Results of a Randomized Hail Suppression Experiment in Northeast Colorado. Part III: Analysis of Hailstone Size Distributions for Seeding and Yearly Effects[1]

EDWIN L. CROW, ALEXIS B. LONG[2] AND JAMES E. DYE

National Center for Atmospheric Research, Boulder, CO 80307

CARLTON W. ULBRICH

Department of Physics and Astronomy, Clemson University, Clemson, SC 29631

(Manuscript received 7 September 1978, in final form 4 October 1979)

ABSTRACT

The hailstone size (diameter) distributions measured by hailpads during the 1972–74 randomized seeding experiment of the National Hail Research Experiment are analyzed statistically for evidence of seeding effects and differences from year to year. Two approaches are taken, one comparing the entire empirical size distributions on seed days and on control days and the other comparing the mean diameters. The latter is based on the consistency with the exponential distribution (truncated at a prescribed minimum diameter), since the exponential distribution can be characterized completely by the difference between the mean diameter and the minimum diameter. Both approaches yield statistically significant results (10% level) only for 1974, when the hailstones were larger on seed days than on control days on the average. This may have resulted from the addition of seeding by rockets in 1974 or from differences in the hailpads used in that year. However, the physical hypothesis for the experiment predicted *smaller* stones on seed days; that tendency did appear in 1973 (though not significantly) and the difference was negligible in 1972.

1. Introduction

In this paper the hailstone size (diameter) distributions measured by hailpads during the 1972–74 randomized seeding experiment of the National Hail Research Experiment are analyzed for evidence of seeding effects. [A preceding paper (Foote and Knight, 1979) in this series describes the experiment. The hailpads differed somewhat from year to year, as detailed by Dye *et al.* (1976).] The physical hypothesis on which the experiment was based predicts that the seeding should reduce the number of large hailstones and increase the number of small hailstones at cloud base, some of which would melt before reaching the ground. Hence the size distributions on both seed and control days are of considerable interest.

The normalized size distributions (probability density functions) for seed and control hail days are constructed for each of the three years separately and examined for differences. Consistency with the exponential form of distribution is tested. On the basis of that consistency, the distributions can be compared and combined by means of a single parameter, the difference of the mean diameter and the minimum measured diameter, or, equivalently, its reciprocal, which is the negative of the slope of the logarithm of the probability density function. A realistic measure of uncertainty is obtained from the variations between days, account being taken of the different numbers of hailstones on different days.

2. Yearly empirical hailstone size distributions

The extent of the data available for estimating hailstone size distributions from the 3-year experiment is summarized in Table 1. Hail was recorded on 815 hailpads in the target area on 28 of the 57 declared hail days. In all, 58 851 hailstone dents were counted and sized, but it is estimated that 143 130 hailstones fell on the pads (parenthetical numbers in Table 1), for an average of 176 per pad and 5112 per day. In 1973 and 1974 the entire pad was analyzed, but in 1972 the numerous smaller hailpad dents [and thereby the smaller stone diameters via calibration (Dye *et al.*, 1976, p. 4)] were counted only in a central square of area 232 cm^2 if there were 10 or more dents in that square. As a result, in that case the total number of stones of a given small size falling on a pad, of area 1135 cm^2 in 1972, is estimated by multiplying the number falling on the central square by $1135/232 = 4.892$.

[1] This research was performed as part of the National Hail Research Experiment, managed by the National Center for Atmospheric Research and sponsored by the Weather Modification Program, Research Applications Directorate, National Science Foundation.

[2] Present affiliation: Department of Meteorology, Texas A&M University, College Station 77843.

0021-8952/79/121559-10$06.50

TABLE 1. Summary of hailpad data from the NHRE randomized seeding experiment.

	Seed days	Control days
Number of days on which hail fell		
1972	9	5
1973	1	4
1974	4	5
Total	14	14
Number of pads hit		
1972	347	121
1973	38	201
1974	64	44
Total	449	366
Number of stones counted		
1972	19608 (86835)*	4903 (21955)*
1973	1678	23215
1974	8118	1329
Total	29404 (96631)*	29447 (46499)*

* Parenthetical numbers are the numbers of stones hitting the entire pads, estimated in 1972 in part from counts on a fraction (232/1135) of most pads except for larger stones.

The terms "size" and "diameter" are used interchangeably herein for the equivalent hailstone diameter derived from the measured minor diameter of each dent. The hailstone diameter is calculated from the dent diameter using a calibration of the hailpad. We note that the size intervals differ from year to year (as did the designs of the pads and the methods of reduction) and that the measurements do not extend to zero diameter.

The numbers of hailstones counted in each diameter category on all hailpads during the experiment are summarized in Tables 2–4. These are derived from Table XIII of Volume II of the NHRE Final Report (Dye *et al.*, 1976) by summing over all pads for each day and all days of each type (seed S or control C). The relative frequency estimates in Tables 3 and 4 are simply ratios to the total number of hailstones, but those in Table 2 are combinations because of the fractional counts mentioned above. For example the first seed day relative frequency in Table 2 is

$$\frac{64 + (1135/232)1603}{2336 + (1135/232)17272} = 0.09105.$$

The standard deviations of the estimates of relative frequencies in Tables 2–4 will be discussed in Section 3.

It was found by Crow *et al.* (1979, Tables 6, 7, 9, 10 and 11) that the mean daily number of hailstones on seed days did not differ significantly from that on control days. Table 7 of that paper shows

TABLE 2. Observed and derived 1972 hailstone diameter distributions. The first line of the third and fourth columns in each category gives the counts on complete pads (1135 cm²); the second (if any), on fractions of pads (232 cm²); the remaining columns apply to the combination.

