

人工降水方案统计设计的统计数值 模拟方法研究

曾光平

张长安

李茂仑

(福建省气象局, 福州 350001)

(吉林省人工降雨防雹办公室, 长春 130062)

摘 要 利用统计数值模拟方法, 通过对各种试验统计方案的功效、准确度和灵敏度的分析, 选择适宜的统计方案、确定试验周期和选择影响效果评价各种因子的最佳值。为人工降水方案统计设计提供可靠的科学依据。分析表明: 在各种随机化试验中, 区域随机交叉试验的功效、准确度和灵敏度高于其他随机化试验; 多因子分层回归分析方案是一种功效、准确度和灵敏度均较高的非随机化试验方案。

关键词: 人工降水; 方案统计设计; 统计数值模拟

1 引言

为了有效地进行人工降水试验研究和业务性抗旱蓄水作业, 必须针对项目的具体目的进行科学的严格的方案设计, 人工降水方案包括物理方案和统计方案二方面, 因此人工降水方案设计包括方案的物理设计和方案统计设计二大方面。

方案的物理设计的目的是使试验或作业的原理、对象、方法和效果在物理上是合理的, 有明确的物理基础。物理设计内容包括提出试验的原理假设(催化原理), 确定合适的试验对象, 科学的催化体制和物理学检验效果方法, 最后还要给出预期的试验效果。

方案的统计设计的目的是要用统计学方法对试验或作业的效果作出客观、定量、科学的评价。通过对资料的收集和分析, 使得原先的假设(预期的效果)能科学地加以检验和证实, 从而得出合理的结论。试验方案的统计设计要针对以下三个问题开展研究:

- (1) 选择试验研究(或外场作业)的统计方案。
- (2) 确定试验研究(或外场作业)的周期(或样本容量)。
- (3) 确定影响方案统计设计的科学性和准确性的各种因子(如区域雨量相关性、样本容量、统计变量变换等)的最佳取值。

迄今几乎所有的大型外场试验研究项目和业务作业计划都把相当长的时间和大量的精力用于方案的统计设计。如著名的以色列人工降雨随机试验^[1]、WMO的“PEP”计划^[2]以及中国古田人工降雨回归试验^[3~5]等都对试验统计方案进行严密的设计。

试验的统计设计中关键的问题是准确地确定自然降水量的估计值。为此, 统计学家和物理学家提出了随机化试验方案和非随机化试验方案两大类, 每类中又有几种不同的

统计方案。由于不同地区天气气候、地理环境、降水概型等差异极大,因此在方案设计时必须进行大量的模拟分析。

采用统计数值模拟方法分别利用福建省古田水库人工降雨试验资料和福建省抗旱业务性人工降雨资料,对随机化试验和非随机化试验方案的统计设计进行研究分析。

2 基本思路和方法

由于影响降水自然起伏的众多因素无法在参量性统计检验中反映出来,使得很难对上述统计设计中必须解决的三个问题给予直接的回答。为了解决人工降雨方案设计中的这些问题,采用统计数值模拟方法通过各种随机试验和非随机试验的方案进行功效、准确度和灵敏度^[4]的分析,选择适宜的统计方案,确定试验周期(样本容量)和选择影响统计方案统计设计的科学性的各种因子的最佳取值。

研究试验的功效、准确度和灵敏度的方法有两种:一种是经典的统计理论分析方法;另一种是统计数值模拟试验方法,本文采用统计数值模拟方法中的自然复随机化试验法^[6]。

本文在研究中三个量分别取以下值:

功效 $p \geq 0.80$;

准确度 相对增值的准确度 $\xi \geq 80\%$ (即失真率 $\eta \leq 20\%$);

灵敏度 (即在一定的相对增雨效果 θ 下的统计检验显著度), $\alpha \leq 0.05$ 。

模拟试验时,从长序列的历史资料中随机抽取一组容量为 k 的样本作为“催化”样本(或任意指定历史资料中一段为“催化”样本),其余为“对比”样本。采用不同的统计方法通过分析试验的功效、准确度和灵敏度,选择适宜的统计方案,确定试验的周期,以及影响统计方案设计科学性的因子的最佳取值。

选取的催化效果 θ 为 0%、10%、20%、30%、50% 和 100%,各催化效果在各自的试验中保持定常。 $\theta = 0\%$ 表示催化效果为零的情形,此时“效果”统计量即是降水自然起伏所引起的“噪声”(或称作假效果)。

3 随机化试验方案的统计设计

研究与生产相结合的项目以及纯研究的项目其统计方案一般采用随机化试验方案。常用的随机化方案有随机序列试验(单区随机试验)、区域随机回归试验、区域随机交叉试验(又称审渡试验)以及区域随机交叉回归试验。利用古田人工降雨试验区 1975~1996 年 4~6 月(试验单元为 3 h)自然降水样本,采用统计数值模拟方法分析这些随机试验方案的功效、准确度和灵敏度,在此基础上选择适宜的统计方案,确定试验周期以及影响因子的最佳取值。

