

等值线图和质量特征图综合分析 及其结果应用的可能性*

张 根 寿

摘 要

本文以分析和解译自然条件图为例,研究了等值线图和质量特征图综合分析的理论和方法。对多幅不同性质、不同表示方法的地图上的地理现象模拟的抽样方法、样点规模和样点布置作了深入研究,特别是将模糊数学综合评分和聚类分析引入信息采集及处理。通过县级农业气候类型区划的实例证实,结果正确且有实用价值。研究表明,模糊数学在多幅不同性质及表示方法的地图分析中有独特的优点。

【关键词】 地图分析;样本模拟;模糊评价;相似聚类速度

引 言

定义1 质量特征图可理解为所有用质底法和区域(范围)法表示的地图,还可引申到某些分级统计图。

等值线图和质量特征图表示了满布或不满布制图区域现象的空间分布特征^[1]。等值线图表示了现象的数量特征,质量特征图则表示了现象的质量特征。单幅或两幅同一表示方法,同一区域、同一性质的等值线图及质量特征图的分析及其分析结果的应用无论是在科学研究、工程规划等方面;还是为了生产新地图从定性分析到定量分析都已经有了较为成熟的方法^[2]。本文运用模糊数学理论,以反映区域气候特点的各幅图——等值线图与地貌类型图、土壤图——质量特征图为例的代表性图件,讨论信息采集处理上的差异及综合分析途径,用模式分析方法将简单变量的分析深化到多种变量多种不同性质地图的综合分析及应用研究,提供一种分析地图的方法。

1 分析自然条件地图中的地理现象模拟

地图上表示的地理现象的数量特征来自统计或实测数据;质量特征来自专业人员通过遥感图象判译及实地考察,研究总结或数据的质量化(如按高度将山地划分为高山、低山……)。

收稿日期:1990-02-16

* 本文是硕士论文的一部分,指导教师为祝国瑞和杜道生副教授。

1.1 变量选择

地图表达的空间现象是多种多样的,而分析地图是在具有明确的使用目的指导下进行的。因此,选择分析变量的重要性体现在它决定了分析结果的可应用性。

(1) 根据地图应用的目的选择分析变量

地理系统是多层次、多变量的复杂系统。对某个问题的分析解决,主要通过选择多个关键变量,从质和数两方面分析变量,总结归纳完成。这往往是一个多元系统,或是多输入单输出系统,如:

$$y = f(x_1, x_2, \dots, x_n)$$

或是多输入多输出系统,如:

$$(y_1, y_2, \dots, y_m) = f(x_1, x_2, \dots, x_n)$$

在研究目的的背景下,考虑各要素之间的关系,依据分析和应用地图的目的选择变量。

(2) 根据地图分析的深度和区域选择变量

对于不同的研究目的,地图分析的深入程度存在差异性。采用统计相关性选择的变量在某些问题研究中,仅获得了影响该问题解决的主要成分,但某些非主要成分也可能是讨论问题的关键变量。如省级综合农业气候区划,依据主成分光、热、水、气、积温、气温等进行,是宏观上大范围的,且全省范围内复杂程度增加,不可能进行深入细致的研究;而县级农业气候区划作为直接指导生产实践的依据要分析得较深入,就一个县的范围来说,气象台站、水文站等监测网点很少,除了气候变量外,还要选择地貌类型、地面高度、植被群落、作物分布、土壤等变量。

1.2 模拟方法

从地图上既不可能也不需要得到某地理现象总体的全部数据和度量,而是用一组样本代替总体。把样本信息即用数量特征作为对某地区进行研究的基础,从而得到这一地区的整体规律。

分析和研究自然条件地图采用的抽样结构一般都是空间性质的。即样本所在的位置起主要作用,样本应在理论上代表整个区域的一切可能位置。自然图的分析不宜采用随机抽样和典型抽样方法,这是因为样本布置易受主观意识的影响(作者多次实验过),样本常常发生凝聚现象,以致把一些重要的空间特性忽略掉,即样本不能保证代表整个研究区域的空间。系统抽样可在地图上抽取所有网络交点(若用点抽样方式),给出了一个均匀的总体覆盖,且快而简单易行。

