

# 用聚类分析划分京郊农业生态系统类型\*

顾 武

(中国农业科学院作物育种栽培研究所, 北京)

## 摘 要

本文从农业生态系统原理出发选择57个适合于北京郊区农业生态系统分类的特征指标,以村作为类型分类的基本单元,对北京郊区有关县区的77个大队进行不重叠的、内在的等级聚合分类(系统聚类),采用了五种聚类分析方法。结果表明,离差平方和法分类效果较好。除最短距离方法外其它方法均可明显分出低集约化山区、中集约化平原、高度集约化近郊和特殊类型四大类型,阈值更小的划分出现相应的一些亚类型。研究农业生态系统类型将为不同类型的农业生态系统采取相应的控制和管理措施提供依据,从而达到扬长避短,因地制宜地分类指导生产,使农业生态系统获得较高的生态经济效益。

**关键词:** 聚类分析, 北京郊区, 农业生态系统。

## 一、研究地区、指标和方法

### 1. 研究地区范围及其特点

北京地区主要有西部山地、北部山地和东南部平原三大地貌单元。山地约占全市面积的61.29%、平原约占38.71%。研究范围主要在围绕市区的大兴、房山、昌平和顺义县境内,以市区为中心,市内三环路为起点向外延伸约62.5公里,绝大多数是在50公里以内。在此范围内选择一定数量的分类单元。由于北部山地、西部山地和东南部平原在离三环路50公里范围内都有一定面积,所以整个研究结果能概括京郊所有农业生态系统类型。

### 2. 分类单元的确定

由于本文涉及范围有限,不宜建立完整的分类等级系统,故仅对农业生态系统村级分类单元进行一些探讨。在国内,对农业生态系统的调控是以行政单元进行的,目前大队成为农村基本核算及管理单位。一般,平原区大队土地总面积在1000—10000亩之间,耕地由几百亩到几千亩,山区大队总面积多至二三万亩。以大队作为分类单元是可行的,有利于直接将类型研究的结果付之于管理和技术的实施。

在昌平、大兴、顺义、房山县及部分农场境内选择了77个大队作为分类基本单元。昌平山区南口镇居庸关大队土地面积最大,达36748亩,双桥农场八里桥大队面积仅有694.6亩。以最畅通的交通距离计算,房山县三合庄大队距三环路85公里,而最近的东风农场六里屯大队距三环路仅2公里(图1)。具代表性单元的选择上着重考虑:尽可能选择不同类型的典型地段,使既有代表性又有差异性<sup>[1]</sup>。

### 3. 指标体系的选择及其依据

农业生态系统是一个复合生态经济系统。选择指标体系应反映农业生产整个过程和纵

\* 本文是在孙鸿良导师指导下硕士论文的一部分,也得到李博先生的热情指导,特此致谢。  
本文于1987年2月24日收到。

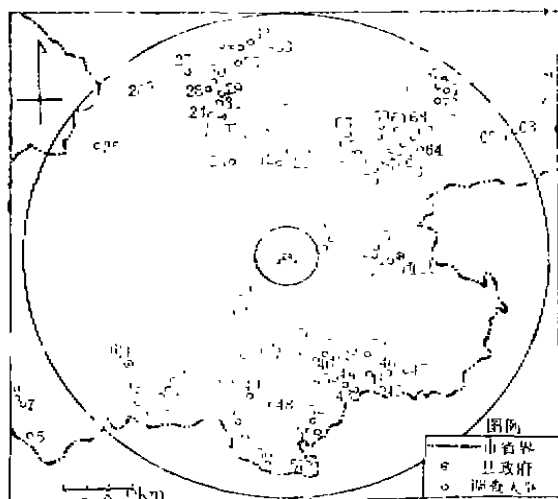


图 1 研究范围和分类单元的分布

Fig. 1 The studying region and distribution of classification units

以上是选取指标的理论依据。实际选取指标时还要考虑分类方法对指标的要求和取得指标数据的可能性。(1) 该指标必须对农业生态系统产生一定影响起到一定的作用。(2) 该指标在选点范围内有明显的差异。(3) 适当舍去与主要指标相关性很大的次要指标。

