Comments on the Climax I and II Experiments Including Replies to Rangno and Hobbs

PAUL W. MIELKE JR.

Colorado State University, Fort Collins, Colorado

26 July 1993 and 9 September 1994

While these comments are stimulated by two articles on the Colorado State University (CSU) Climax I and II wintertime orographic cloud seeding experiments (Rangno and Hobbs 1987, 1993), they provide an opportunity to reflect in general on these experiments. Because Rangno and Hobbs (1987, 1993) question both the randomization and data of the Climax I and II experiments, I shall address these two specific questions and also add a few other comments, which might be relevant and of interest.

I first became aware of the Climax I experiment in January of 1964 (just after completing a Ph.D. in Biostatistics at the University of Minnesota) when F. A. Graybill of the CSU Mathematics and Statistics Department introduced me to L. O. Grant of the CSU Atmospheric Science Department. Consequently, I am not able to claim any credit for having designed the Climax experiments. Distinguished individuals such as F. H. Ludlam, V. J. Schaefer, and J. E. McDonald contributed to L. O. Grant's design of the Climax experiments.

1. Randomization

The randomization scheme of the experimental units in the Climax I experiment involved a sequence of randomly ordered pairs of seed and no-seed decisions (randomized blocks of size two consisting of one seed and one no-seed decision). Both the scheme and its implementation were provided by F. A. Graybill and D. L. Bentley, also of the CSU Mathematics and Statistics Department. Besides the fact that they provided and implemented the randomization scheme of the Climax I experiment, these statisticians played no further roles with the Climax experiments. I did prepare the randomization scheme for the entire Climax II experiment. The sequence of seeded and nonseeded experimental units for the Climax II experiment consisted of varying sized randomized blocks of randomly ordered seed and no-seed decisions with the restriction that the number of seed and no-seed decisions in each block would be equal (about 20–50 decisions constituted each block). This altered randomization sequence for the Climax II experiment was prompted by a discussion with J. Neyman (a University of California statistician) who stressed the importance of allowing for longer sequences of either adjoining seeded or adjoining nonseeded experimental units. The experimental units of the Climax experiments were 24-h periods (the starting times of experimental units near the start of the Climax I experiment differed from the last part of the Climax I experiment and the Climax II experiment; the times were changed to make the overlap between two adjoining experimental units occur during a pronounced diurnal minimum in precipitation, which was observed after the Climax I experiment had already started). The experimental design and related details are described elsewhere (Grant 1986; Grant and Mielke 1967; Mielke et al. 1970, 1971, 1981b). It should be noted that the particular randomization sequence of the Climax experiments has been questioned by Rangno and Hobbs (1987, 1993). Unfortunately, every experiment is at the mercy of its observed randomization sequence, no matter how peculiar it may seem. Criticism at this stage of an experiment is not justified. Replications are necessary to resolve issues regarding the randomization sequence.

The fact that meteorological conditions vary tremendously over a 24-h period has also been a concern. At the time the analyses were accomplished, it seemed important to have the analysis and experimental units be identical (i.e., 24-h periods). I now feel this notion of identical units may have hindered improved analyses. A suggestion by L. O. Grant to imbed eight 3-h analysis units on each 24-h experimental unit would yield a vast improvement since the atmospheric conditions are more uniform during a 3-h period than a 24-h period. (I hope that these alternative analyses will be completed.)

Since the random seed and no-seed decisions of the Climax I and II experiments were imbedded on a sequence of experimental units selected by U. S. Weather Bureau (USWB) duty forecasters in Denver, Colorado,

Corresponding author address: Dr. Paul W. Mielke, Department of Statistics, Colorado State University, Fort Collins, CO 80523.

these experiments could not have been deliberately biased by the Climax experimenters. The experimental design removed any involvement by the Climax experimenters in either the experimental unit declarations or the seed and no-seed decisions. Consequently, the Climax experimenters could not have excluded (or included) specific experimental units in their analyses to enhance the apparent effect of seeding. Many potential experimental units of both the Climax I and II experiments were eliminated because of wrong forecasts by the USWB. In addition, the Climax I experiment was shut down for specific periods of time because of either inadequate funding or interference by commercial seeding operations (Grant and Mielke 1967, p. 127). Even though there were no funding problems, the Climax II experiment was shut down for specific periods because of interference by commercial seeding operations. The Climax experiments were conducted in as honest a manner as any experiments I have ever encountered.

