

79-85

湖泊营养类型的 FUZZY 聚类分析

陈昌齐 谢红珍

(西南农业大学, 重庆 630716)

10343.3

摘要 以总硬度、pH、溶氧、化学耗氧、总氮、总磷、浮游植物、透明度 8 项指标为依据, 参照现行湖泊营养类型划分标准, 应用 fuzzy 聚类分析方法对我国 42 个主要湖泊聚类并划分其营养类型, 同时对各类型的指标特征进行了分析。结果表明: 在相似水平 $\lambda=0.69$ 下, 42 个湖泊共聚为 13 类。其中, 中、富营养型有较强的相似性, 其差异显著的指标为总氮、总磷和化学耗氧量; 贫营养型湖泊相似性差, 其共同特征是指标中一项或数项显著异于一般营养范围。聚类结果与传统分类基本吻合。

关键词 fuzzy 聚类分析 湖泊营养类型

根据水域营养状况将其划分为相应的营养型, 是研究水域生态环境演变、评价水域渔产性能的一种基本方法。就目前湖泊渔业的现状而言, 生产者在大面积养殖工作中的主观能动性, 主要表现在根据水体的自然条件对水体加以最有效的利用。同时, 网箱养鱼等集约化养殖过程中人工投饲、施肥等活动, 又可能影响水域局部乃至整体的生态环境。因而, 作为渔业工作者, 正确地识别湖泊的营养类型, 掌握其演变规律, 更好地为渔业生产决策以及保护湖泊生态环境提供依据。

湖泊营养状况决定于湖泊所在地的气候条件、湖泊本身的形态特点、湖水的理化特性以及生物群落的结构、数量等因素。长期以来, 许多研究者综合上述诸因素或侧重于其中某些因素, 提出了相应的营养型划分标准。这些标准尽管总趋势上保持一致, 但其中某些单项指标尚存在明显差异。湖泊营养型划分方法目前大体分为两种: 一是以单方面指标与相应标准值比较, 直接得其类型归属; 另一是取多方面指标, 以相应标准为根据划分评分等级, 依其综合得分高低判别其类型。以单方面指标判别营养型, 由于仅在诸多相互联系的因素中选取个别加以比较, 其片面性较难避免, 同时, 数值的自然波动(如浮游植物数量的昼夜、季节、区域变化)以及测定误差对判别的干扰严重。此外, 同一水域可能某一指标值属一营养型, 而另一指标值又属另一营养型, 致使判别难以进行。取多方面指标综合评定类型, 其中单项指标的权重(即单项因素对营养状况的影响程度)确定以及评分等级的划分等, 目前尚缺少充分的依据。因此, 湖泊营养类型划分就其指标选择、标准确定以及评定手段, 目前尚未形成统一、规范的方法。本文尝试应用 fuzzy 聚类分析手段, 通过不同湖泊间指标值的比较, 使其自然成类, 旨在探索湖泊营养型划分的某些内在规律。

一、材料与方 法

本文以我国 42 个主要湖泊近期调查结果^[1]为基本资料, 选取数据较完整的 8 个因素作为分析指标(表一)。营养型划分选目前国内较常用的何志辉、王洪道以及世界经济合作与开发组织综合或提出的标准^[1-3], 与本文有关的内容综合列于表 2 中。