Diameter category (cm)		Observed number of hailstones		Number of days with stones		Relative frequency			
						Estimate $\bar{p}$		Standard deviation of $\bar{p}$	
Minimum diameter	Maximum diameter	S	C	S	C	S	C	S	C
0.073	0.249	64 1603	28 610	9	5	0.0910	0.1372	0.0318	0.0202
0.249	0.441	243 5692	99 1469	9	5	0.3235	0.3318	0.0388	0.0358
0.441	0.625	152 3037	39 653	9	5	0.1729	0.1473	0.0170	0.0198
0.625	0.800	191 3439	32 939	9	5	0.1960	0.2107	0.0151	0.0223
0.800	0.980	73 1134	8 237	9	5	0.0647	0.0532	0.0077	0.0105
0.980	1.152	91 1702	12 351	9	5	0.0969	0.0788	0.0131	0.0066
1.152	1.482	54 665	6 122	8	5	0.0381	0.0275	0.0040	0.0073
1.482	1.780	921	185	7	4	0.0106	0.0084	0.0023	0.0042
1.780	2.040	331	69	6	2	0.0038	0.0031	0.0010	0.0022
2.040	2.238	138	28	6	2	0.0016	0.0013	0.0004	0.0008
2.238	2.405	48	10	5	2	0.0006	0.0005	0.0002	0.0002
2.405	2.555	16	5	3	2	0.0002	0.0002	0.0001	0.0002
2.555	2.692	13	1	2	1	0.0001	0.0000	0.0001	0.0000
2.692	2.822	1	0	1	0	0.0000	0.0000	0.0000	0.0000
Totals		2336 17272	522 4381			1.0000	1.0000		

TABLE 3. Observed 1973 hailstone diameter distributions.

Diameter category (cm)		Observed number of hailstones		Number of days with stones		Relative frequency		
						Estimate $\bar{p}$		Standard deviation of $\bar{p}$
Minimum diameter	Maximum diameter	S	C	S	C	S	C	C
0.389	0.560	496	4973	1	4	0.2956	0.2142	0.0037
0.560	0.740	611	8283	1	4	0.3641	0.3568	0.0257
0.740	0.915	308	4865	1	4	0.1836	0.2096	0.0036
0.915	1.090	166	2436	1	4	0.0989	0.1049	0.0071
1.090	1.260	59	1270	1	4	0.0352	0.0547	0.0047
1.260	1.430	21	621	1	4	0.0125	0.0267	0.0026
1.430	1.600	10	371	1	3	0.0060	0.0160	0.0020
1.600	1.765	1	158	1	2	0.0006	0.0068	0.0012
1.765	1.925	5	84	1	2	0.0030	0.0036	0.0009
1.925	2.075	1	48	1	2	0.0006	0.0021	0.0008
2.075	2.220	0	37	0	2	0.0000	0.0016	0.0006
2.220	2.365	0	31	0	2	0.0000	0.0013	0.0005
2.365	2.495	0	18	0	1	0.0000	0.0008	0.0004
2.495	2.615	0	9	0	1	0.0000	0.0004	0.0002
2.615	2.725	0	6	0	1	0.0000	0.0003	0.0001
2.725	2.890	0	5	0	1	0.0000	0.0002	0.0001
Totals		1678	23215			1.0001	1.0000	

that a significant difference in mean numbers of hailstones was found between *years* (significance level of 0.032), primarily due to the smaller daily mean number in 1974. This paper considers the *relative* frequencies of hailstone diameters, and the absolute numbers are not analyzed further. The probability density function (pdf) considered here is the proportion of stones of a given size per centimeter of diameter. It is estimated from the relative frequencies in Tables 2–4 by dividing by the respective category lengths.

The estimated probability density functions of the 1972, 1973 and 1974 seed and control diameter distributions appear in Figs. 1–3. The diameter category intervals are omitted, and the density function estimates are plotted at the midpoints of the categories and are connected by straight lines to suggest the shape of the density function. Maximum likelihood estimates of the density functions under the assumption that the true form is exponential are also shown and will be discussed in Section 4.

3. Analysis of empirical distributions for seeding effects

On the whole, the seed (S) and control (C) distributions have similar shapes in all three years, with an initial increase from the smallest size hail-

TABLE 4. Observed 1974 hailstone diameter distributions.

Diameter category (cm)		Observed number of hailstones		Number of days with stones		Relative frequency			
						Estimate $\bar{p}$		Standard deviation of $\bar{p}$	
Minimum diameter	Maximum diameter	S	C	S	C	S	C	S	C
0.401	0.554	1008	214	4	5	0.1242	0.1610	0.0341	0.0519
0.554	0.709	1855	389	4	5	0.2285	0.2927	0.0218	0.0184
0.709	0.863	2135	363	4	5	0.2630	0.2731	0.0086	0.0183
0.863	1.017	1315	208	4	5	0.1620	0.1565	0.0153	0.0304
1.017	1.171	827	84	4	4	0.1019	0.0632	0.0066	0.0113
1.171	1.325	463	38	4	4	0.0570	0.0286	0.0068	0.0096
1.325	1.480	236	26	3	4	0.0291	0.0196	0.0044	0.0074
1.480	1.634	147	6	3	1	0.0181	0.0045	0.0043	0.0022
1.634	1.788	74	1	3	1	0.0091	0.0008	0.0040	0.0004
1.788	1.942	23	0	2	0	0.0028	0.0000	0.0010	0.0000
1.942	2.096	26	0	2	0	0.0032	0.0000	0.0017	0.0000
2.096	2.251	5	0	1	0	0.0006	0.0000	0.0002	0.0000
2.251	2.405	4	0	2	0	0.0005	0.0000	0.0010	0.0000
Totals		8118	1329			1.0000	1.0000		

