By permission from *WEATHER MODIFICATION, SCIENCE AND PUBLIC POLICY*, edited by Robert G. Fleagle, University of Washington Press, Seattle, 1969.

Evaluation of Weather Modification Field Tests

JAMES E. MCDONALD

THIS SEMINAR on weather and climate modification is one more of many indexes of the current growth of interest in the problem of atmospheric modification. We observe many federal agencies initiating extensive studies of their future missions in this area; we see national advisory groups pondering scientific aspects; we note the broadening of points of view as to what constitutes "weather modification" and a corresponding broadening of attention to myriad side effects—legal, economic, social, and ecological. And we learn of Congressional determination to push this field along with financial support proposed at levels that, to some of us, seem to constitute almost unreasonable munificence when measured against the present primitive state of knowledge in this area or when divided by the scant number of qualified investigators available to wisely utilize such funds.

James E. McDonald is a professor in the Institute of Atmospheric Physics, University of Arizona.

The research described in this essay was carried on with the support of the Office of Naval Research.

Behind all this wave of interest lies a seemingly simple question: can we really modify any weather elements to a useful degree?

Dig into that question, weigh carefully its semantic aspects, examine the present state of the art, and one finds that, although that question is just the kind of straightforward query that a practical person should ask, its seeming simplicity actually conceals bewildering complexity. That complexity becomes still greater when one interprets the phrase "weather and climate modification" in the very broad sense intended in the reported deliberations of two national advisory groups that have recently weighed this and related questions.[1] For it has become the viewpoint of such influential groups that workers in the atmospheric sciences should and must begin to look ahead to the probability that our technologies will permit us to exert at least *perceptible* effects on a much broader range of atmospheric processes than mere precipitation stimulation, a range extending over more than a dozen orders of magnitude in length-scale and almost that great a range of time-scale.

Briefly, there appears to be in progress an almost explosive expansion of the area of serious concern with atmospheric modification possibilities. As that area of concern expands, the area of uncertainty as to the best methods of evaluating the efficacy of modification techniques will expand accordingly. At present we are still groping for really adequate answers about the efficacy of just one modification category, that of "rainmaking." When we turn, in coming years, to try to evaluate the efficacy of tornado modification, hurricane-steering, cyclone manipulation, and climatic modification, our evaluation difficulties will become far greater than they now seem. Or, at least, so it appears to me.

ATMOSPHERIC VARIABILITY—THE BUGABOO OF EVALUATION

Why the difficulty? Why is it likely that each new phase of the future development of weather and climate modification will uncover new and difficult evaluation problems? The answer is that our atmosphere is a physical system characterized by many degrees of freedom and exhibiting enormous variability. And it defies attempts at execution of the type of "controlled experiments" whose effectiveness we

[1] *Weather and Climate Modification: Problems and Prospects* (Washington, D.C.: National Academy of Sciences–Natural Resources Council Publication No. 1350, 1966), Vols. I and II. *Weather and Climate Modification* (Report of the Special Commission on Weather Modification; Washington, D.C.: National Science Foundation Report No. 66–3, 1966).

so admire in many of the physical and engineering sciences. With so broad a spectrum of natural variability, evaluation efforts will forever be plagued by the question so succinctly put in the famous *New Yorker* cartoon about rainmakers, which showed two robed clerics quizzically gazing out of a rain-spattered church window asking each other, "Is it theirs or ours?"

To ask whether it's "theirs or ours" is to ask whether an observed weather process occurring after a modification test is merely one that would have occurred in absence of the modification effort or whether it actually constitutes a significant departure from what was going to happen before we carried out our modification effort. To answer this question involves looking for an artificially stimulated and probably small change against a confusing background of natural fluctuations of wide amplitude. Since we cannot suppress those fluctuations in the classical manner of the laboratory physicist, nor predict them with the surety of the astronomer anticipating an eclipse, we must utilize a category of methods developed over the years by statisticians and by workers in other fields plagued by a high degree of variability in their experimental material.

