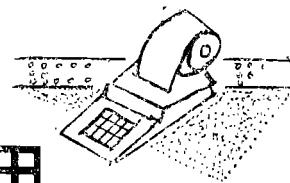
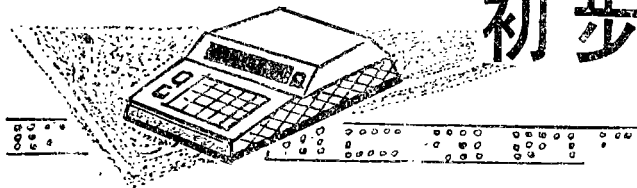


# 因子分析的初步应用



门桂珍

因子分析方法最早由 C. Spearman 提出, 是用来研究一组变量之间的相关性或研究相关矩阵内部结构的一种统计技术。1957 年由克隆拜因 (W. C. Krumbein) 引进地质学, 1962 年, 由英布里 (J. Imbir) 及普迪 (E. G. Purdy) 加以发展。

因子分析以许多变量的相关关系为基础, 将变量加以组合, 构成最少个数的独立的新变量——因子。从而表达变量的总变异, 达到简化变量揭示其变异的目的。因此, 它是研究系统分类与成因分类的有力手段, 其中 R 一型因子分析, 着重研究变量间的相互关系。这一方法应用于沉积学的研究中, 每一个因子可以代表物质来源、水动力条件、生物组合、元素共生组合、矿化强度、沉积环境等。

化, 定不出  $T_2$  与  $T_3$  的具体分界线。

(4) 鉴于其岩性单调, 即使以后将其时代厘定为  $T_{2-3}$ , 相信下杨子湖的沉积相乃至古地理面貌亦不致有太大的变化。

## 2. 下杨子湖的沉积相和海侵方向

以往认为黄马青组为内陆湖泊相, 但是其中含有的瓣鳃类、腕足类化石和几层含铜砂岩均说明它基本上属于滨海湖泊相。黄马青组的瓣鳃类化石, 已知的有 9 属 20 种, 其中有 4 属 8 种见于海相, 4 属 10 种可见于海

## 一、方法概述

假如我们分析了  $N$  个标本的  $n$  个指标 (本例中  $N = 24$ ,  $n = 9$ ), 则将这些数据列成一个数据矩阵 (表 1)。从表 1 中看到每个标本的各个分析数据或不同标本之间的同一指标之间相差无几, 特别是有些指标之间, 可能存在着成比例的增加或减小, 也即存在着一定程度的相关性, 这样, 这批数据提供的信息大部分是重迭的。我们研究问题, 不是直接用  $n$  个单个指标, 而是用它们的线性组合构成的少数综合指标 (因子) 来表达总的变异, 这些线性组合能使样品间的差异尽可能明显, 而且彼此是无关的, 这样, 就相当于把重迭的信息去掉, 减少变量个数, 即缩小空间维数。所以, R 一型因子分析能综合分

相和过渡相, 只有 1 属 2 种为淡水瓣鳃类。

既属滨海湖泊相, 就有海侵问题。下杨子湖虽地临东海, 但其中却混有一些常见于我国西南地区的瓣鳃类: *Myophoria* (*Costatoria*) *goldfussi mansuyi* Hsü, M. (C.) *submultistriata* Chen, M. (C.) *radiata* Loczy, *Eumorphotis* (*Asoella*) *cf. subillyrica* Hsü, 它们提示了下杨子湖与古地中海的某些联系。因此设想: 至少在晚三迭世早期, 下杨子湖可能经鄂东南的狭窄渠道与我国西南地区有过沟通。

