A Monte Carlo Technique for Designing Cloud Seeding Experiments[1]

PAUL T. SCHICKEDANZ[2] AND WAYNE L. DECKER

Dept. of Atmospheric Science, University of Missouri, Columbia

(Manuscript received 1 May 1968, in revised form 26 November 1968)

ABSTRACT

The design of a field experiment in rainfall augmentation requires prior estimates of the duration of the experiment and the density of raingages. A "Monte Carlo" method was developed to generate synthetic climatological rainfall data for various time periods and densities of raingages. The method was applied to a hypothetical cloud seeding experiment. Rainfall data for reporting networks were simulated and the resulting data were used to estimate the change in error variance induced by varying the density in a raingage network and the length of the experiment. The "*t*" test was applied to the simulated nontransformed data which were skewed and to data normalized by a transformation. In addition, the generalized likelihood ratio test was used to test for differences in location parameters of the seeded and nonseeded gamma distributions having a common shape factor.

The applicability and limitations of the method are discussed. With proper consideration of the limitations and with additional research on the problems encountered, it should be possible to obtain a preliminary estimate of the error variance of a proposed experimental design for many areas and various conditions.

1. Introduction

The results of many cloud seeding experiments concerned with increasing precipitation have been inconclusive, whereas others have claimed either substantial increases or decreases in rain. After careful re-evaluation of several *operational* cloud seeding projects in the eastern United States, the National Academy of Sciences (1966) issued a report which concluded that these projects had increased rainfall by 10–20%. The results of two major randomized *experimental* programs in the United States, however, present a different picture. The results of the experiment in southeastern Arizona (Battan, 1966) indicate there is no evidence that airborne silver-iodide seeding of convective clouds increased rainfall or influenced its areal extent. Indeed, it was further concluded that if there was any seeding effect at all, it was to cause a reduction in rainfall. The results of Project Whitetop in Missouri (Decker and Schickedanz, 1966) indicated that less rain fell into seeded areas than fell into the control areas, although the statistical tests employed were not consistent in demonstrating that this difference was significant.

Many times the results of a weather modification experiment are declared inconclusive because the experiment was not conducted for a long enough period of time to obtain significance for a given percentage increase and experimental error. Sometimes it is implied that the network of rainfall stations is not dense enough to detect significant results. Preliminary statistical evaluations of the area in which a cloud seeding experiment is proposed could provide estimates of the error variance and power of the statistical test for various raingage densities and experimental durations. The objective of this paper is to develop a technique whereby such estimates can be made.

[1] Contribution from the Missouri Agricultural Experiment Station, Journal Series No. 5387.

[2] Present affiliation: Illinois State Water Survey, Urbana.

2. The basis of generating data for a synthetic raingage network

a. The requirement for synthetic data

When designing a cloud seeding experiment, it would be advisable to conduct uniformity trials from existing data. Unfortunately, there is not usually available a sufficiently dense network of rainfall reporting stations from which to derive these studies. A method is developed here for generating synthetic data for a dense network which will have the same variability as the existing data.

The method must provide areal average values for days characteristic of the climate of the area and with a variation in the mean rainfall between days identical to that expected for the area. Further, the variation in rainfall between stations should be the same as encountered by networks of corresponding density. It would be similarly beneficial for the generation scheme to account for the correlation between stations. However, for most cloud seeding experiments in which seeded and nonseeded days are randomized over a single target area, the daily rainfall statistic is obtained by a summation or average of the gages on a particular day. This daily statistic becomes the estimated rainfall for the experimental unit in the statistical analysis. When this expression for the experimental unit is used, the correlation between stations in the experimental area does not contribute to the statistical analysis.

b. The mathematical model

In deriving the mathematical model for simulating rainfall data, the following notation was adopted:

X_{ijv} the rainfall recorded for gage v on day j and receiving treatment i.

$\bar{X}_{ij\cdot}=\sum_{v=1}^{m_{ij}} X_{ijv}/m_{ij}$, the average rainfall from gage observations with rain for day j, receiving treatment i

m_{ij} the number of gages with rain, receiving treatment i on the jth day

l_{ij} the number of gages without rain, receiving treatment i on the jth day

$\bar{X}'_{ij\cdot}=\sum_{v=1}^{s_{ij}} X_{ijv}/s_{ij}$, the average rainfall from all gage observations for day j, receiving treatment i

$s_{ij}=m_{ij}+l_{ij}$, the total number of gages for day j, receiving treatment i

$\bar{X}_{i\cdot\cdot}=\sum_{j=1}^{n_i} \bar{X}_{ij\cdot}/n_i$, the daily mean rainfall for days with rain and receiving treatment i, averaged over the $\bar{X}_{ij}$.

n_i the number of days with rain, receiving treatment i

q_i the number of days without rain, receiving treatment i

$\bar{X}'_{i\cdot\cdot}=\sum_{j=1}^{n_i} \bar{X}'_{ij\cdot}/n_i$, the daily mean rainfall for days with rain and receiving treatment i, averaged over the $\bar{X}'_{ij}$.

