

主成分分析在大气质量监测优化布点中的应用

黄玉平¹, 张庆国^{2*}, 汪水兵³, 古今今⁴

(1. 安徽农业大学生命科学学院, 合肥 230036; 2. 安徽农业大学理学院, 合肥 230036;

3. 安徽省环境科学研究院, 合肥 230061; 4. 北京师范大学环境学院, 北京 100875)

摘要: 主成分分析法 (PCA) 以少数的综合变量取代原有的多维变量, 在原始数据信息丢失最小的情况下, 使数据结构得以简化。作者分析了合肥市新站区大气环境监测数据, 采用主成分分析法建立空气质量污染特征因子与污染物之间的数学模型, 再用该模型计算出各点位相对污染程度, 并对监测布点进行分类。以此选出最佳监测点位, 可为大气质量监测优化布点提供方法, 为合肥市新站区环境质量的分区和分级治理提供理论依据。

关键词: 主成分分析; 合肥市新站区; 大气质量; 监测数据

中图分类号: P468.0

文献标识码: A

文章编号: 1672-352X (2011)06-0966-04

Application of principal components analysis in optimizing sites selection of air quality monitoring

HUANG Yu-ping¹, ZHANG Qing-guo², WANG Shui-bing³, GU Jin-jin⁴

(1. School of Life Science, Anhui Agricultural University, Hefei 230036;

2. School of Science, Anhui Agricultural University, Hefei 230036;

3. Research Institute of Environmental Science of Anhui, Hefei 230061;

4. School of Environment, Beijing Normal University, Beijing, 100875)

Abstract: The principal component analysis could substitute for original multi-dimensional variable by a small number comprehensive variable and simplify data structure under the condition of minimizing loss of original data information. The paper was aimed to analyze the air environment monitoring data of Xinzhan area in Hefei city, using the principal component analysis to establish a mathematical model about the characteristic factors of air quality and pollutants. The model was made to calculate each monitoring sites, relative pollutional degree and classifying monitoring sites, thus to select the best sites to monitor the air quality. The result provided an optimized selection of air quality monitoring sites, a theoretical basis for classifying air quality of Xinzhan area of Hefei and governing by different levels.

Key words: principal component analysis (PCA); Xinzhan area of Hefei; air quality; monitoring data

大气质量监测是环境监测的重要方面, 优化布点是环境监测的基础性工作, 对大气质量监测的效果起重要作用。近年来, 随着多元统计分析的普及和应用, 主成分分析法 (PCA) 已成为一种新兴的评价方法, 在环境质量综合评价方面应用广泛^[1-3]。主成分分析是一种把原来多个指标化为少数几个互不相关的综合指标的多元统计方法, 可以达到数据化简、揭示变量之间关系和进行统计解释的目的^[4-6]。作者将主成分分析法应用于合肥市新站区的

大气质量的综合评价, 旨在为合肥市新站区的空气质量管理工作提供参考依据。

1 材料与方法

1.1 数据标准化

为了排除数量级和量纲不同带来的影响, 首先对原始数据进行标准化处理。

设 p 维随机向量 $x = (x_1, x_2, \dots, x_p)^T$, n 个样品 $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$, $i=1, 2, \dots, n$, $n > p$, 构造样本阵, 对

收稿日期: 2011-03-02

基金项目: 国家自然科学基金项目 (70271062) 和安徽高校省级自然科学研究重点项目 (KJ2010A121) 资助。

作者简介: 黄玉平, 女, 硕士。现在怀宁县环保局工作。

* 通讯作者: 张庆国, 男, 博士, 教授。E-mail: qgzhang@ahau.edu.cn

样本阵元素进行如下标准化变换:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad i=1,2,\dots,p,$$

其中 $\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}$, $s_j^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}$, 得标准化样本阵 $Z = (Z_{ij})$ 。

1.2 相关系数矩阵计算

$$R = (r_{ij})_{p \times p} = \frac{Z^T Z}{n-1},$$

其中 $r_{ij} = \frac{\sum_{k=1}^n z_{ki} z_{kj}}{n-1}$, $i, j=1, 2, \dots, p$ 为相关系数。

