

云南高原湖泊群的统计学聚类识别 及水质响应模式研究

蒋青松¹ 梁中耀¹ 赵磊² 李玉照¹ 吴思枫¹ 刘永^{1,†}

1. 水沙科学教育部重点实验室, 北京大学环境科学与工程学院, 北京 100871; 2. 云南高原湖泊流域污染过程与管理
重点实验室, 昆明 650034; † 通信作者, E-mail: yongliu@pku.edu.cn

摘要 为探究湖泊群水质变量的响应模式, 构建适用于监测数据匮乏的湖泊群聚类 and 响应模式识别方法体系 (PCA-BN), 包括4个步骤: 数据预处理、PCA降维与湖泊聚类、贝叶斯网络构建及参数学习、湖泊响应关系模拟。以云南高原湖泊群为例开展研究, 结果表明: 所研究的26个湖泊可分为两类; 由于第一类湖泊受到的人为干扰更严重, 因而叶绿素 a 对总氮和总磷的响应比第二类湖泊更敏感; 第一类湖泊表层水温高, 溶解氧趋近饱和, 随叶绿素 a 变化不显著, 第二类湖泊溶解氧随叶绿素 a 升高而显著升高; 两类湖泊的透明度与叶绿素 a 的关系一致。

关键词 湖泊群; 水质; 聚类; 响应模式; 主成分分析; 贝叶斯网络

中图分类号 X11; X505

Integrated PCA-BN Approach for Identifying the Water Quality Response Patterns for Lakes in Yunnan Plateau

JIANG Qingsong¹, LIANG Zhongyao¹, ZHAO Lei², LI Yuzhao¹, WU Sifeng¹, LIU Yong^{1,†}

1. Key Laboratory of Water and Sediment Sciences (MOE), College of Environmental Science and Engineering, Peking University, Beijing 100871; 2. Yunnan Key Laboratory of Pollution Process and Management of Plateau Lake-Watershed, Kunming 650034;

† Corresponding author, E-mail: yongliu@pku.edu.cn

Abstract An integrated approach of principle components analysis (PCA) and Bayesian network (BN) for identifying the response pattern of different clusters were developed to understand sensitive relationships of water quality in lakes of Yunnan Plateau. The model includes four steps: data preconditioning, lakes clustering with PCA, Bayesian network learning and lake water quality response modeling. The results demonstrate that the 26 lakes can be clustered into two groups; the Chl a concentration responds more significantly to Total Nitrogen (TN) and Total Phosphorus (TP) in the first group, mainly resulting from much higher watershed disturbances; the Dissolved Oxygen (DO) in the first group with higher water temperature is close to saturation and have little change with Chl a increasing, while the second group is not; and there is good consistency on the relationship between Transparency (SD) and Chl a in both groups.

Key words plateau lakes; water quality; clustering; response pattern; principle components analysis; Bayesian network

湖泊在社会经济发展及提供生态服务功能方面发挥着不可替代的作用, 但由于流域污染物的过量排放, 湖泊面临水质下降、富营养化及生态退化等

问题^[1-2]。为了有效地进行湖泊污染防治, 必须以响应关系为基础, 实施水质目标管理^[3-4]。由于地质、气候、水环境特征等条件的差异, 不同湖泊的

响应模式可能不同^[5]。湖泊水质响应关系的研究需要长期持续的监测数据,然而,当前我国的湖泊监测主要集中于大湖泊^[6-9],面积小于 10 km² 的小湖泊缺乏长时间监测数据^[10]。为探究监测数据匮乏的湖泊水质变量的响应模式,可借助湖泊群的研究方法,对湖泊群进行聚类,进而研究不同类别湖泊的水质变量响应模式^[11-12]。

湖泊聚类方法主要包括以下两种。1) 水生态功能区分方法^[13-14],即将具有相同特质的淡水生态系统或生物群体及相关的土地单元划为同一生态区,如美国明尼苏达州拥有上万个湖泊,研究人员根据水生态区确定取样规模,用 10% 的典型湖泊取样代表整个区域的湖泊状况^[15]。然而,即便在相同生态分区中,湖泊的水交换周期、水深、水面积、湖泊系数等也不尽相同,需要采用基于湖泊分层和循环周期的更小尺度的聚类方法^[16]。2) 依据营养盐浓度或富营养化状态划分^[17]。

上述聚类方法存在的主要问题是:考虑的因素单一,不能涵盖湖泊在营养浓度、形态学和分层持续时间等方面的差异^[18]。为了将更多因素纳入聚类方法中,Akbar 等^[19]应用主成分分析(PCA)与 K 均值聚类法,将加拿大 18 个湖泊分为 5 个类别。

湖泊水质变量响应关系的研究方法包括统计模型和机理模型。机理模型需要连续的监测数据和相同的边界条件,不适用于湖泊群响应模式的研究。统计学方法包括回归方法、机器学习方法、贝叶斯网络等。水质变量响应模式包括多条递进因果关系链,一条链的“果”可能是下一条链的“因”,回归和机器学习方法仅能同时处理一条因果关系链,而贝叶斯网络(BN)可以同时处理多条因果关系链。此外,由于在分析非线性和不确定性响应关系方面的优势,贝叶斯网络广泛用于水质变量响应关系的识别中^[20-21]。