I might turn next to quantitative documentation of the nature and extent of atmospheric variability, citing such illustrative points as coefficients of *seasonal* precipitation in arid regions that range up to values near unity, correlation half-distances often of the order of only tens of miles for *monthly* precipitation sums, hurricane frequencies varying by as much as 100 per cent between successive years, and the like. This would be pertinent, and would help to convey to those outside the atmospheric sciences the ticklish nature of our evaluation problem in weather modification. But, instead, I shall ask you to take that on faith or to document it yourself with your own day-to-day and year-to-year experience with the pesky unpredictability of "weather." The time that could appropriately be spent on illustrating our atmosphere's perverse variability, I shall, instead, devote now to another closely related problem, a problem that has puzzled me for years and still continues to puzzle me: why are there so many meteorologists in the antistatistics school?

THE ANTISTATISTICS VIEWPOINT

The modern era of rainmaking dates from the 1946 discoveries of the effects produced by dry ice and silver iodide when introduced into supercooled clouds. Twenty years later one can look back at one peculiar feature of the sequel to those discoveries: certain workers in the atmospheric sciences seem never to have learned, or admitted, that the exploitation of those discoveries raised *practical* questions that could only be answered (in view of the current state of meteorological knowledge) through use of *proper statistical decision-making techniques.* If, today, this were no longer a disputed point it would still warrant scrutiny as a lesson in the history of one science. But there is more than historical interest in this point, for it is clear that we still have with us many meteorologists who really do not fully appreciate this point, and who are still outspokenly insistent that "you'll never learn anything from statistics." The facts are that we know so little about crucial atmospheric processes that when Congress or water resources administrators or economists ask us whether we can usefully modify the weather, we are going to have to rely upon statistical methods for some years (decades?) to come, if we hope to come up with usable answers.

I like to classify antistatisticians within two groups. Type 1 not only wants no part of statistics himself, which is certainly understandable, but he goes much further and condemns and scorns those of his fellow workers who do use statistical techniques. Type 2 is the worker who is not devoting his career to illuminating details of the microphysics of atmospheric processes; he is out trying to modify those processes in field operations of one sort or another; and all the time he is claiming that "you don't need statistics to see the effect." He claims you can see it with your very eyes.

These two types bother and puzzle me. Type 1 blandly overlooks the *practical* questions that are insistently posed by legislators, governmental officials, and others who have to make policy decisions as to whether a given technique of weather modification "works." The decision makers cannot wait the twenty or thirty years it will take the atmospheric physicists to sort out the fascinating complexities of their field. To meet such practical questions, there must be *some* workers in the atmospheric sciences approaching the evaluation of field tests in the one way that will yield present-day answers: statistical hypothesis-testing. Type 2 antistatisticians not only scoff at statistics, but are capable of influencing decisions in high places.

Antistatisticians are not, in my opinion, doing right by science and the public economy when they let their enthusiasm exceed the bounds now set by available statistical evaluations of weather modification. If

I speak to a Congressman or to the public, and say I can "see" the effects of my cloud modification and urge that the reality of those effects is proved beyond doubt, while the statisticians looking over my shoulder are aghast at my neglect of the implications of natural atmospheric variability, I may indeed boost the flow of public funds going into my field. But the time may come when we realize that we should have paid more careful attention to some elementary probabilistic considerations bearing on weather modification.

Of course, on the other hand, the time may come when "farsightedness," or "enthusiastic efforts waged in the face of much professional foot-dragging" will be praised for ushering in a new era of beneficial weather and climate modification. One cannot be sure.

But one can be sure of this: enthusiasm should be based on facts. And baffling variability is a fact of the atmosphere's nature. Detection of anything other than a huge modulation of atmospheric processes requires the use of statistical techniques. For example, the kind of effects now mentioned as conceivable for precipitation stimulation (say 10 to 20 per cent increases or *decreases*), have to be recognized as being very hard to detect against natural variability that may be twice as large as the sought-for effect. With such signal-to-noise ratios, all of us need to be quite tentative, I fear, and quite statistically sophisticated as we slide around on the interface between the atmospheric sciences and public affairs.

