

应用聚类分析对水库营养类型分类  
和 0—1 高指标判别方法

陈燕国 蔡少华

(水利部水库渔业研究所, 武汉 430073)

269-271

摘要 以聚类中正相(Q)分析方法对 19 座水库进行了营养类型的分类, 并与模糊聚类分析进行了比较。同时, 建立了水库营养类型 0—1 高指标的判别方法。最后, 设计了通用的分析和计算程序, 并指出水库营养类型分类中还需进一步解决的问题。

关键词 聚类分析 0—1 高指标判别方法

水库, 水体, 营养类型, 分类

## 一、引 言

使事物条理化是人类认识、改造世界的首要步骤, 聚类则是使事物条理化进行分类的基本方法之一。水体营养类型的分类, 正是基于此种观点, 而成为认识渔业资源特点和从事渔业资源特点和从事渔业实践的起始步骤。

水体营养类型分类问题的研究始于本世纪二十年代的 Thienemann 和 Naumann 对湖泊的营养类型分类研究工作<sup>[1]</sup>。作为人工湖的水库, 由于与湖泊有一定程度上的相似性, 因而在分类方法上也采用了湖泊营养类型分类中的分类方法。

影响水库营养水平的因素包括外部因素, 如集水区土壤、岩石、植被、地区的气候条件等特征, 内部因素如水库本身的理化状况和生物群落结构等, 多种因素形成的有机整体, 综合作用的结果, 但水库仍然被看作为边界明确的生态系统<sup>[2-5, 7]</sup>。水库的营养类型分类是力图依据水库的生态系统结构等内部因素和外部因素对其综合地影响形成的营养状态进行分类。

营养类型的分类指标经历了从单项到多项的发展过程。因各作者在进行营养类型分类时有侧重不同, 所以出现了不同的分类方法, 如依据初级生产力、营养盐的含量、细菌的数量和生物量的分类等<sup>[8, 9, 5]</sup>。后发展为用多指标描述水体的营养水平, 如 Watzel (1975) 给出了不同营养类型湖泊中浮游植物、初级生产力及有关特征将营养类型划分为超贫营养型、贫营养型、贫-中营养型、中营养型、中-富营养型、富营养型、特富营养型、腐植质营养型。

营养类型分类指标的选择由单指标发展为多指标是营养分类方面的一个进步, 因为它在认识客观事物方面克服了片面性, 而趋向全面、整体地认识分析对象。但是, 在对分类指标

湖北省自然科学基金项目(89513)。

本文于 1992 年 5 月 3 日收到, 1993 年 3 月 12 日第二次改回。

的选择和同一营养类型指标的界定值却明显地表现出差异来<sup>[1,10]</sup>。

尽管分类指标由单指标发展为多指标带来了营养类型方面的发展,但有关分类方法学的研究却少见。传统的方法是将实测数据与各营养类型的相应指标的标准值进行对照,来判定该水体的营养类型。因而,这样的判定方法不一致,导致结论的不一致,可比性不强<sup>[5]</sup>。

本文将利用聚类中正相(Q)分析方法对水库营养类型的分类方法进行探讨,并与模糊聚类分析进行比较,比较两种方法的优劣,建立了依据反映水体营养状况和状态变量进行营养等级判定的0-1高指标判别方法。最后,将这些方法设计成通用的计算机程序(程序略)。

## 二、聚类分析的原理、结果及方法比较

聚类分析的方法是各种各样的,但无论采用何种聚类分析方法对实体(Entity)进行分型划类,合理适当的选择分类指标及指标的界定值是分类的第一步骤。分类指标的选择需要考虑所选择的分类指标能综合地描述营养状况。

文中指标的选择,参考了一些研究者对于分类指标的选择,综合了他们共同的分类指标(表1)<sup>[11]</sup>。将进行营养类型分类的19座水库7项的原始数据列成表2(数据来源见文献<sup>[11]</sup>)。

