

Analysis of Air Quality in Environmental Key Cities in China

Jinyu Tang, Xiang Gao*

School of Mathematical Sciences, Ocean University of China, Qingdao Shandong
Email: 18354222836@163.com, *gaoxiangshuli@126.com

Received: May 21st, 2018; accepted: Jun. 6th, 2018; published: Jun. 20th, 2018

Abstract

Air is a condition for human beings and living things to live, so clean air is particularly important for people. However, in recent years, with the development of industry and transportation industry in China, a lot of pollutants have been discharged into the air, and the air quality is getting worse and worse. The air quality problem is also more and more concerned by the government and the public. In this paper, the air quality data of 113 environmental key cities in China Statistical Yearbook of 2017 were selected, and the statistical analysis was carried out by using R and SPSS software. First, the principal component analysis is applied to reduce the dimension of the original data based on the Kaiser-Harris principle. Two principal components are extracted and the main components are ranked in the ranking. Then the 113 cities were classified by cluster analysis, and the 113 cities were divided into 4 types according to the system clustering method. Through the analysis of the source of pollutants according to the regional characteristics of each category, the corresponding control measures of the air pollution situation in China are put forward.

Keywords

Air Quality, Principal Component Analysis, Cluster Analysis

我国环保重点城市空气质量情况分析

唐金玉, 高翔*

中国海洋大学数学科学学院, 山东 青岛
Email: 18354222836@163.com, *gaoxiangshuli@126.com

收稿日期: 2018年5月21日; 录用日期: 2018年6月6日; 发布日期: 2018年6月20日

摘要

空气是人类及生物赖以生存的条件, 因此清洁的空气对人们显得尤为重要。但是, 近年来随着我国工业和交通

*通讯作者。

通运输业的发展,大量的污染物被排放到空气中,空气质量每况愈下,空气质量问题也越来越受到政府和公众的关注。本文选取了《中国统计年鉴2017》[1]中113个环保重点城市的空气质量数据,运用R和SPSS软件进行了统计分析。首先应用主成分分析,对原始数据基于Kaiser-Harris原则[2]进行了降维处理,提取了2个主成分并获取了主成分得分排名;然后用聚类分析方法对113个城市进行了分类,根据系统聚类法将113个城市分为4类,根据每一类的地区特点分析污染物来源,最后对我国空气污染现状提出了相应的治理措施。

关键词

空气质量, 主成分分析, 聚类分析

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

工业的快速发展与城市化导致我国的空气污染形势日益严峻。近年来,我国有越来越多的城市被雾霾问题所困扰,这一现象逐渐引起了国内外媒体与广大学者的关注,我国也开始在治理空气污染方面加大投入。党的十九大报告指出,我们要着力解决突出环境问题。

国内外学者用不同的方法从不同的角度对我国的空气质量情况进行了分析。Kunwar P. Singha [3]等人利用主成分分析和决策树理论对空气质量问题进行了研究;毛宁[4]等人利用全国31个主要城市的空气质量影响因子 SO_2 、 NO_2 、 PM_{10} 进行主成分分析,得出影响城市空气质量的主要因素。赵婉君[5]等人运用主成分分析、因子分析、Bayes判别等统计方法对2009年全国31个主要城市空气质量影响因子 SO_2 、 NO_2 、 PM_{10} 的年平均浓度进行分析,得出31个城市的空气质量等级,对各城市的空气质量进行了比较,并找出了影响各城市空气质量的主要污染物。基于以上研究,本文对113个环保重点城市的空气质量的6个影响因子 SO_2 、 NO_2 、 PM_{10} 、 CO 、 O_3 、 $PM_{2.5}$ 进行主成分分析和聚类分析。

2. 主成分分析

主成分分析[6]也称为主分量分析,利用降维的思想,在损失很少信息的前提下,将多个存在相关关系的变量转化为少数几个综合变量(即主成分)的统计分析方法。

2.1. 相关分析

在进行主成分分析之前,我们先进行变量的相关性分析,相关分析是研究变量间相互关系、测定它们联系的紧密程度的统计方法。为了消除数量级和量纲的不同对结果产生影响,先对原始数据进行了标准。各变量间的相关性如表1。

