Reexamination of Historical Regression Analysis Applied to a Recent Idaho Cloud Seeding Project

Arlin B. Super and James A. Heimbach, Jr.
St. Cloud, Minnesota Springvale, Maine

Abstract. A recently reported analysis of an Idaho operational winter cloud seeding project is examined in detail. The analysis used the traditional historical regression method. It appeared to provide impressive evidence that seeding was effective in increasing seasonal snowpack accumulation during each of four winters with a mean increase near 12%. The analysis was based on a strong relationship with the April 1st control station mean explaining 96% of the target area variance. However, frequent snow melt prior to April 1st was discovered at 4 of 7 control stations and 3 of 10 target stations. Snow melt is shown to have introduced an important but apparently unrecognized variable into the target-control relationship, making it inappropriate for evaluation of seeding effectiveness. Repeating the analysis procedures with March 1st observations from the same stations reduced the "seeding signal" to under 4%. Additional target and control stations were identified and used in comprehensive historical regression analyses. Once a limited number of stations with February melt were discarded, all possible combinations of available control stations were tested to detect target-control relationships which explained the most variance. These relationships did not support the hypothesis that cloud seeding significantly enhanced the seasonal snow water content. Additional testing was done with April 1st observations to demonstrate that a wide range of results can emerge from a large database even with the requirement of a strong target-control association. Recommendations are made for future application of this statistical approach.

1. INTRODUCTION

The historical regression method, used with target and control area measurements, has long been employed in evaluation of cloud seeding projects intended to increase precipitation. Measurements of seasonal snow accumulation, rainfall or streamflow are compared between the intended target area and upwind and crosswind locations (control areas) unlikely to be affected by the seeding. Several nonseeded seasons are needed to establish a historical relationship by linear regression analysis. Control area observations during seeded seasons are used with this relationship to predict expected target area amounts in the absence of seeding. Departures from these predictions are examined for evidence of cloud seeding effectiveness. It is generally assumed that the stronger the target-control relationship, the more successful the method can be in detecting relatively small departures possibly caused by seeding. A linear correlation coefficient (R-value) is typically calculated to express the degree of relationship, or association, between target and control station means, while R^2 is considered the variance explained. This paper will demonstrate that a high R-value does not guarantee a valid means of evaluating seeding effectiveness when snow melt is strongly involved in the target-control relationship.

An early example of this analysis approach was reported by Thom (1957). Dennis (1980) discussed the historical regression method in detail, pointing out possible sources of uncontrolled variance and bias. He noted that some statisticians considered uncertainties sufficient to render useless all such analyses of operational programs. But others believed useful information could be obtained if uncontrolled variance and bias were identified and reduced. Dennis (ibid.) referred to Court (1960) when stating that, "Agreements before a project begins as to which rain gages are to be included in calculating the target and control rainfall, for example, go far toward eliminating both unconscious bias and any temptation to select data to demonstrate a desired result."

Application of this method to nonrandomized observations has potentially serious problems. Therefore, indications of changes associated with seeding should be considered no more than suggestive, and certainly not "proof" in the sense that scientists and statisticians use that term. The authors of this paper used the historical regression approach to seek supporting evidence for their exploratory statistical analyses of the randomized Bridger Range Experiment (Super and Heimbach 1983). Their historical target control regression analyses was not part of the original design, and target and control stations were chosen after the fact, albeit by an objective method. Consequently, the authors were careful to use "suggested" and similar adjectives with indications of possible snowfall increases.

Griffith and Solak (2002), hereafter GS, reported an economic feasibility assessment of winter cloud seeding in the Boise River Basin of southern Idaho. The basis for that assessment was their application of historical regression analysis to operational seeding over four consecutive winters. They concluded that, "An evaluation of the effectiveness of these (cloud seeding) programs indicated an average of a 12% increase in the target area April 1st snow pack water content." The target-control relationship had a high R-value of 0.98, indicating 96% of the variance was explained by the historical regression equation based on 32 pre-seeding winters.

This paper is limited to consideration of the historical regression method and does not address the GS economic assessment. While numerous journal articles have used this method over the years, the GS analysis was chosen for in-depth scrutiny because it was recent and claimed impressive seeding effects for each of four winters, based on a strong target-control relationship. In short, the results were impressive when taken at face value. However, a number of topics were not addressed by GS, such at why particular control stations were selected while other available stations were ignored. The analyses herein will illustrate some of the potential shortcomings that can occur with misapplication of this statistical method. This paper has general relevance concerning this method , but especially as applied to operational winter orographic cloud seeding projects. The intent is to illustrate some complexities and pitfalls of statistical analyses of operational programs.

This paper is organized as follows. The Idaho seeding project and elements of the GS analysis are briefly described in Section 2. Available snow water equivalent observations are discussed in Section 3. Section 4 presents a repeat of the GS analysis to insure data consistency, plus methodology to estimate probability levels. A serious problem with use of April 1st observations; namely, prior frequent snow melt at several stations, is considered in Section 5, divided into 5 subsections. Section 6 presents a comprehensive new analysis of the seeding project with March 1st observations, using all available control and target stations and inclusion of post-seeded seasons. Use of April 1st observations is further explored in Section 7 to better illuminate how such an apparently strong suggestion of seeding effectiveness could be misinterpreted. An attempt is made in Section 8 to select control station groupings to show a desired result. Such "mining" of the data violates the scientific method, but illustrates how false results can be "found" even with strong control-target relationships. Possible reasons for the apparently ineffectual Idaho seeding are briefly considered in Section 9. Section 10 presents a summary and recommendations for future use of the historical regression method applied to winter orographic cloud seeding projects.

2. OPERATIONAL SEEDING PROJECT AND INITIAL ANALYSIS ELEMENTS

Griffith and Solak (ibid.) reported on an operational seeding program that was intended to increase the seasonal snow water equivalent (SWE) in the Boise River Basin target area above Lucky Peak Dam in southern Idaho. Silver iodide (AgI) seeding was conducted between 10 Dec. 1992 - 31 May 1993; 22 Nov. 1993 - 30 April 1994; 15 Nov. 1994 - 15 Apr. 1995 and 1 Nov. 1995 - 31 Mar. 1996. For simplicity, each season will henceforth be identified by the end-of-season year. Thus, seeding operations will be said to have been conducted from 1993-96.

The seeding program used 20 silver iodide (AgI) generator sites shown on Fig. 4 of GS, a map which indicates 13 generators were within the target area boundary while 7 others were outside. The latter followed the western half of the target's perimeter. Generator output was reported as 8 grams of AgI per hour per unit. The ground-based generators were operated by local residents when contacted by project meteorologists. Generators were at relatively low elevations, ranging from 1000 to 5400 ft. All elevations are reported in feet above mean sea level.

Generator locations suggest transport winds were expected to be from approximately south-southwest to north-northwest. This placement is in reasonable agreement with Huggins (1997) who found most 850 mb winds were between 150 to 330 degrees with a median of 238 degrees when liquid water was detected above an adjacent basin. The 850 mb level is near the elevation of the highest AgI generators and should approximate low-level transport. Winds at 700 mb were typically westerly. Reference to Figs. 4 and 5 of GS suggests their control station selection generally minimized contamination. Banner Summit (control station 1), at the eastern end of the Payette Basin immediately north of the target area, is most likely to have been affected if contamination was a problem. It is also the control station nearest to, and most highly correlated with, the target area mean, providing an R-value of 0.973. That value is near R = 0.98 calculated for the mean of all 7 controls used by GS, which explains only 1.4% more variance. Use of contaminated control stations should provide conservative results, assuming any seeding effects resulted in net SWE increases. Some regression results will be given without inclusion of Banner Summit, but the evidence suggests that contamination was insignificant.