表1 我国湖泊、水库营养水平浮游生物和营养物质的一些指标

Tab. 1 Some indicators for trophic level of plankton and nutrition of lake and reservoir in China

指 标	贫营养型	中营养型	富营养型
初级产量 (gO <sub>2</sub> /m <sup>2</sup> d)	0.449 0.174—0.963	1.701 0.534—5.50	4.507 1.970—10.42
浮游植物量 (mg/L)	0.662 0.008—1.280	3.856 0.690—21.0	14.06 3.70—49.58
浮游动物量 (mg/L)	0.959 0.158—2.193	2.103 0.280—17.60	3.591 0.587—9.520
有机物耗氧量 (mg/L)	1.506 0.820—3.650	5.710 1.300—13.58	10.847 2.55—53.70
总 氮 (mg/L)	0.447 0.220—0.840	1.224 0.270—4.120	1.750 0.645—4.770
无机氮 (mg/L)	0.232 0.070—0.656	0.532 0.050—2.057	0.643 0.056—3.374
总 磷 (mg/L)	0.085 0.030—0.250	0.242 0.0003—1.358	0.262 0.008—1.870

### (一) 正相(Q)聚类分析原理及分类结果和分析

聚类中正相(Q)分析是基于多元统计分析之上并应用最广的分析方法,该方法将研究实体作为数值分类的基本单位,依据实体间的诸多属性(attributes)即描述实体数量特征的各个信息项目进行分类。其基本思想是:首先将所研究的每个水库各自看成一实体,然后根

据反映水库间营养水平诸属性的相似程度予以分类。

表 2 各水库营养状态原始数据表  
Tab. 2 Data of nutritive state of the analyzed reservoirs

库名	1 高关	2 杨树档	3 危水	4 卷桥	5 石门	6 陈坡	7 黄坡	8 温峡口	9 青山	10 云溪
有机物耗氧量(mg/L)	8.84	10.40	13.70	15.40	8.88	14.70	15.20	9.10	1.50	6.00
总磷(mg/L)	0.03	0.03	0.03	0.06	0.07	0.09	0.062	0.038	0.022	0.078
总氮(mg/L)	0.90	1.88	4.17	3.76	0.93	1.21	1.08	1.54	0.94	1.21
浮游植物量(mg/L)	2.64	7.59	2.34	5.37	3.24	4.31	3.27	2.44	0.68	0.95
浮游动物量(mg/L)	0.71	0.45	0.36	0.95	0.62	0.22	0.582	0.14	0.812	6.39
初级产量(gO <sub>2</sub> /m <sup>2</sup> d)	1.87	0.828	3.54	1.35	1.48	0.746	3.42	2.71	1.80	8.84
无机氮(mg/L)	0.591	0.671	0.401	0.211	0.835	2.112	0.699	0.099	0.683	0.156
库名	11 石壁	12 富水	13 王英	14 陆水	15 雨山	16 三湖连江	17 熊河	18 先觉庙	19 黑屋湾	
有机物耗氧量(mg/L)	3.30	4.50	3.30	3.60	15.40	5.80	5.69	6.95	6.09	
总磷(mg/L)	0.03	0.071	0.183	0.119	0.118	0.059	0.044	0.076	0.027	
总氮(mg/L)	1.48	1.93	1.75	2.50	2.82	1.24	0.44	0.84	1.06	
浮游植物量(mg/L)	0.18	0.12	0.88	1.46	0.596	0.271	7.07	7.87	24.89	
浮游动物量(mg/L)	3.53	0.69	1.34	1.25	9.12	0.69	0.64	0.81	1.03	
初级产量(gO <sub>2</sub> /m <sup>2</sup> d)	2.51	1.59	2.24	1.86	4.40	7.80	1.12	0.215	0.611	
无机氮(mg/L)	0.185	0.408	0.433	0.786	0.253	0.287	0.411	0.414	0.338	

