

应用聚类分析划分孢粉带(区)的实例

沈 才 明

(中国科学院南海海洋研究所)

唐 领 余

(中国科学院南京地质古生物研究所)

一、序 言

随着计算机的普遍使用,孢粉分析学者对于数值方法在第四纪孢粉分析中的应用有了长足的认识。聚类分析、全成分分析、因子分析、回归分析、时间序列分析等多变量统计分析已用于处理第四纪孢粉学问题。应用这些方法划分孢粉带、孢粉分布区,进而探讨孢粉谱和植被、气候间的关系,定量解决第四纪孢粉学和古生态学的问题。在我国,虽有少数孢粉学者做了这方面的工作,但工作开展得还不够充分。本文是应用Q型聚类分析方法的一个实例。试图用聚类分析方法对珠江三角洲和内蒙古旱海的两个钻孔和南中国海海域表层样品作孢粉带划分和孢粉区划。

系统聚类分析的基本思想是:首先将每个样品看成一类,然后根据样品间的相似程度并类,并计算新类与其它类间的距离,再选择最相似者并类,每合并一次,减少一类,继续这一过程,直至所有样品都合并成一类为止。在计算分析过程中,表示样品间相似性尺度的有欧氏距离、绝对值距离、切比雪夫距离、兰氏距离、马氏距离、夹角余弦、相关系数、斜交空间距离等。聚类方法有最短距离法、最长距离法、中间距离法、重心法、类平均法、可变类平均法、可变法、离差平均和法等。本文采用的相似性尺度是相关系数。聚类方法是最短距离法。

二、孢粉类型的删选

孢粉谱的数据量往往是很大的,例如在珠江三角洲钻孔和南中国海海域的表层样品中都发现有150多种孢粉类型,内蒙古旱海也发现有60多种类型。这样庞大的数据阵用以计算是极为费时的,也是没有必要的。因此需要对孢粉类型进行删选,以较少的孢粉类型来反映原有的孢粉谱的信息特征,并能适合于研究的目的。

孢粉类型的删选应考虑孢粉类型的代表性。由于在第四纪孢粉研究中,钻孔剖面的孢粉带往往是地方性花粉带。而钻孔中地方性花粉带的划分又是为了建立地区性花粉带,因此,在删选过程中,既要选择具有地区代表性意义的木本植物花粉,也要适当选择部分地区性的草本植物花粉和蕨类植物孢子。但要删除仅有局部性质的一些类型,如内蒙古旱海样品中的狐尾藻(*Myriophyllum*)。在有些样品中,当水生植物花粉数量增加至80—90%时,而*Myriophyllum*几乎占水生植物花粉的95%以上,因而在计算中应删除。同时,还应删除那些含量少的地方性类型和偶见的地区性类型,如珠江三角洲钻孔中的Liliaceae, *Typha*, Borraginaceae, *Ceraopteris*等草本植物花粉和蕨类植物孢子以及 *Canarium*, *Ulmus*等木本植物花粉。在南中

国海海域的表层样品中,其孢粉组合具有地区性,因此,孢粉类型的删选则主要依据其含量,同时着重木本植物花粉。此外,孢粉类型的删选,还取决于研究的目的、所研究的地区以及计算机的存储能力等。一般选取 10—30 个孢粉类型。在分析珠江三角洲 ZK₁₅ 钻孔和南中国海海域表层样品时,考虑到两者地处亚热带、热带,孢粉类型丰富,因而选取较多的孢粉类型。珠江三角洲的 ZK₁₅ 钻孔选取 34 个孢粉类型。它们是 *Microlepia*, *Cibotum*, *Cyathea*, *Lygodium*, *Polypodiaceae*, *Pteridium*, *Pteris*, *Pinus*, *Cycas*, *Altingia*, *Liquidambar*, *Araliaceae*, *Cyperaceae*, *Elaeocarpus*, *Euphorbiaceae*, *Castanopsis*, *Castanea*, *Lithocarpus*, *Quercus*, *Gramineae*, *Hamameliaceae*, *Ficus*, *Syzygium*, *Randia*, *Rutaceae*, *Samydaceae*, *Eurya*, *Sterculiaceae*, *Alnus + Carpinus*, *Ranunculaceae*, *Engelhardtia*, *Ilex*, *Gesnertaceae*, *Pterocarya*。南中国海海域表层样中选择 22 个孢粉类型(见分区一节);内蒙古旱海的样品选取 16 个孢粉类型,它们是 *Betula*, *Pinus*, *Ulmus*, *Rhamnus*, *Ephedra*, *Artemisia*, *Chenopodiaceae*, *Compositae*, *Gramineae*, *Polygonum*, *Cyperaceae*, *Cruciferae*, *Ranunculaceae* 以及 *Trees*, *Steppe*, *Swamp + aquatic*。应该说明的是,在南中国海海域和内蒙古旱海的孢粉类型中,我们将乔木(*Trees*)、草本或草原(*Herbs or Steppe*)蕨类或沼泽水生类(*Pteridophyta or Swamp + aquatic*) 3 大类也作为 3 个孢粉类型参加计算,有少数孢粉类型是几个孢粉种类的合并。