April 1st SWE observations were used by GS to evaluate seeding effectiveness. Choice of particular target and control SWE observing stations was briefly explained by GS as follows: "For this program, ten target area and seven control area sites were identified with long-term historical records. Control stations were selected in *close proximity* (emphasis added) to the target stations and at similar elevations in order to obtain as strong correlations as possible." In fact, mean elevations for the target and control stations were 7220 and 6380 ft, respectively, a difference of 840 ft. That difference was significant for snow melt prior to April 1st as discussed below. Higher correlations were found with a number of combinations of the seven control stations, two of which are listed in Table 5, to be discussed in Section 7.

"Close proximity" is a term not everyone might apply to the more distant control stations. Target station 1 was used as a reference point since it was located near the center of the higher elevation portions of the target area, expected to be most impacted by seeding. The distances from that station to control stations E, F and G (also listed as 5, 6 and 7 in Table 1), all well southwest of the target, were 104, 122 and 91 statute mi, respectively. In contrast, the other 4 control stations were from 38 to 67 mi north to north-northwest of the target reference site. All target stations were within 34 mi of target station 1.

GS selected 1961-92 as the historical period, providing a sample of 32 nonseeded winters. There is good reason for the 1961 beginning because a number of stations lacked earlier data. They did not use post-seeding years 1997-2001. Observations from 1997 were discarded because of "contamination concerns" about seeding in the adjacent Payette Basin north of the their target basin. Yet by their approach 1997 had a 13% *positive* departure in the Boise Basin, well above 2 of the 4 seeded seasons. GS noted that the mean of years 1998-2001 was 2% above predictions from the regression equation, near the expected value of zero departure.

GS stated that their target and control stations were selected *after* the first winter of seeding and remained unchanged the following three winters. Therefore, some concern of bias exists, especially since their first winter's evaluation showed a 16.5% SWE increase. That high value was exceeded only by the third winter with 17.4%, and was well above suggested increases of 6.5 and 5.0% for the other seeded winters. Choosing and reporting all observing stations before any seeding obviously enhances credibility of later reported results.

3. SNOW WATER EQUIVALENT DATA

Snow water equivalent measurements were obtained from the Natural Resources Conservation Service (NRCS) Snow Survey Program. Station data are readily assessable from their web site. The 7 control stations and 10 target stations used by GS are given in Tables 1 and 2, respectively, using their number and letter identifiers. These tables also list an additional 7 control and 2 target stations which were found and used later in this study.

A search was made for all NRCS snow course and Snow Telemetry (SNOTEL) locations that met the criteria used by GS for control stations. That is, they were upwind or crosswind of the target basin and no greater than about 120 miles from the center of the target. Because of an obvious March melt problem at lower elevation stations, discussed below, additional stations were required to have a minimum elevation of 6000 ft. Station records were required to extend from 1961 to 2001. The regions searched were the two adjacent areas between 42 deg 30 min to 45 deg 15 min latitude and 116 to 117 deg longitude, and between 44 deg 30 min and 45 deg latitude and 115 to 116 deg longitude. These areas encompass all mountains within about 120 mi distance in directions unlikely to be affected by Boise Basin seeding. Seven additional stations were found besides the 7 used by GS, with distances from target station 1 ranging from 81 to 108 mi, all less than the 122 mi distance to original control station 6. Names, locations and other information concerning the 14 potential control stations are given in Table 1 where stations 1-7 correspond to stations A-G in Table 5 of GS. Both their letter identifiers and the numbers used in this study are included in Table 1. The additional 7 stations added in this study, numbers 8-14, are noted as "new." NCRS identifiers in Tables 1 and 2 which end with "S" indicate a SNOTEL station as of 2001; others are long-term snow courses. The right-most column in Table 1 refers to the R-value between the mean April 1st SWE values of the 10 original target stations of Table 2, and observations from individual control stations, for the 32 winter historical period.

Table 1. Names and locations of **control** stations with distances referenced to target station 1. The 7 stations used by GS are identified both by their letter I.D.s and by numbers 1-7. Stations 8-14 were added in this study. Correlation coefficients were calculated between April 1st SWE accumulation at each station and the mean for target stations 1-10, for the 1961-92 historical period.

Station I.D.	Station Name	NRCS I.D.	Distance (miles)	Elevation feet (msl)	Latitude deg-min	Longitude deg-min	Correlation Coefficient
1-A	Banner Summit	15E11S	38	7040	44 18	115 14	0.973
2-B	Big Creek Summit	15E02S	67	6580	44 38	115 48	0.964
3-C	Cozy Cove	15E08S	42	5380	44 17	115 39	0.861
4-D	Deadwood Summit	15E04S	58	6860	44 33	115 34	0.938
5-E	Mud Flat	16G07S	104	5730	42 36	116 33	0.629
6-F	Red Canyon AM	16G11	122	6650	42 26	116 50	0.588
7-G	Silver City	16F03	91	6400	43 00	116 44	0.815
8-new	Bear Saddle	16E10S	104	6180	44 36	116 58	0.876
9-new	Bogus Basin	16F02	81	6540	42 46	116 06	0.899
10-new	Brundage Reservoir	16D09S	100	6300	45 03	116 08	0.941
11-new	Secesh Summit	15D01S	105	6520	45 11	115 58	0.938
12-new	South Mountain	16G01S	108	6500	42 46	116 54	0.714
13-new	Squaw Flat	16E05S	86	6240	44 46	116 15	0.874
14-new	Succor Creek AM	16F06	92	6100	43 08	116 52	0.375

Table 2. Similar to Table 1 but names and locations of **target** stations with distances referenced to station 1. Stations 1-10 were used by GS while 11 and 12 were added in this study.

Station I.D.	Station Name	NRCS I.D.	Distance (mile)	Elevation feet (msl)	Latitude deg-min	Longitude deg-min
1	Atlanta Summit	15F04S	0	7580	43 45	115 14
2	Camas Creek Division	15F09	34	5710	43 16	115 21
3	Dollarhide Summit	14F08S	30	8420	43 36	114 40
4	Galena	14F17S	30	7440	43 53	114 40
5	Galena Summit	14F12S	27	8780	43 51	114 43
6	Graham Guard Station	15F14S	14	5690	43 57	115 16
7	Moores Creek Summit	15F01	25	6100	43 55	115 40
8	Soldier Ranger Station	14F11S	28	5740	43 29	114 49
9	Trinity Mountain	15F05S	13	7770	43 38	115 26
10	Vienna Mine	14F04S	19	8960	43 48	114 51
11-new	Couch Summit #2	14F18	27	6840	43 31	114 48
12-new	Jackson Peak	15E09S	23	7070	44 03	115 27

A search was also made for any additional stations within the target area basin. The area of interest is approximately between 43 and 44 deg latitude and 114 deg 40 min to 116 deg longitude. Two stations were found besides the 10 used by GS. One is near target station 8 and the other is just within the north boundary of the Boise Basin. Reasons for not including these additional potential control and target stations were not discussed by GS. It may be that once a grouping of stations was found with the high R-value of 0.98, there seemed little point in considering other combinations.

Snow course and snow pillow observations were downloaded from the NRCS ftp site for all stations listed in Tables 1 and 2. Most locations currently use SNOTEL stations equipped with snow (pressure) pillows for SWE measurement. The remaining stations are long-term snow courses manually sampled on or near the first of the month. In fact, SNOTEL observations represent a combination of actual pillow observations since these telemetry stations were installed, plus *estimations* of earlier observations based on nearby snow course measurements. Estimation was done by the NRCS based on snow course and nearby pillow observations for the periods both were obtained. Almost all pillow observations used in this study began in 1981-82, and nearby snow courses were discontinued during 1990-92, providing approximately a decade of co-located measurements from both systems. These changes in measurement method and specific observing locations must introduce some unexplained variance.

All downloaded data were carefully double checked against the original ftp site data to insure that no errors had been made; for example, by downloading observations for the wrong station or month. Once a serious melt problem prior to April 1st was recognized, measurements from February 1st and March 1st were downloaded in addition to April 1st readings.

