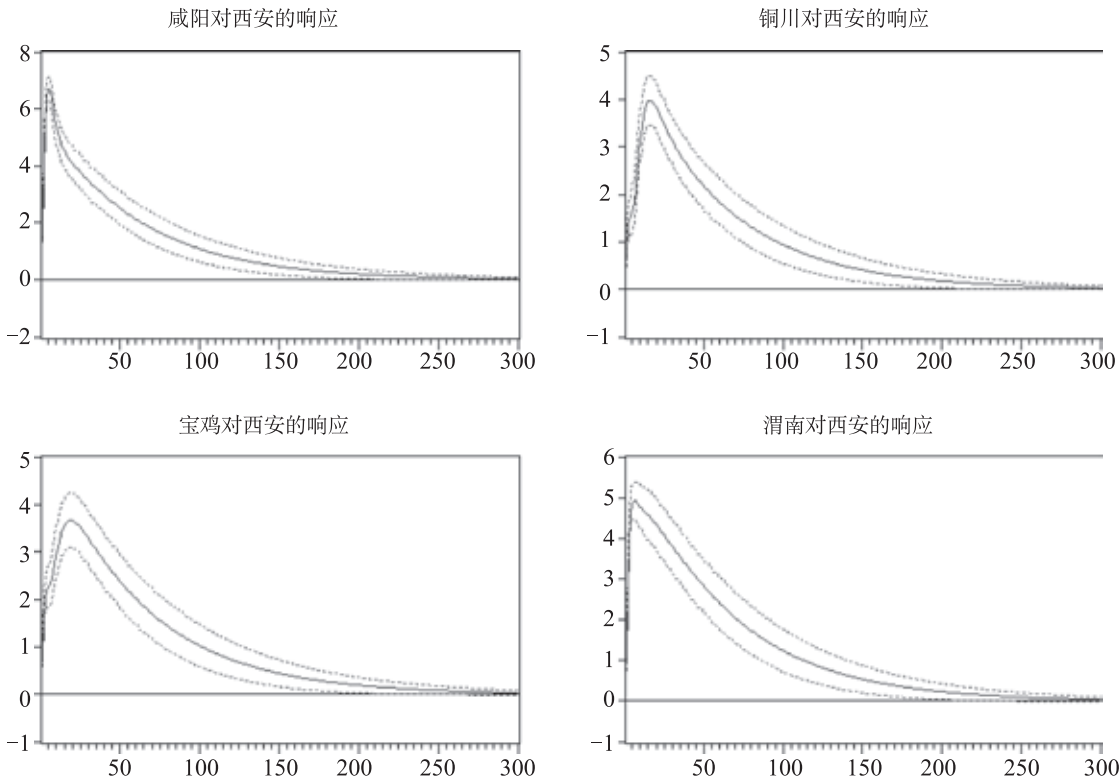


图4 VAR 模型脉冲响应函数

的倒数都位于单位圆之内，表明模型结构稳定，是一个平稳系统。

5.3 脉冲响应分析

对于 VAR 模型，我们重点关注的是系统的动态特征，即利用脉冲响应函数分析对某个内生变量施加的冲击会对其他变量产生怎样的影响。图 4 中的脉冲响应函数图即反映了对西安市（ $PM_{2.5}$ 浓度）施加冲击后其他 4 个城市所受影响的变化情况。由图 4 可知：

- ①西安市 $PM_{2.5}$ 的浓度变化对其他 4 个城市的影响都是正向的，即其他城市的 $PM_{2.5}$ 的浓度会随着西安市的浓度升高而升高。
 - ②冲击对其他 4 个城市的影响有相同的变化趋势，随时间推移，影响都是先增大，达到峰值，然后逐渐衰减趋零。
 - ③影响峰值都出现在冲击后一天（24 小时）之内，但每个城市各有不同，具体表现为咸阳（5 小时）、渭南（6 小时）、铜川（17 小时）和宝鸡（20 小时），结合关中城市群的地图（图 5）我们发现，冲击后影响峰值出现所需的时间随空间距离的递增而递增。
 - ④4 个城市影响的峰值相对大小也有差异，结合图 5 我们发现，随空间距离的递增，峰值大小逐步递减，峰值大小变化具体表现为咸阳（6.7）、渭南（4.9）、铜川（4.0）和宝鸡（3.7）。
- 此外，对其他城市的脉冲响应函数分析也可以得到类似的规律。



图5 关中城市群行政地图

6 结论

本文通过对关中城市群5个城市2015年全年（不包括1月1日）的分时AQI及6项污染物浓度数据的统计和建模分析，挖掘了城市群空气污染的总体情况、日内分时波动规律以及城市间污染的关联规律，得出以下结论：

