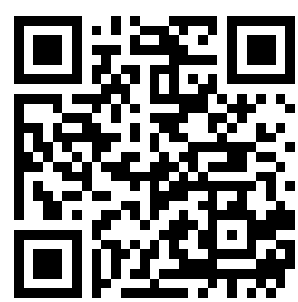


ARO Report 85-2

PROCEEDINGS OF THE THIRTIETH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Sponsored by
The Army Mathematics Steering Committee
on Behalf of

THE CHIEF OF RESEARCH, DEVELOPMENT AND ACQUISITION

U. S. Army Research Office

Report No. 85-2

July 1985

PROCEEDINGS OF THE THIRTIETH CONFERENCE ON THE DESIGN OF EXPERIMENTS

Sponsored by the Army Mathematics Steering Committee

HOST

U. S. Army White Sands Missile Range

White Sands Missile Range, New Mexico

HELD AT

New Mexico State University

Las Cruces, New Mexico

17-19 October 1984

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, North Carolina

FOREWORD

These Proceedings preserve in print many of the invited addresses and contributed papers presented at the Thirtieth Conference on the Design of Experiments in Army Research, Development and Testing. These meetings are sponsored by the Army Mathematics Steering Committee (AMSC) on behalf of the Office of the Chief of Research, Development and Acquisition. Members of this Committee have requested that the guest lecturers be effective researchers who are in frontier fields of live current interest. They feel that the addresses by the principal speakers as well as contributed papers by Army personnel will stimulate the interchange of ideas among the scientists attending said meeting. Noted below is a list of the invited speakers together with the titles of their addresses.

Speakers and Affiliations	Title
Professor John W. Tukey Princeton University and Bell Laboratories	Limited Randomization as the Key to Taking Advantage of Modern Summaries
Professor Roy E. Welsch Massachusetts Institute of Technology	Regression Diagnostics
Professor James Bucklew University of Wisconsin	Quantization
Professor Ulf Grenander Brown University	Recent Work in Pattern Theory
Dr. Ronald L. Iman Sandia Laboratories	Uncertainty Analysis and Sensitivity in Risk Assessment
Professor Bernard Harris University of Wisconsin-Madison	System Reliability

Tne U. S. Army White Sands Missile Range, on 22-24 August 1980, served as host for the Twenty-Sixth Conference on the Design of Experiments, and the meeting was conducted on the campus of its co-host, the New Mexico State University (NMSU) in Las Cruces, New Mexico. This proved to be an ideal arrangement, and members of the AMSC were pleased to hear it could be used for the Thirtieth Conference. This meeting took place on 17-18 October 1984 in the Physical Science Laboratory of NMSU. The attendees would like to thank Mr. Robert A. Voss, the Chairman on Local Arrangements, for all his efforts in planning and conducting this conference.

A couple of days before the start of the Design Conference, a two-day tutorial entitled "The Bootstrap" was presented by Professor Robert Tibshirani of Stanford University. The main purpose of this seminar was to develop, in Army scientists, an appreciation for and the necessary skills needed to handle this statistical method.

Professor Nozer D. Singpurwalla, Department of Operations Research at George Washington University, was selected by the AMSC to receive the Fourth Wilks Award for Contributions to Statistical Methodologies in Army Research, Development and Testing. The citation for his award reads:

> "For singular contributions to reliability theory and life testing methodologies, for professional service to the statistics community, and for invaluable assistance in solving several important testing problems in the Department of Defense."

There are many individuals and things that contribute to the success of these scientific meetings such as the speakers, invited and contributed panelists, the chairpersons, members of the audience, as well as the hosts and their facilities. Sometimes one overlooks the contributions of the members of the Program Committees, so this year the AMSC would like to take this occasion to express its appreciation for their valuable scientific contributions to the stature of these conferences. The names of the committee members for the Thirtieth Conference are:

PROGRAM COMMITTEE

Carl Bates	Carl Russell
Larry Crow	Douglas Tang
Bernard Harris	Malcolm Taylor
Robert Launer	Jerry Thomas
J. Richard Moore	Langhorne Withers

TABLE OF CONTENTS*

*This Table of Contents contains only the papers that are published in this technical manual. For a list of all papers presented at the Thirtieth Conference on the Design of Experiments, see the Program of this meeting.

**This paper was received too late to be placed in the normal order of papers for the Proceedings.

AGENDA

THIRTIETH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING

17-19 October 1984

Host: White Sands Missile Range

Location: Physical Science Laboratory
New Mexico State University
Las Cruces, New Mexico

* * * * * Wednesday, 17 October * * * * *

0815-0915 REGISTRATION

0915-0930 CALLING OF THE CONFERENCE TO ORDER

Robert E. Green, Instrumentation Directorate
The White Sands Missile Range

WELCOMING REMARKS:

0930-1200 GENERAL SESSION I (Auditorium)

Chairman: Robert E. Green, White Sands Missile Range

0930-1030 KEYNOTE ADDRESS

LIMITED RANDOMIZATION AS THE KEY TO TAKING ADVANTAGE OF MODERN SUMMARIES

John W. Tukey, Princeton University and Bell Labs

1030-1100 BREAK

1100-1200 REGRESSION DIAGNOSTICS

Roy E. Welsch, Massachusetts Institute of Technology

1200-1330 LUNCH

1330-1530 CLINICAL SESSION A, PROBLEMS IN MULTIVARIATE ANALYSIS AND EXPERIMENTAL DESIGN

Chairman: J. Richard Moore, Ballistics Research Laboratory

Panelists: George E. P. Box, University of Wisconsin

John W. Tukey, Princeton University and Bell Labs

Stanley L. Sclove, University of Chicago at Illinois

OBJECT CORRELATION IN MULTIPLE SCENES

Robert E. Green, White Sands Missile Range

APPLYING FACTOR ANALYSIS TO REPEATED MEASURES STUDIES

James B. Sampson and Stephen A. Freitas, US Army Natick R&D Center

DESIGN PROBLEMS IN TESTING CHEMICAL PROTECTIVE UNIFORMS FOR PROTECTION AFTER WEAR

Raymond Spring and Jeffrey A. Manickas, US Army Natick R&D Center

1330-1530 TECHNICAL SESSION I, RELIABILITY AND LIFE TESTING

Chairman: Robert H. Turner, White Sands Missile Range

METHODS FOR ASSESSING RELIABILITY GROWTH POTENTIAL

Larry H. Crow, Aberdeen Proving Ground

DETERMINATION OF THE DESIGN ALLOWABLE VALUE USING EXTREME QUANTILE MODELING

Donald Neal, Mark Vangel and Luciano Spiridigliozzi, Army Materials and Mechanics Research Center

BASIC PROGRAMS FOR COMPUTING RELIABILITY AND/OR MEAN LIFE

Donald Rankin, White Sands Missile Range

1530-1600 BREAK

1600-1700 GENERAL SESSION I (Continued) (Auditorium)

QUANTIZATION

James Bucklew, University of Wisconsin

* * * * * Thursday, 18 October * * * * *

0815-1030 TECHNICAL SESSION II, EXPERIMENTAL DESIGN

Chairman: Larry H. Crow, Army Material Systems Analysis Activity

THE APPLICATION OF EXPERIMENTAL DESIGN TO THE EVALUATION OF MULTI-ECHELON STOCKAGE MODELS

Carl B. Bates, US Army Concepts Analysis Agency

LEAST-TIME ANALYSIS: A METHODOLOGY FOR EXPERIMENTAL DESIGN IN LOGISTICS SYSTEMS TESTING AND EVALUATION

Clarence H. Annett, US Army Combined Arms Center

0815-1030 TECHNICAL SESSION III, PHYSICAL MODELING AND TESTING

Chairman: John Robert Burge, Walter Reed Army Institute of Research

COMPARING SHOCK SENSITIVITY OF TWO EXPLOSIVES

Gary W. Gemmill and Audrey E. Taub, Naval Surface Weapons Center

MULTIVARIATE DATA ANALYSIS APPLIED TO SKINFOLD MEASUREMENTS AND THE PERCENTAGE OF BODY FAT FOR BLACK AND WHITE MALE SOLDIERS

Karen P. Hobson, William Beaumont Army Medical Center and Eugene Dutoit, US Army Infantry School

DESERT CAMOUFLAGE PAINT FIELD EVALUATION FOR SAUDI ARABIAN NATIONAL GUARD

George Anitole and Ronald Johnson, US Army Belvoir Research & Development Center

APPLICATION OF THE CONCEPT OF PREDICTION ANALYSIS TO THE EXPERIMENTAL DETERMINATION OF NUCLEAR EFFECTS TEST ENVIRONMENTS

John L. Meason, White Sands Missile Range

0815-1030 TECHNICAL SESSION IV, STATISTICAL INFERENCE

Chairman: Donald Neal, Army Materials and Mechanics Research Center

TETRACHORIC CORRELATION FOR MESOSCALE AREAL PERSISTENCE OF CLOUD CEILINGS

Oscar M. Essenwanger, Red Stone Arsenal

EFFECTS OF BIAS USING COMPUTER PSEUDO-RANDOM NUMBER GENERATOR ALGORITHMS

Raymond Spring and Stephen Freitas, US Army Natick Research and Development Center

NONPARAMETRIC MEDIAN ESTIMATION (WITH APPLICATION TO NUMBER OF SIMULATION REPLICATIONS NEEDED)

James R. Knaub, Jr., US Army Logistics Center

1030-1100 BREAK

1100-1200 GENERAL SESSION II

Chairman: William S. Agee, White Sands Missile Range

RECENT WORK IN PATTERN THEORY

Ulf Grenander, Brown University

1200-1315 LUNCH

1315-1730 SPECIAL SESSION ON RECENT DEVELOPMENTS IN STATISTICS

Chairman: Jagdish Chandra, US Army Research Office

1315-1355 QUANTILE FUNCTIONAL STATISTICAL INFERENCE, UNIFICATION OF DISCRETE AND CONTINUOUS, WORKSHEET DATA ANALYSIS

Emanuel Parzen, Texas A&M University

1355-1435 DAMAGE MODELS

Jayaram Sethuraman, Florida State University

1435-1515 THE DISTRIBUTION OF THE INTENSITY OF ENERGY PENETRATING A MULTINOMIAL FIELD OF RANDOM OBJECTS

Shelemyahu Zacks, State University of New York-Binghamton

1515-1530 BREAK

1530-1610 ON SEGMENTATION OF SIGNALS, TIME SERIES, AND IMAGES

Stanley L. Sclove, University of Illinois at Chicago

1610-1650 APPLICATIONS OF BELIEF FUNCTIONS

Arthur Dempster, Harvard University

1650-1730 EXPERIMENTAL DESIGN IN QUALITY IMPROVEMENT

George E. P. Box, University of Wisconsin, Madison

1830-1930 * * * * * CASH BAR * * * * *

1930-2130 * * * * * BANQUET AND PRESENTATION OF WILKS AWARD * * * * *

* * * * * Friday, 19 October * * * * *

0815-1015 CLINICAL SESSION B, THE ANALYSIS OF SUBJECTIVE DATA AND INDIRECT MEASUREMENTS

Chairman: Carl B. Bates, US Army Concepts Analysis Agency

Panelists: Arthur Dempster, Harvard University

Nozer D. Singpurwalla, George Washington University

DETERMINING COMPLIANCE WITH WEAPON SYSTEM AND EQUIPMENT PERFORMANCE GUARANTEES

Perry C. Stewart, Defense Systems Management College

A PRIORITIZATION METHODOLOGY FOR MATERIEL PROGRAMS

F. Thomas Balzer and Richard T. Maruyama, Hq TRADOC

PROPOSED MODIFICATION OF AFTER HYPOTHESIS TESTING

Paul H. Thrasher, White Sands Missile Range

0815-1015 TECHNICAL SESSION V, PHYSICAL MODELING AND TESTING

Chairman: James R. Knaub, Jr., US Army Logistics Center

FORCE DEVELOPMENT TEST AND EXPERIMENTATION OF THE FIRE SUPPORT TEAM

Jock O. Grynovicki, Ballistic Research Laboratory

A METHOD FOR ESTIMATING DETERMINISTIC WATER WAVES THAT ARE CONTAMINATED WITH RANDOM BACKGROUND NOISE

Michael E. Andrew, Coastal Engineering Research Center

APPLICATION OF A VARIABLE SELECTION TECHNIQUE TO RADAR MEASUREMENT BIAS ESTIMATION

William S. Agee, Robert H. Turner and Andrew C. Ellingson, White Sands Missile Range

1015-1030 BREAK

1030-1230 GENERAL SESSION III

Chairman: Douglas B. Tang, Walter Reed Army Institute of Research

OPEN MEETING OF THE SUBCOMMITTEE ON PROBABILITY AND STATISTICS

UNCERTAINTY ANALYSIS AND SENSITIVITY ANALYSIS TECHNIQUES IN RISK ASSESSMENT

Ronald Iman, Sandia Laboratories

SYSTEM RELIABILITY

Bernard Harris, University of Wisconsin-Madison

1230 ADJOURN

AN INTRODUCTION TO REGRESSION DIAGNOSTICS

Roy E. Welsch

Sloan School of Management

Massachusetts Institute of Technology

ABSTRACT

A regression is constructed using prior knowledge, data, models, and a fitting (estimation) process of some form. It is important to know when the resulting regression depends heavily on a small part of the prior knowledge, on a small part of the data, or on the exact choice of model or fitting process. We want models that are sensitive to the issues of interest, but perform well (i.e., are less sensitive) when assumptions are violated. In this paper we present an overview of the theory, application, and computation of regression diagnostics, especially those related to the analysis of influential data.

INTRODUCTION

Our basic goal in this paper is to learn if our regression is heavily influenced by small subsets of data. A traditional starting point is to look for outliers which may be viewed as observations that appear to be surprising to the investigator, or observations that are not a realization from some target distribution. It is essential that all data used in regression models be examined for outliers. The first step is to look at the response and explanatory variables separately to get a feeling for outliers. At this time, transformations of these variables might be considered. A heavily skewed data series may appear to have lots of outliers. A logarithmic transformation may make the outliers appear much more like the rest of the data. Such transformations are, of course, tentative and need to be considered in light of prior knowledge and subsequent results.

Such a univariate examination does not help us find outliers relative to a particular model and fitting process. In fact, some of the univariate outliers may not look so discrepant in the context of a multivariate model. Conversely, and more commonly, multivariate outliers will arise which cannot be seen in a univariate analysis.

A particularly useful way to detect outliers in the context of a model is to look for overly influential subsets of data. Subsets of data are regarded

as influential if their deletion results in substantial changes to important features of an analysis.

Our discussion of regression diagnostics starts with preliminary steps that are necessary before the decision is made to use a least-squares linear regression model. Then we introduce the idea of adjusted variables and partial regression plots. After a brief discussion of collinearity diagnostics, we define leverage and several different kinds of residuals. We then go on to measure influence and develop plots to summarize influential data diagnostics. We conclude with a brief treatment of diagnostics for generalized linear models and comments about some areas of research in diagnostic methods.

PRELIMINARY STEPS

There are many reasons for performing a regression analysis. Two of the most common are:

(a) fitting an equation or model to data

(b) attempting to describe local averages of y about values of x

$$E(y|x) = g(x). \tag{1}$$

Both of these involve the response data, y, and the regressors X_1, X_2, etc. All too often, the data and (a) and (b) are combined into

$$y = X\beta + \epsilon \tag{2}$$

where X is an nxp matrix of regressors, possibly including the constant carrier, y is nx1, β is px1 and ϵ is nx1. The estimated coefficients, b, are then obtained by a fitting process (usually least-squares) without a great deal of thought.

An important first step is to look at the variables y, and X_1, X_2, etc. separately. These data should be explored using histograms, stem-and-leaf plots, boxplots, etc., (Velleman and Hoaglin, 1981) and granularity (clumps, holes), outliers, and asymmetry ought to be noticed. Outliers need to be tagged, and possible transformations considered. Of course, some outliers may not be so prominent when we consider the multivariate nature of the data. Transformations considered now may also be unnecessary later, but asymmetries, outliers, and large changes of magnitude are clues that some variables may be in the wrong units. In short, take a hard look at the raw data. Do nothing if you wish, but set up a list of things to check as you go further.

The response variable, y, is often supposed to be a random variable with some probability distribution. If it has only two values, looks Poisson, etc., do not try ordinary least-squares regression. You will get stupid results. Consider other models such as the generalized linear models discussed by McCullagh and Nelder (1983). The probability plotting techniques discussed by Chambers et al (1983) are useful for checking these assumptions.

When a thorough univariate analysis has been done, it is time to consider the bivariate (and eventually multivariate) nature of the data. Plots of y versus X_j are always worth making, especially for considering transformations to straighten the plot and computing rough correlations, but can be misleading if used to develop precise models because of the effects of other regressors. Bivariate plots of the regressors are also useful for finding holes, outliers, etc., but the number of plots increases rapidly with p. However, these provide the first clues to the fact that information in the "design" or factor space may be spotty, clumpy, have holes, or be sparse.

Unfortunately, we cannot do this well in higher dimensions--at least not yet. A number of people are working on this (Chambers et al, 1983, Chapter 5).

Usually a tentative model or equation comes with the data. While doing our exploring we should see if the model is sensible. Possible modifications should be noted for later consideration. It is always tempting to make g(x) linear in some proposed coefficients (a so-called linear model). However, g(x) may be quite different--clues to this effect should be noted for they may require different approaches. The model could be nonlinear in the parameters or perhaps a non-parametric approach is needed (Friedman and Stuetzle, 1981). The diagnostics to be discussed below assume that, at least tentatively, the model in (2) is considered reasonable. Diagnostics may open our eyes to further problems, but they cannot take the place of a good preliminary look at the data, structural model (g(x) or, more specifically, $X\beta$) and stochastic model (distribution of y or, in some cases, ε). When diagnostics point to changes we should use these preliminary procedures after making changes and again apply diagnostics to see if we have improved our analysis.

PARTIAL REGRESSIONS AND PLOTS

Most of us are used to thinking about the least-squares estimates as

$$b = (X^T X)^{-1} X^T y. \tag{3}$$

However, it is often more instructive to think about the estimated coefficients in a different way. Denote the residuals found when y is fit by all but the j^{th} regressor by

$$y_{\cdot 12\ldots(j-1)(j+1)\ldots p} = y_{\cdot[j]}. \tag{4}$$

Thus $y_{.[j]}$ is the vector of least-squares residuals obtained by regressing y on all regressors except the j^{th} and is often called the adjusted response variable. Similarly, let $X_{j.[j]}$ denote the residuals obtained by regressing X_j on all of the remaining regressors. These are called the adjusted regressors. It is not hard to show (Mosteller and Tukey, 1977, p. 344) that

$$b_j = \frac{\sum_{i=1}^{n} x_{ij.[j]} y_{i.[j]}}{\sum_{k=1}^{n} x_{kj.[j]}^2} \qquad (5)$$

where $x_{ij.[j]}$ is the i^{th} element of the vector $X_{j.[j]}$ and $y_{i.[j]}$ is the i^{th} element of the vector $y_{.[j]}$. This formula should be compared to that for simple linear regression through the origin. A great deal of information about b_j can be obtained by plotting $y_{.[j]}$ against $X_{j.[j]}$ for each j. These are called partial regression plots (or, in some cases, adjusted variable plots). Useful references are Belsley, Kuh, Welsch (1980) and Chambers et al (1983). Both of these contain interesting examples.

Some properties of these plots are:

(a) The least-squares linear fit to the plotted data has slope = b_j and intercept = 0 (when j is not the intercept variable).

(b) The residuals from the least-squares linear fit are the final multiple regression residuals, y-Xb.

(c) It is relatively easy to see how individual data values influence the estimation of b_j.

(d) Often some information about nonlinearity, heteroscedasticity and unusual patterns can be obtained.

An example of a partial regression plot is given in Figure 1. The simple linear regression line is included and some interesting points have been marked.

Until recently, partial regression plots were thought to be hard to obtain. Velleman and Welsch (1981) show that this is not the case.