4. INITIAL ANALYSIS REVIEW AND P-VALUE CALCULATIONS

4a. Repeat of GS Initial Analysis

The first historical regression analysis done for this paper simply repeated the station combinations and use of the same 32 winter historical period to test that the same observations were used by GS and the current study. In fact, slight differences resulted in the intercept, A, and slope, B, of the regression equation of the form Y = A + B (X). Coefficients A and B were reported by GS as -1.68 and 1.16, respectively, with an R-value of 0.98. The present study calculated A as -1.368 and B as 1.148 with R equal to 0.980. A plot of the two equations showed them to be almost identical. Calculated departures from the regression equations for the 4 seeded winters had minor differences. Values calculated for this paper follow with GS values in parentheses: 1993 = 16.0% (16.5%); 1994 = 7.8% (6.5%); 1995 = 17.6% (17.4%) and 1996 = 5.0% (5.0%). The four-winter mean, stated as 12% by GS, is actually 11.4% using their individual values, very similar to the 11.6% value calculated for the current study. The reason for the slight differences is unknown, and the same values resulted from three independent software programs used in this study. In any event, essentially the same data were used by both studies, a point worth verifying. Unless specifically stated otherwise, all results to follow are based on the data and calculations used in this study.

Figure 1 shows the mean control-target pairs for all individual winters. All nonseeded regression equations calculated in this study are for the same 32 winters used by GS (1961-92) unless otherwise stated. April 1st observations for those winters are shown by solid dots, unlabeled to minimize clutter.

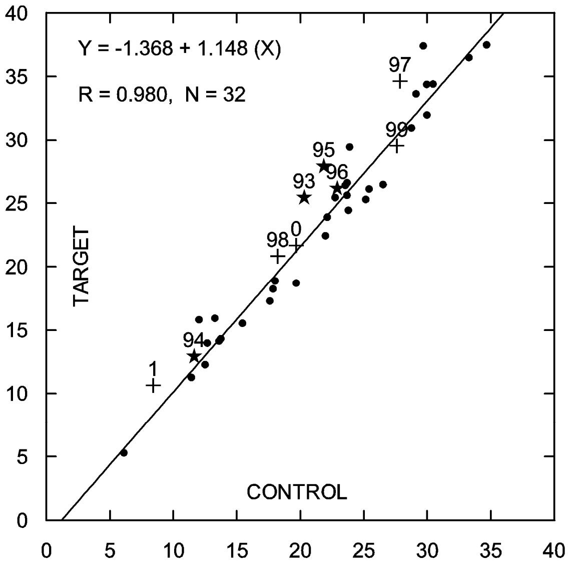

Figure 1. Linear Regression Plot for the Means of 7 Control Stations and 10 Target Stations using April 1st Snow Water Equivalent (inches).

The regression equation is plotted as a solid line. The 4 labeled seeded winters are shown by star symbols while crosses are used for the 5 labeled post-seeded winters, 1997-2001. While the seeded 1993 and 1995 positive departures are among the highest shown, percentage departures for two nonseeded winters, 1963 (unlabeled) and 2001 (labeled as 1), well exceeded those values at 27 and 28%, respectively. Departures of 13% or more also resulted from 1965, 1986, 1988 and 1997 observations. The 1994 and 1996 seeded winter departures were less than those values. Thus, 6 of the total of 37 nonseeded winters (16%) had similar or greater positive departures than the 2 *largest* of the 4 seeded winters. But since all 4 seeded winters had positive departures, and they were the only 4 consecutive winters with positive departures for the entire period 1961-2001, seeding may have been effective from the evidence presented thus far.

An analysis which suggests the presence of a "seeding signal" does not necessarily imply confidence that the signal is real. Analyses of weather modification programs, both research and operational, require searching for a seeding signal which is far smaller, typically by at least an order of magnitude, than the natural variability among response variables. Simply comparing precipitation from seeded units (seasonal SWE accumulation in this paper) to that of nonseeded units can be inconclusive because of the high variance among units. More sensitive methods are required. The historical regression is one means to remove some of the natural variance thereby making any seeding signal more obvious.

4b. Probability Levels From Rerandomization.

It is desirable to determine the probability (P-value) that a suggested seeding effect is real and not due to random selection of test units. Several robust (sensitive) nonparametric tests, which do not require assumptions about the distribution of the response variable, are available for *randomized* experiments. The Wilcoxan rank-sum test has been widely applied.

An effective technique for estimating the P-value of a possible seeding signal, without need to parameterize the precipitation, is to rerandomize the pool of seeded and nonseeded samples. This approach is often called the Monte Carlo technique (e.g., Dennis et al. 1975). The pseudo-statistical results from the rerandomized permutations are compared to the actual results to establish the significance of a statistical result. For example, if the actual residual of a seeded experimental unit is equal to or exceeded by 5% of the rerandomized pseudo residuals, then one can be confident that the null hypothesis can be rejected with a P-value of 0.05 without making a type I error that a seeding effect exists when there is none. This assumes the test is one-tailed; that is, seeding can only increase precipitation, not decrease it.

Monte Carlo testing has been used to estimate P-values for operational seeding (e.g., Griffith et al. 1997). However, it is important to recognize that this approach is not really valid and "rerandomization" is not technically correct in the absence of initial randomizaion. P-values cannot be calculated with precision even during exploratory analysis of randomized experiments (Gabriel 2000). He notes that, "Data from cloud seeding operations must be used very circumspectly, and can never have the same evidential weight as do data from randomized experiments." Gabriel (ibid.) goes on to state that it has been well known and documented from the early days of cloud seeding that, "--- operators cannot produce conclusive evidence of the effect of their operations." Calculating P-values by Monte Carlo techniques does not change the basic fact that nonrandomized operational seeding programs are not randomized experiments, and, therefore, are not protected from unforeseen sources of bias.

With that major caveat, P-values were calculated for this study, *to be used only for general guidance with full realization that that they are not totally valid and are likely overly optimistic.* Their use here is only to provide an approximate framework within which to consider whether the Idaho operational seeding results should be considered even suggestive. For example, if a particular P-value is larger than 0.05, that result can likely be rejected as not even remotely suggestive. A P-value of 0.01 might suggest the particular result is worth further consideration, but does not constitute "proof" of seeding effectiveness in the absence of randomization.

Monte Carlo evaluations used samples from pooled seeded and nonseeded pairs of control-target data from 1961-96 by applying calls to a random number generator. One permutation consisted of randomly drawing pairs of control-target samples from the pool without replacement until all 36 were selected. The first 32 were designated as nonseeded and were used to derive regression coefficients. The four remaining pairs were designated as seeded, and their residuals as well as their mean were derived from the regression equation. This process was repeated 1000 times and the residuals were compared to those from the actual data set. One thousand permutations gives an acceptable approximation to an exhaustive rerandomization of all possible permutations because all are equally likely to occur. The P-value is that portion of the permutations which had mean residuals as large as or greater than the actual observed residual (Dennis et al. ibid.).

One-tailed P-values were calculated for the control-target sites selected by GS for the repeat analysis discussed in Section 4a. The resulting P-values were: 1993 = 0.08; 1994 = 0.19; 1995 = 0.04; 1996 = 0.23 with the mean = 0.01. Only the 1995 winter had a P-value less than 0.05, usually considered "significant" in testing of randomized weather modification experiments. However, the mean P-value of 0.01 suggests, but does not prove, that there might have been a real seeding effect. Based only on these results, GS were perhaps justified in suggesting that Idaho operational seeding increased the April 1st SWE accumulation by an average near 12%.

5. THE SNOW MELT PROBLEM

5a. Pre-April 1st Snow Melt and Ramifications

It was noticed that some control stations had frequent low SWE values on April 1st, and even zero values (no snow existed) for some winters. Controls 5 and 6 and, to a lesser degree, station 7 were especially prone to low readings. For example, station 5 had 7 zero values in the 32 winter historical period while station 6 had 5 zeros, both clear indications that significant snow melt had occurred prior to the April

1st sampling date. Of all 14 potential controls in Table 1, stations 5 and 6 have the lowest R-values with the exception of added station 14. It was suspected that these low R-values were partially caused by late winter melt affecting these control stations much more than the target station mean. These findings suggested that snow melt prior to April 1st was frequent and significant at some control stations and possibly some target stations as well.