$\bar{\bar{X}}_{i\cdot\cdot}=\sum_{j=1}^{t_i} \bar{X}_{ij\cdot}/t_i$, the daily mean rainfall for all days receiving treatment i, averaged over the $\bar{X}_{ij}$.

$t_i=n_i+q_i$, the total number of days receiving treatment i

$\bar{\bar{X}}'_{i\cdot\cdot}=\sum_{j=1}^{t_i} \bar{X}'_{ij\cdot}/t_i$, the daily mean rainfall for all days receiving treatment i, averaged over the $\bar{X}'_{ij}$

$i=1$ implies variable corresponding to this subscript is nonseeded

$i=2$ implies variable corresponding to this subscript is seeded

$P_k(X_{ijv}>0)$ the number of gages with rain in the kth class interval of $\bar{X}_{ij\cdot}$ divided by the total number of gages in the kth class interval, i.e., the probability of rain for the X_{ijv} gage of the kth class interval

$P_k(X_{ijv}=0)$ the number of gages without rain in the kth class interval of $\bar{X}_{ij\cdot}$ divided by the total number of gages in the kth class interval, i.e., the probability of no rain for X_{ijv} gage of the kth class interval

$P(\bar{X}_{ij\cdot}>0)=n_i/t_i$, the probability of rain for day j

$P(\bar{X}_{ij\cdot}=0)=q_i/t_i$, the probability of no rain for day j

The model employed is based on the mixed distribution function,

$$G=P(X<a)=P(X=0)+P(X>0)\cdot P(X<a|X>0), \quad (1)$$

where $P(X<a)$ is the probability of receiving less than a specified amount of rain, $P(X=0)$ the probability of receiving no rain, $P(X>0)$ the probability of receiving some rain, and $P(X<a|X>0)$ the probability of receiving less than a specified amount of rain, given that rain occurs.

The term $P(X<a|X>0)$ is given by

$$F(x)=P(X<a|X>0)=\int_{x>0}^{a} f(x)dx, \quad (2)$$

where the density function $f(x)$ is assumed to be the well-known gamma density

$$f(x)=\frac{1}{\beta^\gamma\Gamma(\gamma)}x^{\gamma-1}e^{-x/\beta},\ x>0,\ \gamma>0,\ \beta>0, \quad (3)$$

where the symbols β and γ are the location and shape factors, respectively, and are estimated by the method of maximum likelihood (Thom, 1958).

Fig. 1 is an illustration of the relationship between $\bar{X}_{ij.}$ (average rainfall for gages with rain) and the corresponding mixed distribution G. The left side of the figure is a representation of $P(X_{ijv}=0)$ (the probability function of zero gage rainfall). As $\bar{X}_{ij.}$ increases in value, the corresponding distribution of $P(X_{ijv}>0)$ becomes less skewed and the probability of zero gage rainfall decreases. There is assumed to be a reversal in the trend of the $P(X_{ijv}=0)$ function for large values of $\bar{X}_{ij.}$. This is due to the nature of summertime convective storms. A heavy rain storm may have a very limited areal extent and only a few gages will be affected. Thus, the average of gages with rain may be large, and the probability of zero rainfall for a gage in the area may be larger than at a smaller value of $\bar{X}_{ij.}$.

In actual practice, there is considerable sampling variability between the distributions and the trend to become less skewed is somewhat masked. To reduce the sampling variability, the values of $\bar{X}_{ij.}$ were partitioned into r class intervals, the X_{ijv} values corresponding to each $\bar{X}_{ij.}$ in the kth class interval being grouped together to form a single frequency distribution. If the $\bar{X}_{ij.}$ amounts for the period of record are divided into r class intervals, there will result a frequency distribution for each class interval. As illustrated in Fig. 1 the distributions associated with the smaller daily means will be skewed the most.

We now consider the frequency distribution of the $\bar{X}_{ij.}$, it being assumed that this distribution is also approximated by the gamma function of the form shown in (3). We then estimate the statistics for this gamma function from the daily mean rain ($\bar{X}_{ij.}$) obtained from the historical climatological record. Random samples from this distribution allow for the selection of daily averages of the area for as many days as desired in a simulated cloud seeding experiment.

It is further desired to select, at random, simulated records from as many raingages as required. Each mean daily rainfall ($\bar{X}_{ij.}$) selected at random is associated with one of r class intervals. The mixed distribution of (1) was estimated for each class interval, and raingage values (X_{ijv}) from this mixed distribution were drawn at random.

