植物群落数量分类的研究

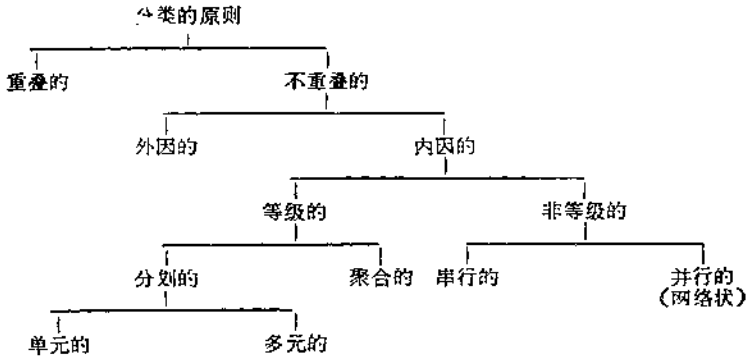
一、关联分析和主分量分析

阳含熙　　卢泽愚　杨周南
（中国科学院综合考察委员会）　（北京市计算中心）

一、前　言

植物群落数量分类的研究是从五十年代开始的，到六十年代电子计算机普遍应用，生物学与地学很多分支都采用了数量分类以后，才迅速地向前发展。许多原来具有不同传统分类观点的学派（如法瑞学派、英美学派、苏联学派等）都进行了数量分类方法的研究，并用这些数量方法去验证他们原来采用分类的结果。现在植物生态学文献中，每年都发表大量数量分类的文章，不断涌现新的方法和技术。

植物群落的数量分类，无论采用那种方法，所遵循的原则大体如下图所示：



迄今为止，植物群落分类都采用不重叠的原则，即一个实体只能属于一类，而不能同时属于两类以上，非此即彼，不能含糊。近年来，很多学者认为，在自然界内，中间类型是很多的，人为地加以割裂会损失许多信息。有些实体只能说在多大程度上属于一类，而又在多大程度上属于另一类，亦此亦彼，界限是模糊的。如在植物分类学中，Bor(1960)研究印缅禾本科分类就将9个种同时放在两个不同的属：芒属（Erianthus）和甘蔗属（Saccharum）之中。数学方面出现模糊数学（Fuzzy mathematics）这一新的领域，就与重叠分类的原则有关。

内因和外因的原则，传统植物群落分类就已存在，植物群落分类用生境因素（如气候、土壤等因素）的很多，但多数人是着重内因的。

等级分类的原则是低级分类从属于高一级的分类，组内的成员要求尽可能相似，而不同组的成员则尽可能不相似。这种等级的分类用数学的术语来表示，需满足如下两个条件：

1. 对任意两个下级组 A_i 和 A_j，必存在一个上级组 A_k，满足 $(A_i \subset A_k) \wedge (A_j \subset A_k)$。

2. 若 A_i 从属于两个不同的上级组 A_j 和 A_k，则 A_j 和 A_k 也必有上下级的关系，即：

$$(A_i \subset A_k) \wedge (A_i \subset A_j) \rightarrow (A_j \subset A_k) \vee (A_k \subset A_j)。$$

等级分类的原则用得最多，大家最熟悉、也最简单，许多传统的植被分类系统都采用这一原则。这种分类最后都是用树状图（dendrogram）表示的；而非等级分类则是用串行或并行的图式表示，如群落的星云状图就是。

分划的原则是从总体的一组实体出发，逐步细分下去；而聚合的原则是从每个实体出发，逐步聚合最后并成整个总体。分划的根据是相异系数，聚合的根据是相似系数。分划可以人为地在任一水平下停止，而聚合则必须全部完成才能停止。因此，聚合的计算量大，如 N 个实体就需要计算 $(N-1)^2$ 次。聚合还有一个困难，在合并过程的初期很可能会把一部分实体放错了位置，而对以后的工作带来麻烦。分划还可以对每个实体进行不同处理以反映它们在分析过程中不同的重要性，也就是"内部加权"的问题，而聚合的方法则不能这样做。

单元分类的原则是根据单项属性而分成有无此属性的两类。它的优点是简单、明确，计算较快，而缺点是如采用了一个无关紧要的属性，就会造成无意义的分划。多元分类是根据全部或多数属性，因而信息多、结果稳定，但计算量大。

单元、聚合的分类很少有人做过，多元、分划的分类计算量太大，至今只有少数尝试。因此，目前绝大多数分类的方法都是单元、分划的，或者是多元、聚合的。

对于植物群落的数据还可采用另一种分析途径，就是以一种或多种属性数据为依据，在属性空间（并非通常的地理位置空间）中排列出实体。这种方法叫做排序（ordination）。我们可以根据外因（如生境因素）、内因（植被的数量特征）或样方的距离来排序。从 1957 年开始，这方面已有大量研究。

本文对两种原始数据应用一个等级分类的方法——关联分析，和一个排序的方法——主分量分析。为清楚起见，先简明地介绍方法，然后讨论分析的结果。

我们采用的原始资料，一种是福建三明光叶红椎（*Castanopsis kawakamii* Hayata）林的调查结果（阳含熙等，1978），立木株数达全林株数的 40.1%，胸高断面积达全林断面积的 62.2%，从传统的分类来看是很均匀的单一群落。另一种是内蒙古呼伦贝尔羊草（*Aneurolepidium chinensis*（Trins.）kitagawa）草原的调查结果（李博等，1978）。原作者将它分为三个群丛组（半湿润草甸草原，半干旱干草原和盐化草原），以及这三者之间的一些过渡类型。我们有意用这两个均匀度相差很大的数据来检验数量分类方法的效果。

