"REVIEWED"

STATISTICAL TECHNIQUES AND KEY ISSUES FOR THE EVALUATION OF OPERATIONAL WEATHER MODIFICATION

Chin-Fei Hsu, Water Survey Div., Ill. Inst. of Natural Resources, Urbana, IL 61801
K. Ruben Gabriel, Dept. of Statistics, Univ. of Rochester, Rochester, NY 14627
Stanley A. Changnon, Jr., Water Survey Div., Ill. Inst. of Natural Resources, Urbana, IL 61801

Abstract. A major factor calling for evaluation of operational cloud seeding projects concerns the growth of its usage. A number of statistical techniques were selected and tested on simulated weather changes in western Kansas, and it was found that principal component regression is consistently one of the more powerful evaluation techniques. Key issues affecting statistical evaluations were identified, they were better comparison, uniform definition of sampling unit, uniform measurement of response, choice of statistical technique, bias-consciousness, starting and stopping effects, and validity of historical comparison.

1. INTRODUCTION

A feature of weather modification in the United States is the existence of operational, or commercial, weather modification projects. Operational projects are generally based on a premise that weather modification works and no proof is needed. Whether these projects can or should be evaluated to gain proof of their efficacy, or to gather scientific knowledge, has been debated for more than 25 years. Certainly, their evaluation has not always been addressed properly or believably. Evaluation of operational weather modification efforts is sufficiently difficult to require a sizeable scientific effort. Not only is atmospheric research needed, but statistical expertise must be involved in the evaluation issue (Changnon *et al.*, 1979).

The key stakeholders in weather modification — the scientists, the weather modification industry, and the user public in the United States — have all come to realize that this issue must be faced. A major assessment of the national program in weather modification completed in 1978 recommended that considerable attention be given to the evaluation of selected future operational projects (Weather Modification Advisory Board, 1978 and Statistical Task Force, 1978).

Well-designed seeding projects offer unique opportunities for learning about the effects of cloud seeding, for testing concepts developed in exploratory research, for transferring technology to user groups, and for encouraging sound weather modification practices. Carefully controlled and randomized projects could help serve as "confirmatory type" experiments, as distinct from exploratory type experimentation.

Such evaluations face a variety of challenges, including: 1) the development of statistical-physcial techniques for evaluation, 2) the evaluation of operational projects to test the techniques, and 3) the planning and control of future operational projects to make them easier to evaluate. A major factor calling for evaluation of operational weather modification concerns the growth of its usage. Figure 1 reveals the extent and number of operational projects in recent years. There were 88 commercial projects in the United States in 1977, with weather modification being applied over 676,000 km^2, 7% of the total area of the United States. Techniques and procedures to evaluate operational projects are being developed, partially as a result of this growing usage.

We have been studying approaches for evaluating operational weather modification projects. The extensive asseesment of inadvertent precipitation modification induced by the city of St. Louis (Changnon *et al.*, 1977) has helped illustrate how evaluations of nonrandomized influences can be achieved. Our primary objective now is to develop statistical-physical evaluation techniques for operational projects, including both the usual nonrandomized operations and those of the future which might employ some degree of randomization ("piggyback type," as recommended by the Weather Modification Advisory Board, 1978). Weather data have been collected for several areas where studies of projects are likely (either simulations or actual operational projects), and potentially useful statistical techniques have been determined (Hsu, 1979a, b; Gabriel, 1979). The statistical techniques are being applied to simulated convective precipitation changes in three areas (Fig. 1) Kansas — warm season rainfall; Montana — seasonal hail; and Illinois — summer rainfall. Because of the experience derived in our research, at Illinois and Rochester, we have also been involved in the statistical assessment of the North Dakota and Utah programs, a part of a NOAA funded, CSU managed effort to address past and future evaluation of these two state projects.

