THE MAGNIFYING GLASS VERSUS THE RUBBER STAMP – THE ROLE OF STATISTICS IN WEATHER MODIFICATION

Tressa L. Fowler*, Barbara G. Brown, and Eric Gilleland
National Center for Atmospheric Research, Boulder, Colorado

1. INTRODUCTION

Weather modification research has been going on for decades. However, it is still in need of "proof" that it works. Researchers rely on statistics for their evidence. However, even if weather modification experiments had been able to produce unequivocal statistical "proof" of an effect, replication of that effect is required before weather modification efforts can be declared a success. A wide variety of tests and analyses have been completed on weather modification experiments, yet no one is completely satisfied with the results.

The problem lies with the use of statistical hypothesis testing as a "rubber stamp of approval". Weather modification experiments are very costly and can only collect (relatively) little information. Additionally, the physical effects of weather modification are not always well understood. Thus, in order to squeeze every last bit of information out this very expensive data, several types of analyses must be completed. However, if each of these analyses is considered a "test" of the efficacy of cloud seeding, then multiplicity issues rule the results. Data from cloud seeding experiments is highly variable, and this reduces the power of even a single test to detect differences. Dividing the allowable error among multiple tests makes detecting differences practically impossible. Conversely, to collect huge quantities of data and spend small fortunes to arrive at a single test statistic is both foolish and wasteful. The recommended solution is to use statistics as a tool for discovery, a mathematical magnifying glass; not as a rubber stamp. Hypothesis testing, modeling (spatial, temporal, Bayesian, etc.), multiplicity issues, and exploratory analyses will be discussed in the sections 2 through 4, respectively. Specific strategies for balancing these analyses are discussed in the conclusions, presented in section 5.

2. HYPOTHESIS TESTING

From the beginning of cloud seeding experiments, use of the standard t-test has been eschewed due to the non-normality of rainfall measurements. Several alternatives have been proposed. Some experiments (e.g. South Africa and Mexico) used estimates of the quartiles to determine if cloud seeding produced significant differences (Mather et al., 1997; Bruintjes et al., 1997). However, quartile estimates have undesirable properties when used on small samples, especially when the distribution of the samples is asymmetric or skew. A few types of ratio statistics were devised in the course of other experiments (Gabriel, 1999). Unfortunately, a ratio of means is asymptotically distributed as Cauchy, which has undefined variance. This is not a desirable property for a statistic, and thus ratio statistics are not recommended.

* Corresponding aurhor address: Tressa L. Fowler, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307 e-mail: tressa@ucar.edu

Use of a standard statistical measure that is able to "cope" with small samples and skewed measurements, the Wilcoxon-Mann Whitney test, is recommended for the confirmatory phase on weather modification experiments. The formula for the WMW statistic and its distribution can be located in almost any statistics text.

The WMW test is a robust test of location based on the sum of the ranks of the observations (Hodges and Lehmann, 1956). Because it is a rank based test, the skewness of the measurements does not have any effect on the test. Small sample sizes have an effect on all statistics. However, tests that incorporate more assumptions require considerably smaller sample sizes than tests that have fewer assumptions *when those assumptions are correct*. The WMW test has fewer assumptions than a standard t-test making it more robust than the t-test. However, if the data are normally distributed, then the WMW test requires a greater sample size to detect the same size difference as a t-test. However, this sample size increase is quite small, usually about an extra 4% (Dixon, 1954). Additionally, when the data are even nominally different from normal, the WMW test is more efficient, i.e. requires a smaller sample size to detect differences than the t-test (Sprent, 1993).

Estimation of required sample sizes can be done when the t-test is employed. Unfortunately, no equivalent procedure exists for the WMW test. However, the t-test method for determination can be employed as a reasonable guess for the necessary sample size. For example, the result of this method, based on the variability estimates from data collected during a field program in the United Arab Emirates (UAE) during summer of 2002, yielded a sample size of 192 cases (96 each seeded and control) to detect a 50% increase in rain mass with 80% power and 5% error.

It has been suggested that spatial and/or temporal modeling be used in the confirmatory phase of weather modification experiments. Spatial and temporal modeling should certainly be used in the examination of weather modification experiments wherever feasible. These types of analyses may provide researchers with better insight into the dynamic effects of cloud seeding as well as any effects from other sources. However, in order to model a source of variation, the nature of the variation must be well understood (Chatfield, 1995). This may not be the case in many weather modification experiments. Additionally, it should be recognized that for some projects, the information that can be gathered may be insufficient to complete a full spatial and/or temporal analysis. It may prove difficult to measure or quantify the many sources of variation in precipitation production. Many projects are being undertaken in areas with relatively sparse coverage of weather instrumentation such as radar, rain gauges, etc. In some cases, seeding experiments are conducted in sparsely distributed locations and/or are not performed regularly. This would certainly make it difficult to do any kind of temporal model, and difficult to apply meaningful spatial analyses as well. These considerations may make it difficult to apply spatial analyses to weather modification experiments.