Let

$$C^T = (X^T X)^{-1} X^T \tag{6}$$

then

$$b = C^T y \tag{7}$$

and

$$b_j = \sum_{i=1}^{n} c_{ij} y_i . \tag{8}$$

Using the normal equations and (5), we can show that

$$b_j = \frac{\sum_{i=1}^{n} x_{ij.[j]} y_i}{\sum_{k=1}^{n} x_{kj.[j]}^2} . \tag{9}$$

The uniqueness of the least-squares estimates implies that

$$c_{ij} = \frac{x_{ij.[j]}}{\sum_{k=1}^{n} x_{kj.[j]}^2} \tag{10}$$

or equivalently

$$x_{ij.[j]} = \frac{c_{ij}}{\sum_{k=1}^{n} c_{kj}^2} \tag{11}$$

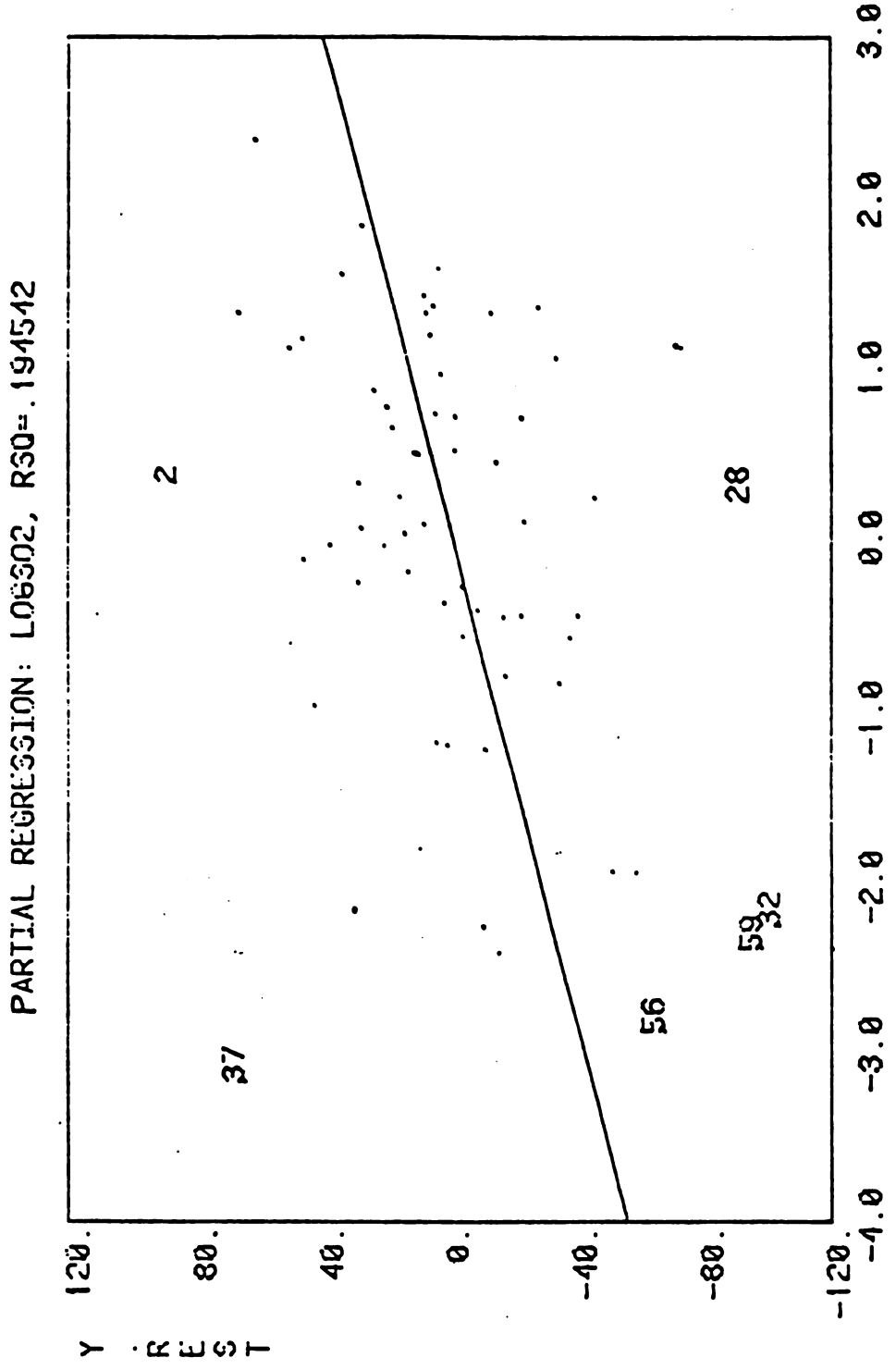
Furthermore, Mosteller and Tukey (1977) have shown that

$$y_{.[j]} = e + b_j X_{j.[j]} \tag{12}$$

where

$$e = y - Xb. \tag{13}$$

FIGURE 1



Therefore b, C, and e are all we need to get partial regression plots. A well-organized regression program can obtain C very easily. There is no excuse for not making these plots a part of every regression analysis. They are an essential diagnostic tool.

COLLINEARITY

Before going further, it is advisable to get some feeling for collinearity. A quick way is to note from (9) that

$$\text{var}\,(b_j) = \frac{\sigma^2}{\sum_{i=1}^{n} x_{ij.[j]}^2} \quad . \tag{14}$$

If the sum in the denominator is small compared to, say, $\sum_{i=1}^{n} x_{ij}^2$, then x_j has been well fit by the other regressors. Since we are used to centering our data, we often compute the squared multiple correlation of X_j on the other regressors,

$$R_j^2 = 1 - \frac{\sum_{i=1}^{n} x_{ij.[j]}^2}{\sum_{i=1}^{n} (x_{ij}-\bar{x}_j)^2} \tag{15}$$

where $\bar{x}_j$ is the j^{th} variable mean. Note that if we are interested in using models with an explicit intercept rather than centered data, the denominator of R_j^2 should not be centered. One statistic often proposed as a measure of collinearity is the variance inflation factor, VIF, found from

$$\text{VIF}_j = \frac{1}{1-R_j^2} \quad . \tag{16}$$

If we just had a simple linear regression on X_j,

$$y = c + dX_j + \varepsilon,$$

then

$$\text{var}(\hat{d}) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2} .$$

Thus

$$\text{var}(b_j) = \text{var}(\hat{d}) \cdot \text{VIF}_j, \qquad (17)$$

and we see that VIF measures the variance inflation due to the presence of additional regressors.

The drawbacks of the above approach (VIF_j large) are that it does not tell us which regressors are involved or how they are involved in the collinear relation with X_j. A very useful way to get this information is described in Chapter 3 of Belsley, Kuh, and Welsch (1980).

If collinearity appears to be a problem, it is wise to reduce it as much as possible before doing influential data diagnostics. When the diagnostics that follow suggest altering or setting aside data it is essential that the new model be rechecked for collinearity.

LEVERAGE

One of the preliminary steps for regression analysis that we have discussed is the making of bivariate plots of the regressors. Generally, the eye will notice outlying points in these plots. However, it is hard to make higher dimensional plots and a variety of tools exist to overcome this

problem. One of the easiest to use measures a type of distance from $\bar{x}$, the row vector of regressor means, to each observation x_i:

$$(x_i - \bar{x})(\hat{X}^T\hat{X})^{-1}(x_i - \bar{x})^T. \tag{18}$$

Here $\hat{X}$ is the X matrix without the intercept column and with column means subtracted off of the remaining columns. Similarly, $(x_i - \bar{x})$ omits the intercept element. Belsley, Kuh, and Welsch (1980) show that

$$h_i - \frac{1}{n} = (x_i - \bar{x})(\hat{X}^T\hat{X})^{-1}(x_i - \bar{x})^T \tag{19}$$

where

$$h_i = x_i(X^TX)^{-1}x_i^T \tag{20}$$

is the diagonal element of the projection or <u>hat matrix</u>,

$$H = X(X^TX)^{-1}X^T, \tag{21}$$

so-called because

$$\hat{y} = Hy . \tag{22}$$

Most modern regression programs now compute h_i and I would not use one that failed to do so.

Often we will want to hypothesize that x_i is possibly erroneous or "strange." Then we may wish to measure the distance from x_i to the <u>rest</u> of the data. One useful way to do this is to compute.

$$d_i = (x_i - \bar{x}(i))(\hat{X}(i)^T\hat{X}(i))^{-1}(x_i - \bar{x}\ (i))^T \tag{23}$$

where $\bar{x}$ (i) and $\hat{X}$ (i) are obtained by assuming that the i^{th} observation did not exist.

Since d_i is related to Mahalanobis distance, it is not hard to show (Belsley, Kuh, and Welsch, 1980) that

$$d_i = \left[\frac{h_i - \frac{1}{n}}{1 - h_i}\right]\frac{n}{n-1} \quad . \tag{24}$$

Both h_i and $h_i/(1-h_i)$ will prove to be useful in what follows.

Hoaglin and Welsch (1978) discuss many properties of h_i. In particular, $0 \leq h_i \leq 1$ (or $1/n \leq h_i \leq 1$ when an intercept is present) and $\sum_{i=1}^{n} h_i = p$ so that the average value is p/n. Let b(i) denote the least-sqaures estimates obtained without using the i^{th} observation. Then simple algebra, e.g. (32), shows that

$$\hat{y}_i = x_i b = (1 - h_i)x_i b(i) + h_i y_i. \tag{25}$$

Thus, $\hat{y}_i$ is a convex combination of the prediction $x_i b(i)$ and the observation y_i. The ratio $h_i/(1-h_i)$ determines the relative contribution of each part. When h_i is one, the i^{th} observation completely determines $\hat{y}_i$ and $\hat{y}_i = y_i$.

There is no general agreement on when h_i is "large." Hoaglin and Welsch (1978) argue that an individual h_i should not be too far from a balanced design (all $h_i = p/n$) and call the i^{th} observation a <u>leverage point</u> when $h_i > 2p/n$ (provided $n > 2p$). Belsley, Kuh, Welsch (1980) show that when the x_i (rows of X) are i.i.d. multivariate Gaussian, the distribution of $h_i/(1-h_i)$ can be related to an F-statistic. This leads to a criterion that calls attention to the i^{th} observation if $h_i > 3p/n$. Note that these leverage criteria depend on p and n.

Huber (1981) uses (25) and suggests that when $h_i > 0.5$, special attention is called for and observations with $h_i > 0.2$ should be noted. These leverage criteria are independent of p and n.

A useful compromise between these two general approaches is to consider h_i, i=1,...,n as a batch of data to be analyzed by exploratory data analysis (Velleman and Hoaglin, 1981). Observations with outlying values of h_i would then be considered leverage points. My own simple rule of thumb is to pay attention when $h_i > \min(0.2, 3p/n)$.

Note that if all of the data is replicated m times, then the new value of h_i is the old value divided by m. Cut-offs that depend on the sample size (such as 3p/n) adjust for this so that replication does not affect those points we determine to be outlying. This seems to us to be a useful property, especially when h_i is related to distance measures.

It is also possible to compute the contribution of the individual regressors to the leverage of each observation. Let

$$h = h_{[j]} + \eta_j \tag{26}$$

where $h_{[j]}$ is the vector of leverage values when X_j is omitted from the regression model. The <u>partial leverage</u>, η_j can be found from

$$\eta_{ij} = \frac{x^2_{ij.[j]}}{\sum_{k=1}^{n} x^2_{kj.[j]}} \tag{27}$$

which is the leverage of the i^{th} point in the partial regression plot for b_j. Data points with large partial leverage for a regressor can exert an

undue influence on the selection of that regressor in most automatic regression model building methods. For some examples, see Henderson and Velleman (1981).

RESIDUALS

While looking for leverage points is a relatively new tool, examining various plots of the residuals, e_i, is not. Surely residuals should be plotted against index (or time), against fitted values, against proposed new regressors (it is best to adjust the new carrier for those already in the model by using the residuals from x_{new} regressed on the current model), etc. Probability plots should also be made. An excellent discussion is contained in Chambers et al (1983).

We feel that the residuals should be properly scaled. Since var $(e_i) = \sigma(1-h_i)$, two useful choices are the internally studentized residual

$$\frac{e_i}{s\sqrt{1-h_i}} \tag{28}$$

and the externally studentized residual

$$e_i^* = \frac{e_i}{s(i)\sqrt{1-h_i}} , \tag{29}$$

where s is the standard error of the regression $(\frac{1}{n-p}\sum_{i=1}^{n}(y_i-x_ib)^2)$ and s(i) is the same but with the i^{th} observation omitted. A simple formula relates s and s(i):

$$(n-p)s^2 = (n-p-1)s^2(i) + \frac{e_i^2}{(1-h_i)} . \tag{30}$$

Under the usual Gaussian error assumptions, e_i^* has a t-distribution with n-p-1 degrees of freedom. If a dummy variable with zero in all positions except for a one in the i^{th} position is added to the current model (X), then e_i^* is a useful diagnostic for seeing if there should be a shift in the intercept for the i^{th} observation. Further details are contained in Belsley, Kuh, and Welsch (1980).

Another form of residual is often useful. The predicted residual is found by computing

$$y_i - x_i b(i) = \frac{e_i}{(1-h_i)} \quad . \tag{31}$$

Hoaglin and Welsch (1978) have noted that when (31) is scaled by its standard error, the result is just e_i^*.

MEASURE OF INFLUENCE

Looking for leverage points and examining various types of residuals form an important step in regression analysis. However, we would like to know if an observation is having a disproportionately large impact on our analysis. An observation is called influential if its deletion would cause major changes in estimates, confidence regions, test and diagnostic statistics, etc. Usually influential observations are outside the patterns set by the majority of the data in the context of a regression model (including the structural model, stochastic model, and fitting procedure). Influential data usually arise from errors in observing or recording data, structural model failure (for example, nonlinear instead of linear) and legitimate extreme observations. Deletion is a way to find procedures to measure influential data. Data should not be

deleted because they are influential, but should be flagged and carefully examined. Alternative fits or forecasts may be needed, one with and one without these data. Judgment or information external to the data will often be necessary.

There are many ways to measure influence. Perhaps the most common is to think of all of the data but the i^{th} observation as "good" and the i^{th} as potentially "strange." We want to find an influence function or measure to see if the i^{th} observation really is a cause for concern. A very useful influence function is

$$b-b(i) = \frac{(X^TX)^{-1}x_i^Te_i}{(1-h_i)} \tag{32}$$

$$= (X^TX)^{-1}x_i^T(y - x_ib(i)) \tag{33}$$

or for each estimated coefficient

$$b_j-b_j(i) = \frac{x_{ij\cdot[j]}(y_i-x_ib(i))}{\sum_{k=1}^{n} x_{kj\cdot[j]}^2} \quad . \tag{34}$$

This can also be stated in terms of (11) rather than adjusted variables.

It is often convenient to scale this measure in some way. Since we are usually interested in changes in the estimated coefficients that are a substantial fraction of the stochastic variability of b, we divide by the standard error of b_j. To estimate the standard error we use $\hat{\sigma}\sqrt{(X^TX)^{-1}_{jj}}$ with $\hat{\sigma} = s(i)$, since we would like an estimate of σ that is not subject to the

"possibly erroneous" i^{th} observation. Other reasons for using s(i) are given in Welsch (1982). All this gives us

$$DBETAS_{ij} = \frac{(b_j - b_j(i))}{s(i)\sqrt{(X^TX)^{-1}_{jj}}} \tag{35}$$

$$= \frac{x_{ij.[j]}}{\sqrt{\sum_{k=1}^{n} x^2_{kj.[j]}}} \cdot \frac{y_i - x_i b(i)}{s(i)} . \tag{36}$$

The first term is the square root of the partial leverage (27) and the second part is related to the predicted residual (31).

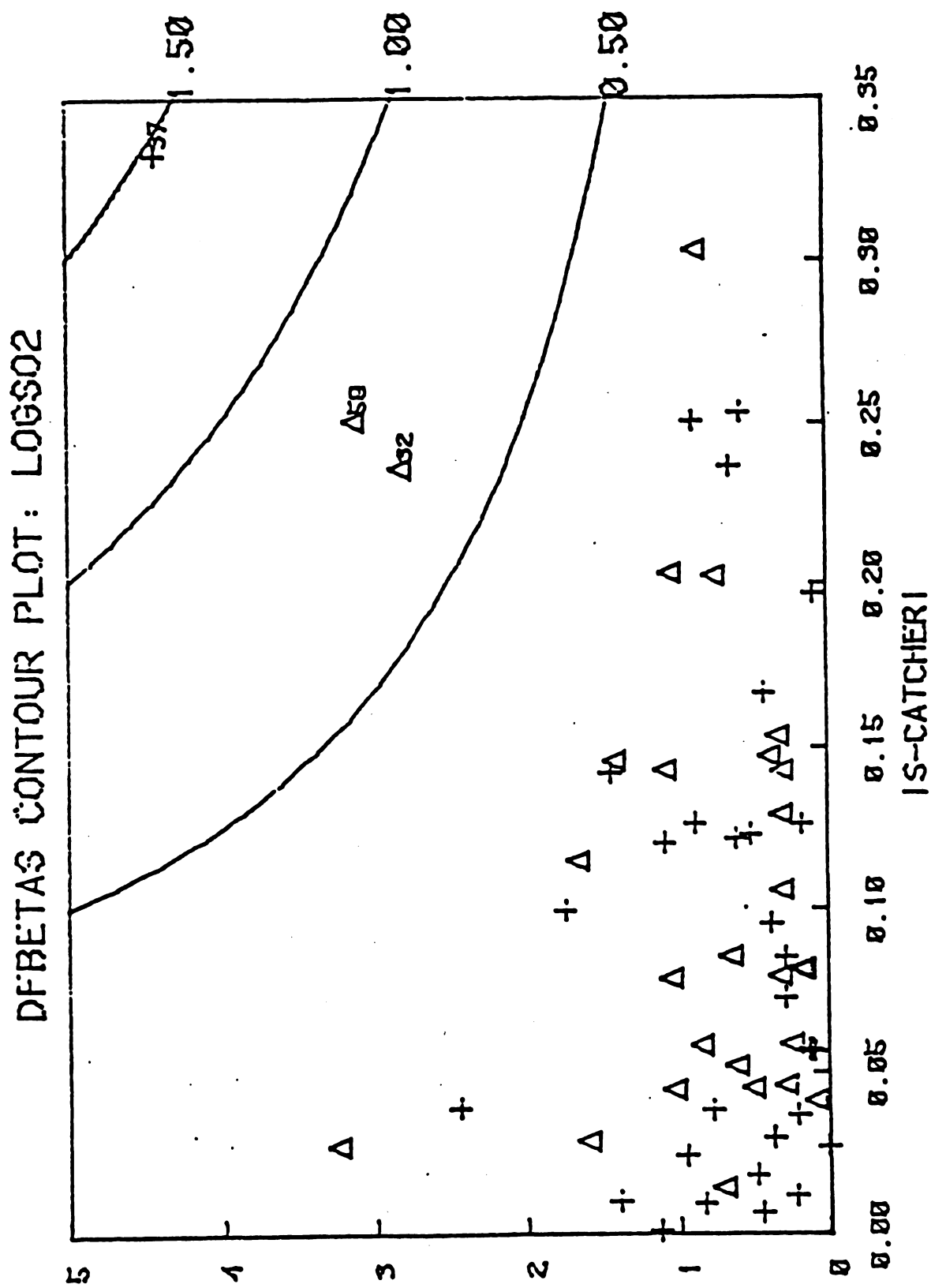
There are three basic ways to decide when |DBETAS| are large. The first is to note which ones are larger than, say, 0.5 or 1. That is, setting aside one observation causes a 0.5 or 1 standard error change in the estimation of β_j. A second method (Belsley, Kuh, and Welsch, 1980) uses the fact that when c_{ij} is constant for all i and the h_i are balanced,

$$\sum_{i=1}^{n} (DBETAS_{ij})^2 \approx 1. \tag{37}$$

When $|DBETAS_{ij}|$ is greater than, say, twice the average value $1/\sqrt{n}$, we take note. A practical rule of thumb is to use min $(0.5, 2/\sqrt{n})$.

The third approach is to look at the DBETAS via exploratory data analysis or contour plots. For a fixed j, $DBETAS_{ij}$ consists of the product indicated in (36). We plot the partial leverage portion on the x-axis and the predicted residual part on the y-axis. Contours of constant influence x.y = c are also plotted. Figure 2 shows such a plot for c = 0.5, 1.0, 1.5, etc. The symbol +

FIGURE 2



denotes a positive DBETAS and Δ a negative DBETAS. Some potentially influential points have been tagged.

Often we are more interested in predictions than coefficients. A prediction is just a linear combination of the estimated parameters, say, $\ell^T b$. It is natural to compare $\ell^T b$ to $\ell^T b(i)$ and scale with a measure of the standard errors of the fit, $s(i)\sqrt{h_i}$. However, we often do not know ℓ so we look for the worst case

$$\sup_{\ell} \frac{(\ell^T b - \ell^T b(i))^2}{s^2(i)\ell^T(X^TX)^{-1}\ell} = \frac{(b-b(i))^T X^T X(b-b(i))}{s^2(i)} \tag{38}$$

$$= \frac{h_i}{1-h_i} \cdot (e_i^*)^2 \quad . \tag{39}$$

From (25) we also note that the difference between the fit, $x_i b$, and the predicted fit, $x_i b(i)$, is just $h_i e_i/(1-h_i)$. When scaled by a measure of standard error of the fit, $s(i)\sqrt{h_i}$, we get

$$\text{DFITS}_i = \frac{x_i b - x_i b(i)}{s(i)\sqrt{h_i}} = \left(\frac{h_i}{1-h_i}\right)^{1/2} \cdot e_i^* \tag{40}$$

which is the square root of (39). Notice that DFITS is the product of a leverage factor (24) and the externally studentized residual (28).