原始数据矩阵

表1

变量 样品	有机总量	粘土岩	黄铁矿	碳酸盐	SiO <sub>2</sub>	Fe <sub>2</sub> O <sub>3</sub>	Al <sub>2</sub> O <sub>3</sub>	CaO	MgO
	1	2	3	4	5	6	7	8	9
1	86.760	12.250	0.000	0.400	46.200	5.260	34.550	3.580	1.300
2	92.520	7.320	0.000	0.160	47.520	3.680	37.700	1.950	0.590
3	96.760	3.070	0.170	0.000	47.860	7.820	36.770	1.380	0.630
4	85.560	13.030	0.940	0.470	45.040	7.440	36.060	3.430	0.650
5	87.750	10.260	1.230	0.760	36.220	24.290	29.250	3.230	1.150
6	75.690	24.060	0.000	2.500	47.400	3.760	40.000	0.970	1.150
7	99.210	0.630	0.160	0.000	86.820	8.870	31.750	2.820	0.960
8	84.500	14.460	1.040	0.000	46.940	14.590	37.350	2.200	0.890
9	94.140	5.860	0.000	0.000	48.660	8.410	38.420	0.510	0.670
10	90.500	6.720	2.780	0.000	35.180	1.580	30.110	0.510	0.590
11	72.050	26.490	1.460	0.000	49.040	5.190	39.280	0.050	0.740
12	98.100	1.710	0.000	0.000	54.220	5.790	32.040	1.330	0.630
13	96.350	3.130	0.260	0.000	44.760	36.540	36.200	0.670	0.700
14	98.480	0.000	1.080	0.440	28.600	17.670	24.370	3.370	5.560
15	88.510	2.540	4.330	0.290	49.120	25.260	18.490	2.460	1.070
16	79.750	19.630	0.000	0.000	53.980	8.270	24.660	4.250	2.370
17	94.540	0.760	2.050	0.310	35.640	34.810	11.180	6.500	0.780
18	91.510	7.880	0.610	0.000	46.060	8.420	38.710	0.820	0.740
19	90.890	3.040	0.230	1.170	62.480	5.110	25.730	0.720	0.700
20	69.790	24.130	5.040	0.000	47.560	10.560	37.990	0.100	1.040
21	78.470	20.600	0.750	0.000	46.960	4.360	38.420	1.280	1.190
22	40.710	29.280	3.580	0.000	61.840	5.190	27.530	0.010	0.850
23	73.030	24.090	2.880	0.000	47.020	8.720	39.420	0.100	0.780
24	86.350	4.990	8.120	0.000	33.480	40.150	13.980	43.00	0.630

注：1~4为煤岩鉴定数据，5~9为煤灰成分化学分析数据

析具有一定相关程度的多指标观测资料，能辨别出原始变量中独立的不相关的线性组合。

我们最终要获得这样一个新的变量

$$Z_j = \sum_{p=1}^m a_{jp} F_p \quad (1)$$

式中， $F_p$ 为公因子， $p$ 为公因子个数， $a_{jp}$ 为因子载荷（或称因子系数），表示第 $j$ 个变量在第 $p$ 个因子中所占比重大小，也就是表示公因子 $F_p$ 的相对重要性。我们的目的就是在 $n$ 个变量的相关矩阵基础上，提取出公因子个数 $m$ ，求出 $A = [a_{jp}]$ 的因子矩阵来。

计算步骤：

1. 将变量标准化，计算其相关矩阵：

$$\text{原始数据矩阵 } X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix}$$

$$\begin{pmatrix} j = 1, 2, \cdots, n \text{ 变量个数} \\ i = 1, 2, \cdots, N \text{ 样品个数} \end{pmatrix}$$

将变量标准化： $z_{ji} = (x_{ji} - \bar{x}_j) / \sigma_{x_j}$

$$\text{其中 } \bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ji} / N$$

$$\sigma_{x_j} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{ji} - \bar{x}_j)^2 / N}$$

计算相关系数，构成相关矩阵 $R = [r_{jk}]$ （表2），

$$r_{jk} = \frac{1}{N} \sum_{i=1}^N z_{ji} z_{ki} \quad (j, k = 1, 2, \cdots, n)$$

2. 用雅可比法求相关矩阵 $R$ 的特征值 $\lambda_p$ 及相应的特征向量 $\alpha_p$ ，（ $p = 1, 2, \cdots, n$ ）。

3. 根据一定原则选择公因子个数，其一是按特征值 $\lambda_p$ 大于1的个数来选定；其二是根据特征值的累积百分比大小来确定（一般百分比 $>80\%$ 即可）。本例中以前者为准，选出三个公因子。

4. 计算因子载荷，构成初始因子矩阵 $A$ （表3）。

九个变量间的相关系数矩阵 R

表 2

变量	1	2	3	4	5	6	7	8	9
1	1.00	-0.89**	-0.36	-0.03	0.14	0.26	-0.13	0.37	0.12
2		1.00	0.14	0.10	0.09	-0.44*	0.43*	-0.41*	-0.11
3			1.00	-0.23	-0.29	0.62**	-0.48*	0.07	-0.10
4				1.00	-0.05	-0.11	0.03	0.03	0.10
5					1.00	-0.38	0.20	-0.25	-0.28
6						1.00	0.77**	0.54**	0.05
7							1.00	-0.68**	-0.22
8								1.00	0.28
9									1.00