It is exciting to see all this public interest in our subject; but to some of us the intensity of that interest seems better described as bordering on the disconcerting. The history of two decades of "enthusiasm" in support of the "threshold of economically competitive nuclear power," a history marked by increasingly disillusioned Congressional and public needling of nuclear power enthusiasts, might serve as an object lesson. But remember that in 1946 the basic scientific prospects for useful exploitation of nuclear energy were far better established than are the scientific prospects for useful exploitation of weather modification techniques in 1967. Now that nuclear power does seem to be coming into its own, the story of that development should be pondered by those of us concerned with the future development of weather modification. Perhaps that history offers hope; perhaps it offers warnings.

While physicists and chemists deal with controlled experiments in which the signal-to-noise ratio is large, atmospheric scientists, in company with other geophysicists, must constantly confront variability. For the time being we must look wistfully at sister fields where one can get all his answers by controlled experiments. If we want answers *now* relative to the practicability of weather modification, we must employ statistical evaluation methodologies in our field tests. In 1977 I suspect that a similar assertion will be in order; by 1987 perhaps we shall have moved on to such new and different types of modification that new and different types of natural atmospheric variability will be plaguing us in exactly the same fashion.

Let me repeat, lest my real point be misunderstood, that I am one of those who would personally prefer to forget many of these practicalities and push on with the many intriguing questions of the underlying physics of weather modification; and I am most emphatically one of those who feel that, in the long run, it will be physics and not statistics that will open up the truly new insights in this field. But, in the meantime, I do wish that Type 1 would stop scorning his fellow-meteorologists who are working at the practical questions of the moment; and I wish that Type 2 would face up more realistically to the harsh facts of atmospheric variability, and temper his enthusiasm with honest admission that it is always extremely difficult to decide, after a weather modification experiment, whether "it's theirs or ours."

EVOLUTION OF EVALUATION METHODS

It is understandable that the earliest of the methods used to evaluate the efficacy of seeding with dry ice and silver iodide was very simple: one seeded and then literally looked for subsequent precipitation. I would hope that even those who are least meteorologically informed will have their thoughts turned by this statement to the principle of the *post hoc* fallacy. Certainly many meteorologists quickly objected to such "evaluation" methods on grounds of the *post hoc* pitfall. But twenty years have not sufficed to drive those objections home in certain quarters, as I have just emphasized in the previous section of my remarks. The seed-and-look approach to assessing treatment efficacy is still found as the substance of papers presented in 1966 at meteorological conferences attended by Type 2 Antistatisticians; but fortunately there *has* been progress in other quarters, so let's push on to better things.

The first slight improvement on seed-and-look was to use as a measure of seeding efficacy the existence of a positive departure from long-term normal precipitation in the target area for the seeded period. This method was accepted by the water-desperate farmer or rancher for a few years around 1950, until critics reminded both the clients and the seeders that, even in absence of all seeding, one must expect positive departures from normal about half the time (more than half the time with the highly-skewed precipitation distributions typical of the very arid regions that sponsored many of the early seeding programs!).

The third generation of evaluation methods was not in wide use in commercial seeding projects before about 1952–53, but by that time a number of statisticians and climatologists had so effectively criticized the per-cent-normal method that the much better method of regression began to appear, which, although a half-century old in statistics, was nonetheless new to many of us. The method of regression is based on solution of a control area as near as possible to the target area without having it subject to contamination from the seeding activities. Often the control area is picked as an area in the direction upwind from the target relative to the prevailing winds; but, needless to say, this is not always possible, so sometimes the two areas lie more or less side by side, separated by a buffer zone. The best control area (assuming contamination has been ruled out) is one that exhibits the highest coefficient of correlation with the target area for the historic period of precipitation record available. One likes to have at least twenty years of record; but often record lengths of only a decade have had to be used. In some projects correlations as high as 0.98 have been encountered, but often values as low as 0.8 or less have had to be accepted, with attendant loss of discrimination. The importance of having high target-control correlation lies in the manner in which this implies a small standard error of estimate. The standard error of estimate is the yardstick with which statisticians estimate the likelihood that an observed departure from the historic regression line is only the play of chance. If the standard error of estimate is small, even quite tiny seeding effects can be detected as statistically significant.

