SOME STATISTICAL TECHNIQUES FOR CLIMATOLOGICAL DATA*

by

Shanti S. Gupta
Purdue University

and

S. Panchapakesan
Southern Illinois University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #80-1

(Revised April 1980)

*This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

SOME STATISTICAL TECHNIQUES FOR CLIMATOLOGICAL DATA*

by

Shanti S. Gupta
Purdue University

and

S. Panchapakesan
Southern Illinois University

ABSTRACT

The need for and the increasing use of statistical techniques in the analysis of climatological data are amply illustrated in the literature. Some known techniques relating to meteorological problems such as weather modification experiments and objective weather forecasting are briefly reviewed here. Also, selection and ranking approach to multiple decision theory is discussed with emphasis on potential applications.

KEY WORDS

Climatological data, statistical models, weather modification, subset selection procedures, normal, gamma, multiple correlations, best predictor variables.

1. Introduction.

The need for statistical methodology in analyzing meteorological data has long been recognized. For example, weather modification provides, as noted by Braham (1979), a "fertile field of interaction and collaboration between meteorologists and statisticians." Satisfactory models have been found to describe meteorological data (see Section 2). Time series data occur

*This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

commonly in climatological studies. Some of the important and interesting problems arise in connection with weather modification experiments, objective weather forecasting and classification of meteorological patterns. Studies in meteorology in general and rain simulation in particular have inspired novel developments in probability and statistics. The concept of characteristic functional first developed by Kolmogorov was later reintroduced by Le Cam (1947) motivated by meteorological studies [see Neyman (1979a)]. The concepts of outlier-prone and outlier-resistant distributions developed in Neyman and Scott (1971) were motivated by cloud seeding experiments.

The objectives of the present paper are to briefly review some important known applications of statistical techniques to meteorological data and to indicate the potential applications of selection and ranking procedures to these problems. No attempt will be made to be comprehensive in the treatment of either objective. Some important distributions that have been satisfactorily used as models in meteorological problems are described in Section 2. The next section deals with weather modification experiments and some related asymptotic optimal tests and nonparametric tests. Section 4 discusses techniques used in a variety of situations other than weather modification experiments. The topics include Markov chain models, the biplot technique, selection of the best predictors in forecasting, and classification of weather patterns. The last section describes some subset selection procedures and discusses the selection of the best regression model under this formulation.

2. Statistical Models.

In this section, we briefly discuss several distributions that have been found useful as models for meteorological data. Any discussion of the techniques for inference will be deferred until later sections.

The gamma distribution has been extensively used as a model for precipitation data. Rain simulation experiments indicate [see Neyman and Scott (1971)] that the distribution of nonzero rainfall per experimental unit (an experimental day or storm) is J-shaped with a long tail and frequent outliers. The gamma distribution answering the above description [Neyman and Scott (1971)] has been found a satisfactory model in practice. The distribution of nonzero rainfall per experimental unit is assumed to have the density

$$(2.1) \qquad f(x) = \frac{\theta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x}, \ x \geq 0, \quad \theta > 0, \alpha > 0.$$

Here, θ is the reciprocal of the scale parameter and α is the shape parameter. It is generally assumed [Neyman (1979a)] that the seeding of the clouds can change the value of the scale parameter but has no effect on the shape parameter. The gamma distribution has been used or verified as a model by Barger and Thom (1949), Mooley and Crutcher (1968), Neyman and Scott (1967a), Schickedanz (1967), Schickedanz and Decker (1969), Simpson (1972), and Thom and Vestal (1968).

Mielke (1973) considered for describing precipitation data the two-parameter Kappa distribution with distribution function

$$(2.2) \qquad F(x) = \begin{cases} [(x/\beta)^{\alpha}/\{\alpha + (x/\beta)^{\alpha}\}]^{1/\alpha}, & x \geq 0 \\ 0 \qquad , & x < 0 \end{cases}$$

where $\alpha > 0$ and $\beta > 0$ denote the shape and scale parameters, respectively.

Wong (1977) made goodness-of-fit comparisons among the gamma, lognormal, three-parameter kappa ($\alpha\theta$ in the place of α in (2.2)), and Weibull distributions using five sets of Alberta hailfall data. He found the Weibull distribution a reasonable alternative to the lognormal and three-parameter kappa distributions for describing precipitation and streamflow data. It should be noted that the lognormal distribution is outlier resistant [Neyman (1979a)] and that Weibull and gamma distributions can be subsumed under the generalized gamma distribution with density

$$(2.3) \qquad f(x) = \frac{\gamma x^{\gamma\alpha-1}}{\beta^{\gamma\alpha}\Gamma(\alpha)} e^{-(x/\beta)^{\gamma}}, \quad x > 0,$$

where α, β, and γ are all positive parameters.

