On the Distribution of a Ratio of Interest in Single-Area Cloud Seeding Experiments

J. NEUMANN AND E. SHIMBURSKY

Dept. of Meteorology, The Hebrew University of Jerusalem, Israel

(Manuscript received 24 June 1971, in revised form 3 November 1971)

ABSTRACT

In some randomized cloud seeding experiments a single-area design was used. The results of seeding were estimated from the ratio S/U (S and U being the average rainfall amounts, respectively, for the time units seeded and unseeded in accordance with some random scheme). It is pointed out in this paper that the ratio S/U can be looked upon as the "result" of a two-area crossover experiment where the "rain" amounts of the "second" area are unity for all time units. This approach makes it possible to use results developed by Gabriel and Feder for the root-double-ratio pertinent to two-area crossover projects. It then follows that while the effect of seeding is estimated as $(S/U)-1$, the statistical significance level of that result is estimated from a curve representing $(S/U)^{\frac{1}{2}}$. A consequence of this is that the significance level is very much lower than is the case for crossover designs for any *given* value of indicated increase (or decrease). The theoretical results are examined in light of data obtained in the 1961–67 Israeli randomized cloud seeding experiment.

1. Introduction

In randomized cloud seeding experiments two statistical designs have figured prominently, i.e., the crossover and the single-area schemes.

The crossover scheme was introduced in Australia (Adderley and Twomey, 1958) and used in the 1961–67 Israeli experiment. This design estimates the effect of seeding by the so-called root-double-ratio (RDR). With a view to studying the distribution of that ratio in randomized trials, Gabriel (see Gabriel and Feder, 1969) has carried out Monte Carlo randomization experiments[1] on the rain data of the two areas involved in the Israeli project; this has enabled him to estimate the statistical significance of the RDR obtained in the Israeli cloud seeding trial. Furthermore, Gabriel and Feder (1969) have shown that the RDR is, under certain conditions, normally distributed in an asymptotic sense with respect to the number of time units (rain days) included in the experiment.[2]

The single-area scheme was used in Project Whitetop (Braham and Flueck, 1970), among others. For the test statistic one can consider the ratio

$$\frac{S-U}{U}=\frac{S}{U}-1,$$

where S is the average rain amount for time units allocated to seeding, U the average rain amount for time units not-to-be seeded, the allocation following some randomized procedure. As far as we know the asymptotic distribution for increasing the number of time units of the ratio S/U (and it is enough to consider that quantity) has not been published.

In this paper we will show that the asymptotic distribution of S/U can be obtained from the asymptotic distribution of the RDR as developed by Gabriel and Feder. In principle, then, we are in a position to judge the statistical significance of any given S/U value obtained in an actual experiment, provided that the number of time units involved in the experiment is sufficiently large. Further, we will consider the results obtained in the 1961–67 Israeli experiment for one of the two areas figuring in the crossover scheme and compare the predictions of the RDR with those of the single-area ratio.

2. Median and mean of S/U

Let $S/U=z$, z assuming non-negative values only. We will now show that z is a random variable. To this end, we introduce a series of random binomial variables, a_i, $i=1, 2, \cdots, N$ (where N is the number of time units in the experiment; here, the number of rain days),

[1] In the Monte Carlo "experiments" one permutes, i.e., shuffles, in a random manner and for a large number of times the rain data for a given period for the area or areas concerned. For example, in the case of a crossover design, one permutes the data of each of the two areas and gets a distribution of the RDR in these randomized computer experiments. One can then judge the statistical significance of the RDR obtained in an *actual* cloud-seeding trial by referring it to the results of the Monte Carlo experiments.

[2] A reviewer has pointed out that the Gabriel-Feder paper neglects to consider cases where the denominator of the RDR is zero. In the present version of our paper we propose to make good this omission.

such that

$$P(a_i)=\begin{cases}\frac{1}{2}, & \text{for } a_i=0\\ \frac{1}{2}, & \text{for } a_i=1\end{cases} \quad (1)$$

If the x_i are the daily rain amounts, then we can write z *formally* with the aid of the binomial variables in the following form $(\sum_i \equiv \sum_{i=1}^{N})$:

$$z=z(a_1,\cdots,a_N)=(\sum_i a_ix_i/\sum_i a_i)/[\sum_i (1-a_i)x_i/\sum_i (1-a_i)].$$

The difficulties in applying the above ratio without further qualifications are twofold:

1) Cases may occur where all the randomly allocated binomial variables are equal. In such cases either the ratio forming the numerator or the one forming the denominator on the right side of the above expression takes the indeterminate form 0/0. The probability for all the N binomial variables to turn out equal in a random experiment is $2[1/(2^N)]=1/(2^{N-1})$ which is small but finite if N is large but finite.