3.1 随机化试验统计方案的选择

通过对不同的随机化试验的功效、准确度和灵敏度分析,选择适合的统计方案。表 1 是各种随机试验方案的功效、准确度和灵敏度数值试验的结果,表中 N 为随机试验

表 1 各种随机试验方案的功效、准确度和灵敏度的数值试验结果 (N 表示样本容量)

方案	θ (%)	N=50			N=100			N=150			N=200		
		p	η (%)	α	p	η (%)	α	p	η (%)	α	p	η (%)	α
J_1	10	0.19	200.75	0.15	0.23	187.89	0.12	0.34	136.05	0.10	0.42	119.63	0.07
	20	0.33	122.43	0.10	0.43	102.48	0.09	0.51	77.11	0.08	0.59	65.25	0.06
	30	0.40	79.41	0.08	0.54	69.56	0.07	0.68	56.34	0.06	0.70	47.13	0.05
	50	0.62	60.32	0.07	0.71	51.24	0.06	0.82	46.42	0.05	0.93	32.63	0.03
	100	0.91	40.11	0.05	0.96	34.60	~0.04	0.98	26.21	0.04	1.00	21.75	0.02
J_2	10	0.24	120.21	0.09	0.39	62.87	~0.08	0.45	54.13	0.07	0.58	45.90	0.06
	20	0.52	63.14	0.07	0.73	34.29	~0.06	0.83	29.07	~0.05	0.97	25.03	0.04
	30	0.71	51.33	0.06	0.98	24.77	~0.05	1.00	22.00	0.04	1.00	18.08	0.03
	50	0.90	34.41	0.06	1.00	17.15	0.05	1.00	14.31	0.03	1.00	12.52	0.01
	100	1.00	21.54	0.05	1.00	11.43	0.03	1.00	9.99	0.02	1.00	8.34	<0.01
J_3	10	0.31	88.33	0.10	0.45	54.23	0.09	0.58	46.76	0.08	0.74	37.95	0.07
	20	0.61	44.16	0.08	0.83	29.58	0.07	0.91	25.36	0.06	0.99	20.70	~0.05
	30	0.80	39.42	0.07	0.99	21.36	0.06	1.00	18.32	0.05	1.00	14.95	0.04
	50	0.97	21.23	0.06	1.00	14.79	0.05	1.00	12.11	0.04	1.00	10.35	0.02
	100	1.00	15.38	~0.05	1.00	9.86	0.04	1.00	7.57	0.01	1.00	6.90	0.01
J_4	10	0.42	63.11	0.08	0.68	35.35	~0.07	0.76	32.13	0.06	0.87	29.58	0.05
	20	0.78	31.56	~0.05	0.96	22.19	0.05	0.99	18.42	0.04	1.00	15.04	0.03
	30	0.87	29.98	0.04	1.00	16.14	0.03	1.00	14.99	0.02	1.00	11.87	0.01
	50	1.00	18.45	0.02	1.00	10.10	0.01	1.00	8.73	0.01	1.00	7.52	<0.01
	100	1.00	10.11	0.01	1.00	6.06	<0.01	1.00	5.41	<0.001	1.00	5.01	<0.001

注: J_1 为随机序列试验, J_2 为区域随机回归试验, J_3 为区域随机交叉试验, J_4 为区域随机交叉回归试验 (下同)

总样本容量 (包括催化样本和非催化样本)。

数值试验结果表明: 四种随机试验方案中, 区域随机交叉回归试验的功效、准确度和灵敏度最高, 如 $N=50$, $\theta=20\%$ 时其功效已接近 0.80, 区域随机交叉试验次之, 达 0.61, 随机序列试验才 0.33; $\theta=30\%$ 时区域随机交叉回归试验已达 0.87, 区域随机交叉试验也达 0.80, 其他两种方案仍未达到 0.80。区域随机交叉回归试验准确度 ξ ($\xi=1-\eta$), 当 $N=100$, $\theta=20\%\sim 30\%$ 时, 达 78%~84% 以上; 随机交叉试验次之, 准确度为 70%~79%; 随机序列试验最低, 低于 30%。区域随机交叉回归试验灵敏度, 当 $N=50$, $\theta=20\%$ 时, α 已接近 0.05, 灵敏度很高; 区域随机回归试验次之, 当 $N=100$, $\theta=30\%$ 或 $N=150$, $\theta=20\%$ 时, α 显著度也达 0.05; 随机序列试验最差, $N=50$ 时, 增雨相对值要达 100% 时才满足显著度等于 0.05 的要求, 若 $\theta=30\%$, 则样本容量要达 200。

上述结果表明, 这四种随机试验方案中区域随机交叉回归试验是一种较好的方案, 在人工降水可能的增雨范围内容易满足统计方案设计的要求。区域随机交叉试验和区域随机回归试验次之, 也是一种可行的统计方案。