表 1 湖泊调查资料统计

Tab. 1 Statistics of the lake parameters

序号	湖泊名称	湖泊所在地	pH值	溶解氧 (mg/L)	化学耗 氧量 (mg/L)	总 氮 (mg/L)	总 磷 (mg/L)	浮 游 植物数 (10 ⁴ 个/L)	总硬度 (德国度)	透明度 (m)
1	洞庭湖	湖南	8.2	9.8	2.66	0.86	0.02	8.7		0.40
2	鄱阳湖	江西	7.2	9.0	1.59	0.42	0.02		1.5	1.20
3	太湖	江苏	8.0	9.5	3.40	0.90	0.02	100.0	4.2	0.50
4	洪泽湖	江苏	7.8	7.0	3.10	0.46	0.10	11.5	1.9	0.30
5	呼伦湖	内蒙古	7.8	8.9	8.29	0.13	0.08	11.6	11.6	0.50
6	黄旗海	内蒙古	9.3	8.1	10.70	0.56	0.04		29.0	
7	乌梁素海	内蒙古	9.1	8.4	3.30	3.32	0.25		48.6	0.45
8	岱 海	内蒙古	8.6	7.8	5.40	1.43		11.2	22.7	7.50
9	洪 湖	湖北	8.8	8.3	4.70	0.23			7.1	1.00
10	果子湖	湖北	7.4	9.0	3.10	2.00			5.1	
11	东 湖	湖北	8.1	8.1	4.25	1.27	0.16	1913.7	7.8	0.40
12	大泊湖	湖北	8.2	8.4	3.60	1.75	0.10		6.7	
13	长 湖	湖北	8.0	6.4	5.30		0.12		5.1	
14	长春南湖	吉林	9.1	10.0	15.00	3.00	0.27	5098.0	9.1	
15	杭州西湖	浙江	8.5	9.4	6.90	2.20	0.10	6920.0	4.2	0.35
16	巢 湖	安徽	7.4	9.2	4.00	1.31	0.12	25.3		0.25
17	镜泊湖	黑龙江	7.6	6.8	9.70	0.19	0.04	40.8		
18	滇 池	云南	8.6	6.8	6.10	1.88	0.11	189.2	7.6	0.50
19	洱 海	云南	7.1		3.40	0.37	0.05		5.6	4.00
20	白洋淀	河北	8.4	7.6	5.40					2.50
21	南四湖	山东	8.4	6.6	4.60	0.60	0.10		12.0	0.30
22	蔚蔚湖	新疆	7.8	9.4	14.20	2.49	0.22		10.5	
23	抚仙湖	云南	8.6	7.2	1.40	0.21	0.02	19.0	8.1	7.00
24	松花湖	吉林	8.0	7.0	8.00	1.60			2.2	
25	淀山湖	上海	8.0	8.6	4.10	1.53	0.06		1.9	1.20
26	扬州瘦西湖	江苏	8.2	8.2	6.70	2.84	0.14			0.90
27	无锡五里湖	江苏	7.8	9.0	4.00	1.56	0.06			0.48
28	南京玄武湖	江苏	8.3	9.0	6.10	2.18	0.17			0.45
29	莫愁湖	江苏	8.2	8.6	6.70	2.98	0.14			0.25
30	嘉兴南湖	浙江	7.5	6.0	4.90	2.08	0.16			
31	阳澄湖	江苏	8.2	6.0	4.00	1.01				
32	新安江水库	浙江	8.0	8.0	1.20					
33	白头山天池	吉林	7.3	6.4	2.90	0.24			2.0	
34	青海湖	青海	9.4		1.40	0.22	0.02	14.6		4.50
35	博斯腾湖	新疆	8.6		4.07	0.40	0.03	2.5		
36	乌伦古湖	新疆	7.8		14.4	0.19	0.05	8.1		
37	高邮湖	江苏	8.5		2.63	0.57	0.03			
38	鄂陵湖	青海	8.5	7.8	3.02	0.42	0.03	3.8	9.5	
39	星方湖	云南			4.20	0.25				
40	杞麓湖	云南			5.56	0.70				
41	新疆天池	新疆	7.9		2.82	0.41	0.01			9.00
42	晋阳湖	山西	8.4		2.60	0.95	0.05			

表 2 湖泊营养类型划分标准(mg/L)
Tab. 2 The standard of the trophic level classification of lakes

