第4卷第2期　　湖泊科学　　Vol. 4, No. 2
1992年6月　　JOURNAL OF LAKE SCIENCES　　June, 1992

73—78

动态聚类分析与太湖水质分区①

黄漪平②　　P 343.3

（中国科学院南京地理与湖泊研究所，南京 210008）

提要　根据1987—1988年太湖实测水质数据，利用系统动态聚类分析方法，将17个指标按营养盐、有机质、有毒有害物质三种污染类型进行综合评估，获得了太湖水域各类污染指标空间分布态势及差异特点，揭示了太湖现存的三种污染模式及地域分布状况，为进一步治理和改善太湖水质提供依据。

关键词　动态聚类分析，水域环境，水质指标　太湖

在水域环境调查中，如何根据有限测点的监测指标，客观地衡量水质特性的空间分布，目前尚未有统一综合分析方法。本文基于实测数据，采用系统动态聚类分析[1]，将诸多指标逐步分类归属和凝聚，对太湖水质指标作相对类别的分区，以了解太湖水质指标空间分布态势和水质分区特点，寻求进一步控制水质污染物质的有效途径。

一、分析方法

聚类分析有系统聚类、动态聚类、图论聚类、模糊聚类等。本文采用动态聚类方法，分别对太湖水域的有机质和卫生指标（包括DO、BOD_5、COD_{Mn}、TOC、大肠杆菌群数），营养元素（包括TN、TP、SiO_2、TFe），有毒有害指标（酚、氰、汞、铬、砷、铜、铅、非离子氨NH_3）等进行归属聚类，以了解太湖水质指标空间分布的特点。

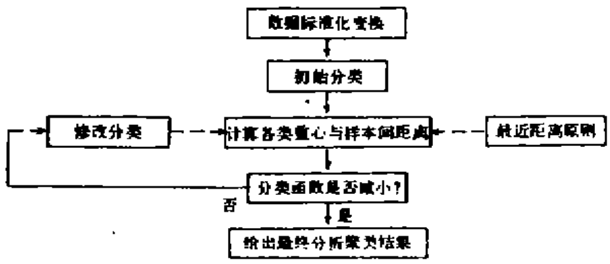

图1　动态聚类分析框备
Fig. 1　Dynamic cluster analysis on water quality indexes

系统的动态聚类就是把各样本个体逐步合并成一些较大的类，直至全部总体都处在一个大群类之下。具体分析程序有以下几个步骤：

① "七五"国家科技攻关课题（75—59—04—06）内容之一。

② 本文承无锡市环保监测研究所、中国科学院南京地理与湖泊研究所潘红玺、范成新、袁静秀等同志协同工作、附图由张辉王清绘在此一并致谢。

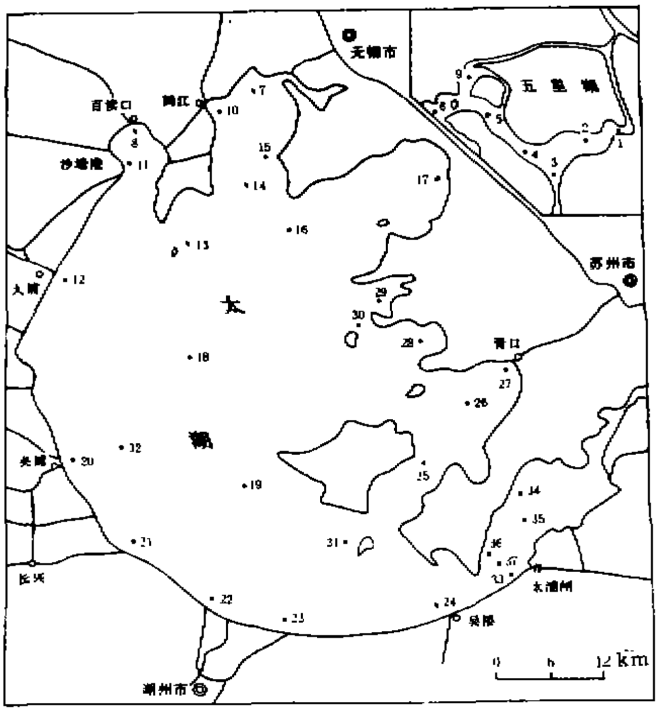

图2　太湖测点位置(1987—1988)

Fig. 2　The locations of measuring points in Taihu Lake(1987—1988)

1. 数据变换处理

在对原始数据的分析处理过程中，首先要对它们进行标准化变换，即对各列数据中心化，再用标准差进行归一化。设有 n 个采样点，m 种指标，则构成 $n\times m$ 的矩阵为：

