REPLY TO THE COMMENTS BY GRIFFITH *ET AL.* ON SILVERMAN'S PAPER ENTITLED "AN INDEPENDENT STATISTICAL EVALUATION OF THE VAIL OPERATIONAL CLOUD SEEDING PROGRAM"

Bernard A. Silverman*, PhD

Consulting Meteorologist, Centennial, CO, USA

Griffith *et al.* (2010) express a number of comments and concerns about Silverman's evaluation of the Vail operational cloud seeding program (Silverman, 2009a). This Reply addresses those comments and concerns in the order in which they were presented in the Comments by Griffith et al.

Griffith *et al.* state that NAWC readily agrees with Silverman's statement, "*Silverman (2007) showed that it is imperative to use as a control or controls,* ***to the extent that available data permits*** ***(emphasis added)****, the streamflow station or stations that yield the most precise results.*", but they go on to say, *"......however, examination of the correlations obtained in this study as provided in Table 4 in The Vail paper (Silverman 2009a) suggests this ideal was not obtained*". They point out that *"The r^2 values obtained were significantly lower than those previously obtained by Silverman in the analyses he has performed on long-term Sierra Nevada programs"*

I hasten to point out that, **to the extent that available data permitted**, I used the control with the highest correlation with each target. Of course I would have preferred to use a control or controls with a higher correlation but none were available. Nevertheless, the controls that I did use improved the precision (reduced the standard error of the estimate) of the evaluations considerably. I also hasten to point out that I did draw the reader's attention (on Page 12) to the fact that "*the target-control correlation coefficients for the Vail targets are substantially smaller than those found for the evaluation of the operational seeding programs in the watersheds of the Sierra Nevada Mountains"* The physical reasons why this is the case is a matter worthy of further investigation.

Griffith *et al.* are concerned that the point estimates of the seeding effect at some of the Vail sub-basins (particularly GBH and GPT) are higher than one would expect according to the Weather Modificaiton Association Capability Statement on Weather Modification (WMA, 2005). They speculate as to the cause by saying "*A couple of explanations for these results could be the short historical periods and low correlations for these two sub-basins (as discussed in the above) and/or possible channeling of seeding material during seeded storms favoring these areas".* They also point out that the Vail sub-basins are considerably smaller in area than the areas of the watersheds in most winter cloud seeding programs and suggest that "...... *it is highly unlikely that cloud seeding over a more typical sized winter cloud seeding target area could produce an average increase of 28%* (as estimated for GBH)*".*

I too was concerned that the point estimates of the seeding effect for some of the Vail sub-basins were much larger than that achieved in similar cloud seeding programs. Therefore, I checked and double-checked the data processing and evaluation calculations to make sure they were accurate. In addition, I followed a suggestion by one of reviewers of the paper and repeated the evaluation of GBH using seven different controls in addition to the primary one that was used (FRR) in order to make sure the result using FRR was not an anomaly. I presented these results in Table 4 of Silverman (2009a). Since the results of the evaluation using the other 7 controls were statistically comparable to the results obtained using FRR as the control, I concluded "*Therefore, it is reasonable to conclude that the estimates of the seeding effect using FRR as the control for all of the targets, as given in Table 3, are statistically credible*".

It is unlikely that the low correlations of GBH and GPT with FRR was the cause of the higher point estimates of the seeding effect than was found in similar cloud seeding programs. The primary effect of lower correlations is to decrease the precision of the estimate (increase its standard error of estimate). Lower correlations can result in an increase as well as decrease in the magnitude of a point estimate (see Table 4 of Silverman, 2009a). In any event, the decrease or increase is very much smaller than the difference between the calculated point estimates and what would expect according to the WMA Capability Statement on Weather Modification (WMA, 2005).

It is also unlikely that the relatively short historical periods for GBH and GPT had an appreciable effect on the point estimates of the seeding effect. The result produced by the bias-adjusted regression ratio method is much less sensitive to the length of the historical period than is the result produced by the

**Corresponding author*: Dr. Bernard A. Silverman, 7038 E. Peakview Place, Englewood, CO 80111; e-mail: silvermanb@aol.com

historical target/control regression analysis method. The target/control regression relationship in the historical target/control regression method is used to predict what would have occurred in the absence of seeding and is solely dependent on the data for the historical period. On the other hand, the target/control relationship in the bias-adjusted regression ratio method is used to reduce the standard error of the estimate and is based on the data for both the historical period and the usually much longer operational period.

