## SOME PROBLEMS IN EVALUATING CLOUD SEEDING EFFECTS OVER EXTENSIVE AREAS

GLENN W. BRIER, THOMAS H. CARPENTER, and DWIGHT B. KLINE

Institute for Atmospheric Sciences
Environmental Science Services Administration

## 1. Introduction

Recently we had occasion to examine a number of operational cloud seeding programs for possible effects of silver iodide ground seeding on the rainfall in areas downwind from the nominal target area. This preliminary report gives the results of the analysis performed to date on 16 operational programs in 11 project areas in eastern United States, involving a total of 62 seeded months.

Cloud seeding operations as objects for statistical evaluation are in contrast to most of the remaining papers in this symposium which are concerned with scientific experiments. However, we feel that the results obtained as well as some of the statistical techniques used may be of sufficient interest and importance to be discussed along with the numerous other problems in weather modification.

## 2. Analysis and results

A common procedure for evaluating the possible effects of cloud seeding on the precipitation in a target area is to compare the actual precipitation during a seeded period with that which would have been expected in the absence of seeding. This expected amount is determined from a regression equation describing the relationship between the average precipitation in the target area and the average precipitation in a nearby control area which presumably is unaffected by the seeding operation. This regression equation can be determined from a historical record during a period when no cloud seeding operations were known to have taken place. A measure of the effectiveness of cloud seeding is the anomaly d, the difference between the actual precipitation (Y) in the target area and that expected  $(\hat{Y})$  if no seeding had been carried out, that is,

$$(1) d = Y - \hat{Y}.$$

Since our concern was more with the pattern of anomalies outside and beyond the nominal target area than with the average over the target, we decided in advance of any data processing to modify the usual procedure by determining precipitation anomalies d for individual stations both inside and outside the original target. Since the control areas were usually to the south or west of target area, we decided to examine the precipitation at stations lying in a semicircle located generally to the east or northeast of the center of the target area. In some cases it was possible to include in this extended target area stations as far as 150 or more miles from the target center.

A number of difficulties arose at this stage that produced a lack of uniformity in the data available for analysis of the 11 project areas. The map in figure 1

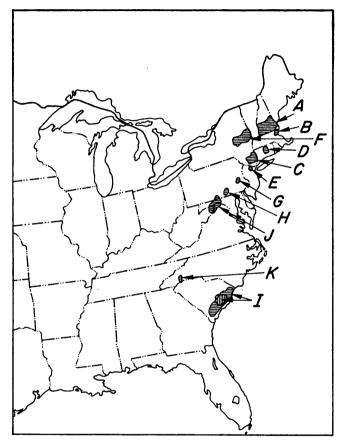


FIGURE 1
Location of projects evaluated.

showing the location of these project areas indicates some of these problems. It is seen that some of the areas are near coast lines and target areas extended downwind would be over the oceans where no data are available. The extended target areas also differed widely in the number and distribution of raingages and the length of historical records available for climatological analysis. The 16 programs

also differed in respect to their duration, some of the seeding operations running for periods as short as one month or less. It was recognized in advance that the noise level for individual projects might be high and some kind of a composite picture would be necessary for meaningful results. We felt that by combining all the projects there might be sufficient data to indicate in a general way how precipitation anomalies were related to distance from the nominal target. An examination of the climatic records for continuity of data, completeness, and so forth, resulted in a total of 784 stations for analysis. These were distributed among the various projects as shown in the third column of table I.

TABLE I

SUMMARY OF PROJECTS, NUMBER OF MONTHS AND STATIONS USED, FOR THE
ANALYSIS OF PRECIPITATION IN THE NOMINAL AND EXTENDED TARGET AREAS

Project	Locale	Number Seeded	of Months Random	No. of Stations	Station Seeded	Months Random
A	New Hampshire	2	2	36	72	72
В	Massachusetts	1	1	12	12	12
$\mathbf{c}$	Southeastern New York	1	1	39	39	39
$\mathbf{D}$	Connecticut	1	1	82	82	82
${f E}$	New Jersey	3	3	84	252	252
${f F}$	Eastern New York	1	1	121	121	121
$\mathbf{G}$	Eastern Pennsylvania	${f 2}$	${f 2}$	104	208	208
$\mathbf{H}$	Southern Pennsylvania	10	10	102	1020	1020
I	South Carolina	5	5	<b>54</b>	270	270
J	Mid-Potomac	24	3	62	1488	186
$\mathbf{K}$	North Carolina	12	12	88	1056	1056
	Total	<b>62</b>	41	784	4620	3318

In order to make the analysis comparable to some others that had been reported, the first step was to make a cube root transformation on all the raw precipitation data, using monthly totals for individual stations and the average for the stations in the control areas. It should be understood that all further discussion is in terms of data that have undergone a cube root transformation.