据以上原则确定出57个划分村级农业生态系统类型的指标体系, 见表1。

表 1 农业生态系统划分的指标体系

Table 1 An attribute system of the village agroecosystematic classification

| 指 标           | 指 标            | 指 标              |
|---------------|----------------|------------------|
| 海拔高度 (m)      | 蔬菜耕地比          | 林业产出比            |
| 土壤类型          | 复种指数           | 农田钾素平衡 (千克/公顷)   |
| 土壤有机质含量 (%)   | 经作耕地比          | 农田腐殖质消长 (千克/公顷)  |
| 土壤全氮含量 (%)    | 家畜饲养量 (标准羊/公顷) | 土地面积总投入 (元/公顷)   |
| 土壤碱解氮含量 (ppm) | 农机总动力 (马力/公顷)  | 农用地农林牧渔收入 (元/公顷) |
| 土壤速效磷含量 (ppm) | 劳力转移程度         | 农业收入占五业比         |
| 土壤速效钾含量 (ppm) | 农业劳力比例         | 林业收入占五业比         |
| 人口密度 (人/亩)    | 人均果树 (株/人)     | 畜牧收入占五业比         |
| 水分充足程度        | 种植多样性 (种类)     | 副业收入占五业比         |
| 有效灌溉率         | 作物亩产           | 农用地纯收入 (元/公顷)    |
| 土地农用地率        | 系统生产力          | 农业收入占四业比         |
| 人均农业用地 (亩/人)  | 种植业产出比         | 林业收入占四业比         |
| 年有效积温 (℃)     | 畜牧产出           | 畜牧业收入占四业比        |
| 农田面积比         | 投能密度           | 蔬菜收入占四业比         |
| 林木覆盖率 (%)     | 无机能投入          | 劳均收入 (元/人)       |
| 离市内距离 (千米)    | 食物能产出比         | 人均收入 (元/人)       |
| 果林面积比         | 输出系统外能量比       | 人均农产品占有量 (万大卡/人) |
| 粮作播面比         | 粮食输出 (千克/公顷)   |                  |
| 水稻播面比         | 农田氮素平衡 (千克/公顷) |                  |
| 小麦播面比         | 农田磷素平衡 (千克/公顷) |                  |

观农业生产的总体。农业生产的自然条件在指标体系中占有一定的地位, 特别对于人口少、经济力量薄弱的地区自然条件对农业生产起决定性作用。土地利用方式与程度、农田格局、林木和作物品种、食物链结构、工业比例等构成了农业生态系统不同的结构形式并产生不同功能和效果。农业生态系统同时又是一个有人干预的经济系统, 通过以经济手段干预自然生产过程, 获得一定的社会经济效果。

农业生态系统各分类单元间所发生的差别必然反映在上述几方面相应的指标上, 聚类分析能根据各种指标数值差异程度大小的综合结果客观地划分出类型。

#### 4. 数据取得

原始数据主要来源于三个方面,一是农业区划资料;二是实地调查,特别是一些生态学参数和实地情况了解;三是各类统计报表。统计资料多数选用 1983 年农村社队年末统计报表,1983 年属略丰水的正常年份,数据能够反映一般年份的当地生产特点,可以作为基础资料来划分农业生态系统类型。

#### 5. 计算方法及步骤

77 个大队的分类单元均对应 57 个综合指标构成  $77 \times 57$  的分类矩阵。首先对原始数据施行属性标准化变换:

$$X_{ij} = \frac{Y_{ij} - \bar{Y}_j}{S_j} \quad \left( \begin{array}{l} i=1, 2, \dots, 77 \text{ 分类单元序号} \\ j=1, 2, \dots, 57 \text{ 指标序号} \end{array} \right)$$

式中  $Y_{ij}$  表示原始数据矩阵中第  $i$  个分类单元第  $j$  个属性 (属性即指标) 的数值,  $\bar{Y}_j$  和  $S_j$  分别表示第  $j$  个属性的平均值和离差。

用标准化后的  $X$  矩阵计算分类单元之间的相似系数,选用欧氏距离的平方,即:

$$D_{ij} = \sum_{k=1}^n (X_{ik} - X_{jk})^2 \quad (i, j=1, 2, \dots, 77)$$

本文定义分类类群之间的相似系数用最短距离法,最长距离法,中线法,类平均法;和离差平方和法<sup>[2]</sup>。画出离差平方和法及最长距离法的树状图。

确定分类等级的划分界限是根据 Demirmen (1972) 提出的分组原则<sup>[3]</sup>,用徐克学等结合线方法<sup>[4]</sup>。为了增加五种方法比较时选择相同划分类型的 机会,选用四处可能出现的有意义的跳跃部分作为阈值。