三、系统聚类分析实例

(一)划带

一个钻孔或其它地层剖面中的孢粉组合的变化是连续不断的,但也是不均衡的。要更好地叙述、对比、解释和讨论地层剖面的变化,就需将孢粉地层划分成更小的单位,即孢粉带。数值方法取得的花粉带,是基于沉积物中所包含的花粉和孢子的种类和数量及指定的数学尺度,不必参考沉积物岩性、推断的气候、古植被或时代,因而就地层学意义来说,数值方法确定的孢粉带是组合带。

聚类分析划带,是在计算出聚类分析结果后,将结果(如表 I),用二维空间的聚类谱系图(插图 2,4)表示出来,然后确定定义类的阈值。阈值的确定主要考虑以下几个方面:

- (1) 所分类重心之间的距离要尽可能大;
- (2) 确定的类中的元素不宜过多;
- (3) 分类的数目必须符合实用目的。

根据确定的阈值将样品分成若干类,在此基础上进行划带,下面是两个实例。

1. 珠江三角洲 ZK₁₅ 孔

珠江三角洲 ZK₁₅ 孔位于广东省中山县。其孢粉类型丰富(插图 1),在聚类分析中,我们对 34 个孢粉类型进行聚类,根据聚类结果(表 I),得出其聚类谱系图(插图 2)。在阈值取 0.65 时, ZK₁₅ 孔 26 个样品可以分成 4 类,这 4 个类相当于 ZK₁₅ 孔的 4 个孢粉组合带(表 II)。在剖面中,在带 2 和带 3 间有一段无孢粉的地层。得出的 4 个孢粉组合带是:

(1) *Pinus-Castanea-Gramineae-Polypodiaceae* 孢粉带: 以 *Pinus*, *Castanea*, *Liquidambar*, 和 *Gramineae* 含量高为特征。

(2) *Castanopsis-Castanea-Elaeocarpus* 孢粉带: *Pinus*, *Castanea*, *Liquidambar*, *Gramineae* 减少,而 *Castanopsis*, *Elaeocarpus*, *Lithocarpus* 等增多。

