On the Design and Evaluation of Cumulus Modification Experiments

JOANNE SIMPSON[1]

Center for Advanced Studies and Deparment of Environmental Sciences, University of Virginia, Charlottesville 22903

JANE C. EDEN

National Hurricane and Experimental Meteorology Laboratory, NOAA, Coral Gables, Fla. 33124

ANTHONY R. OLSEN

Battelle-Northwest, Richland, Wash. 99352

(Manuscript received 28 August 1974, in revised form 21 March 1975)

ABSTRACT

Combination of numerical simulation, many simultaneous measurements, and a large assortment of statistical tools, employed at all stages, have been found useful in design and evaluation of modification experiments on cumulus clouds. A randomized sample is essential, although non-random controls have supplemented it by providing necessary information on natural distributions.

Obstacles to definitive estimates of treatment effects are huge natural variability compounded by the expense and labor involved in obtaining an adequately large data sample. A 26 pair data set from a dynamic seeding experiment on isolated Florida cumuli is used here to illustrate both the problems and the combined approach used to overcome them. In this data set, rain volumes from unmodified single cumuli varied by three orders of magnitude on days screened as suitable. The field phase of the experiment cost above $250,000, requiring instrumented aircraft, calibrated radar, and several radiosondes daily.

Numerical simulation of seeded and unseeded cumulus towers defined the key screening variable "seedability," namely the predicted height difference between seeded and unseeded towers, so that only days on which the physical seeding hypothesis would be expected to work are selected for experimentation. On those days, randomization is between clouds selected by the experimenters as suitable.

Classical and Bayesian statistics are used together in the evaluation, with both univariate and multivariate analyses. Various well-known probability density distributions fitted the seeded and unseeded rainfalls. Among the best were gamma, log-normal, beta-K and beta-P. Seed-control differences were examined with nonparametric and parametric tests (some of the latter after data transformation) and effects of random and systematic measurement errors were considered. In all tests, the seed-control rainfall difference was significant at better than 5%. A multiplicative seeding factor of 2–3 was estimated in several ways (allowing for or getting around the bias problem with ratio estimators related to long-tailed distributions).

1. Introduction

For the past decade NOAA's Experimental Meteorology Laboratory (EML) has been conducting randomized experiments with dynamic cumulus seeding using a combined approach of model simulation, numerous measurement systems, and an increasing repertory of statistical tools both classical and Bayesian. These tools have been inextricably interwoven during every phase of experimentation from early design through successive evaluations.

The modification of tropical convective rainfall is particularly challenging because of its great variability in space and time and the "heavy-tailed" nature of the distributions, but progress is slowly being made in learning how to design and evaluate experiments in these situations, which are of considerable scientific and practical importance.

The problems and procedures are most fruitfully illustrated in the context of the data set of the EML randomized single cloud experiments, from which the seeding effects can be definitively deduced.

2. The data set and some initial results

In 1968 and 1970, a randomized dynamic seeding experiment was conducted on single isolated cumuli over south Florida. The experiment design, model use, and extensive measurements have been published in detail (Simpson *et al.*, 1970, 1971; Simpson and Wiggert, 1969; 1971; Woodley, 1970; Woodley and Herndon,

[1] The contribution of this author was mainly supported under a grant from the RANN Division of the National Science Foundation entitled "Evaluation, Execution and Design of Modification Experiments on Convective Clouds and Systems." Prior to 1 July 1974, she was Director of The Experimental Meteorology Laboratory, NOAA.

TABLE 1. EML single-cloud rain volume data for 1968 and 1970

Date	Cloud number	R (acre-ft)	$R^{1/4}$	$\ln R$
		A. Control Set		
1968				
May 15	1	26.1	2.26027	3.26193
May 16	2	26.3	2.26459	3.26957
May 19	3	87.0	3.05408	4.46591
May 20	4	95.0	3.12199	4.55388
May 20	5	372.4	4.39291	5.91997
May 26	6	0 (0.001)	0 (0.178)	0 (−6.908)
May 27	7	17.3	2.03944	2.85071
May 28	8	24.4	2.22253	3.19458
May 28	9	11.5	1.84151	2.44235
May 30	10	321.2	4.23344	5.77206
1970				
May 25	11	68.5	2 87689	4 22683
May 26	12	81.2	3.00185	4.39691
May 26	13	47.3	2.62250	3.85651
May 26	14	28.6	2.31255	3.35341
May 27	15	830.1	5.36763	6.72155
May 28	16	345.5	4.31134	5.84499
May 28	17	1202.6	5.88885	7.09224
May 28	18	36.6	2.45963	3.60004
July 1	19	4.9	1.48782	1.58923
July 1	20	4.9	1.48782	1.58923
July 2	21	41.1	2.53198	3.71601
July 2	22	29.0	2.32060	3.36729
July 3	23	163.0	3.57311	5.09375
July 16	24	244.3	3.95349	5.49840
July 16	25	147.8	3.48673	4.99586
July 18	26	21.7	2.15832	3.07732
		B. Seeded Set		
1968				
May 15	1	129.6	3.37405	4.86445
May 16	2	31.4	2.36719	3.44681
May 16	3	2745.6	7.23868	7.91775
May 16	4	489.1	4.70272	6.19257
May 19	5	430.0	4.55373	6.06378
May 20	6	302.8	4.17147	5.71307
May 21	7	119.0	3.30283	4.77912
May 26	8	4.1	1.42297	1.41099
May 27	9	92.4	3.10040	4.52613
May 28	10	17.5	2.04531	2.86220
May 30	11	200.7	3.76389	5.30181
May 30	12	274.7	4.07113	5.61568
May 30	13	274.7	4.07113	5.61568
1970				
May 25	14	7.7	1.66580	2.04122
May 26	15	1656.0	6.37918	7.41216
May 27	16	978.0	5.59223	6.88551
May 28	17	198.6	3.75400	5.29129
July 1	18	703.4	5.14992	6.55592
July 1	19	1697.8	6.41906	7.43709
July 2	20	334.1	4.27532	5.81144
July 3	21	118.3	3.29797	4.77322
July 3	22	255.0	3.99609	5.54126
July 3	23	115.3	3.27686	4.74753
July 16	24	242.5	3.94619	5.49100
July 18	25	32.7	2.39132	3.48737
July 18	26	40.6	2.52424	3.70377

1970). Two variables, vertical growth and rain volume, were related to the seeding.

Seedability (km) is defined as the difference in model-predicted top height of a cloud if seeded versus the height of the same cloud if left unseeded. Seeding effect (km) is defined as the difference between measured top height and predicted unseeded top height. For seeded clouds, seedability and seeding effect correlated 0.95, with no relation for the controls. Seeded clouds grew about 3 km taller than controls (difference significant at better than 1%). Their radar echoes were 75% larger in area and lasted 56% longer than unseeded clouds.

The rain volume data are of primary concern here. They were obtained by means of a carefully calibrated 10 cm radar, which was tested against raingage determinations (Woodley, 1970). The data are presented in Table 1. The values can be shown (Woodley and Herndon, 1970) to be comparable in accuracy to the unadjusted radar data from FACE 1973 discussed by Woodley *et al.* (1975). Effects of possible errors in the data upon the deductions regarding seeding effects are examined later in Section 6.

A striking feature of Table 1 is that natural rainfall varies by three orders of magnitude! The average seeded rainfall is 442.0 acre-ft. The average control rainfall is 164.5 acre-ft. The standard deviations are comparable in magnitude to the average values, namely 650.8 and 278.4 acre-ft respectively. The average seed-control difference is 277.5 acre-ft and the ratio r of the average values is 2.68.

Extensive classical statistical testing was applied to the fourth-root transforms of the data (Simpson *et al.*, 1971). The fourth-root transform was used to minimize the effects of non-normality, to reduce heteroscedasity[2], and to make statistical models more appropriate in view of suspected day-to-day variations in seeding effects. Among the tests[3] used were the nonparametric Mann-Whitney-Wilcoxon, with which the null hypothesis was rejected at the 0.5% level. Other tests applied were Student t, covariate regression, analysis of covariance, and analysis of daily means. With all tests, seeded and control populations differed at the 5% significance level or better. In particular, included with the analysis of variance is a calculation indicating a multiplicative seeding effect on rainfall, in the neighborhood of a factor of 3.

