The Florida State University

Department of Statistics

Tallahassee, Florida

⑨ Technical rept.,

⑥

RANDOMIZATION TESTS IN SUPPORT OF SOME STATISTICAL ANALYSES OF A WEATHER MODIFICATION EXPERIMENT

by

⑩ Ralph A. Bradley and Elton Scott

⑭ FSU Statistics Report No. M521,
ONR Technical Report No. 149

TR-149-ONR

⑫ 19

⑪

October, 1979
The Florida State University
Department of Statistics
Tallahassee, Florida 32306

D D C
RECEIVED
NOV 8 1979
A

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

⑮

Research sponsored by the U.S. Office of Naval Research under Contract No. N00014-76-C-0394. Reproduction in whole or in part is permitted for any purpose of the United States Government.

400 277

RANDOMIZATION TESTS IN SUPPORT OF SOME STATISTICAL ANALYSES OF A WEATHER MODIFICATION EXPERIMENT

Ralph A. Bradley and Elton Scott

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

1. INTRODUCTION

A new awareness and interest in statistical problems encountered in the design and analysis of weather modification experiments have developed recently, perhaps stimulated by Brillinger, Jones and Tukey, hereafter BJT, (1978). Sessions on statistics in weather modification have been included in recent meetings of statistical societies, a Tallahassee Workshop on Weather Modification sponsored jointly by the Office of Naval Research and the Florida State University was held in November, 1978, and a two-number, special issue of *Communications in Statistics, Theory and Methods* was devoted to the statistical analysis of weather modification experiments in 1979. Braham (1979) led discussion of field experimentation in weather modification with comments added by a number of workers in the field.

BJT expressed grave concerns about the validity of parametric analyses of weather modification experiments. Gabriel (1979, p. 984, 985) provides two quotations from their report and we quote from his paper to show that he has even more concerns:

> "...Statisticians have indeed been brave to risk using simple parametric models and methods of inference in the analysis of weather modification experiments. These parametric methods are quite unlikely to have been valid when the underlying assumptions were so patently untrue."
>
> "...However, other workers have continued to use parametric methods, evidently preferring to ignore their dubious validity. One may hope that the clear statement by the Task Force (BJT, 1978) might put an end to this."

At the noted Tallahassee Workshop, exploratory parametric analyses of data from Phase I of the Santa Barbara Convective Seeding Test Program were presented, Bradley, Srivastava and Lanzdorf, BSL, (1979). (This test program in two phases was designated Santa Barbara II by BSL.) Several discussants criticized the use of parametric methods with varying degrees of fervor.

The purpose of this article is to respond to criticisms of parametric analyses, particularly if used in approximation to randomization analyses. Our objective is to put the issues in perspective and it is not in any way to obviate the responsibilities of those who devise stochastic models and statistical analyses to examine the validity of assumptions inherent in their methods.

We conclude the article with some comments on multiplicity of analyses, always pertinent in reanalyses of experiments, as discussed by BJT (1978, Section 33), Gabriel (1979, Section 4), and Kruskal (1979, Section 3).

2. THE SANTA BARBARA EXPERIMENT

BSL (1979) give a summary, with references, of Phase I of the Santa Barbara experiment. Winter storms in the region have convective cells grouped into identifiable convective bands, several of which may occur in a storm. The convective band was used as the experimental unit. Criteria for the "seedability" of a convective band proved to be only partially operationally usable. A predetermined randomized decision to seed or not seed an experimental unit was made. A target area and a control area, the latter West of the target area and seeding site, were defined. Precipitation responses attributable to each experimental unit in both areas were measured through use of networks of raingages, not all of which were operative for every experimental unit. Meteorological covariates, some of them measured by radiosonde, were observed for each convective band; unfortunately, radiosonde observations were taken at Santa Barbara Airport in the target area.

