

主成分分析法(PCA)在藻类生长影响分析中的应用

路晓波¹, 赵珊², 周军², 甘一萍², 周律³, 刘晶晶³

(1.山西长治市供水总公司,山西长治 046000 2.北京城市排水集团有限责任公司,北京 100022 3.清华大学环境科学与工程系,北京 100084)

摘要 水华的发生、发展是一个相当复杂的过程,众多影响因素共同导致藻类生长、水华爆发。常用的分析工具难以分析、预测实际情况下藻类的生长。该文利用当今世界上最大型的统计与图表分析软件之一 STATISTICA 对 2008 年和 2009 年奥运森林公园主湖的监测数据进行主成分分析。结果表明 STATISTICA 软件中的主成分分析适用于藻类生长分析,并发现除了已知的藻类生长影响因素(营养盐、温度等)之外,景观湖中底泥的累积效应也应纳入分析之内。

关键词 主成分分析法 再生水 藻类

中图分类号 X524

文献标识码 B

文章编号 :1009-0177(2010)06-0090-03

Application of Principal Component Analysis on Algae Growth Influences

Lu Xiaobo¹, Zhao Shan², Zhou Jun², Gan Yiping², Zhou Lu³, Liu Jingjing³

(1. Changzhi Waterworks Company, Changzhi 046000, China;

2. Beijing Drainage Group, Beijing 100022, China;

3. Department of Environment Science & Engineering, Tsinghua University, Beijing 100084, China)

Abstract Breakout and expanding of algae bloom is complex. Many influence factors cause algae growth and algae bloom. Common analytical tool is difficulty to analyze and predict algae growth. In this paper, the monitoring data of lake in Olympic Park from 2008 to 2009 was measured by using STATISTICA software. It was found principal component analysis system of STATISTICA software can analyze algae growth. The results show cumulative effect must be considered except for the factors such as nutrient and temperature, which are known.

Key words principal component analysis reclaimed water algae

由于水资源短缺,再生水作为第二水源的地位日渐凸现。再生水的回用途径之一就是景观利用。由于再生水污染物本底值较饮用水高,北方景观水体流动缓慢甚至完全静止,水深较浅、阳光易于投射等原因,存在爆发水华的危险。

水华通常是指淡水池塘、河流、湖泊、水库等水体受到污染,氮磷等营养物质大量增加,致使水体达到富营养化或者严重富营养化状态,在一定的温度、光照、水流流速等条件下,某些藻类发生暴发性的繁殖,引起明显的水色变化,并在水面形成或厚或薄的

绿色或其他颜色的藻类漂浮现象^[1]。

一般认为,水温、水文、气候、气象等条件可通过影响湖泊水体的分层、混合以及光照、营养盐的可利用性等,从而直接或间接地影响不同种类藻的细胞密度、种群组成、垂直分布、生命周期等。营养盐的浓度,特别是磷浓度及氮磷比显著地影响或反映了浮游植物的种群组成^[2]。

1 材料与方法

1.1 样品采集

本文的研究对象为奥运森林公园水面,它是世界上迄今为止规模最大的以再生水为主要水源的人工水景,占地面积 60 ha。其中奥运森林公园北区水面约 50 ha,日需补充水量为 6.5 万 t。按照重点考察水华最易爆发和水质最具代表性的布点原则,在奥运森林公园主湖区布置 6 个调查监测采样点,分别为①湿地入口;②湿地出口;③主湖入口;④临水广场(雨水口);⑤主湖船闸;⑥洼里湖死点。

[收稿日期] 2010-04-02

[基金项目] 国家科技支撑计划、奥运景观水系水质保障综合技术与示范(2007BAC22B02);北京市科委资助重大科技项目、北京城市污水处理及再生水质提高关键技术研究及工程示范(D07050601500000)

[作者简介] 路晓波(1963-)男,高级工程师,从事水产和污水处理厂研究和管理。电话 0355-2189090;

E-mail luxiaobo@waterchina.cn。

[通讯作者] 赵珊 电话 010-51352882 E-mail zhaoshan@bdc.cn。

第一阶段采样时间为 2008 年 04 月到 2008 年 10 月, 第二阶段采样时间为 2009 年 05 月到 2009 年 9 月, 共检测到数据 3 000 余个。检测项目及所用方法如表 1 所示。本文重点研究藻类生长重要营养因子——水温、溶解性总磷、氨氮、溶因子——水温、溶解性总磷、氨氮、溶解性总固体对叶绿素的影响, 并利用叶绿素指示藻类生长的情况。

表 1 奥运森林公园主湖区检测项目及方法列表
Tab.1 Detecting Items and Methods of Olympic Lake in Beijing

项目	所用检测方法
溶解性总磷	钼酸铵分光光度法 GB/T 11893—1989
氨氮	纳氏试剂光度法 GB/T 7479—1987
溶解性总固体	重量法(180℃) GB/T 11901—1989
水温	温度计法 GB/T 13195—1991
叶绿素	叶绿素 a, b 和 c 的分光光度法测定(三色法) ^[3]