云南高原湖泊群在自然和人为干扰方面具有独

特性,其流域水资源缺乏,水体交换周期长,富营养化敏感性高^[22],对该区域的研究相对缺乏^[23-24]。本文收集 26 个云南高原湖泊水质调查数据,构建适用于监测数据匮乏的湖泊群聚类和响应模式识别方法体系(PCA-BN),以期识别湖泊群的水质响应模式,分析不同类别湖泊响应模式的异同,进一步探究不同响应模式的机理。

1 研究对象与方法

1.1 研究对象

本文的研究对象是云南重点高原湖泊(滇池、洱海、程海、阳宗海、抚仙湖、泸沽湖,面积>20 km²)和 20 个小湖泊(摆龙湖、碧塔海、草海湿地、茈碧湖、大屯海、海峰湿地、海西海、剑湖、拉市海、纳帕海、普者黑、青海湖、清水海、三角海、属都湖、天池、西湖、月湖、长海、长桥海,面积>1 km²)。小湖泊为 2012 年秋季现场监测的 3 组数据,6 个大湖泊为 2012 年同期常规监测的月数据。为表征湖泊的平均水平,将同一湖泊的监测值取算术平均值。根据数据的可得性,选取水温(WT)、溶解氧(DO)、透明度(SD)、总氮(TN)、总磷(TP)、叶绿素 *a* (Chl *a*)和高锰酸盐指数(COD_{Mn})用于统计分析(表 1)。根据对数正态分布检验结果,SD, TN, TP, Chl *a* 和 COD_{Mn}符合对数正态分布,对其进行对数处理。本文的数据分析基于 R 软件(R version 3.1.3, <http://www.r-project.org>)。

1.2 研究方法

PCA-BN 方法体系包括 4 个步骤: 1) 数据预处理, 2) PCA 降维与湖泊聚类, 3) 贝叶斯网络构建及参数学习, 4) 湖泊响应关系模拟, 如图 1 所示。

1.2.1 PCA

PCA 可以将多个具有相关性的指标转化为主成分^[25]。作为变量降维和筛选的有效方法,PCA 在水环境领域中的主要应用是识别水质变量中的关键

表 1 26 个湖泊水质指标的基本统计量
Table 1 Descriptive statistical analysis of lakes

统计量	WT/°C	DO/(mg · L ⁻¹)	SD/m	TN/(mg · L ⁻¹)	TP/(mg · L ⁻¹)	Chl <i>a</i> /(mg · L ⁻¹)	COD _{Mn} /(mg · L ⁻¹)
平均值	21.39	7.155	2.569	0.544	0.039	0.008	4.626
标准差	3.61	2.020	2.726	0.532	0.042	0.009	2.668
变异系数	0.17	0.282	1.061	0.978	1.070	1.147	0.577
最小值	14.57	2.033	0.367	0.050	0.005	0.0004	0.900
最大值	28.67	13.400	11.933	1.998	0.167	0.035	11.600