[2] When making inferences about the mean value of some response (e.g., rainfall volume) in the presence of treatments applied at several levels (e.g., seed vs no-seed) two situations can occur:

1) The variance of the observations about the within-treatments means can remain unaltered when going from one treatment level to another.
2) The variance of the observations about the within-treatments means can differ from one treatment level to another.

In case 1) the observations are said to be homoscedastic; in case 2) they are said to be heteroscedastic. Almost all traditional statistical inference procedures require homoscedasity.

[3] One-tailed.

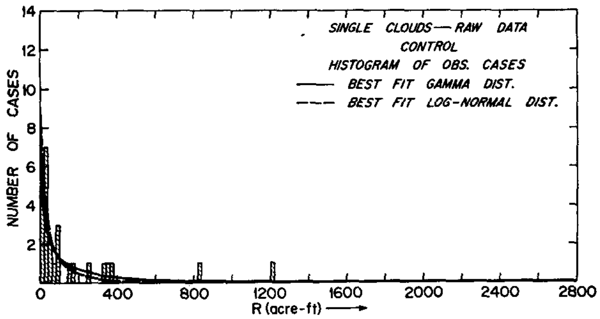

FIG. 1. Histogram of untransformed single-cloud control data compared with best fit gamma (solid) and log-normal (dashed) distributions.

Subsequently, data stratification was undertaken, indicating larger affects of dynamic seeding in Florida on "fair" compared to "naturally rainy" days (defined in terms of radar echo coverage). An additional investigation (Brier *et al.*, 1972) examined and rejected the possibility of selection bias, using covariates and analysis of experiment design.

Simpson (1972) showed that the transformed (fourth-root) data are well fitted by gamma distributions of the form

$$p(R)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}R^{\alpha-1}e^{-\beta R}, \qquad (1)$$

with the shape parameter α invariant between seeded and control populations, which also suggests a multiplicative seeding effect.

It is important to test for any possible bias introduced by the fourth-root transformation. For this test, we use the control rainfall data in Table 1A. The first step (Brier, 1974) consists of comparing the average raw rainfall data $\bar{R}$ with the fourth-power (inverse transform) of $\bar{T}$, the average of the fourth-root transformed data. Since $\bar{R}=164.55$ acre-ft, $\bar{T}=2.90192$ (acre-ft)$^{\frac{1}{4}}$, then $\bar{T}^4=70.9160$. The ratio of $\bar{T}/\bar{R}=0.431$. Clearly, one must be careful with the interpretation of inverse transforms with these data.

Let us further inquire about deduction of seeding effects from this type of data with the fourth-root transform. Suppose that a given seeding method were to multiply every control rainfall in Table 1A by a factor f, an unlikely but illustrative situation. Then let us compare $\bar{T}$ to $\bar{R}$ as a function of f.

$\bar{T}$ may be written

$$\bar{T}=\sum_{i=1}^{n}(fx_i)^{\frac{1}{4}}/n=f^{\frac{1}{4}}\sum_{i=1}^{n}x_i^{\frac{1}{4}}/n,$$

$$\bar{T}^4=f(\sum_{i=1}^{n}x_i^{\frac{1}{4}})^4/n^4,$$

so that

$$\frac{\bar{T}^4}{\bar{R}}=f(\sum_{i=1}^{n}x_i^{\frac{1}{4}})^4/(n^3\sum_{i=1}^{n}x_i), \qquad (2)$$

where x_i is each data bit and n is the number of observations, in this case 26.

The important result is that the ratio is linearly related to f. It is clear from (2) that the desired relationship is

$$\frac{\bar{T}^4}{\bar{R}}=0.431f$$

with the data of Table 1A. This result suggests caution in estimating seeding effects from transformed data, but does not provide a method for reliable correction. In nature, for example, a multiplicative effect of a treatment upon the average of a sample may not imply an identical effect upon each data bit in the sample.

For another illustration, therefore, we made a similar experiment using both seeded and control data in Table 1. The ratio of average seeded to average control rainfalls $r=\bar{R}_s/\bar{R}_c=2.68$. When we average the fourth-root transforms, we find $r_t=\bar{T}_s/\bar{T}_c=1.33670$ so that $(r_t)^4=3.19$. The latter is an over-estimate of r by nearly 20%. The problem of data transformations, how to recognize and correct for their distortions in experimental deductions, is being studied further by our group and others. Thus it is desirable to work with raw data if and when possible and this is done in most of the remainder of the paper.

3. The distribution of the raw (untransformed) data and the number of cases required to resolve seeding factor of various sizes

Histograms of the untransformed (raw) data are shown in Figs. 1 and 2. The best-fit distributions to the data were obtained with our program based on maximum entropy (Simpson and Pezier, 1971; Simpson, 1972) which gives the same results as maximum likelihood. The best-fit distributions (among nine compared) were log-normal and gamma, as shown. Histograms of the transformed (fourth-root) data and best-fit distributions (gamma and truncated normal) are shown in Figs. 3 and 4 (log-normal was a better fit than truncated normal).

Table 2 gives the gamma parameters for raw and transformed data.

The salient feature of the raw rainfall distributions in Figs. 1 and 2 is that they are "heavy-tailed." Since convective rain distribution will always be heavy-tailed,

TABLE 2. Parameters of best-fit gamma distributions for R in acre-ft.

Raw data			Transformed (fourth-root) data	
	α	β	α	β
Control	0.56	0.00341	6.52	2.22
Seed	0.64	0.00145	7.10	1.83

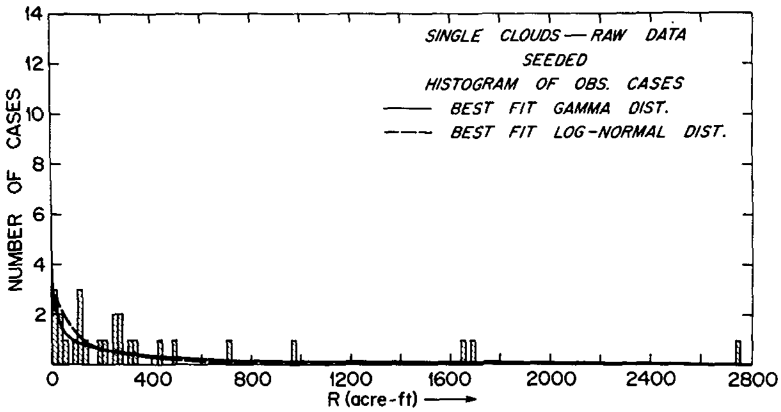

Fig. 2. As in Fig. 1 except for seeded data.

the pitfalls of these distributions and their treatment is worth careful consideration.

In modification experiments, it is important to inquire at an early stage how many experimental cases are required to resolve seeding effects, the available estimators, and what cautions are necessary in their use.

With the EML experiments, two main approaches have been used to predict the sample size needed to estimate seeding factors. An example of one of the two approaches is given by Olsen and Woodley (1975). They study the power of statistical tests as a function of seeding factor and sample size for rainfalls obeying the appropriate gamma distributions characteristic of the EML targets, both with and without measurement error degradation. They showed that the adverse effect of error degradation is small in this context compared to the problem posed by natural variability and small sample. For example, the radar errors reduced the power of the Maximum Likelihood Ratio test by only a few percent for seeding factors in the range 1.2–1.6, while reducing the hypothetical sample size from 50 to 20 pairs almost halved the power of the test. Therefore,

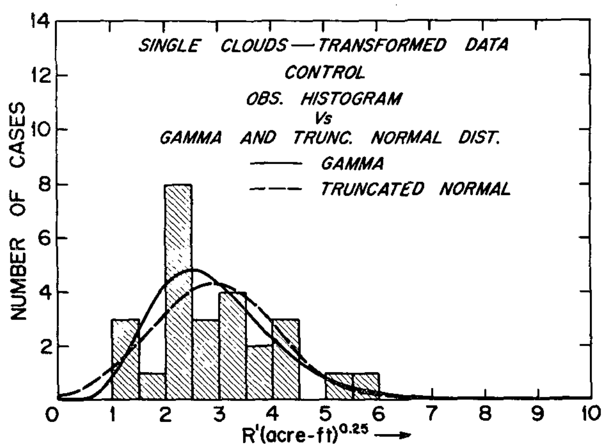

Fig. 3. Transformed (fourth-root) control data versus best-fit gamma (solid) and truncated normal (dashed) distributions.

in the following phase of our analysis, measurement errors are not explicitly considered.