二、方　法

（一）关联分析（Association analysis）

这是一种单元、分划的群落等级分类的方法，即根据一个属性（例如种）将整个样方集合逐次细分而得到不同样方组的方法。它的基本原理是在一个均匀群落中，不同样方中的种是不关联的，因而我们最终将样方分成这样的组，组内各样方中的种间不存在显著的关联，每个组就可以算做一个群落。

这个方法在植物群落分类方面的应用，最早是 Goodall（1953）研究澳州的桉树灌木林。Williams 和 Lambert（1959, 1960）加以进一步改进，并在电子计算机上应用。嗣后，

这种方法在热带稀树草原 (Kershaw, 1968)、热带雨林 (Austin 和 Greig-Smith, 1968)、温带草地 (Gittins, 1965)、温带灌丛 (Harrison, 1970; Goldsmith, 1973) 广泛应用并与其它方法作过比较研究。Proctor (1967) 还用此方法将英伦三岛的苔类分类。一般认为这种方法野外收集数据简便,宜于做初步概查;它可以客观地分成类型,并在野外具体找到,而进行植被制图;进一步又可粗略地与生境因素进行比较分析。在应用过程中也发现一些问题: Hopkins (1968) 指出关联分析每次只用一个指标,对于草本群落有时只反映季相变化; Noy-meir et al.(1970) 指出种数太少时可能会夸大偶见种的重要性。所以 Goldsmith (1973) 建议淘汰罕见种和含种数过少的样方,以免引起错误的分划。Kershaw (1968) 建议存在度在 98% 以上和 2% 以下的种都不计算,而 John et al. (1977) 则建议存在度在 95% 以上和 5% 以下的种均淘汰。

Williams 和 Lambert(1960) 把用种来区分样地的方法称为正分析(normal analysis);他们(1961)又用样方去区分种组,并称为逆分析 (inverse analysis)。在多变量分析的书籍中正分析和逆分析又被称为 R 分析和 Q 分析 (Cattell, 1952)。正分析的结果可以与法瑞学派的群丛来比较, Beeftink(1972) 用此方法整理了世界盐性沼泽 (Saltmarsh) 的数据。逆分析可以用来和生态种组比较 (Ellenberg, 1956)。

假若我们在抽样植被时,对 N 个样方分别记录了 p 个种的量,得到如下形式的观察数据:

$$\begin{array}{cccccccc} & \text{样方} & 1 & 2 & 3 & \cdots & N \\ \text{种} & 1 & x_{11} & x_{12} & x_{13} & \cdots & x_{1N} \\ & 2 & x_{21} & x_{22} & x_{23} & \cdots & x_{2N} \\ & 3 & x_{31} & x_{32} & x_{33} & \cdots & x_{3N} \\ & \vdots & & \cdots\cdots\cdots & & & \\ & p & x_{p1} & x_{p2} & x_{p3} & \cdots & x_{pN} \end{array} \qquad (1)$$

其中 $X_{ij}(i=1, 2, \cdots, p; j=1, 2, \cdots, N)$ 代表第 j 个样方中第 i 种的量。一般关联分析都用某个种存在(记为1)和不存在(记为0)的二元数据。

从原始数据(1)出发,关联分析(正分析)的步骤如下:

1. 计算关联矩阵 $\boldsymbol{V}$: 首先对 N 个样方的集合,计算 p 个种中两两之间的关联指标,它有各种不同的定义,这里我们采用的指标 $V_{ij}(i, j=1, 2, \cdots, p)$ 是这样计算的:

对 N 个样方,列出种 i 和种 j 的 2 × 2 列联表,

样方数	有种 i(i)	无种 i(i)	合计
有种 i(j)	a	b	a + b
无种 i(j)	c	d	c + d
合计	a + c	b + d	N = a + b + c + d

则

$$V_{ij}=\chi^2_{修正}(i, j)/N=\frac{(|ad-bc|-N/2)^2}{(a+b)(c+d)(a+c)(b+d)}(i, j=1, 2, \cdots, p; i\neq j)。$$

一个种自身的关联指标 V_{ii}, 我们规定为0,就得如下形式的 $p \times p$ 关联矩阵:

$$V=(V_{ij})=\begin{pmatrix}0 & V_{12} & V_{13}\cdots V_{1p}\\ V_{21} & 0 & V_{23}\cdots V_{2p}\\ V_{31} & V_{32} & 0 \cdots V_{3p}\\ \cdots & \cdots & \\ V_{p1} & V_{p2} & V_{p3}\cdots 0\end{pmatrix} \tag{2}$$

这个矩阵是对称的，$V_{ij}=V_{ji}$ $(i\neq j)$，同时主对角线上元素都为 0，因此总共 $p\times p$ 个元素中只有 $p(p-1)/2$ 个需要计算。另外，若 $\chi^2_{种i,种j}$ 的值小于某显著性水平下自由度为 1 的 χ^2 值，或者种 i 或种 j 在所有样方中的数据全为 0 或全为 1，则相应的 V_{ij} 都定义为 0。显著性水平可以人为地取不同值，如 0.10，0.05，0.02，0.01 等，显著性水平太高将损失一部分信息，太低则无必要，会增加工作量，一般多用 0.05。

