

长春市水污染控制指标的因子和聚类分析

高成康¹ ,尚金城² ,蒋大和¹

(1. 同济大学环境科学与工程学院 ,上海 200092 ;
2. 东北师范大学环境科学研究所 ,吉林 长春 130024)

摘要 根据长春市水环境监测的实际数据 ,运用因子分析法 ,选取主要的 8 项代表指标、12 个监控断面 ,并依据因子得分对长春市水环境进行聚类分析。根据聚类结果 ,整体把握长春市水环境的空间特征 ,并针对各区域水环境的污染成因和污染因子提出相应微观控制措施 ,为长春市的宏观规划提供依据。

关键词 水环境 ,因子分析 ,聚类分析 ,微观控制措施 ,长春市

中图分类号 :X52 文献标识码 :A 文章编号 :1004-693X(2004)06-0028-03

长春市的水资源问题制约着整个城市的发展 ,目前多数工作从整治大的江河入手 ,忽视了小河流和水库 ,但它们却形成了一个完整的“活”水系 ,影响着整个市区的环境景观和人民的健康。为此 ,运用因子分析方法来分析长春市水环境的水质状况及其污染成因 ,并用聚类分析方法进行验证和空间特征分析 ,提出各水环境相应的微观控制措施和方向。

1 水环境的因子分析

1.1 因子分析原理

因子分析是通过原始变量的线性组合 ,把多个原始指标简化为有代表意义的少数几个指标 ,以使原始指标能更集中、更典型地表明研究对象特征的一种统计方法。其基本原理为 :①建立 n 个区域 p 个指标的原始数据矩阵 $M = M_{ij} (i = 1, 2, \dots, n ; j = 1, 2, \dots, p)$,并对其进行无量纲化或标准化处理。②计算指标的相关系数矩阵 ,求出特征值 λ ,得到贡献率和累积贡献率 ,选取前 $m (m < p)$ 个主因子 ,使得累积贡献率 $\sum_{j=1}^m \lambda_j / \sum_{j=1}^p \lambda_j \geq 85\%$,然后进行具体分析。

这些运算应用 SPSS 软件即可完成 ,结果结合实际情况分析。

1.2 实例分析

1.2.1 断面选取

长春市水系(不包括伊通河)涉及到整个市区(包括郊区)。所选监测断面有 :兴隆水库(1[#])、杨

家粉房水库(2[#])、梁家水库(3[#])、红旗水库(4[#])、新月水库(5[#])、四间房水库(6[#])、小河沿子水库(7[#])、杨家店(8[#])、鲶鱼沟(9[#])、鸭子湾(10[#])、小白桥(11[#])、长吉高速路口(12[#])。其中因本项目目的是恢复长春市水系 ,有些水库几乎干枯 ,但仍是水系的一部分 ,所以也作为监控断面 ,而伊通河已有治理方案 ,故不做考虑对象。

1.2.2 指标选取

长春市水环境接纳污水为生活污水和有机工业污水 ,水环境监测结果见表 1。

表 1 长春市水环境监测结果								mg/L
监测断面	DO	COD _{Mn}	BOD	TN	NH ₃ -N	TP	SS	石油类
1 [#]	9.78	20.60	6.88	4.70	1.90	0.65	25.20	0.13
2 [#]	0.42	120.80	24.17	21.80	18.00	2.90	29.00	1.25
3 [#]	3.02	424.80	141.60	44.81	41.80	1.56	15.70	1.56
4 [#]	11.86	42.60	10.64	24.50	4.20	1.34	275.00	1.12
5 [#]	0.00	1733.40	577.80	98.10	53.80	6.10	589.00	1.05
6 [#]	0.00	455.70	151.90	59.50	49.30	3.25	76.00	1.68
7 [#]	7.28	11.50	2.87	10.80	2.00	1.45	20.70	1.35
8 [#]	10.71	7.30	1.46	2.10	0.67	1.86	338.00	1.41
9 [#]	0.00	77.50	15.50	21.70	8.76	2.37	10.40	1.28
10 [#]	4.68	17.90	3.58	26.90	9.75	1.30	40.30	1.45
11 [#]	7.07	23.80	4.75	34.40	8.69	7.50	4904.00	1.35
12 [#]	9.67	102.30	25.58	53.20	23.10	1.30	11.70	1.61