The study of predictor variables (covariates) for simulated project areas is also being pursued (Achtemeier and Schickedanz, 1980; Achtemeier and Westcott, 1979; Westcott, 1979). A third project objective of our research is to establish operational criteria for future operational projects so as to allow more meaningful scientific evaluations. We also aim to assess past problems in

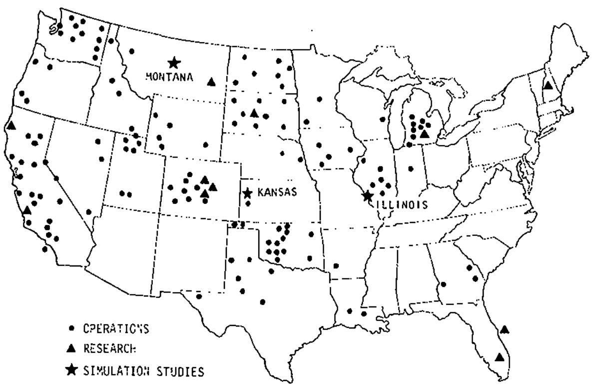

Figure 1. Locations of research and operational weather modification projects in the United States, 1973-1977. Two operational projects were in South-Central Alaska and no projects in Hawaii (from WMAB, 1978a).

evaluations of operational projects and to use these and the project research to set forth recommendations as to the performance of evaluations.

Once we have tested some 20 techniques on simulated weather changes in Montana, Kansas, and Illinois, we will apply the better techniques against a series of past operational projects.

This paper focuses on available results from the simulation testing of techniques and on related issues that introduce bias that affect meaningful statistical assessments.

2. STATISTICAL TECHNIQUES

A number of statistical techniques were selected to undergo Monte Carlo investigation. Techniques for comparing a seeded sample with an unseeded sample on target area data and data for several control areas were considered. They were in the forms of double ratio, target regressions on controls, target regression on first principal component of controls, two target regressions (one to the seeded sample and one to the unseeded sample) and a number of non-parametric rank power statistics.

In the study of 35 years of Kansas rainfall data (see Table 1), the Monte Carol investigation was carried out by several hundred random choices of 5 "seeded" years with the other 30 years designated as "unseeded." The results obtained from this approach can be applied to *both* randomized seed/no-seed comparisons and to operational/historical comparisons under the following two premises: (1) the operational years are to be thought of as outcome of natural randomization, and (2) randomization tests are to be performed on the actual data in order for the findings from the simulation testing to be validly applied.

For multiple regressions and principal component regression, the test statistics used were:

D = average of differences between observed and predicted seeded values;
W = positive rank sum calculated from differences.

Table 1

DATA AND EVALUATION ELEMENTS USED IN THE MONTE CARLO STUDIES

EVALUATION ELEMENT	KANSAS	MONTANA	ILLINOIS
Type	Rainfall Enhancement	Hail Suppression	Rainfall Enhancement
Data	County Rain May-Sept 1936-1970	County Loss-Cost 1948-1976	Areal Rain June-August 1971-1975
Unit	Month	Year	Storm/48-hr
Design*	TC-CH	TC-CH	TCM-R/T-R/TC-CH
No. Seed/ Unseeded	5/30 ----	3/26 6/23	24/122 26/132
Covariate	No	No	Yes/No
Target (sq mi)	750/1500	2000/6000/10,000	300/2000
No. of Targets	1/2	1/2/3	1
Control(sq mi)	800/2000/8000	2000/7000/25,000	300/700/1500/0
No. of Controls	1/3/8	1/3/13	1/3/5/0
Effect	Constant Multiplicative	Constant Multiplicative	Constant/Varying Multiplicative
Run:			
Null	500	1000	500
Each Altern.	100	1000	500

*Acronyms for design: TC-CH = continuous historical target-control design
TCM-R = random moving target-control design
T-R = random target-only design

Double ratio was computed as follows:

DR = $T_S C_{NS}/T_{NS} C_S$, where $T_S(C_S)$ denotes the average of seeded values in the target (control); T_{NS} and C_{NS} are similarly defined for non-seeded values.