项目 类型	何志辉综合的标准				王洪道提出的标准				世经组织提出标准		
	浮植 现存量	耗氧量	总氮	总磷	浮植现存量 (10 ⁴ /L)	耗氧量	总氮	总磷	透明度 (m)	总氮	总磷
贫	<1.5	<1	<0.25	<0.01	5-10	0.12- 0.48	0.02- 0.08	0.0009- 0.005	8-27	0.661 (0.37-1.18)	0.008 (0.005-0.013)
贫中					20	0.96	0.16	0.010	4.4		
中	1-5	1-7	0.25- 1.1	0.01- 0.03	60	1.80	0.31	0.023	2.4	0.753 (0.49-0.17)	0.0267 (0.015-0.048)
中富					100	3.60	0.65	0.050	1.3		
富	5-10	7-15	>1.1	>0.03	200	7.1	1.20	0.110	0.73	1.875 (0.86-4.08)	0.084 (0.048-0.188)
超富	>10	>15			1000	27	4.60	0.560	0.22		

以 x_{ij} 表示第 i 个湖泊第 j 项指标, 则每个湖泊所对应唯一的向量 X_i 可表示为: $X_i = (x_{i1}, x_{i2}, \dots, x_{i8}), i = 1, 2, \dots, 42$ 。运算依聚类分析一般方法^[4]按以下步骤进行。

1. 数据标准化处理 令 $x'_{ij} = (x_{ij} - b_{ij}) / (a_{ij} - b_{ij})$, 其中 $a_{ij} = \max\{x_{ij} | i = 1, 2, \dots, 42\}$, $b_{ij} = \min\{x_{ij} | i = 1, 2, \dots, 42\}$ 。经处理后, 每个湖泊都对应一个新向量 $X'_i = (x'_{i1}, x'_{i2}, \dots, x'_{i8})$ 。为方便计, 不妨仍以 X_i 表示 X'_i , x_{ij} 表示 x'_{ij} 。

2. 统计量标定 令被分类对象间相似程度的统计量为 r_{ij} , 其标定采用最大-最小法。即令:

$$r_{ij} = \frac{\sum_{k=1}^m \min(x_{ik}, x_{jk})}{\sum_{k=1}^m \max(x_{ik}, x_{jk})}, m \leq 8.$$

得相似矩阵

$$R = (r_{ij})_{n \times n} = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \dots & \vdots \\ r_{n1} & r_{n2} & \dots & r_{nn} \end{pmatrix} n = 42$$

3. 聚类 聚类直接由相似矩阵 R 经编网法而完成(参见[4, 6])。

对所聚类型中各指标值按常规统计方法^[5]处理, 并利用 t 检验判别指标在不同类

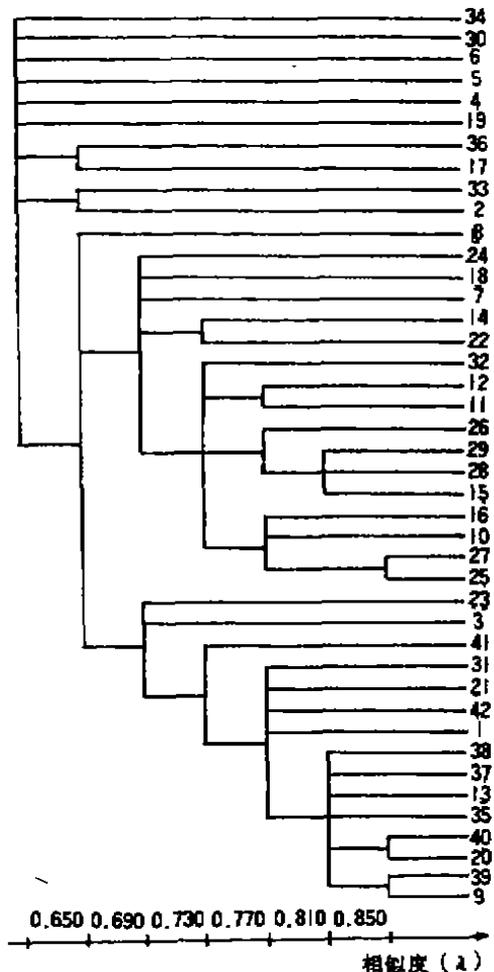


图 1 湖泊营养状况 fuzzy 聚类图
Fig. 1 The fuzzy cluster of trophic classification of the lakes