Presence of significant snow melt prior to the seasonal sampling date can introduce considerable uncertainty into the historical regression method. The main problem is introduction of a second important variable besides snowfall which affects SWE accumulation but is unrelated to seeding. The impact of melt can easily have more impact than seeding effects, stated by GS to average 2.5 inch SWE. When frequent and significant melt occurs prior to the sampling date, the historical regression relationship is no longer valid for evaluation of seeding effectiveness. The relationship is then based not only on widespread snow accumulation but on an unknown snow melt pattern which varies among stations and seasons, even when strong target-control relationships are found. The degree of pre-sampling melt during an individual seeded season will affect departures from the target-control relationship in unknown ways, complicating the search for seeding effects. Melt will be a particular problem when it effects the snow water content differently over target and control areas, as in this study where high altitude control stations were not available.

Snow melt during late winter and early spring is generally dependent upon the prevailing weather, and is enhanced by warm, sunny days or widespread rains. Moreover, melt is a function of several terrain-related parameters, especially elevation, but also slope, aspect, local exposure and relationship to higher terrain (e.g., foehn winds). Thus, significant melt introduces serious unexplained variance into historical regression analysis, degrading the search for possible seeding effects in an unknowable manner. These concerns were partially responsible for Super and Heimbach (1983) rejecting April 1st snow course data, and instead using March 1st observations, as part of their analyses of the Bridger Range Experiment in the colder Montana climate.

5b. Repeating Initial Analysis with March 1st Data

A simple test of potential impacts of March melt is to duplicate the analysis discussed above, but using March 1st observations. If the seeding was truly effective, similar percentage departures would be expected on both March 1st and April 1st as there is no reason to expect exceptionally high cloud seeding potential during March.

The R-value resulting from use of March 1st observations was a respectable 0.983, above the 0.980 value for April 1st data from the same stations. Individual seeded winter departures for 1993-96 were 3.9%, 3.1%, 7.2% and 1.0%, respectively, with a mean of 3.8%. The individual seasonal range from 1-7% indicates considerable variability in seeding effectiveness and/or unstable results from this statistical method. The mean value was only 1/3 of that suggested by April 1st observations. This finding by itself must cast serious doubt on the 12% mean seasonal SWE increase claimed to result from seeding. The P-values estimated for these results by the rerandomization method discussed above were: 1993 = 0.30; 1994 = 0.35; 1995 = 0.14; 1996 = 0.45 and mean = 0.29. None of the P-values begin to approach the usual P = 0.05 criteria so suggestions of seasonal seeding effects up to March 1st cannot be considered to have any statistical significance. In other words, there is a high probability that these individual and mean departures were chance events.

5c. Estimation of March Melt Frequencies

An attempt was made to assess March snow melt by comparing March 1st and April 1st observations for all target and control stations and all 41 winters of record from 1961-2001. Finding less SWE on April 1st is a certain indication of melt during March. It is impractical to adjust March melt by precipitation because most of the stations did not have nearby measurements during much of the historical period. Furthermore, gauges do not indicate whether the precipitation is snow, which enhances SWE accumulation, or rain, which decreases it.

If the unknown snowfall exceeded the water content of any March melt, the April 1st value would be greater even when melt occurred. Therefore, this approach to snow melt estimation is an underestimate. Nevertheless, subtracting the two monthly measurements, and calculating the frequency of lower April 1st readings, provides a conservative first approximation of March melt. The same approach was used to approximate February melt. Resulting melt frequency estimates for February and March are listed in Table 3. For example, Banner Summit had 12% melt for March indicating that 5 of the 41 winters had measurable melt between March 1st and April 1st readings.

Table 3. Approximate and Conservative February and March Snow Melt Frequencies at Control and Target Stations.

Control No.	Elevation (feet)	February Melt %	March Melt %
01-A	7040	5	12
02-B	6580	2	0
03-C	5380	0	44
04-D	6860	2	10
05-E	5730	22	73
06-F	6650	37	78
07-G	6400	8	46
08-new	6180	0	10
09-new	6540	2	20
10-new	6300	0	12
11-new	6520	2	5
12-new	6500	5	37
13-new	6240	2	12
14-new	6100	22	59

Target No.	Elevation (feet)	February Melt %	March Melt %
01	7580	0	2
02	5710	23	66
03	8420	0	0
04	7440	0	7
05	8780	0	2
06	5690	5	46
07	6100	2	12
08	5740	15	78
09	7770	0	7
10	8960	0	0
11-new	6840	5	24
12-new	7070	5	2

A strong relationship between March melt and elevation was found for the target stations of Table 3 with 64% of the melt frequency variance explained by linear regression with elevation. In contrast, only 17% was explained for control stations. Major reasons for this difference are likely the 3270 ft elevation range for target stations, twice that of the 1600 ft range for controls, and the fact that mean and median target elevations, both near 7200 ft, are above the highest available control station located at 7070 ft. Mean and median values for controls are near 6400 ft, or 800 ft lower than corresponding target values. The significantly lower elevations of control stations makes them particularly susceptible to March melt. All stations below 6000 ft had frequent March melt.

Table 3 indicates that several stations, both control and target, had frequent March melt. Some even had frequent melt during February including 2 of the 7 controls and 2 of the 10 target stations used by GS. Four of these original control stations and 3 of the original targets had March melt frequencies in the high 44 to 78% range. This shows that even the conservative March melt estimate was very common at some locations occurring approximately on 1/2 to 3/4 of the seasons. Fifteen of the total of 26 control and target stations had greater than 10% March melt frequencies. The frequent and widespread March melt must have introduced considerable unexplained variance into regression analyses using April 1st observations, no matter what the target-control R-values. Since 4/7 (57%) of the original controls but only 3/10 (30%) of the target station had very frequent March melt, mean control values were more affected than mean target values. This fact raised the concern that winters with positive departures from the regression equation of Fig. 1 might have tended to have higher than usual control station snow melts.

Examination of SWE differences between March 1st and April 1st at control stations 5, 6 and 7, all highly prone to melt, revealed that seasonal positive departures were associated with March melt at those stations. Over the 41 seasons of record, 16 had the same or less SWE on April 1st at all three of those stations. Nine of the ten seasons from 1992-2001 were in that category including all 4 of the seeded years. Eight of these ten seasons had positive departures from the regression equation including all 4 of the seeded years. Possible reasons for this increase in melt frequency with time (global warming, extended drought) are beyond the scope of this paper. The important point is that this evidence of increased melt frequency during the past decade strongly suggests that the historical regression relationship was not constant with time, one of the major assumptions behind applying it. Dennis (1980) expressed the view that, "The most serious difficulty with the historical regression method has to do with the stability in time of the target-control relationship." While the evidence here suggests warmer, dryer winters in the past decade, local effects such as tree growth, forest cutting and disease or fires can also modify target-control relationships over time.

5d. <u>Relationship between March Melt and Target-Control R-values</u>

Correlation coefficients were calculated between the following individual control station variables; their R-value with the April 1st target mean, elevation, and distance from target station 1; all listed in Table 1, and melt frequency during March, given in Table 3. For each set of 14 pairs, a correlation coefficient would have to exceed +/- 0.53 to be significant at the 0.05 level and +/- 0.66 at the 0.01 level. The only variable pair to meet either criterion was between March melt frequency and the target-control R-values, resulting in R = -0.84. That high value indicates their relationship was both strong and statistically significant. Thus, 70% of the variance in April 1st control station R-values with the target mean was explained simply by the frequency of snowmelt during March. In general, the more frequent March melting was at a given control station, the less its correlation with the target area. This is not a surprising result and indicates the need for caution in using control stations, individually poorly correlated with the target, for historical regression analysis.