1.3 特征值和特征向量计算

解样本相关矩阵 R 的特征方程 $|\lambda E_p - R| = 0$,

得 p 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_p$, 根据累积方差贡献率

$\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^p \lambda_j} \geq 0.85$ 确定主成分的个数 m , 使信息的利用率

达 85% 以上, 对每个 λ_j , $j=1, 2, \dots, m$, 解方程组

$Rb = \lambda_j b$, 得单位特征向量 b_j^0 。

1.4 将标准化后的指标变量转换为主成分

$$U_{ij} = z_{ij}^T b_j^0, \quad j=1, 2, \dots, m$$

U_1 称为第一主成分, U_2 称为第二主成分, \dots , U_m 称为第 m 主成分。

1.5 对 m 个主成分进行综合评价

对 m 个主成分进行加权求和, 即得最终评价, 权数为每个主成分的方差贡献率。运用 SPSS 软件中的 Analyze 和 Transform 可以快速准确的进行主成分分析^[7-8]。

2 研究区域概况

合肥新站综合开发试验区是安徽省政府批准设立的省级开发试验区, 享有市级经济管理权限, 位于合肥市区北部, 紧邻合肥老城区。规划范围为合肥新站综合开发试验区管辖范围, 四周界线为: 东起二十埠河, 西至板桥河, 南接合肥市瑶海区, 北至长丰县界, 总面积为 33.67 km²。

3 结果与讨论

监测区共设 7 个监测点, 每个测点的监测指标为 7 项, 每天 24 h 连续监测 5 d, 计算出 5 d 的平均值 (表 1)。对原始数据进行标准化处理后, 求出相关系数矩阵的特征值 (表 2)。由表 2 可见, 前 2 个特征值的累积贡献率已达到 85%。因此, 选取前 2 个主成分即可。至此, 5 d、7 个监测点、7 项指标测定值的全部信息的 87.567% 已被归纳为 2 个主成分。可利用它们对合肥市新站区空气环境质量进行综合评价。

表 1 各监测点监测指标 5 d 平均值

Table 1 The average value of 5 days of monitoring indexes at each monitoring site

监测点 Monitoring site	SO ₂ /mg·L ⁻¹	NO ₂ /mg·L ⁻¹	PM10 /mg·L ⁻¹	TSP /mg·L ⁻¹	苯/ $\mu\text{g}\cdot\text{L}^{-1}$ Benzene	甲苯/ $\mu\text{g}\cdot\text{L}^{-1}$ Methylenzene	二甲苯/ $\mu\text{g}\cdot\text{L}^{-1}$ Dimethylenzene
1	0.030 0	0.039 0	0.119 6	0.167 6	0.012 32	0.001 58	0.001 56
2	0.032 6	0.042 4	0.120 2	0.171 2	0.010 22	0.001 04	0.001 18
3	0.031 8	0.040 8	0.121 0	0.173 0	0.010 30	0.001 14	0.000 96
4	0.031 8	0.041 8	0.123 0	0.175 6	0.010 46	0.002 80	0.001 32
5	0.036 8	0.041 0	0.124 0	0.177 2	0.012 46	0.004 22	0.004 40
6	0.031 8	0.040 4	0.121 2	0.173 0	0.012 36	0.005 04	0.004 12
7	0.039 6	0.040 4	0.124 6	0.178 0	0.012 02	0.005 02	0.002 78

各主成分线性表达式中原始指标的系数取相应特征值对应的正规化单位特征向量, 即可构造出 2 个主成分。但是初始因子载荷矩阵并不满足“结构简单准则”, 各因子的典型代表变量不是很突出, 因而容易使因子的含义模糊不清, 不便于对因子进行解释。因此, 需对因子载荷矩阵进行旋转, 使得因子载荷的平方按列向 0 和 1 两极分化, 达到其结构