①关中城市群的整体空气质量状况较差，且污染特征明显。各城市空气污染全年平均水平都劣于全国平均水平；各城市全年污染波动状况高度相似，呈现出明显的关联性；春、冬季空气污染程度明显严重于夏季，且空气污染的“季节效应”和“集簇性”明显；空气污染主要表现为颗粒物污染， PM_{10} 和 $PM_{2.5}$ 占全年所有分时首要污染物的绝大部分， O_3 污染主要集中在夏季。

②关中城市群日内空气污染规律的季节差异明显。AQI小时指数的变化情况具有明显的季节性，春、冬季和夏季呈现出不同的波动情况，空气质量在春、冬两季表现为下半年相对优于上半年，夏季夜间相对优于白天；但在不考虑四季AQI小时指数图“相位”差异的情况下，四季空气污染日内波动呈现出明显的相似性。

③关中城市群各城市空气污染存在明显的关联规律。一个城市的空气污染冲击会带给其他城市带来正向的污染影响，影响的峰值会在一天之内出现，且该影响会随着空间和时间尺度的增大而逐步衰减。

参考文献

- [1] 胡秋灵, 李雅静. 基于AQI的滇中、黔中和北部湾城市群空气污染统计规律比较研究[J]. 生态经济, 2016, 32(5): 170-174, 185-185.
- [2] 万庆, 吴传清, 曾菊新. 中国城市群城市化效率及影响因素研究[J]. 中国人口·资源与环境, 2015, 25(2): 66-74.
- [3] 方创琳, 关兴良. 中国城市群投入产出效率的综合测度与空间分异[J]. 地理学报, 2011, 66(8): 1011-1022.
- [4] 李养养, 唐小威, 张佳音. 关中城市群环境空气污染特征及对策研究[J]. 环境与发展, 2015, 27(4): 60-62.
- [5] 彭艳, 王钊, 李星敏, 等. 近60a陕西关中城市群大气能见度变化趋势与大气污染研究[J]. 干旱区资源与环境, 2011, 25(9): 149-155.

- [6] KASSOMENOS P, VARDOULAKIS S, CHALOULAKOU A, et al. Levels, sources and seasonality of coarse particles (PM_{10} - $PM_{2.5}$) in three European capitals—Implications for particulate pollution control[J]. Atmospheric environment, 2012, 54: 337-347.
- [7] KIMBROUGH S, BALDAUF R W, HAGLER G S W, et al. Long-term continuous measurement of near-road air pollution in Las Vegas: seasonal variability in traffic emissions impact on local air quality[J]. Air quality, atmosphere & health, 2013, 6(1): 295-305.
- [8] ZHOU H J, HE J, ZHAO B Y, et al. The distribution of PM_{10} and $PM_{2.5}$ carbonaceous aerosol in Baotou, China[J]. Atmospheric research, 2016, 178-179: 102-113.
- [9] XU J S, XU H H, XIAO H, et al. Aerosol composition and sources during high and low pollution periods in Ningbo, China[J]. Atmospheric research, 2016, 178-179: 559-569.
- [10] AZID A, JUAHIR H, TORIMAN M E, et al. Prediction of the level of air pollution using principal component analysis and artificial neural network techniques: a case study in Malaysia[J]. Water, air, & soil pollution, 2014, 225(8): 2063.
- [11] ASSAREH N, PRABAMROONG T, MANOMAIPHIBOON K, et al. Analysis of observed surface ozone in the dry season over Eastern Thailand during 1997-2012[J]. Atmospheric research, 2016, 178-179: 17-30.
- [12] BAI Y, LI Y, WANG X X, et al. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions[J]. Atmospheric pollution research, 2016, 7(3): 557-566.
- [13] KABURLASOS V G, ATHANASIADIS I N, MITKAS P A. Fuzzy lattice reasoning (FLR) classifier and its application for ambient ozone estimation[J]. International journal of approximate reasoning, 2007, 45(1): 152-188.
- [14] 贾瑾. 基于空气质量数据解析大气复合污染时空特征及过程序列[D]. 杭州: 浙江大学, 2014: 1-135.
- [15] 韩超. 关中地区环境空气污染物浓度的统计特征及与气象要素的关系[D]. 西安: 长安大学, 2012: 1-73.
- [16] 王振波, 方创琳, 许光, 等. 2014年中国城市 $PM_{2.5}$ 浓度的时空变化规律[J]. 地理学报, 2015, 70(11): 1720-1734.