The above comment on regression methods has been too brief to cover all important statistical subtleties, but it may convey the basic logic. It is, however, necessary to emphasize rather strongly one subtlety about which past criticisms and controversy have centered: the so-called "storm-types" error. In the early and mid-fifties it was objected, principally by Neyman and his associates,[2] that the user of simple regression methods laid himself open to serious error, either positive or negative, by virtue of the ever-present possibility that at or near the time separating the "historic period" from the "seeded period" a shift in the general circulation might lead to a shift in the predominant "storm-type" governing precipitation in both the target area and the nearby control area. If this did occur unnoticed by the evaluator (who in those days was generally the seeder), then it would be equivalent to a shift to some new regression line that might lie either above or below the historic regression line, depending on whether the new storm type favored target or control.

Although the "storm-types" error has a strong appeal to the instincts of most meteorologists, there has never been an extensive climatological analysis of its quantitative aspects. Herbert Thom[3] found little evidence for it in the West Coast projects he examined, whereas Neyman and associates reported evidence that suggested that the error might be important at times. Some of the large sums of money that appear to be headed in the weather modification direction ought to be spent on settling this point, if for no other reason than its basic climatological interest.

The reason that a certain degree of neglect of the storm-types error is excusable here is that, after dissecting that error and certain other categories of error in the regression method, most statisticians have for about a decade been insisting that simple regression methods of evaluation must be replaced by, or combined with, *randomized designs*. In a randomized cloud-seeding project, one proceeds basically in the following fashion: a *seedable* unit (day, hour, cloud, and so forth) is first selected according to some pre-arranged, fixed, objective scheme; then a coin is flipped (at least in principle), and if it comes up heads the unit is seeded, while if it comes up tails, the unit goes into the control population. The coin-flipping (randomization) step takes entirely out of the hands of the experimenter the bias-laden task of deciding which cases shall constitute his treated population and which cases his control population. It is such a simple device that the novice tends to miss its profound implications; the experienced statistician

[2] J. Neyman *et al., Weather Modification Operations in California* (California State Water Resources Board Bulletin No. 16, June, 1955).

[3] *Final Report of the Advisory Committee on Weather Control* (Washington, D.C.: Gov. Printing Office, 1958), Vols. I and II.

who has been called in to advise on a wide range of experiments knows how beautifully this device, at a single stroke, cuts the Gordian knot of both recognized and unsuspected biases inherent in experimentation. Statisticians love randomization with a fine passion, or at least most of them seem to. A bit of this passion has slowly rubbed off on meteorologists, so that today a majority of those who have responsibilities in the design of cloud modification experiments think in no other terms than randomized seeding trials.

There are so many possible experimental designs for seeding in randomized manner that even a listing would become tedious. Unfortunately there is no book or monograph in existence which adequately discusses experimental design; such a work is definitely needed in view of present mounting interest in weather modification.

It seems advisable, before closing this brief discussion of the history of modification evaluation methods, to note that a few statisticians warn that meteorologists may have been oversold on randomization. They counsel that randomization is no panacea. And the possibility must be kept in mind that hidden or otherwise subtle effects may sometimes trip up the user of randomized trials. For example, the "carry-over" effect of decreasing effectiveness of seeding may be attributed to the persistence of ice-nucleating powers of AgI particles or to other indirect effects. Nevertheless, I believe that the wide variability of all atmospheric processes plus the ever-present dangers of unconscious bias in experimentation speak strongly for incorporation of some form of randomization into all future modification designs—until arrival of that millenium when we have achieved perfect predictability for all weather processes.