(3) *Castanopsis-Lithocarpus-Elaeocarpus* 孢粉带: *Castanopsis* 高含量, 还有较多的

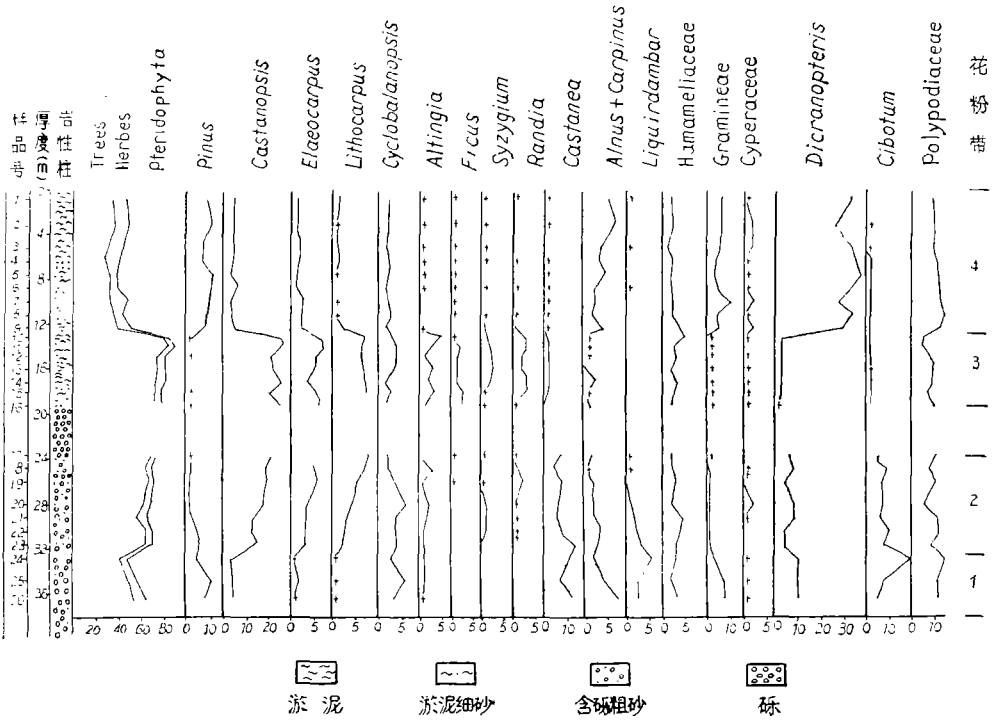


插图1 珠江三角洲 ZK₁₅ 孔孢粉图式
Pollen diagram for core ZK₁₅ in the Zhujiang Delta

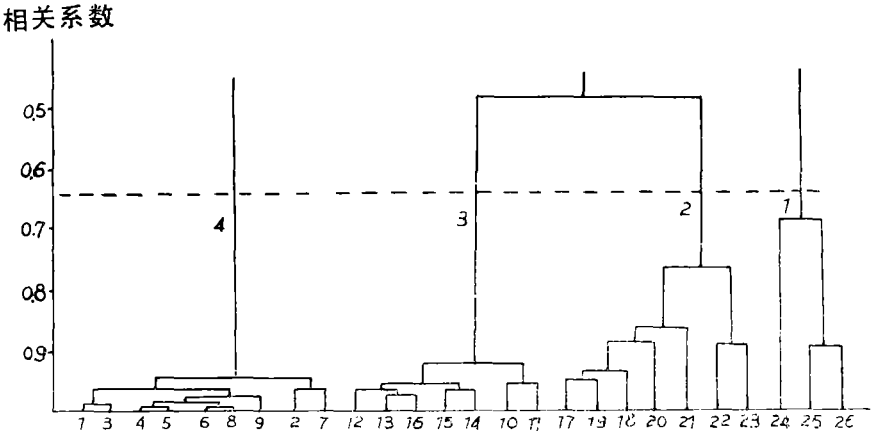


插图2 珠江三角洲 ZK₁₅ 孔聚类谱系图
Dendrogram of cluster for core ZK₁₅ in the Zhujiang Delta

Lithocarpus, *Elaeocarpus*, *Altingia*, *Euphorbiaceae*, *Randia*, *Ficus*, *Cycas*, *Araliaceae*, *Syzygium* 等。

(4) *Pinus*-*Castanopsis*-*Gramineae*-*Dicranopteris* 孢粉带：以 *Dicranopteris* 的高含量和较多的 *Pinus*, *Castanopsis*, *Gramineae* 为特征。