The three-parameter Weibull distribution was used by Stewart and Essenwanger (1978) as a model for wind speed near the surface. Tackle and Brown (1978) have used the distribution function

$$(2.4) \qquad F(x) = \begin{cases} F(0) + (1-F(0))\ (1-\exp\{-(x/\theta)^{\beta}\}), & x \geq 0 \\ 0 & , \ x < 0 \end{cases}$$

where F(0) is the probability of observing zero wind speed.

Luna and Church (1974) have found the lognormal distribution as a satisfactory model for wind speed at many sites. Yao (1974) found the beta distribution as a satisfactory model for frequency distributions of relative humidity observations. The beta distribution has also been used by Mielke (1975).

Bivariate normal distribution is used by Wu, Williams and Mielke (1972) in the analysis of continued-covariate and cross-over designs that arise in cloud seeding experiments. For some other distributions that have been considered in connection with meteorological data, see Mielke (1979).

Associated with all these distributions are the obvious problems of estimation. The several methods of estimation applied to these distributions are of general interest and not restricted to meteorological problems; as such, relevant references can be amply found in the statistical literature. It suffices here to mention a few recent papers motivated by meteorological applications, namely, Crow (1977, 1978), Flueck and Holland (1976), Mielke (1973, 1976), Mielke and Johnson (1973), and Wong (1977). Other problems of inference are discussed in subsequent sections.

3. Weather Modification Experiments.

Early scientific weather modification experiments are attributed to Vincent Schaefer (1946) and Barnard Vonnegut (1947) who showed that pellets of Dry Ice and minute particles of silver iodide would nucleate ice crystals in supercooled clouds. Early days of weather modification are discussed by Byers (1974) and Elliot (1974). One of the important experiments, known as Project Whitetop, was carried out by Professor Braham and his colleagues at the University of Chicago during the summers of 1960 through 1964. The data of this experiment have been reanalyzed by Professor Neyman and his associates at Berkeley. The details of Project Whitetop, controversies regarding its conclusions, and relevant references can be found in the paper by Braham (1979) and the comments by Dawkins and Scott (1979) and Neyman (1979b). A categorized bibliography of weather modification experiments is given by Hanson et al (1979).

Weather modification experiments are getting increasing attention of statisticians as evidenced by the papers in Volume V of the Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, 1967) devoted entirely to this subject and two special issues of Communications in Statistics - Theory and Methods (Volume A8, Numbers 10 and 11, 1979). In the rest of this section, we briefly describe some of the problems and techniques.

A class of asymptotic tests that is routinely used in testing hypotheses regarding the effect of seeding the clouds is called the optimal $C(\alpha)$ tests. These tests were developed by Neyman (1959) and are applicable to testing composite hypotheses that are frequently encountered in practice. These tests are applied by Neyman and Scott (1967b) for evaluating single rain simulation experiments. The basic assumption is that, whether or not seeded, there corresponds to each experimental unit (a fixed duration like 24 hours) a positive probability, say 1-p, of the rainfall being zero. Two mechanisms are introduced, one governing the change in p due to seeding and the other governing the effects of seeding per wet day. The effect of each mechanism is a change in the value of p in either direction. On each experimental day (a day considered 'suitable' for seeding), a randomized decision is made whether or not to proceed with seeding. As a measure of the effect of seeding, Neyman and Scott (1967b) use $\xi = (p_s - p_c)/p_c$, where the subscripts s and c denote "seeded" and "control", respectively. Neyman and Scott (1967b) provide three test criteria, labeled Z_1, Z_2, and Z_3, of which the first two are optimal $C(\alpha)$ tests of hypotheses H_1 and H_2, that cloud seeding does not affect the frequency of wet days, and that it does not affect the rainfall per wet day, respectively. The criterion Z_3 is not a $C(\alpha)$ test; it is a linear combination of Z_1 and Z_2 so chosen as to be sensitive to departures from H_3 that the seeding does not affect the target precipitation averaged per experimental unit, whether wet or dry. The specialization of the conditional density of the target precipitation given that it is not zero, joint with the predictors if such are available, determines several different cases. For some recent work on the detection of variable response to cloud seeding, see Neyman (1979a).

Efficient methods for summary evaluations of several independent experiments are important in view of "the notorious frequency with which rain simulation experiments fail to yield statistically significant results." Davies and Puri (1967) discuss two related but distinct problems specializing certain earlier results concerning $C(\alpha)$ tests.