TABLE 1. All experimental units of the Climax I and II experiments.

	Sample sizes (seed, nonseed)	One-sided P values			Double ratio
		C0	C1	C2	
Climax I					
CRG-1	(104, 115)	0.245	0.459	0.523	0.989
CRG-2	(99, 113)	0.217	0.386	0.466	1.000
HAO	(101, 114)	0.365	0.166	0.134	1.120
ESO	(120, 131)	0.377	0.247	0.213	1.092
Climax II					
CRG-1	(149, 155)	0.622	0.490	0.490	0.948
CRG-2	(147, 154)	0.716	0.649	0.641	0.915
HAO	(170, 170)	0.165	0.065	0.055	1.098
ESO	(182, 190)	0.108	0.162	0.179	1.056
Climax I and II					
CRG-1	(253, 270)	0.270	0.542	0.521	0.967
CRG-2	(246, 267)	0.675	0.563	0.574	0.957
HAO	(271, 284)	0.168	0.041	0.029	1.109
ESO	(302, 321)	0.200	0.145	0.115	1.073

2. Target data

Another concern I wish to address involves precipitation data from USWB Station 05-1660, the Climax, Colorado hourly recording gauge (CRG). It was necessary to take the data directly from the raw output of CRG since it would have taken a number of years before the "official" USWB (now National Weather Service) data would be available from Asheville, North Carolina. Each precipitation value considered here is a 24-h period corresponding to a 24-h experimental unit of the Climax I and II experiments. Let CRG-1 denote the "official" CRG data reduced by USWB personnel, let CRG-2 denote the CRG data reduced by CSU personnel, and let HAO denote the snowboard at the High Altitude Observatory (1 m from CRG). The CRG and HAO data were collected by the same observers of the National Center for Atmospheric Research and the University of Colorado during Climax I and the first part of Climax II. A USWB observer from Leadville obtained the CRG and HAO data during the last part of Climax II after the High Altitude Observatory was closed. Also let ESO denote either (i) the average of CRG-2 and HAO when neither is missing, (ii) either CRG-2 or HAO when the other is missing, or (iii) the average of the Climax experiment snowboards 9 and 11 (located 1 mile to the south and north of CRG and HAO at the summit of Fremont Pass, respectively). Case (iii) was invoked for only 11 of the 623 experimental units of the Climax I and II experiments when both CRG and HAO were missing (Table 1 contains the frequencies of existing values for CRG-1, CRG-2, HAO, and ESO). An examination of Table 1 shows that 100, 110, and 68 of the total 623 experimental units are missing for CRG-1, CRG-2, and HAO, respectively.

It is important to note that Rangno and Hobbs (1987, 1993) require that analyses must be based on the "official" CRG-1 data. I consider it extremely risky to rely on data of which over 10% are missing (over 16% of the data are missing for CRG-1 and CRG-2). This is especially so because missing data are almost never lost randomly (this implies that any analysis based strictly on either CRG-1, CRG-2, or HAO has the possibility of being severely flawed). An inspection of the CRG raw data shows that most of the missing CRG-1 and CRG-2 values arise due to carrying over data from adjacent 24-h periods (i.e., a large portion of experimental units were missed when snow actually occurred). A specific example involving existing data occurred on 13 April 1969 when the precipitation values in inches of water were 0.60, 0.21, and 0.19 for CRG-1, CRG-2, and HAO, respectively (an inspection of the data indicates that about two-thirds of the CRG-1 value is carried over from the previous day; the carryover approach used by USWB personnel is good for climatological purposes, but bad when the data are used to analyze specific 24-h periods). Another feature, which was overlooked for a long time with the CRG-2 and HAO data, was that the values were interchanged for four of the five years during the Climax II experiment and, thus, affected the results pertaining specifically to analyses involving CRG-2 and HAO in two papers (Mielke et al. 1971, 1981b). While the HAO and CRG-2 sites were only 1 m apart, the days with missing values were not the same (the ESO data remained the same due to its definition). The question regarding CRG-1 unfortunately raises an additional concern involving some subsequently discussed control data that were also based on hourly recording gauge data (being "official" does not imply that the values are correct).