2) Cases may occur in which the sum $\sum_i (1-a_i)x_i$ in the denominator of the above expression is zero. This can happen if in $\sum_i (1-a_i)x_i$ all the nonzero rain amounts turn out, by chance, in association with a_i values all equal to unity. If n is the number of dry days in the series of N days, then the probability of the event $\sum_i (1-a_i)x_i=0$ is $2^n/2^N$. If $n=\epsilon N$, $0<\epsilon<1$, then the aforementioned probability is $1/[2^{(1-\epsilon)N}]$; if ϵ is fixed and N is large but finite, this probability is small but finite. The assumption that ϵ is a fixed quantity is a reasonable assumption for the purposes of an approximation.

Case 1) is mentioned in the Gabriel-Feder paper (p. 151), with the authors arbitrarily assigning to z the value 1 in such a case. In case 2), we propose to arbitrarily assign to z the value N. The consequences of such an arbitrary value are unimportant if the fraction of dry days in the series of N days is fixed and N is large.

Since we have already referred to the series of N days as a series of N "rain days," it will be in order to point out that in the 1961–67 Israeli experiment a day was classified as a rain day in the experiment area if a measurable amount of rain occurred on that day in the so-called "buffer zone" separating the two experiment areas (Gabriel, 1966; Neumann *et al.*, 1967) of the crossover design. The buffer zone, where rainfall is well correlated with that of the experiment areas, was not supposed to be seeded at any time. Hence, the classification of a day as a rain day in the experiment areas depended neither on the conditions there nor on the effect of seeding, if any. This procedure reduced the number of actually dry days in the series of N rain days to about 2% of N. It was pointed out earlier [case 2)] that the probability of getting $\sum_i (1-a_i)x_i=0$ in randomized computer experiments is $1/[2^{(1-\epsilon)N}]$, which is extremely small if N is large and ϵ is small and independent of N (ϵ is 0.02 for the Israeli experiment). We therefore believe that assigning the value N to the single-area ratio z when $\sum_i (1-a_i)x_i=0$ will have little influence on the evaluation of the statistical significance of single-area cloud seeding results by permutation trials for that purpose.

We thus define z as follows:

$$z=z(a_1,\cdots,a_N)=\begin{cases}1, & \text{if } a_1=\cdots=a_N \quad \text{(2a)}\\ N, & \text{if } not \text{ all the } a_i\text{'s are equal and if } \sum_i (1-a_i)x_i=0 \quad \text{(2b)}\\ 1/N, & \text{if } not \text{ all the } a_i\text{'s are equal and if } \sum_i a_ix_i=0 \quad \text{(2c)}\\ (\sum_i a_ix_i/\sum_i a_i)/[\sum_i (1-a_i)x_i/\sum_i (1-a_i)], & \text{otherwise.} \quad \text{(2d)}\end{cases}$$

We have added, arbitrarily, (2c). This was done in order to preserve a symmetry property of z represented by (4) below. The symmetry property, $P(\log z)=P(-\log z)$ [see (4)], would have been thwarted by the addition of (2b) alone; (2c) counteracts the effect of (2b). The probability of the event $\sum_i a_ix_i$ is the same as that of $\sum_i (1-a_i)x_i$, that is, $1/[2^{(1-\epsilon)N}]$.

In (2) the *random* binomial variables are weighted by the rain amounts. Hence z itself is a random variable.

In the space of random vectors $(a_1, \cdots, a_N)$, each permutation of these vectors has the same probability, $1/2^N$. There is a one-to-one correspondence between $(a_1, \cdots, a_N)$ and $(1-a_1, \cdots, 1-a_N)$. Further, we note that

$$z(1-a_1, \cdots, 1-a_N)=\frac{1}{z(a_1,\cdots,a_N)}.$$

Consequently,

$$P[z(a_1,\cdots,a_N)]=P\left[\frac{1}{z(a_1,\cdots,a_N)}\right]. \quad (3)$$

We conclude from (3) that the median of the distribution of z is 1. It also follows that

$$P(\log z)=P(-\log z), \quad (4)$$

or, $\log z$ is a random variable symmetrical about 0, i.e., its expected value is 0.

