

文章编号: 1000-2022(2005)06-0739-07

## 聚类分析在人工增雨效果检验中的应用

房彬<sup>1,2</sup>, 肖辉<sup>2</sup>,  
王振会<sup>1</sup>, 孙海燕<sup>1</sup>, 黄美元<sup>2</sup>

(1 南京信息工程大学 电子工程系, 江苏 南京 210044

2 中国科学院 大气物理研究所, 北京 100029)

**摘要:** 在前人研究工作的基础上, 对非随机区域历史回归试验进行了改进, 提出了一种新的试验方案——基于聚类的浮动对比区历史回归人工增雨效果统计检验方法 (CA-FCM 方法), 并与浮动对比区历史回归统计检验方法 (FCM 方法) 进行了比较。通过对河南省春季 6 次作业进行评估得出: 由于 CA-FCM 方法采用了聚类分析, 提高了对比区和影响区的相关系数, 及引入了整层大气可降水量作为协变量, 所以提高了作业区自然降水量估计值的准确性, 比 FCM 方法有更高的效果评估效率。

**关键词:** 聚类分析; 浮动对比区; 历史回归; 整层大气可降水量; 增雨效果

**中图分类号:** P481 **文献标识码:** A

客观、科学、定量地评价人工增雨效果是人工影响天气学科的重要组成部分。几十年来, 根据国内外大量的外场试验研究, 云物理工作者设计出两大类试验方案——随机化试验和非随机化试验。

随机化试验周期长, 并且要求放弃一半的可作业机会, 一般在业务化作业中很少采取这种试验方案<sup>[1]</sup>。非随机化试验对业务化作业来说是一种花费少、易获取增雨效果的方法。它通常分为序列实验、区域对比试验、区域历史回归试验和区域控制模拟试验等<sup>[2]</sup>。现在增雨作业多采取非随机化试验, 这种试验存在作业范围广、作业目标区不固定、作业区 and 对比区的选择存有主观性等问题。为了解决这些问题, 国内外科学家提出了许多方法, 如移动目标区方法<sup>[3]</sup>、浮动控制法 (FCM)<sup>[4,6]</sup>、非固定目标区增雨评估法<sup>[7]</sup>等。这些方法都在播云试验中得到了应用, 并取得了一定的效果, 但是仍然存在对比区和作业区相关性差、功效偏低等问题。

为此, 本文在这些方案的基础上, 利用聚类分析, 对非随机区域历史回归试验进行了改进, 并引入物理协变量作为控制因子和雨量网格插值计算降水量, 提出了一种新的试验方案——基于聚类的浮动对比区历史回归人工增雨效果统计检验方法 (CA-FCM)。并根据河南省近 20 a 的历史资料, 利用 CA-FCM 方法, 对河南省春季人工增雨作业效果进行评估。

收稿日期: 2004-02-10 改回日期: 2004-05-24

基金项目: 国家自然科学基金资助项目 (40333033); 国家自然科学基金资助项目 (40175001); 国家“十五”科技攻关计划项目 (2001BA610A-06-05); “奥运科技”专项 (2001BA904B09); 中国科学院大气物理研究所创新项目 (8-4605)

作者简介: 房彬 (1977-), 女, 辽宁沈阳人, 硕士。E-mail: fangbindem@163.com.

# 1 CA-FCM 方法及应用

## 1.1 CA-FCM 方法

基于聚类的浮动对比区历史回归人工增雨效果统计检验方法 (CA-FCM) 假设作业期作业区与对比区雨量的统计相关关系与历史上同类天气条件下雨量的区域相关性相同<sup>[8-9]</sup>。

### 1.1.1 聚类分析用于影响区和对比站点划分

#### (1) 区域聚类

进行人工增雨作业的区域一般面积较大, 区域内的自然条件和物理条件差异也较大。在过去的非随机化历史回归试验中较少考虑这个问题, 选取的对比区和影响区相关性不高, 从而降低了效果评估的效率。为了克服这个问题, 本文根据作业区域内各气象站的长期雨量资料, 利用统计学中的聚类分析方法, 将作业区域分成若干个副区, 然后分别对每个副区进行效果评估。