Suppose that the distribution of the nonzero precipitation is gamma with density given in (2.1). It is assumed that the effect of seeding is to change θ to $\xi\theta$ (i.e. effect is multiplicative). The interest is to test H: $\xi \geq 1$ against A: $\xi < 1$. Note that $\xi < 1$ corresponds to increased average nonzero rainfall. The results of several cloud seeding experiments indicate [Neyman and Scott (1967c)] a value of α in the interval (0.45, 0.75). One can use likelihood ratio tests or $C(\alpha)$ tests. However, it is a simplistic assumption that the changes induced by cloud seeding can be adequately represented by a simple scale or location parameter shift. Thus, nonparametric techniques are useful in testing for a change due to seeding in the distribution of precipitation amount. Commonly used nonparametric tests are Wilcoxon, Kolmogorov-Smirnov, and median tests. Another test which is applicable is due to Taha (1964) and is based on the statistic $L = \frac{1}{n} \sum_{i=1}^{n} s_i^2$, where the s_i are the ranks of the "seeded" observations in the combined sample of 2n observations. In the sense of asymptotic efficiency, this L test is found superior to Wilcoxon test. James (1967) has made some numerical comparisons of the Pitman efficiency of Wilcoxon, gamma scores, exponential scores and L tests for small values of α coming out in favor of the exponential scores test.

Tamura (1963) proposed a class of tests based on the statistic $A_r = \sum_{j=1}^{N} j^r Z_j$, where $r > 0$, and $Z_j = 1$ or 0 if the jth ordered observation in the pooled sample of size N is a seeded or a non-seeded observation. A similar class of two-sample nonparametric tests is considered by Mielke (1972, 1974) to treat the same problem but with the cross-over design.

Multivariate nonparametric and permutation procedures are useful when a number of measured responses are obtained from each experimental unit. Mielke, Berry and Johnson (1976) have considered multi-response permutation procedures, special cases of which have been earlier suggested by Mantel and Valand (1970). For some further discussion of these procedures, see Mielke (1979).

Weather modification experiments are carried out in a natural environment subject to much variability. Covariates are used in analysis in order to reduce the experimental error. Bradley, Srivastava and Lanzdorf (1979) have discussed covariance analyses effected through the use of multiple regression methods. They have also reviewed the original results of an experiment conducted by North American Weather Consultants and discussed a multivariate analysis without use of covariates or transforms.

4. Statistical Techniques for Other Meteorological Problems.

In this section, we briefly discuss applications of certain statistical techniques to meteorological problems other than the weather modification. The examples are chosen to indicate the scope and the nature of applications.

In several situations we need more sophisticated models than those discussed in Section 2. An important problem in meteorology is the determination of the characteristics of hourly temperatures. Hansen and Driscoll (1977) developed a stochastic model for hourly temperatures for Big Spring, Texas. These temperatures are produced by harmonics representing both diurnal and annual variations, and a Markov chain expression incorporating adjustments for several variations such as seasonal variation of the serial correlation coefficient.

Markov chain models have been used to describe the daily occurrence of precipitation. Gabriel and Neumann (1962) considered a model for daily rainfall occurrence at Tel Aviv. Another model was introduced by Todorovic and Woolhiser (1975). Recently, Katz (1977) proposed a more general model and discussed the distribution of the maximum amount of daily precipitation and the distribution of the total precipitation.

The biplot is a graphical display of a two-dimensional approximation to a matrix obtained by least squares using the first and second singular value components of the matrix. It is related to principal component analysis and multivariate analysis of variance (MANOVA). Its usefulness in the display and analysis of meteorological data is demonstrated by Gabriel (1972) with two sets of data. In one of the examples, the biplot is an approximation to simultaneous tests of different subhypotheses in the one-way layout MANOVA. For mathematical and computational details of the technique, see Gabriel (1971).

In forecasting the state of atmosphere at grid points, we have the problem of obtaining vector-valued estimates of meteorological parameters at a grid point based on multivariate information from several sources. In other words, our estimator Z, a vector of n components, is given by $Z = A_1X_1 + \ldots + A_mX_m$, where the X_i are information vectors (each of n components) and the A_i are nxn matrices. Here the X_i have some joint distribution. The problem is to find the "best" linear combination of the information vectors. Thiebaux (1974a) has considered the criterion of minimizing the variances of the components of Z. An example of this situation is given in Thiebaux (1973). In another paper, Thiebaux (1974b) has discussed a related problem regarding the estimation of covariances of meteorological parameters using local-time averages.

McCutchan and Schroeder (1973) have used stepwise discriminant analysis of eight meteorological variables to classify the days during their study period at a southern California location into one of five types.