简化的目的^[9]。主成分的特征向量值见表 3。根据表 3, 可得到主成分方程 y_1 和 y_2 :

$$y_1 = 0.882x_1 + 0.323x_2 + 0.959x_3 + 0.978x_4 + 0.203x_5 + 0.742x_6 + 0.565x_7$$

$$y_2 = 0.124x_1 - 0.827x_2 + 0.054x_3 - 0.094x_4 + 0.965x_5 + 0.581x_6 + 0.693x_7$$

表 2 相关系数矩阵的特征值
Table 2 Eigenvalues of correlation coefficient matrix

序号 No.	特征值 Eigenvalue	贡献率%Contribution rate	累积贡献率%Accumulated contribution rate
1	4.084	58.342	58.342
2	2.046	29.225	87.567
3	0.525	7.505	95.072
4	0.215	3.074	98.147
5	0.088	1.256	99.403
6	0.042	0.597	100.000
7	-2.06×10 ¹⁶	-2.95×10 ¹⁵	100.000

表 3 两个主成分的特征向量
Table 3 Eigenvector of two principal components

主成分 Principal component	特征向量Eigenvector						
	SO ₂	NO ₂	PM10	TSP	苯 Benzene	甲苯 Methylenzene	二甲苯 Dimethylenzene
I	0.882	0.323	0.959	0.978	0.203	0.742	0.565
II	0.124	-0.827	0.054	-0.094	0.965	0.581	0.693

表 4 空气质量分类标准（日平均）
Table 4 Classification standards of air quality (average by day)

类别 Classes	SO ₂ /mg·L ⁻¹	NO ₂ /mg·L ⁻¹	PM10/mg·L ⁻¹	TSP/mg·L ⁻¹	苯/μg·L ⁻¹ Benzene	甲苯/μg·L ⁻¹ Methylenzene	二甲苯/μg·L ⁻¹ Dimethylenzene
一类 I type	0.05	0.08	0.05	0.12	0.01	0.01	0.01
二类 II type	0.15	0.08	0.15	0.30			
三类 III type	0.25	0.12	0.25	0.50			
三类 III type	0.25	0.12	0.25	0.50			

表 5 合肥市新站区空气质量主成分分析结果
Table 5 The analysis results of air quality of Xinzhan area in Hefei by PCA

监测点 Monitoring site	第一主成分得分F ₁ Score by the first principal component	第二主成分得分F ₂ Score by the second principal component	综合得分F Comprehensive score	污染程度排名 Rank for pollution degree	空气质量分类 Classification of air quality
1	-1.589 29	1.065 80	-0.703 10	7	三类 III type
2	-0.475 39	-1.267 65	-0.406 05	5	三类 III type
3	-0.537 99	-0.775 89	-0.617 39	6	三类 III type
4	0.293 31	-1.016 71	-0.143 93	4	三类 III type
5	1.074 21	0.565 98	0.904 58	2	三类 III type
6	-0.073 79	1.088 95	0.314 30	3	三类 III type
7	1.308 93	0.339 51	0.985 37	1	三类 III type

各主成分得分与对应的方差贡献率乘积的总和即为综合得分。计算出 7 个监测点及三类大气环境质量标准的各主成分得分及综合得分，给予各监测点大气环境质量状况以定量化描述，得分越大，表明污染程度越严重，由此可依据空气质量标准 GB3095-1996（日平均）对各监测点的大气环境质量状况进行排序和分级，分级标准见表 4。

对照表 4 的分类标准，从表 1 可以看出各个监测点的空气质量应属于不同的类别，但表 5 的分析

结果却显示 7 个监测点的空气质量全部属于三类，这主要是由于 7 个监测点中的苯全部超过《空气质量标准》GB3095-1996 中的三级标准所致（表 4）。

在第一主成分 y₁ 中，SO₂、PM10、TSP、甲苯的绝对值较大，对空气质量起主导作用；在第二主成分 y₂ 中，NO₂、苯、二甲苯的绝对值较大，是影响空气质量的主要因子。由于各主成分相互独立，因此，2 个主成分反映了两种污染机制：第一主成分的贡献率为 58.342%，且主导因子主要是反映空