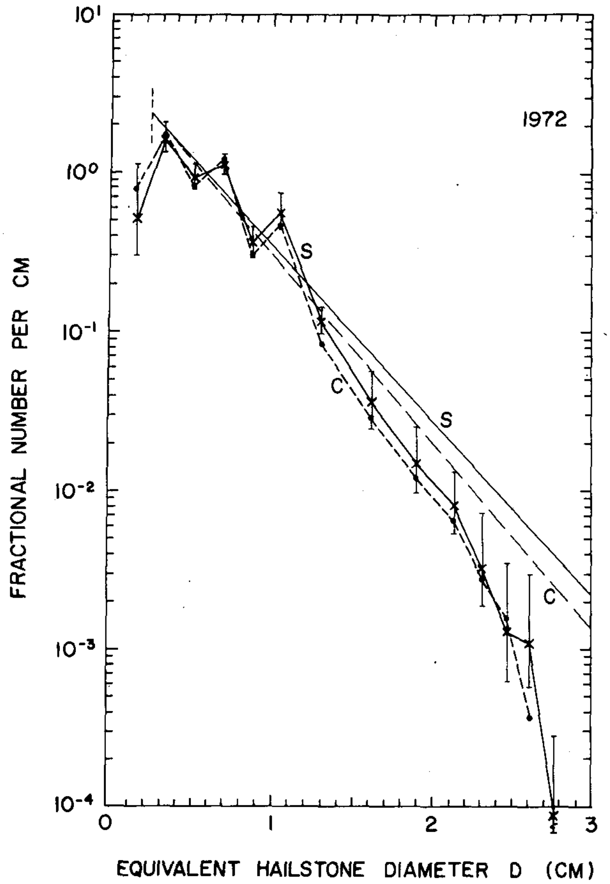

FIG. 1. Normalized hailstone size distributions (probability density functions) for 1972, obtained by dividing the mean daily relative frequencies in Table 2 by the widths of the diameter categories. The straight lines are the maximum likelihood best-fitting exponential distributions for $D \geqslant 0.249$ cm. The 90% confidence intervals based on a gamma approximation within each diameter category are shown for the seed case only.

stone and then a usually steady decline except for three initial sawteeth in 1972 common to the S and C distributions. The sawteeth seem to be best explainable as unnatural aberrations possibly resulting from a tendency to classify stones in even-numbered categories, or resulting from categories being longer or shorter than specified. It is reasonable to smooth them out. The sawteeth for large diameters of the S distributions only in Figs. 2 and 3 can be attributed to random fluctuations in the small samples of those diameters, as indicated by the plotted confidence intervals to be discussed further below.

The estimated probability density functions for seed days in 1972 and 1974 (Figs. 1 and 3) lie above those for control days for diameters larger than about 1 cm except for one point in 1972, but the opposite is true in 1973 (Fig. 2). Whether the differences are statistically significant, in view of the amount of data, needs to be tested. A standard test of the difference between two categorized sample distributions is the chi-squared test of a $2 \times k$ contingency table, where k is the number of size categories. Such a test was performed for each of the three years and was found to result in values of chi-squared of ~100–500, which correspond to extremely high significance (level beyond 10^{-14}). This result was predictable from the huge numbers of hailstones recorded, which the test treats as independent observations without reading error. Chi-squared tests no doubt would also indicate significant differences from day to day or often even from pad to pad on the same day because of the large numbers of hailstones recorded.

For an example we consider the pads at sites 340 and 343 on 7 August 1974, an average day in that 5549 hailstones were counted on 40 pads, though it was the largest of 1974. The two pads were ~3 km apart. The following numbers of hailstones were counted on them in the diameter categories listed in Table 4 from 0.401 cm to 1.634 cm:

Pad 340:	19	38	20	7	3	0	0	0	Sum = 87
Pad 343:	1	9	37	23	11	9	5	4	Sum = 99.

The relative frequencies are obviously very dif-

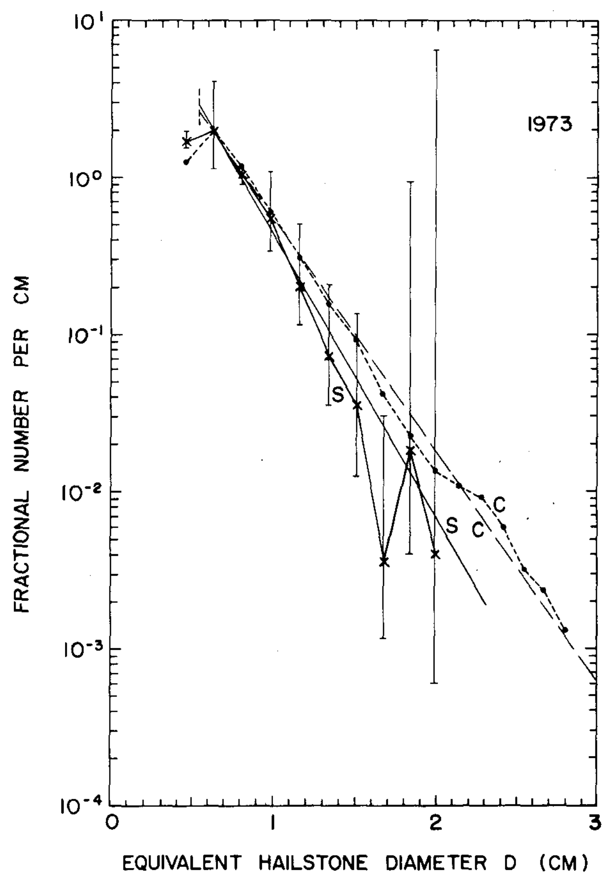

FIG. 2. As in Fig. 1 except for 1973 and Table 3. The straight lines are the maximum likelihood best-fitting exponential distributions for $D \geqslant 0.560$ cm.

ferent, and it is no surprise that χ^2 is large, 70 with 7 degrees of freedom (d.f.), significant at even the 0.001% level.