3.2 试验研究周期 (样本容量) 的确定

以 $p\geq 0.80$ 、 $\xi\geq 80\%$ (或 $\eta\leq 20\%$)、 $\alpha\leq 0.05$ 为标准, 分析区域雨量相关系数 $r=0.85$ 时, 在人工降水可能相对增雨值范围内 ($\theta=20\%\sim 30\%$), 不同随机试验方案所需的样本容量 (即试验研究周期)。

综合考虑功效、准确度和灵敏度,单区随机试验所需的样本容量大于 300 个,可见单区随机试验方案的试验周期过长,很难采用。区域随机交叉回归试验所需的样本容量是区域随机回归试验的一半。人工催化效果一般在 20%~30%之间,对古田水库人工降雨试验区,当 $r=0.85$ 时随机交叉回归试验仅需 150~90 个试验单元,远小于其他随机试验方案所需的试验周期。区域随机回归试验和区域随机交叉试验试验周期大致相当。由于区域交叉试验采用非参量性的秩和检验法检验效果的显著性,其灵敏度较低,如能提高区域随机交叉试验的灵敏度,则有望大大缩短其试验周期。

3.3 影响方案的统计设计科学性因子的分析

影响随机化试验方案的统计设计科学性因子有增雨效果、样本容量和区域降水量相关性等,不同方案对这些因子的反应程度各有所异。

3.3.1 增雨效果对方案统计设计的影响

分析表 1 可以看出增雨效果显著地影响方案的统计设计的科学性。

(1) 在相同的增雨效果下,区域随机交叉回归试验的功效、准确度和灵敏度高于其他三种随机试验方案。区域随机交叉试验的功效、准确度次之,只是灵敏度低于区域随机回归试验;随机序列试验的功效、准确度和灵敏度最低。

(2) 无论哪一种方案,当 N 一定时,各种统计方案的功效、准确度和灵敏度都随着增雨效果的提高而提高。

3.3.2 样本容量对方案统计设计的影响

随机试验的样本容量指随机试验的催化样本和非催化样本容量之和。从表 1 可以看出:

(1) 在相同的样本容量下,区域随机交叉回归试验的功效、准确度和灵敏度最高;区域随机交叉试验的功效、准确度次之,但其灵敏度却低于区域随机回归试验(这是由于区域随机交叉试验的显著性检验采用非参量性的秩和检验法,而区域随机回归试验采用 t 检验法,后者检验灵敏度较高,下同);随机序列试验的功效、准确度和灵敏度最低。

(2) 各种统计方案的功效、准确度和灵敏度都随样本容量的增大而提高,但是提高的速度不一样。区域随机交叉回归试验提高最快,而且很快趋于稳定值,随机序列试验功效、准确度和灵敏度随样本容量变化最为缓慢。

3.3.3 区域降水量相关性对方案统计设计的影响

人工降水是人们试图通过对云体发展过程中某一环节进行有意识的干预而达到增雨目的,因此对地面降水量的资料收集、分析是方案统计设计的极为重要部分。目标区和对比区的区域降水量相关性是直接影响统计方案设计科学性极为重要的因子。以下采用区域随机回归试验方案通过对 5 组区域降水量相关性不同的资料($r_1=0.50$, $r_2=0.70$, $r_3=0.85$, $r_4=0.89$, $r_5=0.98$)的分析来探讨区域降水量相关性对效果评价的功效、准确度和灵敏度的影响。

表 2、表 3、表 4 分别是不同区域降水量相关条件下区域随机回归试验方案功效、准确度和灵敏度的数值试验结果。图 1 给出了不同样本容量和作业效果时功效与区域降水量相关性的关系曲线。

分析表 2、表 3、表 4 和图 1,可以看出:

(1) 功效、准确度和灵敏度与区域雨量相关性关系极为密切,在样本容量和相对

表 2 不同区域降水量相关条件下区域随机回归试验方案的功效值

N	r	增雨效果 (%)				
		10	20	30	50	100
30	r_1	0.090	0.133	0.184	0.318	0.623
	r_2	0.098	0.190	0.299	0.594	0.859
	r_3	0.148	0.291	0.443	0.742	0.996
	r_4	0.258	0.522	0.678	0.899	1.000
	r_5	0.533	0.937	1.000	1.000	1.000
100	r_1	0.132	0.230	0.345	0.616	0.958
	r_2	0.150	0.307	0.588	0.863	1.000
	r_3	0.388	0.730	0.981	1.000	1.000
	r_4	0.465	0.842	0.999	1.000	1.000
	r_5	0.853	1.000	1.000	1.000	1.000
200	r_1	0.164	0.446	0.655	0.861	1.000
	r_2	0.229	0.630	0.943	0.986	1.000
	r_3	0.578	0.966	0.991	1.000	1.000
	r_4	0.764	0.993	0.998	1.000	1.000
	r_5	0.993	1.000	1.000	1.000	1.000
320	r_1	0.196	0.623	0.897	0.950	1.000
	r_2	0.327	0.721	0.954	1.000	1.000
	r_3	0.757	0.998	1.000	1.000	1.000
	r_4	0.883	0.999	1.000	1.000	1.000
	r_5	1.000	1.000	1.000	1.000	1.000