Many examples of statistical prediction schemes in climatology are available. The prediction is based on a number of predictor variables. While the prediction can be made more accurate by bringing in as many relevant predictor variables as possible, some of them may be highly correlated among themselves and the contribution of some may be very marginal. The problem of selecting the best set of predictor variables arise in various situations. Stringer (1972, pp. 132-133) has cited some examples from literature regarding prediction of precipitation and visibility. Martin et al (1963) have considered an example dealing with forecasting of the 24-hour movement and change of central pressures of North American winter antincyclones. Lund (1971) has discussed a problem of estimation of precipitation involving almost 4500 potential predictors.

Several criteria for defining the best set of predictor variables and various techniques for selecting the best set have been discussed in a nice expository paper by Hocking (1976). Also, a brief review and evaluation of significant methods have been given by Thompson (1978). Martin et al (1963) applied forward type stepwise procedure. Lund (1971) has illustrated a method of blending stagewise and stepwise procedures.

It should be noted that these techniques for selecting the best set of predictor variables are not designed to produce a best set with a guaranteed probability. We will come back to this point in the next section.

5. Ranking and Selection Procedures.

In dealing with weather data, one may want to compare different sites (weather stations) on the basis of appropriate characteristics of the meteorological variables involved. For example, we may want to compare these locations on the basis of mean temperature, or mean nonzero precipitation amount, or variability of temperature for a fixed duration. One may be interested in ranking the sites in terms of the values of the characteristic or just in selecting the site with the largest (smallest) value of the characteristic.

Formallyspeaking, we have k independent populations (sites) $\pi_1,\ldots,\pi_k$, where π_i is characterized by the distribution function $F(x;\theta_1)$ and θ_i is an unknown parameter which represents the "worth" of the population. For example, $F(x;\theta_i)$ may be the distribution function of the 24-hour nonzero precipitation amount at the ith site and θ_i may be the mean of the distribution. Let $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ denote the ordered θ_i. To be specific, let us say that π_i is "preferable" to π_j if $\theta_i > \theta_j$ so that the best population is the one associated with the largest θ_i. Ranking and selection problems have been generally formulated using either the <u>indifference zone approach</u> or the <u>subset selection approach</u>.

Let us consider the simple problem of selecting the best population. Under the indifference zone formulation of Bechhofer (1954), we want a procedure R which will select the best population with a minimum guaranteed probability P^* $(1/k < P^* < 1)$ whenever $\delta(\theta_{[k]}, \theta_{[k-1]}) \geq \theta^*$, where $\delta(\theta_{[k]},\theta_{[k-1]})$ is an appropriate measure of the distance between the populations associated with $\theta_{[k]}$ and $\theta_{[k-1]}$, and the quantities θ^* and P^* are specified in advance. In the cases of location and scale parameters, the natural choices for $\delta(\theta_{[k]},\theta_{[k-1]})$ are $\theta_{[k]}-\theta_{[k-1]}$ and $\theta_{[k]}/\theta_{[k-1]}$, respectively. Consequently, $\theta^* > 0$ in the first case and $\theta^* > 1$ in the second. Suppose we want a procedure R based on samples of equal sizes. The problem is to determine the minimum sample size needed to meet the probability requirement.

In the subset selection approach, our goal is to select a non-empty subset of the k populations so that the best population is included in the selected subset with a minimum guaranteed probability P*. Selection of any subset which includes the best population is called a correct selection (CS). The general approach is to evaluate the infimum of P(CS|R), the probability of a correct selection using the procedure R, over the parameter space $\Omega = \{\underline{\theta}: \underline{\theta} = (\theta_1,\ldots,\theta_k)\}$ and obtain the constants involved in defining R so that

$$\inf_{\Omega} P(CS|R) \geq P^*. \tag{5.1}$$

The condition (5.1) is referred to as the P*-condition or the basic probability requirement. In order to meet this requirement, one determines the parametric configuration $\underline{\theta}_0$, the Least Favorable Configuration (LFC), for which the infimum in (5.1) is attained. In general, there may not be a unique LFC. The expected size of the subset selected is one of the measures generally used as performance characteristics of a procedure.

For an extensive survey and bibliography of ranking and selection theory and related topics the reader is referred to the recent book of the authors (1979). Other books in this area are Bechhofer, Kiefer and Sobel (1968), and Gibbons, Olkin and Sobel (1977).

In the rest of this section, we describe briefly subset selection procedures for normal populations in terms of means, for gamma populations in terms of the scale parameter, for multivariate normal populations in terms of multiple correlations coefficients and discuss selection of best predictor variables.