SOME RECENT EVALUATIONS OF CLOUD MODIFICATION EFFICACY

As a result of suggestions made in 1957 by the President's Advisory Committee on Weather Control,[4] a randomized test of silver iodide seeding of orographic winter storm clouds in the vicinity of Santa Barbara was undertaken, but its implications were never entirely clear for reasons that have been discussed by Neyman and others.[5] Results of earlier randomized trials by several groups had given no indications

[4] *Ibid.*

[5] J. Neyman, E. L. Scott, and M. Vasilevskis, "Statistical Evaluation of the Santa Barbara Randomized Cloud Seeding Experiment," *Bulletin of the American Meteorological Society*, XLI (1960), 531.

of positive effects of much practical significance, nor had any significant negative results emerged. Subsequent to 1957, but prior to about 1963, published evaluation reports, with only few exceptions, suggested no marked seeding effects from the various types of AgI seeding employed. By contrast, a number of commercial cloud-seeders continued to claim positive effects of economic significance. Hence there existed a low-key dispute that was not clarified by definitive discussions in the literature. (The commercial seeders published almost none of their seeding analyses in the open scientific literature, so their claims went untested by the normal scientific procedures.) A check carried out by the Advisory Committee on Weather Control on selected commercial projects had tended to support claims about the efficacy of commercial seeding in certain West Coast winter orographic storms (10–15 per cent increases suggested by Thom's analyses of this category, as reported in the 1958 ACWC report), but bitter controversy had obscured those findings and no one did the research required to penetrate those controversial clouds. In retrospect, that failure to clear up the post-1957 controversies over the ACWC results ought to strike many of us as slightly embarrassing.

In late 1963 a new review of the present status and future outlook of the subject was undertaken by the Panel on Weather and Climate Modification, appointed by the Committee on Atmospheric Sciences of the National Academy of Sciences. Their concern was not with just the single question of precipitation stimulation, but with the long-term outlook for all plausible modes of atmospheric modification. Their preliminary report,[6] released in late 1964, sketched a generally pessimistic view of precipitation stimulation—even more pessimistic than that held by some meteorologists outside the ranks of the commercial operators. After further deliberations by the NAS panel, it was decided that some attempt should be made to examine the commercial reports despite their not being available in the scientific literature. This examination expanded into a quantitative check of the results of fourteen short-term projects in the eastern United States and four longer-term winter orographic snowpack seeding projects in the West. None of these was a randomized project, a significant shortcoming; but they extended over a sufficiently long total period of time and were carried on in a sufficient range of geographic areas for

[6] *Scientific Problems of Weather Modification* (Washington, D.C.: National Academy of Sciences–Natural Resources Council Publication No. 1236, 1964).

the panel to reach the conclusion after lengthy discussions that their positive indications could not reasonably be laid to the storm-types error. Independent reviews by meteorologists and statisticians at the Rand Corporation and at the U.S. Weather Bureau tended to support the same conclusion—namely, that these commercial operations gave promising, even if not firm, indication of positive effects. A discussion of the foregoing will be found in the final report of the NAS panel.[7]

The NAS panel also studied a number of other seeding tests and seeding programs whose results were just being reported. Although the evidence was conflicting, with several projects indicating no significant treatment effects, there were enough cases in which the evidence of positive effects argued a more optimistic view of cloud-seeding potentialities that the panel's final report had a distinctly different tone from its preliminary report of one year earlier. It is important to emphasize that the language in which this more optimistic view was expressed was *very* carefully framed, in the course of several sessions involving not only panel members but also members of other groups asked to offer a critique. The key statement in the panel's report was this: "There is increasing but still somewhat ambiguous statistical evidence that precipitation from some types of cloud and storm systems can be modestly increased or redistributed by seeding techniques. The implications are manifold and of immediate national concern."

That statement was followed by a number of carefully drafted qualifications emphasizing that a portion of the evidence on which the panel was basing its cautious optimism comprised nonrandomized commercial operations and warned that without randomization the possibility of misinference could not be fully excluded. The panel also warned that "the theoretical basis for seeding effects is still very crude because we still do not have an adequate understanding of the physical details of many basically important cloud processes." With these and other provisos, the panel hoped to preclude unduly enthusiastic inferences from being drawn while, at the same time, it hoped to persuade both scientist and government officials that a stronger push now seemed in order.