The original statistical analyses were based on application of the Wilcoxon-Mann-Whitney, two-sample, rank test separately for the data from each raingage. Both precipitation measurements and ratios of precipitations at the raingage divided by average corresponding control area precipitations were used with similar results. These analyses were reported by Elliott and Thompson (1972) and Brown, Thompson and Elliott (1975). Covariates were used only in a minor way to consider classified subsets, or separations into subsets, of the experimental data. Concerned with both the validity of the assumption of independence of successive observations for the nonparametric tests and dependencies among the tests, Elliott and Brown (1971) conducted a limited Monte Carlo study using their data with randomized relabelling of the seeding indicator. They report:

> "At the 0.05 significance level for all bands, 29 stations (raingages) in the original test sample were found to show a positive difference between seeded and not-seeded cases; and three Monte Carlo runs (out of 50) were found to have as high or higher counts of stations with a positive difference at this significance level."

Research on statistics in weather modification at the Florida State University had two objectives. Independent reanalysis of the Santa Barbara data was desired because of the economic importance to both agriculture and energy generation of enhanced precipitation from West Coast winter storms. Improved statistical methodology or new use of existing procedures could improve similar future experimentation.

Summarization of data from a raingage network was considered first. BSL (1977) used response surfaces, but found that their "integrated rainfall" was so highly correlated with the simple average of rainfall responses over raingages in a target or control area that the average was the appropriate summary measure. Scott (1979) used principal components for similar summarization with limited success and difficulties in application. His first component was highly correlated with the average and, although two additional components could be identified and interpreted, they contributed little in explanation of variability among raingages in a target or control area. BSL (1979) examined the usefulness of meteorological covariates, such as air-mass characteristics, control area precipitations and band passage times, or subsets of them, along with other possible sources of variation, for example, storm effects. They used parametric analyses after data transformation; an examination of plots of standard deviations of raingage readings against corresponding means for each experimental unit or band showed a linear relationship and suggested the transformation, $z = \log(y + \Delta)$, where y is the precipitation at a raingage for a band and Δ is a constant for a defined target area. The use of covariates essentially eliminated any apparent effect due to seeding and covariate measures were suspected to have been affected by seeding. Simple analyses of the effects of seeding for the defined target area and subsets of it were consistent with the suggested significance reported in the quotation above from Elliott and Brown (1971). Serfling (1977) used the covariate data to group responses for bands and applied nonparametric methods to study the effect of seeding for bands within groups with inconclusive results.

BSL (1979) provided analyses of variance and covariance for various target areas based on weighted analyses of data transformed as indicated above. Target areas were defined in their Table I. The model used, (4.1) in the reference, had

$$z = \beta_0 + \sum_{i=1}^{p} \beta_i V_i + \delta Z + \varepsilon,$$

where z is the mean of the transformed precipitation responses for raingages in a designated target area, β_0 is a general constant, V_i is the observation on the i^{th} covariate, with β_i being the corresponding regression coefficient, $i = 1, \ldots, p$, the number of covariates used (all may be omitted), $Z = 1$ or 0 as the experimental unit (band) was or was not seeded, δ is the parameter measuring the additive effect of seeding, and ε is a random error. Weighted least squares analyses minimized

$$\sum_{\alpha=1}^{N} w_\alpha (z_\alpha - \beta_0 - \sum_{i=1}^{b} \beta_i V_{i\alpha} - \delta Z_\alpha)^2,$$

where w_α was the number of raingage observations contributing to z_α in band α and N was the total number of bands, some seeded, some not. Tables I and II below are taken from

Tables V and A-II of the reference. They are analyses for Target Area (i), the main planned target area; other target areas involved subsets of data used for Target Area (i) and gave very similar results. No covariates were used in Table I and the covariates for Table II were those of Model (6) in (4.5) of the reference. In the next section, we report on randomization analyses to check the validities of the analyses in Tables I and II. Note that, in the reference, the authors avoided the probably untenable parametric assumptions of independence and normality of the random errors ε in the model. In addition, the weights w_α were

TABLE I

Analysis of Variance without Covariates for Target Area (i), Transformed Data

Source of Variation	d.f.	Mean Squares	F-Ratio
Seeding	1	110.29	2.77
Residual	104	39.84	--

TABLE II

Analysis of Covariance, Reference Model (6), for Target Area (i), Transformed Data

Source of Variation	d.f.	Mean Squares	F-Ratio
Seeding (adj.)	1	0.00	0.00
Covariates	7	374.40	22.24
Residual	97	16.83	--

used as approximately proportional to the variances of the ε_α, although no one would assume that precipitations at nearby raingages are likely to be independent. The question is whether or not the parametric analyses are adequate approximations to randomization analyses for practical purposes. The question is not whether or not parametric assumptions are correct.