5e. Effects of High Frequency Melt Control Stations

An attempt was made to better quantify the effects of using April 1st observations from control stations 5, 6 and 7. The percentage departure of all 41 individual winters was calculated from the target-control regression equation plotted in Fig. 1, the latter being essentially the same analysis reported by GS. These percentage departures were correlated against the mean difference, in inches, between April 1st and March 1st SWE at controls 5, 6 and 7. A relationship significant at the 0.01 percent level was found with an R-value of -0.49, explaining about 1/4 of the variance. While 3/4 of the variance in not explained by this relationship, it is an important result, based as it is on only 3 of the 7 control stations. Those 3 had some of the lowest correlations with target means as listed in Table 1.

Plotting on a large-scale topographic map revealed that stations 5 and 6 are on the broad Columbia Plateau with terrain uncharacteristic of the target area mountains. Station 7 is within a small mountain range rising from the plateau which likely explains its somewhat higher R-value. A figure similar to Fig. 1 was plotted (not shown) with the X-axis based on the mean of control stations 5, 6 and 7. The resulting R-value was only 0.771 with much scatter of individual season data pairs. Seeded seasons 1993-96 had departures of 18, -21, 50 and 16%, respectively. It will be recalled that the original analysis shows especially large positive departures near 17% for 1993 and 1995 but values less than half that for 1994 and 1996. These findings suggest that inclusion of high melt frequency stations 5, 6 and 7 contributed to the large positive departures reported by GS for 1993 and 1995 and to the 5% positive departure found for 1996. A number of nonseeded seasons with high positive departures in Fig. 1 also had high departures when only stations 5, 6 and 7 were used as controls.

Another regression (not shown) which used the remaining controls 1, 2, 3 and 4, resulted in a respectable R-value of 0.971. Three of the 4 seeded seasons had positive departures and one was negative but none were outside the general scatter of data points about the regression equation. That is, regression analysis based on the 4 of the original 7 controls nearest the target, and with the highest individual associations with the target, provided no evidence of seeding effectiveness.

It can be concluded that March melt at control stations 5, 6 and 7 had a significant effect on the regression analysis using April 1st observations. The impact was certainly substantial enough to confuse any attempt to detect SWE changes due to seeding.

6. COMPREHENSIVE ANALYSIS USING MARCH 1ST OBSERATIONS

Based on the above evidence, the authors concluded that April 1st SWE observations are generally unsuitable for historical regression analyses of winter orographic cloud seeding effectiveness in the climate of southern Idaho. The only possible exceptions might be the limited number of stations in Table 3 with very low March melt frequencies. Melting at locations with generally lower elevations and more exposure will reduce their April 1st SWE even while SWE values are increasing at higher elevation sites, especially when the latter have more local protection from winds and solar radiation. Attempts to adjust the April 1st data, or to severely limit the number of stations used, would create their own serious uncertainties.

It is physically more reasonable to base regression analyses on March 1st rather than April 1st measurements, especially if stations with frequent February melt are discarded. Admittedly, this approach reduces the fraction of the operational seeding season evaluated. But if the seeding method was truly effective, the "signal" should be in the March 1st observations at least as predominately as in observations a month later, complicated by the serious melt issue and possibly by a higher altitude of SLW cloud sufficiently cold for AgI to be effective.

It was decided to expand analysis of March 1st observations to all available control and target stations in Table 3, with one caveat: use of stations with frequent February snow melt would create some of uncertainties discussed in the previous section. Accordingly, a somewhat arbitrary rule was established to use only stations with no more than 5% February melt frequency. That resulted in discarding target stations 2 and 8 and controls 5, 6, 7 and 14. Table 1 shows that 3 of these 4 controls were among the lowest April 1st correlations with the target area, and that only one other control station (12) had an R-value lower than station 7.

The historical period was also expanded to 36 seasons by including post-seeded seasons 1998-2001. The 1997 season was not included because of seeding in the adjacent Payette River Basin. The intension was to conduct the most physically reasonable and comprehensive historical regression analysis possible with existing SWE observations.

Mean target area March 1st SWE was calculated for the 10 remaining stations of Table 3, which included "new" stations 11 and 12, all having February melt frequencies no higher than 5%. These means were used in regressions with groupings of the remaining 10 control stations listed in Table 3 which met the same criteria.

R-values and percentage departures from linear regression equations were calculated for every possible combination of control stations. In order to determine the highest R-value available from all 10 remaining control stations, the means of all possible groupings of 1, 2, 3 --- 8, 9 and 10 control stations were correlated against the target means. This comprehensive approach provided 1230 permutations of target-control relationships.

Table 4 shows the resulting highest R-values for each grouping by number of stations. Individual and mean departures are given as percentages for the 4 seeded winters. In addition, the bottom row provides values for the highest R-value (0.981) achieved without use of control station 1, most likely to be contaminated. The number of combinations (regressions) possible within each grouping by number of stations is listed.

The mean positive departures of Table 4 are seen to be in the range 1-3% with the exception of the grouping without control station 1 which was 3.7%. Negative departures are seen for seeded winter 1994 except for regression with a single control station. Groupings of 3, 4, 5 and 6 stations all provided maximum R-values greater than the GS analysis value of 0.98. The highest R-value is 0.983, and the other three groupings are 0.981-0.982, all indicating strong relationships. The grouping with the highest R-value explained almost 97% of the variance in target mean SWE accumulation on March 1st. The results of that case are shown in Fig. 2 where only seasons not included in the regression equation are labeled to reduce clutter. Figure 2 provides no apparent evidence of seeding effectiveness.

Table 4. Summary of all available combinations of 10 control stations against the mean of 10 target stations after discarding stations with February melt frequencies higher than 5%. March 1st SWE observations were used from 36 nonseeded seasons. Results are shown for the highest R-value in each grouping by number of stations. Percentage departures are given from individual regression equations for each seeded winter and for their mean. The bottom row lists results for the highest R-value attained without inclusion of control station 1.

No. of Stations	Combin-ations	Max. R -value	Control Stations	1993 (%)	1994 (%)	1995 (%)	1996 (%)	Mean (%)
1	10	0.972	1	4	1	7	-11	0.2
2	45	0.979	1 2	5	-1	3	-4	0.9
3	120	0.982	1 2 3	5	-1	5	5	3.3
4	210	0.982	1 2 3 9	3	-10	5	11	2.2
5	252	0.983	1 2 3 9 11	6	-6	4	7	2.8
6	252	0.981	1 2 3 9 11 13	5	-7	5	8	2.7
7	210	0.980	1 2 3 4 9 11 12	2	-3	2	5	1.6
8	120	0.979	1 2 3 4 9 11 12 13	2	-4	2	6	1.7
9	10	0.978	1 2 3 4 9 10 11 12 13	4	-4	5	6	2.6
10	1	0.976	1 2 3 4 8 9 10 11 12 13	2	-4	3	8	2.2
4		0.981	2 3 9 11	6	-7	4	12	3.7

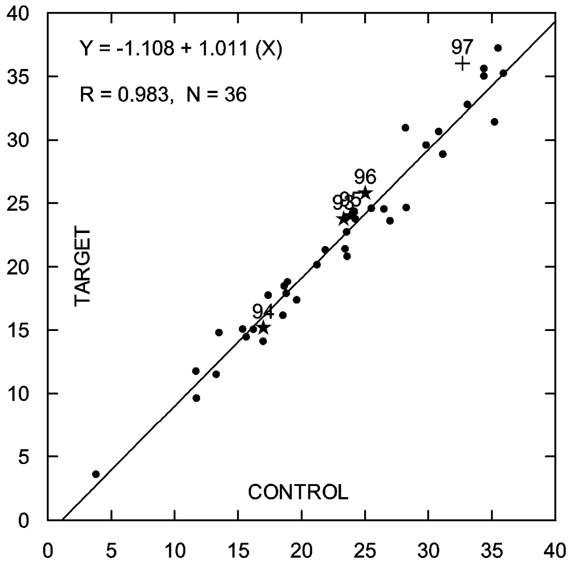

Figure 2. Linear Regression Plot for the Means of low-melt stations, 5 Control and 10 Target, using March 1st Snow Water Equivalent (inches).