An explanation of the large pad-to-pad differences is given by Huggins *et al.* (1979). They found from two-way analysis of variance of the counts on 23 pads by 11 persons in NHRE that the between-person standard deviation was about three times the sampling standard deviation predicted by the Poisson distribution. The chi-squared test illustrated above includes only the Poisson variation, whereas much of the observed variation is additional to the Poisson variation, arising from the difficulty of sizing and counting the dents on the pad. Thus the chi-squared test would tend to find significant differences between pads, days and treatments that were due simply to reader variations rather than some systematic physical difference. The test is therefore not correctly applicable here.

The standard deviations of estimated relative frequencies in Tables 2–4 provide a realistic measure of uncertainty for testing for seeding effects. These standard deviations are derived from the differences among the individual daily relative frequencies taking account of the differing amounts of data by application of the classical Gauss-Markov theorem on least squares (David and Neyman, 1938). We may restrict attention to any particular hailstone size category and let p be the probability that a stone falling on the NHRE target area is in that size category; p is assumed to be the same for all declared hail days in a given year but may differ on seed and control days. We let k_i be the number of stones in that category on the ith day of the sample ($i = 1, 2, \ldots, n$) and m_i be the total number of stones of all sizes on that day. Then the relative frequency of stones of that size on the ith day is k_i/m_i, which has mean p and variance $p(1-p)/m_i$ over the set of hypothetical repetitions of the randomized seeding experiment with m_i fixed if k_i is counted without error. This follows from the binomial distribution if we restrict the repetitions to those with stone totals equal to the same m_i. There are further terms in the variance of k_i/m_i due to reader and calibration errors; we assume they are also proportional to $1/m_i$. Then the Gauss-Markov theorem states in effect that the "best" (uniformly minimum variance) linear unbiased estimate (BLUE) of p can be obtained by weighted least squares with weights equal to the m_i. The theorem also gives a formula for estimating the variance of the estimate. The easily derived results in this case are that the BLUE of p is

$$\bar{p} = \sum m_i(k_i/m_i)/\sum m_i = \sum k_i/\sum m_i, \quad (1)$$

where the summations are over the n days of the sample, and the unbiased estimate of the variance of $\bar{p}$ is

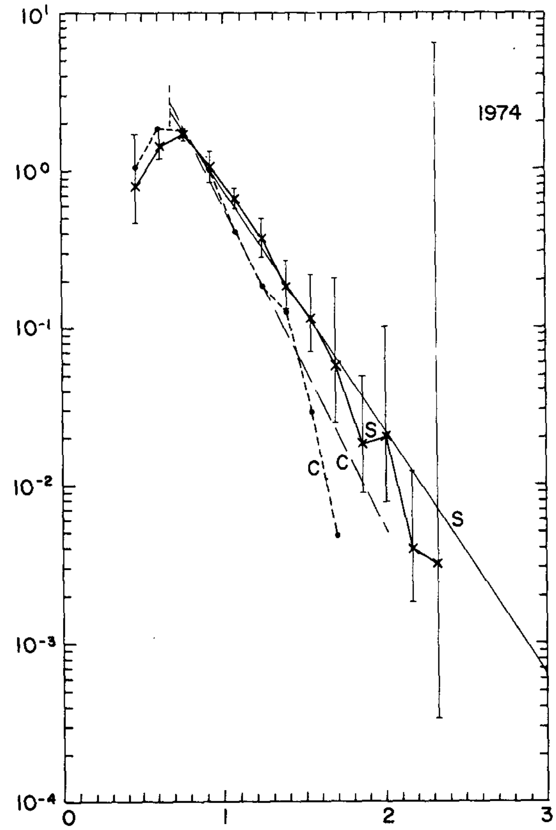

FIG. 3. As in Fig. 1 except for 1974 and Table 4. The straight lines are the maximum likelihood best-fitting exponential distributions for $D \geq 0.709$ cm.

$$s_{\bar{p}}^2 = \frac{\sum m_i(k_i/m_i - \bar{p})^2}{(n-1)\sum m_i} \quad (2)$$

$$= \frac{(\sum m_i)(\sum k_i^2/m_i) - (\sum k_i)^2}{(n-1)(\sum m_i)^2}. \quad (3)$$

The columns labeled standard deviation of $\bar{p}$ in Tables 2–4 are calculated from Eq. (3). For example, the first one in Table 3 comes from the four control days (21 May and 9, 21 and 28 July 1973) with respective k_i values of 2941, 225, 1741 and 66 and m_i values of 13584, 1017, 8224 and 390.[3] [In 1972 the m_i varied from 448 to 9786 on C days and from 418 to 31113 on S days. In 1974 the m_i values ranged from 26 to 685 on C days and from 327 to 5549 on S days (Dye *et al.*, 1976).]

[3] The Gauss-Markov theorem could have been applied *within each day* also with k_i and m_i coming from the ith pad. The purpose of this would be to estimate a weight for each day, which would be the inverse of the variance calculated from (3), based on pads rather than stones. While this would give proper weight to each pad for very large numbers of stones on each pad, the weights are poorly estimated from the numbers available. Hence the weights based on daily total numbers of stones are evidently preferable.

In applying the binomial distribution and Gauss-Markov theorem above and in the inferences to be made below, it is assumed that the counts of stones on different days are mutually independent. This is an acceptable assumption because all of the days on which hail was recorded on pads are separated from one another by at least one other day (at least two other days except in two cases), and the available data indicate this is enough separation for approximate independence (Crow *et al.*, 1979, Section 4b). However, the method yields an unbiased estimate of the true hail size distribution only if any systematic reader and calibration errors are negligible, and this is not necessarily so.