表 3 不同区域降水量相关条件下区域随机回归试验方案失真率 η (%)

N	r	相对增雨效果 θ (%)				
		10	20	30	50	100
30	r_1	310.40	169.31	122.28	84.66	65.57
	r_2	221.61	120.88	87.50	60.44	43.34
	r_3	156.32	85.27	61.58	42.65	28.42
	r_4	107.66	59.94	46.98	30.71	20.56
	r_5	48.76	26.60	19.21	13.30	10.13
100	r_1	185.08	100.95	72.91	50.48	33.65
	r_2	106.70	69.11	42.97	30.55	20.04
	r_3	62.87	34.29	24.77	17.15	11.43
	r_4	53.60	28.51	20.82	15.89	9.93
	r_5	31.55	17.21	12.43	8.60	5.74
200	r_1	130.21	71.02	45.29	35.51	23.65
	r_2	90.09	49.14	22.30	20.57	16.38
	r_3	45.90	25.03	18.08	12.52	8.34
	r_4	42.30	23.31	16.55	11.37	7.34
	r_5	20.84	11.37	8.21	5.05	3.79
320	r_1	98.67	59.27	34.40	20.10	11.91
	r_2	64.06	30.40	20.78	14.71	8.98
	r_3	34.80	18.98	13.71	9.49	6.33
	r_4	30.11	16.63	11.99	7.87	5.38
	r_5	17.93	9.78	7.06	4.86	3.26

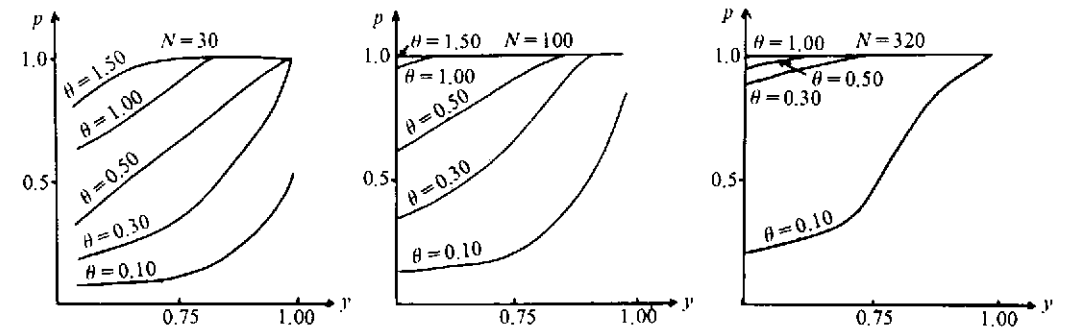


图 1 功效与区域降水量相关系数关系图

表 4 不同区域降水量相关条件下区域随机回归试验方案的灵敏度

N	r	相对增雨效果 θ (%)				
		10	20	30	50	100
30	r_1	>0.10	0.09	0.09	~0.08	0.08
	r_2	0.10	0.08	0.07	0.06	~0.07
	r_3	0.09	~0.07	0.07	~0.06	0.06
	r_4	~0.07	0.07	0.06	~0.05	~0.05
	r_5	0.06	~0.05	0.05	~0.04	0.04
100	r_1	0.10	~0.09	0.09	0.08	0.07
	r_2	~0.09	0.08	0.07	~0.06	0.06
	r_3	~0.07	0.07	~0.06	0.06	0.05
	r_4	~0.06	0.06	~0.05	0.05	0.04
	r_5	~0.05	<0.05	~0.04	0.04	0.03
200	r_1	0.09	~0.08	0.08	~0.07	0.07
	r_2	~0.07	0.07	~0.06	0.06	0.05
	r_3	0.06	~0.05	<0.05	~0.04	0.04
	r_4	~0.05	0.05	0.04	~0.03	0.03
	r_5	0.04	0.04	0.03	0.03	0.02
320	r_1	0.08	0.07	0.06	0.04	~0.03
	r_2	~0.06	0.06	0.05	0.04	0.03
	r_3	0.05	~0.04	0.04	0.03	0.02
	r_4	~0.04	0.04	0.03	0.02	0.01
	r_5	~0.03	0.03	0.02	0.01	0.01

增雨不变情况下，随着区域降水量相关性的提高，功效、准确度和灵敏度也相应提高。

(2) 当相关系数 r 较小时，功效、准确度和灵敏度随 r 增大而提高十分明显，当相关系数 r 较大时，提高变得缓慢，即功效、准确度和灵敏度与相关系数的关系曲线在 r 较小时曲线较陡，随 r 增大曲线趋于平缓。