数据来自 2002 年的监测结果 ,均为年平均值。

1.2.3 计算结果

因子分析法是处理多种变量数据的统计分析方法 ,可将相互间有联系的多变量降维简化 ,分析高维空间中样本分布规律 ,而且可提取综合性好、代表性强、信息损失量较小的综合因子。原始数据来自实

表 2 特征值和累计贡献率

因子	旋转前						旋转后		
	初始的			提取的			提取的		
	特征值	变量/%	累计贡献率/%	特征值	变量/%	累计贡献率/%	特征值	变量/%	累计贡献率/%
1	4.360	54.496	54.496	4.360	54.496	54.496	4.138	51.729	51.729
2	1.663	20.790	75.286	1.663	20.790	75.286	1.180	22.450	74.179
3	1.157	14.460	89.746	1.157	14.460	89.746	1.245	15.567	89.746
4	0.572	7.145	96.892						
5	0.172	2.146	99.038						
6	5.962×10^{-2}	0.745	99.783						
7	1.717×10^{-2}	0.215	99.998						
8	1.762×10^{-4}	2.202×10^{-3}	100						

表 3 载荷矩阵

指标	因子			旋转因子		
	1	2	3	1	2	3
TN	0.942	0.000	0.000	0.904	0.189	0.192
COD	0.930	-0.121	-0.297	0.969	0.000	-0.137
BOD	0.926	-0.119	-0.303	0.966	0.000	-0.144
NH ₃ -N	0.910	-0.231	0.159	0.900	0.000	0.310
DO	-0.681	0.167	-0.274	-0.652	0.000	-0.386
SS	0.000	0.988	0.000	-0.142	0.981	0.000
TP	0.627	0.759	0.000	0.438	0.880	0.000
石油类	0.225	0.000	0.935	0.000	0.000	0.959

测结果(表 1),所选断面和指标可建立 12×8 的原始指标矩阵 $x=(x_{ij})_{np}$ 。虽然所选指标量纲相同,但为消除因数据过分悬殊导致的噪音影响,需对原始数据进行标准化处理,而且对于不同类型的指标采用了不同的标准化方法^[1]。根据相关矩阵可知,各因子间的相关性较好,适合用因子分析法提取综合因子^[2]。旋转前后各因子的特征值和累计贡献率见表 2。

因子分析中选取因子的原则:特征值大于 1 的因子,因子的累计贡献率大于或等于 85%,即 $\sum_{i=1}^m \frac{\lambda_i}{p} \geq 85\%$ 。从表 2 和图 1 看出,选取前 3 个因子可满足 $\sum_{j=1}^m \lambda_j / \sum_{j=1}^p \lambda_j = 89.746\% \geq 85\%$ 。

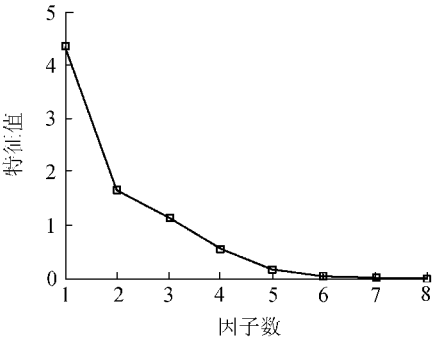


图 1 因子碎石

提取的 3 个因子代表了 8 个因子的综合信息,因而很难命名。为此对其旋转,使因子载荷值向两

极端趋近,以明确各因子代表的含义,采用 Varimax with Kaiser Normalization 因子旋转法对初始因子载荷矩阵施以 25 次的正交旋转。从表 3 可看出因子分类极其明确。同时从表 2 又可知旋转前后综合因子代表的信息量大于或等于 85%,为此满足要求,可认为信息量没有损失;所提取的综合因子的相关性极小即独立性很强^[3]。