2. 决定临界种并据此进行一次分划：在算出的关联矩阵 $\boldsymbol{V}$ 中，对 p 个列（或行）求和，得到：

$$V_{\cdot j}=\sum_{i=1}^{p}V_{ij}\ (j=1,2,\cdots,p)$$

并从中决定使 p 个和达到最大值的种，比如 j_0，即：

$$V_{\cdot j_0}=\max(V_{\cdot 1},V_{\cdot 2},\cdots,V_{\cdot p})$$

我们称种 j_0 为临界种。在样本集合中，它比所有别的种表现出最大的关联。我们可把它当做样本集合分划的标准属性，而将 N 个样方分成含有种 j_0 的 (J_0) 组和不含种 j_0 的 (j_0) 组。令 (J_0) 组有 N_1 个样方，(j_0) 组有 N_2 个样方，显然 $N_1+N_2=N$。

3. 再分划：对上述分划得到的两个样方组，分别重复上述两步进行再分组，这样反复分划下去直到分得的每个子组其内部的种间关联 V_{ij} 全部为 0（即均低于显著性水平）时为止。这就将原 N 个样方分成了种间均不关联的样方组，达到了分成若干个均匀群落的目的。

4. 结果表示：最后将上述逐次进行的组分划过程用树状图表示出来（如图 1 和图 2），其中每个节点表示一个中间组分成两个子组，该节点的水平高度（纵坐标）表示此次分划所依据的 χ^2/N 值。

（二）主分量分析（Principal component analysis，简作 PCA）

它是一种群落的排序方法。对每个样方，以其中 p 个种的量做为坐标，就可看成 p 维空间中的一个点。N 个样方就构成 p 维空间中的 N 个点。主分量分析就是要在较低维的空间中，特别是在直观的二、三维空间中排列出这 N 个点，而尽量少损失一些信息，即尽量保持它在原 p 维空间中的重要特性。

Goodall（1954）第一次将 PCA 用于植物群落的排序研究。因为 PCA 有比较严格的数学基础，在电子计算机应用普遍之后，它就很快得到广泛的应用，并经常用来与其它方法进行比较（Jeffers 1972; Kershaw 1973; Chapman 1976; Gauch Jr. et al. 1977）。

大量的研究证明 PCA 是一个非常有效的方法，最先三个主分量通常可反映原来数据的方差达 40—90%。但是也有两个不足之处：一是当数据的方差非常接近，相关又很小时，就找不出明显的的主分量；二是当数据的分布不是超椭球体，而是诸如马蹄形等非线性情况，则无法应用 PCA。对于非线性数据，解决的办法是缩小数据范围使数据在小范围内大致是线性的，或者进行平方根变换（Kershaw，1968）。

PCA 可作为一种统计方法。对于随机抽样、正态分布的属性进行排序，找出不同属性的输入量（Loading），然后对输入量大的属性与实体作单回归分析，其效果与计算步骤要比逐步回归好而简便。

假若我们已有形如(1)的抽样数据，它表示在每个种为坐标轴(分别记为 $x_1, x_2, \cdots, x_p$) 的 p 维空间中的 N 个样本点。本文所用的 PCA 步骤如下：

1. 调整坐标原点：对(1)的原始数据每一行(种)求平均值：

$$\bar{x}_i = \frac{1}{N}\sum_{j=1}^{N} x_{ij} (i = 1, 2, \cdots, p)$$

如果 p 个平均值不全为 0，表示 N 个样本点对坐标原点是偏离的，首先需将坐标原点平移到形心 $(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_p)$，即将(1)的每一行 N 个数据分别减去该行平均值，得到如下数据矩阵：

$$\boldsymbol{X} = \begin{pmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_1 & x_{13}-\bar{x}_1 & \cdots & x_{1N}-\bar{x}_1 \\ x_{21}-\bar{x}_2 & x_{22}-\bar{x}_2 & x_{23}-\bar{x}_2 & \cdots & x_{2N}-\bar{x}_2 \\ x_{31}-\bar{x}_3 & x_{32}-\bar{x}_3 & x_{33}-\bar{x}_3 & \cdots & x_{3N}-\bar{x}_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{p1}-\bar{x}_p & x_{p2}-\bar{x}_p & x_{p3}-\bar{x}_p & \cdots & x_{pN}-\bar{x}_p \end{pmatrix} \quad (3)$$

下面将 X 矩阵的数据当做 N 个样本点的坐标进行分析。

2. 计算离差指标的矩阵 S：我们用 S_{ij} 表示第 i 种与第 j 种的离差指标，

$$S_{ij} = \sum_{k=1}^{N}(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j) \quad (i, j = 1, 2, \cdots, p)$$