表 3 聚类正相(Q)分析结果  
Tab. 3 The result of normal cluster analysis

顺号	类号	实体间距离	顺号	类号	实体间距离
1	5 1	0.6409325	10	16 10	2.554243
2	1P 17	1.027341	11	13 12	2.853592
3	4 3	1.520863	12	9 1	3.500211
4	14 12	1.611229	13	15 10	4.758955
5	11 9	1.795197	14	12 1	5.011744
6	7 1	1.832665	15	19 1	5.496967
7	8 2	1.853108	16	6 1	5.629235
8	2 1	2.119609	17	3 1	6.033927
9	17 1	2.49008	18	10 1	7.281196

本文将原始数据标准化,采用欧几里德距离(Euclidean distance)来度量水库之间营养水平上的相似性,用离差平方和聚类。

根据程序中的设计,程序中和各变量的取值分别为:

$$N=19 \quad M=7 \quad BA=-0.5 \quad Ty=2 \quad T_1=5 \quad T_2=8$$

$$Y_4=4 \quad Y_1 \neq 1 \quad Y_2 \neq 2 \quad Y_3 \neq 3 \quad Y_7 \neq 7$$

通过在 IBM-PC/XT 上的计算,得到 19 座水库的正相聚类结果,从并类结果(表 3)和

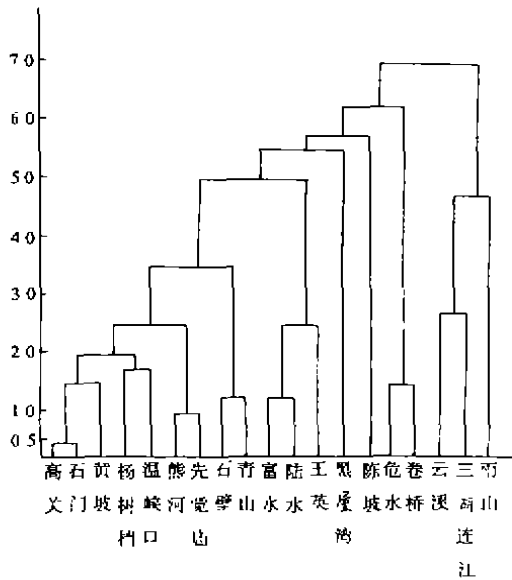


图 1 正相(Q)聚类分析谱系图  
Fig. 1 The hierachical clendrogram of normal cluster analysis

聚类谱系图谱(图 1)可以看到:

1. 实体(水库)按有关营养水平诸多属性的相似程度分类,揭示每一水库属于某一类,19 座水库被分成以下 9 个基本的子类:

(1)高关水库、石门水库、黄坡水库;(2)杨树档水库、温峡口水库;(3)熊河水库、先觉店水库;(4)石壁水库、青山水库;(5)富水水库、陆水水库、王英水库;(6)黑屋湾水库;(7)陈坡水库;(8)危水水库、卷桥水库;(9)云溪水库、三湖连江水库、雨山水库。同时,揭示同一类型的水库间的相似性程度,例如在高关水库、石门水库、黄坡水库所形成的一类中,石门水库同高关水库的相似性程度就高于黄坡水库同高关水库的相似性程度,也就是,在水库的营养水平方面,石门水库比黄坡水库更接近于高关水库。因此,聚类分析能定量地揭示各水库在营养水平上差异程度的高低,以利于比较及认识各水库的营养水平。