TABLE 1. Means, variances and χ^2 goodness-of-fit tests of randomized permutation experiments for the North experiment area of the 1961–67 Israeli Randomized Cloud Seeding Project.*

	$N=381$				$N=517$
	Number of permutations in Monte Carlo experiments				
	300	500	750	1000	1000
1. Sample mean of $(S/U)^{\frac{1}{2}}=z^{\frac{1}{2}}$	0.99150	1.01340	1.03090	1.03500	0.97070
2. Theoretical mean of $(S/U)^{\frac{1}{2}}$ corrected for finiteness of sample [Eq. (9)]	1.00250	1.00250	1.00250	1.00250	1.00140
3. Sample variance of $(S/U)^{\frac{1}{2}}$	0.00562	0.00543	0.00539	0.00529	0.00274
4. Asymptotic variance [Eq. (8a)]	0.00502	0.00502	0.00502	0.00502	0.00283
5. Asymptotic variance corrected for finiteness of sample [Eq. (10)]	0.00504	0.00504	0.00504	0.00504	0.00282
6. χ^2 test of goodness-of-fit to Eq. (8) (18 degrees of freedom)	37.6	20.8	45.1	18.7	13.7

* $N=381$ rain days, 1961–67; $N=517$ rain days, 1949–59.

The foregoing results enable us to draw an inference concerning the expected value of z itself. Let $\xi=\log z$. As e^{ξ} is a convex function, the following theorem [published in Loève (1955, p. 159)] applies to it:

$$e^{E(\xi)} \leqslant E(e^{\xi}), \tag{5}$$

or

$$e^{E(\log z)}=e^0=1\leqslant E(e^{\log z})=E(z). \tag{6}$$

So far we have found that z is a random variable whose median is 1 and whose mean value is greater than or equal to 1. This indicates that the distribution of z is not necessarily symmetrical about 1.

3. The single-area ratio as a double ratio

One can re-arrange (2d) to read

$$z_N=[\sum_i a_i x_i/\sum_i (1-a_i)x_i][\sum_i (1-a_i)/\sum_i a_i], \tag{7}$$

where the subscript N emphasizes that we are still dealing with a sample of N days, $z_N=z(a_1, \cdots, a_N)$. Written in the form of (7) z_N can be interpreted as the double ratio treated in the paper by Gabriel and Feder with the difference that in (7) the rain amounts of the second area are all unity. Hence, $z_N^{\frac{1}{2}}$ is a special case of the Adderley–Twomey (1958) RDR, and all that is said in the Gabriel–Feder paper about the RDR applies to our $z_N^{\frac{1}{2}}$ as well. In particular, it follows, under the conditions for which the Gabriel–Feder results are valid, that

$$\frac{z_N^{\frac{1}{2}}-1}{V_N^{\frac{1}{2}}} \xrightarrow[N\to\infty]{} \text{Normal } (0,1), \tag{8}$$

where

$$V_N=\sum_i \left(\lambda_i-\frac{1}{N}\right)^2, \quad \lambda_i=\frac{x_i}{X}, \quad X=\sum_{j=1}^{N} x_j, \tag{8a}$$

V_N being the variance and where, for sufficiently large N, $z_N^{\frac{1}{2}}$ will be normal with a mean of 1 and a variance represented by (8a). Since the Gabriel-Feder paper does not consider the cases where $\sum_i (1-a_i)x_i=0$, we will treat in the Appendix the asymptotic normality of the distribution of $z_N^{\frac{1}{2}}$ covering all the four cases (2a)–(2d).

4. Comparison with the results of Monte Carlo experiments

In order to check the predictions resulting from (8) against observations, we have examined daily rainfall data for the experiment area designated North in the 1961–67 Israeli randomized cloud seeding experiment. This project is described and discussed in a number of papers and reports as, for example, in Gabriel (1966), Neumann *et al.* (1967) and Gabriel (1970). The observational data comprise daily averages of rainfall at 26 stations in the North for 381 rain days.

Since the predictions of the foregoing sections are of an asymptotic nature and require large values of N, we have also considered the daily rain data for the same 26 stations for the years 1949–59, involving 517 rain days defined as in Section 2. For both of the above periods we have carried out Monte Carlo experiments on the Hebrew University's CDC 6400 computer. These experiments are very much like the earlier Monte Carlo computer experiments conducted by Gabriel (Gabriel and Feder, 1969). Gabriel's randomization experiments relate to both areas of the 1961–67 crossover project; in this paper we confine our main attention to the data of the North area as if the 1961–67 Israeli project were a single-area trial conducted in the North area.