5.1 Normal Populations. Let $\pi_1, \ldots, \pi_k$ be k independent normal populations with unknown means $\mu_1, \ldots, \mu_k$, respectively, and a common variance σ^2. Let $\overline{X}_i$, i=1, ..., k, be the sample means based on samples of size n. The best population is the one associated with the largest μ_i. When σ^2 is known, the procedure R_1 proposed by Gupta (1956) selects the population π_i if and only if

$$(5.2) \qquad \overline{X}_i \geq \max(\overline{X}_1, \ldots, \overline{X}_k) - \frac{d_1\sigma}{\sqrt{n}}$$

where $d_1 = d_1(k, P^*) > 0$ is the smallest constant such that the condition (5.1) is satisfied. The LFC is given by $\mu_1 = \ldots = \mu_k$. This implies that d_1 is given by

$$(5.3) \qquad \int_{-\infty}^{\infty} \Phi^{k-1}(x+d_1)\, \phi(x)\, dx = P^*,$$

where $\Phi(x)$ and $\phi(x)$ are the standard normal cdf and density, respectively. The values of d_1 are tabulated for several values of k and P* by Gupta (1963a)and Gupta, Nagel and Panchapakesan (1973).

When σ^2 is not known, the procedure R_2 of Gupta (1956) is the same as R_1 with σ replaced by s, where s^2 is the usual pooled estimator of σ^2 based on $\nu = k(n-1)$ degrees of freedom. Here again, the LFC is given by $\mu_1 = \ldots = \mu_k$. The values of the constant d_2 (used in the place of d_1) are tabulated by Gupta and Sobel (1957) for selected values of k, ν, and P*.

The procedures R_1 and R_2 can be modified in the case of the population with the smallest μ_i being defined the best. For procedures involving unequal sample sizes, see Gupta and Huang (1976), and Gupta and Wong (1976).

5.2 Gamma Populations. Let π_i have the associated density

$$(5.4) \qquad f(x, \theta_i) = \begin{cases} \dfrac{x^{r-1}}{\Gamma(r)\theta_i^{\,r}} \exp(-x/\theta_i), & x > 0, \ \theta_i > 0 \\ 0 & \text{otherwise.} \end{cases}$$

As we can see, it is assumed that the populations have the same shape parameter $r(> 0)$. Further, r is assumed to be known. Our interest is selecting the population associated with the largest (smallest) θ_i. The gamma distribution not only serves as a model for certain types of measurement, but also includes the case where the observations come from normal populations and the interest is in selecting the population associated with the smallest variance.

For selecting the population associated with the largest θ_i, Gupta (1963b)investigated the procedure R_3 which selects π_i if and only if

$$(5.5) \qquad \overline{X}_i \geq b \max(\overline{X}_1, \ldots, \overline{X}_k)$$

where $\overline{X}_1, \ldots, \overline{X}_k$ are means based on samples of equal size n, and the constant b $(0 < b < 1)$ is chosen so that the P*-condition is met. Gupta (1963b) has shown that $P(CS|R_3)$ is minimized when $\theta_1 = \ldots = \theta_k$ and that the constant b is given by

$$(5.6) \qquad \int_0^\infty G_\nu^{k-1}(x/b)\, g_\nu(x)\, dx = P^*,$$

where $G_\nu(x)$ is the cdf of a standardized gamma random variable (i.e. with $\theta = 1$) with parameter $\nu/2$ where $\nu = 2nr$. Thus the constant b depends on n and r only through ν and its values are tabulated by Gupta (1963b) for selected values of k, P^*, and ν.

For selecting the normal population with the smallest variance, an analogous procedure is given by Gupta and Sobel (1962a) and the appropriate constant can be obtained from the tables in their companion paper (1962b).

5.3 Multivariate Normal Populations. Let $\pi_1, \ldots, \pi_k$ be k independent p-variate normal population where π_i is $N(\underline{\mu}_i, \sum_i)$. Let $\underline{X}_i' = (X_{i1}, X_{i2}, \ldots, X_{ip})$ be a random observation vector from π_i, i=1, ..., p. The populations are ranked in terms of the ρ_i, where ρ_i is the multiple correlation coefficient of X_{i1} with respect to the set $(X_{i2}, \ldots, X_{ip})$. We are interested in selecting a subset containing the population associated with the largest ρ_i. Let R_i denote the sample multiple correlation coefficient between X_{i1} and $(X_{i2}, \ldots, X_{ip})$. Two cases arise: (i) The case in which $X_{i2}, \ldots, X_{ip}$ are fixed, called the conditional case; (ii) The case in which $X_{i2}, \ldots, X_{ip}$ are random, called the unconditional case. In either case, Gupta and Panchapakesan (1969) proposed and studied the rule $\mathcal{R}$ which selects π_i if and only if

$$(5.7) \qquad R_i^{*2} \geq c \max_{1 \leq j \leq k} R_j^{*2}$$

where $R_i^{*2} = R_i^2/(1-R_i^2)$, and $0 < c = c(k, P^*, p, n) < 1$ is chosen to satisfy the P*-requirement. In this case, the infimum of PCS is attained when $\rho_1 = \rho_2 = \cdots = \rho_k = 0$ and the appropriate constant c is given by