The approach illustrated here assumes that the seed and control samples are from gamma distributions with the same α (here $\alpha=0.6$) and different β's so that the actual seeding factor is defined as

$$F\equiv\frac{\langle R_s\rangle}{\langle R_c\rangle}=\frac{\beta_c}{\beta_s}, \tag{3}$$

where $\langle\ \rangle$ denotes "expected value." The gamma distribution has the properties

$$\left.\begin{aligned}\mu_1&=\langle R\rangle=\alpha/\beta\\ \mu_2&=\sigma^2=\alpha/\beta^2\end{aligned}\right\}. \tag{4}$$

The sample ratio $r=\bar{R}_s/\bar{R}_c$ is now called $\hat{F}$, the sample seeding factor estimator.

It can be shown (Simpson *et al.*, 1973a) that the sampling distribution of $\hat{F}$ is β_c/β_s times a Snedecor's F-distribution, so that

$$E(\hat{F})=\frac{\beta_c}{\beta_s}\frac{m\alpha}{m\alpha-1}, \tag{5}$$

$$\text{Var }(\hat{F})=\frac{\beta_c^2}{\beta_s^2}\left(\frac{m\alpha}{m\alpha-1}\right)^2\frac{n\alpha+m\alpha-1}{n\alpha(m\alpha-2)}, \tag{6}$$

where c stands for control, s for seed, n is the size of the seeded sample, and m the size of the control sample. Flueck and Holland (1973) have derived a more general expression for this result and have discussed its implication for modification experiments.

An important point is that $\hat{F}$ or r is a positively biased estimator. If we sample from two identical populations with $\beta_s=\beta_c$ (no seeding effect) and $\alpha=0.6$, then with $m=n=26$

$$\left.\begin{aligned}E(\hat{F})&=1.07\\ \text{Var }(\hat{F})&=0.14\end{aligned}\right\}. \tag{7}$$

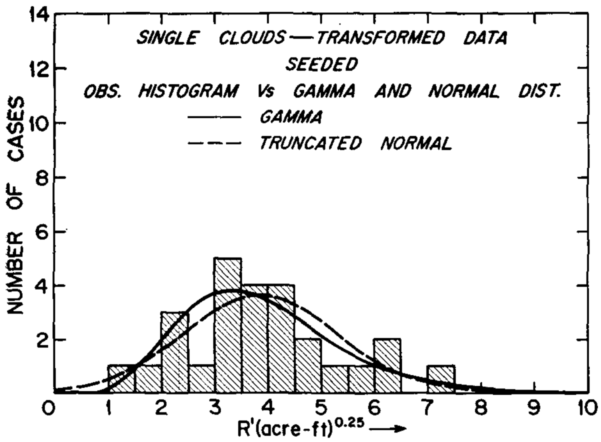

FIG. 4. As in Fig. 3 except for seeded data.

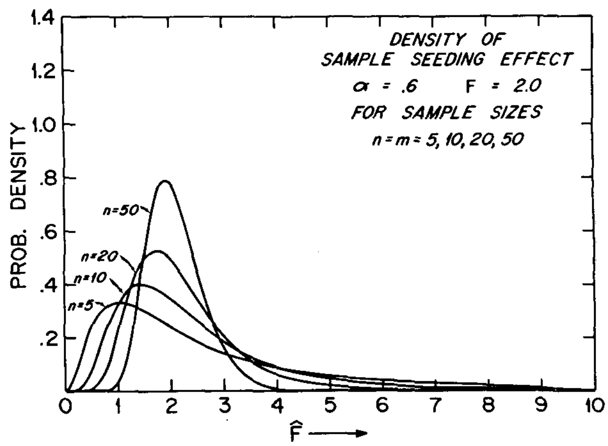

FIG. 6. As in Fig. 5 except for $F=2.0$.

We shall return to this point later. First we calculate the probability density function of $\hat{F}$ as a function of F and sample size in order to examine sample size requirements in a modification experiment. The procedure is described in Appendix II of the report by Simpson *et al.* (1973a). Results pertinent to the single-cloud experiment are shown in Figs. 5–7.

The figures show that with the existing sample $m=n=26$, resolution of a seeding factor of 2–3 appears feasible, but if the seeding effect were smaller, a larger data sample would probably be required. The sample size requirement is relaxed as the gamma shape parameter increases, as it apparently does with larger rainfall units, such as the floating and total target rainfalls in the EML area experiment (see Simpson and Woodley, 1975). The shape parameter is inversely related to the degree that the distribution is "heavy-tailed."

Concerning the positive bias in the estimator $\hat{F}$ or r, in view of the large relative variance, it would be unwise to correct the EML single cloud data ratio by 7% reduction and assume that the result was a good estimator of the multiplicative seeding effect. This point is illustrated by the probability density distributions for $\hat{F}$ shown in Figs. 5–7, using the gamma distribution properties.

Next we use a Monte Carlo experiment to bring the point out more forcefully. We constructed a computer program called RAIN (listed and tested by Simpson *et al.*, 1973a) which draws observations at random from a gamma distribution. For this example, we took a single distribution with $\alpha=0.6$, β corresponding to the control distribution in Table 1, and drew 100 sets of n pairs of cases. The ratio of each term in two matrices containing 100 "seeded" and 100 "control" samples was taken; results are shown in Table 3.

This table presents the frequency of various ratios of the "seeded" sample average to the "control" sample average. This kind of spurious result could arise from an attempted modification experiment where the treatment had no effect.

Table 3 suggests extreme caution in drawing inferences from seeding experiments with small samples

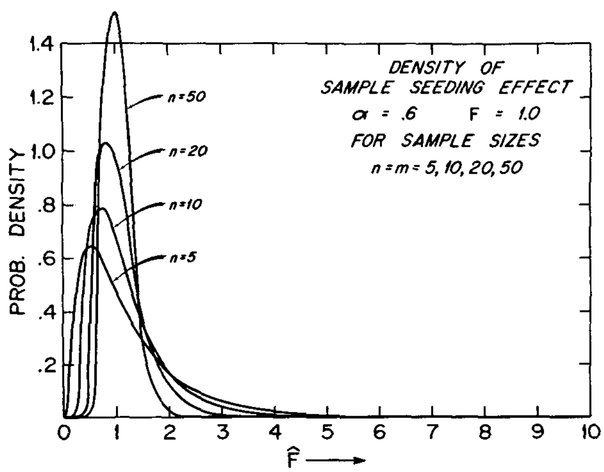

FIG. 5. Probability density of $\hat{F}$ (ratio estimator) versus sample size when $F=1.0$, $\alpha=0.6$.

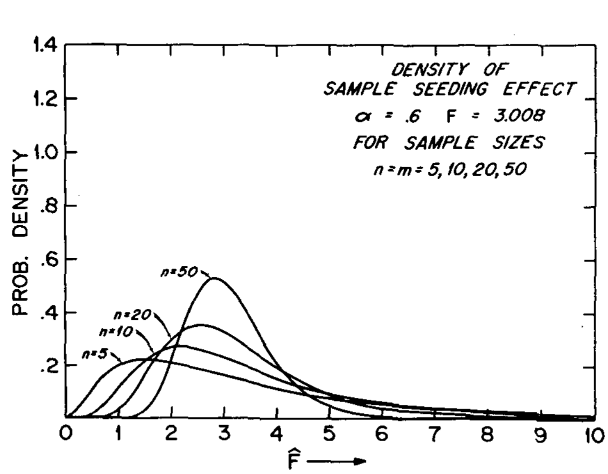

FIG. 7. As in Fig. 5 except for $F=3.008$.

of data. It also suggests in order to specify natural distributions adequately for resolution of seeding effects of a factor of 2–3, that 20–50 cases are a necessary minimum. If we are concerned with expected seeding factors of 50% or less, it is plain that 50 cases are not adequate, with this type of data distribution.