TABLE 2. 500-mb warm temperature partition ($-20°C \leq T \leq -11°C$).

	Sample sizes (seed, nonseed)	One-sided P values C0	C1	C2	Double ratio
Climax I					
CRG-1	(32, 39)	0.216	0.160	0.169	1.254
CRG-2	(29, 37)	0.111	0.071	0.087	1.318
HAO	(30, 37)	0.062	0.013	0.009	1.616
ESO	(35, 41)	0.153	0.038	0.021	1.542
Climax II					
CRG-1	(63, 57)	0.061	0.189	0.238	1.041
CRG-2	(67, 61)	0.003	0.026	0.081	1.029
HAO	(69, 62)	0.001	0.005	0.017	1.130
ESO	(73, 68)	0.003	0.020	0.056	1.094
Climax I and II					
CRG-1	(95, 96)	0.057	0.131	0.147	1.135
CRG-2	(96, 98)	0.003	0.018	0.047	1.156
HAO	(99, 99)	0.0003	0.0004	0.001	1.308
ESO	(108, 109)	0.002	0.004	0.007	1.260

3. Control data

The eight control stations of the Climax I and II experiments (Grant and Mielke 1967, p. 128) were selected by the end of the third year of the Climax I experiment (Grant and Mielke 1967, p. 119). The selections were based on a stepwise regression algorithm using standard USWB stations southwest, west, and northwest of the target. Even though data from the eight control stations were used by Grant and Mielke (1967), the identification of these eight control stations was given later (Mielke et al. 1981b). The oversight of not identifying the eight control stations in the original paper was my error. The initial analyses of the Climax I and II data (Mielke et al. 1970, 1971) were obtained without the control data (another retrospective blunder on my part). I was presented some material a few years later by J. Owen Rhea, which raised severe concerns in my mind about the possibility (Mielke 1979). This concern then led to the reanalysis paper (Mielke et al. 1981b), which finally accounted for the control data and ended up with much stronger statistical results (lower P values) than had been obtained in the previous analyses (Mielke et al. 1970, 1971). The Mielke et al. (1981b) paper reflects my present view on this topic.

4. Partitions

The partitioning criteria (e.g., 500-mb temperatures and 700-mb wind directions) of the Climax I and II experiments were based on objective procedures that did not depend on the target response variables (specific names of individuals who performed this important task included Charles F. Chappell and J. Michael Fritsch). Analyses of the type contained in Mielke et al. (1981b) for CRG-1, CRG-2, HAO, and ESO are given (despite the possibility of obvious flaws involving CRG-1, CRG-2, and HAO results due to missing data) for comparative purposes in Tables 1, 2, and 3 for all experimental units, the 500-mb warm temperature partition, and the 700-mb southwest wind direction partition, respectively. (The CRG-2 and HAO Climax II data have been corrected.) The choice of the warm 500-mb temperature partition ($\geq -20°C$) was based on physical criteria due to F. A. Ludlum before any statistical tests were performed on data of the Climax I experiment (Grant and Mielke 1967, p. 128). The results given in these analyses for ESO are consistent with my present view involving the Climax I and II experiments. The CRG-1, CRG-2, and HAO results must naturally be questioned due to the fact that a large amount of possibly pertinent data are missing. However, the ESO results account for every one of the 623 Climax I and II experimental units.