For two regressions, several statistics were used:

T_1 = test statistic for slope parameter (Bernier, 1965);
T_2 = test statistic for intercept with same slope (Mielke *et al.*, 1977);
T_3 = test statistic for intercept and slope (Bernier, 1965);
TP = non-parametric statistic for slope (Potthoff, 1974);
VCOS = non-parametric variance ratio similar to F-statistic (Quade, 1967);
VBRS = Bross test (Bross, 1964).

For sum of rank power test, the test statistics used were:

$A_r = \Sigma\ (R_i)^r$, for $r = 1, 2, 3$;
$B_r = \Sigma\ |D_i|^r$, for $r = 1, 3$;
$C_r = \Sigma\ \mathrm{SIGN}(D_i)|D_i|^r$, for $r = 2, 3$;

where summation is over the seeded sample; R_i is the rank of the 1-th seeded target-control ratio; $D_i=R_i-(N+1)/2$, with N the total number of observations; and SIGN(a)=1,0, or -1, according to whether a is >0, =0, or <0.

3. MONTE CARLO STUDIES

Test statistics were computed to form the null and alternative distributions for positive 10%, 20%, 30%, and 40% seeding effects. Power curves were then derived by comparing the null and the alternative distributions. Table 2 shows powers of several statistics at the 5% level using average target and control values. Test statistics were ranked by their powers at the 5% nominal significance level for each seeding effect imposed.

"REVIEWED"

Table 2

POWER (IN %) OF SELECTED STATISTICAL TECHNIQUES AT 5% NOMINAL SIGNIFICANCE LEVEL, AVERAGED TARGET, WEST-CENTRAL KANSAS, 100: 500 SAMPLES

		DR	MR		PCR		2 REG		SRP	
	δ		W	D	W	D	T_2	T_3	ANR(1)	ANR(2)
May	1.1	16	17	20	17	17	17	18	11	11
	1.2	44	26	33	30	40	37	42	23	26
	1.3	66	36	53	43	65	65	66	42	43
	1.4	85	47	71	55	86	84	86	56	60
June	1.1	22	20	23	17	18	18	18	10	10
	1.2	52	46	38	53	34	48	39	35	26
	1.3	81	66	54	73	65	74	69	53	48
	1.4	94	76	75	93	82	91	90	78	74
July	1.1	17	20	13	35	16	13	14	17	16
	1.2	38	37	24	65	35	37	39	39	27
	1.3	64	53	46	82	60	64	63	66	58
	1.4	79	63	70	96	80	80	80	81	73
Aug	1.1	23	25	24	18	22	16	18	22	18
	1.2	51	40	54	28	56	46	48	47	48
	1.3	71	55	74	39	78	71	73	62	64
	1.4	88	61	83	46	89	83	85	80	85
Sept	1.1	16	13	22	16	29	27	29	15	17
	1.2	42	20	44	27	52	51	51	38	36
	1.3	71	33	60	43	72	70	71	56	59
	1.4	82	49	71	57	82	76	79	72	78
Season Average	1.1	42	31	45	34	52	50	50	33	31
	1.2	89	64	83	80	87	83	85	81	85
	1.3	98	88	99	91	98	98	98	97	98
	1.4	100	100	100	100	100	100	100	100	99

Table 3 summarizes the findings obtained in the simulation study of west central Kansas, usign 2 center counties as targets; 8 surrounding counties as the control. Findings indicate that, overall, the principal component regression is uniformly one of the most powerful among the statistical techniques investigated. Target-wise, PCR is one of the most powerful techniques for the target average, and for target 10; whereas it is the most powerful technique for target 9. In most cases, PCR is one of the most powerful techniques. An interesting finding is the case of target average in seasonal average, in which most techniques work equally well. The variation of rainfall becomes smaller and the seeding effect tends to be more easily detectable when the average is used.