2. 内蒙古旱海 H₁ 孔

表 1 珠江三角洲 ZK₁ 孔聚类结果
Result of cluster for core ZK₁ in the Zhujiang Delta

聚类步骤	聚类样品号		相关系数
1	4	3	0.997247003
2	8	6	0.995437162
3	3	1	0.994718916
4	6	5	0.994137009
5	9	5	0.983097184
6	16	13	0.976975925
7	13	12	0.97537111
8	5	1	0.974029339
9	15	14	0.973049403
10	7	2	0.971136899
11	14	12	0.966871366
12	11	10	0.964367893
13	19	17	0.953264892
14	2	1	0.951400366
15	18	17	0.947325835
16	12	10	0.92959717
17	26	25	0.904951397
18	23	22	0.902413619
19	20	17	0.898859087
20	21	17	0.875065675
21	22	17	0.76939718
22	25	24	0.691318448
23	17	10	0.593360872
24	24	1	0.355927274
25	10	1	4.61081911E-03

表 II 珠江三角洲 ZK₁ 孔聚类结果和孢粉带
Result of cluster and spore-pollen zones for core ZK₁ in the Zhujiang Delta

系统聚类		孢 粉 带	
类 号	样 品 号	样 品 号	带 号
4	1,2,3,4,5,6,7,8,9	1,2,3,4,5,6,7,8,9	4
3	10,11,12,13,14,15,16	10,11,12,13,14,15,16	3
2	17,18,19,20,21,22,23	17,18,19,20,21,22,23	2
1	24,25,26	24,25,26	1

旱海位于内蒙古自治区凉城县,靠近岱海的东边。H₁ 孔为旱海中心浅钻,仅 2.5m 深。共采样 18 块(插图 3)。根据聚类分析结果,得出其聚类谱系(插图 4)。取阈值为 0.8 时,样品可分成 3 类,在利用类别进行划带时,根据样品的顺序性,作适当调整,可得到 5 个孢粉组合带(表 III)。

旱海 H₁ 孔孢粉组合带分别是:
(1) *Chenopodiaceae-Artemisia-Compositae-Cyperaceae* 优势带。伴有 *Pinus*, *Gramine-*

表 III 内蒙古旱海 H₁ 孔聚类结果和孢粉带
Result of cluster and sporo-pollen zones for core H₁ from Hanhai
Lake in Nei Monggol

系统聚类		孢 粉 带	
类 号	样 品 号	样 品 号	带 号
3	1,2,3,4,5,9,12,13,14	1,2,3,4,5	5
2	6	6	4
1	7,8,10,11,15,16,17,18	7,8,9,10,11	3
		12,13,14	2
		15,16,17,18	1