## 二、结果与分析

### 1. 分类结果

五种等级聚合方法对 77 个分类单元以 57 个指标聚类分析结果见树状图 (图 2)。仅将离差平方和法的结合线表示出来 (图 3)。

### 2. 结合线跃变特点

在 60 次结合以前,在五种方法中结合线都未发生大的飞跃。类间距离相差不大,70 次结合以后类间距离急剧加大,可选择的划分等级几乎全部在 70 次结合以后的结合中,那么至多只有 10 种左右的类型可供划分。5 种方法均可找到 4 个或 4 个以上的阈值,较大阈值上划分的大类群差别是十分明显的而类内差别较小,这是较为清晰的结果。所用距离公式是欧氏距离平方加之用离差标准化增强了这种差异程度,到后期一单差值增大,平方后数值显著增大。

### 3. 群化倾向

聚类开始六次结合都是单元与单元的合并,这时聚类与聚合策略无关而五种方法效果相同。聚类始于顺义县 (71) 蒋各庄和 (70) 陈各庄,接着是顺义县 (65) 东马各庄和 (66) 张喜庄,房山县 (2) 七里店和 (3) 下坡店等,初始阶段就明显出现了相邻分类单元首先聚在一起的所谓“群化”倾向,与樱井明久数值区划结果十分相似<sup>[5]</sup>。初始聚类的单元都是平原以农为主的大队,近郊或在县城附近的大队在后期才能聚合在一起,反映了近郊农业生产的复杂和多变性,农区的生产相对稳定和单一。