注：置信度 $\alpha=5\%$ ， $R_\alpha=0.4061$ ， $R_\alpha>0.4061$ 者划\*；置信度 $\alpha=1\%$ ， $R_\alpha=0.517$ ， $R_\alpha>0.517$ 者划\*\*。  
表示显著相关性，否则不显著相关，忽略不计。

$$a_{jp} = \alpha_{jp} \sqrt{\lambda_p} \quad (j=1, 2, \dots, n) \\ (p=1, 2, \dots, m, m < n)$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}$$

和，叫做公因子 $F_p$ 的方差，式如

$$\sum_{j=1}^n a_{jp}^2 = V_p$$

它相当于某一个公因子 $F_p$ 对所有公因子方差的贡献。显然 $V_p$ 越大，说明 $F_p$ 公因子越重要。

### 5. 计算正交因子解及斜交因子解

理论上已经证明，所求得的初始因子解不是唯一确定的。为了得到确定的因子解，可将初始因子轴进行一定角度的旋转，或者说将初始因子矩阵进行一个适当的线性变换，如果旋转后的两个因子轴仍彼此正交，则叫正交旋转，其结果叫正交因子解。如果旋转后的两个因子轴是斜交的，则为斜旋转，其结果为斜交因子解。通常认为，斜交因子解最能符合自然模型，因为因子轴之间或多或少会有些相关性。这样能使因子结构达到简单化，所谓简单结构是指因子载荷在同一因子轴上数值向两极分化，有尽可能多的载荷接近于0。

本例中以“方差极大法”准则求得正交因子解（表4），以“四次幂极小法”准则求得斜交因子解（表5~6）。其最终以斜交因子解作地质解释。其计算步骤从略。

初始因子矩阵 表 3

因子 变量	$F_1$ $\lambda_1 = 3.25$	$F_2$ $\lambda_2 = 1.95$	$F_3$ $\lambda_3 = 1.32$
1	0.5381	-0.7832	-0.1626
2	-0.6986	0.5874	0.2491
3	0.3696	0.8509	-0.0385
4	-0.1098	-0.1692	0.6021
5	-0.4422	-0.2071	0.2567
6	0.8638	0.3322	-0.2441
7	-0.8408	-0.2725	-0.0053
8	0.7728	-0.0939	0.1947
9	0.2820	-0.1579	0.7754

因子矩阵 A 中，每一行的因子载荷的平方和叫做各变量的公因子方差，式如

$$\sum_{j=1}^m a_{jp}^2 = h_j^2$$

它代表着所有公因子对第j个变量总方差中所占的比例， $h_j \leq 1$ 。

因子矩阵 A 中每一列的因子载荷的平方

正交旋转后的因子载荷矩阵 表4

因子 变量	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
1	0.0541	-0.9611	0.0522
2	-0.2988	0.8970	0.0349
3	0.7639	0.5172	-0.1054
4	-0.2207	0.0453	0.5937
5	-0.5003	0.1054	0.2071
6	0.9227	-0.2100	-0.1430
7	-0.8559	0.1939	-0.1052
8	0.5931	-0.4247	0.3342
9	0.1063	-0.1045	0.8267

斜交因子结构矩阵 表5

因子 变量	F <sub>1</sub> $\lambda_1 = 3.25$	F <sub>2</sub> $\lambda_2 = 1.95$	F <sub>3</sub> $\lambda_3 = 1.32$
1	-0.1222	-0.8850	0.0155
2	-0.1279	0.8345	0.2915
3	0.8431	0.5107	-0.0214
4	-0.1937	0.2191	0.5812
5	-0.4671	0.1291	0.1841
6	0.8645	-0.1752	-0.1029
7	-0.8081	0.0843	-0.1414
8	0.5131	-0.2471	0.3368
9	0.1060	0.1783	0.8217

斜交因子模型矩阵 表6

因子 变量	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
1 有机总量	0.1128	-1.0728**	0.4000
2 粘土	-0.3660	1.0462**	-0.2975
3 黄铁矿	0.7611**	0.4208*	-0.2624
4 碳酸盐类	-0.2642	0.0538	0.5854**
5 SiO <sub>2</sub>	-0.5341**	0.2038	0.1549
6 Fe <sub>2</sub> O <sub>3</sub>	0.9724**	-0.4097*	-0.0342
7 Al <sub>2</sub> O <sub>3</sub>	-0.8884**	0.3955	-0.2125
8 CaO	0.6189**	-0.6118**	0.5200**
9 MgO	0.0690	-0.1939	0.8921**