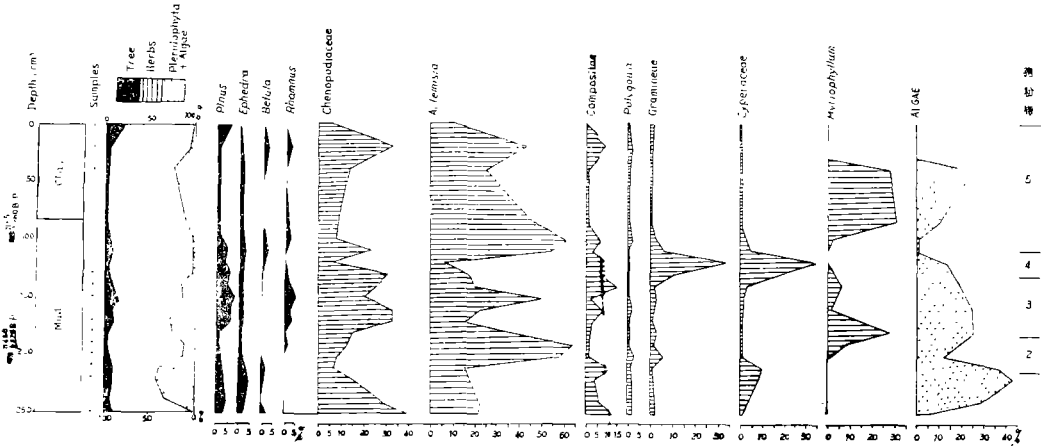


插图 3 内蒙古旱海 H₁ 孔孢粉图式
Pollen diagram for core H₁ from Hanhai Lake in Nei Monggol

- ae, *Ephedra*, *Ranunculaceae*, *Rhamnus*, *Polygonum* 等。
- (2) *Artemisia*-*Chenopodiaceae*-*Gramineae*-*Myriophyllum* 优势带。并含有较多 *Polygonum*, *Compositae*, *Cruciferae*, *Cyperaceae*, 及少量 *Betula*, *Ephedra*。
- (3) *Chenopodiaceae*-*Artemisia*-*Compositae*-*Rhamnus* 优势带。其他成分有: *Pinus*, *Gramineae*, *Myriophyllum*, *Polygonum*, *Cyperaceae*, *Cruciferae*, *Liliaceae*。
- (4) *Cyperaceae*-*Gramineae*-*Compositae*-*Pinus* 优势带。并含有较多 *Chenopodiaceae*, *Artemisia*, *Cruciferae*, *Ephedra* 等。
- (5) *Artemisia*-*Chenopodiaceae*-*Gramineae*-*Myriophyllum* 优势带。其他成分还有 *Compositae*, *Polygonum*, *Ephedra*, *Rhamnus* 等。

从聚类分析的结果可以看出,带 1 和带 3,带 2 和带 5 反映了相似的孢粉组合。在纵向上可以相互对比,可能反映了相似的植被面貌和气候环境。

(二)分区

对海洋表层样品作区划,主要是为了探讨研究区的孢粉源、孢粉的搬运、沉积及其它影响海洋孢粉谱的因素。根据南中国海海域表层22个主要孢粉类: *Pinus*, *Podocarpus*, *Dacrydium*, *Elaeocarpus*, *Mallotus* + *Macaraga*, *Lithocarpus* + *Castanopsis*, *Eugenia* + *Syzygium*, *So-*

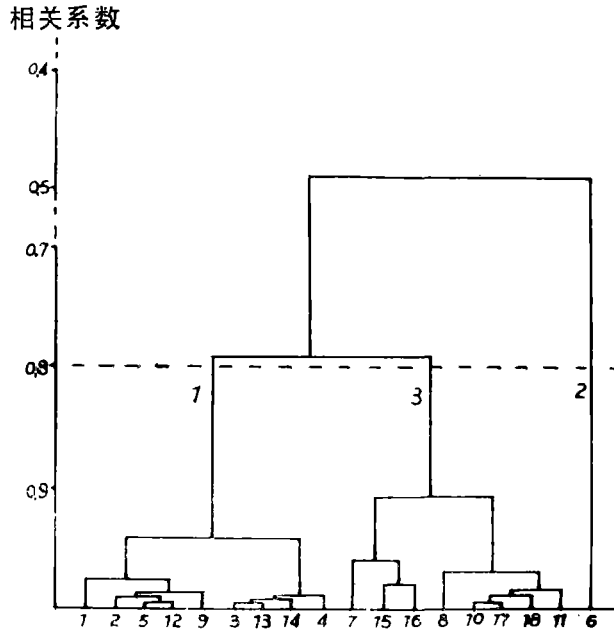


插图 4 内蒙古旱海 H₁ 孔聚类谱系图
Dendrogram of cluster for core H₁ from Hanhai Lake in Nei Mongol

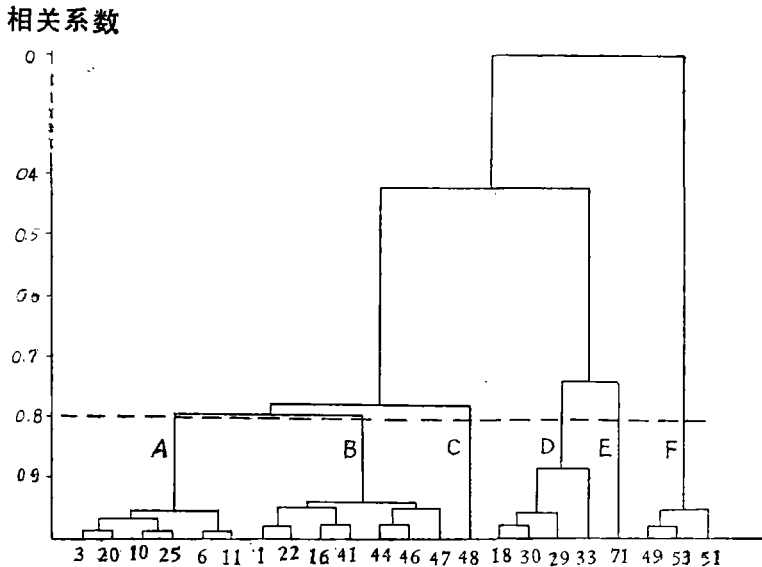


插图 5 南中国海海域表层样品聚类谱系图
Dendrogram of cluster for samples of surface beds in the South China Sea