3.3.4 影响方案统计设计因子的取值

从上面功效、准确度和灵敏度分析可以看出，增雨效果、样本容量和区域相关系数影响了方案统计设计的科学性。若以人工降水可能的增雨值为 20%~30%，可以确定功效 ≥ 0.80 、准确度 $\geq 80\%$ 、灵敏度 $\alpha \leq 0.05$ 下影响统计方案设计因子的最佳取值。表 5 列出不同样本容量情况下区域降水量相关系数的取值。

从表 1 和表 6 可以看出：

表 5 不同试验方案二区降水量相关系数取值

增雨值相关系数	20.0%					30.0%				
	50	100	150	200	250	50	100	150	200	250
J_2	>0.95	0.95	0.93	0.89	0.87	0.93	0.90	0.85	0.83	0.80
J_3	>0.95	0.95	0.93	0.89	0.87	0.92	0.89	0.85	0.83	0.79
J_4	0.95	0.90	0.85	0.75	0.65	0.90	0.85	0.74	0.67	0.55

表 6 各种非随机试验方案的功效、准确度和灵敏度的数值试验结果

	θ	$N=50$			$N=100$			$N=150$			$N=200$		
	(%)	p	$\eta(\%)$	α	p	$\eta(\%)$	α	p	$\eta(\%)$	α	p	$\eta(\%)$	α
序列 试验	10	0.14	223	>0.5	0.20	241	>0.5	0.29	198	>0.4	0.38	139	>0.5
	20	0.25	176	>0.5	0.38	188	>0.4	0.44	157	>0.4	0.51	110	>0.4
	30	0.34	155	>0.3	0.47	143	>0.4	0.49	120	>0.3	0.58	95	>0.3
	50	0.55	81	>0.2	0.63	70	>0.2	0.71	67	>0.2	0.77	42	>0.15
	100	0.84	47	~0.1	0.86	43	~0.1	0.90	31	~0.1	0.94	28	~0.1
区域 对比 试验	10	0.08	175	>0.6	0.01	168	>0.5	0.07	110	>0.5	0.05	99	>0.5
	20	0.10	134	>0.5	0.13	110	>0.4	0.19	88	>0.5	0.21	72	>0.4
	30	0.22	77	>0.5	0.32	65	>0.3	0.41	43	>0.4	0.40	40	>0.4
	50	0.37	52	>0.3	0.41	39	>0.3	0.52	31	>0.3	0.49	27	>0.3
	100	0.63	37	>0.2	0.74	30	>0.2	0.81	25	>0.3	0.78	22	>0.2
区域 回归 试验	10	0.17	76	>0.3	0.28	58	>0.2	0.39	43	>0.2	0.46	39	>0.1
	20	0.35	55	>0.2	0.43	32	>0.1	0.54	30	>0.1	0.77	26	~0.1
	30	0.66	42	>0.2	0.75	29	>0.1	0.84	25	~0.1	0.97	20	~0.05
	50	0.88	23	>0.1	0.92	21	~0.1	1.00	17	~0.1	1.00	18	<0.05
	100	1.00	18	~0.05	1.00	16	>0.05	1.00	15	>0.05	1.00	13	<0.01
多因子 分层回 归试验	10	0.31	40	>0.5	0.49	37	>0.3	0.57	29	>0.20	0.97	19	~0.1
	20	0.58	31	~0.1	0.78	26	~0.05	0.88	20	~0.05	1.00	16	<0.05
	30	0.89	25	<0.05	0.99	19	<0.05	1.00	16	<0.05	1.00	14	<0.05
	50	1.00	17	<0.05	1.00	15	<0.05	1.00	11	<0.05	1.00	9	<0.05
	100	1.00	9	<0.05	1.00	7	<0.05	1.00	7	<0.05	1.00	6	<0.05

(1) 在方案统计设计过程中, 二区降水量相关性是个非常敏感的因子。无论是随机回归试验还是随机交叉回归试验, 在功效、准确度和灵敏度给定的情况下, 二区相关性提高使试验的周期很快缩短。因此在人工降水方案设计时选择合适的影响区和目标区是试验成败的关键条件之一。因此在方案统计设计时, 要根据人工降水可能增雨值、二区相关性和试验周期的关系确定他们的取值范围。

(2) 在二区相关性相同的情况下, 随机交叉回归试验试验周期要远短于随机回归试验。

(3) 虽然随机交叉试验的功效和准确度高于随机回归试验, 但是由于该方案灵敏度分析是采用非参量的秩和检验致使其灵敏度低于随机回归试验, 因而在综合考虑功效、准确度和灵敏度时这两种方案基本上无多大差异。