Again, there are a number of approaches to deciding when |DFITS| is large. We can use a fraction of standard error, like 0.5, 1, etc., or note that

$$\sum_{i=1}^{n} \text{DFITS}_i^2 \approx p \tag{41}$$

when $h_i \sim p/n$. We do not want any observation to stray too far from the average influence so we would single out observations with $|\text{DFITS}_i| > 2\sqrt{p}/\sqrt{n}$. A reasonable rule is to use min $(1, 2\sqrt{p}/\sqrt{n})$ as a cutoff. Cook and Weisberg

(1982) develop a statistic similar to $DFITS_i^2$, namely

$$D_i^2 = \frac{1}{p}\left(\frac{h_i}{1-h_i}\right) \cdot \frac{e_i^2}{s^2(1-h_i)} \quad . \tag{42}$$

They suggest that D_i^2 may be considered large when it exceeds $F_{p,n-p}^{(.5)}$ or, approximately, one. This seems to be an unduly conservative cutoff in practice. D_i^2 is also troublesome because it uses s^2 instead of $s^2(i)$ and hence is not robust to errors in the i^{th} observation. See Welsch (1982) for further discussion of D_i^2.

We prefer to look at contour plots with $[h_i/(1-h_i)]^{1/2}$ on the x-axis and $|e_i^*|$ on the y-axis. Constant influence contours may be plotted as before. Figure 3 provides an example.

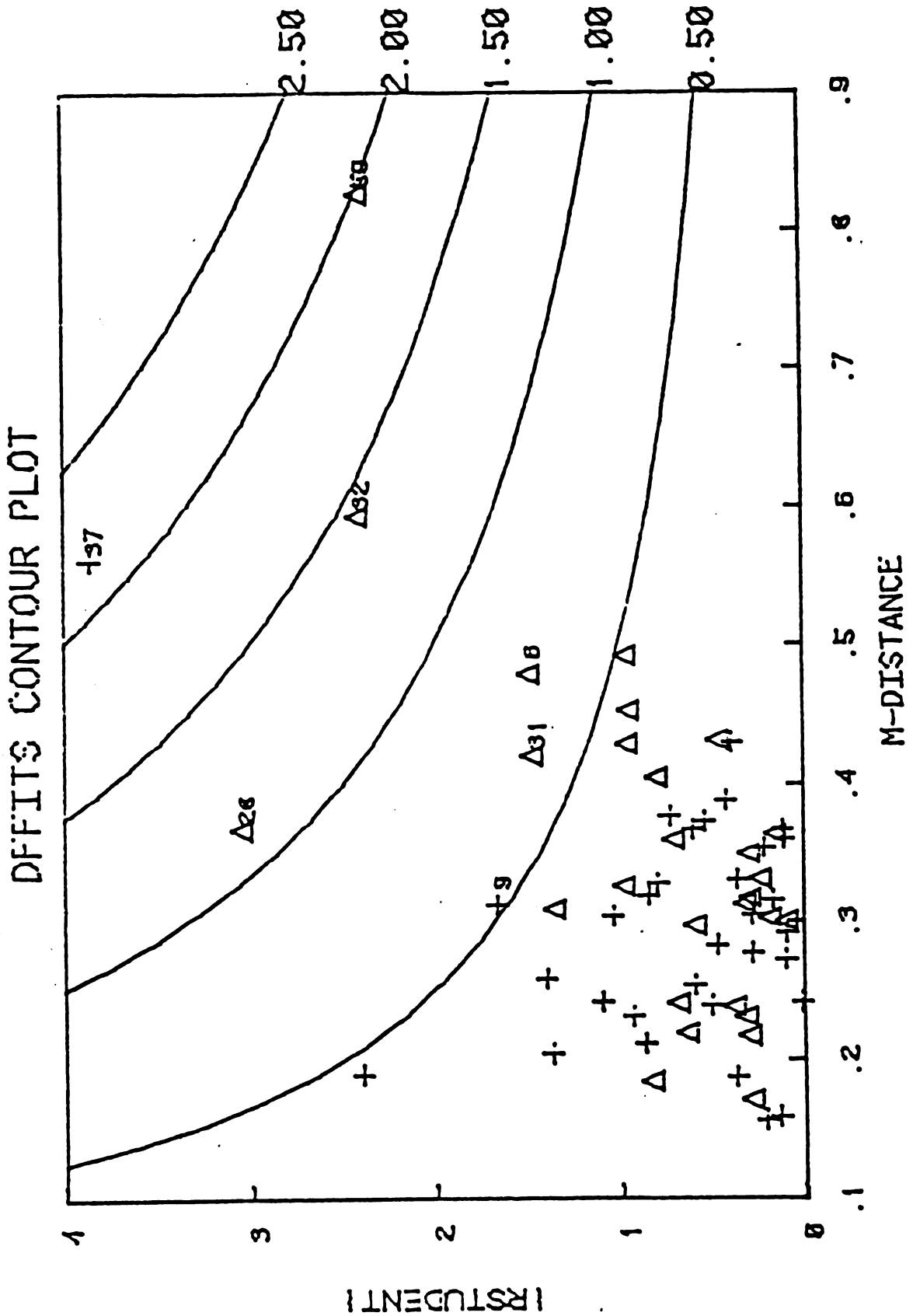
Cut-offs based on sample size adjust for the fact that as sample size increases, variance decreases so that bias caused by erroneous data becomes the major factor in determining the mean square error. To control this bias relative to the decreased variance requires cut-offs that are reduced as sample size is increased.

Often in inference we are interested in confidence intervals or regions. A confidence region consists of a center, a shape, and a scale. For example, in regression a confidence region might be all ß satisfying

$$\frac{(b-ß)^T X^T X(b-ß)}{s^2} \leq a$$

where a is based on the F statistic and p. Here b is the center, X^TX is the shape, and s^2 is the scale.

FIGURE 3



So far we have looked at diagnostics for the center, b. To see what happens to X^TX when the i^{th} observation is considered suspect, we can look at

$$\text{trace } [X^T(i)X(i))(X^TX)^{-1}] = p-h_i \tag{43}$$

or a ratio of volumes

$$\frac{\det[X^T(i)X(i)]}{\det(X^TX)} = 1-h_i. \tag{44}$$

These equations just provide more reasons to look at h_i in its own right.

As for scale, we note again that

$$(n-p)s^2 = (n-p-1)s^2(i) + \frac{e_i^2}{(1-h_i)}$$

so that

$$\frac{s^2}{s(i)^2} \approx 1 + \frac{(e_i^*)^2}{n-p} .$$

Again we have already looked at e_i^* extensively.

These measures can be combined by looking at the ratio of covariance matrix determinants.

$$\text{COVRATIO} = \frac{\det s^2(i)(X^T(i)X(i))^{-1}}{\det s^2(X^TX)^{-1}} \tag{45}$$

$$= [\frac{s(i)}{s}]^{2p} \frac{1}{1-h_i} \tag{46}$$

Contour plots are possible here as well. Velleman and Welsch (1981) note that h_i and e_i cannot vary completely independently since

$$h_i + \frac{e_i^2}{(n-p)s^2} \leq 1 \tag{47}$$

and, of course, when $h_i = 1$, $e_i = 0$.

An important point to note is that observations with large h_i decrease the size of a confidence region while observations with large $|e_i^*|$ increase it. Our goal should be to insure that we are alerted to potentially influential observations. As we can see from the above, influential observations can be both useful and harmful. How we treat them will depend on the purposes of our analysis and their relation to the rest of the data and our models. The best rule of thumb is that there may be more than one good and valid analysis of a data set. Sometimes an analysis with an influential observation and one without are the only way to adequately summarize the data.

GENERALIZED LINEAR MODELS

When the response variable y is Bernoulli, binomial, Poisson, etc., generalized linear models (GLM) are appropriate. A detailed discussion is contained in McCullagh and Nelder (1983). Many of the ideas discussed above can be carried over to these models as well. Basic references are Pregibon (1979, 1981).

The essential idea is to find an influence function (b-b(i)), for the parameters of the model. In the GLM case, this cannot be done exactly since the computation of b requires the solution of a system of nonlinear equations via iterative procedures. However, b-b(i) can be approximated, usually by taking one iteration away from b (the fully iterated solution) with the i^{th} observation removed in an appropriate way. Various kinds of residuals can be defined as well as useful plots.

This is an extremely active area of research at the present time, especially generalizations to survival analysis (Hall et al, 1982), proportional hazards and censured data (Cain and Lange, 1984), Cox models (Storer and Crowley, 1985), matched case control studies (Pregibon, 1984, and Moolgavkar, et al, 1984) and logistic regression (Johnson, 1985). The bibliographies in these papers provide a good overview of work in this area.

INFLUENTIAL SUBSETS OF DATA

If there are two or more outliers in a clump, then influence functions based on setting aside one of the observations will not work well because we will see little change until the entire clump is set aside. The methods discussed above generalize to subsets of data (Belsley, Kuh, Welsch, 1980; Cook and Weisberg, 1982; and Welsch, 1982), but very large amounts of computation are required.

To overcome the computational problems, we have developed a technique called bounded-influence regression (Krasker and Welsch, 1982). A bounded-influence estimator can be viewed as a procedure to find data-dependent

weights (for use in weighted least-squares) so that no small subset of the data is overly influential. The weights and related statistics then become useful diagnostic tools. Examples and associated contour plots are given in Krasker and Welsch (1983). Computational details are discussed in Peters, Samarov, and Welsch (1982).

Another promising approach uses cluster analysis to reduce the computational burden. These ideas are explored more fully in Gray and Ling (1984). Kempthorne (1984) combines clustering with direction searches to attack these problems.

These ideas can be extended to generalized linear models in a number of ways. Some basic references are Samuels (1978), Krasker (1979), Reid (1981), Reid and Crepeau (1985), Pregibon(1982), and Accomando and Pagano (1983). Much more work needs to be done in this area.

COMPUTATION

Computational details of many of the above methods are treated in Velleman and Welsch (1981). They also discuss how to use package programs such as SAS and MINITAB to obtain various diagnostics. The plots used here were made on the TROLL system, a large data analysis and modeling system available under license from M.I.T.

There is no reason why good diagnostics should be omitted from a packaged program. They are essential in my view. We can all demand that they be a part of the new generation of software for personal computers and

workstations. Diagnostics are particularly effective on these devices because graphical tools are readily available.

ACKNOWLEDGEMENTS

The support for this research was provided, in part, by National Science Foundation Grants MCS8116778 and IST8218759 , IBM, and United States Army Research Office Grant DAAG29-84-K-0207.

This paper is a slightly modified version of a paper by the same title in Multiple Regression Analysis: Applications in the Health Sciences. American Association of Physicists in Medecine Symposium Proceedings Number 5, edited by Donald E. Herbert. To be published in 1985. Used by permission.

REFERENCES

Accomando, W.P. and Pagano, M. (1983), "Analyzing a Large Observational Data Set with Categorical Outcomes," Technical Report No. 335Z, Department of Biostatistics, Dana-Farber Cancer Institute, Boston, MA 02115.

Belsley, D.A., Kuh, E., and Welsch, R.E. (1980), Regression Diagnostics, New York: Wiley.

Cain, K.C., and Lange, N.T. (1984), "Estimating Case Influence for the Proportional Hazards Regression Model with Censored Data,", Biometrics, 40, 493-499.

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

Cook, R.D., and Weisberg, S. (1982), Influence and Residuals in Regression, London: Chapman & Hall.

Gray, J. B., and Ling, R.F. (1984), "K-Clustering as a Detection Tool for Influential Subsets in Regression," Technometrics, 26, 305-330.

Friedman, J.H., and Stuetzle, W.Z. (1981), "Projection Pursuit Regression," Journal of the American Statistical Association, 75, 817-823.

Hall, G.J., Rogers, W.H., and Pregibon, D. (1982), "Outliers Matter in Survival Analysis," Technical Report P-6761, Rand Corporation, Santa Monica, CA.

Henderson, H., and Velleman, P.F. (1981), "Building Multiple Regression Models Interactively," Biometrics, 37, 391-411.

Huber, P.J. (1981), Robust Statistics, New York: Wiley.

Hoaglin, D.C., and Welsch, R.E. (1978), "The Hat Matrix in Regression and ANOVA," The American Statistician, 32, 17-22.

Johnson, W. (1985), "Influence Measures for Logistic Regression: Another Point of View," Biometrika, 72, 59-66.

Kempthorne, P.J. (1984), "Identifying Derivative-Influential Groups of Observations in Regression with a Direction Search," Memorandum NS-496, Department of Statistics, Harvard University, Cambridge, MA.

Krasker, W.S. (1979), "Efficient Bounded-Influence Estimation in Logit Models," unpublished paper, December 1979, Department of Economics, University of Michigan, Ann Arbor. (William Krasker is now at the Harvard Business School, Cambridge, MA.)

Krasker, W.S., and Welsch, R.E. (1982), "Efficient Bounded-Influence Regression Estimation," Journal of the American Statistical Association, 77, 595-604.

Krasker, W.S. and Welsch, R.E. (1983), "The Use of Bounded-Influence Regression in Data Analysis: Theory, Computation, and Graphics," Proceedings: Computer Science and Statistics, Fourteenth Symposium on the Interface, New York: Springer-Verlag, 45-51.

McCullagh, P., and Nelder, J.A. (1983), Generalized Linear Models, London: Chapman and Hall.

Moolgavkar, S.H., Lustbader, E.D., and Venzon, D.J. (1984), "A Geometric Approach to Non-linear Regression Diagnostics with Application to Matched Case-Control Studies," Annals of Statistics, 12, 816-826.

Mosteller, F., and Tukey, J. (1977), Data Analysis and Regression, Massachusetts: Addison-Wesley.

Peters, S., Samarov, A., and Welsch, R.E. (1982), " Computational Procedures for Bounded-Influence and Robust Regression," Technical Report No. 30, Center for Computational Research in Economics and Management Science, MIT, Cambridge, MA 02139.

Pregibon, D. (1979), "Data Analytic Methods for Generalized Linear Models," unpublished Ph.D. thesis, University of Toronto.

_____ (1981), "Logistic Regression Diagnostics," Annals of Statistics, 9, 705-724

_____ (1982), "Resistant Fits for Some Commonly Used Logistic Models with Medical Applications," Biometrics, 38, 485-498.

_____(1984), "Data Analysis Methods for Matched Case Control Studies," Biometrics, 40, 639-651.

Reid, N. (1981), "Influence Functions for Censored Data," Annals of Statistics, 9, 78-92.

Reid, N. and Crepeau, H. (1985), "Influence Functions for Proportional Hazards Regression," Biometrika, 72, 1-10.

Samuels, S.J. (1978), "Survival Analysis from the Viewpoint of Hampel's Theory for Robust Estimation," Institute of Statistics Mimeo Series No. 1163, Department of Biostatistics, University of North Carolina at Chapel Hill.

Storer, B.E., and Crowley, J. (1985), "A Diagnostic for Cox Regression and General Conditional Likelihoods," Journal of the American Statistical Association, 80, 139-147.

Velleman, P.F., and D.C. Hoaglin (1981), Applications, Basics, and Computing of Exploratory Data Analysis, Massachusetts: Duxbury Press.

Velleman, P.F., and Welsch, R.E. (1981), "Efficient Computing of Regression Diagnostics," The American Statistician, 35, no. 4, 234-242.

Welsch, R.E. (1982), "Influence Functions and Regression Diagnostics," Modern Data Analysis, edited by R. Launer and A. Siegel, New York: Academic Press, 148-169.

FURTHER READING

Beckman, R.J., and Cook, R.D. (1983), "Outliers," Technometrics, 25, 119-163.

Breiman, L., and Friedman, J.H. (1982), "Estimating Optimal Transformations for Multiple Regression and Correlation." To appear in Journal of the American Statistical Association.

Hocking, R.R. (1983), "Developments in Linear Regression Methodology: 1959-1982, Technometrics, 25, 219-250.

Krasker, W.S., Kuh, E., and Welsch, R.E. (1983), "Estimation for Dirty Data and Flawed Models," Handbook of Econometrics, 1, edited by Z. Griliches and M.D. Intrilligator, North-Holland, Amsterdam, 651-698.

OBJECT CORRELATION IN MULTIPLE SCENES

Robert E. Green
Programs Management Office
Instrumentation Directorate
US Army White Sands Missile Range
White Sands Missile Range, New Mexico

ABSTRACT. Recent changes in the character of missile systems has created the situation where a large number of objects may be located in a relatively small space volume. These randomly located objects are identified by photographing the space volume from two or more different locations. The total number of objects in the space volume is unknown and the number of objects in the intersection of any pair of scenes is unknown. The objective is to identify the objects that are in the intersection of two or more scenes. It is also desired to correlate the objects identified in two or more sequential observations.

I. INTRODUCTION. The purpose of test range instrumentation is to collect data that can be used to evaluate the performance of the object being tested. Photography has been a standard method of collecting performance data since the very early days of missile development. A common method of using photography to collect performance data is to record sequential photographic images from a telescope mounted on a precision tracking mount. The direction of a line connecting the location of the tracking mount and the target is determined by correcting the direction the mount is pointing (azimuth and elevation) by the amount the target is offset from the center of the field-of-view of the telescope. This yields azimuth and elevation angles for the target. If two or more telescope equipped tracking mounts observe a target, then the position of the object in space can be estimated by computing the point that minimizes the sum-of-squares of the distances from the lines defined by the azimuth and elevation angles from each mount.

Recent developments in self-contained munitions have created situations where large numbers of objects are expected to be in the field of view of two or more telescopes. It is possible to estimate the direction of the lines connecting the tracking mount and each target detected in the field of view. A large number of objects in the field-of-view tends to complicate the process of producing data that can be used to evaluate the performance of the objects being tested.

II. Object Correlation. Techniques for performing two types of object correlation are desired. First, an efficient technique for identifying the location in space of the objects in the field of view for a single scene. The images observed in a photograph may not represent a single object since it is possible that two or more objects may share the same direction from an individual mount. Due to differences in aspect angle, range, or pointing direction, the objects photographed from one mount may not all appear in the field-of-view from another mount. An algorithm for this application must consider these factors.

After the objects have been identified at each individual point in time, it is necessary to correlate individual objects as a function of time in order

to estimate velocity. We are given a large number of objects traveling in different directions and at different velocities. We sample the location of these objects at discrete intervals in time. The number of objects identified at each time interval may be incomplete. A procedure is needed that correctly identifies each individual object in each scene for the purpose of estimating its flight performance. Bounds on the expected range of velocities and directions of the objects of interest can be assumed.

III. Summary. Methods that can be used to develop estimates of the flight performance of a large number of objects from time sequential observations of space position as sensed by photo optical techniques is required. These methods need to be sufficiently robust to function when some of the data is incomplete.

AN INTRODUCTION TO FAILURE MODE STRATEGY AND EFFECTIVENESS FACTORS

Larry H. Crow
US Army Materiel Systems Analysis Activity
Aberdeen Proving Ground

Abstract

The growth potential for a system design is defined as the maximum reliability that can be attained for a particular management strategy. The management strategy toward reliability places failure modes into two groups, those that are fixed when seen and those that are not fixed when seen. The management strategy also determines how effective the fixes are. Consideration of the growth potential focuses attention on the impact of not fixing certain failure modes and quantifies the management strategy and engineering effort in terms of attaining the reliability objectives.

In this paper, we discuss a management strategy framework for considering the growth potential and show that a number of other reliability values of interest during a development program can also be addressed within the same framework. These various reliability parameters are all shown to be related and this perspective can be very useful in the planning and management of a reliability program.

Introduction

In the development of a complex system, the initial prototypes will generally have reliability problems. Consequently, the system is typically subjected to a development testing program to find problems and incorporate appropriate corrective actions. This process may involve making reliability predictions and constructing planned growth curves. For the planned growth curve, an initial reliability value must be determined. In addition, such terms as inherent reliability, growth potential, the requirement, and current reliability values all play a role in the design and reliability growth for the system.

In this paper, we give a practical real world structure for putting into perspective initial, inherent, predicted, growth potential, requirement, and current reliability values. All of these values are shown to be related in terms of parameters which are functions of the design strategy, the management strategy for failure modes, and effectiveness factors. When used with reliability predictions and Failure Modes, Effects and Criticality Analysis (FMECA), these results are useful for initializing and constructing planned growth curves, for setting realistic requirements and for developing a viable management approach to reliability.