当 $i = j$ 时，它是 N 个样本值对 x_i 轴的离差平方和，有 $S_{ii} = N V_{ar}(x_i)$（我们是将 N 个样本值当总体看待）；当 $i \neq j$ 时，它是对 x_i 和 x_j 两个轴的交叉积之和，有 $S_{ij} = N \mathrm{Cov}(x_i, x_j)$。显然 $S_{ij} = S_{ji}$，于是得到一个对称的离差指标矩阵：

$$\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}' = \begin{pmatrix} S_{11} & S_{12} & S_{13} & \cdots & S_{1p} \\ S_{21} & S_{22} & S_{23} & \cdots & S_{2p} \\ S_{31} & S_{32} & S_{33} & \cdots & S_{3p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ S_{p1} & S_{p2} & S_{p3} & \cdots & S_{pp} \end{pmatrix} \quad (4)$$

其中 X' 是 X 的转置矩阵。

PCA 除了用 $\boldsymbol{S}$ 矩阵以外，也可用方差—协方差矩阵 $\left(\frac{1}{n}\right)\boldsymbol{S}$，它们都是没有标准化的原始数据；还可用相关矩阵 $\boldsymbol{R}$，就相当于标准化了原始数据。无论用什么形式的矩阵，分析方法都是相同的。

3. 求出坐标刚性旋转的矩阵 $\boldsymbol{U}$：现在要找出原坐标轴的一个刚性旋转，即原变量 x 的线性组合：

$$y_{ij} = \sum_{k=1}^{p} u_{ik}(x_{kj} - \bar{x}_k) \qquad (i = 1, 2, \cdots, p;\ j = 1, 2, \cdots, N)$$

其中 y_{ij} 是第 j 个样本点在坐标旋转后新坐标轴 y_i 上的坐标。或者写成矩阵的形式：

$$\boldsymbol{Y} = (y_{ij}) = \boldsymbol{U}\boldsymbol{X} \quad (5)$$

这里

$$\boldsymbol{U}=\begin{pmatrix} u_{11} & u_{12} & u_{13}\cdots u_{1p} \\ u_{21} & u_{22} & u_{23}\cdots u_{2p} \\ u_{31} & u_{32} & u_{33}\cdots u_{3p} \\ \cdots & \cdots & \cdots \\ u_{p1} & u_{p2} & u_{p3}\cdots u_{pp} \end{pmatrix} \tag{6}$$

是在直角坐标系下刚性旋转的变换矩阵，因而它是酉矩阵，满足于 $\boldsymbol{U}'=\boldsymbol{U}^{-1}$（$\boldsymbol{U}$ 的逆矩阵）。

我们要求这样的刚性旋转：变换后的 N 个点在新的第一坐标轴 y_1 上有最大的离差平方和，在第二轴 y_2 上有次大的离差平方和，……，在第 p 轴有最小的离差平方和，即有 $N\,\mathrm{Var}(y_1)\geqslant N\,\mathrm{Var}(y_2)\geqslant\cdots\geqslant N\,\mathrm{Var}(y_p)$，我们称这 p 个轴分别为第一主分量、第二主分量、……、和第 p 主分量。同时，N 个点对 y_i 和 y_j $(i\neq j)$ 两个轴的交叉积之和均为 0。也就是说，要求

$$\boldsymbol{YY}'=\begin{pmatrix} N\mathrm{Var}(y_1) & 0 & \cdots & 0 \\ 0 & N\,\mathrm{Var}(y_2)\cdots\cdots & & \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & & N\mathrm{Var}(y_p) \end{pmatrix}=\boldsymbol{\Lambda} \tag{7}$$

其中 Λ 是对角线矩阵。对于这样的旋转就保证了将 N 个点投影到较低维空间中去，将保留最多的信息，即发生的畸变最小。

现在的问题归结到求出满足 (7) 的变换矩阵 $\boldsymbol{U}$。这是容易的，因为 S 是对称的，故存在着酉矩阵 U 将 S 变换成对角线矩阵 Λ，即：

$$\boldsymbol{\Lambda}=\boldsymbol{USU}'=\boldsymbol{UXX}'\boldsymbol{U}'=\boldsymbol{YY}'$$

或者，

$$\boldsymbol{US}=\boldsymbol{\Lambda U}$$

这说明 Λ 矩阵的 p 个对角线元素正是 S 的 p 个按大小次序的特征根：λ_1，λ_2，$\cdots$，λ_p；而 U 的每一行向量就是相应的特征向量。因此，求出了 $\boldsymbol{S}$ 的特征值和特征向量就得到了 $\boldsymbol{\Lambda}$ 及变换矩阵 $\boldsymbol{U}$。

4. 排列 N 个样方：知道了变换矩阵 $\boldsymbol{U}$，由(5)求出 $\boldsymbol{Y}$，就得到 N 个点在新坐标系下的坐标。我们希望选取较少的主分量个数，在较低维空间中排出 N 个样方。如果只选前 $k(<p)$ 个主分量，而忽略后面 $p-k$ 个分量，那末上述 Y 的前 k 行就是 N 个样方在这 k 个轴上的坐标，它保留原 p 维空间的信息的百分比为：

$$\sum_{i=1}^{k}\lambda_i\Big/\sum_{i=1}^{p}\lambda_i$$

选择多少轴需权衡简单与准确两个方面。因为 λ_i 值是依大小次序排列的，取前二、三轴就可能占总信息的 40% 以上。选取二、三轴就能在平面上或立体中画出 N 个样方的排序位置，如图 3 和图 4 是二维排序的图象表示。