$$(5.8) \qquad \int_0^\infty F_{2q,2m}^{k-1}(x/c)\, f_{2q,2m}(x)\, dx = P^*,$$

where $q = \frac{1}{2}(p-1)$, $m = \frac{1}{2}(n-p)$, $F_{r,s}$ denotes the cdf of an F random variable with r and s degrees of freedom, and $f_{r,s}$ denotes the corresponding density. The values of c are tabulated by Gupta and Panchapakesan (1969) for selected values of k, m, q, and P*.

5.4. Selection of Best Set of Predictor Variables. In Section 4, we referred to the techniques that have been commonly used for selecting the best predictor variables. We pointed out that these procedures are not designed to guarantee a minimum probability of obtaining the best set. Recently this problem has been investigated by Arvesen and McCabe (1973, 1975), McCabe and Arvesen (1974), and Gupta and Huang (1977) under the subset selection formulation described earlier in this section. Investigations along these lines continue to be of interest in view of their practical importance.

5.5. Other Procedures and Related Problems. There are several parametric and nonparametric procedures available in the literature to suit many contexts that commonly arise. There are single-stage, double-stage, and sequential procedures. There are several modifications of the basic problem. Also important are the related problems of estimating the ordered parameters. Many of these are areas of current research. For an extensive survey and bibliography, see Gupta and Panchapakesan (1979).

REFERENCES

Arvesen, J. N. and McCabe, G. P. (1973). Variable selection in regression analysis. In Proceedings of the University of Kentucky Conference on Regression with a Large Number of Predictors (Ed. W. O. Thompson and F. B. Cady), Dept. of Statist., Univ. of Kentucky, Lexington.

Arvesen, J. N., and McCabe, G. P. (1975). Subset selection problems of variances with applications to regression analysis. J. Amer. Statist. Assoc., 70, 166-170.

Barger, G. L. and Thom, H. C. S. (1949). Evaluation of drought hazard. Agron. J., 41, 519-526.

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist., 25, 16-39.

Bechhofer, R. E., Kiefer, J. and Sobel, M. (1968). Sequential Identification and Ranking Procedures. The University of Chicago Press, Chicago.

Bradley, R. A., Srivastava, S. S. and Lanzdorf, A. (1979). Some approaches to statistical analysis of weather modification experiment. Commun. Statist.-Theor. Meth., A8(11), 1049-1081.

Braham, R. E. (1979). Field experimentation in weather modification. J. Amer. Statist. Assoc., 74, 57-68.

Byers, H. R. (1974). History of weather modification. In Weather and Climate Modification (ed. W. N. Hess), New York: John Wiley & Sons, pp. 3-44.

Crow, E. L. (1977). Minimum variance unbiased estimators of the ratio of means of two lognormal variates and of two gamma variates. Commun. Statist.-Theor. Meth., A6(10), 967-975.

Crow, E. L. (1978). Confidence limits for seeding effect in single-area weather modification experiments. J. Appl. Meteor., 17, 1652-1660.

Davies, R. B. and Puri, P. S. (1967). Some techniques of summary evaluations of several independent experiments. In Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (ed. L. M. Le Cam and J. Neyman), Vol. V, Los Angeles and Berkeley: University of California Press, pp. 385-388.

Dawkins, S. M. and Scott, E. L. (1979). Comment on the paper by R. R. Braham. J. Amer. Statist. Assoc., 74, 70-77.

Elliott, R. D. (1974). Experience of the private sector. In Weather and Climate Modification (ed. W. N. Hess), New York: John Wiley & Sons, pp. 45-89.

Flueck, J. A. and Holland, B. S. (1976). Ratio estimators and some inherent problems in their utilization. J. Appl. Meteor., 15, 536-543.

Gabriel, K. R. (1971). The biplot graphic display of matricies with application to principal component analysis. Biometrika, 58, 453-467.

Gabriel, K. R. (1972). Analysis of meteorological data by means of decomposition and biplots. J. Appl. Meteor., 11, 1071-1077.

Gabriel, K. R. and Neumann, J. (1962). A Markov chain model for daily rainfall occurrence at Tel Aviv. Quart. J. Roy. Meteor. Soc., 88, 90-95.

Gibbon, J. D., Olkin, I. and Sobel, M. (1977). Selecting and Ordering Populations. New York: John Wiley & Sons.

Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Ph.D. Thesis (Mimeo. Ser. No. 150), Institute of Statistics, University of North Carolina, Chapel Hill.