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

二、结 果

1. 依次取相似度 $\lambda = 0.650, 0.690, 0.730, 0.770, 0.810, 0.850$, 由相应水平的聚类结果绘制出聚类谱系图(图 1)。在 $\lambda = 0.690$ 水平上分析, 42 个湖泊分为 13 类, 第 I 类 {1, 3, 9, 13, 20, 21, 23, 31, 35, 37, 38, 39, 40, 41, 42} 共 15 个湖泊; 第 II 类 {7, 10, 11, 12, 14, 15, 16, 18, 22, 24, 25, 26, 27, 28, 29, 32} 共 16 个湖泊; 余下 11 个湖泊各自成类。

2. 计算第 I、II 类中湖泊各营养指标的均值 \bar{x} 、标准差 s 、变差系数 Cv 及该两类湖泊间各指标 t 值, 将 t 值与相应自由度下不同置信度的值 $t_{0.10}, t_{0.05}, t_{0.01}$ 比较, $t_{0.10} < t \leq t_{0.05}$ 差异较显著, 记为“+”; $t_{0.05} < t \leq t_{0.01}$ 差异显著, 记为“++”; $t_{0.01} < t$ 差异极显著, 记为“+++”。结果列于表 3。

表 3 第 I、II 类湖泊指标特征

Tab. 3 Characteristics of the parameters of mesotrophic and eutrophic lakes

项 目	I 类				II 类				显著性比较					
	n_1	\bar{x}_1	S_1	Cv	n_2	\bar{x}_2	S_2	Cv	N	t	$t_{0.10}$	$t_{0.05}$	$t_{0.01}$	显著性
pH	13	8.346	0.270	0.032	16	8.169	0.490	0.060	27	1.164	1.700	2.050	2.770	
溶氧量(mg/L)	9	7.689	1.326	0.172	16	8.569	0.845	0.099	23	2.032	1.710	2.060	2.800	+
化学耗氧量(mg/L)	15	3.761	1.234	0.328	16	6.078	3.769	0.620	29	2.264	1.700	2.040	2.760	++
总氮(mg/L)	13	0.577	0.285	0.494	15	2.127	0.662	0.311	26	7.820	1.710	2.060	2.780	+++
总磷(mg/L)	10	0.043	0.037	0.860	13	0.146	0.067	0.459	21	4.292	1.720	2.080	2.830	+++
浮游植物(10^4 个/L)	5	26.800	41.431	1.546	5	2817.24	3052.96	1.084	8	1.972	1.860	2.310	3.360	+
总硬度(德国度)	6	7.667	2.872	0.375	10	10.370	13.724	1.323	14	0.475	1.760	2.150	2.980	
透明度(m)	7	2.957	3.572	1.208	10	0.523	0.300	0.574	15	2.174	1.750	2.130	2.950	++

三、讨 论

1. 聚类结果与营养类型

分析湖泊营养指标特征并与表 2 给出的标准比较, 聚类结果与营养类型具有如下关系。

(1) 第 I 类湖泊 8 项指标中较为稳定的有 pH($\bar{x} = 8.346, Cv = 0.032$)、溶氧($\bar{x} = 7.689, Cv = 0.172$)、耗氧($\bar{x} = 3.761, Cv = 0.328$)、总氮($\bar{x} = 0.577, Cv = 0.494$)、总磷($\bar{x} = 0.043, Cv = 0.860$)。总硬度、浮游植物量、透明度离散大, 其最大值与最小值可相差几倍乃至几十倍。按表 2 标准比较耗氧、总氮、总磷三指标均值, 该类湖泊具有明显的中营养型特征。该类湖泊分布于湖北、河北、吉林、江苏、云南等省区, 无明显的地域性。