当然系统抽样绝非完美无缺,它具有片面性,但只要有足够的样本就可使其片面性达到极小而忽略。

1.3 样点规模和分布

定义 2 将样本概念限制在点抽样方式的样本定义为样点。

定义 3 点抽样方式抽取的整个区域上样点的总体称其为样点规模。

定义 4 点抽样方式抽取的整个区域内某类目(分量)的全部样点称其为样点数。

(1) 样点分布

对多种面状分布现象的多幅图综合分析,采用等距网络交点覆盖分析区域。区域界线外最近一个样点也作为分析参加者。多幅图的样点分布应保证每个样点地理位置在各幅图上完全相同,才能保证样点的客观性及分析结果的精确性。

(2) 样点规模

地图分析变量的状态由样点决定(空间面状分布现象)时,首先解决的问题是选择多少个样点

才能代表总体。样点规模与选取的置信度,选取变量的标准误差(指示要素在空间或时间分布的差异程度的指标),对变量估值所要求的精度有关。在质量特征图及某些其它种类图中,一般地说都表达出某种特征在总体中所占的比例,在确定样点规模中有独特的作用。计算样点规模的公式为

$$n = \frac{Z^2 \cdot P \cdot (1 - P)}{d^2} \quad (1)$$

式中 d 是允许误差(精度控制量); P 是属于某地理类目的比例; Z 是根据置信度水平在正态表中查得。式(1)的提出基于下面的理论:

点抽样的均值误差允许范围是

$$d = \pm Z \frac{\sigma}{\sqrt{n}}$$

以试探性样点的标准差 S 代替总体标准差 σ , 则 $n = \frac{Z^2 \cdot S^2}{d^2}$ 依据地图表示空间自然现象的特点,用 $P \cdot (1-P)$ 代替 S^2 。 $(1-P)$ 表示不属于的百分比(原意从试探性样点方差代替总体比例方差)。

各类目的样点数由下式计算:

$$n_i = \frac{n \cdot P\%}{100\%} \quad (2)$$

式中 $P\%$ 为某一类目(分量)比例数; n 为样点规模; 100% 为变量总比例。

1.4 样点距的确定

(1)先在某一幅图上,利用样点总体比例较小的 1~2 个类目的样点数求得样点距。将毫米等距网格覆盖在预选图幅上,使得那些占总体比例较小的类目中所落的样点数为由式(2)求得的 n_i , 那么 n_i 相邻点距为样点规模的样点距,称为先验样点距。再反算落在其它类目中的样点数是否与由式(2)算得的 n_i 之差小于等于预先给定的允许误差 d , 若满足,即可定为后验样点距。

(2)参与分析的各幅图比例尺一般是不同的,以样点空间位置模拟地理现象,必须取得同一地理位置的不同变量值,即不同比例尺图幅上样点距不同,地理位置应相同,采用下式

$$d_B = d_A \cdot M_A / M_B \quad (3)$$

式中 M_A, M_B 分别是 A 图和 B 图比例尺分母; d_A 为已知 A 图上样点距。

当有数幅图进行综合分析,应顾及各图幅的类别实际,用最小二乘法求得 n, n_i 及样点距。样点距尽可能取为整数,方便样点布设,提高点位精度。

2 自然条件地图上信息采集

等值线图表示自然现象的空间分布,或是用间隔尺度,或是用比率尺度反映空间连续分布现象的定量特征。一般都采用内插值法或外推法采集信息^[2],其重复劳动的工作量很大。

质量特征图上表示的地理现象是定性类别概念,或为只有次序关系的有序尺度(如仅以平缓、较陡、陡、很陡表示地势起伏),或是用文字、符号表示现象的存在与否的名义尺度(如语言分布、地质构造、岩石分布等),一般都采用二元(即 0, 1)数据结构、权值法等获取信息,完成定性向定量的转换。而二元数据结构仅限于只有一幅质量特征图时,权值法要量算单元及类目的面积,劳动强度也很大。