3. THE RANDOMIZATION TESTS

The randomization distributions of F were sampled by Monte Carlo methods separately for the data of Tables I and II. The number of Monte Carlo runs for each table was 500. The sampling mimicked the original experimental randomization; a value from a symmetric Bernoulli distribution was generated for each run and thus a value of Z in the model was determined. Accordingly, the numbers of declared seeded and not seeded experimental units were observations from a symmetric binomial distribution with sample size 106, the number of bands in the experiment. For each run, analysis like that of Table I or Table II was done and the resultant value of F was recorded together with the sign of the estimated seeding effect.

Sample frequency distributions of t, the signed square root of F, are given in Table III for the data of Tables I and II. It is clear that these distributions are approximately symmetric about the origin. They approximate Student distributions with 104 and 97 degrees of freedom respectively, although these Student distributions are not distinguishable for practical purposes. The mean, variance, and coefficients of skewness and kurtosis of t for the data from Table I are in order 0.029, 1.107, -0.049 and -0.046 and the corresponding values for the Student distribution with 104 degrees of freedom are 0, 1.020, 0, and 0.060. Similarly, in the same order for Table II, the sample values are -0.040, 1.033, 0.142, and 0.312 whereas, for the Student distribution with 97 degrees of freedom, the values are 0, 1.021, 0, and 0.065.

In assessment of significance of the data of Table I, a one-sided test for seeding effect seems appropriate. The 500 Monte Carlo runs resulted in 60 values of F not exceeded by the observed value of 2.77, 37 of them with apparent positive effect of seeding and 23 of them with apparent negative effect. Since we know that the complete randomization distribution of t must be symmetric about the origin, the proper, estimated, one-sided significance level is ½(60/500) = 0.060; the corresponding 0.95 confidence interval is (0.039, 0.081).

TABLE III

Frequency Distributions of t for 500 Samplings of the Randomization Distributions of the Data of Tables I and II

	TABLE I		TABLE II	
Class Interval	Frequency	Relative Frequency	Frequency	Relative Frequency
Less than-3.0	2	0.4	2	0.4
-3.0 to -2.5	2	0.4	1	0.2
-2.5 to -2.0	13	2.6	9	1.8
-2.0 to -1.5	14	2.8	27	5.4
-1.5 to -1.0	49	9.8	47	9.4
-1.0 to -0.5	78	15.6	68	13.6
-0.5 to 0	91	18.2	109	21.8
0 to 0.5	89	17.8	104	20.8
0.5 to 1.0	69	13.8	60	12.0
1.0 to 1.5	48	9.6	38	7.6
1.5 to 2.0	31	6.2	22	4.4
2.0 to 2.5	11	2.2	6	1.2
2.5 to 3.0	3	0.6	6	1.2
Gr. than 3.0	0	0.0	1	0.2
Totals	500	100.0	500	100.0

The parametric, one-sided significance level from tables of the Student distribution with 104 degrees of freedom is 0.050. It appears that the parametric method has given a satisfactory approximation to the significance level of the appropriate randomization test.

The mean square for seeding after adjustment for the covariates in Table II is remarkably small. The covariates have accounted for a major portion of the experimental variation and the residual mean square has been reasonably reduced. This was discussed in the reference where the authors conjectured that the covariates had been affected by treatment (seeding); recall that covariates were measured in the target area. In the 500 Monte Carlo runs with the data of Table II, 12 values of F (for seeding) were recorded that did not exceed the value of Table II, zero when rounded as recorded. This study seems simply to support the earlier conjecture.

