

辽宁汛期降水模糊聚类分析

杨文艳 (南京气象学院研究生部 南京 210044; 盘锦市气象局 盘锦 124010)

摘要 运用1957~2000年辽宁24个测站的月降水资料,对辽宁汛期降水进行模糊聚类分析,将汛期降水异常分成若干相关较好的区域,并进行了信度检验。

关键词 汛期降水 模糊聚类 信度检验

近年来,关于对区域气候异常特征的研究不断深入,就辽宁范围而言,经向和纬向的气候特征存在着很大的差异。本文使用辽宁24个测站的月降水资料,用模糊聚类的方法将汛期降水进行分区分类,以揭示辽宁汛期降水异常的小区域气候特征,为短期气候预测服务提供良好的依据。

1 方法

1.1 资料处理

将1957年1月—2000年12月24个测站近44a的月降水资料按月份分成1~12月资料,提取7~8月的资料进行标准化处理,计算出7~8月平均降水标准化距平序列。

1.2 计算

用模糊聚类方法对近44a 7~8月24个测站平均降水标准化距平序列进行分析。

1.2.1 计算所有测站两两相关(似)系数 r'_{ij} , 生成相关(似)矩阵 \tilde{R}

$$\tilde{R} = (r'_{ij}), i, j = 1, 2, \dots, n \quad (1)$$

相关矩阵是 n 阶对称矩阵,其元素变化在 $[-1, 1]$ 之间。

1.2.2 将相关(似)的矩阵 \tilde{R} 改造成模糊矩阵 R

$$R = (r_{ij}) \quad (2)$$

$$r_{ij} = 0.50 + \frac{r'_{ij}}{2}, i, j = 1, 2, \dots, n$$

R 的元素已变化在 $[0, 1]$ 之间,是模糊矩阵。

1.2.3 将模糊矩阵 R 改成分类关系

$$R \cdot R = R^2, R^2 \cdot R^2 = R^4 \quad (3)$$

若 R 不断自乘(用模糊矩阵运算规则),最后结果已不再改变,设为 R^∞ ,则 R^∞ 是分类关系。

1.2.4 由分类关系 R^∞ 得到截模糊矩阵 R_λ^∞

$$R_\lambda^\infty(i, j) = \begin{cases} 1 & R^\infty(i, j) \geq \lambda \\ 0 & R^\infty(i, j) < \lambda \end{cases} \quad (4)$$

λ 根据 R^∞ 中的元素从大到小选取。这一步可以得到不同 λ

水平的截模糊矩阵。得到变量(样本)分类结果。

2 聚类分析

2.1 代表站

24个代表站(代号1~24)依次为:阜新、黑山、朝阳、叶柏寿、绥中、兴城、锦州、盘山、大洼、沈阳、鞍山、营口、丹东、岫岩、熊岳、庄河、瓦房店、大连、清原、开原、桓仁、章党、本溪、彰武。

2.2 聚类结果

经过模糊聚类分析,选取24个测站7~8月降水分成4类的结果(图1),经模糊聚类将辽宁24个代表站的汛期降水分成4类,4类降水区域由东向西基本呈经向分布。

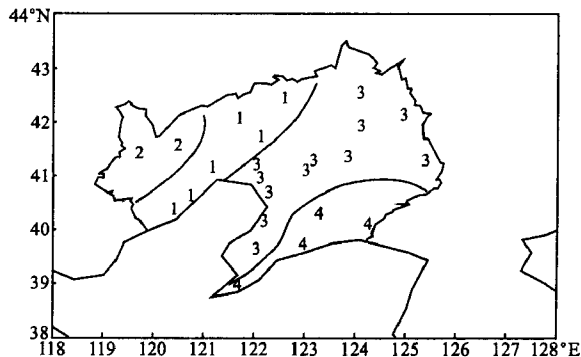


图1 辽宁7~8月降水模糊聚类结果

辽西西部(2类):朝阳、叶柏寿;辽西东部(1类):阜新、彰武、黑山、绥中、兴城、锦州;辽宁中部(3类):盘山、大洼、沈阳、鞍山、营口、熊岳、瓦房店、清原、开原、桓仁、章党、本溪;辽宁东南(4类):丹东、岫岩、庄河、大连。

2.3 结果

表1中数值大小无意义,数值相同为一类。在分析中用相关系数来衡量2个样本相似程度,数值相同的站点之间相关性高于数值不同的点,相关系数如表1。

表1中相关系数为 $r_{20} = 0.848\ 716\ 9$,说明聚成一类的

表1 模糊聚类过程各级相关系数

级别	相关系数	级别	相关系数	级别	相关系数	级别	相关系数
1	0.927 748 1	7	0.890 958 8	13	0.874 418 0	19	0.850 384 2
2	0.925 245 2	8	0.883 134 4	14	0.869 839 5	20	0.848 716 9
3	0.911 144 4	9	0.880 626 2	15	0.869 440 3	21	0.845 281 2
4	0.901 471 2	10	0.877 610 3	16	0.868 497 1	22	0.841 850 8
5	0.895 911 0	11	0.874 599 7	17	0.867 800 4	23	0.839 917 5
6	0.892 384 6	12	0.874 436 4	18	0.863 866 4		