从表1可以看出有些变量之间存在较强的相关性,如 $PM_{2.5}$ 和 PM_{10} 。因此,接下来我们用主成分分析对变量进行降维处理。

2.2. 确定主成分

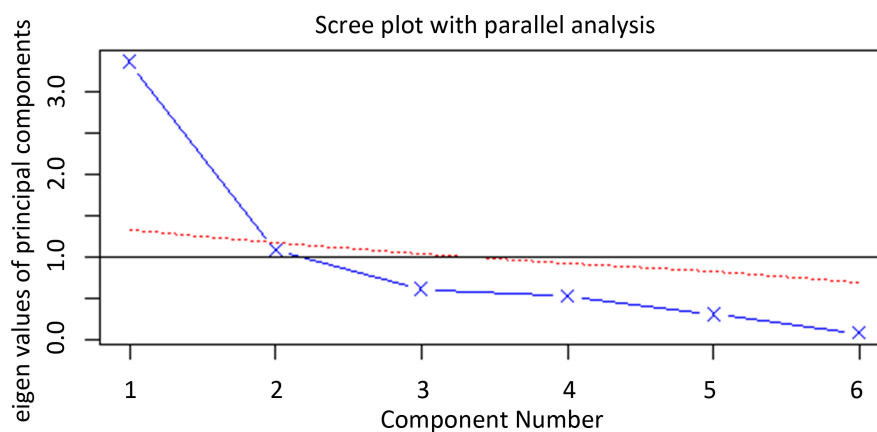
图1展示了基于观测特征值的碎石检验,按照Kaiser-Harris准则,通过碎石图可以看出前2个主成分特征值均大于1,而且前2个主成分以后图形趋于平缓,所以选取前2个主成分作为6个变量的概括。接下来对2个主成分进行提取,计算原始的因子载荷矩阵,结果如表2所示。

Table 1. Correlation matrix**表 1.** 相关系数矩阵

	SO ₂	NO ₂	PM10	CO	O ₃	PM2.5
SO ₂	1	0.291	0.597	0.656	0.281	0.475
NO ₂	0.291	1	0.676	0.574	0.529	0.698
PM10	0.597	0.676	1	0.754	0.529	0.899
CO	0.656	0.574	0.754	1	0.288	0.688
O ₃	0.281	0.529	0.529	0.288	1	0.549
PM2.5	0.475	0.698	0.899	0.688	0.549	1

Table 2. Factor load matrix**表 2.** 因子载荷矩阵

	Comp.1	Comp.2
SO ₂	-0.34128	-0.60183
NO ₂	-0.39993	0.356023
PM10	-0.47646	-0.03315
CO	-0.42283	-0.38653
O ₃	-0.32287	0.589905
PM2.5	-0.46211	0.112043

**Figure 1.** Scree plot**图 1.** 碎石图

可以得出两个主成分与原始变量之间的关系为:

$$F_1 = -0.34SO_2 - 0.4NO_2 - 0.48PM10 - 0.42CO - 0.32O_3 - 0.46PM2.5$$

$$F_2 = -0.6SO_2 + 0.36NO_2 - 0.03PM10 - 0.39CO + 0.59O_3 + 0.11PM2.5$$

当提取了多个成分时, 对选取的主成分进行正交旋转可使所得的结果更容易解释。

图 2 是方差极大旋转的主成分分析的结果, 从上述结果可以看出, 第一主成分与第二主成分与 PM10 和 PM2.5 的相关性都比较大, 不同的是第一主成分与 NO₂ 和 O₃ 的相关性很大, NO₂ 和 O₃ 主要来源于汽车尾气, 汽车排放的氮氧化物, 在合适的气象条件下会生成臭氧, 因而第一主成分可以命名为尾气污染