nneratia, *Bruguiera* + *Rhizophora*, *Carallia* + *Ceriops*, *Cyclobalanopsis*, *Gramineae*, *Cyperaceae*, *Acrostichum*, *Angiopteris*, *Cyathea*, *Lycopodium*, *Dicranopteris* + *Microlepis*, *Polypodiaceae*, *Trees*, *Herbs*, *Pteridophyta*。对 22 个表层样品作聚类分析(表 IV;插图 5)。当阈值取 0.8 时,全部样品可分成 6 类,48 和 71 号样品单独成类,但从谱系图和聚类结果看, 48

表 IV 南中国海海域表层样品聚类结果
Result of cluster for samples of surface beds in the South China Sea

聚类步骤	聚类样品号		相关系数
1	25	10	0.993174637
2	11	6	0.992944641
3	20	3	0.990737207
4	22	1	0.987163126
5	53	49	0.983823322
6	46	44	0.980278801
7	41	16	0.980154266
8	30	18	0.979130844
9	10	3	0.972527597
10	29	18	0.965721014
11	6	3	0.957577026
12	16	1	0.954647314
13	51	49	0.949953775
14	47	44	0.948533279
15	44	1	0.939006333
16	33	18	0.879810478
17	3	1	0.792561034
18	48	1	0.779341603
19	71	18	0.739557765
20	18	1	0.416282286
21	49	1	0.0182596523

和 B 类、71 和 D 类较为接近(插图 5)。所以在根据聚类结果划区时，将 48 和 71 号分别归入由 B 类组成的 I 区和由 D 类组成的 III 区(表 V;插图 6)。

从分区结果看,导致各区孢粉组合特征产生明显差异的主要因素有： 由地理位置决定的孢粉源地、水深、离岸距离、海流以及沉积物特征等。

四、 讨 论

在对 ZK₁₅ 孔作数值划带时,除了作 Q 型系统聚类外,还进行了模糊聚类、对应分析,所得结果彼此间非常一致,和我们未作数值划带时所划的孢粉带也相符合。但需指出,在我们比较系统聚类划带和传统孢粉带划带结果中,也出现不一致的现象,造成这种不一致性是由于:

1. 在进行孢粉类型删选时,既选择了地区性花粉类型,也选择了地方性花粉类型,而传统孢粉带划分时,较注重地区性花粉类型,因而在地方性花粉类型含量较高时,会影响数值划带,导致两者的差异。

2. 传统孢粉带划分有时采用指示种法,主要基于某些花粉类型的首次出现或大规模扩张。而聚类分析划带主要考虑孢粉谱中大多数孢粉成分的相似性。从地层意义来说,前者是延续带,而后者则是组合带,因此,我们不能希望两者能达到一致。

比较聚类分析划带和传统孢粉带划分,聚类分析划带的优点是显而易见的:
1. 一个剖面的孢粉数据是庞大且复杂,人们在进行传统划带时,虽借助于孢粉图式,力图简化这种复杂性,但要抓住数据的结构要点还是相当困难的。应用系统聚类划带,可以利用计