Failure Mode Strategy

The reliability values of interest in this paper are all directly related to the system design strategy and the reliability management strategy. In this section, we discuss the management strategy in terms of failure mode classification and the effectiveness of the fixes.

Among other things, the reliability management strategy determines what problems seen during test will or will not be fixed. The management strategy also determines how effective the design fixes are and when the fixes are incorporated into the system. Although the management strategy toward reliability growth may not be clearly defined or formally stated, it will in fact exist. The management strategy is determined by how management acts in regard to reliability. The system design and management strategy will determine if it is possible to meet the reliability requirement. The management strategy should be considered early in the development program. In addition, the impact of the management strategy on reliability can be measured from the test data and changed if necessary. Numerical examples are given in a later section for evaluating the management strategy.

When the system is tested and failure modes observed, management can make one of two possible decisions, either not fix or fix the failure mode. Therefore, the management strategy places failure modes into two categories called Type A and Type B modes. Type A modes are all failure modes such that when seen during test no corrective action will be taken. This accounts for all modes for which management determines that it is not economically or otherwise justified to take corrective action. Type B modes are all failure modes such that when seen during test a corrective action or fix will be attempted.

The management strategy, therefore, partitions the system into an A part and a B part. Each part has a corresponding failure rate and mean time between failure (MTBF). See Figure 1.

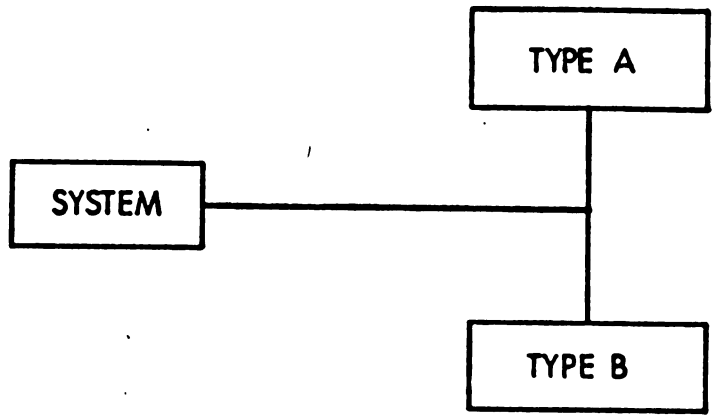


Figure 1. Management Strategy Partitions the System into Two Parts.

Initial Reliability

At the beginning of development testing, the initial system failure rate [S] is

$$[S]_{Init} = [A] + [B]$$

where

[A] - that failure rate due to type A failure modes.

[B] - that failure rate due to type B failure modes.

See Figure 2. The actual value of [S] is determined by the system design. The partition into the A and B parts is determined by the management strategy.

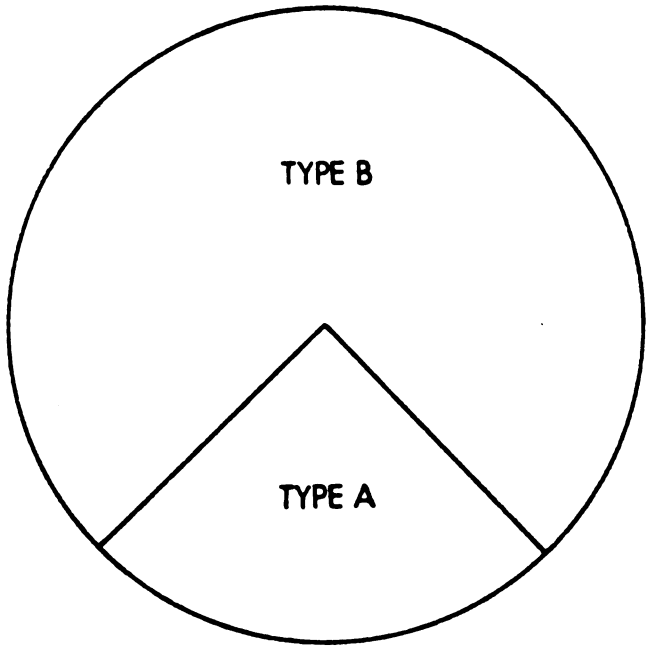


Figure 2. Partition of Total System Failure Rate.

The system MTBF is 1/Failure Rate. The initial MTBF is generally low relative to the requirement and the objective is to achieve reliability growth through finding problems and taking subsequent correction actions.

Current Reliability and Effectiveness Factors

Reliability growth is achieved by decreasing the failure rate [S]. The failure rate [A] for Type A failure modes will not change. With the management strategy, reliability growth can only be achieved by decreasing the Type B failure rate [B]. It is also clear that, in general, we can only decrease that part of the Type B mode failure rate that has been seen during testing. See Figure 3.

It is very important to note that once a Type B failure mode is in the system it is rarely totally eliminated by a corrective action. After a Type B mode is found and fixed, a certain percent of the failure rate will be removed, but a certain percent of the failure rate will generally remain.

A fix effectiveness factor (EF) is the percent decrease in a problem mode failure rate after a corrective action has been made. A recent study by the US Army Materiel Systems Analysis Activity on EFs showed that an average EF d was about 70 percent. That is, on the average about 30 percent, i.e., 1-d percent, of the Type B mode failure rate remained in the system after a corrective action.

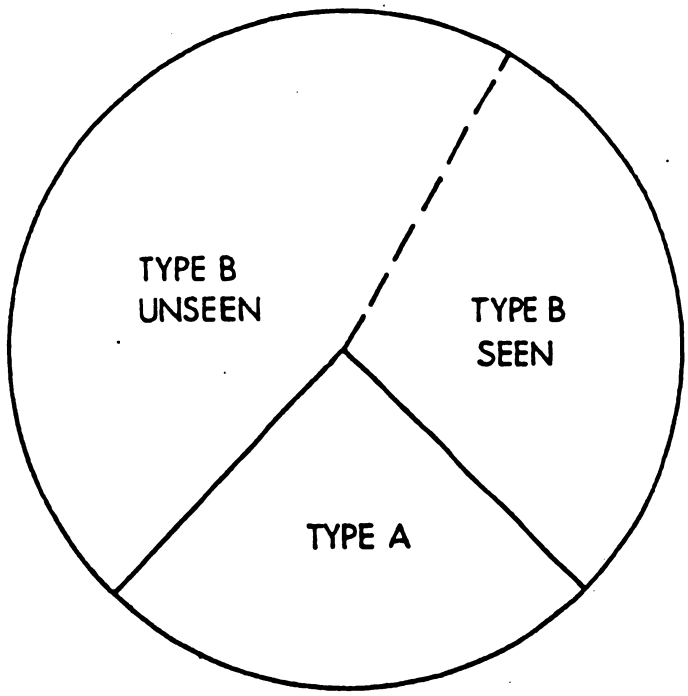


Figure 3. System Failure Rate Partition at End of Test.

Management controls the resources for corrective action. Consequently, fix effectiveness factors are part of the management strategy. For the Type B mode failure rate that has been seen during development testing, we will remove d percent and leave 1-d percent in the system. For illustrative purposes, we will frequently use an average EF d to be 70 percent in this paper. Therefore, after the corrective actions have been made, the current system failure rate consists of the Type A mode failure rate plus the failure rate for the unseen Type B modes, plus 30 percent of the failure rate for the Type B modes which have been seen. See Figure 4.

An important management question is: Can the requirement ever be attained with the way we are doing business i.e., with the management strategy? If the requirement can be met, then usually we also want know how long it will take. This can be answered by addressing the growth rate.

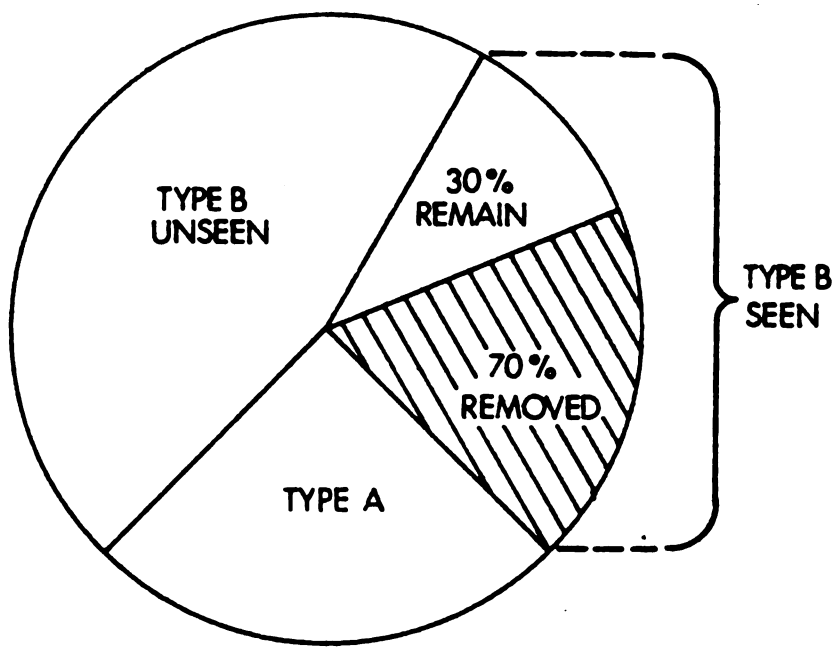


Figure 4. System Failure Rate After Corrective Actions.

Growth Potential and Requirement

The growth potential is the maximum reliability that can be attained with the system design and reliability growth management strategy. The growth potential will have been attained when all Type B failure modes have been found and a fix incorporated into the system. For the system design and management strategy this is the limiting reliability. The growth potential reliability may never actually be achieved in practice.

Figure 5 portrays the growth potential for a system with an average EF of 70 percent. The growth potential failure rate, assuming an average EF of 70 percent, consists of the Type A failure rate plus 30 percent of the Type B failure rate.

The initial failure rate $[S]$ for the system at the beginning of development testing consists of the failure rate $[A]$ for the Type A modes and the failure rate $[B]$ for the Type B modes. That is, $[S]_{Init} = [A] + [B]$.

The growth potential failure rate is the most that the initial failure rate can be reduced with the management strategy. It is the best failure rate attainable. For an average EF of .70, the growth potential failure rate is expressed by

$$[S]_{GP} = [A] + (.30)\ [B]$$

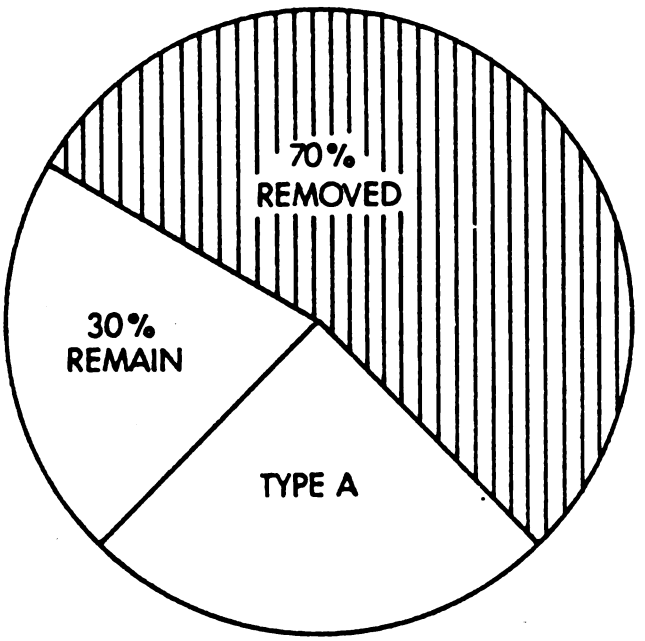


Figure 5. Failure Rate Growth Potential.

The initial MTBF and the growth potential MTBF are shown in Figure 6. An important consideration is whether or not the requirement is below the growth potential MTBF. If the requirement is not below the growth potential, then the requirement cannot be attained with the current system design and management strategy. The methods presented in Ref [2] may be used to easily estimate the growth potential from test data.

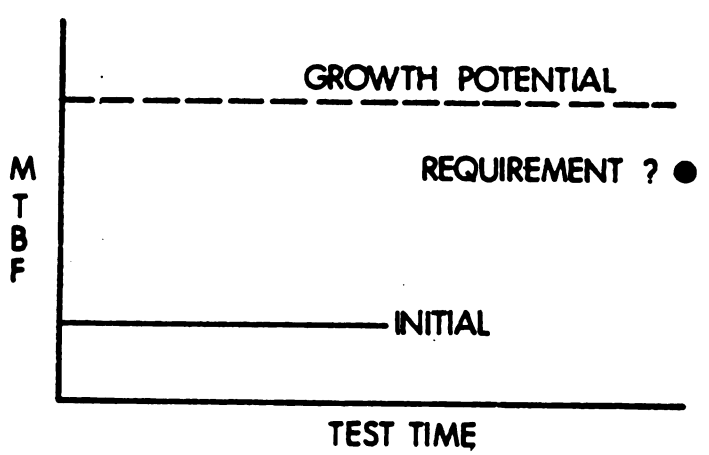


Figure 6. Initial, Requirement and Growth Potential MTBF.

Example

Suppose a system is tested for 400 hours with 42 failures. According to the management strategy, 10 failures are due to failure modes that will not receive a corrective action. That is, these are Type A failures. Also, according to the management strategy, 32 failures are due to failure modes that will receive a corrective action. These are Type B failures. See Figure 7.

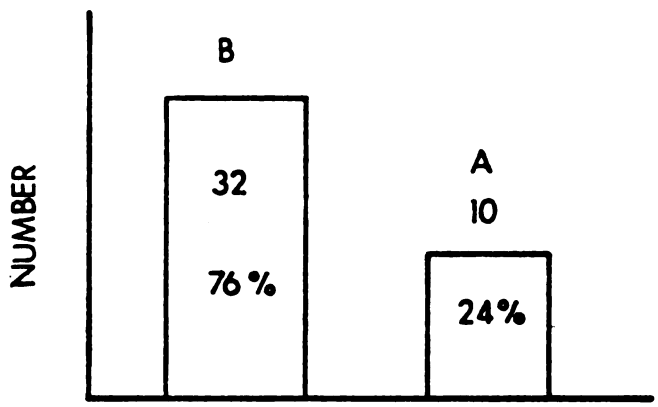


Figure 7. Management Strategy for Type A and B Modes.

We next estimate the initial failure rate and the failure rate and the failure rates for the A and B partition.

$[S] = [A] + [B]$.

The estimates for the Type A and Type B failure rates are the respective number of failures for that type divided by 400, the number of test hours. Consequently with the above management strategy the Type A failure rate is

$[A] = 10/400$

and the Type B failure rate is

$[B] = 32/400$.

The estimate of the initial system failure rate is
$[S]_{Init} = 32/400 = 42/400$ or an estimate of the initial MTBF of 9.5 hours.

If we assume an average EF of 70 percent, then under the above management strategy the growth potential failure rate is estimated by

$[S]_{GP} = [A] + (.30)\ [B]$

or

$\boxed{S}_{GP}$ = 10/400 + (.30) 32/400.

This gives an estimate of the growth potential MTBF of 20.4 hours.

With this management strategy and these data, the initial MTBF of the system is estimated to be 9.5 hours and the very best MTBF that may be attained is estimated to be 20.4 hours. If the requirement is less than 20.4 hours then there is the possibility of it being attained. However, if the requirement is greater than 20.4, say, 25 hours, then it is very unlikely the goal can ever be reached with the present management strategy, regardless of how much testing is conducted. See Figure 8.

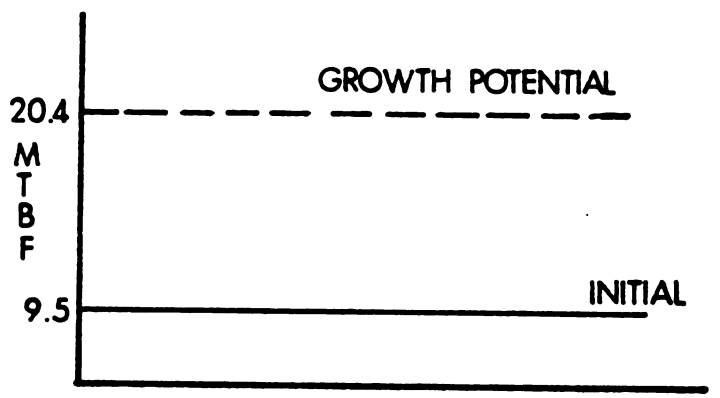


Figure 8. Estimated Initial and Growth Potential MTBF.

If a requirement of 25 hours MTBF is to be attained with the system design, then the management strategy must be changed. This would entail fixing more problems, i.e., increase the Type B modes, and/or increase the effectiveness of the fixes. However, it is noted that an average EF of about 70 percent appears to be typical for many types of systems, although individual EFs may be larger. Therefore, a very large average EF may not be warranted.

Example

Suppose that an average EF of 70 percent is assumed but we desire to change the management strategy so as to increase the growth potential MTBF above 25 hours. Again, suppose that the system was tested for 400 hours with 42 failures. With the new management strategy, 39 failures are due to modes which will receive a corrective action, i.e., Type B modes, and 3 failures are due to modes which will not receive a corrective action, i.e., Type A modes. See Figure 9.

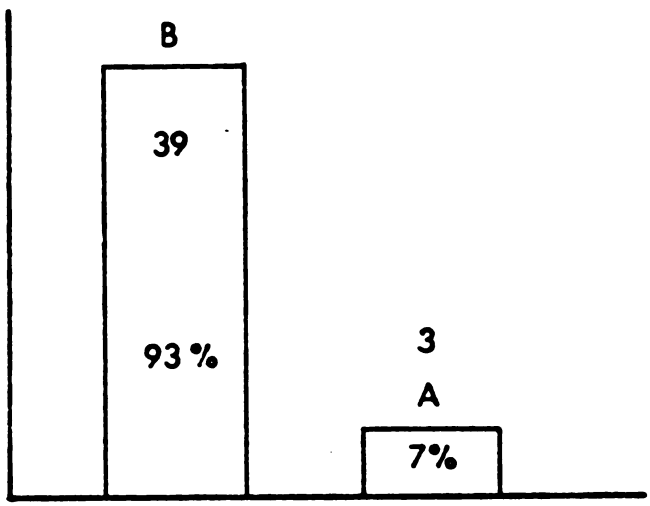


Figure 9. Management Strategy for Type A and B Modes.

The growth potential failure rate for this management strategy is estimated to be

$$[S]_{GP} = 3/400 + (.30)\ 39/400$$

or an estimated growth potential MTBF of 27.2 hours. This is a valid estimate of the growth potential if this management strategy is maintained. This includes classifying, in the long term, at least 93 percent of the failures to Type B modes and no more than 7 percent of the failure to Type A modes. Also, an average EF of at least 70 percent must be achieved.

It is important to note that the growth potential does not estimate the current reliability - it estimates the maximum reliability that will be achieved when all Type B failures modes have been found and fixed by a corrective action. It is showned in Ref [1] that the current failure rate, after corrective action of Type B modes seen during test, equals the growth potential failure rate plus a correction term for unseen Type B modes still in the system. An estimated of the current MTBF utilizing EFs is obtained by applying the projection model and methods discussed in Ref [1] and Ref [2].

Example

Assume again the situation in the first example. In this case, 32 failures will receive fixes at the end of 400 hours of testing. The various failure times for the Type A and Type B modes in this example were generate by computer simulation and discussed in Ref [1]. In Ref [1], these data were used to obtain an estimate of 14.7 hours for the current MTBF. See Figure 10.

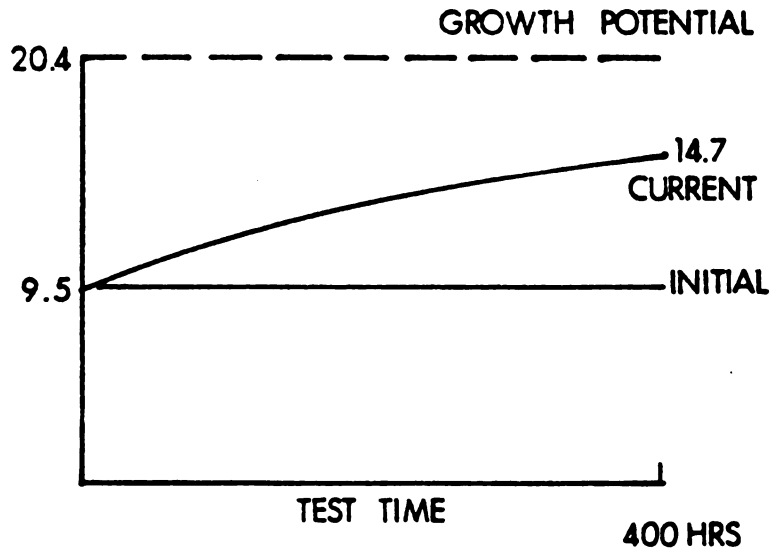


Figure 10. Estimated Initial, Current and Growth Potential MTBF.