The differences between the seed and control estimated relative frequencies in Tables 2–4 and the standard errors of the differences can be calculated immediately. For example, from the first line of Table 2 the difference in relative frequencies for the diameter category 0.073–0.249 cm in 1972 is 0.0462, and its standard error is 0.0377. The ratio of the two, 1.226, suggests that the difference is not significant if the number of days involved, 14, is considered large enough to invoke the Central Limit Theorem and thus conclude the approximate normality of the difference in estimates.

Actually the distributions of the estimates of relative frequencies and their differences are unknown and may not be well approximated by the normal distribution for the very small frequencies of large stones. Since the frequencies cannot be negative, their distributions are probably better approximated by gamma or log-normal distributions. The 90% confidence intervals plotted in Figs. 1–3 are based on the gamma approximation, but those based on normal and log-normal approximations were also calculated for comparison. These, especially the gamma, are described further in the Appendix.

The confidence intervals are shown in Figs. 1–3 only for the hail size distribution on seed days, in order to avoid clutter. In 1972 and 1974 the intervals for the control day distribution have about the same length as the corresponding intervals for the seed day distribution. In 1973 there was only one S day with hail, so that standard deviations were not directly calculable for S days, but the reasonable assumption (verified in 1972 and 1974) was made that the C standard deviation in Table 3 could be applied to the S estimate by multiplying by the square root of the ratio of numbers of hailstones observed, $(23215/1678)^{1/2}$ or 3.72. [This ratio is derived by noting that the expected value of the right-hand side of Eq. (2) is $p(1-p)/\sum m_i$ and that $\sum m_i$ is 23 215 for C days and 1678 for the S day.]

The procedure adopted to test for seeding effect on the size distributions is as follows:

1) Within each size category for each year, observe whether the 90% confidence interval for the S pdf includes the C pdf estimate; likewise with the C interval and the S pdf; if either interval includes the other pdf estimate, the two pdf estimates for that size and year do not differ significantly at the 10% level.
2) If neither above interval covers the other pdf estimate, perform a modified Fisher F test with the ratio of pdf estimates at the 10% significance level, thus accounting for the uncertainty in both estimates. (The modified Fisher F test, described in the Appendix, is based on the gamma distribution approximation used for the confidence intervals.)
3) Count the number of size categories for each year that result in significance in 2).
4) Test the sequence of runs of positive and negative differences between the S and C pdf estimates in each year for a significantly small or large number of total runs.

The results of this series of tests are as follows, and can be confirmed to some extent from Figs. 1–3:

Year	Total number of pdf differences	Number of differences with neither interval covering other pdf (1)	Number of significant results (3)	Total number of + and − runs (4)
1972	13	0	0	6
1973	10	2	1	2
1974	9	4	4	2
Total	32	6	5	10

Thus all but 6 of the 32 differences of S − C pdf estimates can be dismissed as not significant simply on the basis of the uncertainty in one or the other. Of the remaining 6 differences, 5 turn out to be significant at the 10% level. On the average one would expect 3.2 significant results at the 10% level from a total of 32 tests, and 5 could easily result by chance when in fact the two population distributions are identical. It is true that 4 of these occur in 1974, which is quite unlikely by chance under the hypothesis of no real difference; this will be discussed further below. The significance tests were repeated using the normal and log-normal approximations for the distributions of estimates of relative frequencies (i.e., using the Student t test on the relative frequencies or their logs) to determine whether the uncertainty as to the type of distribution is a matter of concern. The only change in the number of significant results was a reduction from 4 to 3 using the log-normal approximation in 1974.

The above numbers of significant results do not take account of possible joint significance of effects that are not individually significant. The total number of runs of positive and negative differences provides such a combination test, although it is not really valid because the differences are not independent, being constrained to add to zero. The number of such runs in 1972 was 6, very close to the number to be expected on the average and thus not significant. The number in 1973 was 2, the minimum possible number and barely significant since the probability of 2 runs in random sampling with 2 positive and 8 negative independent differences is 0.044 (Beyer, 1966, p. 318). In 1974 the number of runs was also 2, but arising from 6 positive and 3 negative differences, so that the corresponding probability is 0.024. Although these 1973 and 1974 results on runs are nominally significant, *the trends in the two years were exactly opposite*. In 1973 the hailstones on the four control days tended to be slightly larger than on the one seed day; in 1974 the hailstones on the four seed days tended to be slightly larger than on the five control days.

4. Testing consistency with exponential distributions

Except for the increase from the category of smallest diameters recorded and a second increase in the case of the 1974 S curve, all six distributions of Figs. 1–3 appear consistent with straight lines, i.e., with an exponential distribution. These inconsistencies can be attributed to systematic undercounting of small dents. Just as the hailstones below a certain threshold (0.073, 0.389 and 0.401 cm for 1972, 1973 and 1974, respectively) were not counted at all, those in the first size category were probably undercounted because of the difficulty in seeing small dents. This was especially true for soft hail and for 1974, when the pads were not covered by foil and deteriorated in the sun (Dye *et al.*, 1976). (Whether or not there is undercounting of small dents does not affect the analyses in Section 3 because seed and control data are both subject to the same undercounting.)