根据 1971—1989 年河南省所有站点 4 月 24 h 累积降水量, 采用欧氏距离平方和最近邻法对河南省进行了聚类分区, 把河南省分成 7 个副区, 图 1 给出了 7 个副区的划分情况。

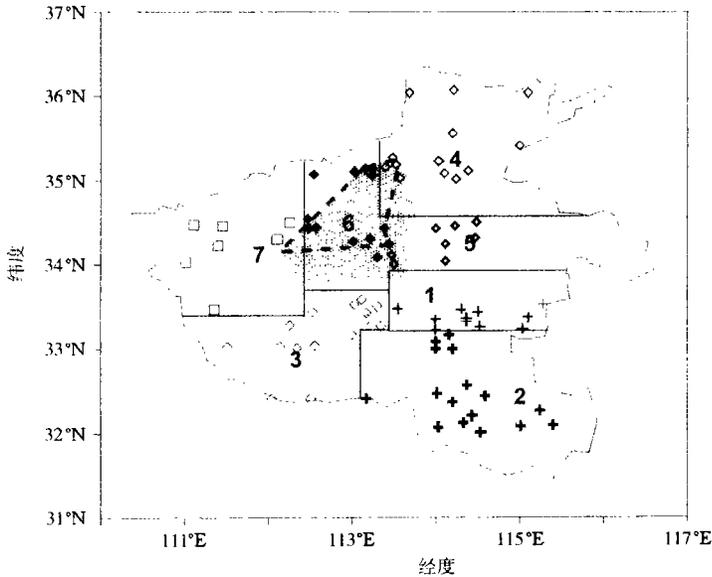


图 1 2002 年 4 月 5 日河南省增雨作业影响区和对比站点划分

Fig 1 Distribution of the control area and target area in the cloud seeding operation on 5th April 2002 in Henan Province

#### (2) 影响区和对比站点划分

总影响区: 以飞机播撒路线上的 700 hPa 高空风 3 h 传输距离以内所覆盖的范围作为总影响区。图 1 中, 粗虚线为作业路线, 阴影区即为总影响区。

副影响区: 总影响区与某副区相重叠的范围。图 1 中, 总影响区 (阴影区) 与标号为 4 的副区相重合的范围即为 4 号副影响区。

对比站点: 选择上述副影响区所位于的副区中未受作业影响的站点和与副影响区所位于的副区相关性高的其他副区中的未受作业影响的站点作为对比站点。

### 1.1.2 协变量和因变量选择

#### (1) 作为协变量的两个条件

协变量与目标区的自然降水量密切相关。

协变量应不受催化作业措施的影响。

为了满足以上条件, 雨日资料应该选择满足降水空间分布比较均匀, 时间变化较为平缓的系统性层状云系降水的资料。本文选择的数据全部为河南省 4 月层状云条件下的气象数据。

#### (2) 协变量的选择

选择的协变量为对比站点 08 00—20 00 的 12 h 累积降水量  $R_1$  和对比站点的整层大气可降水量  $W_1$ 。

整层大气可降水量用如下的经验公式<sup>[10-11]</sup>:

$$W_1 = \exp(b_0' + b_1' t_d) \quad (1)$$

其中,  $W_1$  为测站上空整层大气的可降水量,  $t_d$  为地面露点,  $b_0'$ 、 $b_1'$  为经验系数。

#### (3) 因变量的选择

选择的因变量为副影响区的降水体积  $V_e^{(k)}$ 。在副影响区上迭加上  $5 \text{ km} \times 5 \text{ km}$  网格, 根据雨量站历史资料进行格点插值, 计算出网格交叉点的雨量值。然后, 根据公式<sup>[4]</sup>

$$V_e^{(k)} = 10^{-6} (S_e^{(k)} / n_e^{(k)}) \sum_{i=1}^{n_e} p_i^{(k)} \quad (2)$$

求得副影响区的降水体积  $V_e^{(k)}$  (单位为  $\text{km}^3$ )。其中,  $S_e^{(k)}$ 、 $n_e^{(k)}$  为副影响区的面积和格点个数,  $p_i^{(k)}$  为第  $i$  个格点的降水量,  $k$  为副区序号。