表 V 表层样品区划及其特征
Division of surface samples and its characteristic

区	样 品										区划孢粉特征
I	样品号	1	16	22	41	44	46	47	48	<i>Dacrydium, Sonneratia</i> <i>Lithocarpus, Castanopsis</i> Cyperaceae, <i>Angiopteris</i> <i>Cyathea</i> , Polypodiaceae	
	水深(m)	1759	119	1110	1640	133	100	108	112		
	离岸距离(km)	123.0	84.0	252.0	387.0	276.0	192.0	243.0	291.0		
II	样品号	3	6	10	11	20	25			<i>Pinus, Dacrydium</i> Polypodiaceae, <i>Microlepia</i> <i>Lindsea, Dicranopteris</i>	
	水深(m)	2100	2830	2772	2452	2430	1759				
	离岸距离(km)	870	1350	1530	2190	1470	3210				
III	样品号	18	30	29	33	71				<i>Mallotus, Macaranga</i> <i>Bruguiera, Rhizophora</i> <i>Carallia, Ceriops</i> <i>Acrostichum, Lycopodium</i>	
	水深(m)	95	78	100	53	36					
	离岸距离(km)	450	1500	1800	780	720					
IV	样品号	49	51	53						<i>Pinus, Dacrydium</i> <i>Lithocarpus, Castanopsis</i> <i>Cyclobalanopsis</i> Cyperaceae, Polypodiaceae	
	水深(m)	110	145	1740							
	离岸距离(km)	3450	4080	4590							

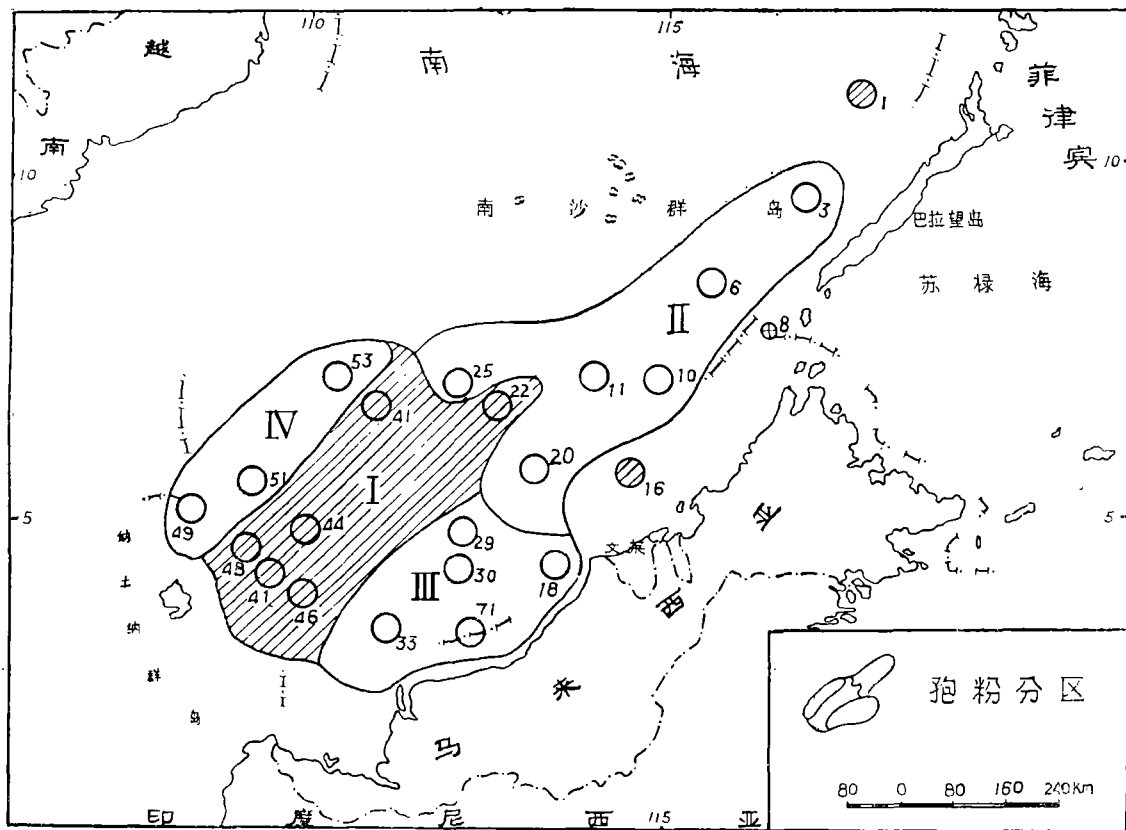


插图 6 南中国海海域表层样品区划

Distributional division of surface samples in the South China Sea

算机帮助我们较容易地权衡大量的信息,明确地认识数据的结构。同时,使用系统聚类划带还能提高我们的工作速度。

2. 应用聚类分析划带,使用的是一个数学尺度,可以避免研究者由于受沉积物岩性、推断的植被、气候等影响所产生的“先入之见”。且即使是不同的孢粉学者,使用同一数据,也会产生相同的结果。