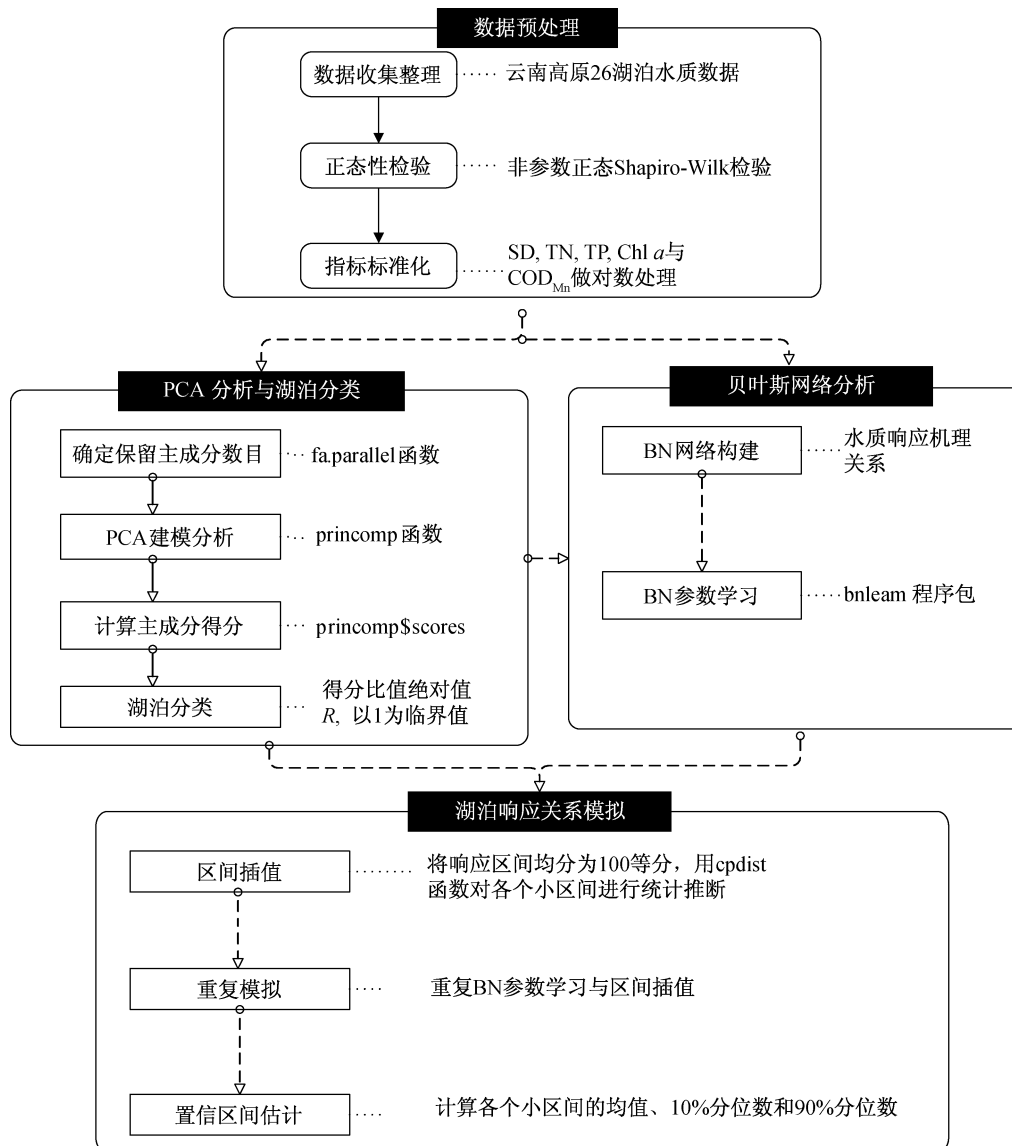


图 1 研究方法体系

Fig. 1 Methodological framework

因素^[26-27]。设样本空间为 $X_{n \times p}$, 其中 n 为样本个数, p 为指标个数。矩阵 $Z = (Z_1, Z_2, \dots, Z_p)$ 满足

$$\text{Cov}(Z_i, Z_j) = 0, i \neq j, \quad (1)$$

$$Z = Q^T X = \sum_{j=1}^p q_j X_j, \text{ 使得 } \sum_{j=1}^p q_j^2 = 1, \quad (2)$$

根据最大化映射的方差, 确定权重向量 Q :

$$\max_{\{q: \|q\|=1\}} \text{Var}(Q^T X) = \max_{\{q: \|q\|=1\}} Q^T \text{Var}(X) Q, \quad (3)$$

Q 的方向由协方差矩阵 $\text{Var}(X)$ 最大特征值 λ_i 的特征向量 γ_i 决定, 变换后矩阵 Z 的协方差 $\text{Var}(Z) = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_p\}$ 满足

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p. \quad (4)$$

PCA 将 p 个原始变量的总方差分解成不相关变量的方差之和, 其中第 i 个主成分 Z_i 的方差 λ_i 占总方差的比例 $\lambda_i / \sum_{j=1}^p \lambda_j$ 称为主成分 Z_i 的贡献率。一般地, 保留的主成分的特征根大于 1^[28]。

1.2.2 BN

BN 表示变量间概率依赖关系的有向无环图, 能够分析变量间的不确定性和非线性关系^[29], 可表示为 $B = \langle N, A, \Theta \rangle$, 其中 N 表示节点或变量全集, A 表示变量间概率依赖关系全集, Θ 表示条件概率分布表(CPT)的参数。这里每个节点 $n \in N$ 表示变

量, 每条边 $a \in A$ 表示变量间概率依赖关系, 同时每个节点都对应 1 个条件概率分布表, 指明该变量与父节点之间概率依赖的数量关系^[30-31]。采用 BN 分析响应关系的步骤是结构学习、参数学习和统计推理, 本文根据水质指标的机理关系构建贝叶斯网络(图 2)。其中, 代表营养盐水平的 TN 和 TP 对代表浮游植物生物量的 Chl a 有显著影响^[32]; Chl a 与 COD_{Mn} 对物理性指标 DO 和 SD 有显著影响^[33]。图 2 中共有 6 个节点和 6 条关系箭头, 分为两种层次关系。

BN 的关键步骤是参数学习, 以最大似然法为例, 观测数据矩阵为

$$\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_m), \mathbf{d}_i = (d_{i1}, d_{i2}, \dots, d_{in}), i = 1, 2, \dots, m,$$

其中, m 为样本数, n 为变量数。在满足样本独立同分布的假设下,

$$P(\mathbf{D}|\theta) = \prod_{i=1}^m P(\mathbf{d}_i|\theta), \quad (5)$$

其似然函数为

$$\ell(\theta|\mathbf{D}) = \log L(\theta|\mathbf{D}) = \log \prod_{i=1}^m P(\mathbf{d}_i|\theta). \quad (6)$$

通过求解式(6)的最大值点, 得到参数 θ^* :

$$\theta^* = \arg \max_{\theta} \ell(\theta|\mathbf{D}). \quad (7)$$

在此基础上, 运用 BN 统计推理方法, 通过联合概率分布公式, 在给定的网络结构和已知数据下, 计算某一事件发生的概率。

1.2.3 响应关系分析

在利用贝叶斯网络识别出水质变量间的非线性和不确定性关系基础上, 对响应关系进行区间模拟。1) 自变量插值模拟。将自变量数值区间均分为 100 份, 在每一个均分的小区间内进行响应变量的估值, 然后取均值代表这个小区间内响应变量的平均水平^[34]; 2) 重复 BN 模拟和区间插值, 计算均