Even with 50 pairs of cases, we see that there is a sizable probability (~20%) of r in the range of 0.51–0.80 and similarly in the range 1.21–1.50 when there is no real seeding effect at all! This is one aspect of the situation meteorologists refer to as a "bad draw" in a seeding experiment, which is an error defined by statisticians as Type I. The converse difficulty occurs when observations are drawn from two distinct populations but the ratio r is sufficiently close to 1 owing to natural variability so that the null hypothesis is accepted. This aspect of the "bad draw" (defined as a Type II error by statisticians) and how to avoid it is being studied further.

An unbiased estimator for the true seeding factor is possible if α is assumed known. If we define

$$\tilde{F}=\frac{m\alpha-1}{m\alpha}\hat{F}, \tag{8}$$

then

$$E(\tilde{F})=\frac{m\alpha-1}{m\alpha}E(\hat{F})=\frac{\beta_s}{\beta_c}=F. \tag{9}$$

Usually, however, α must be estimated. It is possible to substitute an estimate of α in determining $\tilde{F}$.

A point estimator, however useful, does not provide all the information we need because of the large variance in $\hat{F}$. It is desirable to set a confidence limit on F. This is done noting that $\hat{F}/F$ obeys a Snedecor F-distribution with $2n\alpha$, $2m\alpha$ degrees of freedom. Based upon that we wish to obtain

$$Pr\left[a\leqslant\frac{\hat{F}}{F}\leqslant b\right],$$

where $Pr=0.95$, for example.

TABLE 3. Simulated "seeding" experiment—100 samples of n cases drawn at random from the same gamma distribution.

Frequency distribution of ratios $n=$	5	10	20	50
minimum	0.058	0.26	0.24	0.56
⩽0.50	17	7	8	0
0.51–0.80	24	27	19	21
0.81–1.20	22	23	40	53
1.21–1.50	7	15	16	21
1.51–2.0	7	11	12	5
2.01–2.50	10	4	3	0
2.51–3.0	0	10	2	0
⩾3.01	13	3	0	0
maximum	10.70	4.06	2.71	1.81

TABLE 4. All (41) fictitious "seed-control" ratios ⩾2 drawn from control gamma distribution.

Test	Accept	Reject
Optimal $C(\alpha)$	7	34
Maximum Likelihood	6	35
t (fourth-root)	20	21
t (log)	24	17
Wilcoxon rank	23	18
Squared rank	17	24

The 95% equal-tailed confidence interval is found by transforming the probability statement to

$$Pr\left[\frac{\hat{F}}{b}\leqslant F\leqslant\frac{\hat{F}}{a}\right]$$

and placing 2.5% probability in each tail. With $\alpha=0.6$ and assuming 30 degrees of freedom, we obtain from a Snedecor F table that $b=2.07$ and $a=1/2.07$ so that the 95% confidence interval for F is 1.3 to 5.56, in good agreement with values from the Bayesian analysis described later in Table 6. The wide range is, of course, a consequence of the heavy-tailed distribution, which is manifested by the large variance in $\hat{F}$ shown in Eq. (7).

Let us concentrate next on the chance that 26 pairs of cases could differ as much as those in Table 1 and still be from the same population. First, using two nonparametric, two-tailed tests (Mann-Whitney-Wilcoxon and squared rank) and two two-tailed tests depending on the gamma distribution, namely Optimal $C(\alpha)$ and Maximum Likelihood, the null hypothesis is rejected at the 5% significance level with the real data. Next, 1000 pairs of 26 cases are drawn at random from the control gamma distribution, and ratios taken. Of these, 41 or 4.1% exhibit ratios of 2 or greater. These "data sets" were fed into the statistical tests program. Results are shown in Table 4.

We see that the probability of a real positive seeding effect is high with the EML single-cloud rainfall data. However, mere rejection of the null hypothesis is not optimally useful information. Decision makers would like to have an estimate of the magnitude of the seeding effect, which is undertaken next.

4. Estimation of seeding factor using the gamma properties of the raw data

We again assume that the shape parameter α is the same for the seeded and unseeded populations. A probability distribution for F, the multiplicative seeding factor defined in Eq. (3), will be obtained.

Here we use Bayes equation to estimate a probability density distribution for F after seeing the data. A wide range of prior probabilities has been tested, including diffuse or non-informative ones. With the latter, Olsen (1975) has shown that the same results can be obtained (with the same assumptions) using classical statistics.

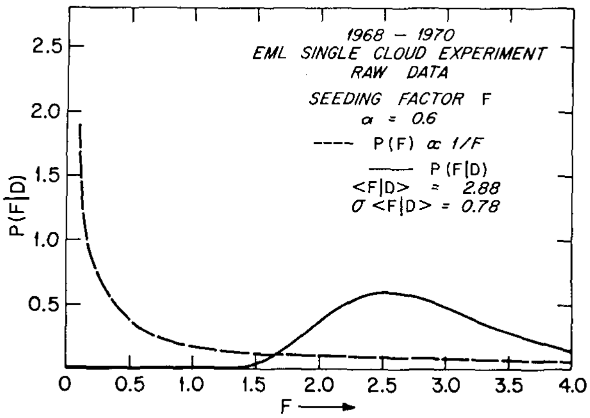

FIG. 8. Probability distribution of seeding factor F when prior F is proportional to $1/F$. The shape parameter $\alpha=0.6$ for seeded and control populations and the control scale parameter is determined from the sample mean.

With the gamma distribution, the Bayesian analysis is easily carried out. Many workers believe this approach improves in usefulness in proportion as sound, informative priors become available. Our philosophy in using numerous classes of priors is based on our desire to test the sensitivity of the calculated posterior probability distribution of F to the choice of prior. In the early stages of experimentation, we place more emphasis upon diffuse priors, which in this context implies as much integrated probability below 1 as above it ($F=1$ means no seeding effect).

The analyses are conducted on three levels of sophistication, the first placing a prior on F alone. The common shape parameter α is assumed equal to 0.6, and the natural distribution is determined from the control sample mean. For this analysis, Bayes equation is written

$$p(F|D)=p(F)\frac{p(D|F)}{p(D)}, \tag{10}$$

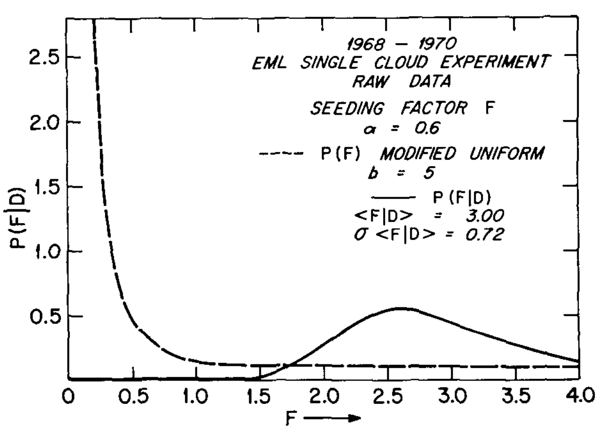

FIG. 9. As in Fig. 8 except that prior distribution on F is modified uniform (equal area above and below $F=1$) with upper limit $F=5$.

TABLE 5. Percent probability that the sample mean of n cases lies within specified limits of $\langle R\rangle$ (from numerical experiments): raw data ($m=1000$); $\alpha=0.6$.

	Specified limit of $\langle R\rangle$					
n	<5%	<10%	<20%	>30%	$<\frac{1}{2}$	>2
5	6.0	12.4	25.0	61.9	20.1	6.0
10	9.8	19.5	37.7	45.9	7.1	2.1
20	13.9	28.6	51.2	28.1	1.8	0
50	20.5	39.1	72.3	7.4	0	0

where D is the set of seeded data in Table 1B. Analyses of this sort have been extensively presented elsewhere (Simpson *et al.*, 1973b; Simpson and Woodley, 1975) so only two examples are illustrated here in Figs. 8 and 9. In Fig. 8, $p(F)$ is proportional to $1/F$. The mean and modal values of posterior F are 2.88 and 2.53, respectively, while equal-tailed 95% confidence limits are 1.73–4.75 and the shortest 95% confidence limits 1.74–4.50.