The 1997 season, when seeding was conducted in the Payette Basin immediately north of the Boise Basin, has one of the largest positive departures in Fig 2. That result suggests 1997 seeding was not effective because control stations 1 and 3 were within the seeded basin and control station 2 was just north of it. Increased Payette Basin snowfall should have placed the 1997 data pair to the right of the regression line.

Rerandomization was used as previously discussed to estimate probabilities that percentage departures for the block of 4 seeded winters were statistically significant. The lowest P-value associated with Table 4 was 0.27 for the grouping of control stations 2, 3 9 and 11. All other P-values ranged between 0.31 and 0.49, the latter equivalent to flipping a coin. Clearly, the results of Table 4 and associated P-values provide no basis for considering the Boise Basin cloud seeding project to have been effective in snowfall augmentation up to March 1st. These results are totally at odds with the findings of GS who used April 1st observations.

7. FURTHER EXPLORATION WITH APRIL 1ST OBSERVATIONS

Previous discussion has demonstrated the futility of using April 1st SWE accumulations frequently influenced by snow melt to assess cloud seeding effectiveness in southern Idaho. But taken at face value, the results reported by GS were quite impressive. All 4 seeded winters had positive departures within the somewhat consistent range of about 5 to 17%, and averaged near 11.5%, a substantial seasonal SWE increase. These results were based on a strong target-control relationship which explained 96% of the variance.

In an attempt to better understand how such an apparently impressive but false suggestion of seeding effectiveness could result, it was decided to analyze the April 1st observations in more detail. The reader should understand that the analyses of this section are flawed by the melt problem. Therefore, any statistical suggestions should be discarded. The purpose of this section is not to evaluate seeding effectiveness but to help demonstrate pitfalls to be avoided in future applications of the historical regression method. It will be further demonstrated that all stations and procedures need to be selected prior to seeding project initiation, and should be *based on physically sound criteria*. Of course, this information must also be made public to enhance the credibility of later statistical suggestions.

Regression equations and seeded season departures were calculated for all possible permutations from the pool of 14 available control stations listed in Table 3. Using all combinations of 1, 2, 3 --- 12, 13 and 14 stations resulted in 16,383 separate regressions. Target means were based on the 10 original stations used by GS after preliminary analyses showed little impact from including the 2 additional stations added in this study (one had serious snow melt). The original 1961-92 historical period was again used. Both steps make these analyses more consistent with the initial GS analysis.

Table 5 lists the combination of stations which provided the highest R-value within each grouping by number of stations, provided that at least one was equal to or greater than 0.980, the value achieved by the GS analysis. R-values are listed to 4 decimal places to better demonstrate how little they varied. The number of individual regressions calculated is listed under "combinations" for each grouping of 3 through 10 stations, given that 14 were available. Percentage departures are listed for individual seeded seasons and the mean of the seeded seasons. The bottom row lists the same information for the single regression providing the highest R-value without inclusion of control 1, most likely to be contaminated. P-values were calculated for 4 season means as previously described, using Monte Carlo techniques, as listed in the right-most column.

Table 5. Summary of highest R-value statistics for indicated groupings by number of stations, using all 14 available control stations and means of the 10 original target stations with April 1st observations. The minimum acceptable R-value was 0.980 which was not achieved by use of fewer than 3 or more than 10 control stations. Percent departures from individual regression equations are given for the four seeded winters and for their mean. Control station numbers are listed for each regression. Listed P-values are for the four season means.

No. of Stations	Combin-ations	Max R-Value	Control Stations	1993 (%)	1994 (%)	1995 (%)	1996 (%)	Mean (%)	P-value
3	364	0.9830	1 2 9	3	-11	11	1	1.0	0.43
4	1001	0.9844	1 2 4 7	9	7	10	-1	6.5	0.08
5	2002	0.9856	1 2 4 5 7	11	10	13	-1	8.2	0.03
6	3003	0.9842	1 2 4 5 7 9	9	-1	14	-1	5.2	0.15
7	3432	0.9833	1 2 4 5 7 9 11	12	4	13	0	7.3	0.04
8	3003	0.9821	1 2 4 5 7 9 11 13	12	6	12	3	8.2	0.05
9	2002	0.9808	1 2 4 7 9 10 11 13 14	6	5	11	2	6.1	0.11
10	1001	0.9796	1 2 4 5 7 9 10 11 13 14	8	6	13	2	6.9	0.09
4		0.9816	2 4 7 9	7	-5	13	1	3.8	0.22

The grouping of all 14 controls yielded an R-value of 0.967, well below the 306 different combinations that equaled or exceeded 0.980. Table 5 shows that the highest R-value of 0.9856 was achieved with a grouping of 5 of the 7 original control stations. Five, but not all 7, of the original controls were "selected" for the grouping which provided the highest R-value with 7 stations. While the mean departure was 8.2% for the regression with the highest R-value, it is difficult to reconcile the -1% departure during 1996 with the positive departures of the 3 previous seeded winters. A P-value of 0.03 was calculated for this particular station grouping *which yielded the highest R-value of all 16,383 regressions*. But, even ignoring the melt problem and lack of randomization, it must be questioned whether that low P-value would mean much in the context of Table 5. For example, very slight R-value differences of 0.9844 and 0.9842 were associated with groupings of 4 and 6 control stations, respectively, compared with the highest value of 0.9856 for 5 control stations. The variance explained differs by less than 0.3% among these 3 highest R-values, yet the P-values for the 4 and 6 station groupings increase to 0.08 and 0.15, respectively. While two other groupings had P-values of 0.05 or lower, another two had values of 0.43 and 0.22. This fact also shows a lack of stability in the results since all four or those R-values are quite similar.

Original controls 1, 2 and 4, are included in most groupings of Table 5. This might be expected since controls 1, 2 and 4 were all highly correlated with the target, having individual R-values of 0.94 or higher (Table 1). Controls 9 and 11, added in this study, were also part of several groupings in Table 5, with individual R-values of 0.90 and 0.94, respectively. Yet poorly correlated station 5 with low R-value of 0.63, and station 7 with a mediocre R-value of 0.82, also show up in most groupings. The latter two stations had high melt frequencies. Including poorly associated control stations which marginally improve the mean group correlation is a suspect approach. It has been shown that frequent melt was an important factor in limiting individual R-values for stations 5 and 7. Inclusion of these two "melt problem" stations apparently had the effect of slightly increasing R-values while dramatically increasing seeded winter positive departures, thereby confusing the search for seeding effects. Super and Heimbach (1983) rejected control snow courses which individually had R-values less than 0.88 in their Bridger Range Experiment analyses.

Similar R-values resulted from each grouping of Table 5. There is probably no real significance in the minor differences since explained variance differs by only about 1%. Slight measurement errors could cause greater changes. Examination of the range of individual winter departures for the five groupings with highest R-values shows extremes from -11 to +14% with means from 1.0 to 8.2%. Admittedly, the next lowest individual departure was -1% which occurred four times. This variability among individual season departures reduces confidence in the mean results. This demonstrates the importance of publishing individual season results, to establish context, besides showing overall project results.

Of the 306 regression equations out of 16,383 permutations which resulted in R-values of at least 0.980, only 8 had mean departures of at least 11.5%. This criterion was used because it rounds to 12%, the mean seeding effect reported by GS. The highest mean among the 8 values was 12.7%. None of these cases produced the highest R-value in their station size grouping and, therefore, are not included in Table 5. One of these cases was, of course, the grouping of 7 original controls. Seven of these cases included original controls 1, 2, and 4 plus 5, 6 and 7 (in 5 runs), or at least 6 and 7. The only exception was a run with stations 1, 2, 5 and 11 which had a mean of 12.1%, but individual winter departures ranged widely from 1 to 19%. It is not surprising that various combinations of the same stations were involved in calculation of similar results.