In addition, weather modification experiments are based on a variety of different designs. Some of these designs would be amenable to spatial/temporal modeling approaches, while others would not. For example, the approach could be very beneficial when the design involves two adjacent areas that are randomly assigned – perhaps on a daily basis – to be treatment or control areas (i.e., much like the Fisher experiments), assuming that adequate datasets are available. However, the approach would not be meaningful in a number of other cases, for example when the experimental unit is a cloud or a storm.

Most importantly, at this point in time, for the purposes of a confirmatory analysis, spatial/temporal modeling is unlikely to satisfy reviewers and critics. The results of such modeling may be sensitive to the inputs, priors, and assumptions. Thus, it may be difficult to show that any positive results are due to seeding and not to model specifications and assumptions (Chatfield, 1995). Additionally, spatial/temporal modeling is a relatively new science and is unlikely to be well understood by the weather modification community. Once these methods demonstrate successful application in the weather modification field and gain acceptance, they may be incorporated into the confirmatory analyses.

3. MULTIPLICITY ISSUES

In their report The Role of Statistics in Weather Resources Management, the Weather Modification Advisory Board's statistical task force claims, "dilution of our data by the asking of multiple questions . . . is guaranteed to keep us from being able to draw firm conclusions" (1978). The allowable error for any experiment is quite small, generally 5%. If several tests are conducted at the 5% error level, then the overall experimental error is much too great. However, if the error is divided between the tests, then little power remains to detect differences.

Many believe that a single question should be the focus of the confirmatory phase of weather modification experiment. Typically, this question is, "Does this type of weather modification effort produce increased precipitation (or decreased hail) in this location during this season?" Certainly other questions could be asked, but the question should be formulated carefully. John Tukey was fond of saying that it is better to have an approximate answer to the right question than the precise answer to the wrong question (Karen Kafadar, personal communication).

A separate exploratory phase of the experiment can include several types of analyses, and ensures that all available information is studied without the issues of multiplicity.

4. EXPLORATORY DATA ANALYSIS

The exploratory phase of a weather modification experiment is arguably the most important. Though weather modification experiments have been carried on for decades, plenty of unknowns remain. To approach the data from a weather modification experiment with the intent of confirming or discrediting a preconceived notion is fine, permitted that is not the extent of the analysis. To limit the analysis in this way is the equivalent of wearing blinders. Instead, it is recommended that the data be analyzed and examined in many different ways, with the intent of seeing what is there rather than looking for what one wants to see.

Certainly, there are as many ways to analyze data as there are people. A small sample of suggested methods is included here. For more methods or greater detail, see Chambers (1983) or Hoaglin *et al* (1983).

A QQ-plot is a plot of quartiles of one distribution or set of data versus the quartiles of another distribution or set of data. It can be used to determine if two sets of data come from the same distribution or it one set of data comes from some known distribution, such as a Gaussian.

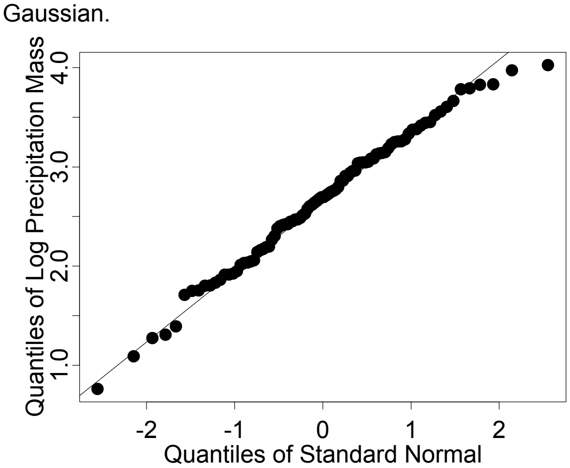

Figure 1: QQ-plot of log precipitation mass measurements versus the quantiles of a standard normal distribition.

If the points fall approximately on a straight line, then the two distributions are the same, otherwise they are no. An example is shown in Figure 1 below, with log precipitation measurements compared to the standard normal. This plot indicates that the log measurements are approximately normally distributed.

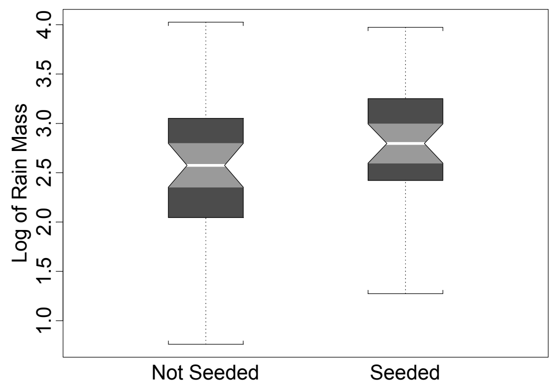

Figure 2: Boxplots showing distibutions of log rain mass measurements for seeded an unseeded cases.