The subsequent use of the control data eliminated my previously stated concern that the Climax results could have been the result of a type I statistical error (Mielke 1979). Since I did not use the control data in the initial analyses (the last few years of the control data were not even entered in the Climax dataset until late in 1979 when we began the reanalysis of the Climax experiments using the control data), the use of the control data was a personal test of my stated concern. The results of the Mielke et al. (1981b) paper demonstrated that the concern cited in my 1979 comment was not justified. Since these data remain available, I do not understand why Rangno and Hobbs (1987, 1993) did not attempt to test statistically the conjectures they have made. (Perhaps the data would support their conjectures regarding the Climax I and II experiments as well as they supported my 1979 type I sta-

TABLE 3. 700-mb southwest wind direction partition ($190° \leq D \leq 250°$).

	Sample sizes (seed, nonseed)	One-sided P values C0	C1	C2	Double ratio
Climax I					
CRG-1	(21, 20)	0.174	0.040	0.031	2.224
CRG-2	(19, 20)	0.168	0.029	0.024	2.258
HAO	(23, 23)	0.190	0.062	0.062	1.484
ESO	(26, 25)	0.168	0.040	0.050	1.544
Climax II					
CRG-1	(43, 36)	0.049	0.051	0.108	1.419
CRG-2	(44, 38)	0.016	0.027	0.058	1.366
HAO	(47, 42)	0.003	0.003	0.007	1.552
ESO	(52, 48)	0.003	0.005	0.012	1.417
Climax I and II					
CRG-1	(64, 56)	0.014	0.006	0.012	1.673
CRG-2	(63, 58)	0.007	0.004	0.008	1.635
HAO	(70, 65)	0.003	0.001	0.003	1.500
ESO	(78, 73)	0.002	0.001	0.004	1.463

tistical error conjecture.) My present conclusion is that the data do support the alternative hypothesis that cloud seeding did increase precipitation amounts in the target area of the Climax I and II experiments during the meteorological partitions associated with Tables 2 and 3.

5. A methodological issue

The final topic of these comments involves a conflict between the data analyses and an initial recognition of a fundamental flaw in the statistical techniques used to implement the Climax analyses. The nonparametric analyses in the Mielke et al. (1981b) paper were based on statistical techniques described in Mielke (1972). During the revision of another manuscript (Mielke et al. 1981a), it came to my attention for the first time that the statistical techniques contained in Mielke (1972) were geometrically inconsistent with the data being analyzed. (This concern still disconcertingly pertains to some of the most commonly used statistical techniques such as the two-sample t test, one-way analysis of variance, the Wilcoxon–Mann–Whitney rank test, and many other routinely used statistical methods.) The latter observation was analogous to developing a physical model for universe A with entirely different physical principles of a universe B. This problem with statistics is a consequence of C. F. Gauss's default contribution of least-squares regression simply because he lacked the computational power around 1805 to implement an alternative regression technique (Mielke 1991; Sheynin 1973). Consequently, two further analyses of the Climax experiments were performed to ensure that the data space and analysis space were congruent with one another (Mielke 1985; Mielke et al. 1982). Since double ratio estimates were commonly used in weather modification and were also seriously affected by this previously mentioned geometric inconsistency, an alternative ratio estimator was also suggested (Mielke and Medina 1983) to overcome this geometric concern, which was on my mind at the time. Hopefully, this discussion explains why these reanalyses continued to appear after the Mielke et al. (1981b) reanalysis paper. (These efforts were driven by purely statistical concerns and without regard to the Climax I and II experiments.)