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

则 $$x'_{ij}=\frac{x_{ij}-\bar{x}_j}{S_j}\qquad \begin{pmatrix} i=1,2,\cdots,n \\ j=1,2,\cdots,m \end{pmatrix} \qquad (1)$$

式中 $$\bar{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij},$$

$$S_j=\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2\right]^{1/2}。$$

通过这种标准化变换后，使得每列数据的平均值为零，方差为1，以便使各样本数据保持相对的稳定性。

2. 选择凝聚点

首先将 n 个样本人为地分为 K 类，再计算每一类的重心，并以此作为各类的凝聚点，对样本进行初始分类。亦即先求出：

$$S(i)=\sum_{j=1}^{m}x'_{ij}\qquad i=1,2,\cdots,n$$

$$M=\max_{1\leqslant i\leqslant n}S(i)$$

$$P=\min_{1\leqslant i\leqslant n}S(i)$$

如欲将全部样本分为 K 类(本文中取 $K=5$)，则对每一个样本 x_{ij} 按下式计算：

$$\frac{(K-1)[S(i)-P]}{M-P}+1$$

假定与上式计算所得数据最接近的整数为 k，则将样本 x_{ij} 归入第 k 类($1\leqslant k\leqslant K$)。

3. 计算分类函数

若 $x_1,x_2,\cdots,x_n$ 表示所分析的 n 个样本点，经初始分级聚为 K 个类别，以 $G_1,G_2,\cdots,G_K$ 表示，它们的凝聚重心分别为 $\bar{g}_1,\bar{g}_2,\cdots,\bar{g}_K$，每一类 G_k 中各含有的样本数记为 $n_1,n_2,\cdots,n_k$，且满足

$$\bar{g}_k=\frac{1}{n_k}\sum_{j=1}^{n_k}\bar{x}_j\qquad k=1,2,\cdots,K$$

$$\bar{x}_j\in G_k$$

如果欲将 K 类样本中的 G_p 与 G_q 合并为新类 $G_r(p、q、r=1,2,\cdots,K)$ 且各自含有 n_p, n_q, n_r 个样本数 $(n_r=n_p+n_q)$，则三者的重心 $\bar{x}_p, \bar{x}_q, \bar{x}_r$ 又须满足如下重心分配原则：

$$\bar{g}_r = \frac{1}{n_r}(n_p\bar{g}_p + n_q\bar{g}_q)$$

对于每一个样本 x_i，我们考察它与各类 $G_k(k=1,2,\cdots,K)$ 间的"距离"D_{ik}：

$$D_{ik}^2 = (\bar{x}_i - \bar{g}_k)^T \cdot (\bar{x}_i - \bar{g}_k) = \sum_{j=1}^{m}(x_{ij}' - \bar{g}_k)^2$$

$$(i=1,2,\cdots,n;\quad k=1,2,\cdots,K)$$

式中 T 表示对 $(\bar{x}_i - \bar{g}_k)$ 的转置运算，最后再定义分类出数为：

$$f(G_1,G_2,\cdots,G_K) = \sum_{i=1}^{n} D_{ik}^2$$

上述分类函数的判别原则是使它逐渐减小，直至达到其最小值为止。

4. 逐级聚类

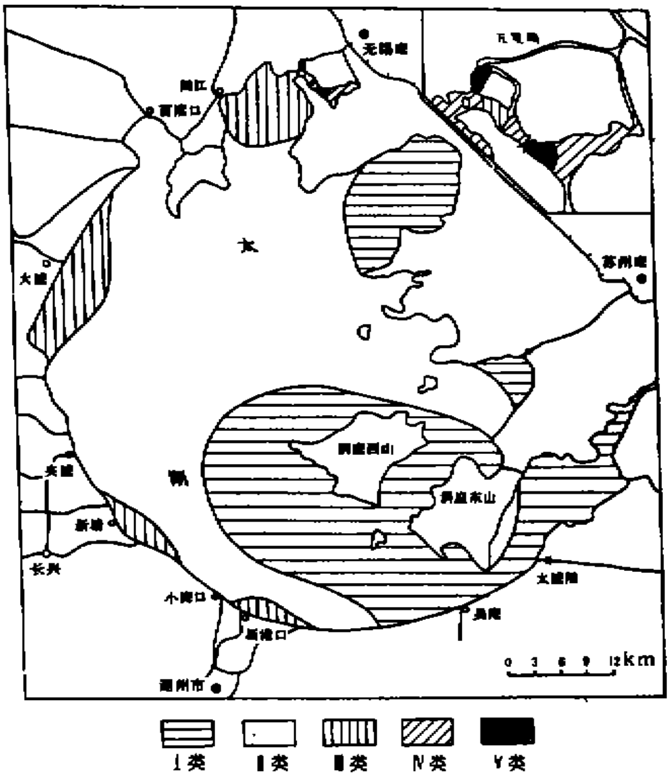

图 3　太湖有机质卫生指标聚类分布(1987—1988)