5. 种对主分量的输入量 (Loading)：上述 $k(<p)$ 个主分量虽然在较低维的空间中排序了样方，但由 $\boldsymbol{Y}=\boldsymbol{UX}$ 可知，每个主分量都是 p 个种的线性组合，不能解释单个种对主分量，从而对整个排序所起的作用。

因为，$\boldsymbol{\Lambda}=\boldsymbol{UXX}'\boldsymbol{U}'=\boldsymbol{UXY}'$，所以 $\boldsymbol{XY\Lambda}^{-\frac{1}{2}}=\boldsymbol{U}'\boldsymbol{V}^{\frac{1}{2}}$，我们令：

$$L = (l_{ij}) = U'\Lambda^{\frac{1}{2}} = \begin{pmatrix} \sqrt{\lambda_1}u_{11} & \sqrt{\lambda_2}u_{21}\cdots\sqrt{\lambda_p}u_{p1} \\ \sqrt{\lambda_1}u_{12} & \sqrt{\lambda_2}u_{22}\cdots\sqrt{\lambda_p}u_{p2} \\ \cdots & \cdots \quad \cdots \\ \sqrt{\lambda_1}u_{1p} & \sqrt{\lambda_2}u_{2p}\cdots\sqrt{\lambda_p}u_{pp} \end{pmatrix} \tag{8}$$

由于 $Y'\Lambda^{-\frac{1}{2}}$ 实际上是标准化的主分量，可见 L 是原始的种向量（X_i）与标准化主分量的离差矩阵。如果 X 也是标准化的（即用相关矩阵 R 做主分量分析），那末 L 就是种与主分量间的相关系数矩阵。因此 l_{ij} 的符号及数据大小反映了第 i 种对第 j 主分量的相关正负及作用大小，一般称 L 为种对主分量的输入量矩阵。

我们可对选择的前 k 个主分量，将 L 的前 k 列数值按种次序和分量次序列出表格形式（如表 1 和表 2）。不仅表中 l_{ij} 值反映了 i 种对 j 分量的作用，而且还可看出每列 l_{ij} 的平方和：

$$\sum_{i=1}^{p} l_{ij}^2 = \sum_{i=1}^{p} \lambda_j u_{ji}^2 = \lambda_j \quad (j = 1, 2, \cdots, k)$$

也就是说，每个主分量的各输入量之平方和正好是该分量的方差（特征值）。同时表中元素每行（种）的平方和：

$$h_i^2 = \sum_{j=1}^{k} l_{ij}^2 = \sum_{j=1}^{k} \lambda_j u_{ji}^2, \quad (i = 1, 2, \cdots, p)$$

反映了第 i 种对所考虑 k 个主分量的作用大小。

三、结果和讨论

光叶红椎林的关联分析结果如图 1 所示。在采用 5% 的临界值水平时，仅能按甜槠的出现与否分划为两组。这一结果和我们过去对光叶红椎和甜槠的关联系数研究是符合的。光叶红椎和甜槠的 λ^2 值是 4.469，在 $p = 0.05$ 水平上显著，这两个种的出现是相关的（阳含熙等，1978）。从野外观察来看，甜槠多数出现在山顶、风势较大、土层浅石砾多的立地。如果用 1% 的水平，则一次也不能分划，这正与光叶红椎基本上是一片纯林的结论一致，它不能显著地再分成异质的群落。

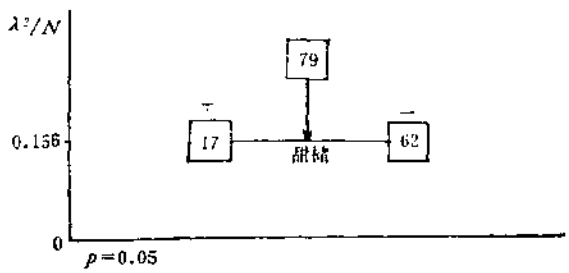


图 1　福建三明光叶红椎林的关联分析分类（11 种 79 个样方）

羊草草原的关联分析结果如图 2，显然可见分划的等级比光叶红椎林多得多，在 $p = 0.05$ 水平下分划的次数竟有 9 次之多，异质情况是很明显的，羊草草原 40 个样方记录了 32 种，原作者用传统方法将样方分成群丛组：第 1—10 个样方是半湿润草甸草原群丛组，第 16—25 个样方是半干旱草原群丛组，11—15 个样方是两者的过渡类型，第 31—40 个样方是耐盐植物的群丛组，第 26—30 个样方是后两个群丛组的过渡类型。

图 2 中就 1% 水平而言，大致可分成四个组：①组 12 个样方是以含有日阴菅种为依据的，它包括了原作者半湿润性草甸草原群丛组的全部 10 个样方和两个过渡类型的样

方;②组 8 个样方是由不含日阴菅的样方中以包含碱蒿种分出来的,它们全属于原耐盐植物群丛组,原组中的另外两个样方被分到了④组;③组的 10 个样方中有 8 个样方属于原半干旱群丛组,另两个样方属过渡类型,原半干旱组中的另两个样方也分到了④组;④组共 7 个样方,除上述混入的 4 个样方外,其余 3 个是过渡类型。分划中还有几个零星样方均属过渡类型。