(2) 第 II 类湖泊 8 项指标中较为稳定的有 pH($\bar{x} = 8.169, Cv = 0.060$)、溶氧($\bar{x} = 8.569, Cv = 0.099$)、耗氧($\bar{x} = 6.078, Cv = 0.625$)、总氮($\bar{x} = 2.127, Cv = 0.311$)、总磷($\bar{x} = 0.146, Cv = 0.459$)。总硬度、浮游植物数量离散大。与第 I 类比较, 耗氧、总氮、总磷均

有显著增加, 透明度趋于稳定且明显降低, 显示出营养状况由中变富的明显趋势。按表 2 所给标准(耗氧、总氮、总磷和透明度四指标均值), 该类湖泊具明显的富营养型特征, 主要分布在长江中、下游地区, 具有较强的地域性。

(3) 序号为 2、6、8、17、19、33、34、36 的 8 个湖泊, 其彼此间各指标值犬牙交错, 相似性较差, 在相似度 0.690 下难以聚为一类。比较表 2 标准, 该部份湖泊总氮、总磷普遍较低, 透明度相对较高; 几乎每个都有一项或多项指标值明显异于一般正常营养范围, 成为其营养状况的限制因素及贫营养型湖泊划分的主要标志。如白头山天池、鄱阳湖总硬度、pH 值明显偏低; 黄旗海、青海湖、岱海总硬度、pH 值显著过高等。该部份湖泊基本分布于蒙新高原和东北高寒地区, 具一定的地域性。

(4) 除上述中、富、贫营养型外, 余下序号为 4、5、30 的 3 个湖泊。这 3 个湖泊各项营养指标值一般在或接近在中、富营养型范围内, 不具备贫营养型特征。但同一湖泊各营养指标相对均匀地交叉落入中、富营养型中, 使该湖泊与中、富营养型湖泊相似性都不高, 未能聚入中、富营养型。如呼伦湖, 耗氧(8.29)明显高于富营养型均值(6.038), 总氮(0.13)远低于中营养型均值(0.577); 洪泽湖浮游植物数(11.5), 远低于中营养型均值(26.8), 而透明度(0.3)较富营养型均值(0.523)更小。产生这种现象是湖泊自身固有特性所致, 还是指标值测定等过程人为误差的影响, 尚等进一步探明。

2. 各营养指标在营养型划分中的意义

用于类型划分的指标应具有的特点是: 在同一类型中相对稳定(变差系数值较小), 在不同类型间差异显著($t > t_{\alpha, 0.05}$)。由表 3 可知, 具有上述特点的指标有总氮、总磷、耗氧, 该三项指标对中、富营养型的划分显然最为重要; pH、总硬度、溶氧在中、富二营养型间无显著差异, 不适宜用于中、富二营养型的划分, 然而该三值过高或过低可作为贫营养型湖泊的标志, 浮游植物数量、透明度在同一类型湖泊中离散性太强, 作为类型划分仅有参考价值。综上所述, 上述 8 项湖泊营养状况调查中较常见的指标, 在综合反映湖泊营养状况本质特征方面尚难尽人意。平均水温、平均水深、细菌量、浮游植物初级产量与种群结构、浮游动物量、无机氮、活性磷、碱度等直接或间接与湖泊营养状况有关的因素在湖泊营养型划分中的作用, 由于目前资料不足, 未能加以比较分析, 有进一步探讨的必要。

3. 聚类结果与传统分类的比较

何志辉以浮游植物初级产量(或现存量)为主要依据的营养型划分^[2]和王洪道综合耗氧、总氮、总磷、透明度、浮游植物量及优势种 6 项指标等权分级评分的营养型划分^[1]与 fuzzy 聚类结果比较列于表 4。

由表 4 知, 聚类所成的中、富营养型与传统分类基本吻合, 有明显差异的仅 4 个湖泊, 其序号为 2、6、17、36。该 4 个湖泊在传统分类中划为中或富营养型, 而聚类结果将其排除于中、富营养型外并归入贫营养型。将该 4 个湖泊营养指标与表 2 标准比较, 除黄旗海外, 其余 3 个湖泊均具有较明显的贫营养型特征, 故可认为聚类结果有其合理性。