Table 3

HIGH POWER STATISTICS FOR SIMULATED RAIN MODIFICATION PROJECT IN WEST-CENTRAL KANSAS

MONTH	TARGET #1	TARGET #2	AVERAGE
May	PCR,SRP	PCR,DR	DR,PCR, 2 REG
June	PCR	PCR	DR,PCR
July	DR,MR,PCR	SRP	PCR
Aug	PCR, 2 REG	PCR	PCR
Sept	PCR,SRP	PCR,MR	PCR, 2 REG
Season Average	PCR	MR, 2 REG	DR,PCR, SRP,MR, 2 REG
Target-wise	PCR,SRP	PCR,SRP, MR	PCR,DR, 2 REG

Studies using Montana hail data and Illinois raingages data are currently being pursued at the Illinois State Water Survey (see Table 1). They cover a broad range of project types, evaluation designs, and areal sizes. Both constant and varying models of seeding-induced effects are used (Huff and Changnon, 1972). In addition, covariates are used in the Illinois study (Achtemeier and Westcott, 1979).

4. KEY ISSUES AFFECTING STATISTICAL EVALUATIONS

There has been much controversy over the validity of evaluations of operational cloud seeding. Thom's (1957) extensive analyses were subjected to scathing criticism by several statisticians (Brownlee, 1960; Neyman and Scott, 1961), who drew attention to a variety of biases which could affect such evaluations. The opinion seemed to be that reliable evidence could be obtained only from randomized experiments. And yet, 20-odd years later, the evidence from such experiments is still inconclusive and one is loath to ignore altogether the large amount of data available from many commercial or state seeding projects. It therefore is appropriate to offer some principles for such evaluations and to consider the problems encountered with the use of these principles.

The fundamental principle of any evaluation of seeding effect is that of *comparison*: precipitation under seeding must be compared with precipitation without seeding - *other things being equal*. The rub is in the last phrase - how do we ensure that other things are at least close to equal?

Comparison of precipitation in a seeded target area with that in an unseeded area is inadequate. No two areas exist which always have equal natural precipitation. Nor can one reduce variability by increasing sample sizes: It is quite inconceivable that a large sample of seeded areas be available for comparison with another large sample of unseeded areas.

Viable comparisons must be made of *occasions*, be they storms, days, 12-hour periods, weeks, or entire seasons. Of course, such occasions differ from one another even more than areas do, but it is quite feasible to reduce that element of variation by taking sizeable samples of occasions for seeding and for control.

In an experiment, the samples of seeded and control occasions are obtained by randomly allocating each occasion to be seeded or to serve as a control. In an operational project, the seeded occasions are chosen by the operator and/or his client and the *only comparison possible* is with similar occasions prior to the operation, to the extent that historical records of similar occasions are available, i.e., of occasions that would have been chosen for seeding by the operator/client's criteria.

Valid comparisons with "other things being equal" can be obtained only if one avoids certain obvious biases, advertent or inadvertent. We list some precepts that must be adhered to.

Occasions must be *defined in a uniform way*, independently of whether they are seeded or control. Selection biases can result when the operator defines the occasions that he will seed and the control occasions that will be used for comparison (Brownlee, 1960, discusses this point eloquently). To avoid such biases, some experiments are designed so that neither seeder nor observer can tell which occasions are seeded (High Plains Cooperative Program, 1979). Historical comparisons can avoid selection biases only if the records allow retrospective definition of occasions in a manner equivalent to the definition used during operations. This excludes practically all short term occasions defined on synoptic grounds by the operators and forces one to rely on chronological definitions such as all days or entire months or seasons. Thus, for example, the evaluation of Santa Clara County seeding was done in units of entire seasons -- December through March -- irrespective of the actual amount or timing of seeding (Dennis and Kriege, 1966).