Our Monte Carlo experiments involve 200, 300, 500, 750 and 1000 permutations of the ratio $(S/U)=z$ for the two periods described earlier in this Section. Results of the permutations are set out in Table 1, but before considering the table, we should point out the following:

1) The figures in line 2 are corrections to the theoretical mean $(=1)$, taking into account the finiteness of the sample. Denoting the expected value so corrected

by the subscript 1, the corrected value reads as follows:

$$E_1(z_N^{\frac{1}{2}})=1+\tfrac{1}{2}V_N, \qquad (9)$$

V_N being the variance occurring in (8).

2) The figures in line 4 are based on a theoretical prediction and are not a function of the number of permutations.

3) Like line 2, the figures in line 5 are corrections to the asymptotic variance V, taking into account the finiteness of the sample. The corrected variance V_1 is

$$V_1(z_N^{\frac{1}{2}})=V_N E_1(z_N^{\frac{1}{2}}). \qquad (10)$$

4) The figures in line 6 give the result of a χ^2 test of goodness-of-fit to the normal distribution (8) of the distributions obtained in the permutation experiment. Here

$$\chi^2_{(f-2)}=\sum_{i=1}^{f}\frac{(n_i-np_i)^2}{np_i},$$

where f is the number of class intervals adopted (=20), n_i the number of $(S/U)_N^{\frac{1}{2}}=z_N^{\frac{1}{2}}$ ratios falling into interval i, p_i the theoretical probability density, and n the number of permutations (n=300, 500, 750 and 1,000). We have ignored the result for n=200.

We note in Table 1 the improvement in the measure of agreement between the "observed" values and the theoretical predictions as we pass from N=381 to N=517. Generally speaking, the variances agree better than the means. The observed means indicate a certain amount of bias, although this bias virtually vanishes for the case N=517.

The line for χ^2, N=381, brings out the important point that it is not enough to increase the number of permutations at a given N in order to reach good agreement with the theoretically predicted normal distribution. As theory indicates, a large N is required.

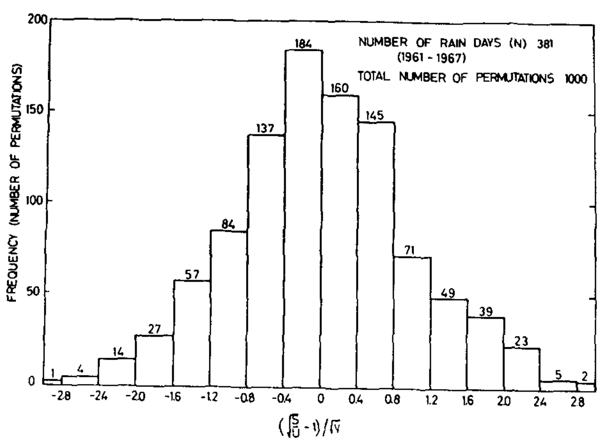

FIG. 1a. Histogram of $[(S/U)^{\frac{1}{2}}-1]/V_1^{\frac{1}{2}}$ in 1000 random permutations of the daily rain data for the North experiment area of the 1961–67 Israeli Randomized Cloud Seeding Project. Daily data are averages for 26 stations. Number of "rain days" N=381. For meaning of V_1 refer to Eq. (10).

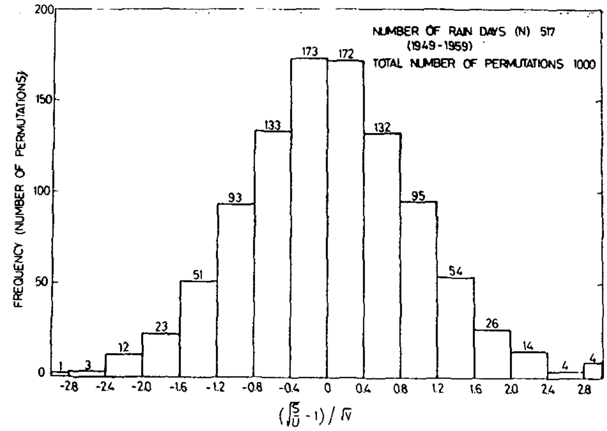

FIG. 1b. Same as Fig. 1a except for the period 1949–59. Number of rain days N=517.