However, these results are based on the assumption that the natural distribution of rainfall is known. We have already glimpsed the pitfalls associated with small samples from heavy-tailed distributions. In the preceding analysis, we used $\bar{R}_c$ and $\alpha=0.6$ to obtain the natural distribution. Table 5 shows our chances of obtaining sample means within specified percentages of the expected value of a prescribed gamma distribution. We therefore see that with 26 cases, we have only somewhat better than an even chance of estimating $\langle R\rangle$ to within 20%.

The next class of Bayesian analysis therefore places a joint prior on F and β_c, the control scale parameter, and both seed and control data sets are involved in determining the posterior distributions.

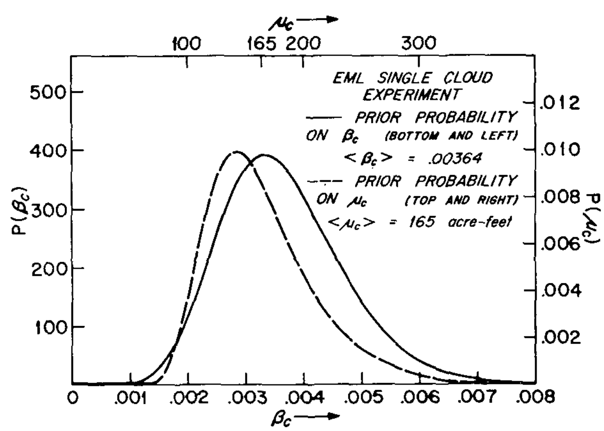

FIG. 10. Prior probability of β_c (solid), the control scale parameter used in the Bayesian analysis shown in Fig. 12. Since β_c is a gamma function, μ_c, the control mean, is an inverse gamma function (dashed) here with an expected value of 165 acre-ft, chosen to agree with the data sample mean.

Table 6. Bayesian analysis for single-cloud seeding factor F: raw data; $\alpha=0.6$; priors on F, β_c (μ_c unknown); R in acre-ft

	Prior F Inverse gamma $K_1=1$; $K_2=0.5$		Prior F $1/F$	
	Gamma prior on β_c $K_1=12$; $K_2=3300$	Gamma prior on β_c $K_1=12$; $K_2=3833.33$	Gamma prior on β_c $K_1=12$; $K_2=3300$	Prior on β_c
Prior $\langle\beta_c\rangle$	0.00364	0.00313	0.00313	$1/\beta$
Prior $\langle\mu_c\rangle$	165 acre-ft	191.67 acre-ft	191.67 acre-ft	
Posterior F				
$\langle F\rangle$	2.39	2.24	2.68	2.87
Mode	2.05	1.92	2.27	2.36
σ	0.77	0.72	0.90	1.08
Confidence limits on Posterior F				
95% equal tailed	1.25–4.23	1.17–3.96	1.37–4.84	1.31–5.49
95% shortest	1.10–3.93	1.03–3.67	1.19–4.47	1.10–5.02
Cumulative probabilities on F				
$F\leqslant 1$	0.003	0.006	0.001	0.004
$F\leqslant 1.2$	0.02	0.03	0.008	0.014
$F\leqslant 1.5$	0.09	0.13	0.05	0.05
$F\leqslant 2$	0.34	0.42	0.23	0.21

Bayes equation is written in the form

$$p(F,\beta_c|\mathbf{x},\mathbf{y})=p(F,\beta_c)\frac{p(\mathbf{x},\mathbf{y}|F,\beta_c)}{p(\mathbf{x},\mathbf{y})}, \qquad (11)$$

where $\mathbf{x}$ is the control data sample and $\mathbf{y}$ the seeded data sample. The analysis for determining $p(\mathbf{x},\mathbf{y}|F,\beta_c)$ has been developed by Olsen (1975). It is facilitated if the prior on β is gamma and the prior on F is an inverse gamma function. Three cases are illustrated in Figs. 10–14, with the pertinent accompanying data in Table 6.

We can see from Figs. 10 and 11 that the priors on μ_c allow for considerable uncertainty in the control distribution and, particularly in Fig. 11, are unfavorable for F, since $\bar{R}_c\approx 165$ acre-ft. The priors on F are also unfavorable, particularly those in Figs. 12 and 13 which have virtually no area above 2. Nevertheless, we still deduce from the data a positive seeding factor of 2 or larger.

The next important question to be addressed is the degree of accuracy with which gamma function parameters can be determined by sampling from a gamma distribution as a function of n, the number of observations. Here we generate data from our RAIN program corresponding to the control distribution. We select n observations at random, pretend the "data" set is a set of rainfall data from n clouds, and apply the maximum entropy program that determines the relative probabilities of several distributions and their best fit parameters. In Table 7 we have repeated this procedure m times for each value of n and then examined the

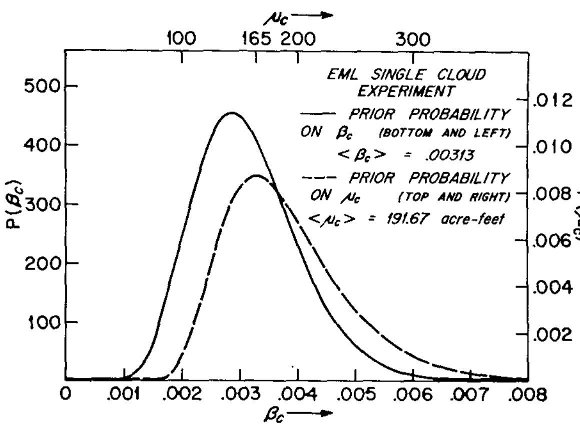

Fig. 11. As in Fig. 10 except for a more unfavorable prior since $\langle\mu_c\rangle=191.67$ acre-ft, larger than the mean of the control data sample.

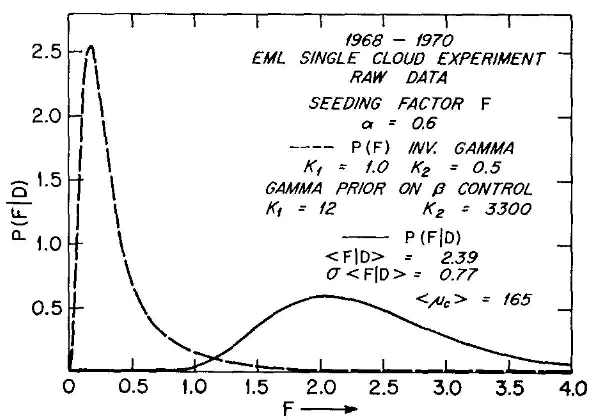

Fig. 12. Probability of seeding factor F corresponding to Fig. 10, with prior F an inverse gamma function (dashed).

TABLE 7. Recovery of gamma parameters from data drawn at random from a gamma function corresponding to raw control data.

	α	$\beta \times 10^2$	Probability gamma	$\bar{R}$ (acre-ft)
A. $n=20$				
Maximum value	1.1092	1.1870	0.4888	327.1277
Minimum value	0.3844	0.1730	0.0441	73.9079
Mean	0.6677	0.4769	0.2689	154.4864
Variance	0.0275	0.0392	0.0058	2384.9482
Standard deivation	0.1658	0.1979	0.0762	48.8359
B. $n=50$				
Maximum value	0.9947	0.6910	0.6408	265.3409
Minimum value	0.3959	0.2120	0.1034	91.8385
Mean	0.6137	0.3779	0 3679	168.6239
Variance	0.0124	0.0116	0.0121	880.1806
Standard deviation	0.1115	0.1075	0.1102	29.6678
C. $n=100$				
Maximum value	0.8516	0.5620	0.8166	218.2472
Minimum value	0.4659	0.2500	0.0762	118.1907
Mean	0.6026	0.3672	0.4596	167.2139
Variance	0.0055	0.0052	0.0334	498.9574
Standard deviation	0.0743	0.0722	0.1827	22.3374

statistics of the recovered parameters and their departures from the "real" parameters. We take $n=20$, 50 and 100 successively; the reason for the choice is that these are the numbers of control cases we might expect to obtain in single or multiple cumulus experiments in 2–10 years of work.