3. Method for simulation of rainfall records

a. The nonseeded sample

The first step in the actual generation is to determine the number of days n_i with rain and the number of days q_i without rain in the experimental area. These days can be determined by using the probability of rain for day j, $P(\bar{X}_{ij.}>0)$, and the probability of no rain, $P(\bar{X}_{ij.}=0)$. The procedure is to select a number T_j at random ($0\leq T_j\leq 1$) and to compare it to the probabilities of rain and of no rain. If $T_j\leq P(\bar{X}_{ij.}>0)$, the first day is designated as a rain day, and if $T_j>P(\bar{X}_{ij.}>0)$, the day is considered to be a day without rain. This procedure is continued until all of the nonseeded t_i

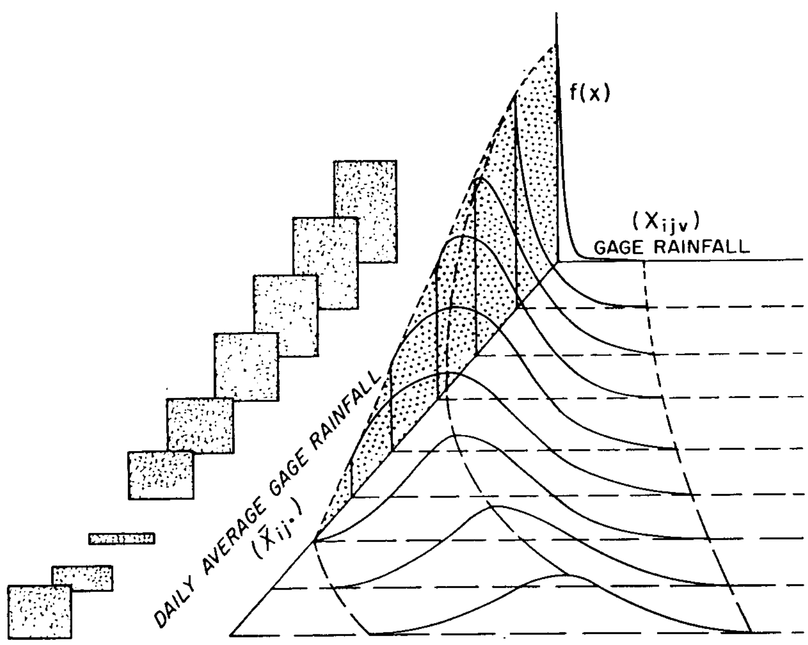

FIG. 1. The mixed probability function for the daily average gage rainfall $\bar{X}_{ij.}$. As $\bar{X}_{ij.}$ increases, the corresponding distribution of X_{ijv} becomes less skewed and $P(X_{ijv}=0)$ (the probability of zero gage rainfall) decreases.

days have been partitioned into days with and without rain.

To determine the daily areal mean rainfall for a day with rain, one must consider another set of random numbers T_j. These numbers are uniformly distributed on the interval $(0 \le T_j \le 1)$. If $(0 \le a < b \le 1)$, then the probability function of T_j can be expressed as

$$P(a < T_j < b) = \int_a^b 1 dt_j = b - a. \quad (4)$$

The cumulative probability values from any distribution of a continuous variate are uniformly distributed on the interval (0,1). Designating $F(x)$ to be a cumulative probability from a given distribution, it is obvious that the selection of the random number T_j is equivalent to selecting a value of $F(x)$ at random. Now the inverse function of any probability function can be expressed as

$$F^{-1}(x) = x = G[F(x)], \quad (5)$$

where $G[F(x)]$ is a function of cumulative probabilities. Therefore, a random variable X can be selected from any distribution by using the above relationships.

Having selected the $\bar{X}_{ij\cdot}$, the gage values for various days can now be generated. Let u be the number of days in the kth class interval, m_{ij} the number of gages with rain, and s_{ij} the total number of gages to be generated. Then a uniformly distributed random number T_v is selected and compared to the $P(X_{ijv} > 0)$ for the distribution of X_{ijv} associated with the kth class interval. Again, if $T_v \le P(X_{ijv} > 0)$, an observation of rain is determined, while if $T_v > P(X_{ijv} > 0)$, an observation of no rain is determined. This procedure is continued until the s_{ij} observations for day j have been established according to rain and no rain cases.

Next, m_{ij} numbers are selected at random and the cumulative probabilities $F(x)$ of (5) are determined. The sample value of X_{ijv} is then determined from the inverse function for each $F(x)$ until a random sample of size m_{ij} has been acquired. This sample combined with the zero values constitutes the sample of s_{ij} gages for the day j of the distribution of the class interval k. This procedure is then continued until samples have been generated for u days of the kth class interval. The same method is then used to acquire the raingage values for all class intervals. The generated gage observations for each day are then averaged to obtain the $\bar{X}_{ij\cdot}'$ values for each day.

In order to determine the effect of different sized samples on the variation of the treatment mean, the sample is changed to other values and the generation scheme for the X_{ijv} repeated until all desired samples have been obtained.

b. The seeded sample

To simulate a weather modification experiment, it is necessary to obtain an estimate of the $\bar{X}'_{ij\cdot}$ values for days receiving a treatment. This estimate can be obtained by assuming that seeding has no effect on the shape factor γ of the gamma distribution, i.e., that $\gamma_1 = \gamma_2$. Then, for certain percentage increases in $\bar{X}_{1\cdot\cdot}$, a corresponding change in β can be computed from the relation $\beta = \bar{X}/\gamma$. By this method, sets of β_2 and $\bar{X}_{2\cdot\cdot}$ would be computed for various increases in the mean precipitation. Using β_2 and $\bar{X}_{2\cdot\cdot}$, the generation procedure for the X_{2jv} would be identical to the method used for generating the raingage values in the nonseeded case.