Fig. 3　The dynamic clustered distribution of organic and hygienic indexes in Taihu Lake (1987—1988)

由上述计算使每一个样本 x_i 与各类 G_K 间的"距离"逐级接近，直至新计算的重心与开始时的凝聚点完全相同，则计算过程终止，否则将重复上述计算过程，迭代的结果会使分类函数 f 逐渐减小而趋于定值。这就是利用最小距离原则的一个优点，通过它来判断迭代是否终止。

太湖水域水质指标动态聚类分析框图如图 1 所示：

按逐级聚类要求编制了计算程序，输入前述 17 项指标，输出则是对它的聚类分析结果。

二、动态聚类划区分析结果

将用上述方法，分别对 1987 年 5 月—1988 年 3 月四次太湖水质调查中有机质卫生指标、营养盐、有毒有害指标进行归属聚类。根据上述指标，预先划定聚合成五类，则可分别求出上述三种类型指标的类别以及各类别所包含的水域范围。最终分类结果见表 1—3，并绘出相应的三类指标划区分布图 3—5。

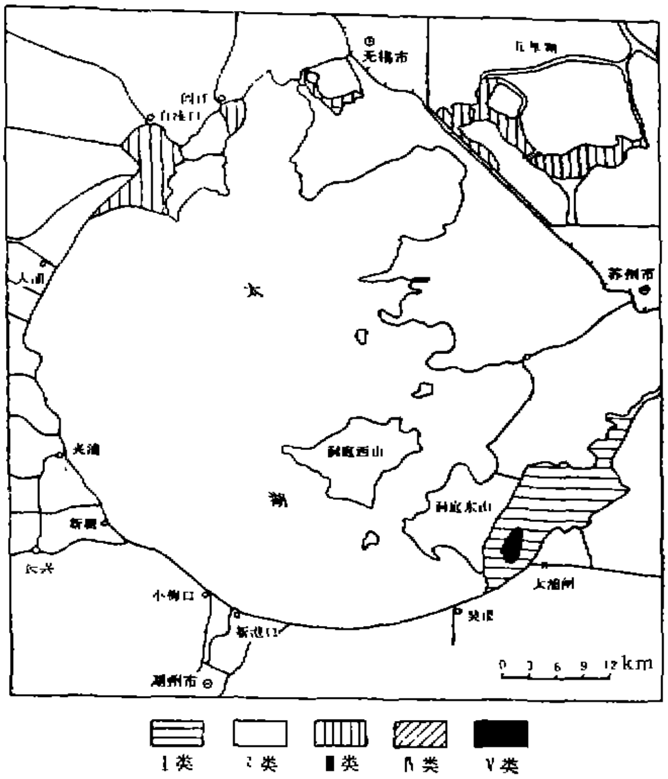

图 4　太湖营养元素指标聚类分布(1987—1988)

Fig. 4　The dynamic cluster distribution of nutritive elements indexes in Taihu Lake (1987—1988)

从图3中可以看出五里湖有机质卫生指标污染较重，北部梅梁湖、西部入湖河口大浦口、新塘港、新港口次之，中部开敞区又次之，东部湖湾、东太湖及洞庭东、西山周围水质最佳。因此，水质在区域上存在明显差异。由图4看出，除五里湖、闾江口局部水域，西北湖湾(竺山湖)属第三类别外，绝大部分湖区属第2类，东太湖属1类，个别类属5类，说明全湖营养状况就局部虽有差异，但从总体上而言则相近。从图5有毒有害物质分类来看，大部开敞湖区及东太湖属一类；西部、南部沿岸及洞庭东、西山之间属2类；梅梁湖、胥口附近及五里湖属3类；五里湖东部最差属5类。

为了简化分析计算步骤，也可按三类聚类进行分析，其方法和过程与按五类聚类完全相同，其结果也类同，只是某些类别所含测点范围有所增减，其差异将较五类聚类分析更小。

表 1　按有机质卫生指标五类动态聚类分析测点归属统计

Tab. 1　Classification of measuring points of 5-type dynamic cluster analysis according to the organic and hygienic indexes

分类	Ⅰ类	Ⅱ类	Ⅲ类	Ⅳ类	Ⅴ类
测点归属	19 24 25 27 29 31 33 34 35 36 37	8 11 13 14 15 16 17 18 20 22 26 28 30 32	5 7 10 12 21 23	1 2 3 6	4 9

表 2　按营养盐指标五类动态聚类分析测点归属统计

Tab. 2　Classification of measuring points of 5-type dynamic cluster analysis according to the nutritive elements indexes