从概念化理论角度来看,质量特征本身是一个模糊概念,很难用确定的数字表示。如地貌形态

类型(图 2)是以质量概念表达地表外貌的分布及特征。但质量特征图表达的定性变量及等值线图用数值表达的变量的空间分布及地域差异性,依据变量的分类原则反映了分量(类目)某种应用价值的优劣性,即属于某概念的程度——隶属度。这是将模糊数学中综合评分方法^[3]引入定性质量特征向定量描述转化的过程及应用的依据。模糊综合评分的实质是用模糊相似选择法对定性特征或定量分级依据某法则转换为模糊定量数据,即模糊变换。

定义 5 设 X 到 Y 的一个映射 f , 不仅值域 $Y^* = Y$, 并且 X 中每一个元素 x 有一个且只有一个 y 与之对应, 则称映射 f 是一一对应的

$$f: X \rightarrow Y$$

这种映射是可逆的

$$f^{-1}: Y \rightarrow X$$

具体来说, 图幅中的“有序”数据集为 X , 则 Y 是定义在实数域内的数集, 步骤为:

(1) 确定各类目及各级评分标准。在具体问题中, 基本原则是, 认为“质量好”的、“优”的评分高, 即隶属度高的评分高。

(2) 建立样点值矩阵。个别样点或许正好落在两种或几种类目或级别的分界线或界线交点上, 这些特殊点的评分值遵循如下定义。

定义 6 设 m 为样点所覆盖的分量(类目)数, 则该样点评分值为

$$q_i = \frac{1}{m} \sum_{i=1}^m Q_i \quad (4)$$

Q_i 为第 i 个分量的评分值。

[例] 地貌类型的模糊评分法数量化过程。若考虑其农业可利用程度, 对应法则确定见表 1。

表 1 地貌类型与对应的评分

地貌类型	中山	低山	丘陵	岗地	平原
评分标准	1	2	3	4	5

3 模糊聚类及分辨性

3.1 数据的模糊化预处理

在进行数据分析组合前, 先标准化同时消去量纲, 模糊数学称为建立隶属函数, 本文采用线性函数

$$\mu_A(x) = \begin{cases} 0 & x = a_1 \\ \frac{x - a_1}{a_2 - a_1} & a_1 < x < a_2 \\ 1 & x = a_2 \end{cases}$$

3.2 聚类过程实现

定义 7 设给定模糊关系矩阵 $R = (R_{ij})$, 如果它满足 $R(I, I) = 1, R(I, J) = R(J, I)$, 则称 $R = (R_{ij})$ 为模糊相似矩阵, 它满足自反性和对称性^[3]。

设 $W = \{X_1, X_2, \dots, X_n\}$ 是所有被分类样本的全体。 T_1, T_2, \dots, T_n 为 W 的子集, 是 W 的一个分类且满足:

$$\bigcup_{i=1}^k T_i = W, \quad T_i \cap T_j = \emptyset \quad (i \neq j) \quad (5)$$

在 T_1, T_2, \dots, T_k 中选 a_i 作为 T_i 的代表元。在聚类过程中, 一个类是多次在 W 中选元而确定的。每次选元过程中作为选择标准的元, 称它为值勤员。具体步聚为:

(1) 取 $a_1=1$, 作为 T_1 的代表元, 以 $a_1=1$ 为值勤员, 对所有满足 $R(1, J) \geq \lambda$ 的元都选入 T_1 类;

(2) 取 T_1 中未担任过值勤员的元 I_1 做值勤员, 对所有满足 $R(I_1, J) \geq R(1, I_1)$ 元都选入 T_1 类;

(3) 循环(2)的过程, 直到 T_1 中的元都担任值勤员, T_1 类无法再按(2)扩大为止;

(4) 取 $a_i \in W$ 或 $(\bigcup_{i=1}^{k-1} T_i)$ 作为 T_i 类的代表元, 在 W 或 $(\bigcup_{i=1}^{k-1} T_i)$ 中进行(1)、(2)的过程, 将 T_i 类确定下来;

(5) 循环(4)的过程, 直到 W 中的元全被分到某一类中去为止。

上述讨论得出如下结论:

若 $\forall \lambda \in [0, 1]$, 且 λ_1 和 λ_2 满足 $0 \leq \lambda_1 < \lambda_2 \leq 1$, 则按水平 λ_2 分出的每一类必是按水平 λ_1 分出的某一类的子类。