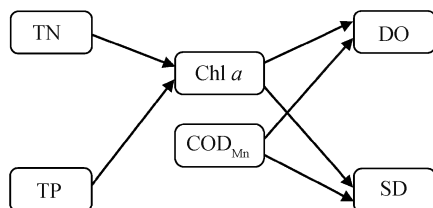


图 2 贝叶斯网络结构
Fig. 2 Structure of Bayesian networks

值、10%分位数和 90%分位数。

2 模型结果

2.1 PCA 结果与湖泊聚类

调用 psych 程序包中的 fa.parallel 函数, 对数据进行平行分析, 确定主成分数目。结果表明, 前两个主成分特征根大于 1, 所以主成分的数目确定为 2; 第一主成分与 TN, TP, Chl a 和 COD_{Mn} 负相关, 与 SD 正相关; 第二主成分与 DO 正相关。前两个主成分的方差贡献率分别为 58.7% 和 18.1%, 累积方差贡献率达到 76.8%, 能有效地解释原始数据的总方差(表 2)。第一主成分表征受人为干扰较强的变量, 第二主成分表征湖体的自然状态。

根据载荷矩阵, 可以计算每个湖泊的主成分得分(即变换后的 Z_i 值)。图 3 为主成分因子载荷图, 箭头长短代表水质指标与相应主成分的相关

表 2 主成分因子载荷矩阵与贡献率
Table 2 The loading matrix and proportion of principal components

主成分载荷表	主成分 1	主成分 2
DO		0.915
SD	0.438	
TN	-0.438	
TP	-0.454	-0.237
Chl a	-0.431	0.225
COD _{Mn}	-0.464	-0.207
方差贡献率	0.587	0.181
累积方差贡献率	0.587	0.768

说明: 空白处表示相关性太小, 可忽略。

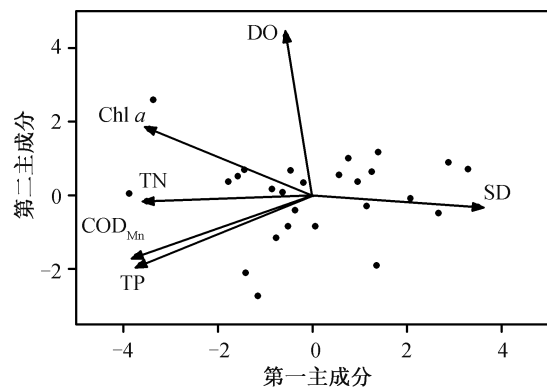


图 3 主成分因子载荷图
Fig. 3 Biplot graphical display of principal component analysis

表 3 云南高原湖泊群聚类结果
Table 3 Result of clustering of 26 lakes in Yunnan Province

类别	<i>R</i>	特征	湖泊聚类结果
一	>1	人为干扰影响较大	滇池、阳宗海、洱海、抚仙湖、泸沽湖、摆龙湖、三角海、大屯海、海西海、普者黑、清水海、茈碧湖、月湖、草海湿地、长桥海
二	<1	人为干扰影响较小	碧塔海、海峰湿地、剑湖、拉市海、纳帕海、青海湖、蜀都湖、天池、西湖、长湖

性强弱。根据得分,湖泊群聚类标准设定为:计算湖泊第一主成分得分与第二主成分得分之比的绝对值 R ,如果 R 大于 1,则判定为第一类;如果 R 小于 1,则判定为第二类(表 3)。第一类湖泊主要受人为干扰影响,而第二类湖泊受人为干扰影响较小。

Astel 等^[28]在对斯特鲁马河的水质分析中,比较了 PCA, CA 和 SOM 等 3 种聚类算法,发现基于主成分得分的聚类不适合大样本(1104 组样本)数据集。本研究以主成分得分之比的绝对值 R 来反映分析结果,这种聚类方法建立在 PCA 的几何意义之上。在二维坐标系中,若数据的分布形式为任意的椭圆分布,经过 PCA 线性变换(即将坐标系进行正交旋转到另一方向),使其椭圆长轴和短轴为坐标轴,变换后的两坐标轴不相关,且数据点的方差大部分被横轴解释,而短轴上解释的方差较小,这样新的坐标系就能保留大部分原坐标的信息,这在多维平面上体现得更为突出^[25]。本文的 PCA 将六维平面降低到二维平面,第一主成分是原数据方差变化最大的方向,在两个主成分形成的二维坐标系(图 3),数据越靠近第一主成分,该点与第一主成分的关联就越比第二主成分大,反之亦然。由此将湖泊分为两类,同类湖泊中水质指标的绝对大小并不代表响应关系的敏感性,而基于 R 值确定的类别能够表征某一类湖泊的主要影响因子。

2.2 贝叶斯网络(BN)响应关系拟合

根据 BN 构建的响应关系概念图进行模拟,结合置信区间得到非线性的响应关系(图 4)。由于每类湖泊数量较少,会造成 BN 学习参数的误差,所以通过 200 次 BN 模拟得到响应区间。在 Chl a 浓度与 TN 的响应关系中,两类湖泊具有相同的变化趋势,但第一类湖泊的 Chl a 浓度变化上限是 0.010~0.015 mg/L,而第二类湖泊的 Chl a 浓度变化上限低于 0.005 mg/L;在相同的 TN 区间值下,第一类湖泊中 Chl a 随 TN 的增加而增加,对 TN 的响应敏感,而第二类湖泊无显著变化(图 4(a))。在 Chl a 浓度与 TP 的响应关系中,第一类湖泊的总磷变化