### 1.1.3 建立副影响区自然降水体积 $V_e^{(k)}$ 的多元回归预报方程

首先, 选择未作业的副影响区的降水体积  $V_e^{(k)}$  为因变量, 对比站点的降水量  $r_i$  (取原始降雨量  $R_1$  的正态化变量  $\lg R_1$ ) 和整层大气可降水量  $w_i$  (取原始可降水量  $W_1$  的正态化变量  $W_1^3$ ) 为自变量, 然后进行逐步回归。最后, 得到副影响区的自然降水体积  $V_e^{(k)}$  的多元预报回归方程, 即

$$V_e^{(k)} = d + \sum_{i=1}^{n_1} a_i r_i + \sum_{j=1}^{n_2} a_j r_j + \sum_{i=1}^{n_1} b_i w_i + \sum_{j=1}^{n_2} b_j w_j \quad (3)$$

其中  $r_i$ 、 $w_i$  和  $n_1$  为副影响区所位于的副区内的对比站点的降水量、整层大气可降水量、对比站点的个数。  $r_j$ 、 $w_j$  和  $n_2$  为与副影响区所位于的副区相关性高的副区中的对比站点的降水量、整层大气可降水量、对比站点的个数,  $k$  为副区序号。

### 1.1.4 作业期总影响区自然雨量的估计值

把回归方程 (3) 作为副影响区自然降水体积的预报方程, 将作业期作业时段对比站点的各个物理量代入方程, 得出每个副影响区的自然降水体积期望值  $V_e^{(k)}$ , 然后, 根据所有副影响区的自然降水体积期望值的和, 可以得出总影响区的自然降水体积期望值  $V_e$ 。

### 1.1.5 催化作业效果

绝对增雨量<sup>[1]</sup>为

$$\Delta V_e = V_e - V_{e_0} \quad (4)$$

相对增雨量<sup>[1]</sup>为

$$R = \frac{V_e - V_{e_0}}{V_{e_0}} \times 100\% \quad (5)$$

其中  $V_e$  为作业日作业时段总影响区实测降水体积,  $V_c$  为作业日作业时段总影响区自然降水体积期望值。

最后, 利用  $t$  检验对每次催化作业相对增雨量  $R$  进行评估, 当  $t$  值的显著水平小于 0.05 时增雨效果显著; 利用成对试验符号秩和检验法对平均相对增雨量  $\bar{R}$  进行评估, 当秩和显著性水平小于 0.05 时, 增雨效果显著。

## 1.2 CA-FCM 方法个例评估

### 1.2.1 个例评估

利用河南省未开展人工增雨作业的 1971—1989 年 4 月的站点气象数据作为历史统计资料, 对 1998—2002 年河南省 4 月的 6 次飞机增雨作业个例 (表 1) 用 CA-FCM 方法进行效果检验。表 2 给出了河南省 6 次作业总影响区的增雨情况。

表 1 1998—2002 年河南省 6 次飞机增雨作业路线

Table 1 The seeding routes of six cloud-seeding operations in Henan Province in the April of 1998—2002

作业号	作业日期	作业起始时刻	作业结束时刻	作业路线
1	2002-04-05	10:35:00	12:10:00	新郑—洛阳—焦作—辉县—新乡—郑州
2	2000-04-01	09:55:00	11:11:00	焦作—泌阳—孟津—伊川—偃师—新密
3	2000-04-14	11:00:00	12:20:00	延津—通许—临颖—许昌—登封
4	1999-04-08	06:40:00	08:05:00	焦作—安阳—内黄
5	1999-04-11	08:54:00	09:50:00	新乡—濮阳—开封—郑州—洛阳
6	1998-04-11	15:34:00	16:45:00	温县—卢氏

表 2 CA-FCM 方法检验河南省 6 次增雨作业总影响区的增雨效果

Table 2 Results of the effect evaluation of the six cloud seeding operations in Henan Province by CA-FCM