Since the method described above requires a great deal of computer programming and processing of additional data, an approximate estimate was used. By this method the number of gages and days required to obtain significance were determined in terms of a particular experimental design and statistical test. First, the components of the test for the nonseeded sample were computed. Then, with certain inferences concerning the seeded sample, the differences required for significance can be obtained through algebraic relations.

4. The experimental model

a. The design of the experiment

Consider an experimental plan which allows for randomization of seeded and nonseeded days over a single target area (the days define the experimental units and the gages are the sampling units). Out of the total number of days in which the experiment is to be conducted, t_1 days will be nonseeded and t_2 days will be seeded. In this simulated experiment the seeded days are selected at random but restricted so that one-half of the days are seeded and one-half are nonseeded.

If the above experiment were to represent a factorial arrangements of seeded conditions and days, both treatment effects (i.e., seed and nonseed) would have to appear on each day. Since this is virtually impossible to achieve, the experimental design proposed is an extended, completely randomized design with subsampling (Ostle, 1954, Steele and Torrie, 1960) which utilizes the nested or hierarchical classification analysis of variance for estimation. This appropriate analysis of variance is shown in Table 1, together with expected values of the mean squares.

It is seen that $\sigma_i'^2$, the variance between days treated alike (i.e., the experimental error), is composed of two components of variation. These are $\sigma_\epsilon'^2$, the variance among gages, and $\sigma_D'^2$, the variance among days adjusted for treatment effects. However, in most seeding experiments, the rainfall for a particular day is determined from the sampling units (i.e., rain recorded in the gages) for the day and the subsequent analysis is based on the daily average gage rainfall. Also, in many instances due to non-normality, the mean rainfall for the experimental units (days) is transformed while the data from the sampling units are not transformed.

TABLE 1. Analysis of variance for the cloud seeding experiment.

Source of variation	Sum of squares	Degrees of freedom	Mean square	Expected mean square
Between treatments	$SS=st\sum_{i=1}^{r}(\bar{\bar{X}}'_{i..}-\bar{\bar{X}}'_{...})^2$	$r-1$	$SS/(r-1)$	$\sigma_\epsilon'^2+s\sigma_D'^2+\frac{st}{r-1}\sum_{i=1}^{r}\tau_i^2$
Among days treated alike	$SS=s\sum_{i=1}^{r}\sum_{j=1}^{t}(\bar{X}'_{ij.}-\bar{\bar{X}}'_{i..})^2$	$r(t-1)$	$SS/r(t-1)$	$\sigma_\epsilon'^2+s\sigma_D'^2$
Among gages on days treated alike	$SS=\sum_{i=1}^{r}\sum_{j=1}^{t}\sum_{v=1}^{s}(X_{ijv}-\bar{X}'_{ij.})^2$	$rt(s-1)$	$SS/[rt(s-1)]$	$\sigma_\epsilon'^2$
Total	$SS=\sum_{i=1}^{r}\sum_{j=1}^{t}\sum_{v=1}^{s}(X_{ijv}-\bar{\bar{X}}'_{...})^2$	$rts-1$		

Math model: $X_{ijv}=\mu+\tau_i+D_{ij}+\epsilon_{ivj}$, where

μ = overall mean estimated by $\bar{\bar{X}}'_{...}$
τ_i = effect of treatment i
D_{ij} = effect of the day j to which treatment i is applied
ϵ_{ijv} = effect of gage v on day j subject to treatment i
$t=t_1=t_2$ and $s=s_1=s_2$

Therefore, the direct relation between $\sigma_D'^2$, $\sigma_\epsilon'^2$ and the experimental error is not obvious. Hence, the usual procedure is to compute the relation

$$\sigma_i'^2=\sum_{j=1}^{t}(\bar{X}'_{ij.}-\bar{\bar{X}}'_{i..})^2/(t-1),$$

where it can be shown that $\sigma_i'^2$ must decrease as the number of gages increases.

b. Transformations to normal distributions

The frequency distributions associated with mean daily rainfall are positively skewed (mode and median less than the mean), and should not be used in tests of significance which require normality. The large number of zeros at the lower bound of the distribution causes an even greater departure from normality than is present in data containing only rain days. These zero values tend to make tests of significance, which are based on the normal distribution, even less valid. Also, the likelihood ratio test, as presented later, cannot be used when zeros are present in the data. Therefore, the zero rain days should be removed from the data before proceeding with further evaluation. When the estimate of variance is based on rain days only, the prime notation is removed from the symbols for the variance.