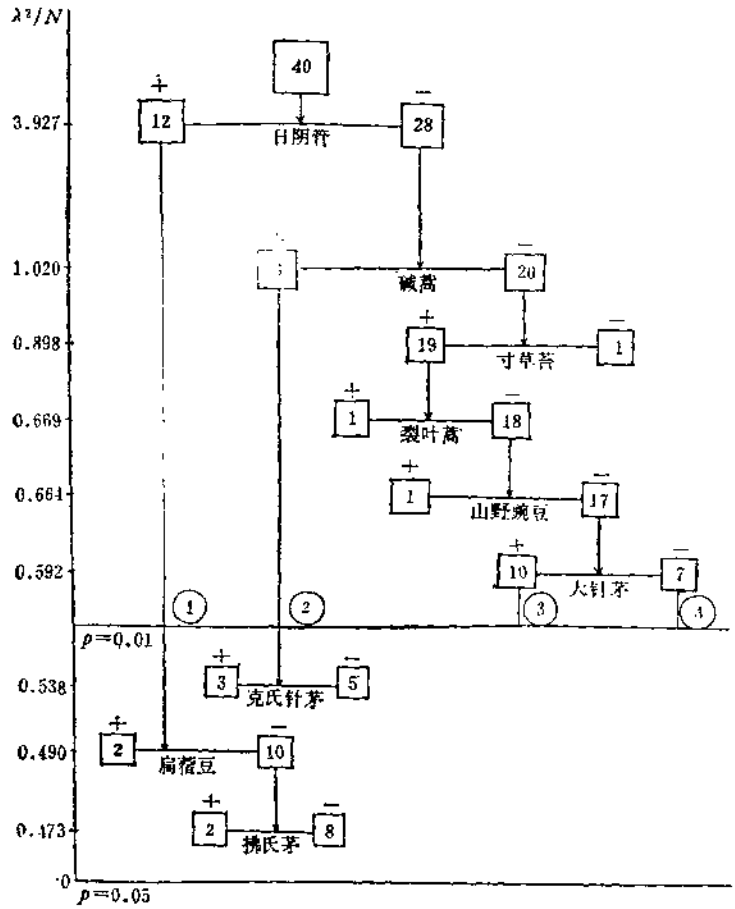


图 2　内蒙古呼盟羊草草原的关联分析分类(40 个样方 32 个种)

由此可知,关联分析分类的结果与原来定性分析分类的结果基本上是吻合的,少数样方的例外分划可能是由于过渡样方或个别偶见种所致。倘若用 0.05 显著性水平,将引起进一步的分划,目前还没有野外观察和其它分析来验证其生态意义。

对光叶红椎林进行主分量分析,图 3 表出了 79 个样方在前两个主分量上的二维排序。图中可见样方是相当均匀地排列的,不形成任何较明显的集团。

表 1 是 11 个种对前三个主分量的输入量表。从中可以看出毛相桐、齿草树、中华鼠刺和木荷对这三个主分量的作用较大。三维主分量只占全部信息的 44.4%,二维只占 31.3%,从计算机算出的结果可知,考虑了前七维才占总信息的 81.1%。在只考虑 11 个种的十一维空间中得到这样的排序效率是相当低的,加之,个别种对前几维排序的作用不

是特别突出，二、三维的排序看不出明显的样方集团，这些结果都进一步印证了光叶红椎林是相当均匀的一片纯林。

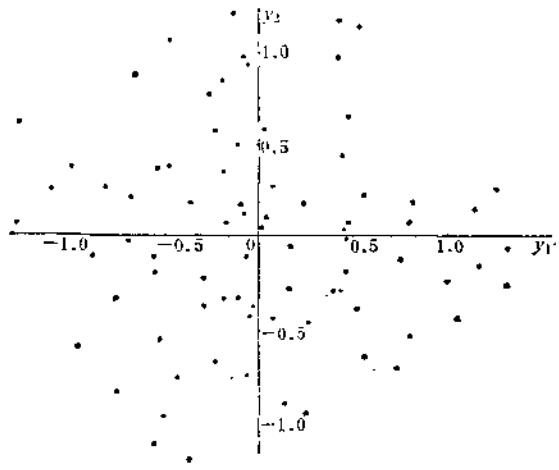


图3　光叶红椎林79个样方 *PCA* 的二维排序(占总信息的31.3%)

对羊草草原主分量分析的一个二维排序图形如图4所示。图中40个样方明显地分成三个集团，与原作者的三个群丛组及关联分析的分组都非常吻合。右边用虚线框出的11个样方，包括了半湿润草甸草原群丛组中全部1—10个样方，另外还包括了本来属于干草原与盐化草原过渡型的第26号样方，这是因为日阴菅偶然出现的缘故。在关联分析中，第①组也同样包含了该样方，这是一个颇有意趣的现象。左上方框出的8个样方均属于半干旱草原群丛组，原组中第20和22两个样方稍远离这一样方集团，而混杂于其它的过渡样方中。左下方框出的样方，包括了耐盐植物群丛组的10个样方。其余没有框出的样方均属于三个集团之间的过渡类型。