注:  $\alpha = 1\%$       $R_\alpha = 0.517$   
 $\alpha = 5\%$       $R_\alpha = 0.406$

二、实际应用

利用湘中测水煤系三、五煤的部分分析测试数据,作R-型因子分析。经过计

算,选出三个大于1的特征值即三个因子,它们说明总方差的73%,也即说明了变量的大部分。从这里可以看到因子分析的一个作用,是从原始的9个变量缩减到3个相互独立的新变量,达到了降低空间维数的目的。其次,我们再来探讨一下三个因子的主要地质含义。

因子相关矩阵 表7

因子	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
F <sub>1</sub>	1.00	0.255	0.097
F <sub>2</sub>		1.00	0.398
F <sub>3</sub>			1.00

总观表4~6,可以发现初始因子解、正交因子解、斜交因子解的结果都极为近似,说明这三个因子极为稳定的代表着变量的总体,我们仅采用斜交因子解来说明。从表7因子相关矩阵来看, $r_{F_1F_2} = 0.255$ , $r_{F_3F_1} = 0.097$ , $r_{F_2F_3} = 0.3975$ ,这三个相关系数都小于置信度为5%的临界值

斜交因子模型简化矩阵 表8

因子 变量	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
有机总量	<div>+++</div>	<div>---</div>	<div>+++</div>
CaO			
Fe <sub>2</sub> O <sub>3</sub>			
黄铁矿		<div>++</div>	
粘土		<div>+++</div>	
碳酸盐			<div>+++</div>
MgO			
SiO <sub>2</sub>	<div>---</div>		
Al <sub>2</sub> O <sub>3</sub>	<div>---</div>		

注: 负载荷值     正载荷值  
---     +++     表示载荷值 > 0.517者  
- - -     ++     表示载荷值 > 0.406者

$R_a = 0.406$ , 因此为显著不相关, 这进一步说明这三个因子是相互独立的, 效果是最佳的, 用它来推断解释是信得过的。

从表 8 和图 1 ~ 2 可以看出,  $F_1$  主要代表黄铁矿 (0.7611)、 $Fe_2O_3$  (0.9724)、 $CaO$  (0.6189) 和  $Al_2O_3$  (-0.8884)、 $SiO_2$  (-0.5341) 等一组数据, 这一组合反映了第一因子主要由黄铁矿、方解石和石英等组成。其中煤化学分析中的  $Fe_2O_3$ , Fe 元素既有二价的, 也有三价的, 只能作为辅助参数使用。与  $F_1$  正轴方向相斥的是  $Al_2O_3$  与  $SiO_2$ , 它们既可以以硅铝酸盐的形式存

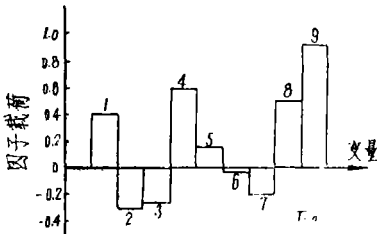
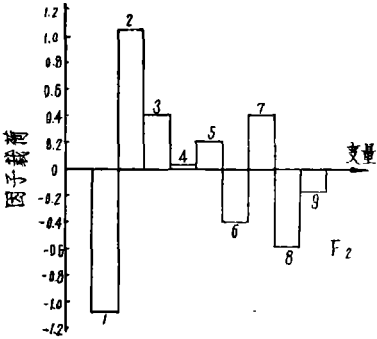
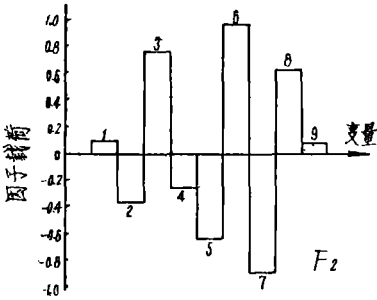


