CORRESPONDENCE

Comments "On Using Historical Comparisons in Evaluating Cloud Seeding Operations"

WALLACE E. HOWELL

U.S. Department of the Interior, Bureau of Reclamation, Denver, CO 80225

20 October 1983 and 23 January 1984

ABSTRACT

Aspects of the finding, by simulation, that historical/operational target/control evaluations of weather modification projects are radical appear questionable. Their form is simulated but not their content. The radical tendency has already been discounted in reviewing bodies. Means are available to reduce the degree of radicalism. Attention to these possibilities may well lead to cost-effective new information from past operational data.

1. Introduction

The core of the finding by Gabriel and Petrondas (1983, hereinafter GP) is that the target/control statistics they calculated for historic periods and applied to simulated operating periods were radical. That is, they overestimated the significance of departures (in either direction) from the historical regression line. This would result in an overoptimistic indication of the significance (though not the magnitude) of a simulated seeding increase.

Before concluding that cloud-seeding evaluations using standard statistical procedures with historical/operational, target/control data have been speciously encouraging, one should seek answers to two questions. First, have GP fairly simulated the content of real evaluations as well as their form? Second, if their conclusions are correct, are the results of past evaluations to be held therefore less encouraging? One may also consider whether the difficulty they studied can be at least partly overcome.

2. The simulation of content

Gabriel and Petrondas note that cycles and trends disturb the stationarity assumed for meteorological time series in the usual statistics of significance. Target and control areas are subject, in general, to different trends, and these affect the statistics through the distance between target and control regions and through the time difference between historical and operational periods. The quality of the statistics is subject to further deterioration to the extent that the data fail to represent the climate faithfully and thus introduce a sort of noise. These things—the space and time separations and the representativeness of the data—qualify the evaluation content.

One of the best-documented sets of real historical/operational, target/control evaluations in the literature is that compiled by Prof. J. E. McDonald for the use of the Panel on Weather and Climate Modification of the National Academy of Science and referred to in Tables 1 and 2 of the Panel's report (1966). Details of most of these evaluations were communicated to this commentator at the time. Together with other published sources, they have been used in what follows.

The content of real evaluations suggests that evaluators sought a control area as near as possible to the target while giving reasonable consideration to climatic similarity and risk of contamination by seeding effects. Within the ground rules for use of traceable data, etc., they enlarged the data bases as much as possible and made them as representative as possible by including all acceptable stations within the respective areas. (Where in one evaluation the commercial operator had characterized a small target by the sole official station within it, McDonald included several others nearby.) Many times the control area embraced the target in a large C. Mean target/control separations were usually less than 55 km and never exceeded 150 km. Individual precipitation values were never averaged over periods longer than one month.

These selections of content tended to yield high coefficients of target/control correlation for the historical periods. In the real evaluations surveyed, the median of the correlation coefficients was 0.92, with very few less than 0.80 and none less than 0.73. It seems that real evaluations have not been made in the face of coefficients less than 0.80, except under unusual circumstances. Considerable efforts, usually successful, have been made to push them to 0.90 and above. One would then expect that the differential climatic trend between target and control was small.

The GP simulations breached these principles with respect to content, sometimes to an extreme degree. They characterized target and control each by a single station, compromising representativeness to an undetermined degree. They paired such different precipitation climates as those of Obir (a high Alpine station under strong orographic influence from Adriatic secondary lows in winter) and Vienna (a broad-valley middle-European location), and paired O'Kiep (a western continental desert location) with Port Elizabeth (a southeast maritime coast). By using annual averages, they may have caused further deterioration, possibly contaminating a whole year with the effects of a brief but strong vacillation.

The GP pairings separated by more than about 250 km must be suspected of unrepresentative differential climatic trends. Those more than 600 km apart are not simulations of real evaluations at all. The pairing of Darwin with Sydney, more than 3000 km away and differing in latitude more than Brownsville, Texas, from Moncton, New Brunswick, is an extreme case.

The GP simulation does conform to the content of real evaluations only respecting the time difference between middles of historical and operational periods. Their spread brackets nearly all real evaluations.

These differences in content, taken together, are enough to render quite inappropriate any quantitative extrapolation of the radical tendency of GP's simulation to the interpretation of real evaluations. In particular, the real-world relationship between this tendency and the correlation coefficient may be questioned. In the present exchange of comments, GP have provided correlation data (GP's reply, Table 1 and Figs. 1–5).

The correlation coefficients for GP's 31 station pairs range from 0.670 to −0.197, with a median of 0.274. None of them explains as much as half the variance and the majority explain less than a tenth. GP, then, have examined the radical tendency of a data body with a content quite useless for real evaluations. At best, their extrapolation can stand only if the differences in correlation coefficients are irrelevant. It is not enough to show that the data do not prove a connection to exist; it must be shown that a connection does not exist.