6. A general appraisal

The patterns of differences between target and control means for seeded and nonseeded experimental units of the warm 500-mb temperature partition in Table 1 of Mielke (1985) is consistent with my present feeling about the effects of orographic cloud seeding: 1) some cases, which would naturally have no precipitation, will have small amounts, 2) some cases with small amounts will be increased, and 3) little or no increases occur with cases involving larger storms with an adequate natural seeding mechanism. Another feature, which has seemed peculiar to me, is that so much attention has continually been placed on the warm 500-mb temperatures when the southwest 700-mb wind directions yield extremely consistent positive results for all the analyses of both Climax I and II (Mielke et al. 1971, 1981b). Perhaps this arises because the 700-mb wind directions were not mentioned in either Grant and Mielke (1967) or Mielke et al. (1970); this may be attributed in part to an emphasis at the time on J. W. Tukey's notion of exploratory and confirmatory experiments. An exploratory experiment is the usual type of experiment that is intended to explore possible effects, which might be attributed to a specific treatment. (Such effects may not be anticipated in advance of the experiment.) In contrast, a confirmatory experiment is designed to confirm one or more specific effects of a treatment, which have been well documented prior to the experiment. When a substantial effect is demonstrated in two or more carefully replicated independent experiments, I feel the effect has been confirmed (neglecting to state an effect in advance eliminates being confirmed in the rigid sense of a confirmatory experiment). Since cloud seeding experiments involve such complex phenomena, a sequence of well-designed sets of replicated independent experiments is recommended. When serious doubts arise for a given set, then additional replications may be needed. While this is admittedly a very time consuming process, technological development and improvement in a discipline such as weather modification require this type of effort.

Acknowledgments. This work has been supported by National Science Foundation Grant 5-3 3277. The author also wishes to express his sincere appreciation to the numerous individuals who helped implement the Climax I and II experiments.

REFERENCES

Grant, L. O., 1986: Hypotheses for the Climax wintertime orographic cloud seeding experiments. *Precipitation Enhancement—A Scientific Challenge, Meteor. Monogr.,* No. 43, R. R. Braham, Jr. Ed., 105–108.

——, and P. W. Mielke, 1967: A randomized cloud seeding experiment at Climax, Colorado, 1960–1965. *Proceedings on the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 5, University of California Press, 115–131.

Mielke, P. W., 1972: Asymptotic behavior of two-sample tests based on powers of ranks for detecting scale and location alternatives. *J. Amer. Statist. Assoc.,* **67**, 850–854.

——, 1979: Comment on field experimentation in weather modification. *J. Amer. Statist. Assoc.,* **74**, 87–88.

——, 1985: Geometric concerns pertaining to applications of statistical tests in the atmospheric sciences. *J. Atmos. Sci.,* **42**, 1209–1212.

——, 1991: The application of multivariate permutation methods based on distance functions in the earth sciences. *Earth-Sci. Rev.,* **31**, 55–71.

——, and J. G. Medina, 1983: A new covariate ratio procedure for estimating treatment differences with applications to Climax I and II experiments. *J. Climate Appl. Meteor.*, **22**, 1290–1295.

——, L. O. Grant, and C. F. Chappell, 1970: Elevation and spatial variation effects of wintertime orographic cloud seeding. *J. Appl. Meteor.*, **9**, 476–488. [Corrigenda, **10**, 842; **15**, 801.]

——, ——, and ——, 1971: An independent replication of the Climax wintertime orographic cloud seeding experiment. *J. Appl. Meteor.*, **10**, 1198–1212. [Corrigendum, **15**, 801.]

——, K. J. Berry, P. J. Brockwell, and J. S. Williams, 1981a: A class of nonparametric tests based on multiresponse permutation procedures. *Biomertika*, **68**, 720–724.

——, G. W. Brier, L. O. Grant, G. J. Mulvey, and P. N. Rosenzweig, 1981b: A statistical reanalysis of the replicated Climax I and II wintertime orographic cloud seeding experiments. *J. Appl. Meteor.*, **20**, 643–659.

——, K. J. Berry, and J. G. Medina, 1982: Climax I and II: Distortion resistant residual analyses. *J. Appl. Meteor.*, **21**, 788–792.

Rangno, A. L., and P. V. Hobbs, 1987: A re-evaluation of the Climax cloud seeding experiments using NOAA published data. *J. Climate Appl. Meteor.*, **26**, 757–762.

——, and ——, 1993: Further analyses of the Climax cloud-seeding experiments. *J. Appl. Meteor.*, **32**, 1837–1847.

Sheynin, O. B., 1973: R. J. Boscovich's work on probability. *Arch. Hist. Exact Sci.*, **9**, 306–324.