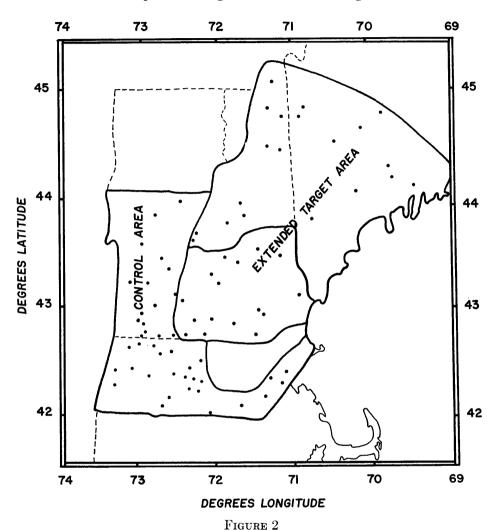
The next step in the statistical analysis was to derive for each project a regression equation for each station in the target and extended target areas. This equation was of the form

$$(2) Y = a + bX$$

where X is the mean monthly precipitation total for stations in the control area and the coefficients a and b were determined from the historical data using the observed monthly precipitation totals. Figure 2 illustrates the distribution of the stations in the control and target areas for Project A (New Hampshire). Figure 3 illustrates a scatter diagram showing a typical regression line and the transformed data (Bradford, New Hampshire).

In addition to deriving the regression coefficients for each station, the correla-

tion coefficient r was computed. Maps of  $r^2$  were then plotted, as shown in figure 4. In this figure the value of  $r^2$  for each station of Project A is plotted in terms of station coordinates expressed in degrees of latitude and longitude from the center



Location of stations used in control, target and extended target areas for Project A.

of the nominal target area. This procedure was followed for all the projects to facilitate the comparison and combining of results. The maps similar to those shown in figure 4 were helpful in detecting possible errors and in addition have an interest from a meteorological or climatological viewpoint.

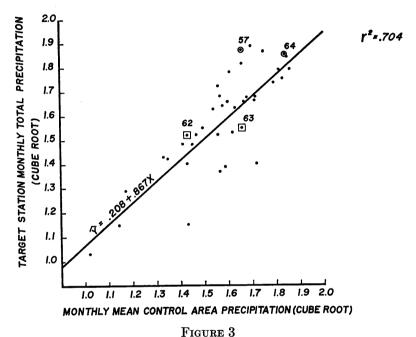
After the regression equations were computed it was possible to determine for each station and for each seeded month an anomaly

$$(3) d = Y - \hat{Y},$$

where Y is the observed rainfall index and  $\hat{Y}$  is the rainfall estimated from the regression equation. Also determined was the statistic

$$t = \frac{d}{E}$$

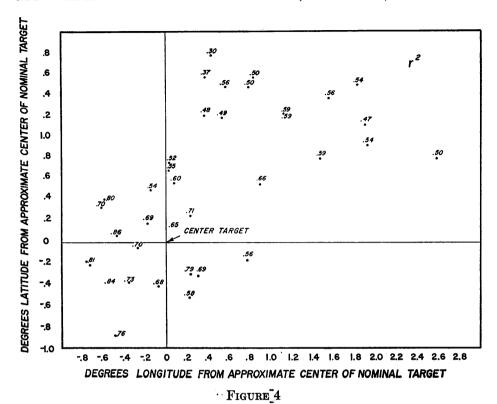
where E is the error estimate of d as given by classical regression theory. These



Regression line and data for typical station in Project A (Bradford, N. H).

values were computed for every station in each project for each seeded month. Thus, the basic experimental unit used here was a project month; in those cases when the seeding operation ran for only a fraction of a month an adjustment was made to make it comparable to an entire month.