Transformations of data from skewed distributions are used widely in practice, one such transformation being that where each $\bar{X}'_{ij.}$ is replaced by $\log_{10}(1+\bar{X}'_{ij.})$. The means and variances were computed for the logarithmic transformed data to show the effect of transformations on the data. The following notation will be adopted to differentiate between nontransformed and transformed data:

- $\bar{X}'_{ij.}$ the nontransformed mean of all gages for day j and receiving treatment i
- $\bar{\bar{X}}'_{i..}$ the mean of the $\bar{X}'_{ij.}$ over rain days with the same treatment
- $\hat{\sigma}_i^2$ the variance of the $\bar{X}'_{ij.}$ about their mean
- $\bar{X}'_{Lij.}$ the logarithmic transformed mean of all gages for day j and receiving treatment i
- $\bar{\bar{X}}'_{Li..}$ the mean of the $\bar{X}'_{Lij.}$ over rain days with the same treatment
- $\hat{\sigma}_{Li}^2$ the variance of the $\bar{X}'_{Lij.}$ about their mean

The relation between each variance ($\hat{\sigma}_i^2$, σ_{Li}^2) and the sample size s was established by fitting an equilateral hyperbola asymptotic to lines parallel to both axes. This curve allowed $\hat{\sigma}_i^2$ to approach infinity for extremely small samples and decrease to a constant value as sample size approaches infinity. The equation of the equilateral hyperbola is

$$1/Y=a+b/X,$$

where the constants a and b are determined by a least-squares procedure. The actual fitting of the curve was obtained by a simple linear regression by setting $Y'=1/Y$ and $X'=1/X$. Substituting $\hat{\sigma}_1^2$ for Y and s_1, the number of raingages, for X, estimates of the variance of the daily areal mean rainfall for the nonseeded days can be obtained for use in further analysis. The equilateral hyperbola was also fitted to the sample values of γ and β for various raingage densities. The estimates of γ and β, which are obtained from the fitted curve, are used later in the maximum likelihood ratio test and in deriving estimates of the seeded variances.

In order to compare simulated seeded increases in the mean on the nontransformed scale to those on the transformed scale, a relation between the nontransformed and transformed means is required. From the method of Neyman and Scott (1960) a comparison can be made between the mean values of the nontransformed and logarithmic transformed variate by using

$$\bar{X}'_{i\cdot\cdot} = \exp[m^2\hat{\sigma}_{Li}^2/2) + (m\bar{X}'_{Li\cdot\cdot})] - 1, \quad (6)$$

where m is a constant equal to 2.3026.

Since the means and variances of skewed data are not independent, the variances of the treated and nontreated data will not be the same. Since the seeded sample was not generated, an estimate must therefore be obtained for the seeded variances on the nontransformed and transformed scales. Since the data are gamma distributed, estimates of the seeded nontransformed variance can be obtained if the assumption of common shape γ is made. The mean of the gamma distribution is $\gamma\beta$ and the variance is $\gamma\beta^2$. Hence, for arbitrary increases in $\bar{X}'_{i\cdot\cdot}$, the corresponding change can be calculated for $\hat{\sigma}_i^2$ assuming that $\hat{\gamma}$ does not vary.

The variances of the nontransformed and log-transformed data are related by (Aitchison and Brown, 1957)

$$m^2 \exp[\hat{\sigma}_{Li}^2] = [\hat{\sigma}_i^2/(\bar{X}'_{i\cdot\cdot}+1)^2] + 1. \quad (7)$$

Using (7), estimates of the seeded transformed variance can be obtained for various increments of increase in the nontransformed data. With the use of (6) the corresponding difference in the transformed means can be obtained.

c. The tests of significance

Since the distribution of $\bar{X}'_{ij\cdot}$ is skewed, the use of the "t" test is questionable (Mood and Graybill, 1963). The "t" test for nontransformed data will be applied in order to compare it to other tests. The form of the "t" test used was

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_d},$$

where $\hat{\sigma}_d$ is given by

$$\hat{\sigma}_d = \left(\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}\right)^{\frac{1}{2}},$$

$\bar{X}_1$ and $\bar{X}_2$ are the means, and $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are the variances of the nonseeded and seeded days. The "t" test should also be applied to the transformed data for various increases in the mean. The power of the "t" test for nontransformed and transformed data can then be computed.