The actual response, especially in Congress, was rather overwhelming to some panel members, including myself, for the proposed levels

[7] *Weather and Climate Modification* (NAS-NRC Publication No. 1350), Vols. I and II.

of spending in the area of cloud seeding alone ran to many tens of millions of dollars per year in some of the bills. Speaking only for myself, I would have to emphasize that I cannot see how we can beneficially spend such funds in the *immediate* future, nor do I feel that the panel findings warrant such expenditures. A principal difficulty, in my opinion, is that we lack sufficient trained manpower to spend that kind of money on cloud seeding without waste and boondoggling. On this point there was some division of opinion within the NAS panel; some argued that engineers, chemists, physicists, and so forth, would quickly be drawn into the field of weather modification in the same way that such persons were drawn into the space program once large funds became available. My own dimmer view is that there are always people ready to spend such funds, but I would hate to see us duplicate the wastefulness that has marked some of the space-spending of recent years.

CLOSING REMARKS

I wish to be counted as one who is enthusiastically in favor of federal spending on good, sound programs designed to assure benefit to the general public. But I also wish to be counted as one who deplores wasteful federal spending, and I wish to be counted as a scientist who fears that over-fat budgets will, in the long run, destroy some of the highest values of science. If we do not avoid any and all misrepresentations of the prospects and potentialities of science and technology, when called upon by legislators or administrators to assess the state of an art, then sooner or later the collective respect we once deserved will vanish. We scientists may then be viewed as one more special-interest group trying to elbow into a favorable position at the public trough. For these reasons, I have been somewhat alarmed at the official response to the report of the NAS panel. I view the prospects for useful weather modification as interesting, and I think they warrant increased attention and federal support—but at a level almost one order of magnitude below what has recently been proposed. If we do show that we can profitably spend more, then let us all go out and try to get more. But an explosive rise in expenditures, a step-function jump in support for weather modification, does not seem to me to be warranted by available scientific information.

Not all of my professional colleagues share my concern on this point. Some say you have to get "big money," and you have to expect an uncomfortable portion of it to be wastefully spent, if you want to push up to an entirely new level of activity in weather modification. I suppose I would have to reluctantly agree if I were only more strongly convinced that we had grounds for telling Congress that we will deliver truly useful modification techniques once we get the money. That kind of conviction is not, in my view, discernible in the "cautious optimism" carefully spelled out in the final report of the NAS Panel on Weather and Climate Modification.

But possibly that is just fuddy-duddy conservatism and lack of imagination and enthusiasm. Such lack of imagination has held up technologic progress in the past more than once. Because I have been concerned about this problem of how one sets about deliberately stimulating a scientific field, I recently read A. C. Clarke's free-wheeling discussion of the shape of future technology, *Profiles of the Future*.[8] This changed some of my notions about the long-term future of weather and climate modification; not because of any meteorological insights presented by Clarke, but because of cogent general analysis of how technologies unfold. It is a book that might seem frothy in a science-fiction sense to some; but I commend it as collateral reading for this series of papers. It made me lose a little of my prior conservatism but (as I hope I have made clear) not all of that conservatism, relative to weather modification potentialities.

We need much more good work directed toward evaluation of weather modification field tests. We should all do everything in our power to see that substantial portions of the federal funds that seem to be heading into the area of weather modification are spent on pinning down more conclusively the actual potential for useful modification of atmospheric processes. We should, I believe, work to encourage international studies of climate modification, studies that are going to require years, if not decades, even to block out in general terms. And we should work to insure that it will be the general welfare of all mankind that benefits from scientific research in this area, and that weather modification is not allowed to be exploited for military purposes or for special interests. If associated with the broad goals of society, research and field trials in all aspects of weather and climate modification should prove exciting and worthwhile.

[8] A. C. Clarke, *Profiles of the Future* (paperback ed.; New York: Bantam Books, 1964).