4 非随机化试验方案的统计设计

通常的非随机试验方案有顺序试验、区域对比试验、区域回归试验和多因子分层回

归试验。采用统计数值模拟方法,分析这些方案的功效、准确度和灵敏度,在此基础上选定试验方案并确定影响因子的取值。

多因子分层回归试验是选择一些与降水过程有密切关系的物理现象或天气现象,通过分层建立回归关系。由于探测手段局限,福建省夏季人工降雨时只考虑了天气形势、降水类型和 12 h 雨强三项因子。根据人工降雨季节天气和降水特征,天气形势分为:副热带高压控制(含副热带高压边缘)和锋面系统二类;降水类型分为对流性降水、混合型降水和稳定型降水三类;降水强度分为小于 2 mm/12 h、2~10 mm/12 h 和大于 10 mm/12 h 三类。考虑到样本容量和自然降水的稳定性,取锋面系统影响下的稳定性降水,雨强区间取 2~10 mm/12 h。由于其他分层取值回归时样本容量少,降水起伏大,二区区域雨量相关性差,导致对比单元与催化单元条件相似的假设的依据不充分,回归分析的功效、准确度和灵敏度很低,使回归方程代表性很差,回归有效性差。因此本文对其他分层暂不作分析。

模拟试验资料是福建省 1961~1990 年 7~9 月自然降水资料。试验单元为 6 h,自然降水样本容量为 1360 个,区域雨量相关系数为 0.5981 ($\alpha < 0.01$)。

4.1 非随机化试验统计方案的选择

表 6 列出各种非随机试验统计方案数值模拟的试验结果。表中 N 为非随机化试验催化单元样本容量。

分析表 6 结果可以看出,虽然不同方案的功效、准确度和灵敏度随样本容量和增雨效果的变化各有特点,但还是可以得出四种非随机试验方案中,多因子分层回归试验的功效、准确度和灵敏度均高于其他方案。如 $N=50$, $\theta=0.30$ 时,多因子分层回归试验的功效 $p=0.89$,区域回归试验 $p=0.66$,其他两种方案 p 均小于 0.35;多因子分类回归试验失真率 $\eta=25\%$,区域回归试验次之为 42%,序列试验最差为 155%;多因子分类回归试验灵敏度 $\alpha < 0.05$,区域回归试验次之,区域对比试验最差 $\alpha > 0.5$ 。当 $N=100$, $\theta=30\%$ 时,多因子分类回归分析法的功效 $p=0.99$ 、准确度 $\xi=81\%$ 、灵敏度 $\alpha < 0.05$,已达到规定标准;区域回归试验 $p=0.75$ 、 $\xi=71\%$ 、 $\alpha > 0.1$,未达到规定标准,但较接近规定标准,其他两种方案甚至到 $N=200$, $\theta=100\%$ 时,还达不到规定标准。

上述结果表明四种非随机试验方案中,多因子分层回归试验在分层取值情况下其功效、准确度和灵敏度均高于其他三种方案,区域回归试验方案次之,序列试验和区域对比试验方案难以满足方案统计设计的要求。

4.2 试验研究周期(样本容量)的确定

表 7 列出以 $p \geq 0.80$ 、 $\xi \geq 80\%$ 、 $\alpha \leq 0.05$ 为标准,二区降雨量相关系数为 0.70 时,对应于不同增雨值各种非随机试验方案的试验周期。

从表 7 可以看出,在人工降水可能的增雨范围内(20%~30%),在分层取值情况

表 7 各种非随机试验方案试验的周期

相对增雨 (%)	10	20	30	50	100
序列试验	250	200	170	150	135
区域对比试验	210	180	130	100	80
区域回归试验	190	130	80	65	50
多因子分层回归试验	150	100	50	40	30

下, 多因子回归试验方案的试验周期最短, 催化作业样本数为 100~50, 区域回归试验次之为 130~80。其他两种方案试验周期更长。

4.3 影响方案统计设计科学性因子的分析

与随机化试验方案一样, 影响方案统计设计的科学性因子有区域降水量相关性、催化作业样本容量、对比单元样本容量及统计变量的变数变换等, 但主要是前二项。以下着重讨论区域降水量相关性和催化单元样本容量对方案统计设计科学性的影响, 采用统计数值模拟方案分析这些因子对多因子分层回归试验方案的影响, 并确定这些因子的最佳取值和试验周期。数值试验结果列于表 8、表 9。

表 8 $\theta = 20\%$ 时多因子分层回归试验数值模拟结果

区域降水量 相关系数	N=30			N=50			N=100			N=150		
	p	$\eta(\%)$	α	p	$\eta(\%)$	α	p	$\eta(\%)$	α	p	$\eta(\%)$	α
0.34	0.19	67	>0.5	0.27	59	>0.5	0.45	46	>0.1	0.59	40	>0.05
0.52	0.41	53	>0.2	0.47	42	>0.1	0.61	35	>0.05	0.72	27	~0.05
0.61	0.48	47	>0.1	0.54	33	~0.1	0.77	29	>0.05	0.91	23	<0.05
0.70	0.57	38	~0.1	0.72	27	<0.1	0.89	21	~0.05	0.99	17	<0.05
0.86	0.73	35	>0.05	0.84	21	<0.05	0.97	16	<0.05	1.00	09	~0.01