3. 系统聚类有时还能指示上下孢粉带间的联系,便于研究者叙述和比较同一地层剖面的孢粉带,推断古植被和古气候。

但是,由于在阈值取不同值时,聚类分析会得到许多分类结论,而要确定哪一分类是最恰当的,显然存在着不可避免的人为性。因此对于聚类分析的客观性不能期望过高。但如果我们能同时使用多种数值方法和系统聚类一起对同一地层剖面作孢粉带划分,则会得到较为客观的结果。

参 考 文 献

王开发、卞玉清、张玉兰、蒋 辉、孙煜华,1982: 东海沉积物中孢粉组合的因子分析。海洋学报,4(5)。

王宗仁,1982: 地质数据的多变量统计分析。科学出版社。

罗积玉,1986: 微机用多元统计分析软件。四川科学技术出版社。

Boulter, M. C. and R. N. C. B. Hubbard, 1982: Objective paleoecological and biostratigraphic interpretation of Tertiary palynological data by multivariate statistical analysis. Palynology, 6: 55—68.

- Gordon, A. D., and H. J. B. Birks, 1974: Numerical methods in Quaternary paleoecology, 1: Zonation of pollen diagram. *New Phytol.*, 71: 961—979.
- Malmgren, B. A., and J. P. Kennett, 1973: Recent planktonic foraminiferal distribution in high latitudes of the South Pacific: A multivariate statistical study. *Palaeogeogr. Palaeoclimat. Palaeoecol.* 14: 127—139.
- Mello, J. F., and M. A. Buzas, 1968: An application of cluster analysis as a method of determining biofacies. *J. Paleontol.*, 42: 747—758.
- Oltz, D. F., 1971: Cluster analysis of Late Cretaceous-Early Tertiary pollen and spore data. *Micropaleontol.* 17: 221—232.
- Valentine, J. W., and R. G. Peddicord, 1967: Evaluation of fossil assemblages by cluster analysis. *J. Paleontol.* 41: 502—507.
- Webb, T., III, 1973: A comparison of modern and presettlement pollen from Southern Michigan (U. S. A.). *Rev. Palaeobot. Palynol.* 16: 137—156.

[1988 年 8 月 3 日收到]

APPLICATION OF CLUSTER ANALYSIS TO PALYNOLOGICAL ZONATION WITH EXAMPLES

Shen Cai-ming

(*South China Sea Institute of Oceanology, Academia Sinica*)

Tang Ling-yu

(*Nanjing Institute of Geology and Palaeontology, Academia Sinica*)

Summary

Based on the results of Q-system cluster analysis, two pollen stratigraphical sequences from the cores in the Zhujiang Delta and the Hanhai Lake in Nei Monggol are divided into zones, while the surface samples from the South China Sea are separately attributed to different distributive areas, with a summary of the basical procedure of the system cluster analysis for palynological zonation. The zones derived from the system cluster analysis correspond closely to the concept of an assemblage zone.

Using the system cluster analysis in palynological zonation can help us hold and balance large amounts of information, and gain insights into the structure of the data. In terms of mathematical criteria, this method avoids the possibility of bias on the part of an investigator; the similarity and difference among the zones in a sequence can be found by the system cluster analysis. However, due to the various decisions associated with the system cluster analysis with different threshold values, and the inevitable uncertainty about whether the decisions taken are the appropriate ones, it is suggested that one should not pitch too high any claim for the 'objectivity' of the system cluster analysis. But when several methods such as system cluster, fuzzy cluster and correspondence analysis are applied to the same data, similar results are obtained, thus providing the investigators with more confidence that the results suggested are not solely an artifact of the method employed.