3.3 分辨性

模糊聚类过程类的划分是依据 λ 水平截矩阵而进行, 给出不同 λ 水平, 就划分出类别及类别的量。模糊数据矩阵 R 的元是客观实际的模拟量, 反映了所研究事物的内在联系; 其次 λ 水平的选择是根据 R 矩阵中元显示的数据分布特征所决定。这里, 数据分布的特征则是所研究的客观事物内在本质的再现。聚类是按地理现象本身的相似性和差异进行, 而分类又经进一步检验得到。故聚类结果是客观的。

4 自然条件图综合分析及其结果应用*

4.1 变量(图幅)选择

当阳县农业气候区划选取变量为: $\geq 10^\circ\text{C}$ 积温 ($\geq 10^\circ\text{C}$ 积温分布图); 年降水量 (年降水量图); 年平均气温 (年平均气温图); 总辐射量分布 (总辐射量分布图); 无霜期分布 (无霜期分布图); 地貌类型 (地貌类型图), 它以形态成因分类, 隐含有地面高度因素^[6]。

当然还应选择植被群落、作物分布等自然变量。在表示方法上它们是以质底法、范围法制作的地图, 其实质仍是质量特征的定性变量; 其次植被群落, 尤其是作物分布在当阳县已受到人类的强烈干预, 自然分布特征已不显著, 故未选入。尽管如此, 用水、热、光、地貌来作为气候类型区划变量仍具有重要的实用价值。

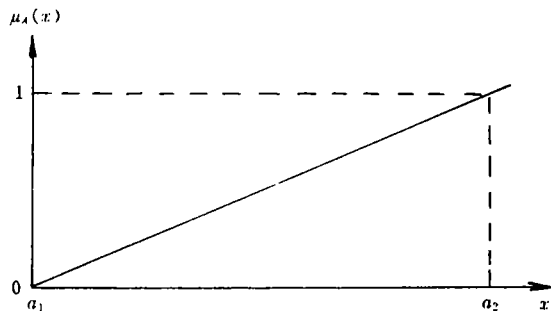


图 1

* 以当阳县农业气候类型区划为例。

4.2 信息采集

气候类型区划是分析区域内空间现象的平均分布特征的结果,不需顾及区划单元。因此采用将栅格网覆盖在各幅图上,用点群模拟面状现象。

地貌类型图上表示出了各地貌形态占区域总面积的比例,依据该图核算有关数据。

(1)设允许误差为 $d=4\%$,置信度取99.9%,查得 $Z=3$ 。用(1)式算得样点规模见表2。经分析,决定用137作为模拟空间的样点规模,因其它对气候类型区划来讲数量显得太大。

表2 样点规模

地貌形态	中山	低山	丘陵	平原、岗地
占面积% ^(%)	2.5	5.1	52.5	39.9
样点规模	137	272	1402	1371

(2)地貌形态类型各类目样点数由式(2)求得,列在表3及图2中。

表3 地貌形态类型各类目样点数

地貌形态	样点数	占样点规模比例	面积比例	误差
中山	3	2.13	2.5	-0.37
低山	7	5.00	5.1	-0.10
丘陵	70	50.03	52.5	-2.47
平原	60	42.90	39.9	+3
Σ	140	100	100	+0.06