营养水平。

2. 由原始数据(表 2)的有关营养类型的分类指标(表 1),可以看到被分类的 19 座水库都为非贫营养水库。

3. 程序化的正相聚类分析方法可以随时将得到的有关水库营养状态的参数输入计算机进行分类,得到新的谱系图。

4. 通过正相聚类分析,可以定量地分析出一组水库中不同营养水平的水库所形成的诸子类之间的相互关系,从而知道这组水库在营养水平上的整体联系。

(二) 正相聚类分析和模糊聚类分析比较

为了将正相聚类分析和模糊聚类分析进行比较,对同样的 19 座水库用模糊聚类进行了

表 4 模糊聚类分析结果  
Tab. 4 The result of fuzzy cluster analysis

顺 号	类 号	实体间距离	顺 号	类 号	实体间距离
1	1 5	0.9723241	10	3 4	0.870249
2	13 14	0.9593271	11	2 17	0.8660762
3	1 8	0.9514745	12	1 2	0.8275196
4	12 13	0.9381523	13	1 9	0.8137996
5	17 18	0.9347538	14	1 3	0.8049802
6	9 12	0.9230922	15	10 16	0.7882448
7	9 11	0.914816	16	1 10	0.7756888
8	6 7	0.8812958	17	1 15	0.6671078
9	4 6	0.8711188	18	1 19	0.3813048

分析。实体相似性用相关系数度量。

1. 分类结果 利用模糊聚类分析,仅得到19座水库的7个类型( $0.30 \leq \lambda \leq 1.0$ ) (表4、图2)。

(1)高关水库、石门水库、温峡口水库;(2)熊河水库、先觉庙水库、杨树档水库;(3)王英水库、陆水水库、富水水库、青山水库、石壁水库;(4)陈坡水库、黄坡水库、卷桥水库、危水水库;(5)云溪水库、三湖连江水库;(6)雨山水库;(7)黑屋湾水库。

2. 两种方法的比较 从两种分类结果可以看到,模糊聚类分析同聚类正相分析的结果存在差别。其原因在于模糊聚类考虑到实体间的模糊性差异而减少了子类的数量,而水库营养类型的分类一旦将指标及界定值确定之后,无需考虑模糊性现象,因此,为了给出更为细致的分类结果,正相聚类分析更适合于营养类型分类。

### 三、水库营养类型的0-1高指标判别方法及营养等级排序

聚类分析可将一组水库依据它们状态变量的差异在营养水平上分成若干类,并能得到不同营养水平上的差异大小和属于同一子类的水库之间差异的大小。但聚类分析不可能对水库的营养等级作进一步地判断。例如,高关水库、石门水库和黄坡水库是同一个子类,且高关水库和石门水库的类间距离为0.6409,高关水库和黄坡水库的类间距离为1.833,因而石门水库与黄坡水库相比更接近于高关水库的营养水平。但我们还需要进一步地知道由高关水库、石门水库和黄坡水库应隶属于贫营养型、中营养型、富营养型,还是它们之间的过渡类型,其差异是因为它们的各指标都(或有多少)高于石门水库相应的指标产生的,还是因为它们的各指标都(或有多少)低于石门水库相应的指标产生的。为此,我们建立了以下方法,并称其为0-1高指标判别方法:

1. 分别以表1中的贫、中、富营养型指标的界定值与某一水库相应指标的测定值进行对照,以0表示该水库的某一指标小于表1中对应指标的界定值,而1则表示该水库的某一指标大于或等于表1中的对应指标的界定值。如果某水库实际测得的各指标完全符合表1中的贫(中或富)营养型的相应指标,则称该水库为标准贫(中或富)营养型水库。

2. 根据上述的表示方法,分别以表1中的贫、中和富营养的指标与表2中各水库的指标进行对照,然后分别统计每一水库高指标出现的频率(表5-7)。以贫(中或富)营养型的指