由表 3 和图 2 可看出,旋转使因子载荷值向两极端趋近更为明显,利于综合因子的命名。第一主因子代表 BOD、TN、NH₃-N、DO,其与 DO 呈负相关性,说明 DO 愈大该因子的贡献率愈小,占原始方差的 54.496%,命名为有机工业污染因子;第二主因子代表 SS 和 TP,占原始方差的 20.790%,这些污染物多为生活污水产生的,可命名为生活污染因子;第三因子只代表石油类,可直接称之为石油因子,占原始方差的 14.46%。

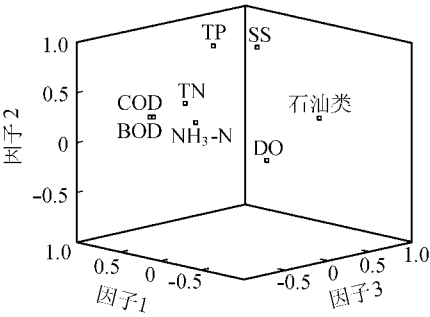


图 2 旋转因子载荷量

2 水环境的聚类分析

由上述计算,得到各断面 3 个综合因子的得分见表 4,它说明了不同污染因子对各断面水体的贡献程度,由此提出具有针对性的、合理化的微观优化控制措施和方向。为了更直观地观察结果,现对其进行聚类。以断面的因子得分为变量,采用样本聚类法(hierachical cluster),在聚类过程中,聚类方法为最远邻近法(Further neighbor),距离测度的方法采用欧式距离平方作为类间距离(squared euclidean

distance)。图 3 以综合因子得分为变量做聚类的三维图,图 4 用统计软件 SPSS 做聚类分析所得出的垂直树状分类图。从图 3、4 中可看出二者结论是一致的,11#、5#、1# 断面都明显地优先分离出来,所有断面形成不同污染类型。

表 4 因子得分

监测断面	有机工业污染因子	生活污染因子	石油因子
1#	-0.52027	-0.55897	-2.52643
2#	-0.05849	-0.22679	0.34070
3#	0.55519	-0.59693	0.78965
4#	-0.60940	-0.23448	-0.62394
5#	2.72183	0.54946	-0.99762
6#	0.87096	-0.24963	1.26843
7#	-0.65769	-0.38676	0.06364
8#	-0.87444	-0.11241	0.04180
9#	-0.19914	-0.31449	0.37457
10#	-0.39712	-0.46638	0.51492
11#	-0.67949	3.03201	0.17220
12#	-0.15195	-0.43463	0.66567