上限是 0.18 mg/L, Chl a 随 TP 升高而增加;第二类湖泊的 TP 变化上限是 0.12 mg/L, Chl a 浓度随 TP 升高而无明显变化。因此,第一类湖泊中 Chl a 浓度对 TP 的响应关系更敏感(图 4(b))。在 DO 浓度与 Chl a 的响应关系中,第一类湖泊的 Chl a 变化范围为 0~0.035 mg/L, DO 浓度随着 Chl a 的增加而先增加后趋于平缓;第二类湖泊 Chl a 的变化范围为 0~0.012 mg/L, DO 随 Chl a 的增加而先增加后趋于平缓(图 4(c))。在 SD 浓度与 Chl a 的响应关系中,两类湖泊的 SD 均随 Chl a 浓度的增加而减小,逐渐趋近平缓(图 4(d))。

从同类湖泊的对比来看,第一类湖泊 TN 引起 Chl a 的变化范围为 0~0.015 mg/L,而 TP 引起 Chl a 的变化范围为 0~0.03 mg/L(图 4(a)(b)),表明第一类湖泊 Chl a 对 TP 的变化更敏感。第二类湖泊 TN 引起 Chl a 有微弱上升的趋势,而 TP 没有引起 Chl a 增加,因此第二类湖泊对 TN 更敏感,但不明显。两类湖泊的 DO 随 Chl a 浓度的增加而增加,SD 随 Chl a 浓度的增加而下降。

3 讨论

本文所研究的湖泊以贫营养与中营养为主,滇池、大屯海和三角海为富营养湖泊。根据综合富营养状态指数(TLI)计算公式^[35]得到,第一类湖泊平均 TLI 为 36.6,与第二类湖泊平均 TLI (38.4)无显著差别(F 检验的 p 值为 0.723),但第一类湖泊营养化水平的内部差异显著大于第二类湖泊(F 检验, $p<0.05$)。引起湖泊富营养化的主要因素是气候(包括太阳辐射、季风、气温等)、水温、湖泊形态和营养盐含量等。水温是重要的影响因素,直接影响湖泊的生化反应^[36]。两类湖泊的平均海拔分别是 1830 m 和 2450 m,按垂直气温梯度 6°C/km 估计,大气温差约为 3.7°C,实际上两类湖泊的平均水温分别为 22.7°C 和 19.6°C,相差 3.1°C,经过 F 检验,发现两类湖泊的平均温度显著不同(p 值为 0.031)。在适当范围内,温度的升高能提高浮游植物的生长