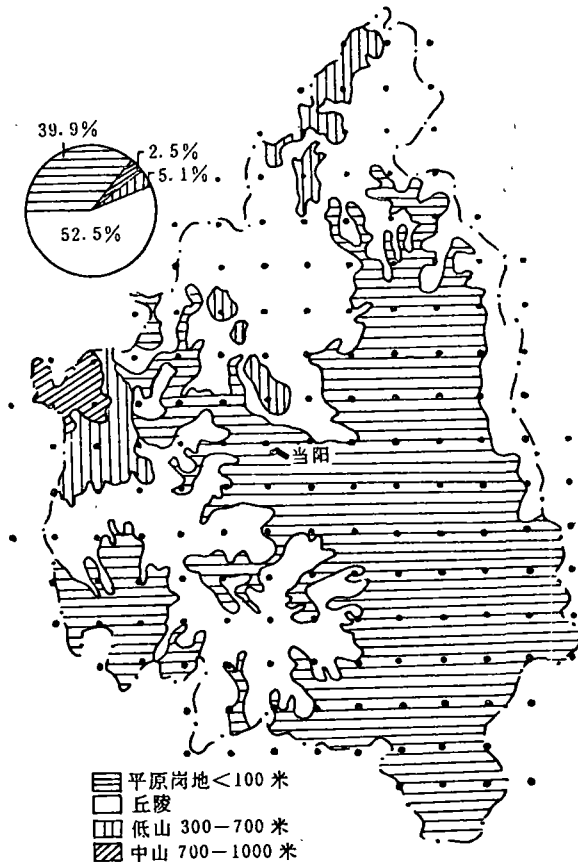


图2 当阳县地貌类型图及样点布置

(3)分别采用插值法和模糊评分法获取样点数值。对各变量的评分标准确定为对农业的利用程度贡献,越有利评分越高。如降水量见表4。插值法采集的信息称为数据1——DAT.1,见表5;采用模糊评分法采集的信息称为数据2——DAT.2,见表6。

表4 降水量评分标准

分级(mm)	≤950	950~1000	1000~1050	>1050 ^[3]
评分	1	2	3	4

区域界线外有31个样点。因数据量较大(171个)仅列出两个原始数据的首尾部分。

表5 DAT.1(插值法)

样点	变量 积温 ≥10℃ (℃)	年平均 降水 (mm)	年平均 温度 (℃)	无霜期 日数 (天)	总辐射量 千卡/cm ²	安全生长期 积温 (℃)
1	4500	1030	15.8	248	106.5	3350
2	4600	1030	15.7	248	106.7	3350
3	4700	1030	15.6	248	106.8	3350
4	4600	1030	15.6	248	106.1	3320
5	4750	1030	15.7	249	106.4	3350
⋮	⋮	⋮	⋮	⋮	⋮	⋮
167	4970	970	16.3	263	106.0	3530
168	4900	970	16.3	263	106.5	3530
169	4900	970	16.3	261	105.4	3530
170	4880	970	16.3	261	105.7	3530
171	4860	970	16.3	262	106.3	3530

表6 DAT.2(模糊评分法)

样点	变量 地貌	积温 ≥10℃	年平均 降水	年平均 温度	无霜期 分布	总辐射量	安全生长期 积温
1	2	2	3	5	5	6	4
2	3	2	3	5	5	6	4
3	3	2	3	5	5	6	4
4	2	2	3	5	5	6	4
5	2	2	3	5	5	6	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
167	4	3	2	6	8	6	6
168	4	3	2	6	8	6	6
169	4	3	2	6	8	5	6
170	4	3	2	6	8	5	6
171	4	3	2	6	8	6	6

4.3 建立模糊相似矩阵及聚类

(1)对 DAT.1、DAT.2 用极差标准化^[3];

$$x'_{ij} = \frac{x_{ij} \min_{1 \leq j \leq n} \{x_{ij}\}}{\max_{1 \leq j \leq n} \{x_{ij}\} - \min_{1 \leq j \leq n} \{x_{ij}\}} \quad \begin{cases} i = 1, 2, \dots, m \\ j = 1, 2, \dots, n \end{cases}$$

$$0 \leq x'_{ij} \leq 1$$

(2)用欧氏距离公式^[3]:

$$D_{ij} = 1 - \sqrt{\frac{1}{m} \sum_{k=1}^m (x'_{ki} - x'_{kj})(x'_{ki} - x'_{kj})}$$

建立了 $\tilde{R}_1, \tilde{R}_2, \tilde{R}_3$ 三个模糊相似矩阵,形式为:

$$\tilde{R} = \begin{bmatrix} D_{1.1} & D_{1.2} & D_{1.3} & \cdots & \cdots & \cdots & D_{1.170} & D_{1.171} \\ & D_{2.2} & D_{2.3} & \cdots & \cdots & \cdots & D_{2.170} & D_{2.171} \\ & & D_{3.3} & \cdots & \cdots & \cdots & D_{3.170} & D_{3.171} \\ & & & \ddots & & & & \vdots \\ & & & & \ddots & & & \vdots \\ & & & & & \ddots & & \vdots \\ & & & & & & & D_{171.171} \end{bmatrix}$$