气无机污染的指标,这说明合肥市新站区的空气污染主要是由无机物造成的;其中主要是由于合肥现阶段城市建设过程中建筑工地产生的粉尘污染及扬尘,以及城区工厂生产中煤和石油燃烧产生的 SO_2 和甲苯。第二主成分的贡献率为 29.225%, 主要影响因子是苯、 NO_2 和二甲苯,它们反映的主要是空气有机物污染。其原因是由于机动车尾气排放引起的。概括起来合肥市新站综合开发区的空气污染主要为三类污染:一是扬尘污染,二是煤烟污染,三是汽车尾气污染。前两者为主要污染要素。

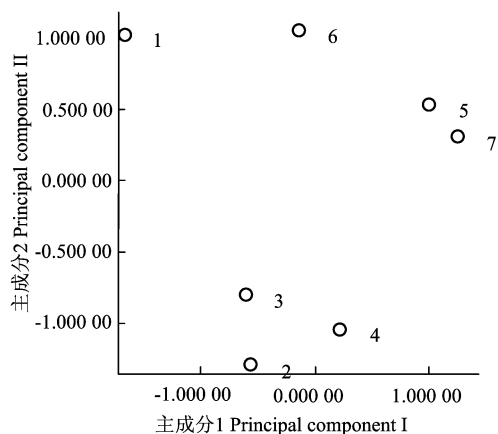


图 1 环境监测点的主成分分类

Figure 1 Classification of environmental monitoring sites according to the PCA

以第一主成分为横坐标,第二主成分为纵坐标。将 7 个测点在坐标图中标出,便得到图 1 所示的分类图。由图 1 可见,7 个测点被分成 4 类:(2、3、4), (5、7), (1) 和 (6)。这 4 类在二维平面上是按照一定的方向和顺序依次排列的。从左至右,空气污染程度逐渐增强;自上而下,污染程度逐渐减弱。从各特征值的贡献率可知,第 7 种污染物的贡

献率很小,对环境质量影响不大,因此可以不监测。根据 2、3、4 号监测点的现场情况,挑选 4 号监测点作为最优监测点;5、7 号监测点选择 5 号监测点作为最优监测点。这样监测点的数量从原来的 7 个减少到 4 个,而监测效果和信息采集程度与原来区别不大,大大地节约了监测成本。

采样是环境监测工作的首要问题。环境数据的采集既要降低采样成本,又要能反映环境污染状况,因此监测点的数量确定、分布优化是一个重要的研究课题^[10-11]。本文的分析结果表明,主成分分析法可为环境监测点的优化选择提供理论依据。

参考文献:

- [1] 蒲文龙, 郭守泉. 主成分分析法在环境监测点优化中的应用[J]. 煤矿开采, 2004, 9(4): 6-7.
- [2] 冯利华. 环境质量的主成分分析[J]. 数学的实践与认识, 2003, 33(8): 32-35.
- [3] 鲁斐, 李磊. 主成分分析法在辽河水水质评价中的应用[J]. 水利科技与经济, 2006, 12(10): 660-662.
- [4] 王晓鹏. 河流水质综合评价之主成分分析法[J]. 数理统计与管理, 2001, 20(4): 49-52.
- [5] 方开泰. 实用多元统计分析[M]. 上海: 华东师范大学出版社, 1989.
- [6] 吴翊, 李永乐, 胡庆军. 应用数理统计[M]. 长沙: 国防科技大学出版社, 1995.
- [7] 袁志发, 周静芋. 多元统计分析[M]. 北京: 科学出版社, 2006.
- [8] 卢文岱. SPSS for Windows 统计分析[M]. 北京: 电子工业出版社, 2003.
- [9] 罗积玉, 邢瑛. 经济统计分析方法及预测[M]. 北京: 清华大学出版社, 1985.
- [10] 常凤池. 某黄金矿山开发的环境污染问题及评价方法[J]. 有色金属矿产与勘查, 1994, 3(5): 317-322.
- [11] 古今今, 张庆国, 汪水兵, 等. 基于综合主成分及径向基网络的环境质量评价[J]. 环境科学与技术, 2010, 33(7): 196-200; 205.