分类	Ⅰ类	Ⅱ类	Ⅲ类	Ⅳ类	Ⅴ类
测点归属	32 33 34 35 36	3 7 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	1 2 4 5 6 8 9 10 11		37

表3　按有毒有害指标五类动态聚类分析测点归属统计

Tab. 3　Classification of measuring points of 5-type dynamic cluster analysis according to the toxicity and harmful index

分类	Ⅰ类	Ⅱ类	Ⅲ类	Ⅳ类	Ⅴ类
测点归属	11 13 16 17 18 19 20 29 31 32 35 36 37	3 5 8 12 14 15 21 22 23 24 25 26 28 33	2 6 7 9 10 27 34	4	1

三、结　语

通过对太湖不同污染类型17个指标的水质动态聚类分析，揭示出1987—1988年全湖水域各类指标类别分级的空间分布趋势和差异：

1. 太湖三种污染类型不同指标的空间分布特点是北部高于南部，湖西部、西南及南部入湖河口附近水域水质各项指标均较东部及开敞湖区为高；

2. 营养盐指标除五里湖、闾江口局部水域和西北竺山湖区较差外，大部湖区属Ⅰ类，差异较小。

应该说明的是动态系统聚类与一般水质评价的分级完全无关。

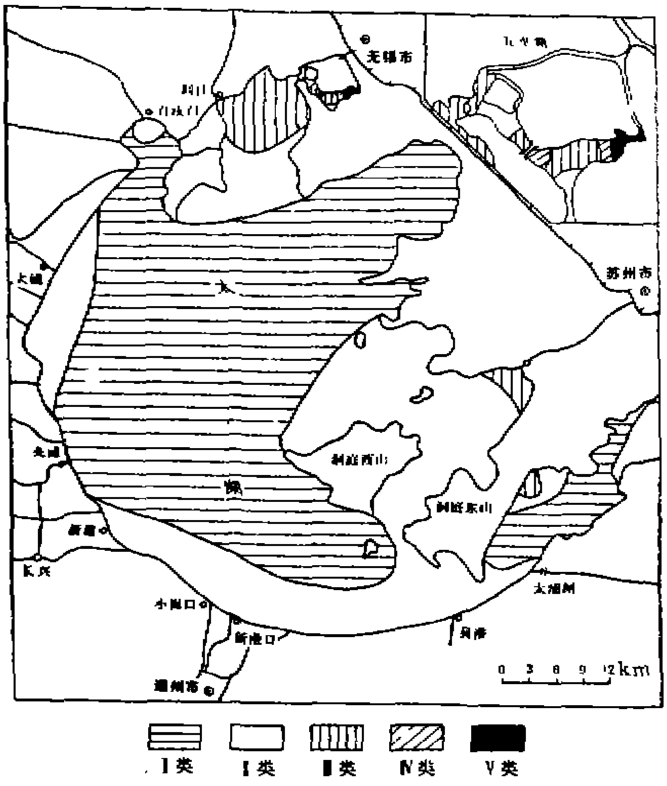

图5　太湖有毒有害指标聚类分布(1987—1988)

Fig. 5　The dynamic cluster distribution of indexes of toxic and harmful materials in Tainu Lake (1987—1988)

从上述动态分析获得的分区结果中，还可以从中探求改善水质的途径，如从上述图3—5不同类型污染物的分布中，污染程度受点污染源、入湖河道及梁溪河等出湖河道倒灌入湖的污染负荷量的影响较大。因此，为改善太湖水质，必须从控制入湖河道及出湖河道倒灌的污染负荷量入手，并根据不同水情严格控制入湖河道及工矿废水排放标准，使污染负荷量不超过规定的水环境容量，以使太湖水质不再恶化并逐步改善。

采用动态聚类分析方法，藉PC-1500微机计算即可获得湖泊水域水质监测指标理想的时空分布结果，这对水域水质指标空间分布的研究不失为是一种简便易行的方法。

参 考 文 献

[1] 罗积玉、邢瑛。经济统计分析方法及预测。北京，清华大学出版社，1987：192—197。

DYNAMIC CLUSTER ANALYSIS AND DISTRIBUTIONS OF WATER QUALITY IN TAIHU LAKE

Huang Yiping

(Nanjing Institute of Geography and Limnology, Academia Sinica, Nanjing 210008)

Abstract

According to the date of water quality collected from Taihu Lake in 1987—1988, seventeen indexes in three pollution types of organic matter, hygienic index, nutrients, toxic and harmful materials were evaluated by means of a systematic dynamic cluster analysis. The spatial distribution of pollution index in Taihu Lake and the differences of index character were obtained, and the results revealed three type pollution models and their distribution status. This paper offers an important information for further improving water quality and controlling water pollution of Taihu Lake.

Key Words dynamic cluster analysis, water area environment, index of water quality, Taihu Lake