Thus the hail size distributions are truncated at some lower bound, say $D_{\min}$. The left-truncated exponential probability density function has the form

$$f(D) = \theta e^{-\theta(D-D_{\min})}, \quad D \geq D_{\min}. \tag{4}$$

The maximum likelihood estimate of θ is

$$\theta_{ml} = \frac{1}{\bar{D} - D_{\min}}, \tag{5}$$

where

$$\bar{D} = \sum_{j=1}^{k} \bar{p}_j D_j, \quad \sum_{j=1}^{k} \bar{p}_j = 1, \tag{6}$$

$\bar{p}_j$ is the relative frequency of observations in the jth diameter category, D_j is the midpoint of that category, and k is the number of categories. We note that $\bar{D} - D_{\min}$ is the important dimension, not simply $\bar{D}$. Thus, the left-truncated exponential distribution is characterized by one informative parameter. This makes it easy to compare hail size distributions if they are all exponential distributions.

Since the pdf estimates for the category of smallest stones in Figs. 1–3 are inconsistent with the exponential form, and in any case we wish to estimate the ultimate slope, $D_{\min}$ was taken as the lower boundary of the next category in 1972 and 1973. In 1974, $D_{\min}$ was taken as 0.709 cm, the lower boundary of the third category. The values of $D_{\min}$ and $\bar{D}$ (cm) calculated from Tables 2–4 (where the relative frequencies have to be renormalized for $D \geq D_{\min}$) are

	$D_{\min}$	$\bar{D}_S$	$\bar{D}_C$
1972	0.249	0.643	0.620
1973	0.560	0.799	0.850
1974	0.709	0.989	0.920

The corresponding numerical slopes (cm^{-1}) are

	$\theta_{ml,S}$	$\theta_{ml,C}$
1972	2.54	2.70
1973	4.18	3.45
1974	3.57	4.75

The lines are plotted in Figs. 1–3, their level being adjusted so that they fit the pdf histogram estimates previously plotted before discarding the category of smallest stones, i.e., to yield unit area including all categories rather than only those beyond $D_{\min}$; this avoids a superficial replotting of points. For example, the adjusted 1973 C pdf is multiplied by the ratio 18242/23215, yielding the equation

$$f(D) = 2.71 e^{-3.45(D-0.560)}, \quad D \geq 0.560 \text{ cm}.$$

The question of whether the data are consistent

TABLE 5. Mean diameters $\bar{D}$ of the NHRE hailstone size distributions and their standard errors $s_{\bar{D}}$, both values in centimeters. The first diameter category is deleted. In 1974 the second category is also deleted.

	Seed		Control		Both	
Year	$\bar{D}$	$s_{\bar{D}}$	$\bar{D}$	$s_{\bar{D}}$	$\bar{D}$	$s_{\bar{D}}$
1972*	0.954	0.026	0.931	0.021	0.950	0.019
1973	0.799	0.072	0.850	0.018	0.847	0.017
1974*	0.840	0.021	0.771	0.016	0.805	0.035
1972–74*	0.864	0.046	0.850	0.046	0.867	0.043

* $\bar{D}$ is adjusted to the 1973 minimum diameter, 0.560 cm, by adding 0.311 cm for 1972 and subtracting 0.149 cm for 1974.

with the exponential form of distribution can be answered by noting what proportion of the 90% confidence intervals for individual pdf values previously calculated cross the fitted distribution. We delete the category of smallest stones, simply noting their inconsistency with the distributions fitted to the other categories. We obtain the following counts, using 90% confidence intervals for both S and C distributions:

	Total number of intervals		Number not crossing exponential line	
Year	S	C	S	C
1972	13	12	10(8)	2(0)
1973	9	15	0	1
1974	11	7	1	1
Total	33	34	11(9)	4(2)

The numbers in parentheses apply if the intervals associated with the sawteeth in Fig. 1 are subtracted out. Whether or not this is done, the 1972 seed distribution has many more significant deviations from the maximum likelihood-fitted exponential than can be expected by chance, but it is the only one that does. The maximum likelihood-fitted straight line is more influenced by the thousands of stones at small diameters than the relatively few at large diameters and does not maximize the number of confidence intervals that cross a fitted line. A straight edge can be laid on Fig. 1 that misses only the first and last confidence intervals and those of two upper projecting sawteeth. It is more appropriate to test exponentiality using the confidence intervals against such a line than against a less favorable line.

In summary, except for small stones that may be systematically undercounted, the distributions of hailstone diameters deduced from hailpads show essential consistency with the exponential form, as illustrated by the tendency to linearity in Figs. 1–3.

5. Testing the reduced mean diameter for seeding and yearly effects

If the distribution of hailstone diameters is a truncated exponential, the estimated distribution is completely specified by the reduced mean diameter, $\bar{D} - D_{\min}$, or its reciprocal, the slope, aside from the instrumental constant $D_{\min}$. Even if the distribution is not quite exponential, it is convenient to take $\bar{D} - D_{\min}$ as a primary characteristic. It is a linear function of the observed relative frequencies, and its variance can be estimated by applying the Gauss-Markov theorem to daily mean diameters $\bar{D}_i$, just as it was applied to the daily relative frequencies k_i/m_i in Eqs. (1)–(3).

The resulting values of $s_{\bar{D}}$ are summarized in Table 5 along with the values of $\bar{D}$ given in Section 4 adjusted to the same $D_{\min}$, 0.560 cm, by adding 0.311 cm for 1972 and subtracting 0.149 for 1974. In other words, the values of $\bar{D}$ in Table 5 are (aside from sampling fluctuations) the average hail diameters that would have been obtained in 1972 and 1974 if the data had been truncated at 0.560 cm rather than 0.249 or 0.709 cm, respectively, and were in fact samples from the fitted exponential distributions.