Our primary objective was to determine how the anomalies d varied with distance from the target center. This was done by classifying all the stations in a project according to distance, using intervals of  $0.4^{\circ}$ . This class interval was a compromise since a smaller interval left too many classes with small numbers of stations and use of a larger class interval presented the danger of smoothing



Plot of  $r^2$  for stations in Project A.

out details that might be of interest or importance. These various distance classes were called "rings" and defined as follows:

Ring	Distance in Degrees
1	0.000 to 0.399
<b>2</b>	0.400 to 0.799
3	0.800 to 1.199
4	1.200 to 1.599
5	1.600 to 1.999
6	2.000 to 2.399
7	2.400 to 2.799
8	2.800 to 3.199
9	3.200 to 3.599
10	3.600 to 3.999
11	4.000 and greater.

The results of averaging the d anomalies by ring, denoted here as D, are shown in table II according to project months. The last column in the table shows the average anomaly D for the project month, giving equal weight to each ring. The

TABLE II  $\begin{tabular}{ll} Average Anomalies $D$ According to Distance Rings for the \\ 62 Seeded Project Months \\ \end{tabular}$ 

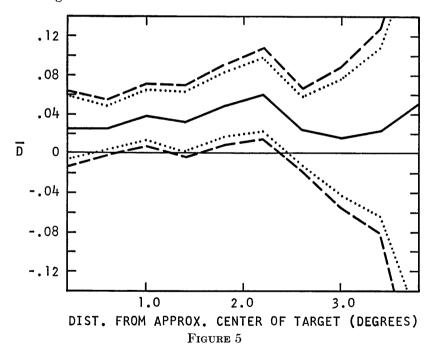
Proj-		V	1	2		4	Ri	ngs	7		0	10	11	Mea
ect	Month	Year	1		3	4	5	6	7	8	9	10		
A	Nov.	1964	.05	.10	.20	.12	.15	.09	.16					.1
A B	Nov. Nov.	1957 1964	.13 .04	.12 .02	.14 .06	.11	.15	.14	.17					.1 .0
č	Aug.	1964	.08	.08	.09	.09								<i>.</i> :
D	Aug.	1964	.04	.05	04	03	04	01						0
E	Oct.	1964	.09	.08	.12	.14	.12	.07	.09					.1
E	Nov. Dec.	1964 1964	$12 \\ .05$	12 .03	09 .05	07 .07	01 .06	06 .05	04 .14					). — ).
F	Jul.	1964	09	.04	.18	.19	.13	.03	.19	.27	.28			
G	Aug.	1954	.12	.12	.13	.21	.24	.33						.!
G H	Jul. Sep.	1955 1957	00 $21$	06 $27$	.01 02	.19 03	.38 .00	.58 13	04					! !
H	Oct.	1957	04	02	02	03	13	13 11	08					0
H	Nov.	1957	00	02	03	01	05	06	10					(
H	Dec.	1957	01	.01	.00	.02	.03	.05	.07					).
H H	Jan. Oct.	1958 1954	.08 .10	.08 .08	.07 .00	00 $03$	.01 14	.03 —.19	02					). - ,
Ĥ	Nov.	1954	.01	.08	.11	.13	.13	.33	.38					
H	Oct.	1963	29	23	12	24	19	25	18					:
H H	Nov.	1963	.05	.01	.03	.00	.02	.01	01					). ). —
I	Dec. Feb.	1963 1957	01 .14	02	04 .10	06 .13	10 .08	06 .28	09	.33	.33	.37	.35	
Î	Mar.	1957	.08	.05	.04	.11	.02	07	01	.09	.09	01	.05	
I	Apr.	1955	03	09	05	.09	.25	00	14	49	57	32	36	
Ī	May	1955	02	01	.04	.02	07	.13	05	04	.10	08	05	9
I J	Jun. Jun.	1955 1957	02	02	05	.10 10	03 09	00 $10$	.09 03	09	22	.29	.24	-:
J	Jul.	1957	02 06	02	10	10 07	03 11	21	14	13	32			_:
J	Aug.	1957	.02	.06	.31	.39	.32	.28	.15	.12	.11			.:
J	Jun.	1958	.10	.15	.24	.11	.03	.12	.08	06	12			
J J	Jul. Aug.	1958 1958	05 .01	04 $01$	09 .02	.01 .12	.00 .14	.15 .23	.10 .09	02 $00$	01 .00			). ).
j	Jun.	1959	.21	01 .24	.23	.14	.23	.11	.13	08	.30			
J	Jul.	1959	.26	.04	.09	.16	.19	.20	.10	.22	.15			
Ĵ	Aug.	1959	.04	.10	.12	.14	.14	.15	.20	.17	.35			
J J	Jun. Jul.	1960 1960	.10 07	.23 01	.05 .08	.01 .22	.04 .29	05 .36	.00 .30	$.02 \\ .24$	$21 \\ .34$			
Ĵ	Aug.	1960	.26	.26	.31	.34	.16	.12	.13	.17	.16			
J	Jun.	1961	.10	.06	.06	06	.09	.14	.10	05	11			
Ĵ	Jul.	1961	.10	06	01	07	.21	.06	.18	.23	.46			_:
J J	Aug. Jun.	1961 1962	34 $01$	18 .06	10 .03	.01 .03	05 .16	04 .17	.00 .06	03	.08 08			<u> </u>
Ĵ	Jul.	1962	.05	.01	05	12	25	15	19	16	23			
J	Aug.	1962	27	07	22	20	.10	.20	.13	.28	.27			
J	Jun.	1963	.21	.24	.21	.18	.19	.14	10	16	17			_:
J J	Jul. Aug.	1963 1963	08 03	18 .03	16 .06	30 .08	18 $03$	18 .08	05 $02$	02 $.00$	.15 10			
Ĵ	Jun.	1964	16	14	32	28	16	20	22	18	16			
J	Jul.	1964	.03	.09	.02	19	.10	.18	.02	10	.33			
J	Aug.	1964	10	09	01	13	17	17	16	31	60			-:
K K	Oct. Nov.	1954 1954	.00 .24	.09 .14	.21 .10	.19 .04	.47 .04	.34 00	01 $18$					
K	Dec.	1954	.04	02	01	00	01	00 .04	.06					:
K	Jan.	1955	00	.02	02	.07	.09	.04	.01					
K	Feb.	1955	.14	.17	.11	.10	.12	.01	09					
K K	Mar. Apr.	1955 1955	13 $06$	09 $13$	.14 05	.02 03	01 07	.32 .18	.43 .26					
K	Apr. May	1955	06 .25	13	03	03	.03	07	24					
K	Jun.	1955	.10	.01	00	08	08	10	.09					_
K	Jul.	1955	.11	.09	19	09	16	16	11					-
K K	Aug. Sep.	1955 1955	29 .06	07 .09	01 .06	03 01	07 .15	05 .04	01 03					-
	Sep. Mean 1		.02	.03	.04	.03	.05	.04	.03	.01	.02	.05	.05	