Inherent and Predicted

The Type A failure modes are determined by the management strategy. The Type A group accounts for all failure modes for which management determines that it is not economically or otherwise justified to take corrective action. Economically justified is, of course, relative to the design strategy. For example, what is economically justified for one car type may not be for another, depending on the respective design strategies and objectives.

An inherent failure mode will exist whenever the failure rate for that mode cannot be economically reduced further by corrective action. An inherent mode failure rate will remain in the system. Ideally, the group of Type A failure modes should consist only of inherent failure modes. If the Type A group includes modes which are not inherent - i.e., can be economically corrected - then, this is indicative of bad reliability management. Also, if the failure rate for the initial group of inherent modes at the beginning of reliability testing is large relative to the requirement, then this indicates a bad system design for the requirement. When all system failure modes are inherent - cannot be economically reduced further - then the system has attained its inherent reliability.

A valid reliability prediction should generally be expected to address the inherent system reliability. This is very important because of the relationship between inherent reliability and the growth potential reliability. If the management strategy is sound such that all observed failure modes which are not fixed are, in fact, already at their inherent failure rates, and those failure modes which are fixed have EFs that reduce their failure rates to the inherent, then the growth potential failure rate equals the inherent failure rate. This observation can be very useful for developing planned reliability growth curves.

Conclusions

In this paper, we have discussed the initial, current, growth potential, requirement, inherent and predicted reliabilities and their relationships in terms of the design strategy and management strategy for the system. In particular, the initial failure rate is partitioned into two parts.

$$[S]_{Init} = [A] + [B]$$

For an average EF d the growth potential and initial failure rate are related by

$$[S]_{GP} = [A] + (1-d)[B]$$

or

$$[S]_{GP} = [S]_{Init} - d[B].$$

In addition, the current system failure rate, when all Type B modes which have been seen during test have received a corrective action, is related to the growth potential by

$$[S]_{CUR} = [S]_{GP} + \text{correction term for unseen Type B modes.}$$

In term of MTBF, the requirement must be below the growth potential MTBF if it is to be attained. Also, the predicted reliability should generally be expected to address the inherent reliability. Therefore for planning purposes, we may equate the two, i.e.,

$$[S]_{Inh} = [S]_{Pred}.$$

In general, the inherent MTBF will be greater than the growth potential, depending on the management strategy. Under a sound management strategy they will be the same, i.e.,

$$[S]_{Inh} = [S]_{GP}.$$

Therefore, it follows that given a realistic prediction and assuming a sound management strategy, then

$$[S]_{GP} = [S]_{Pred}.$$

These relationships are very useful for developing planned reliability growth curves and for evaluating the realism of attaining the requirement with the management strategy. Figure 11 summarizes these relationships.

References

1. Crow, L.H., Reliability Growth Projection From Delayed Fixes, 1983 Proceedings of the Annual Reliability and Maintainability Symposium, pp.84-89.

2. Crow, L.H., Methods for Assessing Reliability Growth Potential, 1984 Proceedings of the Annual Reliability and Maintainability Symposium, pp.484-489.

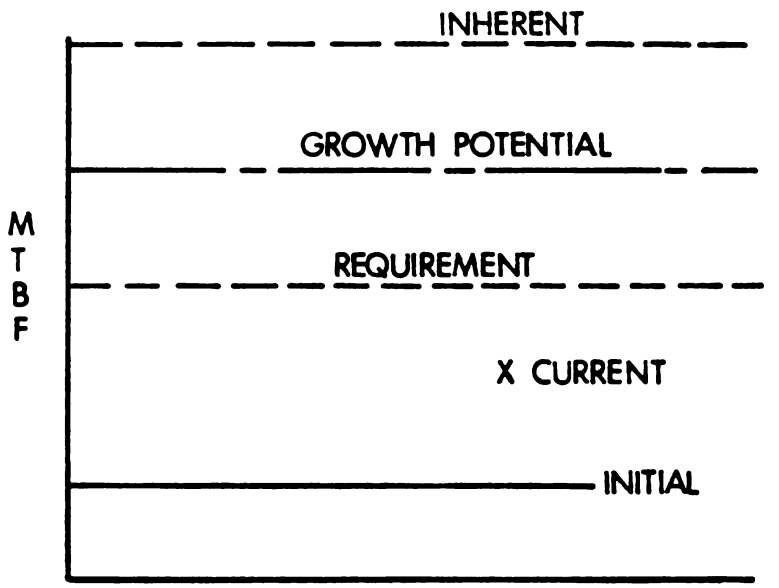


Figure 11. Comparison of Reality Values.

Determination of the Design Allowable Value Using Extreme Quantile Modeling

Donald Neal, Mark Vangel, and Luciano Spiridigliozzi
Army Materials and Mechanics Research Center
Watertown, Massachusetts 02172

ABSTRACT

This paper describes a methodology for obtaining the A and B material design allowable values using an extreme quantile modeling process. The allowables represent a value determined from a specified probability of survival with 95 percent confidence in the assertion. The survival probabilities are .99 for the A and .90 for the B allowable. The required representation of the small portion of the data in the left hand tail of the distribution was obtained from the following models: the two parameter Weibull, two parameter exponential and the Bootstrap method.

Development of the exponential model is presented in detail. The primary effort involved determining an unbiased estimator of the LCB for the exponential sample order statistic corresponding to the required quantile value. This was obtained in terms of the LCB of the same ordered value from a uniform distribution and the MLE of the exponential parameters.

Applying the Weibull censored data analysis presented in Lawless [1] provided an effective method for weighting the lower ordered values where at least seventy-five percent right censoring was considered. The Bootstrap [2] method was applied in order to obtain the variance at the one and ten percentile

values. In the case where data was limited or contaminated an extrapolation process involving an extreme value model was introduced. Excellent correlation was obtained from among all three models.

INTRODUCTION

The inability to obtain exactly the same structural properties from all specimens obtained from a manufactured material results in a relatively large variability in strength measurements when a large number of specimens are considered. In the case of designing an aircraft structure it is required to design such that a minimum stress value exists in critical locations and these values do not exceed the minimum guaranteed material properties (strength). Obtaining minimum strength values will reduce possibility of some production components containing weaker material than that from the laboratory test element. This guaranteed minimum strength value is defined as the design allowable by aircraft design engineers.

Usually the measured value is considered acceptable in estimating the population parameters for predicting population percentiles. In the case of the design engineer it is advisable to have a prediction which will determine the accuracy of the percentile estimate at a high degree of statistical confidence. This is the correct interpretation of an allowable value. For example, certain military standards, e.g., MIL-HDBK-5[3] require material property data to be presented on an A or B allowable basis. A and B allowables are defined by the probability statement providing a 95 percent confidence of the assertion that probability of surviving the A allowable value is <.99 and <.90 for the B allowable.

AMMRC is involved in the development of a statistical chapter for the MIL-17 HDBK on composite material in aircraft structural design. The chapter will include methods for determining the design allowable values. The inability to identify the proper statistical model from limited or multi-modal data motivated the authors to find new robust models.

This paper presents original methods for obtaining an accurate measure of the above mentioned design allowables involving lower tail modeling. Breiman, Stone and Gins [4] have discussed the difficulties existing in model identification when very small tail probabilities are required. This is the result of parameter estimates that usually are obtained from data in the central portion of the distribution where most failures occur leaving the tail region limited in representation. This is unfortunate, since the relatively small amount of data in the tail region is of prime importance to the allowable computation. The lower tail modeling presented in the paper avoids central region modeling by modeling lower ordered values of the distribution. The exponential and Bootstrap methods involve truncating the tail. The Weibull model uses a censoring process for the high ordered strength values.

When small samples or multi-modality prevents reasonable model identification it is necessary to either apply conventional non-parametric methods or devise some scheme similar to those advocated by the authors. The primary difficulty in the extreme quantile modeling technique involved determining tolerance bounds on the quantile values in the allowable computation.

The Exponential Extreme Quantile LCB Estimator

Breiman, Stone and Gins introduced the exponential extreme quantile model which is the foundation of our research. The method for estimating the LCB of an extreme quantile presented here, however, was developed by the authors. A Monte Carlo investigation of the original (Breiman,et.al.) LCB estimator disclosed a bias of about 3% when sampling from an exponential distribution. Since an exponential distribution satisfies the extreme quantile model exactly, it is desirable that the LCB estimator be unbiased in this situation. This paper presents a new LCB estimator which is unbiased when the underlying distribution is exponential.

Consider a sample of n independent, identically distributed random variables from the two parameter exponential distribution

$$F(X;\tau,a) = \begin{cases} e^{-(\tau-X)/a} & 0 \le X \le \tau \\ 0 & X > \tau \end{cases} \tag{1}$$

Define the lower tail as

$$\{X_{(1)}, X_{(2)}, \ldots, X_{(m)}\}, m \le n/2,$$

where $X_{(j)}$ is the jth sample order statistic. Assuming that $F(0;\tau,a)$ is approximately zero, the joint density of the lower tail values may be approximated as

$$f_{X_{(1)},\ldots,X_{(m)}}(X_1,\ldots,X_m) = \frac{n!}{(n-m)!}a^{-m} e^{-\sum_{1}^{m}(X_m-X_i)/a} \cdot e^{-m(\tau-X_m)/a}(1-e^{-(\tau-X_m)/a})^{n-m} \tag{2}$$

for $\tau > X_{(m)} > \ldots > X_{(1)}$. Based on the lower tail values, the maximum likelihood estimators of the exponential parameters are

$$\begin{cases} \hat{a} = \frac{1}{m-1}\sum_{i-1}^{m-1}(X_{(m)}-X_{(i)}) \\ \hat{\tau} = X_{(m)} + \hat{a}\ln(n/m) \end{cases} \tag{3}$$

where a has been weighted with m-1 rather than m in order to yield an unbiased estimator. The maximum likelihood estimator of the qth quantile $\hat{X}_q$, $0 < q < 1$, is

$$\hat{X}_q = \hat{\tau} - \hat{a}\ln(1/q) = X_{(m)} - \hat{a}\ln(m/nq). \tag{4}$$

This is obtained by solving $F(X_q;\tau,a)$ for X_q and replacing parameters with their MLE's. The lower 95% confidence bound on X_q with q =.1 and q=.01 determines the B and A design allowables respectively.

If the quantile of interest corresponds to a sample order statistic, then the true lower confidence bound on the quantile may be obtained by applying the inverse probability transform to the LCB of the corresponding order statistic from a uniform distribution; it follows that for this situation the true allowable may be readily obtained.

In what follows, we have restricted attention to the B allowable, i.e. the lower 95% tolerance limit on the 10% point. Consideration of the A allowable (tolerance on the 1% point) may be made in a similar manner. Further, we have considered primarily samples of size 10k for k integer, since in this case the required limit corresponds to the LCB on the kth order statistic.

Let $X_1,\ldots,X_n$ be iid $F(\cdot;\tau,a)$ as in equation 1. The density of the kth order statistic is

$$f_{X_{(k)}}(t;\tau,a) = \frac{n}{a}\binom{n-1}{k-1} e^{-(n-k+1)(\tau-t)/a}\,(1-e^{-(\tau-t)/a})^{(k-1)} \qquad (5)$$

and the true B allowable may be expressed as

$$\int_0^{X_b} f_{X_{(k)}}(t;\tau,a)\,dt = .05. \qquad (6)$$

with change of variable

$$\begin{cases} u = e^{-(\tau-t)/a} \\ u_b = e^{-(\tau-X_b)/a} \end{cases} \qquad (7)$$

equation 6 can be written as

$$\int_0^{u_b} n\binom{n-1}{k-1} u^{n-k}(1-u)^{k-1}du = .05. \qquad (8)$$

The upper limit on the integral is therefore the lower 5% point of the Beta(k,n-k+1) distribution, i.e.:

$$\sum_{s=0}^{k-1}\binom{k-1}{s}u_b^s\,(1-u_b)^{k-s} = .95. \qquad (9)$$

The allowable estimator

$$\hat{X}_b = X_{(m)} + \hat{a}\ln(10u_b m/n) \qquad (10)$$

may be obtained from equation 7 by replacing parameters with their MLE'S.

Since the MLE's are unbiased in this case, we have derived an estimator which will provide unbiased allowable estimates when the underlying distribution is exponential. this estimator is computationally very simple; the percentiles of the beta cdf (equation 9) are easily obtained on a computer. We have, for ease of reference, called this new estimator the reduced bias estimator, or RBE.

Monte Carlo Investigation of Reduced Bias Estimator

A Monte Carlo study using the two parameter Weibull distribution was performed in order to provide a preliminary assessment of the RBE. Since the problem of determining allowables is particularly acute for small samples, we

considered sample sizes of 10, 20,and 30, with tail sizes of 3, 6 and 10 respectively. The Weibull shape parameters (α) considered are in the range $2 < \alpha < 100$, a range of values consistent with those expected from test results on most materials. This investigation is summarized in Figure 1.

The exponential tail allowable estimator is a linear combination of the sample order statistics. For the order statistics of a Weibull sample, the mean and covariance matrix may be determined exactly [5]. Unfortunately, the covariance matrix is quite complex, therefore a simulation was used to estimate the variance of the allowable estimator from Weibull samples.

The exact expression for the mean of the kth order statistic is

$$E(X_{(k)}) = \Gamma(1+1/\alpha)\cdot\beta\cdot n\binom{n-1}{k-1}\sum_{s=0}^{k-1}\binom{k-1}{s}\left(\frac{(-1)^s}{(n-k+s+1)}\right)^{(1+1/\alpha)} \qquad (11)$$

Although equation 11 appears to be relatively simple to evaluate, machine roundoff error prevented computation of the sum. the reason for the difficulty is that equation 11 is proportional to the (k-1)st backward difference of $(1/n)^{(1+1/\alpha)}$ that is

$$\nabla^{k-1}((1/n)^{(1+1/\alpha)}) = \sum_{s=0}^{k-1}\binom{k-1}{s}\left(\frac{(-1)^s}{(n-k+s+1)}\right)^{(1+1/\alpha)} \qquad (12)$$

where ∇ is defined by

$$\nabla f(x) = f(x-1) - f(x) \quad .$$

The difference operator may be expressed in terms of the derivatives of the function and Stirling numbers of the second kind [6].

The results of the Monte Carlo study are presented in Figure 2. The estimated bias is expressed as a percentage of the allowable estimate. As a measure of the variance, the estimated 90%

confidence interval is also given as a percentage of the estimate. These results indicate that the bias is unacceptably large for small shape parameter but decreases sharply as the parameter increases. This is also true for the confidence interval as a percentage of the estimate. The mean square error decreases with increasing sample size.

In Figure 3, the estimated B allowable for the Weibull distribution is plotted vs the the shape parameter. Note the nearly vertical slope of the curve for small α. The estimated allowable curve agrees closely with the true curve over the entire range; the distance between the curves measured along a normal is always small. When measuring the bias and confidence for small α, we are cutting the curves obliquely in a region of rapid change in the allowable with increasing shape parameter.

In order to compare the RBE method, a standard nonparametric estimator was applied - the first ordered value from a sample of 30. That is, if the unspecified continuous underlying distribution is $F(\cdot)$, then

$$P(X_{(1)} \leq X_{.1}) = P(F(X_{(1)}) \leq F(X_{.1})) = P(F(X_{(1)}) \leq .1)$$

$$= 1 - .9^{30} = .958 \qquad (13)$$

Since minimal assumptions are made with respect to distribution, this nonparametric estimator generally is conservative. In Figure 4, the RBE bias and confidence for a sample of 30 are repeated, along with the corresponding exact results for the first ordered sample value. Except for very small shape parameter, the proposed estimator has much smaller mean square error than the other nonparametric estimator.

Determining a Tail Truncation Point

A problem to be addressed is the choice of a tail truncation point (m value). In the simulation just described the tail is

approximately 1/3 of the data. This choice was made subjectively before the analysis began. To examine the sensitivity of bias to m value, the expression for the exact expected order statistics (equation 11) was used to calculate the expected bias in the RBE for the sample sizes and Weibull populations of figure 1. These results are presented in Figure 5 a, b, and c for sample sizes 10, 20 and 30 respectively. These plots show considerable fluctuation in bias for small shape parameters. The bias decreases steadily with increasing modulus, in good agreement with the Monte Carlo results. From Figure 5, it appears that the optimal tail lengths are 3, 6, and 8 for sample sizes of 10, 20, and 30 respectively. For small shape parameters, a longer tail greatly reduces the expected bias. Figure 5d is a magnification of Figure 5c: it illustrates the expected bias for a sample of size 30 and a modulus of less than 10. If one takes a tail of 12 points, the expected bias may be reduced to 5% or less for $3 < \alpha < 10$. A more complete treatment of the tail truncation problem will be presented in a forthcoming AMMRC report.

Allowable Estimate From the Bootstrap Method

The Bootstrap method used in determining B design allowable values involves letting $\hat{F}$ be the empirical distribution of observed values (strength test results),i.e. the probability distribution with mass 1/n for each X value (X_1, X_2, $X_3 \ldots X_n$). The Bootstrap sample is obtained by selecting a random set of n new values, independently with replacement from $\hat{F}$ (note, some values will be repeated once, twice, etc). From the ordered set of the Bootstrap sample the 10 percent point (j ordered value $X_{(j)}$ obtained consistent with the 90 percent survival number is used for the allowable computation.

N Bootstrap samples are obtained with corresponding $X_{(j)}$ value (10 percent points). The resulting sequence $X^1_{(j)}$, $X^2_{(j)}$, $X^3_{(j)}, \ldots X^N_{(j)}$ estimates the distribution of the 10 percent point of F. The 5 percent point of the Bootstrap samples results

in the B allowable estimate of F. N usually is greater than 500. If sample size is small, preventing adequate representation of percentile values relative to the Bootstrap process, then a smoothing method is suggested. Currently, the authors are applying the 2 parameter Weibull tail model to the extreme quantile region of the empirical distribution. A schematic of this procedure is shown in Figure 6. Applying the regression model as shown in the figure will provide generation of any sample size including extrapolation to lower quantile values (eg. .01 for A allowable). The smoothing techniques has been very effective in smoothing tail region data that contains contaminated data (eg. outliers and bi-modality). The flexibility of the Weibull model provides a good tail representation.

Weibull Censored Data Allowable Estimate (Lawless)

The method developed by [1] for obtaining tolerance bounds for Weibull extreme quantile values was applied to censored data. The data was right censored in some instances by 80 percent in order to preserve a homogenous lower extreme quantile region. This scheme is extremely effective if two or more modes existing in the data and only data from lowest mode is modeled with remainder of data censored - (See Figure 7). The B allowable of 1.30 agrees with non-parametric solution while complete sample solution of 1.06 is too conservative. This technique is equally effective for unimodal data.

Results and Discussion

In Figure 8a the complete ranked sample is displayed. The allowable estimates for the functional representation are tabulated at right of figure. The Weibull, normal and lognormal agree quite well for the B allowable estimate of 44.0 KSI. The tail region is not adequately represented by the three functions,

therefore reliance on the 44.0 KSI value could result in an unsatisfactory B allowable estimate. The third ordered number of 44.0 KSI represents 90 percent survivability value exclusive of the tolerance bound. This indicates an obvious overly non-conservative allowable estimate.

Figure 8b displays the Weibull censored data result for the complete sample shown in Figure 8a. Within the figure is tabulated the Design B 38.8. This number represents the B alllowable estimate agreeing with the conventional non-parametric solution of 38.5 which is the first ordered value when sample size is 30. The reliability of the non-parametric solution is usually good except when there is limited amount of dispersion in the data. In this case, the results will be overly conservative. The 12 percent coefficient of variation for the sample in Figure 8a is acceptable, therefore the non-parametric method is acceptable and consequently the censored data result. The direct modeling of only the first four ordered values from the sample of 30 with 26 censored number has provided surprisingly accurate allowable estimates. The censored data method has consistently provided accurate results for most data sets for arbitrary tail sizes greater than 4.

The ranked data of 18 values is displayed in Figure 9. Note, the obvious bi-modality existing in the data, a result limiting selection of an adequate model for obtaining the allowable value. The results from conventional functional representation Weibull, normal, and lognormal are 43.0, 42.3, and 42.8 respectively. These values appear to be non-conservative in that the 90 percent survival value is 43.0 not including reduction due to the tolerance estimate related to the sample size. The obvious poor representation of the tail region has resulted in this error.

In the tabulation to the right of Figure 9, the results of censored data solution, Bootstrap method and the exponential model are shown to be 38.2, 36.1, and 36.5 respectively. The agreement among the three techniques is quite good. Only the first mode (first five ordered values) were modeled except for the Bootstrap process. Finally, the non-parametric [7] result is

included in order to show lowest possible allowable value considering all models. The result is somewhat conservative due to the limited sample size..