Since the data are gamma distributed, a test which involves gamma distributed data would seem to be the most logical. We thus use the generalized likelihood ratio test, which is applicable to large samples, and which assumes common shape factors of the gamma distributions, in the form[3] (Schickedanz, 1967)

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \frac{\left(\dfrac{1}{\hat{\beta}^{\hat{\gamma}}\Gamma(\hat{\gamma})}\right)^{n_1+n_2} \prod_{i=1}^{n_1} x_{1i}^{(\hat{\gamma}-1)}}{\left(\dfrac{1}{\hat{\beta}_1{}^{\hat{\gamma}'}}\right)^{n_1}\left(\dfrac{1}{\hat{\beta}_2{}^{\hat{\gamma}'}}\right)^{n_2}\left(\dfrac{1}{\Gamma(\hat{\gamma}')}\right)^{N} \prod_{i=1}^{n_1} x_{1i}^{(\hat{\gamma}'-1)}} \times \frac{\prod_{j=1}^{n_2} x_{2j}^{(\hat{\gamma}-1)} \exp[-(\sum_{i=1}^{n_1} x_{1i} + \sum_{j=1}^{n_2} x_{2j})/\hat{\beta}]}{\prod_{j=1}^{n_2} x_{2j}^{(\hat{\gamma}'-1)} \exp[-(\sum_{i=1}^{n_1} x_{1i}/\hat{\beta}_1 + \sum_{j=1}^{n_2} x_{2j}/\hat{\beta}_2)]}, \quad (8)$$

where n_1 is the number of observations in the sample from the first distribution, n_2 the number of observations in the sample from the second distribution, $(x_{11}, x_{12}, \cdots x_{1n_1})$ the observation in the sample from the first distribution, $(x_{21}, x_{22}, \cdots x_{2n_2})$ the observations in the sample from the second distribution, and $\Gamma(\hat{\gamma})$, $\Gamma(\hat{\gamma}')$ the incomplete gamma functions for $\hat{\gamma}$ and $\hat{\gamma}'$, respectively.

In addition:

$$\hat{\gamma} = \frac{1+\sqrt{1+4A/3}}{4A},$$

$$A = \ln\bar{\bar{x}} - (\sum_{i=1}^{n_1} \ln x_{1i} + \sum_{j=1}^{n_2} \ln x_{2j})/N,$$

$$N = n_1 + n_2,$$

$$\bar{\bar{x}} = (\sum_{i=1}^{n_1} x_{1i} + \sum_{j=1}^{n_2} x_{2j})/N,$$

$$\hat{\beta}_1 = \bar{x}_1/\hat{\gamma}',$$

$$\hat{\beta}_2 = \bar{x}_2/\hat{\gamma}',$$

$$\hat{\beta} = \bar{\bar{x}}/\hat{\gamma},$$

$$\hat{\gamma}' = \frac{1+\sqrt{1+4A'/3}}{4A'},$$

$$A' = (n_1\ln\bar{x}_1 + n_2\ln\bar{x}_2)/N - \frac{1}{N}(\sum_{i=1}^{n_1} \ln x_{1i} + \sum_{j=1}^{n_2} \ln x_{2j}).$$

Since the seeded sample was not generated, the effect of seeding was expressed as percentage changes in the mean; the substitutions

$$\sum_{i=1}^{n_1} x_{1i} = n_1\bar{x}_1, \quad \sum_{j=1}^{n_2} x_{2j} = n_2\bar{x}_2$$

[3] The maximum likelihood estimates were approximated by truncating a series expansion of the digamma function. Recently, Thom (1968) has discussed the error involved due to truncation of this series.

were therefore made in (8). Also, since the individual observations of the seeded sample were not available, the product terms involving $\hat{\gamma}'$ could not be estimated, and it was assumed that $\hat{\gamma}' = \hat{\gamma}$. Under these conditions (8) reduces to

$$\lambda = \frac{(1/\hat{\beta}^{\gamma})^N \exp[-(n_1\bar{x}_1+n_2\bar{x}_2)/\hat{\beta}]}{(1/\hat{\beta}_1^{\hat{\gamma}})^{n_1}(1/\hat{\beta}_2^{\hat{\gamma}})^{n_2}\exp[-(n_1\bar{x}_1/\hat{\beta}_1+n_2\bar{x}_2/\hat{\beta}_2)]}. \quad (9)$$

Finally, for ease in computation it is best to take natural logarithms of both sides of (9) giving

$$\ln\lambda = -N_\gamma \ln\hat{\beta} - (n_1\bar{x}_1+n_2\bar{x}_2)/\hat{\beta} + (n_1\hat{\gamma}\ln\hat{\beta}_1 + n_2\hat{\gamma}\ln\hat{\beta}_2) + (n_1\bar{x}_1/\hat{\beta}_1+n_2\bar{x}_2/\hat{\beta}_2). \quad (10)$$

The quantity $-2\ln\lambda$ is approximately distributed as Chi square with 1 degree of freedom. The null hypothesis is $\beta_1=\beta_2$ and the alternative hypothesis is $\beta_1 \neq \beta_2$. Tests of significance are made from a Chi square table. The power of the test was computed using tables of the noncentral Chi square (Fix, 1954).

The testing of the likelihood ratio using (8) rather than (9) will be performed at the Illinois State Water Survey. This study will include comparison of transformed and nontransformed data for a wide spectrum of shape and location parameters. The power of the test under a wide range of conditions will also be included.

5. An example of a simulated cloud seeding experiment

To demonstrate the application of this method in planning a cloud seeding experiment, a 5625 mi² area was selected in central Missouri. The hypothetical experiment was to operate only during the 1–15 June period. Three questions were to be answered by the simulation: 1) How many days would the experiment need to be run to detect a given increase in precipitation? 2) How dense a raingage network is required to detect a given increase of precipitation? 3) Which of the available statistical tests were the most powerful?