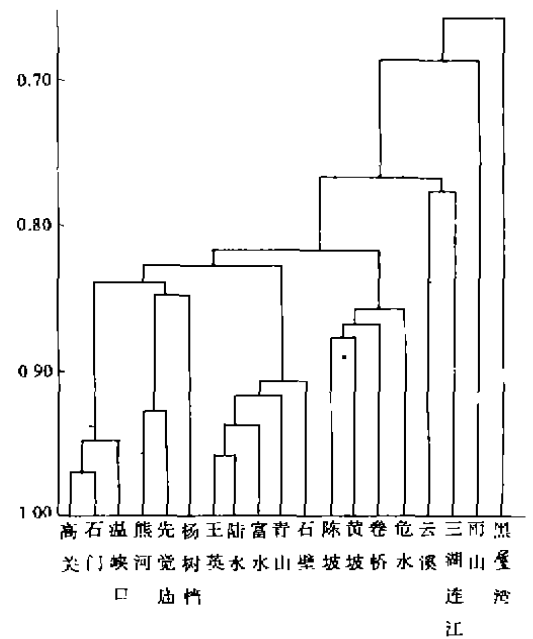


图2 模糊聚类分析谱系图

Fig. 2 The hierarchical dendrogram of fuzzy cluster analysis

标来衡量某一水库实际测得的相应指标,如果有 50% 的指标大于或等于贫(中或富)营养型的指标,则称该水库为强贫(中或富)营养型。进一步地,根据高指标出现的频率大小,我们定义下列术语以区别它们的等级差异,如果高指标出现的频率在 50%—60% 之间则称 0—1 强贫(中或富)营养型;依次类推有,在 60%—70% 之间称为 1—强贫(中或富)营养型;在 70%—80% 之间称为 2—强贫(中或富)营养型;80%—90% 之间称为 3—强贫(中或富)营养型;在 90—100% 之间称为 4—强贫(中或富)营养型。另一方面,如果以贫(中或富)营养型的指标界定值来衡量某一水库相应的指标,高指标出现的频率如果在 40%—50% 之间称为 0—弱贫(中或富)营养型,30%—40% 之间称为 1—弱贫(中或富)营养型,20%—30 之间称为 2—弱贫(中或富)型,10%—20% 之间称为 3—弱贫(中或富)营养型,0—10% 之间称为 4—弱贫(中或富)营养型。

表 5 各水库的营养参数高出贫营养界定值的频率

Tab. 5 The higher index occurrence frequency of trophic parameters of reservoirs comprised of the oligotrophic limitation

库名	有机物耗氧量 (mg/L)	总磷 (mg/L)	总氮 (mg/L)	浮游植物量 (mg/L)	浮游动物量 (mg/L)	初级产量 (gO <sub>2</sub> /m <sup>2</sup> d)	无机氮 (mg/L)	指标高出频率(%)
高关	1	0	1	1	0	1	1	71
杨树档	1	0	1	1	0	1	1	71
危水	1	0	1	1	0	1	1	71
卷桥	1	0	1	1	0	1	0	57
石门	1	0	1	1	0	1	1	71
陈坡	1	1	1	1	0	1	1	86
黄坡	1	0	1	1	0	1	1	71
温峡口	1	0	1	1	0	1	0	57
青山	0	0	1	1	0	1	1	57
云溪	1	0	1	1	1	1	0	71
石壁	1	0	1	0	1	1	0	57
富水	1	0	1	0	0	1	1	57
王英	1	1	1	1	1	1	1	100
陆水	1	1	1	1	1	1	1	100
雨山	1	1	1	0	1	1	1	86
三期连江	1	0	1	0	0	1	1	57
熊河	1	0	0	1	0	1	1	57
先觉庙	1	0	1	1	0	0	1	57
黑屋湾	1	0	1	1	1	1	1	86

3. 由上面的分析判定方法,如果以表 1 中的贫、中、富营养指标的只是定值来分别衡量 A、B 两水库,且最后 A 水库的高指标出现的频率高于 B 水库指标出现的频率,我们用符号“>”表示 A 水库的营养水平高于 B 水库的营养水平。

4. 根据上述方法得到营养等级的判定和排序如下:

(1) 高关水库、石门水库、黄坡水库的营养等级分别为:0—弱中营养型、2—弱中营养型和 2—弱强营养型,它们之间在营养水平上可排为:黄坡水库 > 高关水库 > 石门水库;

(2) 杨树档水库为 0—强中营养型水库,温峡口水库为 0—弱中营养型,它们之间在营养水平上可排序为:杨树档 > 温峡口水库;

表6 各水库的营养参数高出中营养界定值的频率

Tab. 6 The higher index occurrence frequency of trophic parameters of reservoirs comprised of the mesotrophic limitation

库名	有机物耗氧量 (mg/L)	总磷 (mg/L)	总氮 (mg/L)	浮游植物量 (mg/L)	浮游动物量 (mg/L)	初级产量 (gO <sub>2</sub> /m <sup>2</sup> d)	无机氮 (mg/L)	指标高 出频率(%)
高关	1	0	0	0	0	1	1	43
杨树档	1	0	1	1	0	0	1	57
危水	1	0	1	0	0	1	0	43
卷桥	1	0	1	1	0	0	0	44
石门	1	0	0	0	0	0	1	29
陈坡	1	0	0	1	0	0	1	43
黄坡	1	0	0	0	0	1	1	43
温峡口	1	0	1	0	0	1	0	43
青山	0	0	0	0	0	1	1	29
云溪	1	0	0	0	1	1	0	43
石壁	0	0	1	0	1	1	0	43
富水	0	0	1	0	0	0	0	14
王英	0	0	1	0	0	1	0	29
陆水	0	0	1	0	0	1	1	43
雨山	1	0	1	0	1	1	0	57
三湖连江	1	0	1	0	0	1	0	43
熊河	0	0	0	1	0	0	0	14
先觉庙	1	0	0	1	0	0	0	29
黑屋湾	1	0	0	1	0	0	0	29

表7 各水库的营养参数高出富营养界定值的频率

Tab. 7 The higher index occurrence frequency of trophic parameters of reservoirs comprised of the eutrophic limitation

库名	有机物耗氧量 (mg/L)	总磷 (mg/L)	总氮 (mg/L)	浮游植物量 (mg/L)	浮游动物量 (mg/L)	初级产量 (gO <sub>2</sub> /m <sup>2</sup> d)	无机氮 (mg/L)	指标高 出频率(%)
高关	0	0	0	0	0	0	0	0
杨树档	0	0	1	0	0	0	1	29
危水	1	0	1	0	0	0	0	29
卷桥	1	0	1	0	0	0	0	29
石门	0	0	0	0	0	0	1	14
陈坡	1	0	0	0	0	0	1	29
黄坡	1	0	0	0	0	0	1	29
温峡口	0	0	0	0	0	0	0	0
青山	0	0	0	0	0	0	1	14
云溪	0	0	0	0	1	1	0	29
石壁	0	0	0	0	0	0	0	0
富水	0	0	1	0	0	0	0	14
王英	0	0	0	0	0	0	0	0
陆水	0	0	1	0	0	0	1	29
雨山	1	0	1	0	1	0	0	43
三湖连江	0	0	0	0	0	1	0	14
熊河	0	0	0	0	0	0	0	0
先觉庙	0	0	0	0	0	0	0	0
黑屋湾	0	0	0	1	0	0	0	14

(3) 熊河水库、先觉庙水库分别为 3-弱中营养型, 2-弱中营养型, 它们之间在营养水平上可排序为: 先觉庙水库 > 熊河水库;

(4) 石壁水库、青山水库分别为 0-弱中营养型和 2-弱中营养型, 它们之间在营养类型水平上可排为: 石壁水库 > 青山水库;

(5) 富水水库、陆水水库、王英水库分别为 0-强贫营养型、0-弱中营养型和 2-弱中营养型, 它们之间在营养水平上可排序为: 陆水水库 > 王英水库 > 富水水库;

(6) 黑屋湾水库为 3-强贫营养型水库, 且距中营养型水平较远;