Having established the statistical credibility of the result for GBH I speculated that "*The fact that the seeding effect changes rapidly over the very short distances between seeding targets suggests, as one possible explanation, that the silver iodide nuclei from the ground generators are channeled by the terrain into a focused plume, and not widely dispersed as intended"*. Finally, I agree with the speculation by Griffith et al that it is highly unlikely that cloud seeding over a more typical sized winter cloud seeding target area could produce an average increase as large as that for GBH (28%). I agree, not because it is not scientifically possible, but because it is logistically difficult to seed a more typical sized winter cloud seeding target area as efficiently as the small area of GBH was apparently seeded.

Griffith *et al.* promote the use of double mass plots in the evaluation of cloud seeding programs and illustrate its usefulness in the Vail evaluation. They apply it to the Vail sub-basin area of Upper Gore Creek (GUP) and show how the "breaks" in the plot suggest major changes in seeding effectiveness. They speculate that the double mass approach is more sensitive in suggesting differences in seeding effectiveness than Silverman's technique of plotting the "time evolution of seeding effect". They also illustrate the utility of the double mass plot in selecting target and control sites when using a historical target/control evaluation technique.

I agree that the double mass plot is a useful tool in a target/control evaluation of cloud seeding programs. It can provide useful qualitative information of the type illustrated in the Comment by Griffith *et al.*; however, it cannot provide accurate quantitative estimates of the seeding effect and its statistical characteristics. It should be applied within the limits of its capabilities. I do, however, question whether the double mass plot is more sensitive in suggesting meaningful differences in seeding effectiveness than Silverman's technique of plotting the "time evolution of seeding effect". In comparing the two types of plots, it should be recognized that the double mass plot reflects the year-to-year changes in seeding effectiveness while the time evolution plot reflects how the cumulative evaluation is affected by those year-to-year changes in seeding effectiveness. A potential change in seeding effectiveness or its physical cause cannot be very important if a "break" on the double mass plot signals a possible change in seeding effectiveness and that possible change in seeding effectiveness is not apparent on the time evolution plot. Consider the following example - A double mass plot and time evolution plot for the Pitman Creek (PIT) sub-basin of the San Joaquin operational cloud seeding program is given in Fig. 1 and Fig. 2, respectively. The result for PIT was chosen because ground seeding started in 1951 and was supplemented by aircraft seeding in 1975. Both of these events should show up on the double mass plot and the effect of adding the aircraft seeding should show up on the time evolution plot. Other major changes in seeding effectiveness should show up on both plots. I leave it to the reader to decide which of the two plots best reveals the meaningful differences in seeding effectiveness. In any event, I suggest that both plots should be used. All the tools in our arsenal of analysis techniques should be used to maximize the amount of information that can be obtained.

Griffith *et al.* allege that I make contradictory statements about the estimates of the seeding effect and their statistical meaning in Silverman, (2009a) and in Silverman's other statistical evaluations of long-term, non-randomized winter cloud seeding programs (Silverman, 2007, 2008, 2009b). They also allege that *"Silverman's analysis provides* ***"estimates"*** *of the results of cloud seeding but does not provide* ***"statistical proof"*** *of the significance of these estimates as strongly implied in this paper"*. Furthermore, they claim that my use of the term "Statistical Evaluation" in the titles of 2 of my 4 papers is misleading.

There are no contradictory statements in my presentation of the results. Simply stated, in each of my evaluation papers that they cite, I did the following: 1) I evaluated the operational cloud seeding program(s) using a statistical methodology (ratio statistics) that was empirically tailored (bias-adjusted regression ratio) to provide valid inferences for operational/historical comparisons (non-randomized data) [Note: for Silverman (2009b) I used the Monte Carlo permutation (re-randomization) method which is inherently valid], 2) I presented the resulting estimates of the seeding effect indicating which of the estimates of the seeding effect were statistically significant based on the statistical methodology that was used, and 3) I discussed the caveats associated with each set of results. I did not imply that the results provided **"statistical proof"** of the significance of the estimates of the seeding effect; rather, I concluded each paper with the statement "*From a rigorous statistical viewpoint, the suggested effects that are indicated must be confirmed through new, a priori, randomized experiments specifically designed to establish their validity"* Finally, given the fact that I conducted statistical analyses in my 4 evaluation papers, I fail to understand why anyone would think that the use of the term "Statistical Evaluation" in the titles of 2 of my 4 papers is misleading. A review of the literature will show that it is common practice to use the term "Statistical Evaluation" in the title of papers that describe cloud seeding programs that have been subjected to statistical analyses and evaluations.