Conclusions

Three methods for estimating lower confidence bounds on extreme quantiles have been discussed in this paper: an exponential tail estimator developed by the authors, an application of the Bootstrap to extreme quantiles, and a censored data method due to Lawless. All of these methods are based on the idea of estimating confidence on an extreme quantile from a tail model, as opposed to a model involving the complete sample. Consequently, all three are potentially useful in situations where the tail appears well behaved but the underlying distribution is in doubt.

The method of Lawless and a Bootstrap procedure were introduced briefly into the presentation only for comparison purposes. Lawless's method has consistently provided good results in obtaining the allowable values. At present, considering all available models it is the "best". The primary drawback of this censored data method is its computational complexity. Preliminary results from the Bootstrap method have been good but inconclusive. More effort is needed in developing this method for the allowable application.

The exponential tail estimator was presented in detail. This new estimator for confidence on extreme quantiles has the advantage of simplicity and intuitive appeal. A Monte Carlo study for small samples from Weibull populations revealed acceptable bias except in the case of small shape parameters. Preliminary investigation of tail truncation points suggests that modeling a longer tail reduces expected bias and variance in the case of small Weibull modulus.

There is general agreement among the three extreme quantiles models in determining allowable estimates. This suggests that these models can provide effective alternatives to inadequate conventional models where model uncertainty exists.

References

1. Lawless, J. (1975) "Construction of Tolerance Bounds for the Extreme Value and Weibull Distributions ", Technometrics, 17, 255-261

2. Efron, B. (1979)" Bootstrap Methods: Another Look at the Jackknife", Annals of Statistics, 7, 1-26

3. Military Standardization Handbook, Metallic Materials and Elements for Aerospace Vehicle Structures, Volumes I and II, MIL-HDBK-5B, 15 August 1974, Available from Superintendent of Documents, Government Printing Office, Washington, DC 20013, USA

4. Breiman, L., Stone C.J., Gins, J.D. (1981), "Further Developments of New Methods for Estimating Tail Probabilities and Extreme Value Distributions", Final Report on U.S. Air Force Office of Scientific Research Contract No. F49620-80-C-0037, Technology Services Corporation, Santa Monica, California

5. Lieblein, J. (1955), "On moments of Order Statistics from the Weibull Distribution", Annals of Mathematical Statistics, 26, 330-333

6. Hildebrand, F.B. (1974), Numerical Analysis, McGraw-Hill: New York

7. Billings, G.B.,(1967) "Improving Binomial Reliability Estimates - A moderately Distribution Free Technique For Small Sample Reliability Estimation", Proceedings of the Twelfth Conference on the Design of Experiments in Army Research, Development, and Testing (ARO-D Report 67-2)

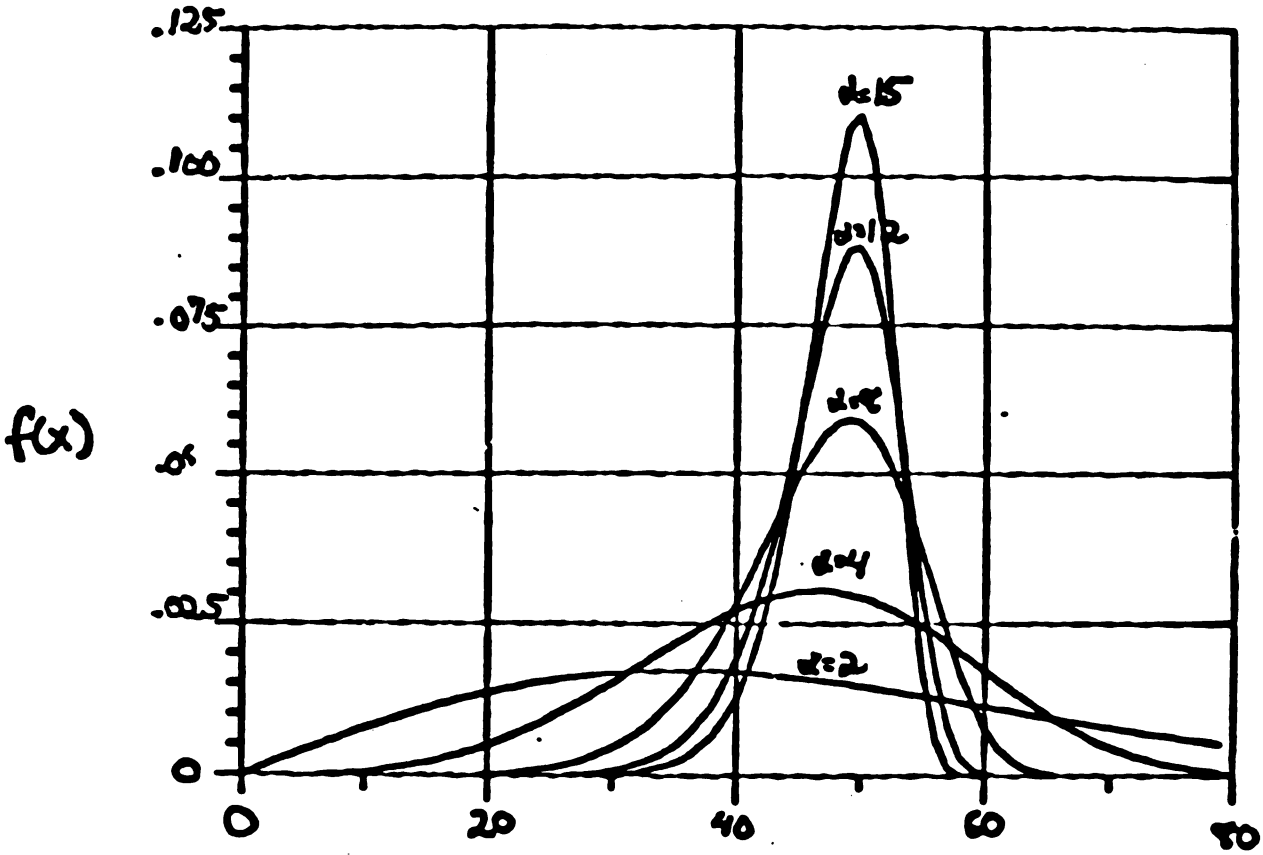
Acknowledgement

To Sharon F. DeMarco of AMMRC for her excellent work in the preparation of this manuscript. To E.M. Lenoe of AMMRC for supporting this research work.

FIGURE I

Monte Carlo Simulation of Reduced Bias Estimator

For the Weibull Distribution



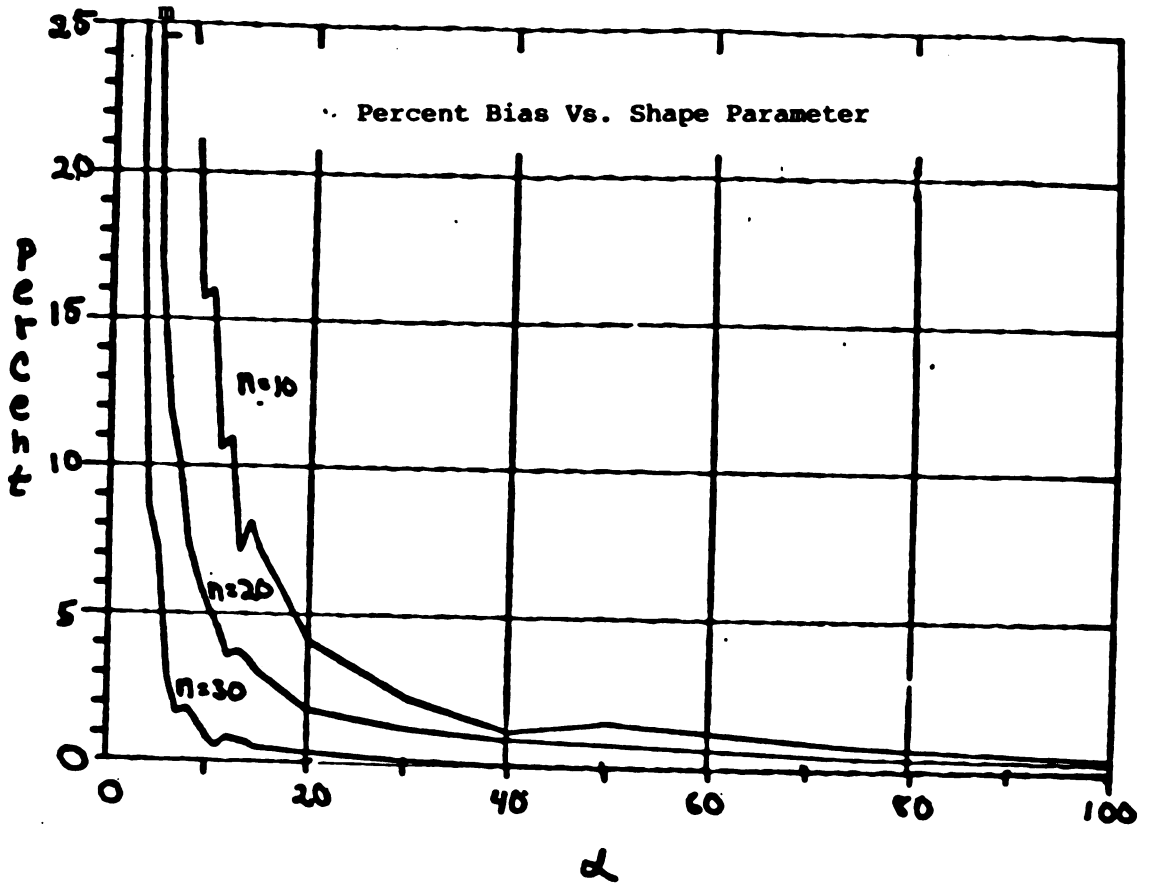
$$f(x) = (\alpha/\beta)^{\alpha} x^{\alpha-1} e^{-(x/\beta)^{\alpha}}$$

$\beta = 50$
$\alpha = 2,3, \ldots, 15,20,30,40,50,75,100$
$n = 10,20,30$
$m = 3,6,10$

2500 simulations were performed for each (n,α) combination.

Figure 2

Monte Carlo Estimation of Bias and Variance (RBE)



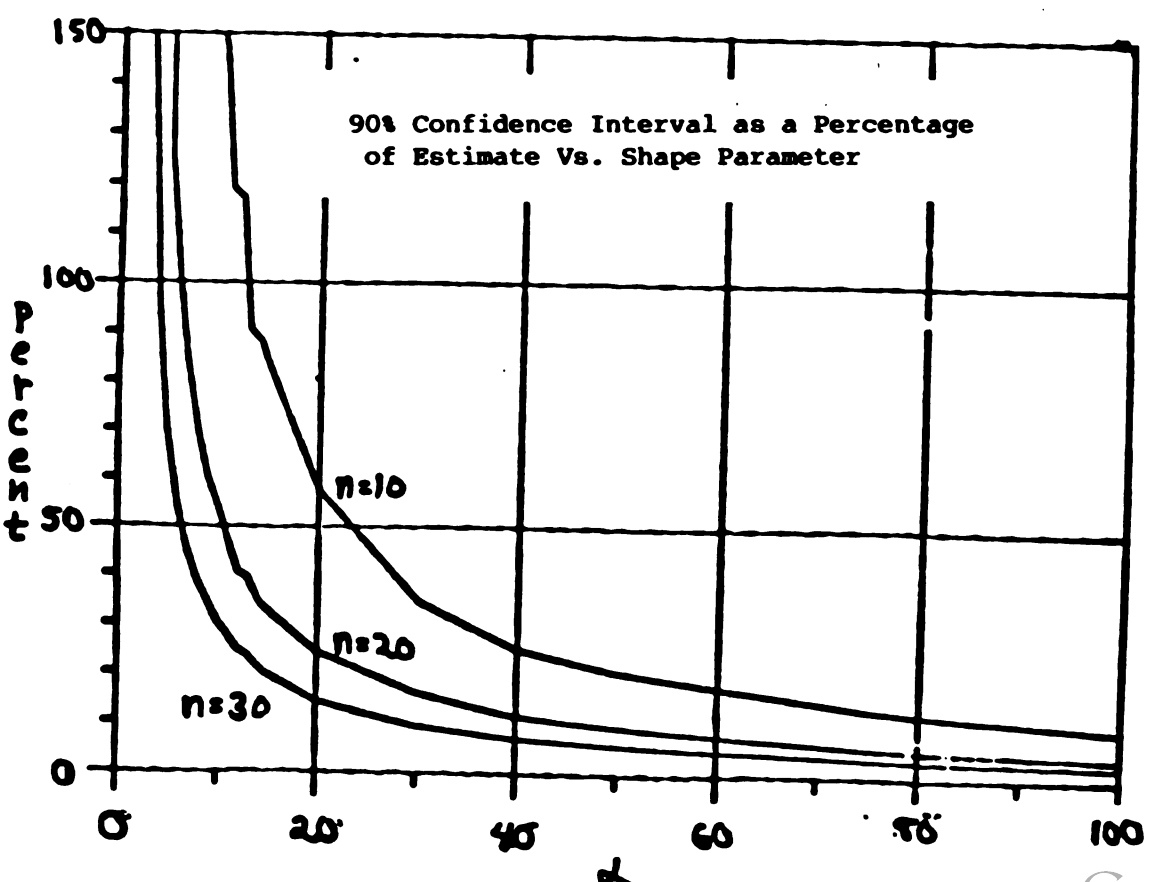


Figure 3

Exact 'B' Allowable, Estimated Allowable and 90% Confidence Interval for N=30

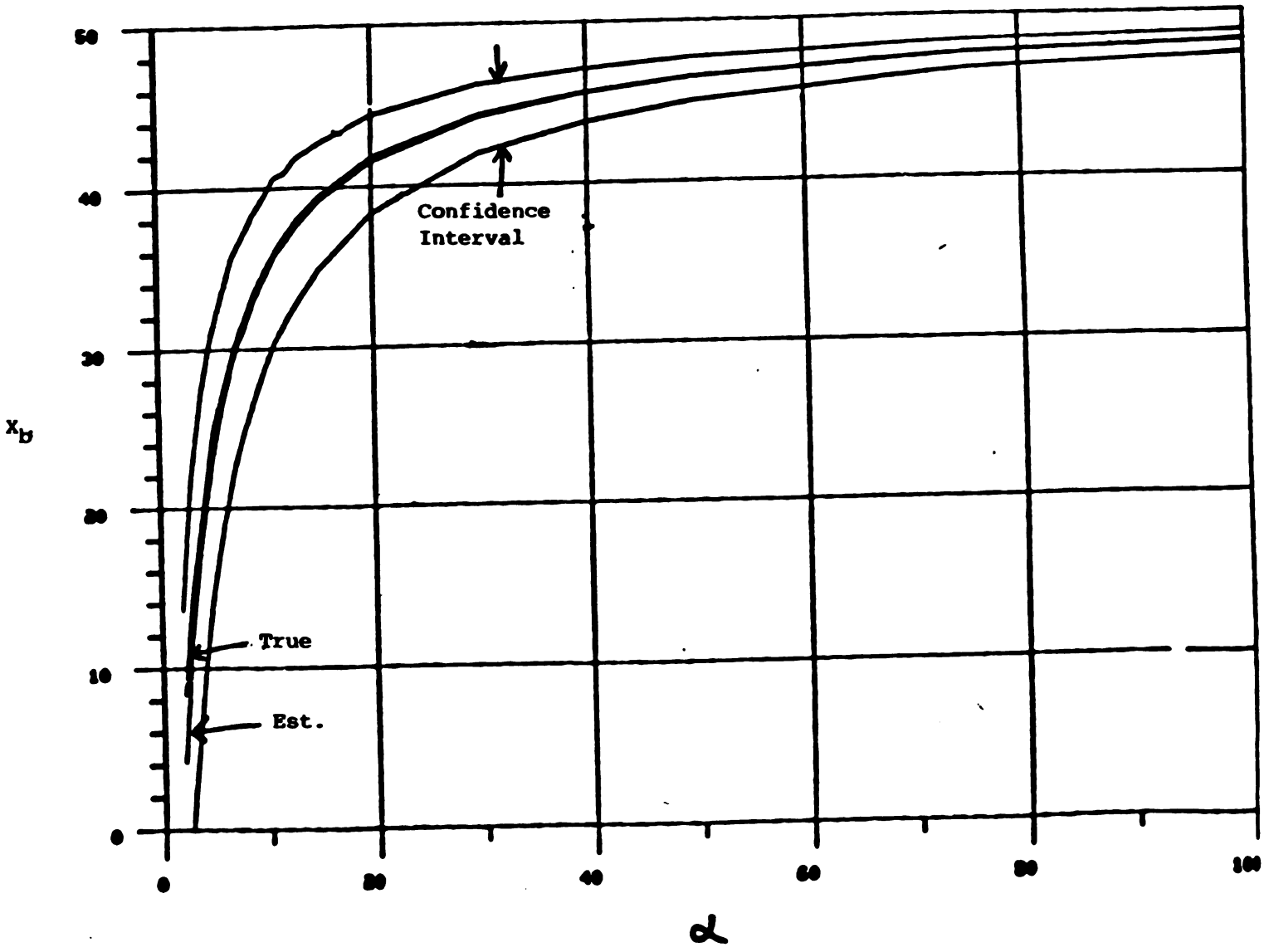
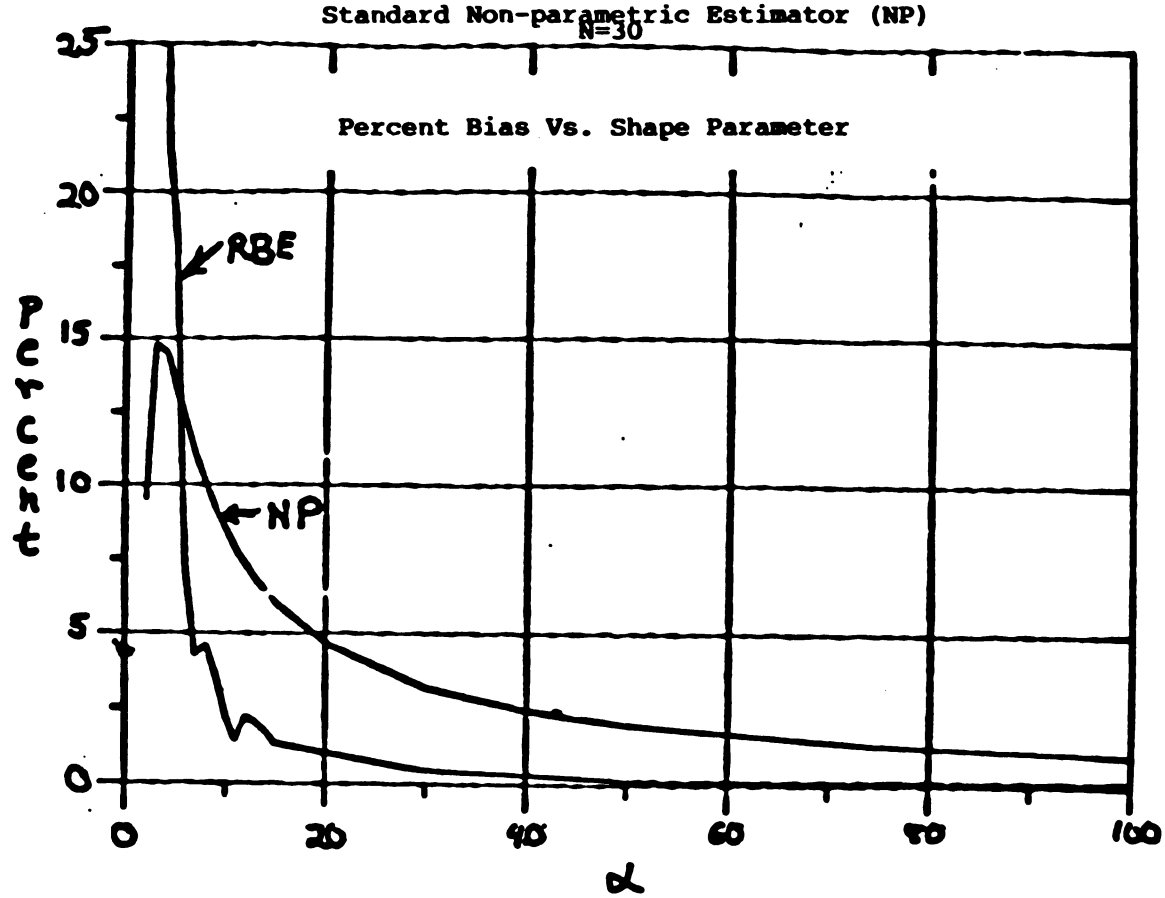