Boxplots are another method of comparing samples. A boxplot shows the distribution of the measurements for each sample. The top and bottom of the box denote the 75th and 25th percentiles of the data, respectively. The center line in the box indicates the placement of the median and the notches around that center line show an approximate 95% confidence interval on the median. The top and bottom of the whiskers extend to the highest and lowest values of the data that are not outliers. Outliers, when there are any, are indicated with a mark above or below the whisker. An example boxplot is shown in Figure 2.

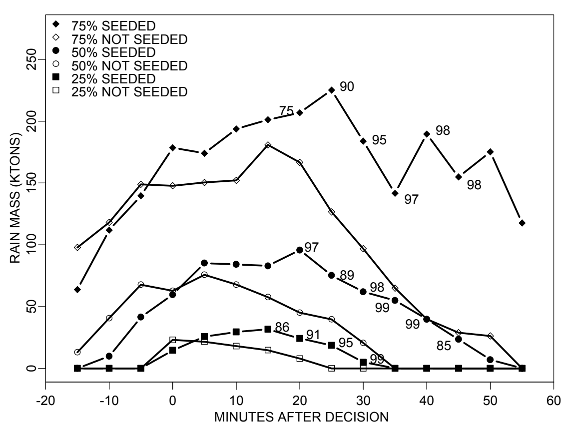

Figure 3: Time series plot of rain mass quartiles for seeded and unseeded cases. In some places, the plot has been annotated with p-values.

Most weather modification data have a time component, especially when measurements are taken by radar. When possible, it is recommended that the time series be examined for possible trends. Time series of individual storms can be interesting but may be overwhelming. Thus, time series of summary statistics such as averages, quartiles and variance are recommended instead. An example time-series plot using the quartiles is shown in Figure 3.

5. CONCLUSIONS AND STATISTICAL ISSUES

Statistics can be a great tool for uncovering the unknown from batches of data. While standard hypothesis testing tells us little, its use in the confirmatory stage of the experiment is nearly essential at this time. The exploratory phase of a weather modification experiment is limited only by time, money and imagination. Thus, it can be quite complete and thus is more likely to uncover interesting facts than the confirmatory analyses.

For the confirmatory phase of the experiment, the WMW test is recommended for its combination of efficiency and robustness. In the exploratory phase, a variety of exploratory graphical analyses are recommended.

Spatial and temporal analyses, whether Bayesian or not, should be conducted as part of the exploratory analyses of weather modification experiments where feasible and meaningful. Perhaps in the future, these analyses will be incorporated into the confirmatory phase of the experiments when it is appropriate to do so, but until then, accepted methods should continue to be used.

ACKNOWLEDGEMENTS

The views expressed are those of the authors and do not necessarily represent the official policy of the U.S. Government.

REFERENCES

Bruintjes, R. T., D. W. Breed, B. G. Foote, M. J. Dixon, B. G. Brown, V. Salazar, and H. R. Rodriguez. 1999: Program for the Augmentation of rainfall in Coahuila (PARC): overview and design. *Seventh WMO Scientific Conference on Weather Modification*, Thailand. 53-56.

Chambers, J.W., W.S. Cleveland, B. Kleiner, and P. A. Tukey, 1983: Graphical Methods for Data Analysis. Duxbury Press, Boston.

Chatfield, Chris, 1995: Model Uncertainty, Data Mining and Statistical Inference, Journal of the Royal Statistical Society. Series A (Statistics in Society), Vol. 158, No. 3. (1995), pp. 419-466.

Dixon, W. J., 1954: Power under normality of several non-parametric tests. *Ann. Math. Stat.*, **25,** 610-614.

Gabriel, R. 1999: Ratio statistics for randomized experiments in precipitation simulation. *J. Appl. Meteor.*, **41**, 505-518.

Hoaglin, D. C., R. Mosteller, and J. W. Tukey, editors 1983: Understanding Robust and Exploratory Data Analysis. John Wiley and Sons, New York.

Hodges, J. L., Jr., and E. L. Lehmann, 1956: The efficiency of some non-parametric competitors of the *t*-test. *Ann. Math. Stat*., **27,** 324-335.

Mather, G.K. D.E. Terblanche, F.E. Steffens, and L. Fletcher, 1997: Results of the South African cloud seeding experiments using hygroscopic flares. *J. Appl. Meteor*., **36**, 1433-1447.

Sprent, P., 1993: *Applied Nonparametric Statistical Methods*. Chapman and Hall, London, UK, 342pp.

Weather Modification Advisory (1978): The Role of Statistics in Weather Resources Management, Volume II. U.S. Government Printing Office, Washington, D.C.