TABLE III

Number of Stations Used in Precipitation Analysis for Each Project
According to Distance Class

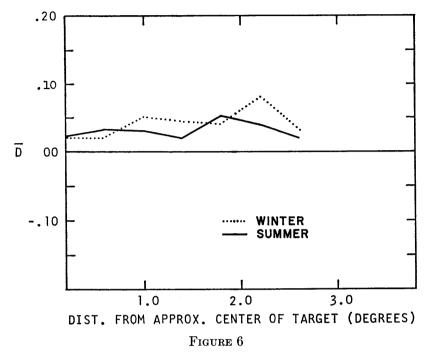
	Ring													
Project	1	2	3	4	5	6	7	8	9	10	11	Total		
A	6	14	2	3	6	4	1					36		
${f B}$	4	4	4									12		
$\mathbf{C}$	8	9	14	8								39		
$\mathbf{D}$	8	16	20	20	13	5						82		
${f E}$	11	10	14	14	16	13	6					84		
$\mathbf{F}$	5	7	14	18	23	15	14	20	5			121		
$\mathbf{G}$	8	8	33	37	15	3						104		
$\mathbf{H}$	5	12	17	26	21	14	7					102		
1	2	7	7	8	3	5	9	<b>2</b>	1	3	7	<b>54</b>		
J	3	8	7	6	8	6	13	9	<b>2</b>			62		
$\mathbf{K}$	8	19	13	17	20	9	<b>2</b>					88		
Total	68	114	145	157	125	74	<b>52</b>	31	8	3	7	784		

last line in the table shows the average for each column or ring. Table III shows the number of stations in each ring and it is noted that the projects differ widely in terms of the number of stations analyzed and the distance of these stations from the target center.