Measurements of responses must also be uniform over all occasions, seeded and unseeded. In an experiment one may arrange for observers to be blind ot the seeding and ensure unbiasedness. But historical comparisons must be restricted to standard instrumentation that was available and unchanged during historical and operational years. This may be easier to achieve with precipitation records than with radar data or hail damage reports, but it always has to be checked carefully.

Subtle biases can creep *into the choice of the particular measurements* e.g., raingages in the target and in a nearby area to be used as a covariate. These choices should obviously not be influenced by the data observed on seeded and control occasions. In an experiment this can be done by making these choices before the experiment is run. In operational/historical comparisons this is only partly possible. Some such biases might be avoided by choosing before the operations begin (if the operators have the foresight to plan for evaluation at that stage - see Dennis and Kriege, 1966). But some biases of this kind are inevitable with such comparisons - though perhaps minor in importance - since the historical data will necessarily be available at the time choices of measurements are made (Gabriel, 1980).

The *method of statistical analysis* should similarly be decided on as far as possible prior to the occasions. Adapting a technique of analysis to the data at hand is an obvious source of bias, as is concentration on the analyses of special phenomena observed during the seeded and unseeded occasions (Court, 1980). This affects the multiplicity of ex-post analyses of experimental data (e.g., the several breakdowns of Whitetop data) as much as the ex-post decisions on which hail suppression years had been genuinely operational and which had consisted of mere "tooling up" (Peterson, 1975).

The only way to avoid this *multiplicity* bias completely is to adhere rigorously to a *preordained protocol* of measurement and evaluation. However, the bias is likely to be much smaller when a very few principal tests are carried out than when an army of students spend years in sifting an experiment's data for possible clues.

Cloud seeding operations are often initiated as a result of several years of drought or bad hail loss. Inclusion of these years among the unseeded occasions would clearly bias comparisons. One would do well, therefore, to exclude the few years preceding operations. No such "starting biases" can occur with randomized experiments.

Both operations and experiments tend to be *discontinued* if their initial seasons do not show promising results. Also, they may well be *terminated and evaluated* when apparently sufficient positive evidence of effects has accumulated. With experimentation, one might hope that there would be some attempt to plan the length beforehand, but funding needs and the wish to publicize findings may violate well-laid plans. As to operations, it is not even conceivable that there should be binding a priori decisions to continue them for exactly so many years, no more and no less. It is difficult to see what one can do to control either of these possible sources of bias, that of discontinuation and that of termination. One can only hope they are minor, both for experiments and for operations.

Another criticism of operational evaluations had addressed the *validity of historical comparisons*. There is some evidence that meteorological phenomena do not vary independently from day to day, week to week, or even year to year - that long term components of variation may exist in addition to short term occasion-to-occasion variability (Brier and Enger, 1952). It is not clear how grave this matter is and how much it might affect analysis of operations. One would hope for careful studies of time components of variation of precipitation which would tell us how far we might go wrong in analyzing the comparison of a seeded decade with an unseeded decade as though we were comparing two independent samples of 10 years each. As of now, we can only warn of this difficulty with historical/operational comparisons, but not gauge its magnitude.

5. REFERENCES

Achtemeier, G. L., and P. T. Schickedanz, 1979: On the temporal decay of the relationship between rainfall and environmental covariates for the High Plains. *J. Appl. Meteor., 18,* 1679-1683.

____, and N. E. Westcott, 1979: Environmental covariates derived from surface field analyses. *Preprints, 7th Conf. on Inadvertent and Planned Wea. Mod.,* Amer. Meteor. Soc., Boston, 207-208.

Bernier, J., 1967: On the design and evaluation of cloud seeding experiments performed by Zlectricite de France. *Proc., 5th Berkeley Symp. on Math. Statist. and Prob., 5,* 35-54.

Brier, G. W., and I. Enger, 1952: An analysis of the results of 1951 cloud seeding operations in central Arizona. *Bull., Amer. meteor. Soc., 33*(5), 208-210.