Gupta, S. S. (1963a). Probability of integrals of the multivariate normal and multivariate t. Ann. Math. Statist., 34, 792-828.

Gupta, S. S. (1963b). On a selection and ranking procedure for gamma populations. Ann. Inst. Statist. Math., 14, 199-216.

Gupta, S. S. and Huang, D. Y. (1976). Selection procedures for the means and variances of normal populations: unequal sample sizes case. Sankhyā Ser. B, 38, 112-128.

Gupta, S. S. and Huang, D. Y. (1977). On selecting an optimal subset of regression variables. Mimeo. Ser. No. 501, Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Gupta, S. S., Nagel, K. and Panchapakesan, S. (1973). On the order statistics from equally correlated normal random variables. Biometrika, 60, 403-413.

Gupta, S. S. and Panchapakesan, S. (1969). Some selection and ranking procedures for multivariate normal populations. In Multivariate Analysis - II (Ed. P. R. Krishnaiah), New York: Academic Press, pp. 475-505.

Gupta, S. S. and Panchapakesan, S. (1979). Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations. New York: John Wiley.

Gupta, S. S. and Sobel, M. (1957). On a statistic which arises in selection and ranking problems. Ann. Math. Statist., 28, 957-967.

Gupta, S. S. and Sobel, M. (1962a). On selecting a subset containing the population with the smallest variance. Biometrika, 49, 495-507.

Gupta, S. S. and Sobel, M. (1962b). On the smallest of several correlated F-statistics. Biometrika, 49, 509-523.

Gupta, S. S. and Wong, W. Y. (1976). Subset selection procedures for the means of normal populations with unequal variances: unequal sample sizes case. Mimeo. Ser. No. 473, Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Hansen, J. and Driscoll, D. M. (1977). A mathematical model for the generation of hourly temperatures. J. Appl. Meteor., 16, 935-948.

Hanson, M. A., Barker, L. E., Bach, C. L., Cooley, E. A. and Hunter, C. H. (1979). A bibliography of weather modification experiments. Commun. Statist.-Theor. Meth., A8(11), 1129-1153.

Hocking, R. R. (1976). The analysis and selection of variables in linear regression. Biometrics, 32, 1-49.

James, B. R. (1967). On Pitman efficiency of some tests of scale for the gamma distribution. In Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (ed. L. M. Le Cam and J. Neyman), Vol. V, Los Angeles and Berkeley: University of California Press, pp. 389-393.

Katz, R. W. (1977). Precipitation as a chain-dependent process. J. Appl. Meteor., 16, 671-676.

Le Cam, L. M. (1947). Un instrument d'etude des fonctions aleatoires: La fonctionnelle characteristique. C. R. Acad. Sci. Paris, 224, 710-711.

Luna, R. E. and Church, H. W. (1974). Estimation of long-term concentrations using a "universal" wind speed distribution. J. Appl. Meteor., 13, 910-916.

Lund, I. A. (1971). An application of stagewise and stepwise regression procedures to a problem of estimating precipitation in California. J. Appl. Meteor., 10, 892-902.

Mantel, N. and Valand, R. S. (1970). A technique of nonparametric multivariate analysis. Biometrics, 26, 547-558.

Martin, F. L., Borsting, J. R., Steckbeck, F. J. and Manhard, A. H. (1963). Statistical prediction methods for North American winter anticyclones. J. Appl. Meteor., 2, 508-516.

McCabe, G. P. and Arvesen, J. N. (1974). A subset selection procedure for regression variables. J. Statist. Comput. Simul., 3, 137-146.

Mielke, P. W. (1972). Asymptotic behavior of two-sample tests based on powers of ranks for detecting scale and location alternatives. J. Amer. Statist. Assoc., 67, 850-854.

Mielke, P. W. (1973). Another family of distributions for describing and analyzing precipitation data. J. Appl. Meteor., 12, 275-280.

Mielke, P. W. (1974). Squared rank test appropriate to weather modification cross-over design. Technometrics, 16, 13-16.

Mielke, P. W. (1975). Convenient beta distribution likelihood techniques for describing and comparing meteorological data. J. Appl. Meteor., 14, 985-990.

Mielke, P. W. (1976). Simple iterative procedures for two-parameter gamma distribution maximum likelihood estimates. J. Appl. Meteor., 15, 181-183.

Mielke, P. W. (1979). Some parametric, nonparametric and permutation inference procedures resulting from weather modification experiments. Commun. Statist.-Theor. Meth., A8(11), 1083-1096.

Mielke, P. W., Berry, K. J. and Johnson, E. S. (1976). Multi-response permutation procedures for a priori classifications. Commun. Statist.-Theor. Meth., A5, 1409-1424.