贫营养型湖泊由于彼此相似性差而难以聚为一类, 是 fuzzy 聚类分析方法用于湖泊营养型划分的不足之处。这一不足能否通过选取更能反映湖泊营养本质特征的指标作为聚类依据, 或改进聚类分析方法本身来加以克服, 尚待进一步探讨。

表 4 聚类结果与传统分类比较

Tab. 4 Compared with other classification method

营养分类 湖泊序号	何志辉[2]	王洪道[1]	聚类结果	营养分类 湖泊序号	何志辉[2]	王洪道[1]	聚类结果
38	贫	贫中	中	36		中	贫
23	贫	贫中	中	37		中	中
17	富	中	贫	18		富	富
24	富		富	19		贫中	贫
7	富		富	11		富	富
27	富	富	富	15		富	富
28	富	富	富	26		富	富
42	富		中	8		贫中	贫
21	中	中富	中	6		中富	贫
3	中	中富	中	39		中	中
34		贫中	贫	40		中富	中
2		中	贫	22		富	富
1		中	中	31		中富	中
35		中	中	25		中富	富
16		中富	富				

传统分类,出于选择单项比较指标或确定不同指标的权重以及确定评分等级的需要,必须预先肯定各种因素对湖泊营养状况影响的大小和选定类型划分标准。由于目前对湖泊营养状况演变规律认识尚有不足,处理上述问题难免带有一定的主观性,这是同一湖泊被不同人划分为不同营养型的原因。采用 fuzzy 聚类分析方法,直接从已知的湖泊营养指标数值出发,借助数学手段使其自然归类,然后由所聚类型的指标特征去确定其营养型(事实上可以由所聚类型的指标特征出发去找到或制定出适合我国湖泊营养状况特点的类型划分标准),具有较强的客观性。fuzzy 聚类分析方法,有望成为了解湖泊营养状况演变规律,准确划分湖泊营养类型的新途径。

参 考 文 献

- [1] 王洪道等. 中国湖泊资源. 北京, 科学出版社, 1989, 163—165, 248—251.
- [2] 何志辉. 中国湖泊和水库的营养分类. 大连水产学院学报, 1987, (3): 1—10.
- [3] 胡传林等. 保安湖渔业生态和渔业开发技术研究文集. 北京, 科学出版社, 1991, 108.
- [4] 贺仲雄. 模糊数学及其应用. 天津, 天津科学技术出版社, 1983, 152—176.
- [5] 中国科学院数学研究所统计组. 常用数理统计方法. 北京, 科学出版社, 1979, 3—4, 23—25.
- [6] 许仁忠. 模糊数学及其在经济管理中的应用. 成都, 西南财经大学出版社, 1987, 204—217.

FUZZY CLUSTER ANALYSIS OF LAKE TROPHIC TYPES

Cheng Changqi Xie Hongzhen

(Southwest Agricultural University, Chongqing 630716)

Abstract

By using fuzzy cluster analysis on 8 parameters including total-hardness, pH, dissolved-oxygen, COD, total-nitrogen, total-phosphorus, phytoplankton and diaphaneity of 42 main lakes in China, a trophic classification method is discussed for comparison with other classification methods in present publications. As the similarity value is " $\lambda=0.690$ ", 42 lakes can be classified into 13 trophic types, among which parameters belonging to mesotrophic lakes are of significant similarity, as well as eutrophic ones, except total-nitrogen, total-phosphorus and COD with remarkable differences, however, among which parameters belonging to oligotrophic lakes aren't of significant similarity, and a few parameters differ from the common parameter value. The results of fuzzy cluster analysis of lake trophic types are similar to those of the traditional methods, showing that fuzzy cluster analysis may be used in classification of lake trophic types.

key words fuzzy cluster analysis, trophic types of lake