作业日期	$V_e / (10^6 \text{m}^3)$	$V_c / (10^6 \text{m}^3)$	$\Delta V_e / (10^6 \text{m}^3)$	$\frac{\Delta V_e}{V_c}$	$t$	$\alpha$	标准差 $/(10^6 \text{m}^3)$	$\frac{\Delta V_e}{V_c}$ 秩次
2002-04-05	1 299.656	1 063.242	236.414	0.222	2.062	< 0.05	114.653	4
2000-04-01	1 667.427	1 704.936	-37.509	-0.022	-0.320	> 0.05	117.352	-1
2000-04-14	253.836	211.144	42.692	0.202	0.27	> 0.05	158.924	3
1999-04-08	978.520	912.346	66.174	0.073	0.41	> 0.05	163.212	2
1999-04-11	1 685.447	1 314.521	370.926	0.282	2.15	< 0.05	172.457	5
1998-04-11	2 691.662	1 763.643	928.018	0.526	7.48	< 0.05	124.112	6
合计	8 576.548	6 969.832	1 606.715					-1
平均	1 429.425	1 161.639	267.786	0.231				

由表 2 可见, CA-FCM 方法对河南省 6 次作业进行评估, 得到总影响区催化后 12 h 平均绝对增雨体积为  $0.27 \times 10^9 \text{m}^3$ , 相对增雨效果为 23.1%。应用  $t$  检验法对每次作业相对增雨量进行显著性检验, 得到有 3 次作业  $t$  值的显著水平  $\alpha < 0.05$ , 增雨效果显著。应用成对试验符号秩和检验法对平均相对增雨量进行显著性检验, 得到秩和为 -1, 显著性水平  $\alpha = 0.031$ , 表明增雨效果是很显著。

### 1.2.2 空白试验

通常, 增雨播云总是选择降水条件好的云区进行。若不催化, 影响区的雨量是否也大于预计的雨量? 为了对此进行验证, 需要进行“空白试验”。在实际作业中, 由于各种原因, 总会漏掉或放过一些可试验 (或作业) 的机会, 可把它们集中起来进行不催化的空白试验。应用 CA-FCM 方法, 对河南省 1998—2002 年 4 月 6 次没有作业的降水事件进行空白试验的效果评估, 结果见表 3。

表 3 CA-FCM 方法检验河南省 6 次空白作业总影响区的增雨效果

Table 3 Results of the effect evaluation of the six “blank” cloud-seeding operations in Henan Province by CA-FCM

作业日期	$V_e / (10^6 \text{m}^3)$	$\bar{V}_e / (10^6 \text{m}^3)$	$\Delta V_e / (10^6 \text{m}^3)$	$\frac{\Delta V_e}{V_e}$	$t$	$\alpha$	标准差 $/(10^6 \text{m}^3)$	$\frac{\Delta V_e}{V_e}$ 秩次
2002-04-28	1 429.104	1 303.071	126.033	0.097	1.099	> 0.05	114.653	1
2000-04-14	28.919	36.975	-8.055	-0.218	-0.07	> 0.05	117.353	-4
2000-04-08	67.773	52.011	15.762	0.303	0.10	> 0.05	158.924	6
1999-04-18	75.079	66.485	8.594	0.129	0.10	> 0.05	163.212	3
1999-04-27	16.706	22.584	-5.878	-0.260	0.04	> 0.05	172.457	-5
1998-04-05	90.473	102.818	-12.346	-0.120	0.10	> 0.05	124.112	-2
合计	1 708.054	1 583.944	124.110					10
平均	284.676	263.990	20.685	0.078				

由表 3 可见, 这 6 次空白试验的平均“增雨”效果仅为 7.8%。应用  $t$  检验法对每次空白试验相对增雨量进行显著性检验, 得到这 6 次作业  $t$  值的显著水平  $\alpha > 0.05$ , 增雨效果不显著。应用成对试验符号秩和检验法对平均相对增雨量进行显著性检验, 得到秩和为 10, 显著性水平  $\alpha = 0.281$ , 表明空白试验的增雨效果不显著。