To summarize, only 8 regressions out of the 16,383 possible combinations with the 14 available controls provided the following: (1) somewhat consistent positive departures for each seeded winter, (2) a mean of at least 12%, almost the highest observed, and (3) a high R-value of at least 0.980. Results presented by GS were from one of these 8 combinations, and the only such regression based on their 7 controls. It is possible that they made a fortuitous choice of control stations which provided a high R-value. Apparently, they were unaware of the seriousness of the March melt problem.

The issue of long-term stability of the target-control relationship was raised in Section 5c, where evidence was presented that melt was unusually frequent at some control stations during the past decade. Almost all SNOTEL stations used in this study were in operation by 1982 so records no longer had to be estimated from nearby snow courses. Accordingly, a regression was calculated with the original target and control stations and April 1st observations, but for a more recent "historical" period chosen as 1982-92 plus 1998-2001. These 15 seasons provided a remarkable R-value between target and control means of 0.988, the highest seen in this study. Departures of the seeded winters were: 1993 = 12.8%; 1994 = 2.2%; 1995 = 14.7%; 1996 = 2.5% with mean of 8.0%. One nonseeded season, 2001, had a departure larger than 1993 and 1995 at 15.6%. Three other nonseeded winters ranged from 10 to 11%. The 1994 and 1996 seeded seasons were well within the general scatter with several nonseeded seasons exhibiting larger positive departures. P-values calculated by rerandomization follow: 1993 = 0.10; 1994 = 0.29; 1995 = 0.06; 1996 = 0.28 and 4 season mean = 0.045. Thus, only one of the four individual seasons was near the 0.05 significance level, and the mean was just below it at 0.045. The apparent significance of the mean departure is likely due in part to the seeded winters presenting the only 4 consecutive positive departures in the data set. But only 2 of 4 seeded seasons had large positive departures, making the results difficult to reconcile with consistent seeding effectiveness. Moreover, the reader is reminded of the major caveat of Sec. 4b regarding P-values applied to nonrandomized projects.

In summary, the mean departure was reduced to 2/3 of the 12% value based on the longer historical period, and all individual seasons were also reduced, especially 1994 and 1996 where departures were about 1/3 and 1/2 of initial results. These findings, based on a shorter but likely more representative nonseeded period bracketing the seeded years, suggest serious lack of stability from application of this method. Moreover, overly optimistic P-values from rerandomization were not impressive.

8. STATION SELECTION FOR DIFFERENT RESULTS

Table 6 was developed to demonstrate that a wide range of results can be found when testing a large number of control stations combinations. A search was made to find the case with the *minimum* mean departure in each grouping up to 14 control stations which had R-values of at least 0.980; that is, within the 306 regressions discussed above. It will be recalled from Table 5 that groupings with less than 3 or more than 10 stations did not have R-values that high. It is seen that groupings with 3 through 7 stations have high R-values of 0.982 or 0.983, similar to values in Table 5. Admittedly, no run showed a negative mean departure. But it might be pondered whether the suggestions of seeding ineffectiveness in Table 6 are less valid than the limited positive suggestions of Table 5 or the original analysis reported by GS. The (unlisted) P-values for the means in Table 6 were unimpressive, ranging from 0.08 to 0.53. One could certainly make the argument that control station 14 should not be included in analyses, as it is poorly associated with the target area. Yet that station is repeatedly found in Table 6. But the same argument can be made against use of control station 5, and to a lesser extent station 7, which repeatedly appear in both Tables 5 and 6.

The major point of the preceding discussion and Table 6 is that one can "mine" the results from large numbers of permutations to either show evidence to support the hypothesis that seeding was effective, or to reject it. Of course, such selection of results is in total opposition to valid science. But the approach used illustrates one of the serious pitfalls in interpretation of claimed results from the historical regression method applied to cloud seeding. Unless *both* target and control station selection, and the

Table 6. Similar to Table 5 but chosen by the *lowest* mean departure within each grouping by number of stations, for R-values of 0.980 or greater. April 1st SWE observations were used. See text for details.

No. of Stations	R-value	Control Stations	1993 (%)	1994 (%)	1995 (%)	1996 (%)	Mean (%)
3	0.983	1 2 9	3	-11	11	1	1.0
4	0.983	1 2 4 14	1	10	3	-3	2.8
5	0.983	1 2 4 9 14	1	-2	6	-3	0.4
6	0.982	1 2 4 7 9 14	0	-3	9	-1	1.2
7	0.982	1 2 4 7 9 13 14	1	-1	9	2	2.9
8	0.980	1 2 4 5 7 9 10 14	4	0	14	-1	4.2
9	0.980	1 2 4 5 7 9 10 13 14	5	2	13	2	5.3
10	0.980	1 2 4 5 7 9 10 11 13 14	8	6	13	2	6.9

objective reasons for station selection, are made public *prior* to seeding operations, post-project analysis results should be viewed with considerable skepticism.

9. POSSIBLE REASONS FOR SEEDING INEFFECTIVENESS

This brief section was added after each reviewer suggested inclusion of some discussion about why revised analyses of the Idaho operational program failed to demonstrate increased snowfall. Because of the absence of supporting physical observations in Idaho, the following discussion is somewhat speculative. However, comprehensive physical and modeling evaluations of a similar Utah program were summarized by Super (1999). Anyone interested in why this type of operational seeding is unlikely to succeed is referred to that summary, but especially to the references therein to many publications by various authors resulting from several years of cooperative state-federal research in Utah. The fundamental problem was simply failure to routinely and adequately seed the clouds. Super (ibid.) stated that, "The main problem with the Utah operational program is that observed effective AgI ice nucleus concentrations are too low for significant snowfall enhancement from the mildly supercooled clouds. Much of the time, the entire SLW layer is too warm for effective seeding with AgI, which begins to nucleate ice crystals near -6˚C, but which is ineffective in concentrations observed until the SLW cloud is colder than about -9˚C." Low observed concentrations of effective ice nuclei and seeded ice crystals were related to frequent stable layer trapping of valley-released AgI. Moreover, when AgI was transported into the mountain-induced SLW zone, volume filling with AgI was often limited because of widely spaced generators, typically 10 mi apart.

It has long been recognized that failure to achieve transport and dispersion of adequate concentrations of effective AgI is the major problem of winter orographic cloud seeding (Rangno 1986). Reynolds et al. (1989) reported on 1,681 individual silver-in-snow samples from 3 long-term operational project targets plus the target of one special generator network. *Less than 15% of the samples indicated any silver greater than background*. Reynolds et al. (ibid.) concluded that, "These are disturbing results, even if one considers only scavenging, in that the AgI must not have passed over large regions of the target during precipitation events." Similar poor targeting results were reported for two large target areas by Warburton et al. (1995). While finding "seeding silver" in target snowfall does not prove seeding was effective, failure to find it certainly demonstrates serious problems with project design. At least seasonal silver-in-snow sampling should be considered a minimum physical evaluation for any winter operational project.

Too many winter operational projects have released AgI from a limited density of often low-elevation sites and simply assumed, without any physical verification, that all necessary subsequent links in the chain of physical processes proceeded as hoped for. Such links include transport and dispersion, loading of SLW cloud with sufficient AgI concentrations *effective* at prevailing temperatures, sufficient time/distance for snowflake growth, etc., with the anticipated end product of useful snowfall enhancement. Numerous observational and modeling studies exist in the scientific literature which indicate that such an approach is naïve at best.

10. SUMMARY AND RECOMMENDATIONS

Application of the historical regression method to a recent operational winter orographic cloud seeding was examined in detail. Analysis by Griffith and Solak (2002) indicated a mean seeding effect of 12% based on April 1st snow accumulation observations.