收稿日期:2003-07-31;修订日期:2003-12-15

站点间相关系数都大于 r_{20} 。

聚成 4 类的过程: 首先, $r_1 = 0.927\ 748\ 1$, 站点 12 和 15 聚成一类, $r_2 = 0.925\ 245\ 2$ 时, 站点 5 和 6 聚为一类, 越先聚成一类的站点相关性越好, 降低标准 $r_{19} = 0.850\ 384\ 2$, 24 个站点聚成五类, 再降低标准 $r_{20} = 0.848\ 716\ 9$, 五类中的第 1 和第 2 类聚成一类, 24 个站点分成四类。

降低相关标准, 当取 $r_{23} = 0.839\ 917\ 5$ 时, 所有站点归为一类, 说明 24 个测站汛期降水相关系数都在 0.839 917 5 之上, r 越大, 分类越细, r 越小, 分类越粗, r 由大到小, 样品逐步聚合成一类。

2.4 聚类分析过程图解

以第 3 类聚类分析过程为例, 聚类结果分析过程可表示如图 2。图 2 中左侧为所取的 24 个测站代号, (1)~(11) 序号为聚类的先后顺序, 代表不同的相关标准, 其中先聚为一类的站点汛期降水相关性较后聚成一类的要好。

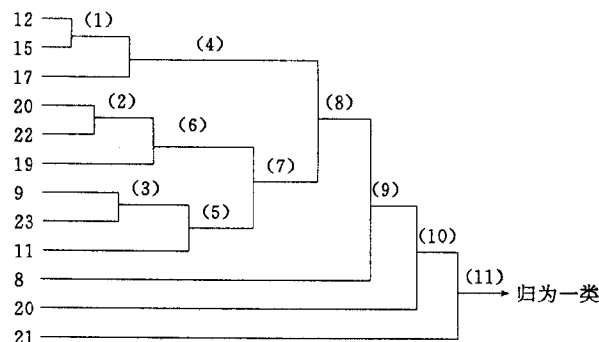


图 2 模糊聚类分析过程示意

3 信度检验

为了进一步表明所分成的四类降水区域在降水异常上同区内具有较好的一致性, 计算同类降水代表站的两两相关系数, 结果第 1 类大于等于 0.90, 第 2 和第 4 类大于等于 0.85, 第 3 类大于等于 0.82, 相关系数的最小值为 0.82, 大部在 0.85 以上, 单序列相关性的 T 检验^[2]远超过 0.01 统计检验标准, 即所分各类降水异常同类的站点间存在显著的相关关系, 聚类结果在统计学上认为是有意义的。为验证所分不同类间降水异常特征差异显著, 计算各类间 44 a 降水序列的相关系数列在表 2。

表 2 四类降水异常类间相关系数

类别	第 1	第 2	第 3	第 4
第 1	1.00	0.66	0.40	0.30
第 2	—	1.00	0.67	0.45
第 3	—	—	1.00	0.64
第 4	—	—	—	1.00

表 2 中数值表示所在行与列的类间的相关系数。由表 2 可见, 各类间相关系数都在 0.67 以下, 最小为 0.30, 计算差值 T 检验全部通过了显著性水平为 0.01 的检验标准, 说明不同类的均值存在显著差异, 所分类别具有统计学意义。

4 结语

模糊聚类分析方法是一种多元统计分析方法, 它的优点是在样本的类型和类型数目未知条件下, 通过分析确定所给

秋的收获始于春的耕耘

——贺《辽宁气象》创刊二十周年

马虹旭

曾几何时

有人问我: 你最爱什么

我回答说

我爱天空朵朵的白云

我爱海边徐徐的晚风

我爱早晨灿烂的朝霞

我爱雨后美丽的彩虹

然而

我最爱的还是编辑、作者为之奋斗的一方热土

——《辽宁气象》

秋的收获始于春的耕耘

二十年的业绩如星河灿烂辉煌

风雨中, 气象人播种着幸福

雷鸣中, 气象人耕耘着希望

全国优秀气象期刊一等奖的历史令我赞叹

连续多年全省一级期刊的荣誉令我神往

忘不了二十个春秋

编辑们艰苦的创业

忘不了七千个日夜

作者们不断的追求

啊, 《辽宁气象》

当历史的车轮滚滚向前的时候

你大胆地改革开放

当二十一世纪的太阳向你招手的时候

你勇敢地扬起时代的臂膀

看, 起伏的山峦是气象人不屈的意志

听, 奔涌的流水是气象人进军的号角

腾飞吧, 《辽宁气象》

满天星斗已为你照亮前进的征程

前进吧, 《辽宁气象》

电闪雷鸣已把成功的凯歌奏响

样本应该划分的类型和数目。本文用模糊聚类的方法将辽宁地区汛期降水分成四类, 聚成一类的降水异常具有良好的一致性, 不同类间降水异常差异显著。四类降水在地理分布上的特征为: 明显呈东西向分布, 经向差异显著大于纬向差异, 可以作为进一步有针对性的对小范围的气候特征研究的基础。

参考文献

- 1 施能. 气象科研与预报中的多元分析方法. 北京: 气象出版社, 1995.
- 2 魏凤英. 现代气候统计诊断预测技术. 北京: 气象出版社, 1999.