## 2 FCM 方法及应用

### 2.1 FCM 方法

浮动对比区历史回归人工增雨效果统计检验方法 (FCM) 与 CA-FCM 方法的主要差别在于它没有进行聚类分析。

首先, 选取对比站点 08:00—20:00 的 12 h 积累降水量  $R_1$  和影响区总的降水体积  $V_e$  作为基本资料。然后, 选择相似天气条件下未作业情况的影响区总的降水体积  $V_e$  为因变量, 对比站点的降水量  $r_i$  (取原始降雨量  $R_1$  的正态化变量  $\lg R_1$ ) 为自变量, 逐步回归得到自然降水体积的多元预报回归方程<sup>[5]</sup>

$$\bar{V}_e = \sum_{i=1}^{n_1} a_i r_i + c \quad (6)$$

其中,  $r_i$  为对比站点的降水量,  $n_1$  为对比站点的个数,  $a_i$  和  $c$  为回归系数。最后, 把多元逐步回归方程 (6) 作为影响区自然降水体积的预报方程, 将作业期作业时段对比站点的降水量  $r_i$  代入方程, 求出影响区作业时段的自然降水体积估计值  $\bar{V}_e$ 。

### 2.2 FCM 方法个例评估

同样利用河南省未开展人工增雨作业的 1971—1989 年 4 月的站点气象数据作为多元回

归的历史统计资料,对河南省 1998—2002 年 4 月的 6 次飞机增雨作业案例(表 1),用 FCM 方法进行效果检验。表 4 列出了这 6 次作业的增雨效果。

表 4 FCM 方法检验河南省 6 次增雨作业总影响区增雨效果

Table 4 Results of the effect evaluation of the six cloud seeding operations in Henan Province by FCM

作业日期	$V_e / (10^6 \text{m}^3)$	$V_e' / (10^6 \text{m}^3)$	$\Delta V_e / (10^6 \text{m}^3)$	$\frac{\Delta V_e}{V_e}$	$t$	$\alpha$	标准差 $/(10^6 \text{m}^3)$	$\frac{\Delta V_e}{V_e}$ 秩次
2002-04-05	1 496. 532	1 270. 401	226. 131	0. 178	1. 56	> 0. 05	145. 329	3
2000-04-01	1 667. 427	1 765. 639	- 98. 212	- 0. 056	- 1. 14	< 0. 05	86. 215	- 2
2000-04-14	271. 395	213. 132	58. 263	0. 273	1. 12	> 0. 05	190. 082	5
1999-04-08	978. 520	820. 254	158. 266	0. 193	1. 82	> 0. 05	86. 819	4
1999-04-11	2 261. 426	2 295. 697	- 34. 271	- 0. 015	0. 13	> 0. 05	261. 426	- 1
1998-04-11	2 684. 305	1 934. 536	749. 769	0. 389	6. 41	< 0. 05	116. 979	6
合计	9 359. 605	8 299. 659	1 059. 946					- 3
平均	1 559. 934	1 383. 277	176. 658	0. 128				

由表 4 可见, FCM 方法对河南省 6 次作业进行评估,得到总影响区催化后 12 h 平均绝对增雨体积为  $0. 177 \times 10^9 \text{m}^3$ , 平均相对增雨量为 12. 8%。应用  $t$  检验法对每次作业相对增雨量进行显著性检验,得到有 4 次作业  $t$  值的显著水平  $\alpha > 0. 05$ , 增雨效果不显著。应用成对实验符号秩和检验法对平均相对增雨量进行显著性检验,得到秩和为 - 3, 显著性水平为  $\alpha = 0. 078$ , 表明增雨效果不是很显著。显著性不高的原因主要是由于没进行聚类分区,从而导致区域相关性不高。

### 3 CA-FCM 和 FCM 两种评价方案比较

FCM 方法的平均相对增雨量为 12. 8%, 秩和显著性水平为  $\alpha = 0. 078$ , 增雨效果不是很显著,这是由于对比区和影响区相关性差造成的。

CA-FCM 方法评估效果很好,去掉自然降水起伏因素的影响,总影响区仍可得到 15. 3% 的相对增雨效果。这是由于 CA-FCM 方法采用聚类分析(CA)方法改进对浮动对比区的分区方案,提高了对比区和影响区的相关性;采用网格插值技术提高了雨量的计算准确度;引入了整层大气可降水量作为协变量提高了作业区自然降水量估计值的准确性,因而从整体上提高了非随机化作业效果的评估效率。