图1.  $F_1$ 、 $F_2$ 、 $F_3$  因子的变量组合图

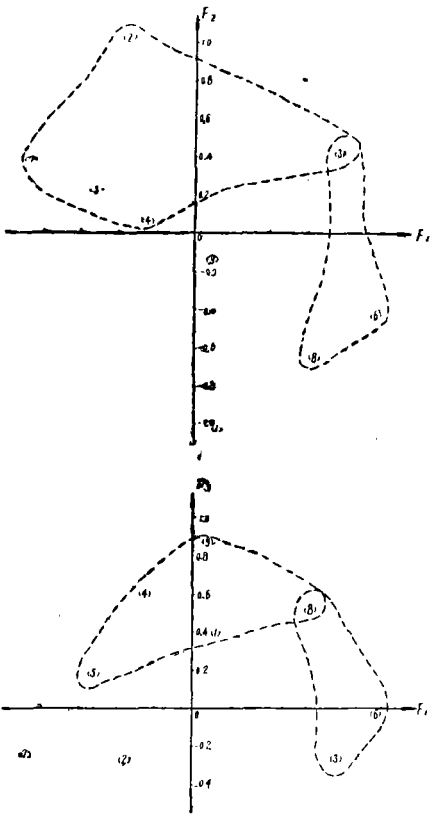


图2. 变量在因子平面上的聚类图

在, 也可以  $SiO_2$  单独呈石英颗粒形式存在, 正负方向是一个共同的组合, 之所以对立, 可以解释为成煤时期同沉积阶段的成因不同或物质来源不同所致。对比一下三五煤的矿物组成 (表 9), 可以看出, 这一因子完全与五煤的矿物成分相当。因此, 第一因子可以是五煤层的强还原环境的成因因子。

三、五煤层矿物成分表\* 表 9

煤层 矿物	三 煤	五 煤
主要的	高岭石	石 英
次要的	伊利石 黄铁矿	高岭石 黄铁矿
少量的	方解石、白云石、石英、绿泥石、锰菱铁矿	方解石、白云石、菱铁矿、蒙脱石、叶腊石、绿泥石

\* 本表摘自煤炭科学院地质勘探研究所湘中测水课题研究报告。

# 早古生代腐泥无烟煤及炭质岩中 菌藻类化石简易分离方法

朱 丽 英

1972年,我们在对于早古生代煤的成因和性质的研究工作中,在显微镜下(光片和薄片),发现了微古菌藻类化石,并见到了藻植体向煤基质过渡的中间类型。随后,经过努力进一步攻克了复型关,使之能在电子显微镜下观察(照片1、2)。

然而,由于镜下观察受切片方向的限制,很难看到全面的、完整的藻体形态。因此,必须设法从煤或顶底板岩石中把古藻类的遗体分离出来。但是由于煤和岩石的变质,从高于焦煤变质阶段的腐植煤中分离出孢子尚且不易作到,要想从变质程度更高的早古生代腐泥无烟煤中把藻类分离出来就更加困难。以往国内外也未有过成功经验的报导。1978年,笔者经过反复试验,终于找到了从上述煤与顶底板岩石中分离出藻类化石的方法。所分离出的藻类原植体绝大多数色

彩鲜明,结构、形态清晰。

这一分离方法的成功,使得古藻类得以进一步研究鉴定,从而为研究生物与煤及伴生矿产的成因关系提供了可能性,也必将促进早古生代生物学、地层学的研究工作。现将该方法介绍于后。

## 1. 方法原理

从早古生代煤和顶底板岩石中分离出古藻类化石方法的实质在于,通过化学处理使腐泥基质和矿物杂质溶解除掉,使较稳定的古藻类遗体保留,并进一步分离出来。

早古生代煤及顶底板岩石中的矿物质主要为二氧化硅和碳酸盐矿物,因此,可以用盐酸、氢氟酸煮沸溶解除去。腐泥基质可以用浓硝酸和浓硫酸混合液加热氧化除去,为使藻类细胞壁不致因强酸作用发生收缩,所

$F_2$ 代表粘土类(1.0462)、黄铁矿(0.4208)、 $Al_2O_3$ (0.3955)、及有机总量(-1.0728)、 $CaO$ (-0.6118)、 $Fe_2O_3$ (-0.4097)等一组数据。可以看出,第一二因子都强烈地受黄铁矿的影响,说明成煤环境中的还原条件是普遍的控制因素。但第二因子中更重要的是以高粘土类为特征,而粘土类与有机总量处于相斥状态,这是容易理解的。 $CaO$ 、 $Fe_2O_3$ 等还可以以方解石、菱铁矿等形式存在,如果对比一下表9,则不难发现,这一因子反映了三煤层的矿物成份。因此,可以认为第二因子是三煤的成因