It would appear in the case of the North area, treated as a single-area design, that a value of N something like 500 is large.

Figs. 1a and 1b show the histograms for the case of 1000 permutations. The enhanced agreement with a normal distribution as we pass from N=381 to N=517 days is obvious.

5. Estimating the level of significance of single-area ratios

One can estimate the level of significance of any given single-area ratio $(S/U)_0=z_0$, say, such as would be obtained in a single-area randomized cloud seeding experiment, by considering the area under the curve $z^{\frac{1}{2}}$ for $z\geqslant z_0$.

In Table 2 we list the levels of significance, i.e., areas, for a series of z values relating to the North experiment area based on the permutaton experiments for the actual cloud seeding experiment of 1961–67 as well for the years 1949–59 preceding the experiment. In both cases the columns marked "observed level of significance" give the number of permutations greater than or equal to the specified value of z, the number of such

TABLE 2. Levels of significance (percent) of hypothetical single-area ratios $(S/U)=z$ based on an asymptotic normal distribution [Eq. (8)] and on the results of 1000 randomized permutations for the North experiment area of the 1961–67 Israeli Randomized Cloud Seeding Project.*

		N=381		N=517	
S/U	$(S/U)^{\frac{1}{2}}$	Theoretical level of significance [Eq. (8)]	Observed level of significance in 1000 permutations	Theoretical level of significance [Eq. (8)]	Observed level of significance in 1000 permutations
1.05	1.024	37.8	39.4	31.9	32.5
1.10	1.048	18.1	18.1	17.6	17.5
1.15	1.072	16.4	16.6	8.3	8.2
1.20	1.095	9.7	10.1	3.4	3.7
1.25	1.118	5.3	7.4	1.2	1.3
1.30	1.140	2.7	4.3	0.4	0.5

* N=381 rain days, 1961–67; N=517 rain days, 1949–59.

permutations being expressed as percentages of the total number of permutations. The columns called "theoretical level of significance" are based on the normal distribution (8).

We note in Table 2 the fair-to-good measure of agreement between the observed and the theoretical values, the agreement being rather good for $N=517$. However, we are more concerned with some of the negative aspects of Table 2.

According to Gabriel (1970, Table 4), the RDR for the 1961–67 crossover experiment is 1.15 (i.e., a 15% increase of rain is indicated for days seeded), and the statistical significance level of the foregoing result is 0.9% based on the randomized permutation trials. If we now take the single-area ratio S/U for the North area, the indicated increase again is 15%, but, according to Table 2, the significance level of that result is about 16%—a very poor showing in comparison with the crossover experiment. Of course, it was the inefficiency of the single-area design that prompted Adderley and Twomey (1958) to introduce the crossover scheme, but we believe that the single-area design and its poor efficiency have not received adequate attention in the literature.

If one wishes to test a one-tailed hypothesis concerning positive effects of seeding using a single-area design, one must take note of the fact that while the effect of seeding is estimated as $(z-1)$, the significance level to be assessed from the $z^{\frac{1}{2}}$ curve assuming a large N. Thus, if in a cloud seeding experiment the single-area ratio z is 1.15 (and if N is large), the significance level is obtained as the area under the $z^{\frac{1}{2}}$ curve beyond $z=1.15$. Of course, this area is much larger than the appropriate area for a true crossover experiment, resulting in a RDR of 1.15. Similar considerations apply to the case where negative effects are examined by a one-tailed test.

Actually, the above direct comparison between single-area and crossover results is possible only if the variance of $z^{\frac{1}{2}}$ for the single-area experiment and the variance of the RDR for the crossover experiment are equal or very nearly so. This condition is met by the 1961–67 Israeli experiment. It is seen from our Table 1 that the sample variance of $z_N^{\frac{1}{2}}$, $N=381$ days, is 0.00529. On the other hand, a reference to Table 2 in the Gabriel–Feder paper shows that the sample variance of the RDR (see line 1 of their Table 2) is 0.0052. The latter is based on the years 1961–66 covering a total of $N=327$ days.

