

我国降水量的模糊聚类分区

徐国昌 姚辉

(兰州干旱气象研究所)

提 要

本文采用模糊聚类中的系统聚类分析法对全国年、季降水量作了分区计算。为了客观地表示区内相关程度，提出了一种计算区内相关系数的方法。并根据实际情况，提出了选择降水分区方案的原则。结果比较令人满意。

一、前言

降水量的客观分区，是天气气候分析研究的基础性工作。中央气象台曾将我国东部地区夏季大范围降水分为三类主要雨型，即黄河流域以北，黄河与长江之间，长江流域以南等*。但是这种分型是经验性的，定性的。全国的降水客观分区还没有人做过。本文采用模糊聚类中的系统聚类分析法对全国年、季降水量作了分区计算。为了客观表示区内相关程度，我们提出了一种区内相关系数的计算方法。同时，根据降水分区的实际情况，提出了选择分区方案的原则。用上述方法作的全国降水分区比较令人满意，可供长期预报降水分区和天气气候分析参考。

本文用的资料主要以中央气象台的中国160站的月降水量资料为主，为加密青藏高原及西北地区台站网密度，又增补了高原18个站的资料，共计178个站，年代长度为35年(1951—1985)。

二、方法

1. 系统聚类分析法是以模糊数学理论为基础的一种客观分类方法。这种方法是从事物本身具有的模糊性出发，以描述事物亲疏程度的相关系数等统计量为依据，在模

糊等价关系的基础上，按不同分类水平逐步进行客观分类。关于这个方法的具体做法请参见文献[1]。

然而，系统聚类的结果不是一个唯一的分类方案，而是一系列的分类方案。每一 λ 分类水平对应一种分类方案， λ 由大到小，分类由细到粗。到底选择其中的哪一个方案，或哪几个方案？模糊聚类并没有给出一条成熟的方法，一般由作者凭经验选定，这显然是不够客观的。这次全国年降水量模糊聚类中共出现了117个分类方案，单凭经验选择显然是很困难的。因此，需要制定一套合理的确定分区的办法。

2. 区内平均相关系数的计算

在不同的 λ 水平下，各区的大小(包括的站数)是不同的。一般区愈大(站愈多)，区内各站之间的平均相关系数愈低。为了选择分区方案，需要确定区内平均相关系数。设某区内共有 P 个站，各站之间的相关系数为 r_{ij} ，相关系数的总个数为组合数

$$C_p^2 = \frac{P!}{2!(P-2)!} = \frac{1}{2}P(P-1)$$

因此，某一区内各站之间的平均相关系数

* 中央气象台，西太平洋副热带高压和我国东部降水的关系，一九七六年全国长期天气预报经验交流会材料选编。

$$\bar{R} = \frac{\sum_{i=1}^p \sum_{j=i+1}^p r_{ij}}{\frac{1}{2}P(P-1)}$$

3. 最终分区方案的选择

(1) 由于我们作的是全国范围的分区, 只选取 ≥ 3 个以上站组成的区。

(2) 区内各站之间的相关应该 是比较紧密的, 因此, 各区内平均相关系数应该达到一定的信度水平。本工作中, 我们规定的信度水平为0.05, 个别情况下, 也考虑0.10水平的结果。

(3) 一般来说, 总是区愈小, 区内相关愈密切, 如果单纯根据区内相关系数的高低来选择, 那么选出的区一般都是很小的。这不符合我们的目的。由于聚类结果中各区满足某一信度水平的往往不止一个结果, 同时考虑到分区所包括的面积在达到给定信度水平条件下应尽可能大。我们规定, 最终选择的分区, 是在0.05信度水平下的最大区域。因此, 最终选择的各个区, 可能是在不同的 λ 水平下选出来的。

(4) 个别情况下, 出现两块地理位置相隔甚远但属于同一聚类分区的情况, 实际上是两块相关比较密切的遥相关区。则作为两个区对待。

根据以上原则, 我们在模糊聚类计算的基础上, 得到了全国月、季、年降水量的客观分区。

三、结果

1. 我国年降水量可划为以下11个区(图1):