Bross, I. D. J., 1964: Taking a covariate into account. *J. Amer. Statis. Assoc., 59,* 725-736.

Brownlee, K. A., 1960: Statistical evaluation of cloud seeding operations. *J. of Amer. Statist. Assoc., 55,* 446-453.

Changnon, S. A., F. A. Huff, and C. F. Hsu, 1979: On the need to evaluate operational weather modification projects. *Bull., Amer. Meteor. Soc., 60,* 770-774.

____, ____, P. T. Schickedanz, and J. L. Vogel, 1977: *Summary of METROMEX Volume 1, Weather anomalies and impacts.* Bull. 61, Illinois

State Water Survey, Urbana, 260 pp.

Court, A., 1960: Evaluation of seeding trials. *J. Amer. Soc. Civil Eng., Irrig. and Drainage Div., 86,* 121-126.

Dennis, A. S., and K. R. Gabriel, 1980: Another look at the evaluation of seeding operations in Santa Clara. Grant Report NSF ATM79-05536.

____, and D. F. Kriege, 1966: Results of ten years of cloud seeding in Santa Clara County, California. *J. Appl. Meteor., 5,* 684-691.

Gabriel, K. R., 1980: Another look at the evaluation of seeding operations in Santa Clara. Report to NOAA State/Federal Project.

____, 1979a: Re-randomization tests on weather experiments. *Preprints, 7th Conf. Inadvertent and Planned Wea. Mod.,* Amer. Meteor. Soc., Boston, J5-J6.

____, 1979b: Some statistical issues in weather experimentation. *Comm. Statist. Theory-Meth., A8*(10), 975-1015.

High Plains Cooperative Program, 1979: *The Design of HIPLEX I.* U.S. Dept. of Interior, Bureau of Reclamation, 30 pp.

Hsu, C. F., 1979a: Monte Carlo studies of statistical evaluation techniques for weather modification. *Preprints, 7th Conf. Inadvertent and Planned Wea. Mod.,* Amer. Meteor. Soc., Boston, J3-J4.

____, 1979b: Two methods of computing statistical powers with application to weather modification. *Proc., Statis. Comp. Section,* Amer. Statist. Assoc., Washington, 243-246.

Huff, F. A., and S. A. Changnon, 1972: Evaluation of potential effects of weather modification on agriculture. *J. Appl. Meteor., 11,* 376-384.

Mielke, P. W., and J. S. Williams, 1977: Covariance analysis technique based on bivariate log-normal distribution with weather modification applications. *J. Appl. Meteor., 16,* 183-187.

Neyman, J., and E. L. Scott, 1961: Further comments on the "Final Report of the Advisory Committee on Weather Control." *J. Amer. Statist. Assoc., 56,* 580-600.

Petersen, T. A., 1975: An analysis of thirteen years of commercial hail suppression in central Alberta. *J. Was. Mod., 7,* 153-170.

Porthoff, R. F., 1974: A non-parametric test of whether two simple regression lines are parallel. *Ann. Statist., 2*(2), 295-310.

Quade, D., 1967: Rank analysis of covariance. *J. Amer. Statist. Assoc., 62,* 1187-1200.

Statistical Task Force, 1978: *The role of statistics in weather resources management.* U. S. Dept. of Commerce, Washington, DC.

Thom, H. C. S., 1957: An evaluation of a series of orographic cloud seeding operations. *Final Report of the Advisory Committee on Weather Control. Vol. II.* Washington, U. S. Government Printing Office, 25-50.

Weather Modification Advisory Board, 1978: *The Management of Weather Resources, Volume 1: Proposals for a National Policy and Program.* Report to the Secretary of Commerce, Washington, DC, 229 pp.

Westcott, N. E., 1979: *Annotated Bibliography of Predictor Variables for Weather Modification Application.* Illinois State Water Survey, Urbana, 120 pp.