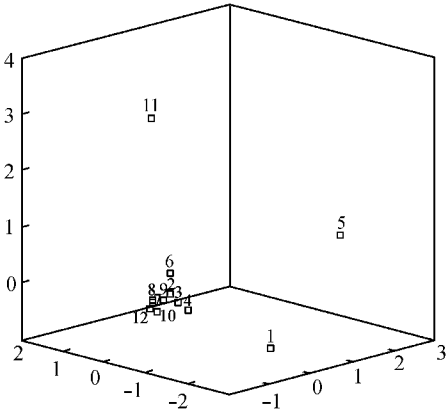


图 3 断面因子得分

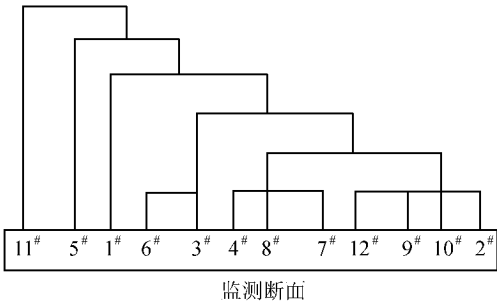


图 4 SPSS 聚类树状分类

据表 4、图 3,可将断面的水质污染类型分为 4 类:第 1 类为小白桥(11#),主要受生活污染因子控制,污染物以 TP 和 SS 为主;第 2 类为新月水库(5#),主要受有机工业污染因子影响,污染物为 BOD、TN、NH₃-N、DO 等,而且生活污染因子对此断面的影响也较大;第 3 类为兴隆水库(1#),各污染因子的贡献都较小,水质相对较好;第 4 类为其他水体,石油因子影响稍大,与其他两类综合因子有负相关性。

3 水环境分析与微观控制措施

3.1 第 1 类水体:小白桥

小白桥在长春市新开河下游与伊通河交汇处,位于城市远郊,附近无工业污水排入,污染多来自于生活污水,TP 和 SS 的质量浓度最大值分别为 7.50 mg/L、4904 mg/L,均属严重超 V 类。针对这种情况可采用一级处理(或物理处理)和深度脱磷的污水处理方法,为从源头控制污染,还应限制使用含磷较高的洗涤剂,据此可采取一些更有效的、经济的具体措施。

3.2 第 2 类水体:新月水库

新月水库是长春市串库和宋家洼子水库的代表水体,尤其是串库,它是由长春市多个水库天然相连所形成的奇特景观,也是本项目治理的主要目标。串库位于长春市西北方,远离市中心且附近多农田用地,地域比较开阔,占地面积约 1493 hm²,其中水面面积为 14 hm²。污染较为严重,除石油外其他因子均超 V 类,但污染物几乎全为有机物。因为水体接纳的主要是生活污水,且工业污水以大成玉米加工公司排放的污水为主,它排放的污染物也是有机物。为此,可采用土地处理系统代替污水处理厂来解决此区域大量的水体污染。串库接纳污水量现为 7 万 t/d,预计 2020 年达 14 万 t/d,拟建土地处理系统规模为 10 万 t/次(5 d 为一个运行周期),采用两个快速渗滤土地处理工艺系统串联的形式,并充分利用自然条件减少工程费用(即把四间房水库作为清水池),提高经济作物的经济效益。

宋家洼子水库与新月水库的污染成因或治理措施都是相似的,现接纳污水量 1.3 万 t/d,2020 年达 3 万 t/d,所以设计土地处理系统的规模为 2 万 t/d,用梁家水库作清水池,以一级、二级之间的处理为主,处理水满足梁家水库和纪家水库及伊通河的部分景观用水^[4]。该项目建成后其经济效益是最为可观的:直接经济效益 1795.8 万元,渔业、农业经济效益分别为 8.4 万元、25.55 万元,人体健康效益 25.35 万元,改善土壤理化性质价值、改善生态环境价值分别为 2.7 万元、2.42 万元,总计该土地处理系统运行后所带来的总效益为 3636.79 万元^[5,6]。

总之,在此区域采取土地处理系统,不仅为长春市远期规划(建立两大公园)提供了必要的水资源,而且使整个区域都在水体、绿化的服务半径内,所以是个优化的可行方案。

3.3 第 3 类水体:兴隆水库

兴隆水库地处高新技术开发区,污染源较少,只是 N、P 略有偏高,多是农用化肥和居民洗涤剂产生的,只要保持好目前的水质, (下转第 34 页)