Since the $\bar{D}$ are approximately normally distributed by the Central Limit Theorem, we can test for differences by using the Welch-Aspin modified Student t-test. For example, in 1972,

$$w = \frac{0.954 - 0.931}{(0.026^2 + 0.021^2)^{1/2}} = 0.69,$$

a nonsignificant difference. The S − C difference in 1973 is also not significant, but the S − C difference in 1974 is significant at the 5% level. This is consistent with the trend exhibited in Fig. 3 and discussed at the end of Section 3.

The absence of significant S − C differences in 1972 and 1973 provides a justification for pooling the S and C data in each year to obtain a better estimate of the yearly mean diameters. Although the S − C difference in 1974 is statistically significant, it is not large compared with the difference between years. To assign an approximate mean diameter to the 1974 experimental results, it is considered reasonable to weight the S and C results equally and calculate a standard error, $s_{\bar{D}} = 0.035$, from their difference. This is about twice the standard errors of the pooled 1972 and 1973 means, as shown in Table 5.

Differences in mean diameter between years might be expected because of differences in pad construction and analysis. Although the Table 5 values are adjusted to the same minimum diameter, all three pairs of yearly values of $\bar{D}$ for control days differ significantly at the 5% level, and the 1972 and 1974 values for seed days differ at the 1% level.

The unadjusted values of $\bar{D}$ differ even more; this emphasizes the importance of carefully selecting the minimum diameter to be observed. Because of the significant differences between years, the mean diameters assigned to the combined three years of the experiment in Table 5 are obtained by weighting all three years equally, and their standard errors are calculated from their equally weighted deviations. The mean diameter from all of the data, 0.867 cm, happens not to lie between the values from the separate seed and control data because of the different weighting (influenced especially by the more numerous seed data in 1972 and the more numerous control data in 1973). However, because of the year-to-year variation it can only be concluded that the mean equivalent hailstone diameter in northeastern Colorado is ~0.87 cm if the minimum equivalent diameter of stone included is 0.56 cm, with a standard error of ~0.04 cm and thus 90% confidence limits (Student t) on the true mean of ~0.75 and 0.99 cm. These limits on the mean diameter can be transformed to limits on the slope θ by subtracting $D_{\min}$ and inverting; thus $2.3 < \theta < 5.3$.

As mentioned earlier, $\bar{D} - D_{\min}$ may well be considered a primary characteristic whether or not the distribution is exponential. Within any given year the test of the difference between $\bar{D}_S$ and $\bar{D}_C$ is identical to that between $\bar{D}_S - D_{\min}$ and $\bar{D}_C - D_{\min}$, and we can just as well perform this test on *all* of the data for each year as on only the data consistent with the exponential form of distribution. This has been done. The mean diameters for all of the data, of course, are considerably less than those for the truncated data, but the only change in the results of the significance testing is that the 1974 seed and control mean diameters are *not* significantly different. Thus none of the three yearly S − C differences in mean diameter of the full observed distributions of hailstones is significant.

6. Concluding remarks

The large amount of data on hailstone sizes from the NHRE randomized seeding experiment has been analyzed for seeding effect and differences between years using statistical estimation, tests and confidence intervals extensively. The methods, especially the application of the Gauss-Markov theorem of Section 3 and the approximate confidence intervals and modified F test based on the gamma distribution outlined in the Appendix, are suggested for future analyses of hailstone size distributions.

The differences between seed and control days were examined for possible significance, both by testing the differences between the empirical hailstone size distributions (Figs. 1–3) and by testing the differences in mean diameter (Table 5). Both approaches yield statistically significant results (10% level) only for 1974, when the hailstones were larger on seed days on the average. The opposite tendency was observed in 1973, and the difference was negligible in 1972. The physical hypothesis on which the experiment was based had predicted a tendency toward *smaller* hailstones on seed days, but physical explanations for an increase have also been propounded. The seeding *was* different in kind and magnitude in 1974, since it was done by rocket as well as flare without reduction in rate (Foote and Knight, 1979). Consequently the dosage achieved per cell in 1974 was about twice those in the previous two years (Foote *et al.*, 1979, Table 3). Whether this was the cause of the significant difference cannot be known; it could be a chance result also.

Except for the first one or two size categories, which are believed to be subject to undercounting, the hailstone size distributions are consistent with the exponential distribution (truncated at some minimum diameter). This makes it possible to characterize the entire distribution by the reduced mean diameter and to test for possible differences with this single characteristic if the first one or two categories are deleted. With such deletion the difference is statistically significant only in 1974, and that significance disappears if no categories are deleted. The reduced mean diameters depend on the minimum diameter observed or accepted, and there are differences between years in the NHRE results for this reason, as well as, presumably, because of differences in the hailpads used.

With hindsight it is apparent that it would have been desirable to achieve a completely satisfactory hailpad design and analysis before the first year, so that hailpad differences between years would not have entered. Standardization should be sought in future experiments but may not necessarily be attained (Lozowski *et al.*, 1978; Morgan, 1978).

Acknowledgments. The authors are indebted to three journal reviewers for many comments that resulted in this substantially revised paper, as well as to colleagues G. Brant Foote and Griffith M. Morgan, Jr., for their suggestions.

APPENDIX

Confidence Intervals and Significance Tests for Probability Density Functions

The confidence intervals for the normalized hailstone size distributions shown in Figs. 1–3 are approximate because the distribution of the estimated ordinate for each size category in repeated sampling is not known. For sufficiently large samples the distribution would be approximately normal by the Central Limit Theorem. On this basis the Student t confidence intervals with degrees of freedom one less than the number of days in the sample could be justified and were actually calculated as one alternative. However, since the estimated ordinates cannot be negative and a few of their standard deviations are substantial fractions of their own values, a better approximation to their distribution than the normal would be a distribution over only positive numbers, like the log-normal or gamma. Confidence intervals based on both of these were also calculated to provide a sensitivity analysis with respect to form of distribution.