During the 31-year period from 1935–1965, there were 10–17 raingages in a 75 mi² area with Columbia, Mo., at the center. All of the days with rain were tabulated. For this area 71% of days during the period 1–15 June had rain while 48% of the raingages reported rain over all 465 days. The mean rainfall was computed for each day, and frequency distributions for the raingage records from 17 class intervals of the mean daily rainfall were tabulated. These frequency distributions were similar to those shown in Fig. 1.

Using the technique outlined in Section 3a, raingage measurements were simulated for experiments of from 25–250 days duration and with raingage densities of from one gage per 2860 mi² to one gage per 22 mi². The mean rainfall, the transformed means and the corresponding variances were computed by the method outlined in Sections 4a and 4b.

The seeding effect was simulated by assuming an increase from 10–100% in the mean daily rainfall. Unless the seeded sample is actually generated, the seeded variances cannot be estimated for the smaller raingage samples. The reason they cannot be estimated is because in the generation of data from the unseeded distribution of gages it is possible to generate all zeros for the gage values even though the randomly selected mean was greater than zero. This is most likely to occur if the probability of rain is low and the sample of gages is small. Thus, the average rain for the day is zero, even though it was designated to be a rain day. (The same effect occurs in nature when some storms are not recorded due to a sparse network of gages.) Since γ and β can only be estimated for samples in which each $\bar{X}_{ij\cdot} > 0$, the estimates of γ and β are not available for the nonseeded or seeded cases. With the relationship presented earlier the means and variances were estimated for each increase where the density of gages was 16 (1 per 352 mi²) or greater. From these values tests of significance were computed. A comparison of day, gage combinations at which the test statistic is significant is made in Table 2. This table lists the day, gage combination at which the statistical tests become significant for various increases in the mean.

Considering the 5% level of significance, the likelihood ratio test detects significant differences at lower gage and day combinations than the other tests. The "*t*" test for nontransformed data detects significant differences at lower gage and day combinations than it does for transformed data.

Test statistics were computed only for selected days and gages. Therefore, it is possible to obtain significance for more than one test at the same day, gage combination. For example, the "*t*" test for nontransformed data and the maximum likelihood ratio test are significant at the same combination for an 80% increase. Therefore, it is of interest to compare the probabilities of the significant test statistic for nontransformed data to the probability of the other test statistics for the same day, gage combination. This comparison is shown in Table 3, along with the power of the test.

The fact that the likelihood ratio test is the most powerful was expected because the distribution employed allows for the skewed nature of the data. The implication that the "*t*" test for nontransformed data

TABLE 2. The day, gage combination for which the test statistic is significant for a two-sided test at the 0.05 probability level.

Test statistic		Per cent of "true" mean increase 10	20	40	60	80	100
"*t*" test for non-	No. of days	—	—	100	75	50	50
transformed data	No. of gages	—	—	48	16	16	16
"*t*" test for log	No. of days	—	—	175	100	75	50
transformed data	No. of gages	—	—	48	24	16	16
likelihood ratio test	No. of days	—	—	100	50	50	25
	No. of gages	—	—	24	32	16	16

TABLE 3. Comparison of the probability and power of the significant nontransformed "t" to the other test statistics for the same combination of days and gages.

Significant test statistic		Per cent of increase in "true" mean 10	20	40	60	80	100
"t" test for nontransformed data	Probability	—	—	0.0497	0.0266	0.0212	0.0110
	Power	—	—	0.502	0.612	0.603	0.721
"t" test at same day, gage combination for log data	Probability	—	—	0.1466	0.0963	0.0918	0.0480
	Power	—	—	0.310	0.383	0.391	0.505
likelihood ratio test at same day, gage combination	Probability	—	—	0.0455	0.0191	0.0174	0.0072
	Power	—	—	0.522	0.649	0.665	0.803

is more powerful than the test for transformed data was somewhat surprising. Two possible explanations are listed below:

1) When applying the "t" test to skewed data, there is no need to transform the data. Indeed, the power of the test may be decreased.

2) The "t" test is completely invalid to use for nontransformed data when the data are highly skewed.

There seems to be no universal agreement concerning the application of the "t" test to skewed data. Mood and Graybill (1963) indicate that the size of confidence intervals may deviate considerably from the desired percentage. It was noted in this paper that the means and variances may not be independent when the data are skewed and that the variances of the treated and nontreated data may thus differ. Under these conditions, Natrella (1963) states that confidence intervals based on the pooled estimate of variances may be seriously distorted. However, Natrella further states that if the two samples involved are of equal size, or approximately equal size, then the type I error of the two sided "t" test will not be seriously increased when inequality of variances exists. The results of this research indicate that the second explanation is the most logical and reasonable conclusion that can be drawn.