Griffith *et al.* state "*NAWC believes that the standard historical target/control analysis technique, when applied correctly, is entirely adequate in providing* <u>*estimates*</u> *of potential increases due to cloud seeding from long-term, non-randomized programs. More sophisticated techniques, such as those proposed by Silverman (which is actually an adaptation of the historical target/control technique) add little to this standard approach since the data are non-randomized and statements on statistical significance and confidence intervals are therefore not valid.*'

The reluctance by Griffith*et al.* to accept a more reliable and more robust statistical methodology is inconsistent with the WMA's recommendation (Boe et al, 2004) that states *"We (WMA) recommend that evaluation techniques presently being applied to operational programs be independently reviewed and as necessary revised to reduce biases and increase statistical robustness to the extent possible. Recognizing that randomization is not considered to be a viable option for most operational seeding programs, we acknowledge there is much room for improvement in most present evaluations, many of which are presently done in-house".* The historical target/control regression analysis technique does not provide reliable estimates of potential increases due to cloud seeding from long-term, non-randomized programs. I refer the reader to Silverman (2007, 2010) for a thorough analysis of the deficiencies of the historical target/control regression analysis method. A summary of these deficiencies are as follows: 1) it is not robust to departures from the assumptions under which it was derived, 2) lack of robustness affects the reliability and accuracy of the estimates of the seeding effect that it produces, 3) it overestimates the effects of seeding, and 4) it tends to produce appreciably more significant results than it properly should. On the other hand, the ratio statistics analysis technique, that Gabriel (1999) derived for randomized data, is considerably more robust and produces estimates of seeding that are statistically comparable to those from re-randomization analysis through the application of an adjustment factor suggested by Gabriel and Petrondas (1983) to compensate for the bias introduced by using data from a non-randomized program (Silverman, 2007, 2010).

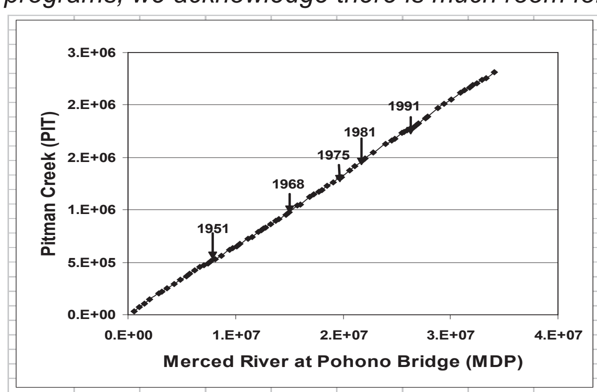

Figure 1. Double mass plot of the target Pitman Creek (PIT) against the control Merced River at Pohono Bridge (MDP) The arrows point to the water year when ground (1951) and aircraft (1975) seeding started, and water years when major changes in seeding effectiveness are evident on the time evolution plot (Figure 2).

The World Meteorological Organization (WMO, 2007) recommends "*Confidence intervals should be included in the statistical analyses to provide an estimate of the strength of the seeding effect so informed judgments can be made about its cost effectiveness and societal significance*". The point estimates of the seeding effect along with their statements of statistical significance and confidence

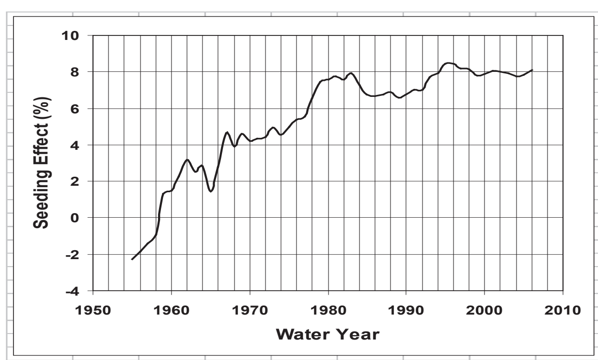

Figure 2. Cumulative year effect of seeding (time evolution of seeding effectiveness) plot for the Pitman Creek (PIT) sub-basin. The seeding is evaluated as a function of the cumulative number of years of seeding, i.e., initially the first 5 seeded years (1951-1955), then the first 6 seeded years (1951-1956), then the first 7 seeded years (1951-1957), ... , and finally all seeded years (1951-2006). The seeding effect calculated for each seeded water year is the value that would have been obtained if the evaluation were done for all seeded years up to and including that water year.

intervals should be provided; however, the limitations of the historical target/control regression technique should be recognized and its associated caveats should be acknowledged. It would be misleading to present a point estimate of the seeding effect without presenting a basis for establishing its statistical significance, i.e., its confidence interval and/or its P-value. This is especially true for the historical target/control regression method which produces point estimates whose reliability and accuracy are questionable. And, of course, statements of statistical significance and confidence intervals are entirely valid if they are determined through re-randomization and those from the bias-adjusted regression ratio are statistically comparable to those from re-randomization.