(1) 长江中下游区(CJ): 主要包括长江中下游流域及江南丘陵地区。

(2) 淮河流域区(HH): 以淮河流域为中心, 南与长江流域区相接。

(3) 河套区(HT): 包括宁夏、陕北、甘肃东部及内蒙巴彦淖尔盟, 乌兰察布盟和伊克昭盟。

(4) 华北区(HB): 包括河北、山西、北京、天津、锡林郭勒盟和昭乌达盟。

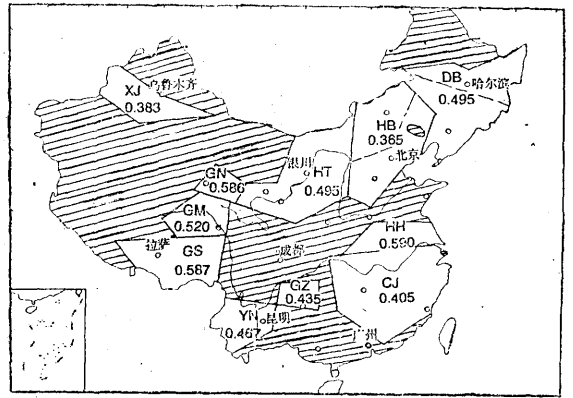


图1 全国年降水量分区图

折线为区界线,

字母为各区代号, 数字为区内平均相关系数。

斜线区为不能分区的地区(下图同)

分二个亚区, 北区为锡林郭勒盟, 其余地区为南区。

(5) 东北区(DB): 分南北两个亚区, 北区主要是黑龙江南部, 南区为吉林和辽宁。

(6) 高原南部区(GS): 主要位于雅鲁藏布江流域及西藏东部。

(7) 高原中部区(GM): 位于青海南部, 长江、黄河源地。

(8) 高原北部区(GN): 位于柴达木盆地东部。

(9) 云南区(YN): 位于云南省。

(10) 贵州区(GZ): 位于贵州省。

(11) 新疆区(XJ): 位于天山南北。

2. 夏季(6—8月)、秋季(9—11月)与年降水量分区比较接近(图3、图4), 春季(3—5月)和冬季(12—2月)相差较大(图2、图5)。这是可以理解的。因为年降水量主要由夏季和秋季决定。

但是夏、秋季与年降水量分区也有某些差异。首先我们注意到秋季我国东部的南方和北方之间的过渡带没有了, 淮河区向北扩大到黄河, 向西扩大到陕西南部, 华北区南

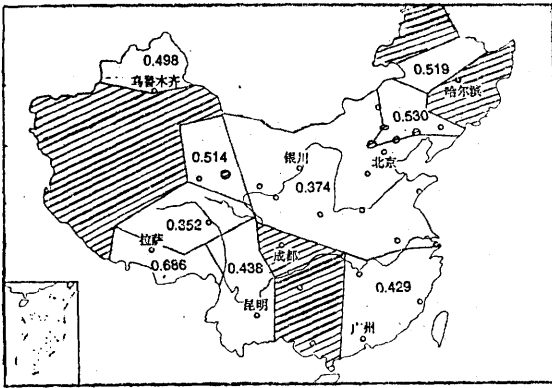


图 2 全国春季降水量分区图

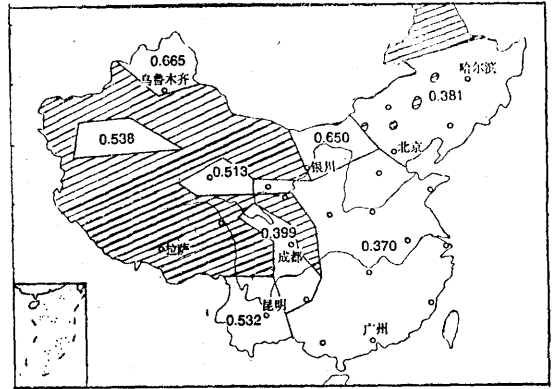


图 5 全国冬季降水量分区图

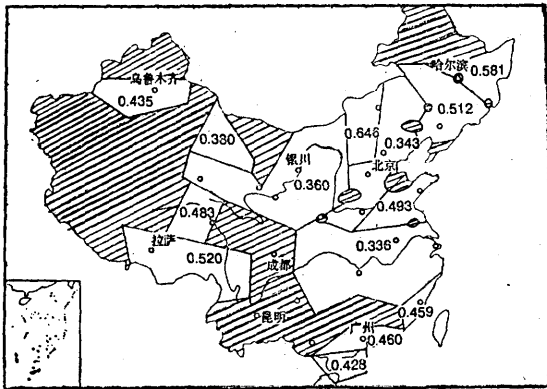


图 3 全国夏季降水量分区图

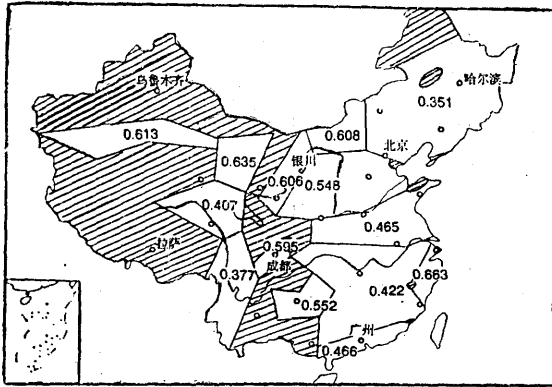


图 4 全国秋季降水量分区图

部和河套区南部合并，东北区向南扩大到华北区北部，在我国东部自南向北排列着4个区。这是因为一般秋季降水主要在9月到10月上半月，此时，西太平洋副高比较完整，且年际南北位置变化较大，相应的雨带年际变化也较大，副高偏北的年份多雨区在 35°N 以北，副高偏南的年份多雨区在长江流域及其以南。形成了南北排列比较规则的几个区。