1.2 分析方法

分析软件选用 STATISTICA, 为当今世界上最大的统计与图表分析软件之一。

主成分分析也称主分量分析, 旨在利用降维的思想, 把多指标转化为少数几个综合指标。主成分分析法(PCA)是一种数学变换的方法, 它把给定的一组相关变量通过线性变换转成另一组不相关的变量, 这些新的变量按照方差依次递减的顺序排列。在数学变换中保持变量的总方差不变, 使第一变量具有最大的方差, 称为第一主成分, 第二变量的方差次大, 并且和第一变量不相关, 称为第二主成分。依次类推, I 个变量就有 I 个主成分^[4]。

该方法的统计学原理是, 寻找 $r(r < n)$ 个反映系统主要特征的新变量, 以压缩原有数据矩阵的规模。 r 个新变量即所谓的“主成分”(Factor), 每个主成分均是原有变量的线性组合, 并在很大程度上体现原有变量的综合效果, 彼此正交独立, 且具有一定的实际含义。主成分的这种“主要性”以特征值的形式表征出来, 即特征值越大, 其反映的系统信息就越多, 因而也就越“主要”。通常选择两个最主要的主成分, 分别作为横纵坐标轴建立二维向量体系, 并将所有变量投影到该二维空间中, 以直观体现各变量之间的相互关系, 其相关程度以变量之间的相对空间位置表征出来^[5]。

2 结果与分析

2008 年、2009 年监测数据如表 2、3 所示。

表 2 2008 年监测数据列表

Tab.2 Detecting Data of 2008

2008 年	水温 /℃	溶解性总磷 / $(\text{mg}\cdot\text{L}^{-1})$	氨氮 / $(\text{mg}\cdot\text{L}^{-1})$	溶解性总固体 / $(\text{mg}\cdot\text{L}^{-1})$	叶绿素 / $(\text{mg}\cdot\text{L}^{-3})$
变化范围	7.8~31.2	<0.025~0.886	0.02~3.56	218~612	1.06~391.8
平均值	24.5	0.05	0.24	428	37.5

表 3 2009 年监测数据列表

Tab.2 Detecting Data of 2009

2008 年	水温 /℃	溶解性总磷 / $(\text{mg}\cdot\text{L}^{-1})$	氨氮 / $(\text{mg}\cdot\text{L}^{-1})$	溶解性总固体 / $(\text{mg}\cdot\text{L}^{-1})$	叶绿素 / $(\text{mg}\cdot\text{L}^{-3})$
变化范围	15.2~30	0.01~0.811	0.03~1.16	355~882	1.81~362
平均值	25.3	0.06	0.31	487	60.0

图 1 为 2008 年影响因子与叶绿素的 PCA 分析。

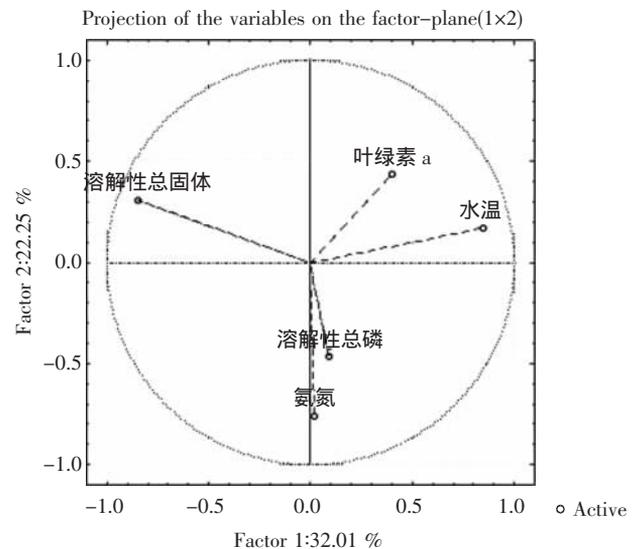


图 1 2008 年影响因子与叶绿素的 PCA 分析

Fig.1 PCA Analysis of Influencing Factors and Chlorophyll in 2008

主成分(Factor)是原有变量的线性组合, 体现原有变量的综合效果, 彼此正交独立。通常选择两个最主要的主成分, 分别作为横纵坐标轴建立二维向量体系, 并将所有变量投影到该向量空间中, 以直观体现个变量之间的相互关系, 其相关程度以变量之间的相对空间位置表征出来。具体计算过程为: 进入 Statistica 软件后, 调入相关数据文件。单击 statistics—multivariate exploratory—Principal components & classification analysis, 进行计算、分析。

图 1 中, 叶绿素与影响因子的空间距离均较远, 没有构建起变量关系群, 表明其相互之间的关系并不密切。这是由于 2008 年奥运森林公园主湖区是为迎接第 29 届奥林匹克运动会, 于 2008 年首次投入使用。虽然采用市政再生水作为水源, 但由于水体中缺少优势藻种、营养盐浓度较低、公园运行维护措施