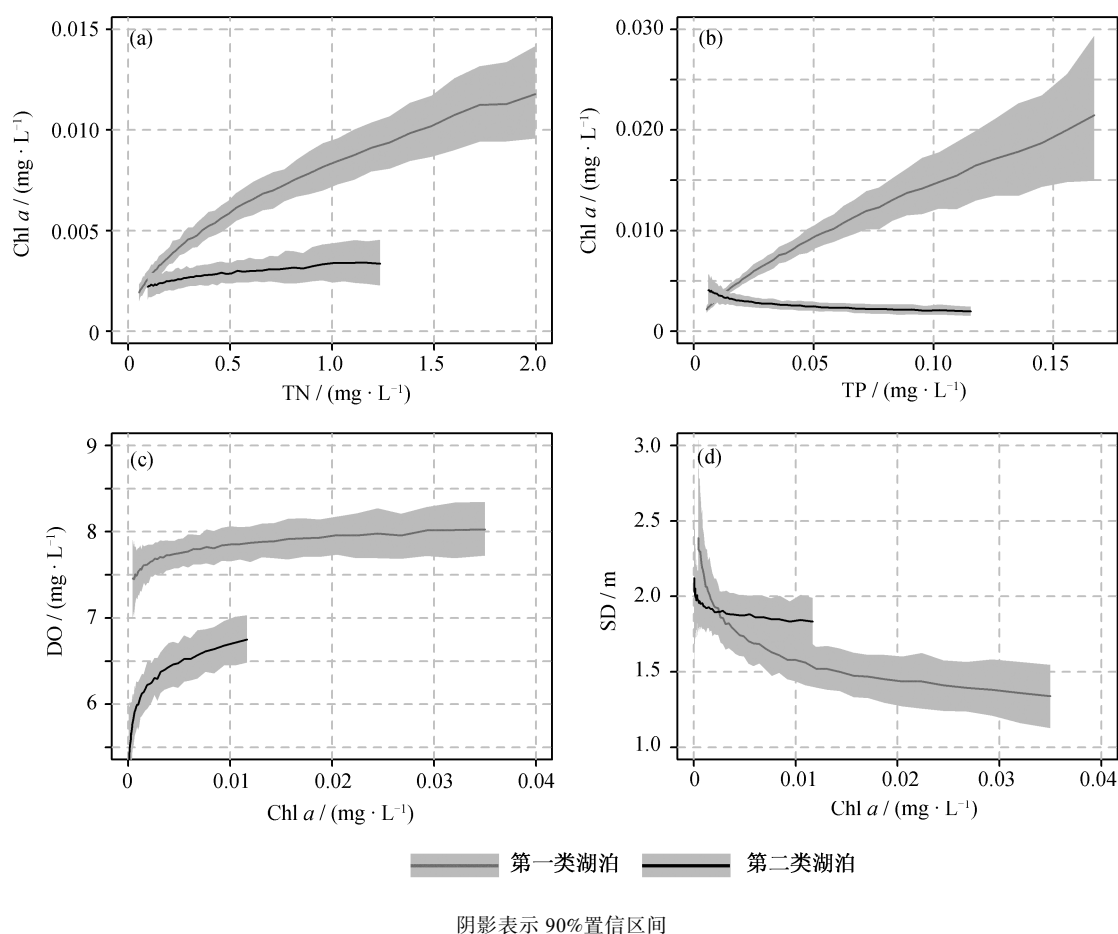


图 4 两类湖泊的响应关系拟合

Fig. 4 Comparison of water quality responses for lakes in two groups

速率,并加快微生物新陈代谢^[18,37],这很可能是第一类湖泊 Chl *a* 变化范围较大的原因。

研究表明,湖泊水深对湖泊分层有促进作用^[38],并且湖泊越小,分层所要求的水深越低^[39]。第一类湖泊平均水深为 20.8 m (包含深达 100 m 的抚仙湖),第二类湖泊的平均水深为 14.0 m。除去第一类湖泊中异常深的抚仙湖,两类湖泊的水深无显著差别(F 检验的 p 值为 0.912),包含滇池、洱海等大湖的第一类湖泊平均湖泊面积(61 km²)高于第二类湖泊的平均湖泊面积(10.5 km²),但在 95%置信度下不显著,其 F 检验的 p 值为 0.130。以上数据表明,除异常值外,两类湖泊形态的差别并不显著,因而对于水质响应将更多地考虑流域内的人为干扰。

一般地,随着海拔的增加,人口密度和人为干扰逐渐减小^[40],湖泊富营养化水平降低^[20];湖泊-流域是一个相互反馈的动态变化系统^[41],流域耕地

与湖泊面积的比值越大,表明单位水体受社会经济的影响越大^[42-43]。本研究中,第一类湖泊的平均海拔显著低于第二类湖泊,同时第一类湖泊的流域耕地-湖泊面积比值高于第二类湖泊(图 5),因此第一类湖泊受到的人为干扰更大,致使其生态系统物质平衡出现扰动,从而出现 Chl *a* 随着 TN 和 TP 的增加而上升更快的响应关系(图 4)^[44]。

在 DO 与 Chl *a* 的响应关系中,两类湖泊都呈现 DO 随 Chl *a* 的增加而增加的趋势,与关于云南湖泊群的已有研究相似^[44-45]。本研究中第二类湖泊的 DO 变化更敏感,水温是造成该现象的主要原因。第一类湖泊平均水温比第二类湖泊高 3.1℃,其饱和 DO 浓度分别是 8.5 和 9.1 mg/L。从响应曲线可以看出,第一类湖泊的溶解氧已达到 8 mg/L,趋于饱和,浮游植物产生的更多的溶解氧不能全部保留在表水层或扩散至深水层,而会逸出水体,进入大气。因此,第一类湖泊呈现 DO 缓慢上升后不

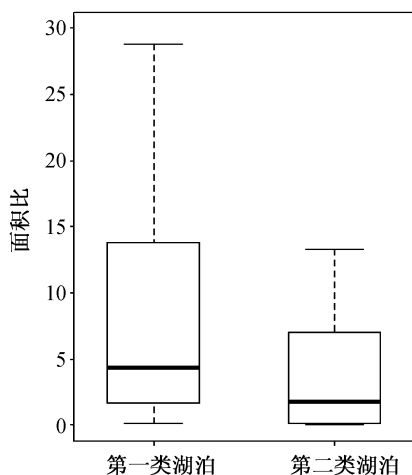


图 5 两类湖泊的流域耕地面积-湖泊面积比

Fig. 5 The watershed cropland: lake area ratio of two clusters

变的趋势。而第二类湖泊离饱和 DO 值差距大, 浮游植物产生的溶解氧能更多地积累起来, 使得第二类湖泊中的 DO 上升更快速。

在 SD 与 Chl *a* 的响应关系中, 两类湖泊均随着 Chl *a* 的增加而降低, 下降趋势一致。此结果与 Liu 等^[46-47]和李蒙等^[45]云南湖泊 SD 与 Chl *a* 呈负相关关系的研究结果一致。湖水的透明度因湖水中悬浮物、浮游藻类对光的吸收和散射而降低。据本研究的监测数据, 两类湖泊中的悬浮物浓度无显著差别(*p* 值为 0.793), 其浓度值均很低(仅大屯海的悬浮物高于 15 mg/L), 表明悬浮物对透明度的影响很小。另一方面, 当浮游植物数量增加时, 通过散射直接降低透明度, 也能吸收某些波长的光间接影响透明度, 所以云南湖泊群中透明度随着叶绿素的增加而降低。