Mielke, P. W. and Johnson, E. S. (1973). Three parameter kappa distribution maximum likelihood estimates and likelihood ratio tests. Mon. Wea. Rev., 101, 701-707.

Mooley, D. A. and Crutcher, H. L. (1968). An application of the gamma distribution function to Indian rainfall. ESSA Tech. Report. EDS 5.

McCutchan, M. H. and Schroeder, M. J. (1973). Classification of meteorological patterns in southern California by discriminant analysis. J. Appl. Meteor., 12, 571-577.

Neyman, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics* (ed. U. Grenander), New York: John Wiley & Sons, pp. 213-234.

Neyman, J. (1979a). Developments in probability and mathematical statistics generated by studies in meteorology and weather modification. *Commun. Statist.-Theor. Meth.*, A8(11), 1097-1110.

Neyman, J. (1979b). Comment on the paper by R. R. Braham. *J. Amer. Statist. Assoc.*, 74, 90-94.

Neyman, J. and Scott, E. L. (1967a). Some outstanding problems relating to rain modification. In *Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (ed. L. M. Le Cam and J. Neyman), Vol. V, Los Angeles and Berkeley: University of California Press, pp. 293-325.

Neyman, J. and Scott, E. L. (1967b). Note on techniques of evaluation of single rain simulation experiments. In *Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (ed. L. M. Le Cam and J. Neyman), Vol. V, Los Angeles and Berkeley: University of California Press, pp. 371-384.

Neyman, J. and Scott, E. L. (1967c). On the use of $C(\tilde{\alpha})$ optimal tests of composite hypotheses. *Bull. Inst. Internat. Statist.*, 41, 477-496.

Neyman, J. and Scott, E. L. (1971). Outlier proneness of phenomena and of related distributions. In *Optimizing Methods in Statistics* (ed. J. Rustagi), New York and London: Academic Press, pp. 413-430.

Schaefer, V. J. (1946). The production of ice crystals in a cloud of supercooled water droplets. *Science*, 104, 457-459.

Schickedanz, P. T. (1967). A Monte Carlo method for estimating the error variance and power of the test for a proposed cloud seeding experiment. Ph.D. Thesis, University of Missouri, Columbia.

Schickedanz, P. T. and Decker, W. L. (1969). A Monte Carlo technique for designing cloud seeding experiments. *J. Appl. Meteor.*, 8, 220-228.

Simpson, J. (1972). Use of the gamma distribution in single-cloud rainfall analysis. *Mon. Wea. Rev.*, 100, 309-312.

Stewart, D. A. and Essenwanger, O. M. (1978). Frequency distribution of wind speed near the surface. *J. Appl. Meteor.*, 17, 1633-1642.

Stringer, E. T. (1972). *Techniques in Climatology*. San Francisco: W. H. Freeman Company.

Tackle, E. S. and Brown, J. M. (1978). Note on the use of Weibull statistics to characterize wind-speed data. *J. Appl. Meteor.*, 17, 556-559.

Taha, M. A. H. (1964). Rank test for scale parameter for asymmetrical one-sided distributions. *Publ. Inst. Statist. Univ. Paris*, 13, 169-179.

Thiebaux, H. J. (1973). Statistical approaches to grid-point estimation of meteorological parameters. In Proc. of the Third Conference on Probability and Statistics in Atmospheric Science, Boston: American Meteorological Society, pp. 202-206.

Thiebaux, H. J. (1974a). Minimum variance estimation of coefficient matrices in a dependent system. Biometrika, 61, 87-90.

Thiebaux, H. J. (1974b). Estimation of covariances of meteorological parameters using local-time averages. J. Appl. Meteor., 13, 592-600.

Thom, H. C. S. and Vestal, I. B. (1968). Quartiles of monthly precipitation for selected stations in the contiguous United States. ESSA Tech. Report EDS 6.

Thompson, M. L. (1978). Selection of variables in multiple regression: Part I. A review and evaluation. Int. Statist. Rev., 46, 1-19.

Todorovic, P. and Woolhiser, D. A. (1975). A stochastic model of n-day precipitation. J. Appl. Meteor., 14, 17-24.

Vonnegut, B. (1947). The nucleation of ice formulation by silver iodide. J. Appl. Phy., 18, 593-595.

Wong, R. K. W. (1977). Weibull distribution, iterative likelihood techniques and hydrometeorological data. J. Appl. Meteor., 16, 1360-1364.

Wu, S. C., William, J. S. and Mielke, P. W. (1972). Some designs and analyses for temporally independent experiments involving correlated bivariate responses. Biometrics, 28, 1043-1061.

Yao, A. Y. M. (1974). A statistical model for the surface relative humidity. J. Appl. Meteor., 13, 17-21.