(3)聚类是在对模糊相似矩阵的元 D_{ij} 进行数据分布特征分析的基础上,选择 λ 水平数据序列聚类。为避免聚类结果的重复输出部分,进一步改进了聚类过程。具体方法为:若在 λ_1 的水平上得到了聚类结果群 C ,对 C 中每个 $c_i = \{i, j\}$ 置零,即 $c_i = \{i, j\} = 0$,将它归入一类,不属 C 的 D_{ij} 原值不变, i, j 编号不变,称为隐划去 $c_i = \{i, j\}$ 。取 $\lambda_2 < \lambda_1$,在 λ_2 水平上聚类过程中,不重复输出在 λ_1 水平上的 $c_i = \{i, j\}$,聚类结果不变。

如在 λ_1 水平上有 $C = \{3, 5\}$,即 3 与 5 归为一类,让 $C = \{3, 5\} = 0$,即 $D_{3.5} = 0$;在 λ_2 水平上有 $C = \{2, 3\}, C = \{2, 5\}, C = \{3, 5\}$ 分别为一类,事实上即 $C = \{2, 3, 5\}$ 共为一类,由于在 λ_1 水平上隐划去了 $C = \{3, 5\}$,虽在 λ_2 水平上不出现 $C = \{3, 5\}$ 归类过程,但通过 λ_2 水平上的 $C = \{2, 3\}, C = \{2, 5\}$ 仍然得到 $C = \{2, 3, 5\}$,见图 3。从而大大加快了聚类过程,节约了时间和物力。DAT. 2 的动态聚类图略。

4.4 聚类结果分析及评价

DAT. 1 和 DAT. 2 虽然是由两种不同方法获取的信息,但聚类结果基本一致,且 DAT. 2 获得的结果更适合客观现象的分布特征,如紫盖寺林场区年降水量小于 970(mm),在 DAT. 1 的气候类型区划图中与其周围毗邻,而在 DAT. 2 的气候类型区划图中与周围分异显著,单独成为一区。

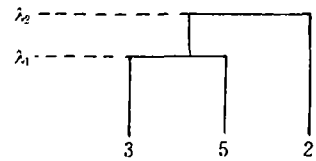


图 3

从 DAT. 1, DAT. 2 得到的气候类型区划图中,区划界线在延伸趋势上雷同。与原地图集^[6]中“农业气候区划图幅”区划界线在从西北向东南的带状分布规律上是吻合的。可以认为模糊聚类结果正确,而且更趋实用合理。从原图上分析,农业气候区划仅考虑了 $\geq 10^\circ\text{C}$ 积温和年平均气温^[6],这是粗略的、不合理的。从实验结果可以看出,模糊评分法获取信息的速度是插值法的 5~10 倍。

从 DAT. 2 得到的气候类型区划图 4 中,类型区划界线大致与地貌形态类型界线吻合,但不是简单的重合,应该认为地貌在气候区划中,尤其在小区区的区划中有极为重要的作用。类型区划界线是多种气候变量和影响因素的综合,不单纯受某种变量绝对控制。