It should be noted, in all references to the number of days required to obtain significance, that the days are the number of rain days per treatment; also, that it took one year to acquire 15 observations in which a portion of these are zero rain days. Therefore, in discussing the number of years required to obtain significance these factors need to be taken into account.

Although an actual cloud seeding experiment would be conducted for longer than 15 days per year, the following inferences can be made. Under the assumption of dependence of means and variances, there is no gain in the power of the test if the data are transformed. However, the likelihood ratio test developed in this research, which tests the location parameters of two gamma distributions with common shape factor, does increase the power of the test. Since the "t" test is only an approximation when applied to gamma-distributed data and because the likelihood test is more powerful, it would seem that the likelihood ratio test is the appropriate test to use.

For the experimental design used in this research, rainfall was summed over an area so that a total or mean value was obtained. Under this condition, increasing the duration of the experiment rather than increasing the number of gages had the greatest effect on the test of significance.

6. Discussion of the results

In applying the method of generation to a given area, several factors should be considered. First, the method described applies only to daily observations of rainfall. Further work needs to be done to see if other types of rainfall observations could be used. Also, research con-concerning other methods of generation should be undertaken.

It should be noted that the example was only applied to the first 15 days of June. In actual practice a cloud seeding experiment would be applied to more than 15 days a year. Therefore, generation functions should be fitted to other periods of record. Once the data were generated for all periods one could proceed with further analysis using the combined data from all periods. This would yield more experimental units per year; hence, the time involved would not be as long as indicated by the data used in this research. It is difficult to state what the reduction in time would be since in combining the data there would be an increase in the error variance. Also, other periods of the summer might have fewer days with rain. This would have a tendency to increase the time involved in the experiment if zeros are excluded from the data.

It would be of interest to actually generate the seeded samples. This would enable one to test for differences in means without making assumptions concerning the relationships between means and variances. Also, the generation of seeded samples would allow one to study the effect of increasing rainfall for selected portions of the distribution functions. This allows for great flexibility in studying some of the more subtle changes in rainfall due to seeding.

The model as developed so far was for testing a single area in which half of the days are seeded and half are nonseeded. Modification of this model and testing procedure to allow for more general randomization and testing between seeded and nonseeded areas should be developed. In this case, the seeded and nonseeded samples may not be independent.

The problem of the presence of a large number of zeros in the data needs to be studied in detail. This has always been a troublesome factor in cloud seeding studies.

Finally, the model needs to be developed and tested for other areas, other seasons of the year, and other choices of experimental units. The testing of some of these features will soon be started at the Illinois State Water Survey.

Acknowledgments. The authors would like to express their appreciation to Dr. Gary F. Krause, Department of Statistics, and to Dr. James D. McQuigg, Department of Atmospheric Science, University of Missouri, for their helpful suggestions and comments. This research was supported by the National Science Foundation under Grants GP 762 and GA 557.

REFERENCES

Aitchison, J., and J. A. C. Brown, 1957: *The Log-Normal Distribution*. Cambridge University Press, 176 pp.

Battan, Louis J., 1966: Silver-iodide seeding and rainfall from convective clouds. *J. Appl. Meteor.*, **5**, 669–683.

Chow, V. T., 1964: *Handbook of Applied Hydrology*, First edition. New York, McGraw-Hill, 1453 pp.

Decker, Wayne L., and Paul T. Schickedanz, 1966: Project Whitetop: A summary of the rainfall analysis. Univ. of Missouri Rept. under NSF Grant GP-762, 362 pp.

Fix, Evelyn, 1954: Tables of the noncentral χ^2. *University of California Publ. Statistics*, Vol. 1, Berkeley, 390 pp.

Mood, Alexander, M., and Franklin A. Graybill, 1963: *Introduction to the Theory of Statistics*, Second edition. New York, McGraw-Hill, 443 pp.

National Academy of Sciences, 1966: Weather and climate modification, problems and prospects. Vols. I and II, NAS-NRC Publ. 1350, Washington, D.C.

Natrella, Mary Gibbons, 1963: *Experimental Statistics*. Natl. Bur. Stnds. Handbook 91, Washington, D. C., 499 pp.

Neyman, Jerzy, and Elizabeth L. Scott, 1960: Correction for bias introduced by a transformation of variables. *Ann. Math. Statist.*, **31**, 643–655.

Ostle, Bernard, 1954: *Statistics in Research*, First edition. Ames, The Iowa State College Press, 487 pp.

Schickedanz, Paul T., 1967: A Monte Carlo method for estimating the error variance and power of the test for a proposed cloud seeding experiment. Ph.D. thesis, University of Missouri.

Steele, Robert G. D., and James H. Torrie, 1960: *Principles and Procedures of Statistics*, First Edition. New York, McGraw-Hill, 479 pp.

Thom, H. C. S., 1958: A note on the gamma distribution. *Mon. Wea. Rev.*, **86**, 117–122.

——, 1968: Direct and inverse tables of the gamma distribution. ESSA Tech. Rept., Environmental Data Service, EDS-2 30 pp.