4 结论

本文以 2012 年云南高原湖泊监测数据为基础, 采用 PCA 方法, 将湖泊群分为两类, 分别构建 BN 模型, 分析同类群湖泊的响应关系, 主要结论如下。

1) 原始数据经 PCA 分析发现, 第一主成分与 SD, TN, TP, Chl *a* 和 COD_{Mn} 具有强相关性, 第二主成分与 DO 有强相关性, 前两成分的累积贡献率达到 76.8%, 能够保留原数据主要的信息。基于对湖泊两个主成分得分比值的绝对值与临界值 $R=1$ 的比较, 将湖泊分为两类。

2) BN 的响应关系模拟结果表明, 第一类湖泊

中 Chl *a* 对 TN 和 TP 的响应比第二类更敏感, 由于第一类湖泊海拔显著低于第二类湖泊, 但流域-湖泊面积比值无显著差别, 因此人为干扰对第一类湖泊的影响更剧烈。在 DO 与 Chl *a* 的响应关系中, 第一类湖泊由于饱和溶解氧低致使 DO 随叶绿素 *a* 上升缓慢, 而第二类湖泊饱和溶解氧高, 浮游植物产生的氧气更容易以 DO 形式存在于水体中。在 SD 与 Chl *a* 的响应关系中, 两类湖泊的悬浮物质含量均很低, 透明度随着浮游植物浓度的升高而降低, 趋势一致, 程度相近。

湖泊水质响应是一个复杂的过程, 其中多种因素相互影响, 本文构建的 PCA-BN 方法可对缺乏长时间序列数据的湖泊群进行分析。但是, 对于具有长时间序列的单个湖泊, 直接建立模型分析水质变量之间的响应关系则更为直接。

参考文献

- [1] 孟伟, 苏一兵, 郑丙辉. 中国流域水污染现状与控制策略的探讨. 中国水利水电科学研究院学报, 2004, 2(4): 242-246
- [2] Conley D J, Paerl H W, Howarth R W, et al. Controlling eutrophication: nitrogen and phosphorus. Science, 2009, 323: 1014-1015
- [3] Baresel C, Destouni G. Uncertainty-accounting environmental policy and management of water systems. Environmental Science & Technology, 2007, 41(10): 3653-3659
- [4] 刘永, 邹锐, 郭怀成, 等. 智能流域管理研究. 北京: 科学出版社, 2012
- [5] Swaney D P, Scavia D, Howarth R W, et al. Estuarine classification and response to nitrogen loading: insights from simple ecological models. Estuarine, Coastal and Shelf Science, 2008, 77(2): 253-263
- [6] Trebitz A S, Brazner J C, Cotter A M, et al. Water quality in Great Lakes coastal wetlands: basin-wide patterns and responses to an anthropogenic disturbance gradient. Journal of Great Lakes Research, 2007, 33(1): 67-85
- [7] 吴丰昌, 孟伟, 宋永会, 等. 中国湖泊水环境基准的研究进展. 环境科学学报, 2008, 28(12): 2385-2393
- [8] Liu Y, Wang Y L, Sheng H, et al. Quantitative evaluation of lake eutrophication responses under alternative water diversion scenarios: a water quality modeling based statistical analysis approach. Science