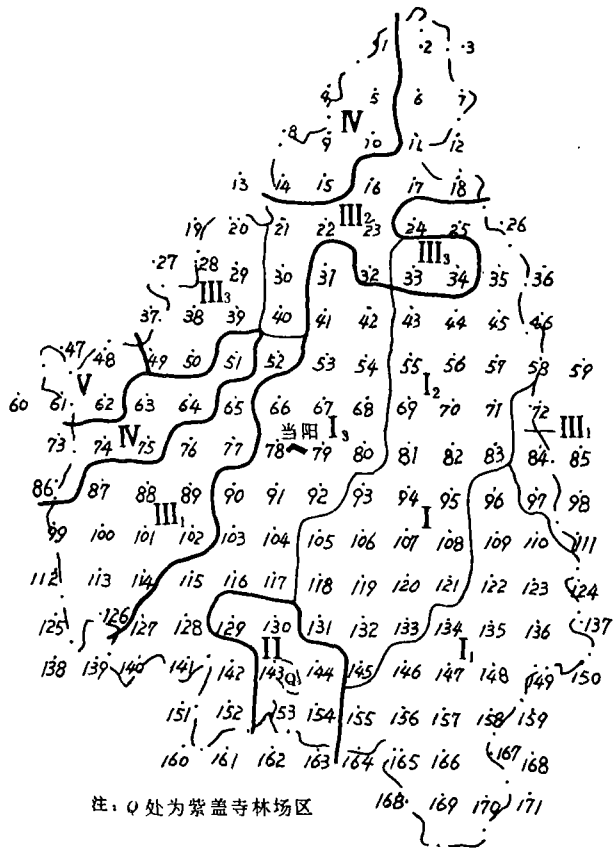


图 4 气候类型区划

据 DAT. 2 的类型区划图 4, 从地域上看

- I —— 中、南平原、岗丘温暖少雨区
- II —— 南丘岗温暖多旱区
- III —— 南北过渡丘陵温和较湿润区
- IV —— 西南、西北低山温凉湿润区
- V —— 西中山温寒湿润区

各区气候要素各项指标及地貌形态类型见表 7。

表 7 当阳县农业气候类型区划指标

指 标 变 量	I	II	III	IV	V	单 位
地貌	平原、岗丘	丘、岗	丘陵	低山	中山	
年平均降水	975	900	1025	1025	1025	mm
≥10℃积温	5150	5300	5000	4800	3500	℃
安全生长期积温	3525	3525	3400	3000	1500	℃
年平均气温	16.25	16.25	15.5	15	11	℃
总辐射量	106	105	104	102.5	103	千卡/cm ²
无霜期分布	262.5	270	250	235	228	天/年

必须指出:本文的气候分区研究既是区划又是类型,所以分类分区在分布上是可以分离的。故这种方法能反映出类单位在空间的分布特征,且分类单位是对具体对象属性的逐级概括。

5 结 论

点抽样方式和系统抽样方法模拟地图上表示的面状分布现象,依据本文方法选择足够的样点和正确的布设样点,就能保证模拟的客观性和分析结果的精度。模糊数学综合评判——评分法的引入,使在等值线图和质量特征地图上采集信息具有速度快、操作简单、适应面广、正确反映客观实际和多幅不同性质图综合分析成为可行的优点。模糊模拟矩阵聚类分析中值勤员和非重复归类及输出方法在机上运行中显示出节约时间、人力和物力的效益。通过当阳县气候类型区划的例子得出地图分析和应用有很好的发展前景和实际应用价值。可以预料随着各种地图的大量生产,地图分析方法的成熟,地图在科学研究和经济建设中将发挥越来越重要的作用。

参 考 文 献

- [1] 张克权,黄仁涛.专题地图编制.测绘出版社,1982.
- [2] 祝国瑞.地国分析.武汉测绘科技大学讲义,1987.
- [3] 汪培庄.模糊集合论及其应用.上海科学技术出版社,1983.
- [4] 董文泉等.数量化理论及其应用.吉林人民出版社,1979.
- [5] 当阳县农业区划办公室.当阳县农业区划图集.1985.
- [6] 张 超等.计量地理学导论.高等教育出版社,1984.

The Combined Analysis of Isoline Maps and Quality Feature Maps and the Feasibility of the Application of the Result

Zhang Genshou

Abstract

This paper, by analysing the maps of physical conditions, studies the theories and methods of the combined analysis of isoline maps and quality feature maps. Deep researches have been made into the sampling methods and size and arrangement of samples of simulating geographical phenomena on the maps of different properties and expressions. Especially, the fuzzy evaluation and fuzzy cluster have been introduced into the information acquisition and processing. Trial for regions of agricultural climate types has proved the result is correct and is able to be used in practice also. It shows that fuzzy mathematics has a unique advantage in the combined analysis of maps of different properties and expressions.

【Key words】 map analysis; natural conditions; sample simulation; fuzzy evaluation; similar cluster speed