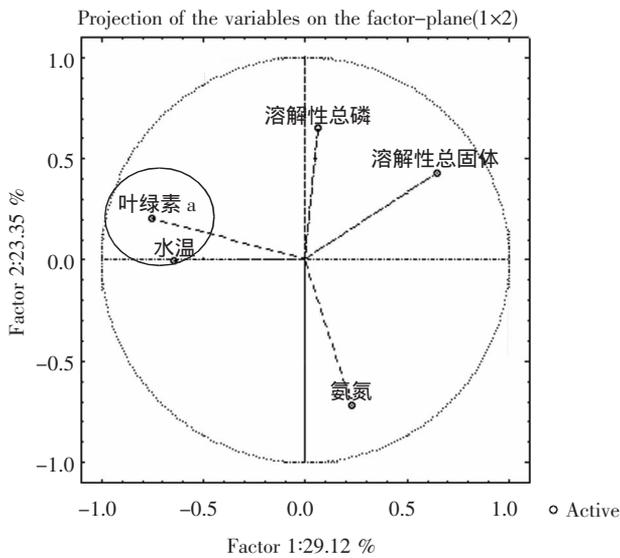


图2 2009年影响因子与叶绿素的PCA分析
Fig.2 PCA Analysis of Influencing Factors and Chlorophyll in 2009

等原因,藻类在2008年没有大量繁殖,叶绿素的平均值只有 37.5 mg/m^3 (一般认为,对于北方浅水景观形水体,叶绿素小于 100 mg/m^3 即可达到景观目的)。藻计数、优势藻等指标和肉眼观察的结果都与上述结论一致。

图2为2009年影响因子与叶绿素的PCA分析。

与图1不同,图2中水温与叶绿素的空间距离较近。且影响因子与叶绿素的空间距离都较为接近,表现出一定程度的正相关性。虽然影响因子的数值除温度外较上一年有所增加,但仍然保持在较低的水平,远远低于城市污水再生利用景观用水水质标准(GB/T 18921—2002)中的要求。此时藻类和水温构成了关系比较密切的变量关系群。

分析产生2008年、2009年不同结果的原因,可以发现除了已知的藻类生长影响因素(营养盐、温度等)之外,景观湖中底泥的累积效应也是影响因素。

为确定图1、图2的五个对系统产生影响的变量组合能否用变量组合1(Factor 1)、变量组合2

表4 2008年影响因子与叶绿素的PCA特征值分析
Tab.4 Eigenvalue Analysis of Influencing Factors and Chlorophyll in 2008

序号	特征值	总方差/%	累积特征值	累积值/%
1	1.600 346	32.006 92	1.600 346	32.006 9
2	1.112 627	22.252 53	2.712 973	54.259 5
3	1.018 704	20.374 08	3.731 677	74.633 5
4	0.879 436	17.588 71	4.611 112	92.222 2
5	0.388 888	7.777 75	5.000 000	100.000 0

表5 2009年影响因子与叶绿素的PCA特征值分析
Tab.5 Eigenvalues Analysis of Influencing Factors and Chlorophyll in 2009

序号	特征值	总方差/%	累积特征值	累积值/%
1	1.456 177	29.123 55	1.456 177	29.123 5
2	1.167 603	23.352 05	2.623 780	52.475 6
3	0.991 025	19.820 51	3.614 805	72.296 1
4	0.759 777	15.195 53	4.374 582	87.491 6
5	0.625 418	12.508 36	5.000 000	100.000 0

(Factor 2)表示,需要验证Factor 1、2的特征值是否显著。在PCA分析模块中,选择特征值分析。

特征值分析结果如表3、表4所示。在5个对系统变化产生影响的变量组合中,变量组合1(Factor 1)和变量组合2(Factor 2)的和,占有系统变化的50%多,特征值显著,可以作为表征系统变化的主成分,由其共同构成的二维空间可以表征系统内各独立变量的相关关系。

3 结论与展望

通过2008年对奥运森林公园主湖区的连续监测和数学分析,得出以下结论:

(1) STATISTICA软件的主成分分析方法可应用于景观水体藻类生长研究,通过相对空间位置直观、简便地得出藻类生长的主要影响因素。

(2)除了已知的藻类生长影响因素(营养盐、温度等)之外,景观湖中底泥的累积效应也应纳入分析之内。

(3) STATISTICA软件是一款强大的分析软件,它不仅能进行定性分析,还可进行定量分析和实验设计。

参考文献

- [1] GAMINI H. Freshwater Algal Blooms and Their Control; Comparison of the European and Australian Experience [J]. Journal of Environmental Management, 1997, 51: 217-227.
- [2] 吕晋, 邹红娟, 马学礼, 等. 武汉市湖泊蓝藻分布影响因素分析[J]. 生态环境, 2008, 17(2): 515-519.
- [3] 戴荣继, 黄春, 佟斌, 等. 藻类叶绿素及其降解产物的测定方法[J]. 中央民族大学学报(自然科学版), 2004, 13(1): 75-80.
- [4] 洪楠. Statistica for Windows 统计与图表分析教程[M]. 北京: 清华大学出版社, 北方交通大学出版社, 2002.
- [5] 刘娟. 污水生物处理系统中原生动物的群落结构及其影响因素分析[D]. 北京: 中国科学院, 2008.
- [6] 金相灿, 刘鸿亮. 中国湖泊富营养化[M]. 北京: 中国环境科学出版社, 1990.