表 9 $\theta = 30\%$ 时多因子分类回归试验数值模拟结果

区域降水量 相关系数	N=30			N=50			N=100			N=150		
	p	$\eta(\%)$	α	p	$\eta(\%)$	α	p	$\eta(\%)$	α	p	$\eta(\%)$	α
0.34	0.21	61	>0.5	0.32	55	>0.5	0.77	38	>0.1	0.87	32	~0.05
0.52	0.44	48	>0.2	0.51	37	>0.1	0.93	30	<0.05	1.00	22	<0.05
0.61	0.68	39	~0.05	0.83	25	<0.05	0.99	17	<0.05	1.00	13	<0.01
0.70	0.77	32	<0.05	0.95	19	<0.05	1.00	11	<0.01	1.00	08	<0.01
0.86	0.81	24	<0.05	1.00	14	<0.01	1.00	07	<0.01	1.00	05	<0.01

分析表明:

(1) 区域降水量相关系数和样本容量明显地影响各种方案的功效、准确度和灵敏度, 并随区域降水量相关系数的提高和样本容量的增加而提高。

(2) 取 $p \geq 0.80$ 、 $\xi \geq 80\%$ 、 $\alpha \leq 0.05$, 对于人工降水可能增雨范围 (20%~30%), 多因子分层回归试验在分层取值情况下相应于 $r=0.86$ 、0.70 和 0.61 的催化作业单元分别为 50~30、100~50、150~100; 当 $r < 0.52$ 时, 无法满足方案统计设计的要求。

综上所述, 在分层取值的情况下, 多因子分层回归分析法较序列试验、对比试验和历史回归试验的功效、准确度和灵敏度均高。在人工降雨可能增雨范围内, 区域降水量相关系数 $r > 0.60$ 所需样本容量 150~100; 当 $r > 0.70$ 所需的样本容量为 100~50。

5 随机化试验和非随机化试验功效、准确度和灵敏度比较

为使试验方案的功效、准确度和灵敏度分析结果具有可比较性, 因此在对随机化试验和非随机化试验的功效、准确度和灵敏度进行比较时均采用古田人工降雨试验自然降水资料。随机化试验方案的数值试验结果见表 1, 非随机化试验方案的数值试验结果列

表 10 各种非随机试验方案的功效、准确度和灵敏度的数值试验结果

	θ (%)	$N=50$			$N=100$			$N=150$			$N=200$		
		p	$\eta(\%)$	α	p	$\eta(\%)$	α	p	$\eta(\%)$	α	p	$\eta(\%)$	α
序列 试验	20	0.28	144	>0.5	0.40	132	>0.3	0.47	101	~0.4	0.55	84	>0.2
	30	0.37	103	>0.2	0.49	84	>0.2	0.56	71	~0.2	0.63	60	=0.2
	50	0.57	76	~0.2	0.67	60	~0.2	0.74	53	~0.2	0.87	37	>0.1
区域对 比试验	20	0.21	136	>0.5	0.27	99	>0.5	0.35	74	>0.5	0.47	51	>0.5
	30	0.35	118	>0.5	0.43	62	~0.5	0.49	42	>0.5	0.68	39	>0.4
	50	0.48	73	>0.3	0.66	41	>0.3	0.75	27	~0.3	0.91	25	>0.3
区域回 归试验	20	0.40	76	>0.3	0.54	43	>0.2	0.71	31	>0.2	0.83	27	~0.2
	30	0.68	54	>0.2	0.77	37	~0.2	0.89	26	>0.1	1.00	23	>0.1
	50	0.88	39	~0.2	0.95	21	>0.1	0.99	20	~0.1	1.00	18	~0.1

于表 10 (表中催化效果只列出 $\theta=20\%$ 、 30% 、 50%)。

分析表 1 和表 10, 可以看出: 随机化试验方案的功效、准确度和灵敏度均高于对应的非随机化试验。

(1) 随机序列试验与序列试验相比较, 当 $\theta=20\%\sim 30\%$, 其功效值大 $3\%\sim 12\%$; 失真率小 $13\%\sim 30\%$; 灵敏度 α , 随机序列试验小于 0.1, 而序列试验均大于 0.1。

(2) 随机交叉试验与区域对比试验相比较, 当 $\theta=20\%\sim 30\%$ 时, 功效高 $32\%\sim 56\%$; 失真率小 $24\%\sim 92\%$; 灵敏度 α , 随机交叉试验为 $0.04\sim 0.1$, 区域对比试验为 $0.4\sim 0.5$ 。