因子, 第二主成分与 SO_2 和 CO 的相关性很大, SO_2 和 CO 主要源自工业燃烧过程, 因此而第二主成分可以命名为工业污染因子。2 个主成分的方差和占总方差的 80%, 即基本上保留了原来 6 个污染指标的信息。选取的 2 个主成分的方差贡献率和累积方差贡献率如表 3 所示。

2.3. 主成分得分排名

由表 4 可以得出 113 个城市中空气质量前 10 名的城市为海口、湛江、北海、厦门、泉州、福州、深圳、汕头、珠海和克拉玛依, 这些城市中的大多数都位于沿海地区, 第三产业的发展迅速; 空气质量后 10 名的城市为安阳、石家庄、保定、临汾、邯郸、唐山、焦作、淄博、洛阳和太原, 这些城市的第二产业发展比较快, 人口偏多。从城市所在的地理位置来看, 我国中东部地区的空气质量比较差, 而东南部沿海地区的空气质量比较好。从地区的产业结构来看, 第三产业发达的城市的空气质量比第二产业发达的城市的空气质量要好。

3. 聚类分析

聚类分析[7]是研究对样品或指标进行分类的一种多元统计方法, 所谓“类”, 通俗地说就是相似元素的集合。以下是用系统聚类法对 113 个城市地区进行分类的结果:

从表 5 可以看出, 第一类中的城市空气质量最差, 第二类中的城市空气质量比第一类中城市的空气质量相对要好一些, 处于一般水平, 这两类中的城市大多属于中西部地区, 人口都比较密集, 第二产业的发展迅速。要想使这些城市的空气质量从根本上得到改善, 就要优化这些城市的产业结构。第三类和第四类中的城市空气质量都比较好, 其中第四类中的城市空气质量最好, 这两类中的城市大多属于长江、珠三角地区和沿海城市, 这些城市的第三产业比较发达, 工业污染相比前两类城市的污染轻。

	RC1	RC2	h2	u2	com
SO_2	0.07	0.88	0.79	0.21	1.0
NO_2	0.80	0.31	0.74	0.26	1.3
PM10	0.65	0.68	0.89	0.11	2.0
CO	0.33	0.85	0.83	0.17	1.3
O_3	0.85	0.04	0.73	0.27	1.0
PM2.2	0.72	0.56	0.84	0.16	1.9
				RC1	RC2
ss loadings				2.43	2.38
Proportion Var				0.40	0.40
Cumulative Var				0.40	0.80
Proportion Explained				0.50	0.50
Cumulative Proportion				0.50	1.00

Figure 2. Component load matrix

图 2. 成分载荷矩阵

Table 3. Contribution rate table

表 3. 贡献率表

主成分	方差贡献率	累积贡献率
1	65%	65%
2	15%	80%

Table 4. Score ranking
表 4. 得分排名

城市	排名	城市	排名	城市	排名
安阳	1	鞍山	39	烟台	77
石家庄	2	铜川	40	绵阳	78
保定	3	呼和浩特	41	赤峰	79
临汾	4	宝鸡	42	上海	80
邯郸	5	日照	43	南昌	81
唐山	6	大同	44	大连	82
焦作	7	无锡	45	岳阳	83
淄博	8	荆州	46	吉林	84
洛阳	9	常州	47	绍兴	85
太原	10	南京	48	九江	86
长治	11	武汉	49	南充	87
阳泉	12	本溪	50	桂林	88
郑州	13	合肥	51	温州	89
济南	14	抚顺	52	广州	90
咸阳	15	自贡	53	张家界	91
西安	16	宜昌	54	柳州	92
渭南	17	哈尔滨	55	宁波	93
三门峡	18	扬州	56	遵义	94
乌鲁木齐	19	苏州	57	牡丹江	95
银川	20	长春	58	齐齐哈尔	96
泰安	21	镇江	59	韶关	97
济宁	22	芜湖	60	贵阳	98
石嘴山	23	马鞍山	61	南宁	99
平顶山	24	湘潭	62	昆明	100
开封	25	连云港	63	曲靖	101
兰州	26	株洲	64	拉萨	102
北京	27	杭州	65	玉溪	103
枣庄	28	重庆	66	克拉玛依	104
天津	29	青岛	67	珠海	105
潍坊	30	南通	68	汕头	106
徐州	31	金昌	69	深圳	107
西宁	32	攀枝花	70	福州	108
锦州	33	泸州	71	泉州	109
沈阳	34	湖州	72	厦门	110
延安	35	长沙	73	北海	111
包头	36	常德	74	湛江	112
秦皇岛	37	宜宾	75	海口	113
成都	38	德阳	76		