Table 7 was constructed with $m=100$. The column entitled "Probability gamma" shows the relative probability of the gamma distribution compared to other common distributions, such as log-normal, Weibull, etc. (see Simpson, 1972). Expert opinions consulted suggest that $m=100$ may be marginal for reproducibility but in analyzing the latter, we are aided by the extensive valuable work of Bowman and Shenton (1968, 1970) on the gamma distribution. They derived asymptotic expansions of the expectations, presenting tables of bias, standard deviations, etc., of the parameters as a function of the parameters themselves and of sample size. They also conducted Monte Carlo experiments with up to 10^5 cases each. Table 8 is reproduced from their figures.

With 26 cases, there is clearly a large standard deviation in α. Hence the optimum Bayesian analysis would be to allow for uncertainty in α, β_c and F, placing a joint prior probability on all three. The only assumptions retained are the gamma distributions, the equality of α between seed and control, and the multiplicative seeding factor. Bayes equation now is written

$$p(F,\mu_c,\alpha\,|\,\mathbf{x},\mathbf{y})=p(F,\mu_c,\alpha)\frac{p(\mathbf{x},\mathbf{y}\,|\,F,\mu_c,\alpha)}{p(\mathbf{x},\mathbf{y})}. \qquad (12)$$

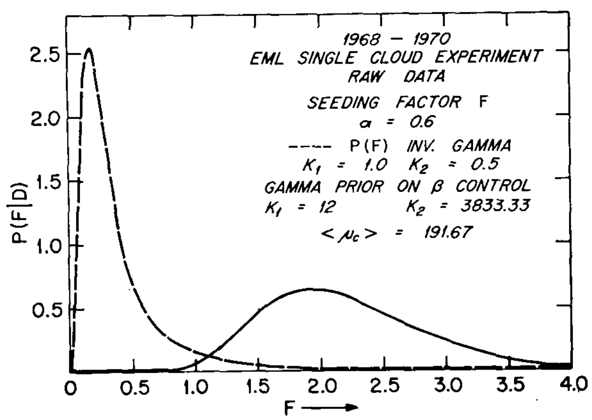

FIG. 13. Probability of seeding factor F corresponding to Fig. 11 with prior F an inverse gamma function (dashed).

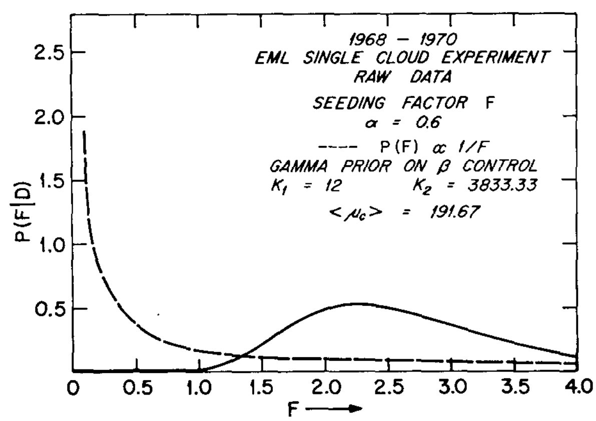

FIG. 14. Probability of seeding factor F corresponding to Fig. 11, with prior F proportional to $1/F$ (dashed).

The mathematics of solving this form of the equation has been presented by Olsen *et al.* (1975). For priors we chose α uniform in the range of 0.1 to 3.0, $p(\mu_c)$ proportional to $1/\mu_c$, and $p(F)$ proportional to $1/F$. These are unfavorable priors for the seeding factor. Results are shown in Figs. 15 and 16. We see that posterior α is quite sharply peaked at 0.6, a gratifying result. For F, the expected and modal values are 2.87 and 2.40, respectively, with a standard deviation of 1.1. Equal-tailed 95% confidence limits are 1.24 to 5.51. Despite all the uncertainties and unfavorable priors, there is little (0.6%) probability that the seeding factor is 1 or less and high probability (77%) that it exceeds 2.

5. Consideration of alternative distributions

The fit of the data to the gamma distribution is examined further in Fig. 17. The seeded and control points are plotted on gamma probability paper with identical shape parameter $\alpha = 0.6$, following the method outlined by Crutcher *et al.* (1973). If the fit were perfect the data should be along the straight line with slope 1. Actually, the highest two points of both distributions are off the scale of the drawing and all four lie above the straight line, suggesting that the distribution may be more heavy-tailed than a true gamma.

Since with the raw data, the relative probability and χ^2 test give a better fit for the log-normal distribution than for the gamma (Simpson, 1972), Fig. 18 is constructed with the data on log-normal paper. All points are depicted. The least-squares best-fit straight lines are shown, which are nearly parallel. They converge very slightly at the wet end, hinting that perhaps the seeding factor may be smaller for the wettest clouds. This may or may not be related to our finding that on "naturally rainy" days (defined in terms of echo area coverage), dynamic seeding effects on single-cloud rainfall in south Florida may be zero or negative (Simpson and Woodley, 1971). The seeding factor is readily estimated from Fig. 18 using the distance between the two straight lines near the middle of the range. The mean difference in $\ln R$ is 1.18 which when exponentiated gives $F = 3.25$, on the upper side of our various estimates.

Mielke[4] has kindly provided the development and results with our data of the Maximum Likelihood test using the log-normal distribution. Again there are two parameters, one relating to the scale and the other to the shape of the distribution. Statistics T_1 and T_2, respectively, are calculated to test whether seeded and control populations (now both assumed to be well fitted by log-normal distributions) differ significantly. T_1 for the scale parameter is 6.33, so that the populations differ at $p = 0.012$. T_2 for the shape parameter is 0.0178, with $p = 0.89$, so that the shape parameter is again shown not to differ between these seeded and control populations.

Mielke and Johnson (1973, 1974) have shown that the seeded data are as well or better fit by their beta-K (also termed a three-parameter K) and beta-P distributions as by either the gamma or log-normal. The desirable computational properties of the beta distributions will lead to their further investigation for this and other data sets from cumulus experiments.

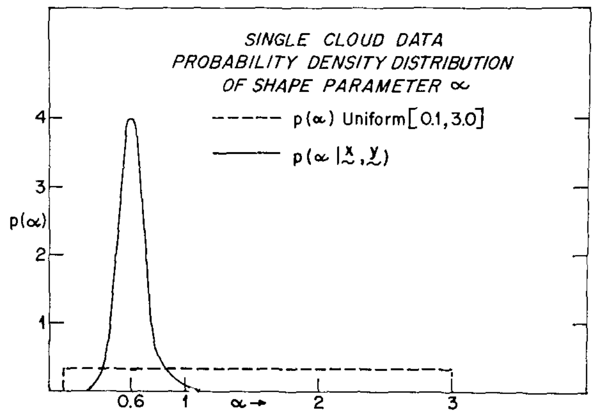

Fig. 15. Probability density of the gamma shape parameter α with prior uniform (dashed) 0.1 to 3.0; posterior (solid).

6. Possible effects of errors in the data

By careful comparison of radar rainfall estimations with gage measurements, Woodley *et al.* (1975) found

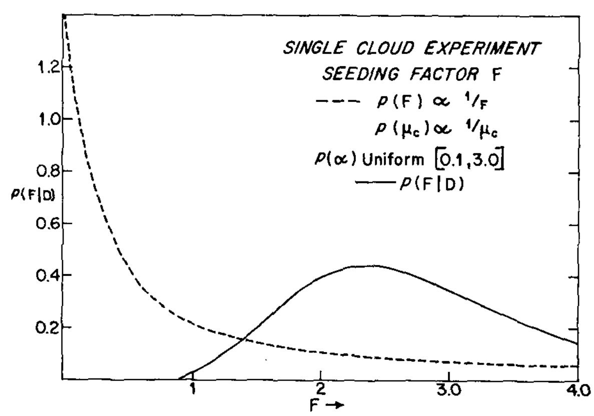

Fig. 16. Probability density distribution of seeding factor F when priors are placed on F (proportional to $1/F$, dashed), μ_c (proportional to $1/\mu_c$), and α (uniform). Posterior F is solid.