Average anomaly D for 62 seeded months according to distance from center of nominal target.

The averages at the bottom of table II are shown plotted in figure 5 along with the 90 and 95 per cent confidence limits as determined from the variances computed from the project month values given in table II. Figure 6 shows the average



Average anomaly D according to distance from center of the nominal target and winter and summer seasons.

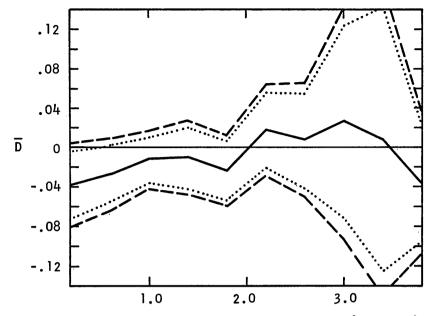
according to the winter and summer seasons. There are indications of positive anomalies, some of them statistically significant, extending to distances of approximately 150 miles.

One might suspect a hidden source of bias or even computational or programming errors. To give some protection against these possibilities, the original plan of the investigation included a set of random or fictitious months chosen for each project for use as a statistical control. Forty one such unseeded months were chosen from periods in which no seeding took place, and independent of the historical data from which the regressions were derived. The results for these random or unseeded months, analyzed in the same way as the operational seeded months, are presented in table IV. Figure 7 shows the averages according to distance from the center of the nominal target together with the 90 and 95 per cent confidence limits. Except for Ring 1, which has fewer stations than Rings 2–5, there is no suggestion of significant departure from the expected. In any case, the suggestion is that any bias in the analysis is on the negative side, which would make the detection of any real positive effects of seeding more difficult.

Proi							Rings							Mean
ect		Year	1	2	3	4	5	6	7	8	9	10	11	D
A	Nov.	1963	05	02	+.03	+.08	+.06	+.05	+.03					+.03
A	Nov.	1962	+.03	+.09	+.11	+.10	+.13	+.16	+.19					+.12
В	Nov.	1963	03	00	+.06									+.01
C D	Aug.	1963	+.07	+.01	+.05	+.12	1 00	1 10						+.06
E	Aug. Oct.	1963 1963	02 $13$	$+.08 \\11$	$^{+.05}_{08}$	$^{+.08}_{02}$	+.09	$^{+.16}_{05}$	+.20					+.07 03
Ē	Nov.	1963	+.11	+.08	+.06	+.03	$02 \\ +.02$	+.00	+.20					+.06
Ē	Dec.	1963	02	+.00	00	+.03	00	03	04					01
F	Jul.	1963	15	15	+.00	+.08	01	+.05	+.17	+.15	05			+.01
Ğ	Aug.	1956	13	09	11	13	14	13	1	1 .10	.00			12
Ğ	Jul.	1957	12	26	23	23	05	00						15
$\mathbf{H}$	Jan.	1960	21	05	+.07	10	10	09	+.02					07
$\mathbf{H}$	Oct.	1960	+.00	02	+.04	+.07	+.05	+.08	+.05					+.04
н	Nov.	1960	06	11	09	01	02	10	08					07
H	Dec.	1960	00	02	02	07	07	14	17					07
H	Sep.	1959	+.09	02	05	04	07	08	15					05
H	Oct.	1962	+.26	+.14	+.06	05	02	+.05	+.13					+.08
H	Nov.	1962	+.00	08	08	12	19	15	23					12
H	Oct.	1964	+.10	+.05	03	09	12	20	29					08
H	Nov.	1964	10	06	07	10	11	12	16					10
H	Dec.	1964	+.01	00	01	+.01	00	+.02	01					+.00
I	Feb.	1956	05	06	04	+.08	+.14	+.09	+.10	+.03	+.20	+.05	+.18	+.07
Ī	Mar.	1956	09 $26$	02	10	09	+.02	00	09	01	20	+.04	+.03	05
İ	Apr. May	1956 1956	$20 \\ +.14$	21 $01$	$09 \\ +.08$	05	07 04	$04 \\ +.24$	07	$10 \\ +.35$	18	06	+.06	10
İ	Jun.	1956	+.00	+.07	03	$+.11 \\ +.04$	+.02	+.00	+.09 07	16	$^{+.42}_{+.28}$	08 $14$	$05 \\ +.10$	$+.11 \\ +.01$
j	Jun.	1956	04	02	04	02	+.02	+.07	+.00	+.06	02	14	+.10	+.00
Ĵ	Jul.	1956	+.19	+.24	+.17	+.16	+.05	+.17	+.14	+.00	02 09			+.11
Ĵ	Aug.	1956	10	15	<b>23</b>	45	23	15	12	12	29			20
K	Oct.	1955	18	04	10	+.03	05	07	05	•••	.20			07
K	Nov.	1955	11	04	01	03	05	09	33					09
K	Dec.	1955	28	24	15	23	32	15	25					23
K	Jan.	1956	16	10	07	04	12	08	+.02					08
$\mathbf{K}$	Feb.	1956	+.01	+.02	+.06	+.07	+.03	+.22	+.30					+.10
$\mathbf{K}$	Mar.	1956	10	05	+.01	+.06	02	+.22	+.37					+.07
K	Apr.	1956	+.05	+.07	+.09	+.08	+.00	+.18	+.12					+.08
K	May	1956	06	01	+.02	06	02	06	20					06
K	Jun.	1956	18	13	10	10	05	+.05	+.20					04
K	Jul.	1956	+.11	+.07	+.06	+.08	+.11	+.17	+.15					+.11
K	Aug.	1956	15	14	+.00	04	03	08	+.17					04
K	Sep.	1956	+.17	+.24	+.16	+.13	+.23	+.19	+.07	1 00		٠.		+.17
	Mean I	,	04	03	01	02	02	+.01	+.01	+.02	+.01	04	+.06	