- of the Total Environment, 2013, 468(7): 219–227
- [9] Zou R, Zhang X L, Liu Y, et al. Uncertainty-based analysis on water quality response to water diversions for Lake Chenghai: a multiple-pattern inverse modeling approach. *Journal of Hydrology*, 2014, 514: 1–14
- [10] 张军莉, 赵磊, 聂菊芬. 云南高原小湖泊水质空间分布特征研究. *环境科学导刊*, 2015(2): 26–34
- [11] Omernik J M. Map Supplement: ecoregions of the Conterminous United States. *Annals of the Association of American Geographers*, 1987, 77(1): 118–125
- [12] Freeman A M, Lamon III E C, Stow C A. Nutrient criteria for lakes, ponds, and reservoirs: a Bayesian TREED model approach. *Ecological Modelling*, 2009, 220(5): 630–639
- [13] Lamon III E C, Malve O, Pietiläinen O P. Lake classification to enhance prediction of eutrophication endpoints in Finnish lakes. *Environmental Modelling & Software*, 2008, 23(7): 938–947
- [14] Hemsley F B, Wright J F, Sutcliffe D W, et al. Classification of the biological quality of rivers in England and Wales // Assessing the biological quality of fresh waters: RIVPACS and other techniques. *Proceedings of an International Workshop held in Oxford, UK, on 16–18 September 1997*. *Freshwater Biological Association (FBA)*, 2000: 55–69
- [15] Heiskary S A, Wilson C B, Larsen D P. Analysis of regional patterns in lake water quality: using ecoregions for lake management in Minnesota. *Lake & Reservoir Management*, 1987, 3(1): 337–344
- [16] Lewis W M. A revised classification of lakes based on mixing. *Canadian Journal of Fisheries & Aquatic Sciences*, 2011, 40(10): 1779–1787
- [17] Naumann E. Några synpunkter angående limnoplanktons ökologi med särskild hänsyn till fytoplankton. *Svensk Botanisk Tidskrift*, 1919, 13(6): 129–163
- [18] Kalff J. *Limnology: inland water ecosystems*. New Jersey: Prentice Hall, 2002
- [19] Akbar T A, Hassan Q K, Achari G. A methodology for clustering lakes in Alberta on the basis of water quality parameters. *Clean-Soilair Water*, 2011, 39(10): 916–924
- [20] Abdi H, Williams L J. *Principal component analysis*. Wiley Interdisciplinary Reviews: Computational Statistics, 2010, 2(4): 433–459
- [21] Ouyang Y, Nkedi-Kizza P, Wu Q T, et al. Assessment of seasonal variations in surface water quality. *Water Research*, 2006, 40(20): 3800–3810
- [22] 曹金玲, 席北斗, 许其功, 等. 地理气候及湖盆形态对我国湖泊营养状态的影响. *环境科学学报*, 2012, 32(6): 1512–1519
- [23] 刘总堂, 李春海, 章钢娅. 运用主成分分析法研究云南湖库水体中重金属分布. *环境科学研究*, 2010, 23(4): 459–466
- [24] Yang Y H, Zhou F, Guo H C, et al. Analysis of spatial and temporal water pollution patterns in Lake Dianchi using multivariate statistical methods. *Environmental Monitoring & Assessment*, 2010, 170: 407–416
- [25] Olsen R L, Chappell R W, Loftis J C. Water quality sample collection, data treatment and results presentation for principal components analysis — literature review and Illinois River watershed case study. *Water Research*, 2012, 46(9): 3110–3122
- [26] Rigosi A, Hanson P, Hamilton D P, et al. Determining the probability of cyanobacterial blooms: the application of Bayesian networks in multiple lake systems. *Ecological Applications*, 2015, 25(1): 186–199
- [27] Loiselle S A, Azza N, Cózar A, et al. Variability in factors causing light attenuation in Lake Victoria. *Freshwater Biology*, 2008, 53(3): 535–545
- [28] Astel A, Tsakovski S, Barbieri P, et al. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. *Water Research*, 2007, 41(19): 4566–4578
- [29] Tsamardinos I, Brown L E, Aliferis C F. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, 2006, 65(1): 31–78
- [30] Nielsen T D, Jensen F V. *Bayesian networks and decision graphs*. New York: Springer Science & Business Media, 2007
- [31] Hart B T, Pollino C A, Hart B T, et al. *Bayesian modelling for risk-based environmental water allocation*. National Water Commission, 2009
- [32] Xu H, Mccarthy M J, Zhu G, et al. Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): the need for a dual nutrient (N & P) management strategy. *Water Research*, 2011, 45(5): 1973–1983
- [33] Lee H S, Lee J H W. Continuous monitoring of short term dissolved oxygen and algal dynamics. *Water Research*, 1995, 29(29): 2789–2796
- [34] Gabriels W, Goethals P L M, Dedecker A P, et al.

- analysis of macrobenthic communities in Flanders, Belgium, using a stepwise input variable selection procedure with artificial neural networks. *Aquatic Ecology*, 2007, 41(3): 427–441
- [35] 王明翠, 刘雪芹, 张建辉. 湖泊富营养化评价方法及分级标准. *中国环境监测*, 2002, 18(5): 47–49
- [36] Ji Z G. *Hydrodynamics and water quality: modeling rivers, lakes, and estuaries*. Wiley-Interscience, 2008
- [37] McQueen D J, Lean D R S. Influence of water temperature and nitrogen to phosphorus ratios on the dominance of blue-green algae in Lake St. George, Ontario. *Canadian Journal of Fisheries and aquatic Sciences*, 1987, 44(3): 598–604
- [38] Kling G W. Comparative transparency, depth of mixing, and stability of stratification in lakes of Cameroon, Westafrica. *Limnology & Oceanography*, 1988, 33(1): 27–40
- [39] Gorham E, Boyce F M. Influence of lake surface area and depth upon thermal stratification and the Depth of the summer thermocline. *Journal of Great Lakes Research*, 1989, 15(2): 233–245
- [40] 孙玉莲, 赵永涛, 曹伟超, 等. 山区人口分布与环境要素关系的定量分析. *安徽农业科学*, 2011, 39(19): 11705–11707
- [41] 宋长青, 杨桂山, 冷疏影. 湖泊及流域科学研究进展与展望. *湖泊科学*, 2002, 14(4): 3–14
- [42] Liu Wenzhi, Zhang Quanfa, Liu Guihua. Effects of watershed land use and lake morphometry on the trophic state of Chinese lakes: implications for eutrophication control. *Clean — Soil, Air, Water*, 2011, 39(1): 35–42
- [43] 秦伯强, 高光, 朱广伟, 等. 湖泊富营养化及其生态系统响应. *科学通报*, 2013(10): 855–864
- [44] Zang C, Huang S, Wu M, et al. Comparison of relationships between pH, dissolved oxygen and chlorophyll a for aquaculture and non-aquaculture waters. *Water, Air, & Soil Pollution*, 2011, 219: 157–174
- [45] 李蒙, 谢国清, 戴丛蕊, 等. 滇池外海水体叶绿素 *a* 与水质因子关系研究. *云南地理环境研究*, 2009, 21(2): 102–106
- [46] Liu W, Li S, Bu H, et al. Eutrophication in the Yunnan Plateau lakes: the influence of lake morphology, watershed land use, and socioeconomic factors. *Environmental Science and Pollution Research*, 2012, 19(3): 858–870
- [47] Liu W, Zhang Q, Liu G. Lake eutrophication associated with geographic location, lake morphology and climate in China. *Hydrobiologia*, 2010, 644(1): 289–299