Table 8. Properties of recaptured gamma parameters as a function of sample size (After Bowman and Shenton, 1970).*

n	Mean bias α	Mean bias $1/\beta$	Fractional variance α	Fractional variance $1/\beta$
20	0.1311	−0.0425	0.1204	0.1503
50	0.0477	−0.0169	0.0344	0.0610
100	0.0231	−0.0084	0.0155	0.0306

* $\alpha = 0.6$ (corresponding to raw single cloud data).

[4] Personal communication.

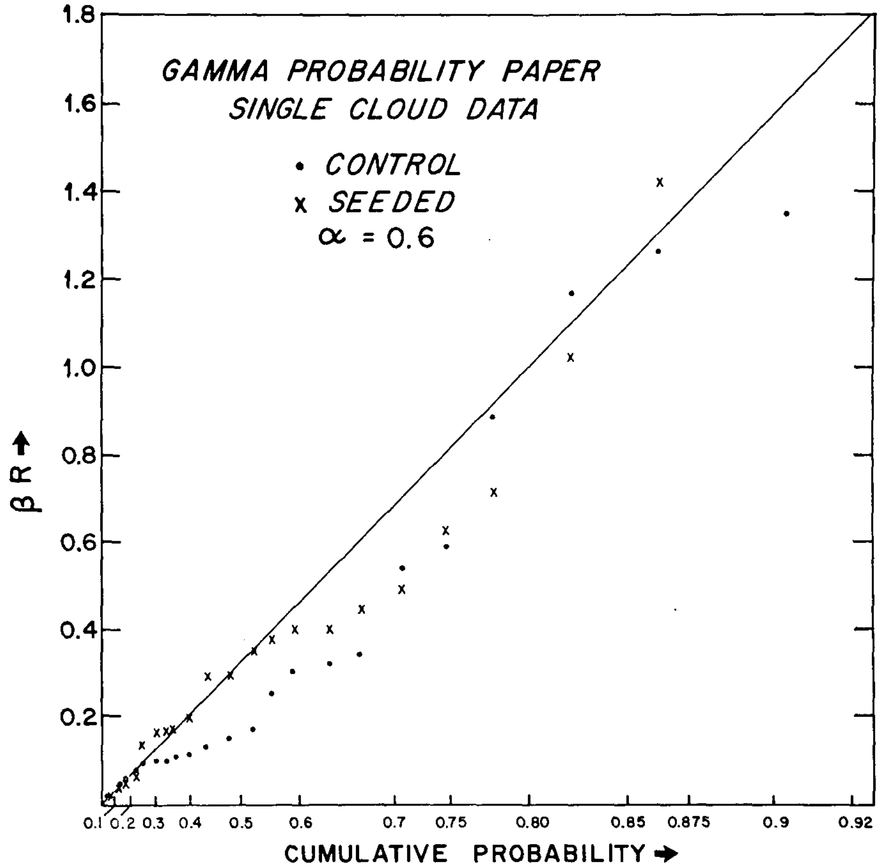

FIG. 17. Plot of the single-cloud data on gamma probability paper: control data ●'s; seeded data x's. If fit were perfect, data would be on the straight line. The two wettest points of both distributions are off the scale of the diagram and above the line, suggesting the data distribution is somewhat more heavy-tailed than a true gamma.

that out of 93 days, 80% of the radar estimates were within a factor of 2 of the cluster standard. The mean factor of difference was 1.51. The daily representation

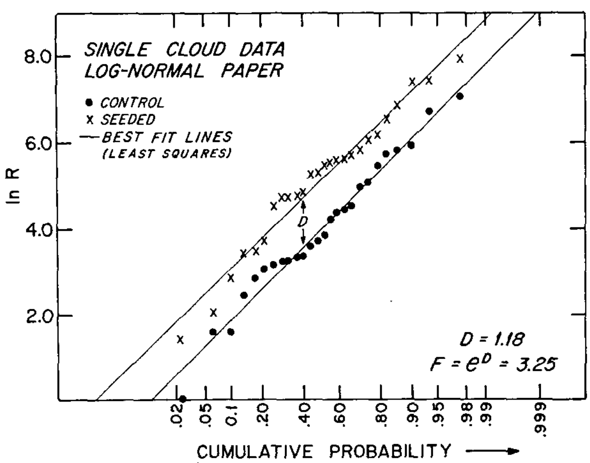

FIG. 18. Plot of the single-cloud data on log-normal probability paper. Control data ●'s; seeded data x's. Lines are least-squares best fits to data.

of rainfall by radar improves if one adjusts it using gages. For the EML mesonetwork in 1973, the radar representation of rainfall was adjusted by the ratio G/R (for each day) of the summed gage to radar rainfalls obtained from peripheral gage clusters. In the mean, this adjustment produced a statistically significant 15% improvement in radar accuracy. The distribution of G/R was well-fitted by a gamma function with $\alpha' = 3.98$ and $\beta' = 3.18$. For example, for the GO days in FACE 1973, G/R varied between 0.57 and 1.79.

The data in Table 1 are unadjusted. A total of 18 days produced the 52 GO clouds. To test the possible effect of measurement errors on seeding effect deductions, the following experiment was conducted: 100 sets of 18 values of G/R were drawn at random from the gamma distribution $\alpha' = 3.98$, $\beta' = 3.18$. These were then applied to the data in Table 1 in such a way that the same value of G/R was applied to all clouds on the same day. The 100 sets of 26 pairs of degraded data were then fed into the statistical test program, which involves six tests. Two tests depend on the gamma distribution, namely the Optimal $C(\alpha)$ and Maximum Likelihood. Two are t-tests using the logarithmic and fourth-root transforms

Table 9 Results of statistical tests with degraded single-cloud data.*

Test	Accept	Reject
Mann-Whitney-Wilcoxon	6	94
Squared rank	10	90
t-test (log)	18	82
t-test (fourth-root)	6	94
Likelihood Ratio	28	72
Optimal $C(\alpha)$	31	69

* Accept or reject null hypothesis at 5% significance level (two-tailed tests).

and two are nonparametric, using raw data. Results of the experiment are shown in Table 9 and Figs. 19 and 20.

The relatively low power of the Optimal $C(\alpha)$ and Likelihood Ratio tests puzzled us and was examined further. We first repeated the same experiment with an additional 100 sets of 26 pairs of degraded data, with virtually the same results as shown in Table 9. Next we examined 100 further cases where 52 different errors were generated and applied, one to each cloud. Results were again similar. The degraded data set summarized in Table 9 was then examined in detail. The sets for which the $C(\alpha)$ and MLR (Maximum Likelihood Ratio) tests accepted the null hypothesis were compared with the sets for which the null hypothesis was rejected by the same two tests. A particular examination was made of those 25 data sets where the null hypothesis was accepted by the $C(\alpha)$ test and rejected by the Mann-Whitney-Wilcoxon test. A combination of three factors appear to lead to the result in question: first, the ratio of "seed" to "control" was 26% lower in these cases than the average of 2.7 for the 100 cases; second, the fit of the gamma distribution was significantly worse for the degraded "control" data sets than for the actual data; and third, the $C(\alpha)$ and MLR tests are known to be more sensitive to small samples than the other tests compared.

Clearly the tests with hypothetical corrections applied here are more stringent than those which could be conducted if we knew the radar errors which were actually applicable to each case. The fact is that the data set in Table 1 is degraded by unknown errors and could be improved if the actually applicable G/R were known for each day. What we have done in the above test is to degrade the data further by arbitrarily applying randomly selected large factors of "correction." Even so, Fig. 19 shows that no "seed-control" ratios are reduced below 1.5 and only 15% are below 2. Furthermore, the most commonly used nonparametric test (Mann-Whitney-Wilcoxon) still detects a seeding effect at the 5% significance level in 94% of the cases.