表 1　　**光叶红椎林 PCA 的输入量表**

种	第一主分量	第二主分量	第三主分量	h^2
卡氏乌饭 *Vaccinium carlesii*	−2.13	1.23	−0.64	6.46
毛 柏 树 *Adinandra millettii*	1.75	−2.91	0.75	12.12
木　　荷 *Schima superba*	2.83	1.03	−0.56	9.37
西 草 树 *Randia densiflora*	−0.30	−1.10	3.23	11.72
中华鼠刺 *Itea chinensis*	2.23	1.94	1.76	11.82
中华杜英 *Eaeocarpus chinensis*	−1.87	−1.42	0.65	5.94
豹 皮 樟 *Litsea chinensis*	−0.37	1.60	0.59	3.05
甜　　槠 *Castanopsis eyrei*	−1.31	1.85	0.38	5.28
钩　　骨 *Tricalysia uvidiflora*	−1.49	0.64	2.12	7.12
马 尾 松 *Pinus massoniana*	−0.80	−0.92	−1.86	4.94
栲　　*Castanopsis hystrix*	1.50	0.49	−0.31	2.58
特征值 (λ)	31.16	25.47	23.77	80.40
占总信息百分比	17.20	14.10	13.10	44.40

表2给出了32个种中输入量较大的13个种对前三个主分量的输入量表，其余次要的种只算总和值而不一一列出。从表中可见，日阴菅是对前三个主分量作用最大的种。在我们共考虑32个种的32维空间中，前三个主分量占了总信息的50.7%，比起光叶红椎林来排序效率要高得多，也反映羊草草原的40个样方具有较强的异质性。

另外，从排序图4还可看出：半湿润草甸草原群丛组全在第一主轴的右边，而半干旱草原群丛组和耐盐植物群丛组几乎全在左边，说明它们间的分划主要是第一主分量的作

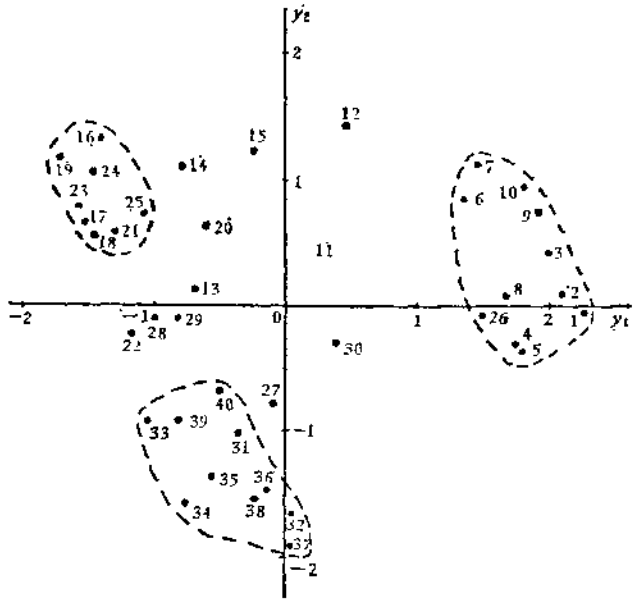


图 4　羊草草原 40 个样方 PCA 的二维排序(占总信息的 44.3%)

表 2　　羊草草原 PCA 的输入量表

种	第一主分量	第二主分量	第三主分量	h^2
贝加尔针茅 *Stipa baicalensis*	2.17	1.36	0.01	6.56
大针茅 *S. grandis*	−1.66	0.93	−0.84	4.33
糙隐子草 *Cleistogenes squarrosa*	−1.92	1.43	1.12	6.99
日阴菅 *Carex pediformis*	2.60	0.80	−0.09	7.41
寸草苔 *C. duriuscula*	−2.24	−0.05	−0.13	5.04
裂叶蒿 *Artemisia laciniata*	2.31	0.42	−0.30	5.61
山野豌豆 *Vicia lamoena*	1.89	0.66	−0.08	4.02
细叶白头翁 *Pursatilla turczaninovii*	2.24	0.84	0.04	5.73
展枝唐松草 *Thalictrum squarrosum*	2.27	−0.39	−0.05	5.30
冷蒿 *Artemisia frigida*	−1.65	0.90	0.74	4.08
阿尔太狗娃花 *Heteropoppus altaicus*	−1.69	1.52	0.77	5.76
柴胡 *Bupleurum scorzonerifolium*	0.33	1.90	0.50	3.97
碱蒿 *Artemisiaanethifolia*	−0.42	−1.81	0.80	4.10
⋮	⋮	⋮	⋮	⋮
特征值 (λ)	61.85	34.35	14.10	110.30
占总信息百分比	28.4	15.9	6.4	50.7

用;由表 2 可知对第一主分量作用最大的种是日阴菅(输入量为 2.60),正与关联分析中相应分划的临界种是日阴菅 (看图 2) 一致。同时,半干旱草原群丛组与耐盐植物群丛组一个在第二主轴上半部,一个在下半部,它们的分划主要依据第二主分量 (看图 4); 由表 2 知对第二主分量成正相关的是柴胡 (输入量为 1.90),而最大负相关的是碱蒿 (输入量为 −1.81),这两个种通常不同时出现,碱蒿只在盐化生境出现,而柴胡则在半干旱组出现。这在第二轴样方的排列清楚地表现出来,同时与关联分析中相应分划的临界种碱蒿也比