APPENDIX

The Asymptotic Normality of the Root of the Single-Area Ratio

According to Loève (1955, p. 168, Exercise 10c), if $\xi_N=\xi(a_1, \cdots, a_N)$ is an asymptotically normal random variable with mean 0 and variance 1 for $N\rightarrow\infty$, then the random variable $(z_N^{\frac{1}{2}}-1)/V_N^{\frac{1}{2}}$, $z_N=z(a_1, \cdots, a_N)$, is similarly normal provided that for any $\delta>0$ the probability

$$P\left\{\left|\frac{z_N^{\frac{1}{2}}-1}{V_N^{\frac{1}{2}}}-\xi_N\right|>\delta\right\}\rightarrow 0 \quad \text{as} \quad N\rightarrow\infty. \tag{11}$$

We now identify ξ_N with the right side of (12) of the Gabriel–Feder paper (where they proved its asymptotic normality) and apply (11) to the distribution represented by (2).

In considering (11) in relation to (2), we have the following four exclusive events:

Event (2a). $a=\{z_N=1\}$

Event (2b). $b=\{z_N=N\}$

Event (2c). $c=\left\{z_N=\frac{1}{N}\right\}$

Event (2d). $d=\{z_N=(\sum_i a_ix_i/\sum_i a_i)/[\sum_i (1-a_i)x_i/\sum_i (1-a_i)]\}$

Let us denote by H the collection of all permutations of $a_1, \cdots, a_N$ such that

$$\left|\frac{z_N^{\frac{1}{2}}-1}{V_N^{\frac{1}{2}}}-\xi_N\right|>\delta.$$

Then, since the events (2a), (2b), (2c) and (2d) are mutually exclusive, their probabilities add, or

$$P(H)=P(H\cap a)+P(H\cap b)+P(H\cap c)+P(H\cap d),$$

where the symbol $\cap$ represents the "intersection" of two groups. Now, according to the theorem on compound probabilities [Feller, 1950, p. 81, Eq. (1.9)],

$$\begin{aligned}P(H)=&P(H|a)P(a)+P(H|b)P(b)\\&+P(H|c)P(c)+P(H|d)P(d)\\&\leqslant(P(a)+P(b)+P(c)+P(d).\end{aligned} \tag{12}$$

But, in accordance with what was said in Section 2, the probability $P(a)=1/(2^{N-1})$, $P(b)=P(c)=1/[2^{(1-\epsilon)N}]$. All three of the aforementioned probabilities $\rightarrow 0$ as $N\rightarrow\infty$ $(0<\epsilon<1,\ \epsilon$ fixed). As to event (2d), we use Gabriel and Feder's Taylor expansion to conclude that

$$|[(z_N^{\frac{1}{2}}-1)/V_N^{\frac{1}{2}}]-\xi_N|$$

converges in probability to 0 as $N\rightarrow\infty$. Here we use the Taylor expansion in the domain where $\sum_i a_ix_i$ and $\sum_i (1-a_i)x_i$ (corresponding, respectively, with T and $1-S$ in the Gabriel–Feder paper) are not equal to 0. We thus conclude that the sum of probabilities in (12) approach 0 as $N\rightarrow\infty$, which means that (11) is satisfied or (2) is asymptotically $N(0,1)$.

Acknowledgments. The writers are indebted to Prof. K. R. Gabriel for his comments on this paper.

REFERENCES

Adderley, E. E., and S. Twomey, 1958: An experiment on artificial stimulation of rainfall in the Snowy Mountains of Australia. *Tellus*, **10**, 275–280.

Braham, R. R., and J. A. Flueck, 1970: Some results of the Whitetop experiment. *Preprints of Papers, Second Natl. Conf. Weather Modification*, Santa Barbara, Calif., Amer. Meteor. Soc., 176–179.

Feller, W., 1950: *An Introduction to Probability Theory and Its Applications*, Vol. 1. New York, Wiley, 419 pp.

Gabriel, K. R., 1966: The Israeli artificial rainfall stimulation experiment. Statistical evaluation for the period 1961–1965. *Proceedings Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 5, University of California Press, 91–114.

——, 1970: The Israeli rainmaking experiment 1961–67. Final statistical tables and evaluation (Tables prepared by M. Baras). Tech. Rept., Hebrew University, Jerusalem, 47 pp.

——, and P. Feder, 1969: On the distribution of statistics suitable for evaluating rainfall stimulation experiments. *Technometrics*, **11**, 149–160.

Loève, M., 1955: *Probability Theory*. New York, Van Nostrand, 517 pp.

Neumann, J., K. R. Gabriel and A. Gagin, 1967: Cloud seeding and cloud physics in Israel: Results and problems. *Proc. Intern. Conf. Water for Peace*, Vol. 2, Washington, D. C., 375–388.