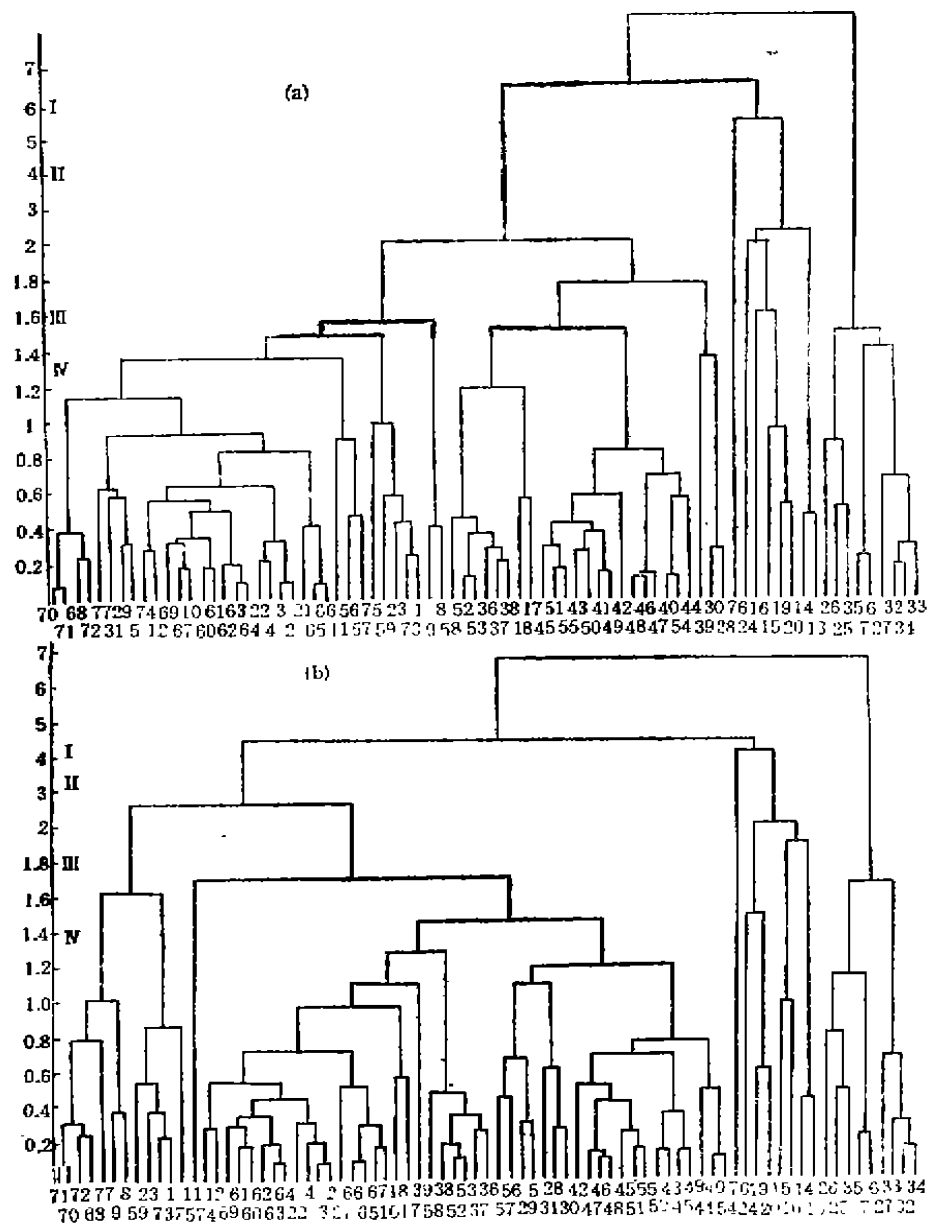


图 2 聚类分析树状图

Fig. 2 The dendrogram of Cluster analysis

(a) 离差平方和法 I = 5.93, II = 3.87, III = 1.61, IV = 1.38

(b) 最长距离法 I = 4.2, II = 3.3, III = 1.78, IV = 1.37

#### 4. 类型划分

类平均法、最长距离法、离差平方和法在第一层次（阈值较大）的划分中得到明显差异的三大类群：（1）山区果林生产类型，（2）平原农业生产类型（主要经营种植业和养殖业）和（3）近郊高度集约化的蔬菜生产类型。

降低划分层次除最短距离外其它四种方法同时分出了四种类型，即在近郊蔬菜生产类

型中又分出(76)顺义胜利大队,该大队情况比较特殊,据调查,该队已基本脱离农业生产,工副业收入较高。1983年耕地又被国家征用,目前仅利用307亩耕地,人均0.4亩。从征用地中得到额外收入783000元计入其它收入项中,使收入结构、土地收入、人均收入等指标与其它类型数值发生极大分异。所以胜利大队可看作为一类受城市强烈影响,农业生态系统结构近于崩溃的特殊系统。因此较大阈值的划分中分为四大类型较为合适。

以上四类中各类型所含分类单元除类平均法的划分结果将近郊菜区中的芦沟桥农场(13)齐庄大队和(14)老庄大队归在平原农业类型所分结果完全相同。服从多数方法划分结果可将这两单元放入菜区类型中,同时也说明了该两个大队有别于其它大队。

进一步更细致的亚类型划分,在不同方法中差别较大。近郊菜区虽然只有7个大队,低阈值划分中出现较多的亚类型,表现了近郊农业生产极大的波动性。平原农业类型聚类距离数值较小,单元间性质和特征的变异远不及前一类型。近郊分类单元属性易出现极大与极小数值而使平原区中的变异的重要指标可区分的信息量减少和分辨率降低,但较小阈值上仍能划分亚类型。以最长距离法和离差平方和法结果比较一致,在Ⅲ、Ⅳ层次划分中已经出现农区亚类型,比其它方法有更好的效果而能使平原类型划分为多组有意义的类型。(59)顺义白各庄、(73)河南村、(23)昌平踩河、(1)房山窦店和(75)顺义前进是一类工副发展规模较大较好的农区类型,是京郊闻名的典型;(45)、(42)等是平原农林果并举的类型;(74)、(69)等为典型的种植业为主的农业生产类型,后两者包含单元在两个方法中互有交叉,反映类型特征有共同之处。山区类型的再划分结果比较分散,实际上山区农业生产由于自然条件限制,其发展类型相对稳定和单一,划分依据中社会性因素如人均耕地、人口密度、商品生产等以人或土地为单位则可能是发生变异的主要原因。