### 4 小 结

本文提出了基于聚类的浮动对比区历史回归人工增雨效果统计检验方法(CA-FCM)。并通过对个例的评估和与浮动对比区历史回归人工增雨效果统计检验方法(FCM)结果的比较,得出 CA-FCM 的评估效果较好的结论。

综上所述,选择合适的物理协变量作为控制因子以及提高影响区与对比区相关性,可以增大自然降水量估计值的准确度,提高非随机化作业效果的评估效率。

## 参考文献:

- [ 1 ] 叶家东, 罗幸贫. 随机试验功效的数值分析 [ J ]. 大气科学, 1984, 42( 1 ): 69-78
- [ 2 ] 曾光平. 非随机化人工增雨试验效果的统计模拟研究 [ J ]. 应用气象学报, 1999, 10( 2 ): 255-256
- [ 3 ] 夏彭年. 内蒙古地区层状云催化的条件和效果——介绍“移动目标区”人工增雨效果评估方法 [ C ]. 人工影响天气 ( 十一 ). 北京: 气象出版社, 1998: 33-40.
- [ 4 ] Shipilov O I, Kobskov B P, Abbas A. Statistical evaluation of cloud seeding operation in Syria ( 1991— 1993 ) [ R ]. 6th WMO Scientific Conference on Weather Modification. Italy Paestrum, 1994: 341-344
- [ 5 ] Abbas A, Mustafa A. Syrian rain enhancement project ( 1991— 1998 ) [ R ]. 7th WMO Scientific Conference on Weather Modification Thailand Chiang Mai 1999: 118-120.
- [ 6 ] Kobskov B P, Melnichuk Y V, Abbas A. Statistical estimation of cloud seeding operations in Syria ( 1991— 1996 ) [ R ]. 7th WMO Scientific Conference on Weather Modification Thailand Chiang Mai 1999: 161-164
- [ 7 ] 段 英, 赵亚民, 赵利品. 飞机人工增雨非固定目标作业效果评估方法 [ C ]. 人工影响天气 ( 十一 ). 北京: 气象出版社, 1998: 76-79
- [ 8 ] 严采蓁, 陈万奎. 统计评估人工增雨效果的一些问题研究 [ J ]. 应用气象学报, 2001, 12( 增刊 ): 81-87.
- [ 9 ] 叶家东, 李铁林. 区域趋势控制回归变量回归分析效果评估方法研究 [ J ]. 气象科学, 2001, 21( 1 ): 64-72
- [ 10 ] 杨景梅, 邱金桓. 我国可降水量同地面水汽压关系的经验表达式 [ J ]. 大气科学, 1996, 20( 5 ): 620-626
- [ 11 ] 杨景梅, 邱金桓. 地面湿度参数计算我国整层大气可降水量及有效水汽含量方法的研究 [ J ]. 大气科学, 2002, 26( 1 ): 19-22.

## Application of Cluster Analysis to the Statistical Assessment of the Effect of Artificial Rain Enhancement

FANG Bin<sup>1, 2</sup>, XIAO Hui<sup>2</sup>,

WANG Zhen-hui<sup>1</sup>, SUN Hai-yan<sup>1</sup>, HUANG Mei-yuan<sup>2</sup>

( 1. Department of Electronic Engineering NU IST, Nanjing 210044 China

2. Institute of Atmospheric Physics Chinese Academy of Science, Beijing 100029 China)

**Abstract** In this paper a new method of evaluating the efficiency of non-randomized artificial precipitation enhancement—Cluster Analysis-based Floating Control historical regression Method (CA-FCM) is presented and compared with the traditional Floating Control historical regression Method (FCM) which does not base on cluster analysis. It is shown from six examples of effect evaluations for cloud seeding operation cases carried out in the central region of Henan Province that the CA-FCM can more effectively evaluate the efficiency of cloud seeding operations than the FCM, because the CA-FCM adopts cluster analysis which highly improves the correlativity between rainfall distributions in the control area and target area and chooses the atmospheric precipitable water as a covariant, thus improving the inferential accuracy of natural rainfall in the cloud seeding operation area.

**Key words** cluster analysis, floating control method, historical regression, atmospheric precipitable water, cloud seeding effect