(3) 当 $\theta=20\%\sim 30\%$ 时, 随机回归试验比区域回归试验功效高 $0\%\sim 21\%$; 失真率小 $2\%\sim 13\%$; 灵敏度 α , 随机回归试验为 $0.03\sim 0.07$, 区域回归试验为 $0.1\sim 0.3$ 。

6 结论

(1) 通过对功效、准确度、灵敏度的数值分析, 研究人工降水方案的统计设计, 结果表明: 随机化试验的功效、准确度和灵敏度均高于相应的非随机化试验。随机化试验中区域随机交叉回归试验的功效、准确度和灵敏度均高于其他三种随机化试验, 是随机化试验中最佳方案; 区域随机交叉试验 (窜渡试验) 次之, 随机序列试验 (单区随机试验) 最差。各种非随机化试验中多因子分层回归试验在分层取值的情况下的功效、准确度和灵敏度较高, 是非随机化试验中较好的方案, 区域回归试验次之。

(2) 区域降水量相关性、样本容量、增雨效果、变量变换均影响分析的功效、准确度和灵敏度, 其中区域降水量相关性最为敏感。因此在方案统计设计时应把大量精力放在前期调研上, 尽量选择相关性好的区域以提高试验的功效、准确度和灵敏度。

(3) 在人工降水可能的相对增雨范围内 ($\theta=20\%\sim 30\%$), 各种随机化试验方案中区域随机交叉回归试验的试验周期最短。古田人工降雨试验在 $r=0.85$ 条件下, 仅需 $150\sim 90$ 个试验单元。各种非随机化试验方案中, 多因子分层回归试验在分层取值情况下的试验周期小于其他非随机化试验方案, 福建省人工降雨多因子分层回归分析 (自然雨强为 $2\sim 10\text{ mm}/12\text{ h}$, 降水为稳定型, 天气形势为锋面系统), 对应于 $r>0.70$ 所需的样本容量为 $100\sim 50$ 。

(4) 多因子分层回归试验方案在国内外一些著名的人工降水试验(如福建省古田人工降水随机试验、以色列的人工降雨随机试验)中均已采用。分层取值范围要考虑到分层后样本容量、降水稳定性及二区相关性。合适的分层可以提高二区雨量相关性,从而提高方案的试验功效、准确度和灵敏度,提高回归分析的有效性。此外,这一方法可能会使某些分层取值内的样本相关性很差,回归分析功效、准确度和灵敏度很低,回归分析的有效性很低。所以在选用多因子分层回归方案时应采用统计数值模拟方法分析不同分层取值范围内样本的功效、准确度和灵敏度,并在方案统计设计中分别说明其功效、准确度和灵敏度。

(5) 由于区域间天气气候、地理环境、降水概型差异极大,各地区开展人工降水时应采用统计数值模拟方法,通过分析功效、准确度和灵敏度来选择适宜当地的统计方法、试验周期以及影响效果评价各种因子的最佳取值,为方案的统计设计提供依据。此外,由于自然降水时空分布十分复杂,且受众多因子的影响,因此应加强催化物理响应的研究,并与统计分析结合起来,使方案设计更科学有效。

参 考 文 献

- 1 W. N. 赫斯(王昂生、徐华英等译), 人工影响天气和气候, 北京: 科学出版社, 1985, 316~343.
- 2 气象科技情报所等, 人工影响天气新进展, 北京: 气象出版社, 1985, 38~48.
- 3 曾光平、吴章云, 人工降水, 福州: 福建省科学技术出版, 1997, 35~49.
- 4 Yeh. Chia-dong, Cheng Keming and Zeng Guangping, Randomized cloud seeding at Gutian, Fujian, China, *WMA*, 1982, 4.
- 5 Zeng Guangping, A Comprehensive Evaluation of the Effect of Artificial Rainfall in Gutian Reservoir Area, China, The Fifth WMO Scientific Conference on Weather Modification and Applied Cloud Physics, May, 1989.
- 6 叶家东、曾光平, 随机试验功效的数值分析, 气象学报, 1984, 42(1), 69~79.

Study on Statistic Numerical Simulation Method of Precept Statistic Design of Artificial Precipitation

Zeng Guangping and Zhang Chang'an

(Fujian Institute of Meteorological Science, Fuzhou 350001)

Li Maolun

(Jilin Artificial Precipitation Office, Changchun 130062)

Abstract By using of statistic numerical simulations, analysis of the efficacy, accuracy and sensitivity of several test statistic precept, a feasible statistic precept is chosen and the test cycle is ascertained. The best value of all kinds of factors influencing the effect estimate are selected. This provides reliable science data for precept statistic design of artificial precipitation. The result indicates that the efficacy, accuracy and sensitivity of area randomization across test is greater than those of other randomization tests. Multi-factorial delamination regression analysis precept is a non-randomization test precept, in which the efficacy, accuracy and sensitivity are greater.

Key words: artificial precipitation; precept statistic design; statistic numerical simulation