夏季从6月到8月，是副高和雨带季节变化最大的季节。一般6月多雨带在长江中下游和淮河，7月到黄河，8月到河套北部和长城内外。各年副高和雨带变化很大，尤其是8月，变化最大，除了副高南北位置变化大外，东西位置变化也很大。使整个夏季降水出现了复杂的情况。但是与年降水分区相比，几个主要区位置没有大的变化。变化最大的是淮河区与长江中下游区合并，在年降水量的过渡区山东出现了一个新区，它可能和副高东西位置的年际变化较大有关。

春季最重要的特点是华北、河套以及淮河流域三个区合并成一个最大的区域。这是因为春季我国北方降水的多少主要由东亚大槽的深浅决定，其作用范围几乎包括华北和西北东部的广大地区。

冬季我国东部的河套、淮河、长江中下游以及华北区南部合成一个大区，东北为另

一个大区。这是因为冬季东亚大槽最深，其深浅的变化几乎决定了我国整个东部地区的降水量。

月降水量的分区与年季相似，这里不再赘述。

在年、季降水量分区图上的斜线阴影区为在0.05的信度水平下，不能聚类出3个站

以上的地区。说明这些地区降水距平情况比较复杂，划不出较大的区来。从年降水分区图(图1)上看出，从山东、河南经陕南到四川有一个很明显的阴影区，其北部边界大约在 35°N ，西部边界在 100°E 左右。这块阴影区实际上是我国南方和北方，我国东部与青藏高原的过渡区。尤其是四川的过渡区，

附表 历年各区年降水量平均值(标准化)

年	区	CJ	HH	HT	HB	DB	GS	GM	GN	YN	GZ	XJ
1951		0.20	-0.22	0.10	-0.38	0.58	0.68	-0.09	9999	-0.38	-0.12	0.17
1952		0.44	-0.03	-0.18	-0.51	-0.25	1.02	-0.10	9999	0.55	0.26	0.31
1953		0.48	-1.16	-0.20	-0.01	0.23	-0.15	-0.47	9999	-0.02	-0.38	0.65
1954		1.34	1.93	-0.17	0.89	0.53	1.26	0.16	9999	-0.42	2.46	0.54
1955		-0.46	-1.01	-0.31	0.37	-0.22	0.69	-1.40	-0.09	0.63	-0.20	-0.54
1956		-0.21	2.04	-0.26	1.06	0.82	-0.14	-1.07	-1.10	-0.72	-0.28	0.07
1957		-0.05	-0.21	-0.79	-0.38	1.35	-0.53	0.21	-0.72	0.66	0.89	-1.19
1958		-0.44	-0.20	1.02	0.39	-0.89	-0.67	-0.13	-0.22	-1.48	-1.06	2.65
1959		0.42	-0.06	0.78	1.69	0.83	-0.24	-1.59	0.72	0.91	-0.07	0.88
1960		-0.36	0.36	-0.33	-0.52	1.27	0.42	-0.44	-0.22	-1.64	-0.53	0.58
1961		0.55	-0.60	1.52	0.04	-0.41	-0.11	0.56	-1.18	0.70	-0.10	-0.25
1962		0.41	0.65	-0.66	-0.45	-0.38	2.40	0.08	-0.67	-0.11	-0.68	-0.59
1963		-1.19	0.51	-0.54	0.64	0.48	0.36	0.76	-0.10	-0.72	-0.16	-0.77
1964		-0.40	0.42	1.64	1.10	0.58	-0.20	-0.19	-0.22	-0.83	0.55	0.69
1965		-0.33	0.36	-1.33	-1.14	0.07	-0.02	0.11	-1.19	1.10	0.29	-1.06
1966		-0.71	-1.55	-0.43	-0.28	-0.06	0.58	-1.08	-0.31	1.34	-1.99	0.26
1967		-0.61	-0.47	1.64	0.15	-0.85	-1.43	0.41	2.61	-0.40	1.05	-1.08
1968		-0.29	-0.46	0.12	-0.69	-0.21	0.49	-0.39	-1.09	0.34	-0.15	-0.99
1969		0.29	0.45	-0.54	0.78	0.27	0.33	-1.18	0.14	-0.67	-0.34	0.22
1970		0.69	0.24	0.37	-0.32	-0.53	0.52	0.30	1.21	1.14	0.87	0.08
1971		-1.16	0.08	-0.61	-0.02	0.30	-0.58	0.16	0.57	0.32	0.32	0.37
1972		-0.16	0.72	-0.87	-1.14	-0.56	-0.64	-0.20	-0.96	-0.66	0.58	0.46
1973		1.29	-0.70	0.49	0.77	0.06	-0.05	-0.47	-0.36	1.56	-0.02	0.53
1974		-0.44	0.72	-0.90	0.03	0.04	0.35	1.25	-0.19	-0.27	0.18	-0.31
1975		1.26	0.75	0.31	-0.25	-0.75	-0.80	0.70	0.06	-0.45	-0.52	-0.42
1976		-0.28	-1.00	0.31	0.02	-0.58	-0.64	0.33	1.16	0.75	0.48	-0.23
1977		0.16	0.08	-0.03	0.35	-0.66	1.00	-0.68	1.78	-0.83	1.10	-0.04
1978		-0.96	-2.04	0.67	0.07	-1.04	0.59	-0.61	-0.54	-0.15	-0.06	-0.29
1979		-0.83	0.14	-0.10	0.51	-1.05	0.16	-0.91	0.53	0.13	-0.61	0.50
1980		0.52	0.38	-0.91	-1.11	-0.27	0.54	0.68	0.10	-0.25	-0.70	-0.46
1981		0.10	-0.48	0.23	-0.39	0.54	-0.85	0.79	0.77	-0.23	-1.63	0.05
1982		0.21	0.04	-1.11	-0.51	-0.76	-1.15	0.98	0.78	-0.47	-0.04	-0.64
1983		1.06	0.20	0.36	-0.23	0.21	-1.51	1.25	0.17	1.00	0.79	-0.24
1984		-0.07	0.21	0.18	-0.57	0.26	-0.66	-1.61	-0.39	-0.48	-0.07	0.35
1985		-0.47	-0.06	0.52	0.03	1.00	0.84	1.62	-0.11	0.36	-0.12	-0.24