Analyses like those of Tables I and II were very similar for other target areas of the reference. Monte Carlo studies can be expected to yield comparable results.

4. REMARKS

Has fortune favored the brave? (See the first quotation from Gabriel.) We think not. The parametric Student test has been shown to be relatively robust. The authors judged that the parametric analyses would be adequate for practical purposes from considerable personal experience. Kempthorne (1975, p. 323), in discussion of the role of normal-law parametric theory, states:

> "This role is to provide useful approximations to the randomization test procedure..."

Miller, Shaw and Veitch (1979) would prefer randomization tests, but in consideration of a cloud-seeding experiment in Tasmania, both consider data transformation and agreements of parametric and randomization tests as reported in their Table II. They state:

> "Table II shows that, in this experiment, the significance levels obtained from the permutation test agree well with those from the F-test without pooling which suggests that the assumptions of the classical test are adequately satisfied."

(We would have preferred a conclusion in terms of adequacy of approximation of the F-test to the permutation test.) Significances obtained from the data of Table I are in reasonable agreement with the small permutation test sampling of Ellictt and Brown (1971).

How should statisticians proceed with the analysis of weather modification experiments? The cautious statement by BJT has our partial endorsement. It was quoted by Gabriel (1979) and is repeated here:

> "...if either theory or substantial experience shows that the tail areas (degrees of unlikeliness) offered by an approximation to a parametric analysis (we usually have to settle for approximations in such analyses) are repeatedly and constantly close to the results of re-randomization, it may be wise to use the approximate, parametric, tail areas even more widely, reserving re-randomization calculations for only the most crucial comparisons."

We would be braver than BJT. Exploratory analyses of the limited and expensive available data from weather modification experiments can be justified as a search for new insights subject to possible confirmation in future experiments. This is the way that science works. Creation of a need to use only re-randomization analyses is likely to stifle such investigations; while modern computer power may make such analyses possible, they are cumbersome and expensive. We believe that parametric analyses should be satisfactory for data exploration and exploratory use of new means of data summarization and we would reserve re-randomization calculations for confirmation of apparently interesting effects suggested by the parametric analyses. In this way, some of the experience required by BJT may accumulate. We do urge care in the parametric analyses and good data handling, including use of transformations when indicated. We would endorse the caution of BJT in preplanned data analyses of confirmatory experiments.

We are in general agreement with the warnings expressed in the references of Section 1 on the matter of multiplicity of analyses. BJT (1978, Section 33) take a moderate approach:

> "We should strive to do as much inquiry into alternate modes of analysis as we can (a) on older data, or (b) during the exploratory phase of the study that most concerns us. A few alternatives, particularly if highly correlated in their results, may survive into confirmatory phases, but we should do whatever we can to keep the number that survive as close to one as we can."

Gabriel (1979, Section 4) elaborates and Kruskal (1979, Section 3) discusses the situation with examples, suggesting that

> "Examples of the high frequency of low frequency events abound."

Gabriel considers also the complementary problem, the artificial inhibition of significance:

> "But we must equally warn of the pitfall of multiplying the number of re-analyses of a successful experiment (i.e., one which has established an effect) until some re-analysis reveals another correlate of the "effect" which may then be used as an alternative explanation or a rejection of the finding that the treatment was effective."

He used analyses later reported by BSL (1979) to illustrate:

> "An apparently significant precipitation difference was found between seeded and unseeded bands. Subsequently, in the course of a large number of further analyses, the seeded/unseeded difference was found to be largely confounded with differences between storms, with no significant difference remaining between seeded and unseeded bands within storms."

Two major points can be made on the issue of multiplicity of analyses. The first is that the degree of dependency or correlation is crucial. On one extreme, repeated application of the same test to the same data would cause no concern, although it would be silly. On the other extreme, searches for some subset of the experimental data for some result of significance without regard to experimental objectives or design would be ludicrous and some low-frequency event would surely be detected. The second point is that experimental design itself may dictate a sequence of analyses as sources of variation and improved experimental precision, explicitly or implicitly designed into the experiment, are considered for their effectiveness.