Using the data of table II and table IV, it is possible to make some comparative tests between the seeded and unseeded months. Table V shows a comparison made using the last columns of tables II and IV, where all stations in the target and extended target areas are used irrespective of their distances downwind. The average anomaly for the seeded cases is positive (0.0334) and differs significantly from zero. The unseeded months have a slightly negative but not significant departure from zero. The standard deviations are practically identical (0.1103 compared with 0.0919). The standard t test for differences between two means shows statistical significance.

Table VI shows a comparison between the seeded and unseeded months by individual project. In only one project of the eleven does the average anomaly for the unseeded months exceed the anomaly for the seeded months. The bi-



DIST. FROM APPROX. CENTER OF TARGET (DEGREES)

FIGURE 7

Average anomaly D for 41 unseeded months according to distances from center of the nominal target.

nomial sign test gives a probability of p = 0.01 for this. The unseeded months had a mean anomaly of 0.00 and the seeded months had a mean anomaly of 0.06. A t test on the paired comparisons, with only 10 degrees of freedom available, showed statistical significance on the one tail test, but not on the two tail test.

	Seeded	Unseeded (random or control)	Comparison Test
Number of Project Months $d = \text{difference}$	62 0.0334	41 -0.0127	(101 DF) difference = 0.0461
S = standard deviation $E = $ standard error of $d$	0.1103 0.0140	0.0919 0.0144	0.1034
$t = \frac{d}{E}$	2.39	-0.88	2.22
probability (two tails)	p=0.020	p = 0.40	p = 0.03
probability (one tail for increase)	p=0.010	p = 0.80	p=0.015