几乎各季都有。说明高原和我国东部天气气候很不一样*。珠江流域和沿海可能受热带天气系统影响,变化比较大。西北干旱区一般都不能形成区,主要是因为降水量太少,变率大,代表性较差。测站较少也是部分原因。另外东北北部也有一块过渡区。

分区后,第k区中(对年降水量 $k=11$)第i站l年降水量标准化值为

$$x'_{kii} = \frac{x_{kii} - \bar{x}_{ki}}{\sigma_{ki}} \begin{matrix} k=1, 2, \dots, 11, \\ i=1, 2, \dots, P, \\ l=1, 2, \dots, 35 \end{matrix}$$

在区内作P个站标准化值算术平均,得k区第l年平均标准化降水量

$$\tilde{x}_{ki} = \frac{1}{P} \sum_{i=1}^P x'_{kii}$$

为了使用方便,我们将1951—1985年全国11个区逐年标准化降水量值列于附表,供参考。

四、小结和讨论

1. 模糊聚类可以提供一系列客观的、可供选择的分区方案,是气象要素客观分区的一种有效的计算方法。但是对于进行全国降水这样应用大量台站资料的分区计算,没

有一套客观的辅助方法是难以得出结果的。本文提供的以计算区内平均相关系数为基础的方法在进行全国月、季、年降水的客观分区中取得了比较满意的结果,这个方法可供进行类似工作时参考。

2. 模糊聚类的结果受台站网密度的影响。选多大的密度要根据需要而定。但是无论用多大的密度,台站的分布要尽量相对均匀。

3. 区内平均相关系数的信度检验,我们仍用单相关系数信度检验的方法。其中可能有些偏差。

4. 这次聚类分区,得到的长江中下游区、淮河区以及华北区与过去的概念是一致的,是对中央台三类主要雨型(1)的一种证实。但是这次分区发现高原上的三个区等,为开展分区长期预报和研究降水区的形成提供了客观事实。

参考文献

- (1) 贺仲雄, 模糊数学及其应用, 天津出版社, 1983, 1.
 (2) 徐国昌, 李梅芳, 青藏高原的雨季, 北方天气 (5) 1984.

Classification of precipitations in China by fuzzy cluster method

Xu Guo chang Yao Hui

(Lanzhou Dryness and Meteorology Institute)

Abstract

In this paper, annual and seasonal precipitations in China are classified by the fuzzy systems cluster method. A method used to find out the correlation coefficient within a certain class is given for objectively expressing of interclass correlation level. A principle for selecting of precipitation classification scheme is put forward. The results are satisfactory.

*徐国昌, 西北干旱的形成和预报, 兰州干旱气候会议文件, 1983.