Figure 4

Comparison of Exponential Tail (RBE) and Standard Non-parametric Estimator (NP)



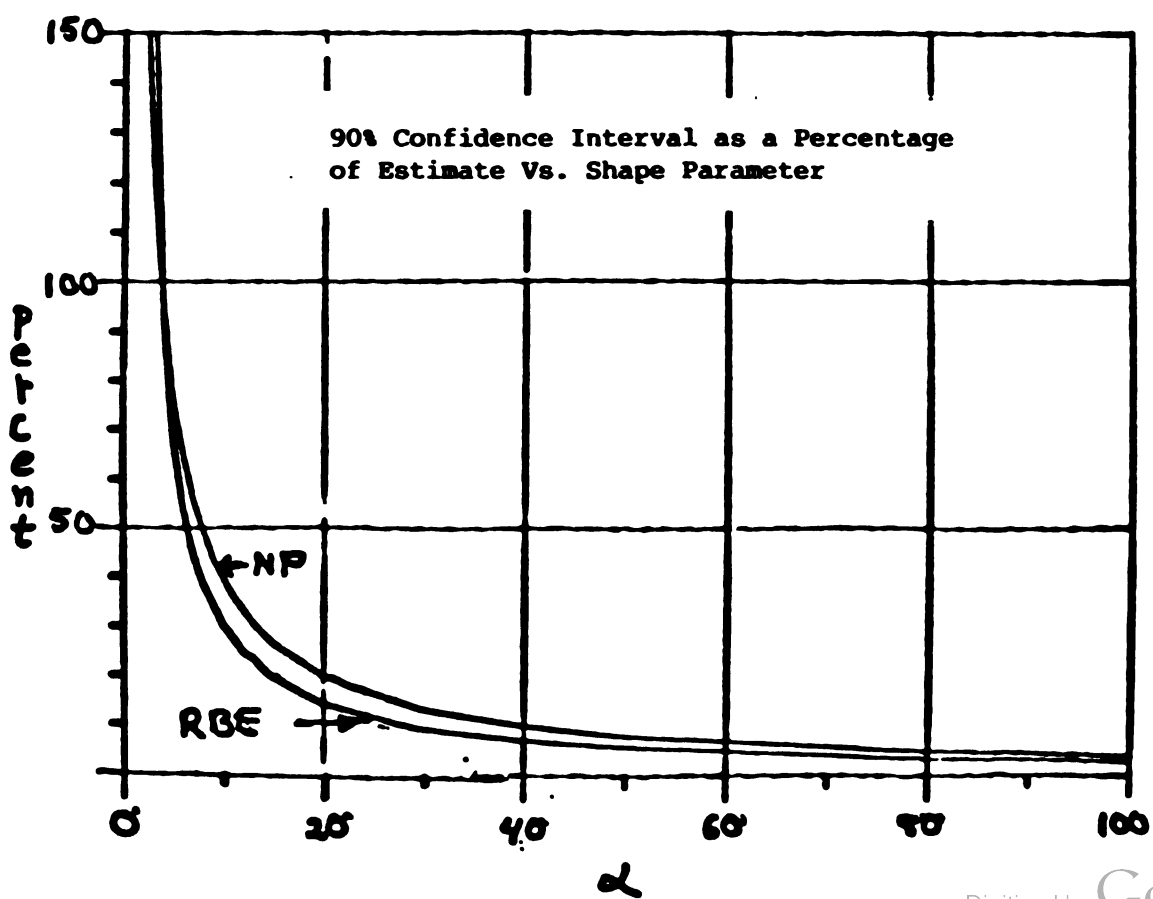
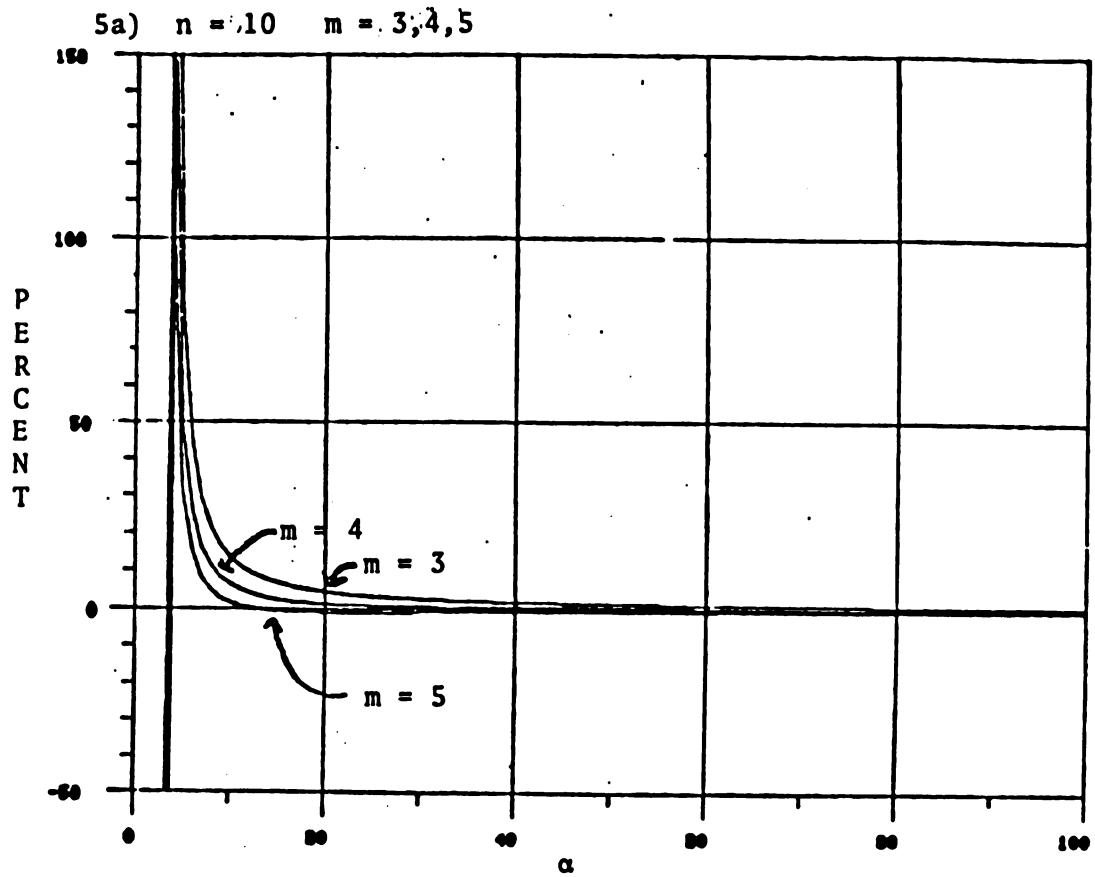


FIGURE 5

RBE Percent Bias Vs. Shape Parameter

$F(x) = 1-e^{-(x/50)^{\alpha}}$ m = tail length



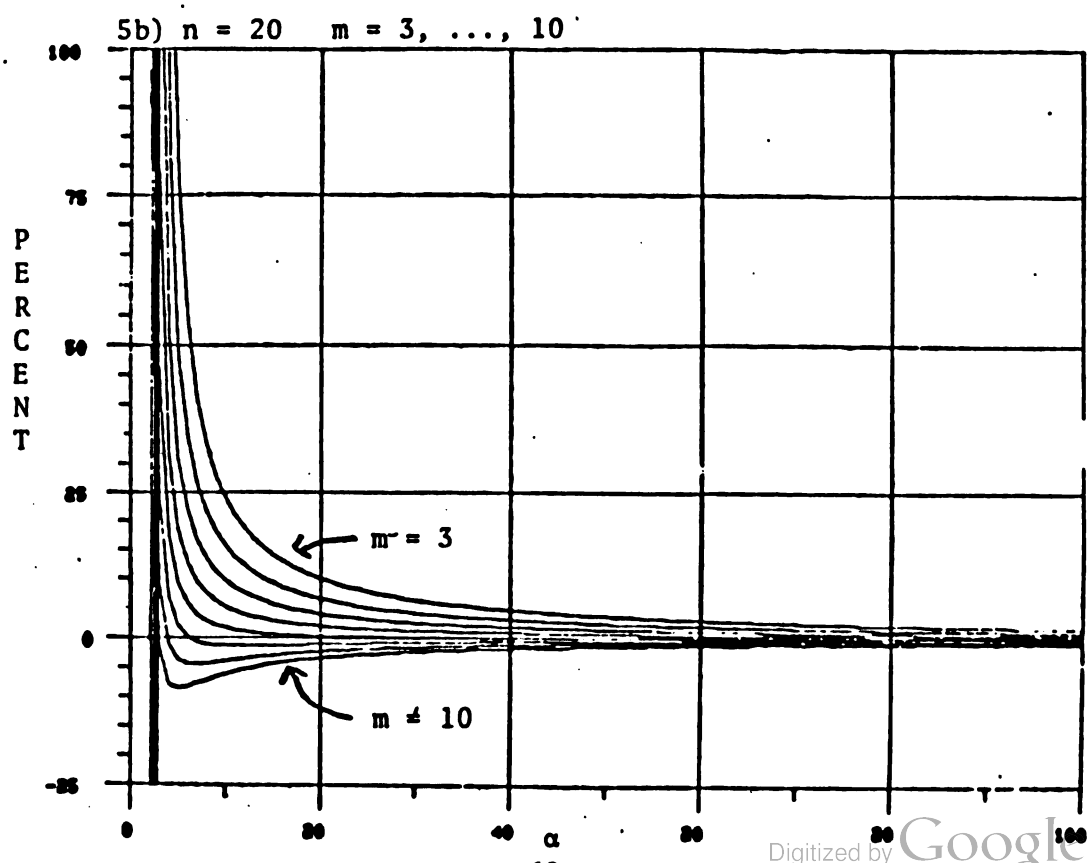
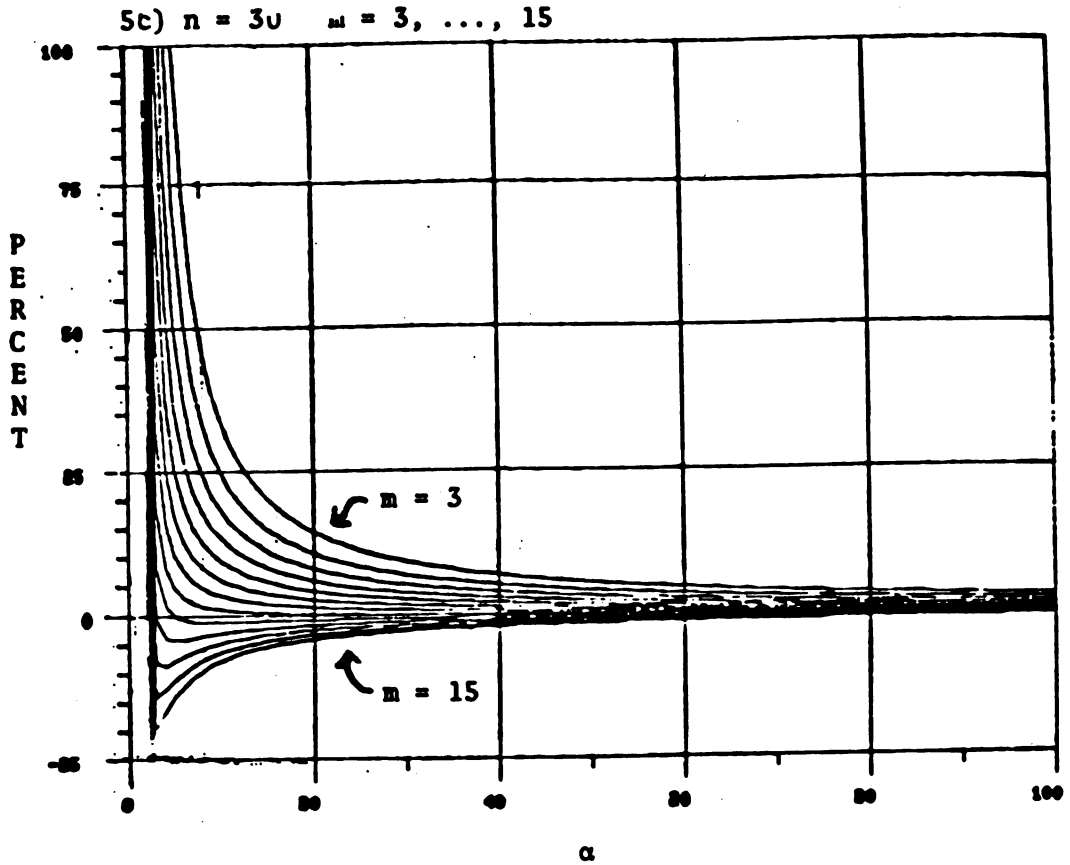


FIGURE 5 (continued)



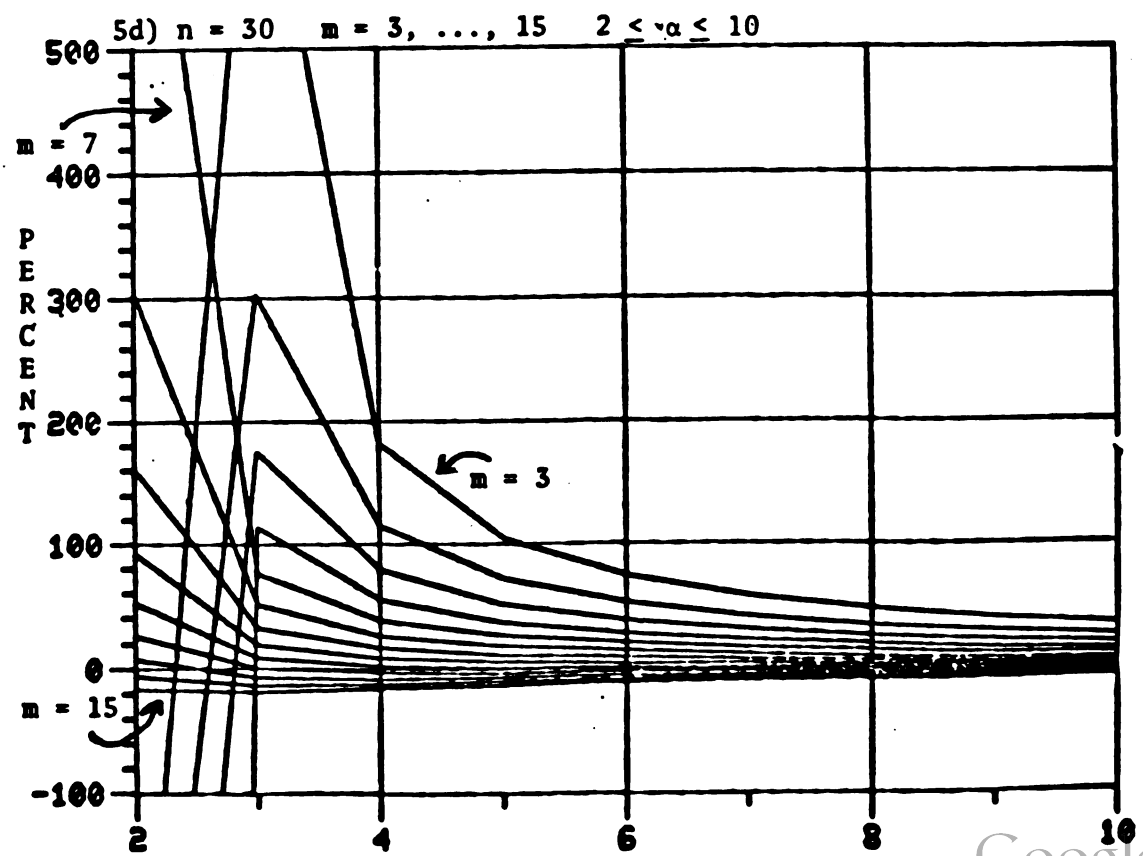
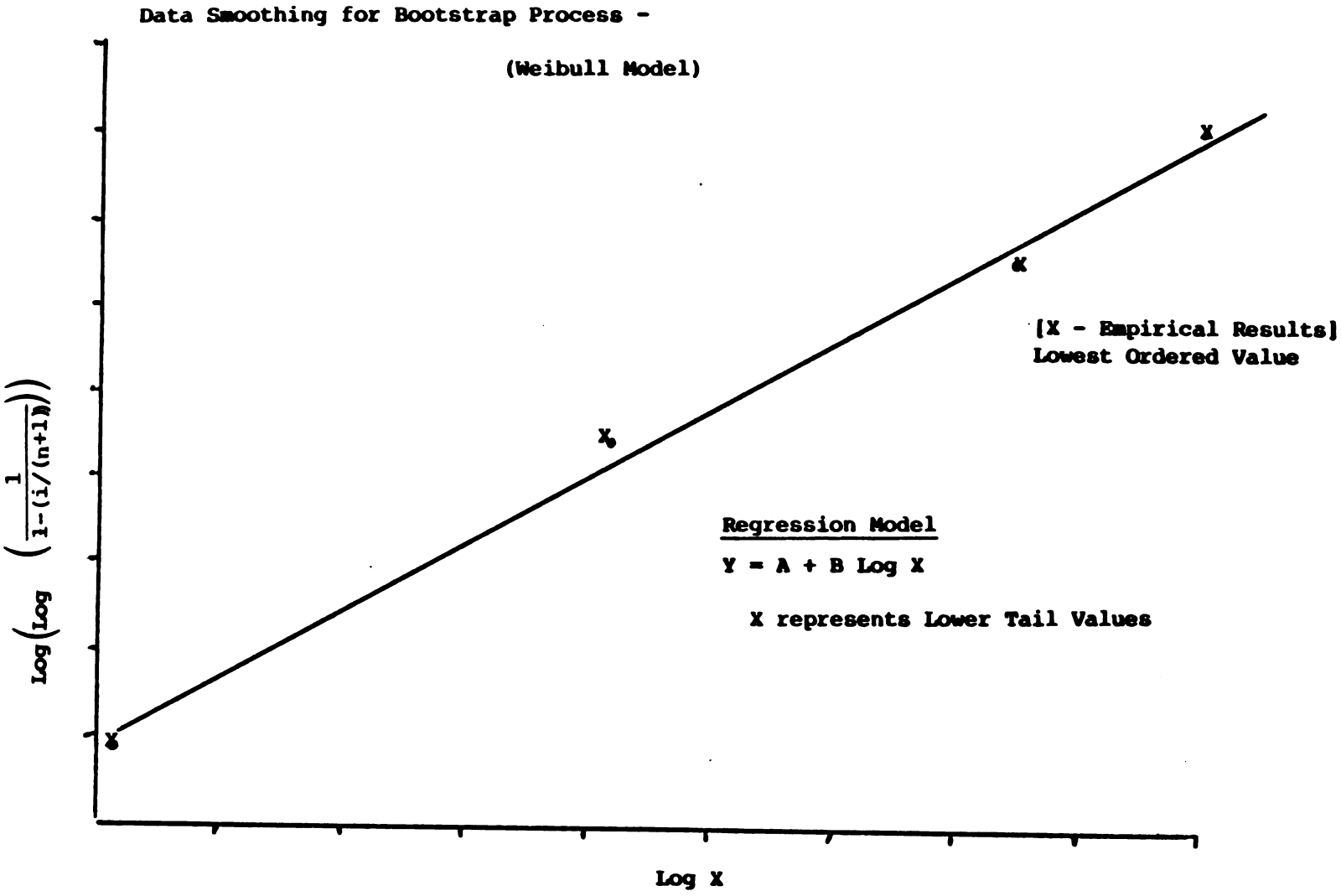