The application of the log-normal assumption involves using $\log(k_i/m_i + \text{a constant})$ as the observations in Eqs. (1) and (2) rather than k_i/m_i. If none of the k_i for a given size category is zero, then the constant was taken as zero. If any of the k_i is zero, then the constant was taken as one-fourth of the minimum possible positive relative frequency on the day or days with zeros, i.e., $1/(4 \times \max m_i)$; this is consistent with replacement of the discrete count by the continuous log-normal variable.

The gamma approximation is more attractive than the log-normal for two reasons. Not only do the zero counts present no problem, but also Eq. (1) yields directly an unbiased estimate of the true relative frequency, and neither of these is true for the log-normal.

It is often convenient, as well as necessary in this case, to estimate the required shape parameter of the gamma approximation from the first two moments of the distribution, i.e., from Eqs. (1)–(3) above (Bury, 1975, p. 168). The estimated equivalent number of degrees of freedom, which is twice the standard shape parameter for the gamma distribution, is given by

$$\bar{\nu} = 2(\bar{p}/s_{\bar{p}})^2. \tag{7}$$

Then the classical confidence limits for the variance of a normal population can be applied using the tabulated percentage points of a chi-squared distribution. This procedure is analogous to using the percentage point z_α of the normal distribution rather than that of the Student t distribution, $t_{n-1,\alpha}$, in the case that p is taken to be normally or log-normally distributed. Since the gamma distribution is shaped more like the log-normal than the normal, an intuitive correction to allow for estimating ν by $\bar{\nu}$ is proposed, namely expanding the above classical confidence limits in the ratio by which the use of $t_{n-1,\alpha}$ instead of z_α expands the limits for p based on the log-normal distribution of $\hat{p}$. Straightforward algebra then results in the "modified chi-squared" confidence interval for p,

$$\bar{p}\left(\frac{\bar{\nu}}{\chi^2_{\bar{\nu},\alpha}}\right)^{t_{n-1,\alpha}/z_\alpha} < p < \bar{p}\left(\frac{\bar{\nu}}{\chi^2_{\bar{\nu},1-\alpha}}\right)^{t_{n-1,\alpha}/z_\alpha}, \tag{8}$$

where $\chi^2_{\bar{\nu},\alpha}$ is the upper 100 α percentage point of the chi-squared distribution with $\bar{\nu}$ degrees of freedom, $\chi^2_{\bar{\nu},1-\alpha}$ is the corresponding lower percentage point, the confidence level is $1 - 2\alpha$, and n is the number of independent days of data. For the 90% confidence level and 5 days of data, $t_{n-1,\alpha}/z_\alpha = 1.296$.

If independent observations of the same gamma distribution were available, then optimal confidence limits for its scale parameter (p in this case) would theoretically be calculable (Engelhardt and Bain, 1977). That is not the situation here. Although the interval (8) is in any case an approximation, it is asymptotically, for large $\bar{\nu}$, the same as the Student t interval obtained when $\bar{p}$ is assumed to have a normal distribution.

Corresponding to the modified chi-squared confidence interval for a single p is a "modified F" test of the hypothesis of equality of p's on seed and control days. It is the same as an F test except that both of the degrees of freedom $\bar{\nu}_1$ and $\bar{\nu}_2$ are reduced in the ratio $(z_\alpha/t_{n_1+n_2-2,\alpha})^2$. It is based simply on the asymptotic variance of F_{ν_1,ν_2}.

REFERENCES

Beyer, W. H., ed., 1966: *Handbook of Tables for Probability and Statistics*. Chemical Rubber Co., Cleveland. 502 pp.

Bury, K. V., 1975: *Statistical Models in Applied Science*. Wiley, New York, 625 pp.

Crow, E. L., A. B. Long, J. E. Dye, A. J. Heymsfield and P. W. Mielke, Jr., 1979: Results of a randomized hail suppression experiment in northeast Colorado. Part II: Surface data base and primary statistical analysis. *J. Appl. Meteor.*, **18**, 1538–1558.

David, F. N., and J. Neyman, 1938: Extension of the Markoff theorem on least squares. *Statist. Res. Mem.*, **2**, 105–116.

Dye, J. E., A. J. Heymsfield, I. Paluch and D. W. Breed, 1976: *Final Report, National Hail Research Experiment Randomized Seeding Experiment, 1972–74:* Vol. II. *Precipitation Measurements*. National Center for Atmospheric Research, Boulder, Colo., 530 pp.

Englehardt, M., and L. J. Bain, 1977: Uniformly most powerful unbiased tests on the scale parameter of a gamma distribution with a nuisance shape parameter. *Technometrics*, **19**, 77–81.

Foote, G. B., and C. A. Knight, 1979: Results of a randomized hail suppression experiment in northeast Colorado. Part I: Design and conduct of the experiment. *J. Appl. Meteor.*, **18**, 1526–1537.

——, C. G. Wade, J. C. Fankhauser, P. W. Summers, E. L. Crow and M. E. Solak, 1979: Results of a randomized hail suppression experiment in northeast Colorado. Part VII: Seeding logistics and post hoc stratification by seeding coverage. *J. Appl. Meteor.*, **18**, 1601–1617.

Huggins, A., E. L. Crow and A. B. Long, 1979: Subjectivity of hailpad data reduction. Submitted to *J. Appl. Meteor.*

Lozowski, E. P., R. Erb, L. Wojtiw, M. Wong, G. S. Strong, R. Matson, A. Long, D. Vento and P. Admirat, 1978: The hail sensor intercomparison experiment. *Atmos.-Ocean*, **16**, 94–106.

Morgan, G. M., 1978: Summary of the panel discussions: Hail instrument calibration. *Atmos.-Ocean*, **16**, 137.