Since the source of the radical tendency is attributed to trend differences, one may start by asking whether GP's data show a difference between better- and worse-correlated data and between those that give the trends a shorter or longer time to work. From Figs. 1–5 of GP's response, 2×2 contingency tables were prepared that divide the data into those with better- or worse-than-median correlation coefficients and again divide them by probabilities higher or lower than the expected value of 0.05.

When the five-year-long groups are combined, the 2×2 contingency table become 52|23||46|26, which yields a one-tailed probability of occurrence of 0.3, using the standard 2×2 contingency test in the form

$$U = M^{0.5}(|A - B| - 1)[(A + B)(2M - A - B)]^{-0.5},$$

where A and B are the first and third members of the table and M is the sum of the first (or last) two members. This outcome is suggestive but quite inconclusive. However, if one asks how the shorter- or longer-year groups contribute to it, the following results are found. Combining the data for the 3- and 5-year pseudo experiments, the 2×2 contingency table becomes 16|14||18|12, with a one-tailed probability of 0.5, while the combined 10- and 20-year data yield a contingency table of 25|5||18|13 and a one-tailed probability of 0.046. This strongly suggests that the radical tendency decreases as the correlation coefficient increases and that this decrease is strongest when the trends have the longest time to work.

Thus it may not be easy to rule out a decrease in the radical tendency for the longer-running historical/operational periods as the correlation coefficients increase. The logic of working with data so poorly correlated as to be entirely useless for the evaluations being simulated is hard to fathom. It might be easier, and it would certainly be more convincing, to repeat the entire experiment using data fully representative of those used in real evaluations.

3. Discouraging or encouraging?

Throughout the past thirty or more years, the presentation and discussion of historical/operational, target/control evaluations have been accompanied by expressions of doubt concerning the validity of their indications. Among the grounds for concern, the questionable validity of applying to the operational period the statistics derived from the historical period has been prominent. These doubts were considered by the Panel on Weather and Climate Modification (1966) and later by the National Weather Modification Advisory Board (1978). Both, nevertheless, characterized the overall results of such evaluations as encouraging.

The work of GP points toward substituting quantitative estimates of correction to the significance levels derived from these evaluations for the qualitative disparagement to which they have so long been subjected. Extension of their work may quite possibly lead to wider appreciation of such evaluations. It would be desirable to have more adequate simulation of content and suitable re-evaluation of a representative sample of well-documented real operations, following GP's path. After all, the indicated seeding-related increases have already proved large enough to survive substantial depreciation and still be called encouraging. It is the vague, qualitative doubts that have robbed them of authority.

4. Onward and upward

If GP's work renews attention to historical/operational, target/control evaluations, it may also call attention to improvements that would reduce the radical tendency they have described, leading to further work to take advantage of improvements in content.

One improvement that can be applied to many past operations (and made part of the planning for future ones) is to use historical data both preceding and following the operational period (e.g., Howell, 1958). This enlarges the data set without increasing the time gap between historical and operational periods. The first-order term of the differential climatic trend then drops out. Operations open to re-evaluation by this approach number in the scores. Perhaps GP's data set, with little change, could achieve some simulation of the improvement to be expected.

Another possible improvement is a method that merges target and control into a single area. This is the method of specific raininess, which has been used in evaluations where areas adjoining the targets were for one reason or another unsuitable as controls (e.g., Lopez and Howell, 1965; Howell and Lopez, 1966). This method stems from Thom's (1958) suggestion that cloud seeding that changed the amount of precipitation should not much affect the number of precipitation events. It looks for a change in the average precipitation per precipitation event in the target from historical to operational periods. It thereby avoids all differential climatic drift except that operating on different parts of the target itself. Of course, a combination of any or all these approaches is possible.

In consideration of the rapidly escalating cost and often inconclusive outcomes of recent scientifically oriented precipitation-stimulation experiments, reprocessing the data from past operations by new approaches may be eminently cost-effective. GP's contribution in that direction may be prophetic beyond their first indications.

REFERENCES

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

Howell, W. E., 1958: A reappraisal of an early cloud seeding evaluation. *J. Meteor.*, **15**, 562–563.

——, and M. E. Lopez, 1966: Cloud seeding in southern Puerto Rico, April–July 1965. *J. Appl. Meteor.*, **5**, 692–696.

Lopez, M. E., and W. E. Howell, 1965: Cloud seeding at Medellin, Colombia, during the 1962–1964 dry seasons. *J. Appl. Meteor.*, **4**, 54–60.

National Academy of Sciences—National Research Council, 1966: *Weather and Climate Modification Problems and Prospects, Vol. II.* NAS-NRC Publ. 1350, 210 pp.

Thom, H. C. S., 1958: A statistical method of evaluating augmentation of precipitation by cloud seeding. Final Report of the Advisory Committee on Weather Control, Vol. II, Tech. Rep. No. 1, 5–24.

Weather Modification Advisory Board, Harlan Cleveland, Chairman, 1978: *The Management of Weather Resources.* U.S. Dept. of Commerce, Washington, DC, 229 pp. (See p. 46.)