(7) 陈坡水库为 0-弱中营养型水库;

(8) 危水水库、卷桥水库都为 0-弱中营养型, 且进一步依据富营养型的指标进行判定, 表明两水库营养水平基本相同;

(9) 云溪水库、三湖连江水库、雨山水库分别为 2-弱富营养型、3-弱富营养型和 0-弱富营养型, 它们在营养水平上的等级关系为: 雨山水库 > 云溪水库 > 三湖连江水库。

## 四、讨 论

从上述研究分析过程中, 我们可以看到:

1. 由于正相聚类分析与模糊聚类分析相比较, 对同一组水库可以得到更为细致地分类结果, 因而, 我们认为正相聚类分析更适用于水库营养类型的分类。

2. 利用聚类分析的方法对水库营养类型进行分类, 可以得到属于同一营养型的水库之间在营养水平上的差异, 同时也可以得到不同营养类型之间在营养水平上的差异。将聚类分析方法程序化之后, 方法的使用也比较方便。但是, 聚类分析不能对水库依据其营养等级的高低予以排序, 也不能判定营养水平的高低。

3. 所建立的 0-1 高指标判别方法, 可以判断各水库所属营养水平, 并可依据它们营养水平的高低进行排序。

4. 对水库营养类型的分类指标的选择及各指标界定值的确定还需要作进一步的研究, 即需要确定能恰如其份地反映水库营养状态的指标及相应的界定值。



## 参 考 文 献

- [1] 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。
- [2] Patten, B. C., *Ecological Systems Analysis and Fisheries Science. Trans. Amer. Fish. Soc.*, 1969, (3):570—581.
- [3] Brylinsky, M. & Mann, K. H., *An analysis of factors governing productivity in lakes and reservoirs, Limnology and Oceanography*, 1973, 18(1):14.
- [4] Brylinsky, M., *Estimating the productivity of lakes and reservoirs. International Biological Programme 22. The Functioning of Freshwater Ecosystems*, edited by Le Cren, E. D. & R. H. Lowe—McConnell, Cambridge University Press, 1980, 411—453.
- [5] 陈敬存。关于水库渔业生产性能评价问题。水库渔业,1982,(1),9—19。
- [6] 陈敬存。湖北省水库渔业发展问题剖析。水利渔业,1986,(3),2—7。
- [7] 何吉湘等。湖南省中型水库渔业性能地理环境指标的初步研究。农业现代化研究,1990,11(1):37—41。
- [8] Wetzel, R. G., *Limnology (second edition)* Saunders College Publishing, New York, 1981.
- [9] 何志辉。湖泊的营养分类。水库渔业,1982,(4):46—51。
- [10] 李思发。水库养鱼和捕鱼。上海科学技术出版社,1988。
- [11] 湖北省水库调查队。湖北省水库渔业生产性能调查。水库渔业,1982,(4):10—47。
- [12] 贺仲雄。模糊数学及其应用。天津科学技术文献出版社,1984。
- [13] 扈垵编。实用模糊数学。北京,科学技术文献出版社1989。

## APPLICATION OF CLUSTER ANALYSIS IN THE TYPICAL TROPHIC LEVEL CLASSIFICATION OF RESERVIOR AND 0—1 HIGHER INDEX DISCRIMINATION

Chen Yanguo      Cai Shaohua

*(Institute of Reservoir Fisheries, Ministry of Water Resources & Chinese Academy of Sciences, Wuhan 430073)*

### Abstract

The trophic levels of 19 reservoirs have been classified by using Q - matrix in cluster analysis and its comparison with fuzzy cluster analysis has also been made, and a new 0—1 higher index discrimination for the trophic level classification of one reservoir has been established. All methods are transferred into computer programmes and the impending problems of the trophic level classification are put forward.

**Key words**      Cluster analysis, 0—1 higher index discrimination