Reanalysis with the same target and control stations showed almost identical results and indicated essentially the same data set was used in the current analysis, a point worth verification. However, a serious problem was found with snow melt prior to April 1st at some stations, especially controls.

Frequent and significant melt before sampling complicates any search for seeding effects by introducing another important variable unrelated to seeding. The historical relationship is then based not only on snow accumulation but on a melt pattern which varies among stations and seasons in unknowable ways. The target-control relationship is no longer valid for seeding evaluation if melt is pronounced before snowpack sampling.

The original GS analysis methodology was repeated with March 1st rather than April 1st data. The result was a slightly higher R-value and reduction of the apparent mean seeding effect to 1/3 of the initial results. This finding is sufficient to raise serious concerns about the initial results unless one had reason to believe March had exceptionally high seeding potential.

March melt frequencies were approximated by subtracting March 1st from April 1st SWE observations, ignoring unknown March snowfalls. This conservative approach showed melt frequencies between 44 and 78% at several control and target stations. Three of the original 10 target stations, and 4 of the 7 controls had March melt during about 1/2 to 3/4 of winters. Several zero readings were found on April 1st at 2 of these controls. It was shown that about 70% of the variance in individual control station correlations with the April 1st target mean was explained by March snowmelt frequency. More frequent snowmelt at a control station lessened its association with target snowpack accumulation. The influence of three control stations with high melt frequencies was shown to have especially complicated initial analysis of seeding effectiveness.

An exhaustive search was made for the strongest relationships among all available March 1st target and control observations for all 36 nonseeded seasons. Stations were excluded if their February melt frequency exceeded 5%. This is considered to be the physically most reasonable approach with existing data. The results do not support the hypothesis that the four season seeding project increased snow accumulation up to March 1st.

Some further consideration was given to use of April 1st data using an additional 7 available control stations. None of the results can be valid, because of the melt problem. However, these analyses helped illuminate the range of possibilities from the large number (over 16,000) of available regressions with all possible groupings of 14 controls. Very few of these many permutations provided apparent seeding effects as high and consistent as the initial analysis even though they had equal or higher target-control associations. To make a point, a successful search was made for equally high relationships that suggested no "seeding effect." While this approach certainly violates the scientific method, it demonstrated that a wide range of possibilities could be found if one "mined" the available database. Selection of measurement stations and methodology prior to seeding eliminates this possibility.

Finally, the original analysis procedures were repeated using a shorter nonseeded historical period bracketing the seeded seasons. The period began in 1982 because almost all SNOTEL stations with pressure pillows became operational by then. It extended to 1992, just before seeding commenced, and included the 1998-2001 post-seeded period. These 15 nonseeded seasons had the highest target-control R-value seen in this study at 0.988. This analysis showed large reductions in the positive departures from the initial results for two of the seeded seasons and for the mean. These findings suggest serious lack of stability from application of this method.

It was concluded that this comprehensive evaluation of the four winter Idaho seeding project, applying the often used historical regression method, *provides no reason to conclude that seeding was effective in snowpack augmentation*. Problems with the initial analysis were discovered which should guide future attempts to use this method.

Recommendations for future applications of the historical regression method to winter orographic seeding programs include:

- Choose and publish all measurement stations and statistical approaches to be used before data becomes available from the first seeded winter. As pointed out decades ago, this practice will add considerable credibility to later analysis results because the possibility of human bias is removed.

- Choose stations on physically reasonable basis. Stations with significant and frequent snow melt prior to sampling are obviously unsuitable choices. Using samples taken as late as April 1st will be unsatisfactory in many mountain and foothill climates. A control station poorly correlated with the target is likely an inappropriate choice because there are physical reasons for the low degree of association. Adding such stations to the control area group because they slightly increase the target-control R-value is not justified. As a "rule of thumb," based on their experience, the authors would be hesitant to use stations with individual R-values below about 0.85 and certainly 0.80.

- Evaluations of operational seeding projects should point out that valid P-values cannot be calculated with any precision in the absence of randomization. Estimates may be made as in this paper, but only with the major caveats noted, and only for the purpose of attempting to determine whether any suggested results *might* have some validity. The more valid alternative is to present no estimation of seeding effectiveness because such estimates without P-values must call into question any interpretation of results. It is understood that sponsors expect some evaluation of whether seeding was successful. But they should be made to understand that evaluation of operational projects may provide suggestions but never scientific proof.

- Sponsors and project managers could gain credibility for their seeding programs by keeping evaluations independent from operations. Independent evaluation of seeding effectiveness by experts with no stake in the results is similar to use of independent auditors in commerce and government.

- Confidence in the results of a statistical analysis can be increased if AgI targeting is verified in some fashion. As a minimum, target seasonal snowpack samples should be collected for silver sampling, an economical approach. Plume tracing of AgI or a tracer gas provides more definitive information. Presentation of statistical results in the complete absence of physical support is a practice unlikely to enhance the scientific credibility of weather modification as discussed in the most recent American Meteorological Society Policy Statement on Weather Modification (AMS, 1998).

REFERENCES:

AMS, 1998: Policy Statement of Planned and Inadvertent Weather Modification, *Bull. Amer. Meteor. Society*, **79**, 2771-2772.

Court, A., 1960: *Jour. of the Irrigation and Drainage Division, Proc. Am. Soc. Civil Engineers*, **86,** No. IR1, 121.

Dennis, A. S., 1980: *Weather Modification by Cloud Seeding*. Academic Press, New York, NY, 267 pp.

Dennis, A.S., J.R. Miller, Jr., D.E. Cain and R.L. Schwaller, 1975: Evaluation by Monte Carlo Tests of Effects of Cloud Seeding on Growing Season Rainfall in North Dakota. *Jour. of Applied Meteorology*, **14**, 959-969.

Gabriel, K.R., 2000: Parallels between Statistical Issues in Medical and Meteorological Experimentation. *Jour. of Applied Meteorology*, **39**, 1822-1836.

Griffith, D. A. and M. E. Solak, 2002: Economic Feasibility Assessment of Winter Cloud Seeding in the Boise River Drainage, Idaho. *Jour. of Weather Modification*, **34**, 39-46.

Griffith, D.A., J.R. Thompson, D.A. Risch and M.E. Solak, 1997: An Update on a Winter Cloud Seeding Program in Utah. *Jour. of Weather Modification*, **29**, 95-99.

Huggins, A.W., 1997: *Payette River Basin Weather Modification Research Program. Part 1: Radiometric and Meteorological Analyses to Assess Cloud Seeding Potential.* Final Report to Idaho Power Company from the Desert Research Institute, Reno, 76 pp.

Rangno, A.L., 1986: How Good are our Conceptual Models of Orographic Cloud Seeding? *Precipitation Enhancement - A Scientific Challenge.* (R. Braham, Ed.) Meteor. Monograph, **21**, No. 43, Amer. Meteorological Society., 115-126.

Reynolds, D. W, J.H. Humphries and R.H. Stone, 1989: Evaluation of a 2-month Cooperative Ground-based Silver Iodide Seeding Program. *Jour. of Weather Modification*, **21**, 14-28.

Super, A.B., 1999: Summary of the NOAA/Utah Atmospheric Modification Program: 1990-1998. *Jour. of Weather Modification*, **31**, 51-75.

Super, A. B. and J. A. Heimbach, 1983: Evaluation of the Bridger Range Winter Cloud Seeding Experiment using Control Gages. *Jour. of Climate and Applied Meteorology*, **22,** 1989-2011.

Thom, H. C. S., 1957: *Final Report of the Advisory Committee on Weather Control*, Vol. II, U.S. Govt. Printing Office, Washington, D.C., pp. 5-25.

Warburton, J.A., R.H. Stone and B.L. Marler, 1995: How the Transport and Dispersion of AgI Aerosols May Affect Delectability of Seeding Effects by Statistical Methods. *Jour. of Applied Meteorology*, **34**, 1929-1941.