TABLE VI

MEAN PRECIPITATION ANOMALY (CUBE ROOT SCALE) BY PROJECT AREA
FOR ALL STATIONS IN NOMINAL AND EXTENDED TARGET AREAS

Project	Locale	Seeded	Unseeded (random or control)	Seeded Minus Unseeded
A	New Hampshire	0.13	0.07	0.06
В	Massachusetts	0.04	0.01	0.03
$\mathbf{C}$	Southeastern New York	0.09	0.06	0.03
D	Connecticut	-0.01	0.07	-0.08
$\mathbf{E}$	New Jersey	0.04	0.01	0.03
$\mathbf{F}$	Eastern New York	0.14	0.01	0.13
$\mathbf{G}$	Eastern Pennsylvania	0.18	-0.13	0.31
H	Southern Pennsylvania	-0.03	-0.04	0.01
Ι	South Carolina	0.06	-0.01	0.07
J	Mid-Potomac	0.03	-0.03	0.06
$\mathbf{K}$	North Carolina	0.02	-0.01	0.03
Means		0.06	0.00	0.06

the primary of the tail p = 0.074, the tail p = 0.007

The results of the analysis to date of the average precipitation anomalies in the target and extended target areas seem to indicate the following:

- (1) the seeded months are different from the historical months;
- (2) the seeded months are different from the unseeded months (chosen independent of the historical record);
- (3) the independent unseeded months show slight but not generally statistically significant anomalies in the negative direction, suggesting that if any bias exists in the statistical analysis, it tends to operate in a direction against finding positive effects of seeding.

If these results had been obtained from scientific experiments, one would ordinarily conclude that cloud seeding had affected rainfall. However, since these were not randomized experiments, a rigorous statement regarding a causal relation between seeding and precipitation anomalies cannot be made. There may be hidden biases and we are continuing to examine the data with such possibilities in mind. For example, we plan to examine the terminal month of the operations to see whether there was any tendency to stop the seeding when higher than normal precipitation occurred in the area. Also, we expect to examine the precipitation anomalies during the period after the cessation of seeding since some investigations have reported evidence of a persistence effect of silver iodide due to contamination of trees, vegetation, and so on.

Another type of bias might be introduced if cloud seeding operations and/or experiments are started in drought periods. This possibility has been suggested by several investigators although no quantitative estimates have been given regarding the direction or magnitude of such possible effects. Our analysis, based on the historical data from unseeded periods, disclosed no evidence that the re-

gression relationship between target and control areas is disturbed during drought periods.

The results, shown in table VII, indicate that the anomalies d expected after

## TABLE VII

Average Anomalies D for the First Three Distance Rings According to Elapsed Time from the Initial Month Chosen on the Basis of Drought in the Target Area

\* indicates beyond 5% significance level; \*\* beyond 1% significance level

							Time	е					
-		1	nitial N	<b>I</b> onth	1	1 Month Later			Year L	ater	2 Years Later		
		N	D	t	N	D	t	N	D	t	N	D	t
Extreme Drought in	Ring 1 Ring 2			-7.68** -7.22**	29 31	-0.00 0.02	-0.15 1.03	45 45	0.02 0.01	1.09	41 40	0.01	0.72 1.31
Target Area	Ring 3			-7.27**	~~			46	0.01	0.38	40	0.02	0.80
Moderate Drought in Target Area	Ring 1 Ring 2 Ring 3	166		-8.27** -7.35** -8.11**	111 111 108		0.14 -0.29 -2.03*	148 151 146	0.01 0.00 -0.00	1.02 $0.21$ $-0.71$	134 135 130		
Mean	<u>'                                     </u>		-0.09	-7.65**		-0.00	-0.50		0.01	0.45		0.01	0.52

a dry period are essentially the same as in the entire sample. Extreme drought months were defined as those falling in the lowest (driest) one tenth of the precipitation frequency distribution. If the month fell in the lower one third of the distribution, it was classified as a moderate drought. The initial months must show significance since they were selected as having deficient precipitation. If there is any bias, it tends to be in the negative direction, making it more difficult to detect positive seeding effects. Another bias which would tend to act in the same direction is due to the contamination of the control area by the ground seeding. Such an effect might be minimized by using a different control area for each wind direction, storm type, season, and so forth. This should also increase the power of a test of seeding effects. We estimate that by using the methods reported in this analysis, a total of 20 seeded months would be needed to give a 90 per cent chance of detecting a 13 per cent increase at the 5 per cent significance level for Project A (New Hampshire, winter). There is good reason to think that a considerable improvement can be made on this by stratifying according to relevant meteorological parameters and by sharpening up the statistical estimates by multiple regression methods, data transformations, and so forth. These and other problems are being investigated not only in order to obtain a better understanding of what might have happened during these particular operations but as part of a general program of searching for more powerful evaluation techniques.