Reanalysis of the Santa Barbara experiment was initiated at the Florida State University because there was concern with the methods originally used and in the hope that new insights for improved future similar experimentation might result. The sequence of analyses performed represents, in our view, a considered, logical and independent reanalysis of the Phase I Santa Barbara data. Further, the Phase II data were available and reserved for possible confirmatory analyses, analyses that are in progress, although experimental design changes lead to new difficulties. The sequence of analyses began with investigation of improved methods of raingage data summarization.

Possible covariates were considered and found suspect, seemingly having been affected by treatment. A transformation was developed for improved parametric analyses. Several target areas were considered and yielded highly correlated analyses; this was necessary to consider in a limited way possible areas of effect and particularly to investigate possible effects far downwind as reported in the literature. It was clear that storm effects were partially confounded with treatment. We believe that this further analysis reported by Gabriel led to very low power,' within-storm comparisons of seeded and unseeded experimental units. The experimental design, the procedure for the randomized assignment of treatment to experimental units, and the possible effects of treatment on otherwise likely covariates, suggest that analyses like those of Table I are the appropriate analyses. The results of sampling the randomization distribution of t reported in this article substantiate the approximate results of the parametric analysis of Table I.

We are gratified that there has been a new interest by statisticians in the design and analysis of weather modification. experiments. Needs for new considerations in the foundations of statistical inference have been revealed or emphasized. There are similar needs for new and improved insights into cloud physics and experimental procedures from meteorologists.

ACKNOWLEDGEMENT

Research supported by the U.S. Office of Naval Research under Contract No. N00014-76-C-0394. Reproduction in whole or in part is permitted for any purpose of the United States Government.

BIBLIOGRAPHY

Bradley, R. A., Srivastava, S. S. and Lanzdorf, A. (1977). Summarization of precipitation data for a weather modification experiment. *Proc. Fifth. Conf. on Prob. and Statist. in Atmosph. Sciences,* 201-205.

Bradley, R. A., Srivastava, S. S., and Lanzdorf, A. (1979). Some approaches to statistical analysis of a weather modification experiment. *Comm. Statist.-Theor. Method. A8,* 1049-1081.

Braham, R. R., Jr. (1979). Field experimentation in weather modification. *J. Amer. Statist. Assoc. 74,* 57-68.

Brillinger, D. R., Jones, L. V., and Tukey, J. W. (1978). *The Management of Weather Resources, Volume II: The Role of Statistics.* Report to the Secretary of Commerce, Statistical Task Force to the Weather Modification Advisory Board, Washington, U.S. Government Printing Office.

Brown, K. J., Thompson, J. R., and Elliott, R. D. (1975). Large scale effects of cloud seeding, final report, 1970-74 seasons. Tech. Report 75-2, Aerometric Research Inc., Goleta, CA.

Elliott, R. D. and Brown, K. J. (1971). The Santa Barbara II project-downwind effects. *Int. Conf. on Weather Mod., Preprint Vol.* 179-184, Canberra, Sept. 6-11, 1971.

Elliott, R. D. and Thompson, J. R. (1972). Santa Barbara convective seeding test program, 1970-71 season and 1967-71 summary. Tech. Publ. 5308, Naval Weapons Center, China Lake, CA.

Gabriel, K. R. (1979). Some statistical issues in weather modification. *Commun. Statist.-Theor. Methods A8,* 975-1015.

Kempthorne, O. (1975). Inference from experiments and randomization. *A Survey of Statistical Design and Linear Models,* (J. N. Srivastava, Editor), Amsterdam: North-Holland, 303-332.

Kruskal, W. (1979). Comment (Article by Braham cited above). *J. Amer. Statist. Assoc. 74,* 84-86.

Miller, A. J., Shaw, D. E., and Veitch, L. G. (1979). Analyzing the results of a cloud-seeding experiment in Tasmania. *Commun. Statist.-Theor. Methods A8*, 1017-1047.