因子,而同时有向 $F_2$ 负轴方向煤质变好的趋势。

$F_3$ 代表碳酸盐类(0.5854)、 $CaO$ (0.5200)、 $MgO$ (0.8921)等一组数据。很明显,这一组合是由方解石、白云石构成的碳酸盐类沉积,它代表着浅海相的沉积环境,与成煤作用毫不相干。因此,第三因子可称作不成煤因子。

从 $F_1$ 、 $F_2$ 、 $F_3$ 三个因子的先后排列次序来看,它反映了三、五煤层成煤环境在垂向上的递变规律。第一因子是五煤的成因因子,五煤是测水早期形成的可采煤层,当

以加入适量的15~25%浓度的冰醋酸。

## 2. 分离步骤

1) 样品洗净、凉干并破碎过筛(孔径为0.25~0.5毫米,视藻类化石大小而定),将筛下的粉样称取20~50克。(以下步骤均须在通风橱内进行)。

2) 按 $\text{HNO}_3:\text{H}_2\text{SO}_4:\text{CH}_3\text{COOH}=4:1:1$ (体积比)配方配好氧化混合液。

3) 将样品置入烧杯中,加入1:1的盐酸水溶液加热煮沸,除去全部碳酸盐矿物,余者再加清水,用离心试管离心数次洗净。

4) 将洗净的样品置于塑料坩锅内,小心缓缓加入氢氟酸并加热煮沸,使全部二氧化硅转变为硅的氟化物,一般加热到在坩锅底部留下粥状物为好,再用10%盐酸水溶液加热煮沸除去氟化硅,最后离心洗净。

经上述步骤,样品中绝大部分矿物杂质已被除掉。

5) 再将样品置入烧杯中,并加入浓硝酸、浓硫酸和冰醋酸氧化混合液,使液面稍高于样品介面,用玻璃棒搅拌均匀后在电热板上加热。据样品变质程度不同,加热温度控制在60°C至混合液沸点之间。加热过程

中随着氧化剂的消耗,要不断加入新鲜的氧化混合液。氧化过程中要不断取样在镜下观察,当分离出的藻类形态结构清楚时,氧化即可停止。最后加水稀释,离心洗净。

6) 将氧化后洗净的样品,干燥脱水,加入甘油,以防止藻类化石的干裂和破坏。

7) 除样品的一部分留作制片外,其他大部分倾入有塞子的小玻璃瓶中,加进一小块百里酚,以防止霉菌生长。再用溶化的石蜡封住瓶口,样品可以长期保存。

## 3. 制片

将在镜下观察到的藻类化石,用蘸有植物胶的解剖针尖粘起,挑置在载薄片的植物胶上,然后微微加热,使其上的植物胶溶化,再盖上盖薄片。用蘸有酒精的纱布清除片上多余的植物胶后,贴上标签,即制成薄片。

所分离出的古藻类原植体,一般呈深浅不同的褐棕、棕褐、棕红、黄褐和褐灰色,形态多种多样,已经南京大学朱浩然教授审定。(照片3~6)。

工作中得到南京大学朱浩然教授和我所廖克光、马晓原等同志热情帮助,在此致谢。

时地势凹凸不平,聚煤盆地面积局限,水动力条件比较活跃,携带的碎屑物多为石英或其他粗碎屑物,或经短距离搬运,沉积在泥炭沼泽中,形成了较稳定的局部可采煤层;第二因子是三煤成因因子,从三煤开始,聚煤盆地相对扩展,水动力条件比较平稳,携带物质经长距离搬运,进入沼泽中则多为细粒粘土类及胶体形式的沉积物,因而形成了稳定的可采煤层。就沉积环境来讲可以说从第一因子的海湾泻湖相过渡到第二因子的滨海泻湖相成煤环境,而至第三因子则为浅海相沉积,就不利于成煤了。

上述分析,我们可以看到,R一型因子分析着重考虑了九个变量之间的内在联系,研究其共生组合,发掘其组合的内在因素,探讨煤的沉积环境,是成因分类的一种行之有效的办法,但应指出的是,由于本例所使用的变量不很全面,仅仅用了煤岩鉴定及化学分析的部分数据,没有将煤质分析数据充分利用,这样必将对成果解释带来一定的局限性与片面性,值得今后注意。同时这是一次初步尝试,错误之处,诚望读者批评指正。