I disagree with the characterization of ratio statistics as "*an adaptation of the historical target/control technique*". Since ratios are widely used in the evaluation of weather modification experiments, Gabriel (1993) derived the randomization distributions of ratio statistics and the means and the standard errors of the asymptotic distributions of these ratios and their logarithms, distributions that are important to the correct application and interpretation of this type of statistics. In view of the above points, the bias-adjusted regression ratio is considerably more reliable and accurate than the historical target/control regression technique and, therefore, adds a lot more to the analysis. Even better yet would be to use re-randomization analysis, a statistical method of unquestioned validity.

Griffith *et al.* state *"......... because the Vail and similar California analyses are a posteriori and are applied to non-randomized projects, the analyses and their indicated results carry the same caveats as similar analyses conducted by others over the years*."

I most certainly agree with their statement; therefore, I discussed the caveats associated with the results in each of my evaluation papers (Silverman, 2007, 2008, 2009a, 2009b). Given their statement, I find it very interesting that Griffith *et al.* (2009) did not feel that they needed to mention the caveats associated with the results of their statistical evaluation of the Utah operational cloud seeding programs using the historical target/control regression analysis technique. Griffith *et al.* have a problem accepting the results produced by the ratio statistics method, a method that is based on sound statistical principles, but they have no problem with the results produced by the historical target/control regression analysis method, results that they accept without any caveats (Griffith *et al.* 2009). They implicitly accept without any qualifications the unsubstantiated assumption of the historical regression method that the target/control regression relationship derived for the historical period predicts with statistical certainty what would have occurred during the operational period in the absence of seeding.

REFERENCES

Boe, B., G. Bomar, W.R. Cotton, B.L. Marler, H.D. Orville, and J.A. Warburton, 2004: The Weather Modification Association's Response to the National Research Council's Report entitled "Critical Issues in Weather Modification Research" Report of a Review Panel. *J. Wea. Mod.*, **36**, 53-82.

Gabriel, K.B., 1999: Ratio statistics for randomized experiments in precipitation stimulation. *J. Appl. Met.*, **38**, 290-301.

Gabriel. K.R. and D. Petrondas, 1983: On using historical comparisons in evaluating cloud seeding operations. *J. Climate Appl. Meteor.*, **22**, 626-631.

Griffith, DA., M.E. Solak and D.P. Yorty, 2009: 30+ winter seasons of operational cloud seeding in Utah. *J. Wea. Mod.*, **41**, 23-37.

Griffith, D.A., D.P. Yorty and M.E. Solak, 2010: Comments on Silverman's paper published in the WMA 2009 Journal of Weather Modification entitled " An Independent Statistical Evaluation of the Vail Operational Cloud Seeding Program". *J. Wea. Mod.* , **42**, 148-154.

Silverman, B.A., 2007: On the use of ratio statistics for the evaluation of operational cloud seeding programs. *J. Wea. Mod.*, **39**, 50-60.

Silverman, B.A., 2008: A statistical evaluation of the Kern River Operational Cloud Seeding Program. *J. Wea. Mod.* , **40**, 7-16.

Silverman, B.A., 2009a: An independent statistical evaluation of the Vail Operational Cloud Seeding Program. *J. Wea. Mod.*, **41**, 7-14.

Silverman, B.A., 2009b: An evaluation of the San Joaquin Operational Cloud Seeding Program using Monte Carlo permutation statistics. *J. Wea. Mod.* , **41**, 15-22.

Silverman, B.A., 2010: Comments on the Griffith et al. report on operational cloud seeding programs in Utah. *J. Wea. Mod.*, **42**, 133-136.

Weather Modification Association, 2005: WMA Capabilities Statement on Weather Modification. WMA, Fresno, California.

WMO, 2007: WMO Weather Modification Statement and Guidelines. CAS-MG2/Doc4.4.1, Appendix C. (Available online at www.wmo.int/pages/prog/arep/wwrp/new/weathermod_new.html)