Scott, E. (1979). A multivariate methodology for the analysis of weather modification experiments. Tech. Report M514, Department of Statistics, Florida State University, Tallahassee, FL.

Serfling, R. J. (1977). Toward a nonparametric covariance analysis of a weather modification experiment. Tech. Report M428, Department of Statistics, Florida State University, Tallahassee, FL.

OTHER TECHNICAL REPORTS ON THIS CONTRACT

Report Numbers	
M388, ONR-110	Hanson, Morgan, Bach, Charles L. and Cooley, Edward A., Bibliography of Statistical and Meteorological Methodology in Weather Modification, September, 1976.
M409, ONR-111	Bradley, Ralph A. & Srivastava, S.S., Correlation In Polynomial Regression, March, 1977.
M410, ONR-112	Bach, Charles L., An Interpretive History of 30-years (1945-1975) of Weather Modification, March, 1977.
M417, ONR-117	Bradley, Ralph A., Srivastava, S.S. and Lanzdorf, Adolf, Summarization of Precipitation Data In a Weather Modification Experiment: I. A Response Surface Approach, June, 1977.
M419, ONR-118	Gleeson, T.A., Data Summarization In a Weather Modification Experiment: II. Concomitant Variables, June, 1977.
M420, ONR-119	Hanson, Morgan A., Rank Tests in Weather Modification Experiments, June, 1977.
M428, ONR-122	Serfling, R. J., Toward a Nonparametric Covariance Analysis of a Weather Modification Experiment, August, 1977.
M440, ONR-126	Hanson, Morgan A., Barker, Lawrence E., and Hunter, Charles H. Bibliography of Statistical and Meteorological Methodology in Weather Modification, II, October, 1977.
M442, ONR-127	Scott, Elton, Data Summarization in a Weather Modification Experiment: III. A Multivariate Analysis, June, 1978.
M467, ONR-133	Bradley, Ralph A., Srivastava, Sushil S., and Lanzdorf, Adolf, An Examination of the Effects of Cloud Seeding in Phase I of the Santa Barbara Convective Band Seeding Test Program, June, 1978.
M469, ONR-134	Hanson, Morgan A., Barker, Lawrence E., Bach, Charles L., Cooley, Edward A., and Hunter, Charles H., A Bibliography of Weather Modification Experiments, July, 1978.
M490, ONR-135	Bradley, Ralph A., Srivastava, Sushil S. and Lanzdorf, Adolf, Some Approaches to Statistical Analysis of a Weather Modification Experiment, January, 1979.
M514, ONR-147	Scott, Elton, A Multivariate Methodology for the Analysis of Weather Modification Experiments, August, 1979.
M521, ONR-149	Bradley, Ralph A., Scott, Elton, Randomization Tests in Support of Some Statistical Analyses of a Weather Modification Experiment, October, 1979.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU No. M521 ONR No. 149	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE RANDOMIZATION TESTS IN SUPPORT OF SOME STATISTICAL ANALYSES OF A WEATHER MODIFICATION EXPERIMENT		5. TYPE OF REPORT & PERIOD COVERED Technical Report
7. AUTHOR(s) Ralph A. Bradley Elton Scott		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0394, Task Order No. 042-338
9. PERFORMING ORGANIZATION NAME & ADDRESS The Florida State University Department of Statistics Tallahassee, Florida 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME & ADDRESS Office of Naval Research Statistics and Probability Program Arlington, Virginia 22217		12. REPORT DATE October, 1979 13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Randomization tests, parametric analyses, weather modification, cloud-seeding, multiplicity of analyses, Monte Carlo methods.

20. ABSTRACT

The validities of analyses of variance and covariance applied to weather modification experimental data are examined through Monte Carlo samplings of the corresponding randomization distributions. It is found that the parametric tests approximate the randomization distributions well. Thus the use of parametric procedures in exploratory analyses of weather modification data, after appropriate transformation, is justified.

The authors comment also on the issue of multiplicity of analyses in data explorations. It is pointed out that analyses that are closely related and explore inherent sources of experimental variation are unlikely to result in misleading conclusions.