$$S_1 = (3.2 \quad 2.4 \quad 1.6 \quad 0.8 \quad 0)^T$$

将 6.26 进行等差间隔, 间差为 1.565, 得到水资源价格向量

$$S_2 = (6.26 \quad 4.695 \quad 3.13 \quad 1.565 \quad 0)^T$$

将 S_1, S_2 代入式 (6), 上海市 2001 年水资源价格计算为

$$W_1 = B \times S_1 = 1.816 \text{ 元/m}^3$$

$$W_2 = B \times S_2 = 3.553 \text{ 元/m}^3$$

4 结 论

考虑水量、水质及社会经济因素的综合影响, 上海市水资源价值等级“高、偏高、一般、偏低、低”的隶属度分别为: 0.308, 0.154, 0.192, 0.192, 0.154, 因此, 简单地将其判断为属于某一个等级是不合适的。根据水资源价值的综合评价结果, 上海市 2001 年生活用水资源本身的价格应为 1.816 元/m³, 如果考虑供水单位的供水成本及正常利润, 其生活用水水价应为 2.805 元/m³, 如果考虑污水处理费用, 水价应为 3.553 元/m³。

从水资源价值系统实际存在的复杂性和模糊性出发, 水资源价值模糊综合评价模型综合考虑了水量、水质、社会经济因素对水资源价值的影响, 评价

结果较之简单的判定水资源价值属于哪一个等级更合理。同时, 水资源价值模糊综合评价模型本身较好地考虑水资源价值随水量、水质及社会经济因素而变化的水资源价值系统的动态性。该模型将水资源价格与水的用途效益联系起来, 使水资源价格更趋合理, 从而有利于促进水资源优化配置目标的实现。

参考文献:

- [1] 王建生. 水资源紧缺成都评价指标[J]. 水利计划, 1998, (2): 27~30.
- [2] 李广贤. 水资源利用工程与管理[M]. 北京: 清华大学出版社, 1998. 22~24.
- [3] 姜文来. 水资源价值模型研究[J]. 资源科学, 1998 (4): 21~23.
- [4] 顾圣平. 水资源模糊定价模型[J]. 水利发展研究, 2002, (1): 17~19.
- [5] 陈鸣钊. 模糊数学及其应用[M]. 南京: 河海大学出版社, 1993. 7~8.
- [6] 齐欢. 数学模型方法[M]. 武汉: 华中理工大学出版社, 1996. 2~4.
- [7] 梁梁, 盛昭瀚, 徐南荣. 一种改进的层次分析法[J]. 系统工程, 1989 (6): 33~74.

(收稿日期 2003-08-25 编辑 舒建)

(上接第 30 页)

即能满足经济发展所需的硬环境。

3.4 第 4 类水体: 其他

此类水体污染因子主要是石油因子, 受工业和交通的影响较大, 尤其是长吉高速公路监测断面, 石油质量浓度高达 1.6 mg/L, 因其地处交通要道, 时常有私人洗车经营, 严重污染了此处水体; 其他水体多位于长春市汽车厂附近, 石油因子也较高。这些水体的治理可针对去除石油污染物来设计污水处理技术, 避免无效投资。

4 结 论

利用因子分析和聚类分析法进行长春市水环境分析, 可明确水环境的污染原因和水环境的空间特征, 针对不同水体提出相应的治理措施, 为本项目水环境的宏观规划提供了科学依据, 节省了大量的无效投资, 并能准确查出主要污染贡献, 以便有针对性地减少污染物排放, 采取必要的微观措施。

参考文献:

- [1] 卢纹岱. SPSS for Windows 统计软件[M]. 北京: 电子工业出版社, 2000. 92~237.

- [2] 王茂军, 张学霞. 中国区域可持续发展水平的时空分异初探[J]. 中国人口、资源与环境, 2001, 11(2): 80~85.
- [3] 王晓鹏. 河流水质综合评价之主成分分析方法[J]. 数理统计与管理, 2001, 20(4): 49~52.
- [4] 国家环境保护局科技标准司. 城市污水土地处理技术指南[M]. 北京: 中国环境科学出版社, 1997. 60~79.
- [5] 郝明家, 王莹. 城市水污染集中控制指南[M]. 北京: 中国环境科学出版社, 1996. 230~237, 292~318.
- [6] 郭扬善. 城市污水处理——投资与决策[M]. 北京: 中国环境科学出版社, 1992. 179~206.

(收稿日期 2004-02-20 编辑 舒建)