较一致。由此可知，判别半湿润草甸草原群丛组的一个重要依据是日阴菅，判别耐盐植物群丛组的重要依据是碱蒿。

综上所述，对于两个均匀程度相差悬殊的植物群落，两种不同方法所得结果是一致的，与传统定性分类的结果也是基本吻合的。过去有些学者认为植被的连续或不连续问题与研究的方法有关；排序容易得出连续的结果，分类则容易得到不连续的结果。我们这次研究再次说明这种意见是不妥当的，植被形成连续或不连续并不是方法引起的假象。

本文所用的两种数量分类方法，都是在电子计算机上实现的，应用的程序保存在北京市计算中心，可供其他同志使用。

参 考 文 献

[1] Austin, M. P. and P. Greig-smith, 1968: The application of quantitative methods to vegetation survey II: Some methodological problems of data from rain forest, J. Ecol. 56(3): 827—844.

[2] Beeftink, W. G., 1972: Übersicht über die Anzahl der Aumnahmen Europäischer und Nordafrikanischr Salzpflanzengesellschaften für das Projekt der Arbeitgruppe für Datenverarbeitung, Grundfragen und Methoden in der Pflanzensoziologie: 371—396.

[3] Bor, N. L., 1966: Grasses of Burma, Ceylon, India and Pakistan.

[4] Cattell, R. B., 1952: Factor analysis: An introduction to essentials Biometries, 21: 190—251.

[5] Chapman, S. B. ed., 1976: Methods in plant ecology.

[6] Ellenberg, H., 1956: Aufgaben und Methoden der Vegetationskunde.

[7] Gauch. L. G. Jr. et al., 1977: Comparative study of reciprocal averaging and other ordination techniques. J. Ecol. 65: 157—174.

[8] Gittins, R., 1965: Multivariate approaches to a limestone grassland community: I. A stand ordination, J. Ecol. 53: 385—401.

[9] Goldsmith, F. B., 1973: The vegetation on exposed sea cliffs at South Stank, Anglessey I. The multivariate approach, J. Ecol. 61(3): 755—818.

[10] Goodall, D. W., 1953: Objective methods for the classification of vegetation: I. The use of positive interspecific correlation. Aust. J. Bot. 1: 39—63.

[11] Goodall, D. W., 1954: Objective methods for the classification of vegetation: III. An essay in the use of factor analysis, Aust. J. Bot 2: 304—324.

[12] Harrison, C. M., 1970: The phytosociology of certain English heathland communities, J. Ecol. 58: 573—589.

[13] Hopkins, B., 1958: Vegetation of the Olokemeji forest reserve, Nigeria. V. The vegetation on the svanna site with special refercnee to its seasonal changes, J. Ecol. 56: 97—115.

[14] Jeffers, J. N. R., 1972: Plotting of multidimensional data, Merlewood Research and Development, pap. 35.

[15] John, D. M. et al., 1977: A quantitative study of the structure and dynamics of benthic subtidal algal vegetation in Ghana (tropical West Africa), J. Ecol. 65. 497—521.

[16] Kershaw. K. A., 1968: Classification and ordination of Nigerian savanna vegetation, J. Ecol. 56: 467—482.

[17] Kershaw, K. A., 1973: Quantitative and dynamic plant ecology.

[18] Noy-meir, I., 1970: Association analysis of desert vegetation, Israel J. Bot. 19: 561—597.

[19] Noy-meir, I., and Austin, M. P., 1970: Principal component ordination and simulated vegeaztion data, Ecology, 61, 551—552.

[20] Proctor, M. C. F., 1967: The distribution of British liverworts —— a statistical analysis, J. Ecol. 55(1): 119—136.

[21] Williams, W. T., and J. M. Lambert, 1959: Multivariate methods in plant ecology: in plant communities. J. Ecol. 47. 83—101.

[22] Williams, W. T., and J. M. Lambert, 1960: Multivariate methods in plant ecology: II. The use of an electronic digital computer of association analysis, J. Ecol. 48. 689—710.

NUMERICAL CLASSIFICATION OF PLANT COMMUNITIES
I. ASSOCIATION ANALYSIS AND PRINCIPAL COMPONENT ANALYSIS

Yang Han-xi
(The integrated survey Commission of natural resources, Academia Sinica)

Lu Ze-yu Yang Zhou-nan
(Computer Centre, Peking)

Summary

First of all, various approaches of classification and ordination are briefly treated. Monotheic and divisive association analysis (AA) is used for classification and more sophisficated principle component analysis (PCA) for ordination in this first paper. Computational procedures are stepwise presented in detail and computer programmes are written and deposited in the Computer Centre, Peking. Two sets of binary data are used: 79 quadrats with 11 tree species of a *Castanopsis kawakamii* forest in Central Fukien and 40 quadrats with 32 species of *Aneurolipidium chinese* grassland in Inner Mongolia. Results of AA and PCA agree closely to each other and also to the qualitative classification of the original authors. *Aneurolepidium chinese* grassland is shown clearly to have 4 discontinuous groups and its 2 and 3 dimensional ordination by PCA can retain 43.7% and 50.2% of the total information content respectively. It is, therefore, postulated that AA and PCA are complementary and very useful methods to a wide variety of ecological context, including interrelation between vegetational and enviromental attributes.