低层次(阈值较小)划分中的亚类型多数都兼有上层同类的特征,所以划分的类型既有差别又有共性。分类层次较低时相互间联系与过渡更为明显。

### 5. 不同方法的评价

不同方法中某些分类单元归宿有争议说明了分类方法尚未统一。五种方法除最短距离法外其它四种方法都能分出有意义的类型。最短距离法由于有相当强的空间压缩性使分类单元有明显链接聚合的趋势,该方法不适于农业生态系统的聚类。中心法聚合过程具有非单调性,逆转现象严重,结果不太理想。离差平方和法效果最好。

## 三、小结与讨论

1. 五种等级聚合方法中离差平方和法效果最好,是村级农业生态系统多指标数值分类较为合适的方法。

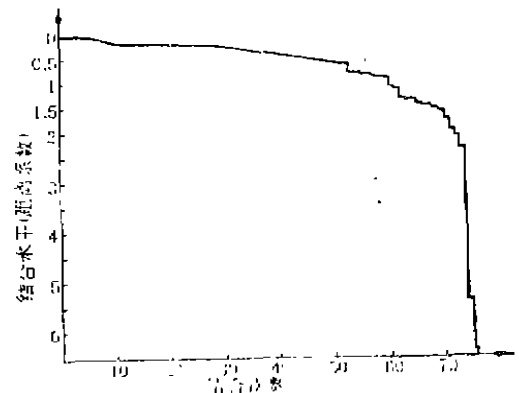


图 3 离差平方和法分类结合线

Fig.3 Linking curve of the ward's hierarchical classification method

2. 用57个指标对77个大队进行等级聚合方法分类, 初步将北京郊区村级农业生态系统分为低集约化山区、中集约化平原和高度集约化近郊农业生态系统三大类型和一种特殊类型。区分明显且易于理解。阈值更小的划分能相应出现一些亚类型。

3. 地理位置相近的分类单元有首先发生聚合的群化现象。平原农业类型表现出单元间差异较小而稳定, 近郊蔬菜类型受城市影响较大, 单元间差异大而表现出不稳定性。

北京郊区不同类型农业生态系统特征指标的数值差异较大, 典型山区和平原类型用数值分类方法分为同一类型的可能性极小, 今后聚类分析中可以将典型的山区类型从分类单元中分出, 分别选择分类指标进行分类, 可建立分类体系和分类系统, 使亚类型的划分有更高的分辨率。

今后可考虑用在标准化之前对数变换的方法在一定程度上减小大数或过小数对分辨率的影响。

### 参 考 文 献

- [1] 王学萌等, 1984, 县级农业区划区域划分的数学方法。农业区划 6:34—42。
- [2] 阳含熙、卢泽愚, 1981, 《植物生态学的数量分类方法》。科学出版社, 第96—120页。
- [3] 方开泰、潘恩沛, 1982, 《聚类分析》, 第96—101页, 地质出版社。
- [4] 徐克学、李德中, 1983, 我国人参属数量分类研究初试。植物分类学报, 21(1):34—43。
- [5] 樱井明久, 1973, 关东中部地区农业区划。农业地理译丛(2):第43—57页。

## THE CLASSIFICATION OF AGROECOSYSTEM TYPES WITH CLUSTER ANALYSIS IN BEIJING SUBURBS

Gu Wu

(Institute of Crop Breeding and Cultivation, Chinese Academy of Agricultural Sciences)

In this paper, 57 specific attributes which are suitable for the classification of the village level agroecosystems are selected according to the principles of agroecosystem. These specific attributes are taken for classifying the agroecosystems of 77 villages (production brigades) in some counties of Beijing suburbs and 5 methods of non-overlapping intrinsic hierarchical agglomerative analysis are adopted. The results are as follow: except the single linking method, the other methods adopted can classify the villages clearly into 4 major types: the less intensive farming mountain areas, the middle intensive farming plain areas, the highly intensive farming environs and the specific type. The Ward's hierarchical classification method is better than the others in this classification. With less threshold values, some subtypes were found. The differences among the units in the plain areas are smaller and the structure of it's units are steady, but the units closed to urban in third type are influenced by the city, the differences are obvious and the structure of it's units is unsteady.

**Key words:** cluster analysis, Beijing suburbs, agroecosystem,