Table 5. Classification results

表 5. 分类结果

类别	城市
第一类	北京、石家庄、唐山、邯郸、保定、阳泉、济南、淄博、枣庄、潍坊、济宁、泰安、 郑州、洛阳、平顶山、安阳、焦作、成都、三门峡、西安、铜川、咸阳、渭南
第二类	太原、长治、临汾、沈阳、锦州、南京、无锡、徐州、常州、苏州、湖州、合肥、日照、 开封、武汉、宜昌、荆州、自贡、泸州、德阳、宝鸡、兰州、银川、石嘴山、乌鲁木齐
第三类	秦皇岛、大同、呼和浩特、包头、赤峰、大连、鞍山、抚顺、本溪、长春、吉林、 哈尔滨、上海、南通、连云港、扬州、镇江、杭州、宁波、绍兴、芜湖、马鞍山、 南昌、九江、青岛、烟台、长沙、株洲、湘潭、岳阳、常德、张家界、广州、 柳州、桂林、重庆、绵阳、南充、宜宾、拉萨、延安、金昌、西宁
第四类	齐齐哈尔、牡丹江、温州、福州、厦门、泉州、韶关、深圳、珠海、汕头、 湛江、南宁、北海、海口、攀枝花、贵阳、遵义、昆明、曲靖、玉溪、克拉玛依

4. 建议

1) 优化产业结构。产业结构能直接影响到这个地区的空气质量。通过上述分析可知, 第三产业发达的城市的空气质量比第二产业发达的城市的空气质量要好。因此要对一些地区的不合理的产业结构进行优化, 大力发展该地区的第三产业, 优化各地区的资源配置。

2) 加大对环境保护的宣传力度。充分利用各种媒体, 大力宣传防治空气污染的相关规定以及空气污染给人们的身体健康带来的危害, 提高公众的防范意识和人们保护环境的自觉性。

3) 完善对企业的管理政策。加强对企业排放污染物的管理, 完善处理污染物的收费制度, 对环保节能产业和现代服务业给予大力支持, 将企业利益与环境问题联系起来, 使企业自觉的合理控制污染物的排放量。

4) 分区域治理。从上面的分析可以看出不同地区的空气污染问题也不同, 各地区应根据本地区的情况进行治理。对于工业发展比较快的地区应着力解决工业污染问题, 提倡使用清洁能源, 改进生产技术以减少污染物的生成。

基金项目

山东省自然科学基金(ZR2018MA006), 山东省研究生教育创新计划项目(SDY15129), 山东省研究生导师指导能力提升项目(SDY17009)。

参考文献

- [1] 中国统计局. 中国统计年鉴[M]. 北京: 中国统计出版社, 2017.
- [2] Kabacoff, R.I. R 语言实战[M]. 王小宁, 等, 译. 第二版. 北京: 人民邮电出版社, 2016.
- [3] Singha, K.P., Gupta, S. and Raia, P. (2013) Identifying Pollution Sources and Predicting Urban Air Quality Using Ensemble Learning Methods. *Atmospheric Environment*, **8**, 426-437.
- [4] 毛宁, 李益祺. 基于主成分分析的全国主要城市空气质量评价[J]. 现代商贸工业, 2014(10): 49-50.
- [5] 赵婉君, 冯良机. 基于环境统计分析的全国主要城市空气质量浅析[J]. 广东化工, 2015, 42(2): 105.
- [6] 何晓群. 多元统计分析[M]. 北京: 中国人民大学出版社, 2105.
- [7] 高惠璇. 应用多元统计分析[M]. 北京: 北京大学出版社, 2005.

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2325-2251，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：sa@hanspub.org