7. Concluding remarks

The EML data set for randomized dynamic seeding of single cumuli has been examined in detail and a positive seeding effect on the rainfall has been established beyond reasonable doubt (to a significance level of 5% or better). The best assessment of the magnitude of the multiplicative seeding factor is between 2 and 3.

Methods of analysis have been applied which take into account the "heavy-tailed" character of the rainfall distributions. Rather large "errors" have been applied randomly to the data, which lower the seed-control difference below the 5% significance level in only 6 out of 100 trials with the Mann-Whitney-Wilcoxon tests.

While this experiment is mainly scientific in motivation, and may not have direct practical application to water resources in south Florida, part of its importance is the demonstration of one of the largest and most definitive seeding factors on cumulus rainfall. Its execution and evaluation have also built up a combined set of measurement, simulation and statistical tools for use in cumulus experiments elsewhere and for the more difficult application of dynamic seeding to many cumulus over a large target area. The ongoing Florida cumu-

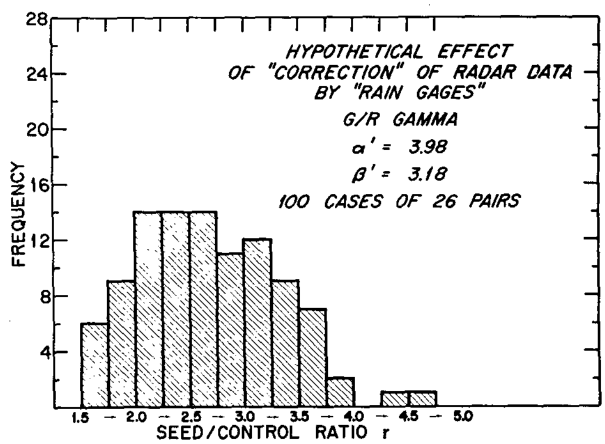

Fig. 19. Histogram of seed-control ratio r after a multiplicative set of "errors" is applied to the data 100 times.

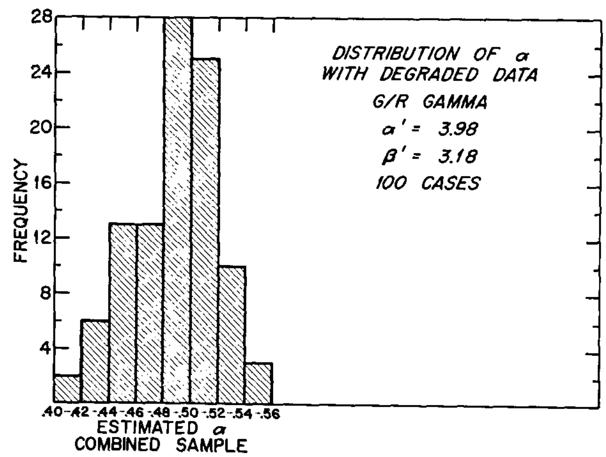

Fig. 20. Histogram of gamma shape parameter after application of multiplicative "errors" to data.

lus program is described elsewhere (Simpson and Woodley, 1975). The single-cumulus experiment illustrates many of the pitfalls encountered in attempts to modify convective rainfall, such as "heavy-tailed" distributions and offers some insights concerning how these pitfalls can be avoided and/or overcome.

Acknowledgments. The writers are grateful to their colleague, William L. Woodley, whose work has contributed greatly to that reported here. Paul Mielke helped with an important calculation and valuable criticisms.

Discussions with Glenn Brier, John Flueck and Earl Johnson were also extremely helpful. Paul Rosenzweig found and corrected an important error in an earlier draft and clarified some points regarding data transformations.

The manuscript was prepared at various stages by Lynn Bird, Michele Ochsner, Christine McKay, Mary Anglin and Mary Morris.

REFERENCES

Brier, G. W., 1974: Design and evaluation of weather modification experiments. *Weather and Climate Modification*, W. N. Hess, Ed., Wiley, 206–225.

——, G. F. Cotton, J. Simpson and W. L. Woodley, 1972: Cloud seeding experiments: Lack of bias in Florida series. *Science*, **176**, 163–164.

Bowman, K. O., and L. R. Shenton, 1968: Properties of estimators for the gamma distribution. CT-1, Computing Technology Center, Union Carbide Corporation, Nuclear Division, Oak Ridge, Tenn., 50 pp.

——, and ——, 1970: Small sample properties of estimators for the gamma distribution. CTG-28, Computing Technology Center, Union Carbide Corporation, Nuclear Division, Oak Ridge, Tenn., 160 pp.

Crutcher, H. L., G. L. Barger and G. F. McKay, 1973: A note on a gamma distribution computer program and graph paper. NOAA Tech. Rept. EDS 11, 92 pp.

Flueck, J., and B. S. Holland, 1973: Ratio estimation problems in meteorological research. *Preprints Third Conf. Probability and Statistics in Atmospheric Science*, Boulder, Amer. Meteor. Soc., 87–90.

Mielke, P. W., Jr., and E. S. Johnson, 1973: Three-parameter kappa distribution, maximum likelihood estimates and likelihood ratio tests. *Mon. Wea. Rev.*, **101**, 701–707.

——, and ——, 1974: Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology. *Water Resources Res.*, **10**, 223–226.

Olsen, A. R., 1975: Development and comparison of Bayesian and classical statistical methods as applied to randomized weather modification experiments. *J. Appl. Meteor.*, **14**, 970–973.

——, and W. L. Woodley, 1975: The effect of natural rainfall variability and measurement errors in the detection of seeding effect. *J. Appl. Meteor.*, **14**, 929–938.

——, J. Simpson and J. C. Eden, 1975: A Bayesian analysis of a multiplicative treatment effect using two samples. *Technometrics*, **17**, 161–166.

Simpson, J., 1972: Use of the gamma distribution in single cloud rainfall analysis. *Mon. Wea. Rev.*, **100**, 309–312.

——, and J. Pezier, 1971: Outline of a Bayesian approach to the EML multiple cloud seeding experiments. NOAA Tech. Memo. ERL OD-8, 43 pp.

——, and V. Wiggert, 1969: Models of precipitating cumulus towers. *Mon. Wea. Rev.*, **97**, 471-489.

——, and ——, 1971: 1968 Florida cumulus seeding experiment: Numerical model results. *Mon. Wea. Rev.*, **99**, 87–118.

——, and W. L. Woodley, 1971: Seeding cumulus in Florida: New 1970 results. *Science*, **172**, 117–126.

——, and ——, 1975: Florida area cumulus experiments 1970–1973. Rainfall results. *J. Appl. Meteor.*, **14**, 734–744.

——, W. L. Woodley, H. A. Friedman, T. W. Slusher, R. S. Scheffee and R. L. Steele, 1970: An airborne pyrotechnic cloud seeding system and its use. *J. Appl. Meteor.*, **9**, 109–122.

——, ——, A. H. Miller and G. F. Cotton, 1971: Precipitation results from two randomized pyrotechnic cumulus seeding experiments. *J. Appl. Meteor.*, **10**, 526-544.

——, J. C. Eden, A. R. Olsen and J. Pezier, 1973a: On the use of gamma functions and Bayesian analysis in evaluating Florida cumulus seeding results. NOAA Tech. Memo. ERL OD-15, 86 pp.

——, W. L. Woodley, A. R. Olsen and J. C. Eden, 1973b: Bayesian statistics applied to dynamic modification experiments on Florida cumulus clouds. *J. Atmos. Sci.*, **30**, 1178–1190.

Woodley, W. L., 1970: Precipitation results from a pyrotechnic cumulus seeding experiment. *J. Appl. Meteor.*, **9**, 242–257.

——, and A. Herndon, 1970: A raingage evaluation of the Miami reflectivity rainfall rate relation. *J. Appl. Meteor.*, **9**, 258–264.

——, A. R. Olsen, A. Herndon and V. Wiggert, 1975: Comparison of gage and radar methods of convective rain measurement. *J. Appl. Meteor.*, **14**, 909–928.