Figure 6



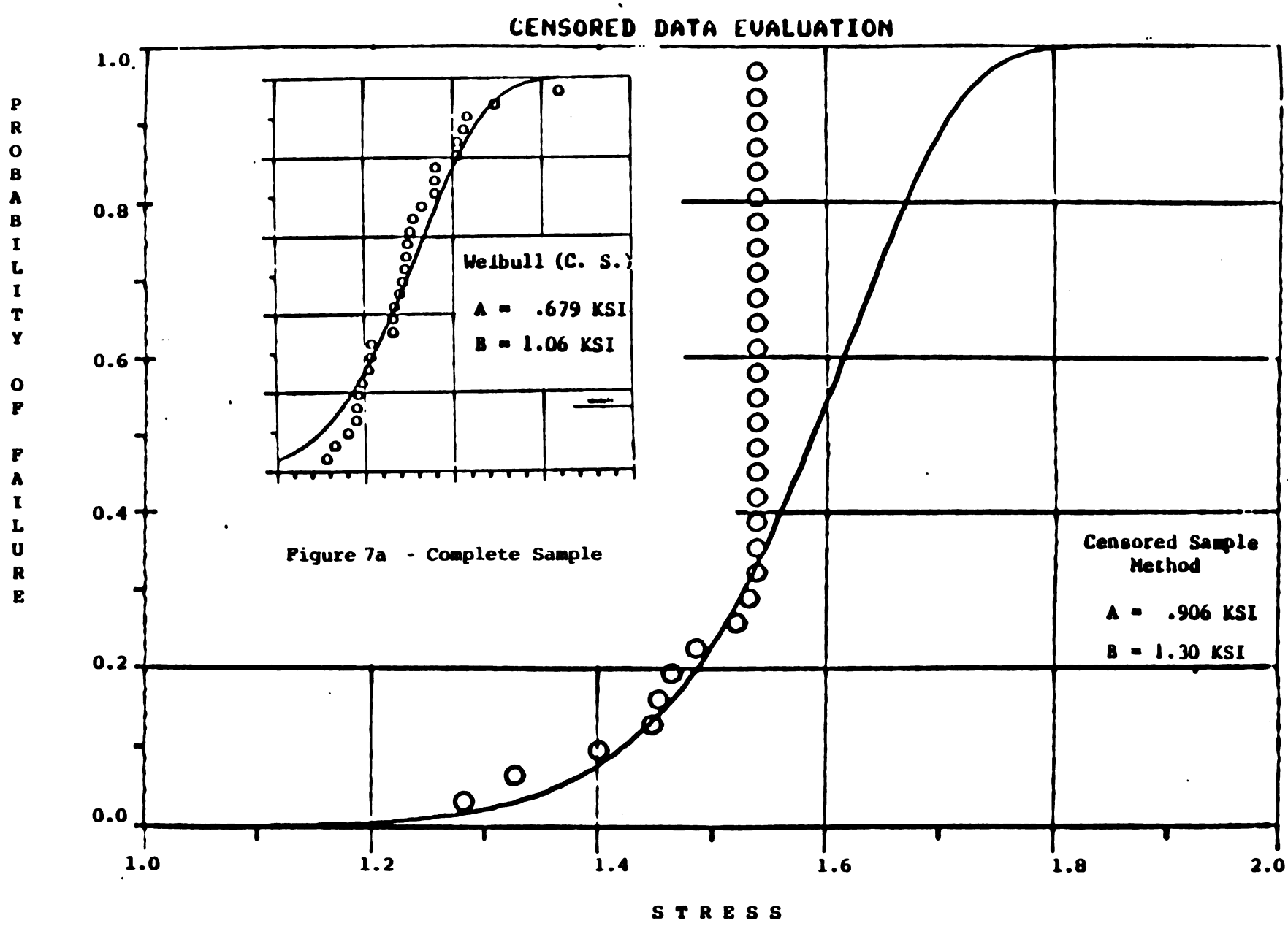


Figure 7a - Complete Sample

Figure 7 Weibull Censored Data Analysis

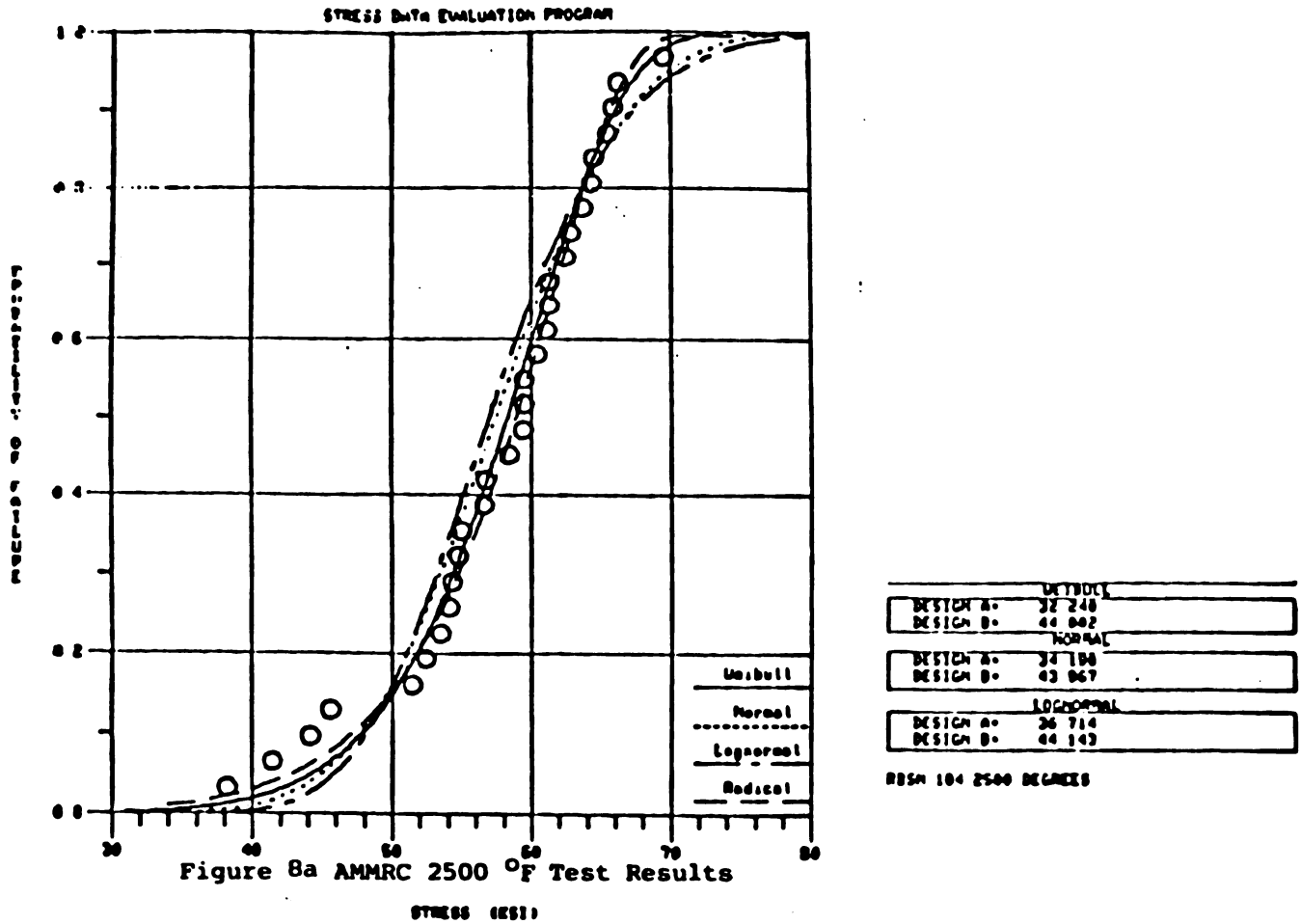


Figure 8a AMMRC 2500 °F Test Results

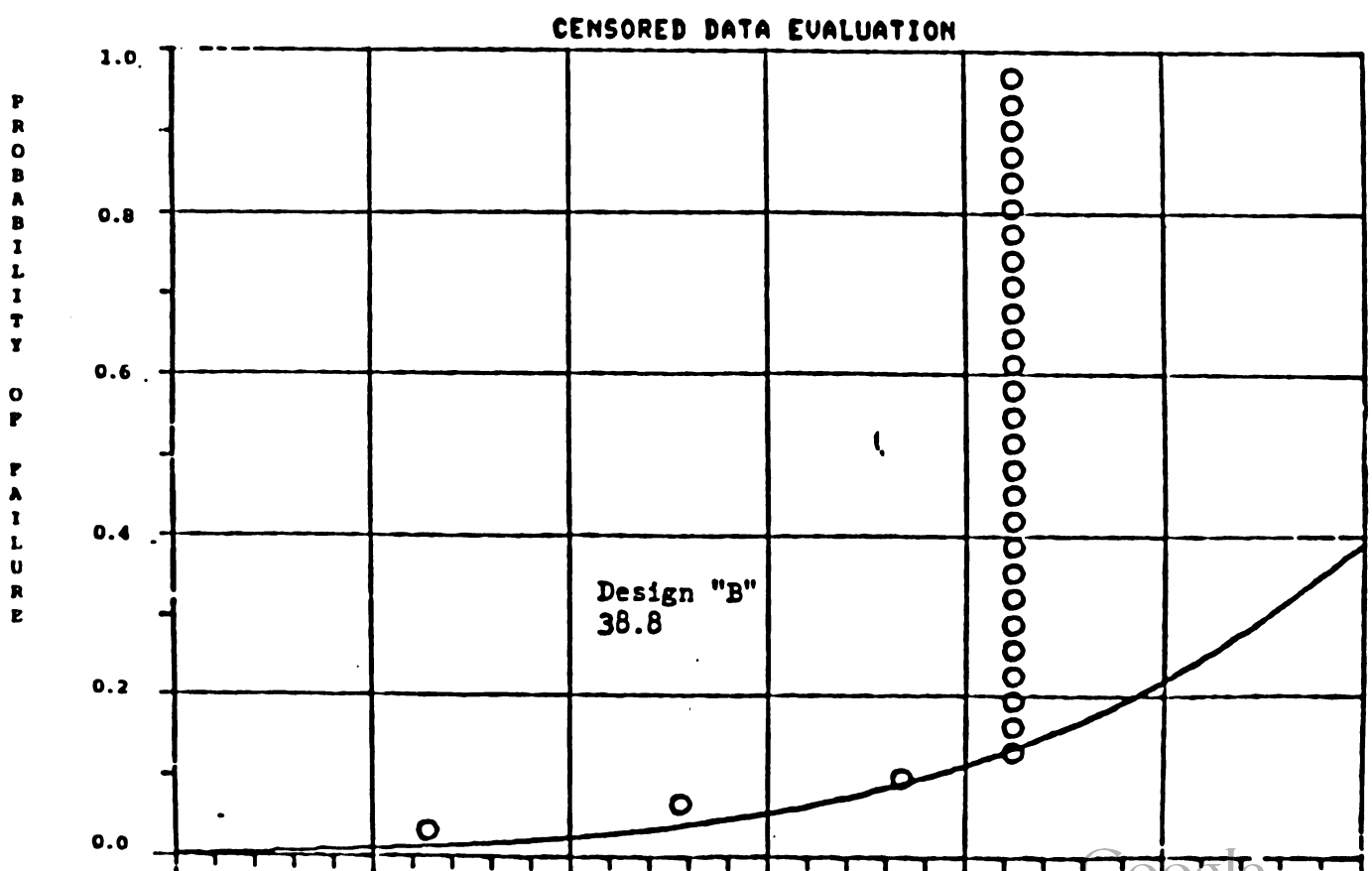
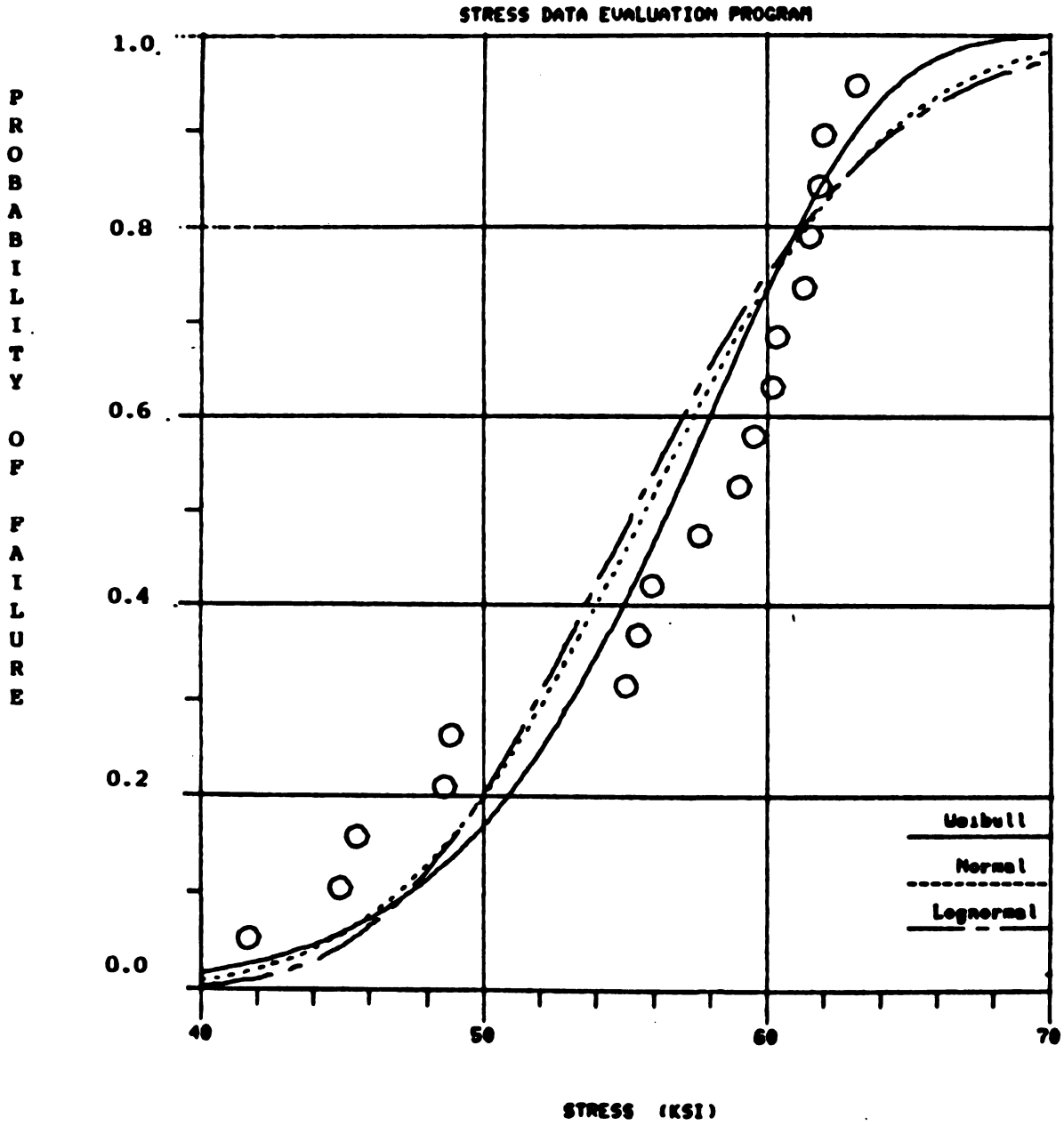


Figure 8b Weibull Censored Data Evaluation of AMMRC 2500 °F data



Design "B" Allowable

Weibull = 43.0
Normal = 42.3
Lognormal = 42.8
Censored Data = 38.2
Bootstrap = 36.1
Exponential = 36.5
Non-Parametric = 27.2
(Billing's Method)

BASIC Programs for Computing Reliability and/or Mean Life

Donald W. Rankin
Lieutenant Colonel
US Air Force (retired)

1. Introduction. In essence, these BASIC programs are used to compute those definite integrals which are associated with certain probability functions and which yield confidence levels. A full mathematical development of the formulae used will be found in two earlier papers. [1] and [2]

2. Binomial (Beta) distribution. (Sampling with replacement.) If L denotes the level of confidence, r the unknown probability of observing a success, n the sample size and k the observed number of failures, the required formula is

$$1-L = \int_{r=0}^{z} f(r)dr \quad \text{where} \quad f(r) = (n+1)C(n,k)r^{n-k}(1-r)^{k} .$$

Obviously, z falls between 0 and 1.

Program 1. This program properly is used to test for compliance with a minimum reliability standard. Values for z, n, and k are given, from which L is computed.

Program 2. This program determines the minimum reliability associated with a stated confidence level. Given are L, n, and k, from which z is computed.

Program 3. This program computes a "Best Estimate" of the reliability consistent with a stated confidence level. It minimizes the difference z_2-z_1 between the limits of integration which span that confidence level.

3. Hypergeometric probability. (Sampling without replacement.) If L denotes the level of confidence, N the population size, n the sample size, k the number of observed defectives and x the unknown number of defectives in the original population, then the required formula is

$$L = \sum_{x=k}^{x=m} P(x) \quad \text{where} \quad P(x) = \frac{C(x,k)C(N-x,n-k)}{C(N+1,n+1)} .$$

It is found that m falls between k and $N-n+k$.

Program 4. This program properly is used to test for compliance with a specified maximum allowable number of failures in the original population. Values for N, n, k and m are given, from which L is computed.

Program 5. This program produces a listing of all the discrete probabilities from k to m. The upper bound can be established by specifying either m or L.

Program 6. This program determines the maximum number of failures in the original population which is consistent with a stated confidence level. Given are L, N, n and k, from which m is computed.

Program 7. This program computes a "Best Estimate" of the number of failures in the original population, consistent with a stated confidence level. All of the included probabilities exceed every excluded one.

Program 8. This program lists all the discrete probabilities associated with the "Best Estimate" of the number of failures in the original population. (See Program 7.)

Program 9. This program is similar to Program 7, but gives in addition the maximum likelihood estimate.

4. Poisson distribution. (Constant failure rate.) "Mean Life" is given by the reciprocal of the constant (but unknown) failure rate. Let T denote the duration of the test in any suitable units, and k the number of failures observed during test T. If x/T denotes the (unknown) failure rate per unit in the whole population, then the probability that x does not exceed some value z is given by

$$P(z) = \int_{x=0}^{z} f(x)\,dx \quad \text{where} \quad f(x) = \frac{x^k e^{-x}}{k!} .$$

This integral, of course, is equivalent to a confidence level; i.e., L = P(z). Once z has been determined, the minimum value of Mean Life is given by T/z.

Program 10. This program properly is used to test for compliance with a specified minimum Mean Life in the original population. Values for T and k are given and T/z is specified, from which L is computed.

Program 11. This program determines the minimum Mean Life which is consistent with a stated confidence level. Given are L, T and k, from which z (and hence T/z) is computed.

Program 12. This program computes a "Best Estimate" of the failure rate in the original population which is consistent with a stated confidence level. Having thus determined z_1 and z_2, the corresponding values of Mean Life are given by T/z_2 and T/z_1.

5. The logarithm of the factorial. The presence of large factorials in the several formulae virtually dictates computation by logarithms. The BASIC syntax normally does not contain a routine for computing the natural logarithm of the factorial of an integer. In view of this deficiency, a representative program has been included which achieves the desired purpose. It can be used directly, or employed as a model when composing a required sub-routine. For single-digit integers, it computes the factorial directly, then passes to the logarithm. For integers of two or more digits, Stirling's complete formula is employed. (See [1], p. 278.) The result is accurate to 11 or 12 significant digits, or to the accuracy with which the machine computes simpler logarithms, whichever is less. It must be remembered that the count of significant digits includes both the characteristic and the mantissa of the logarithm. But the subsequent antilogarithm will contain no more significant digits than does the mantissa of the logarithm.

BIBLIOGRAPHY

[1] Rankin, D. W., ESTIMATING RELIABILITY FROM SMALL SAMPLES; in Proc. 22nd Conf. on D.O.E. in Army R.D.&T., Department of Defense, 1977.

[2] Rankin, D. W., ESTIMATING MEAN LIFE FROM LIMITED TESTING; in Proc. 28th Conf. on D.O.E. in Army R.D.&T., Department of Defense, 1983.

APPENDIX A

BASIC Program Listings

```
5 DISP "The LOGARITHM of the FACTORIAL."
10 DISP "ENTER  Positive Integer  (or Zero)"
15 INPUT N
20 FACTORIAL=1 @ LOGARITHM=0
25 IF N=0 THEN 80
30 IF N=1 THEN 80
35 IF N>9 THEN 65
40 FOR I=2 TO N
45    FACTORIAL=FACTORIAL*I
50 NEXT I
55    LOGARITHM=LOG (FACTORIAL)
60 GOTO 80
65 SERIES=((1/(7*N^2)-.5)/(30*N^2)+.5)/(6*N)
70 STIRLING=LOG (N)*(N+.5)-N+(SERIES+.918938533205)
75    LOGARITHM=STIRLING
80 PRINT N;"  LOGARITHM of the FACTORIAL = ";LOGARITHM
85 PRINT CHR$ (10)
90 END
```

```
2 PRINT "                    Program 1.",CHR$ (10)
4 PRINT "     This program tests for compliance with"
6 PRINT "a minimum reliability standard.",CHR$ (10)
8 DISP "ENTER  Sample Size" @ INPUT SAM
10 DISP "Enter desired reliability standard" @ INPUT Z
12 DISP "Enter number of failures observed" @ INPUT K
14 IF K>1 THEN 22
16 IF K=1 THEN 20
18 CONF=1-Z^(1+SAM) @ GOTO 48
20 CONF=1-Z^SAM*(1+SAM-Z*SAM) @ GOTO 48
22 P=1+SAM @ GOSUB 62
24 U=Q @ P=SAM-K @ GOSUB 62
26 V=Q @ P=K @ GOSUB 62
28 FZ=EXP (K*LOG (1-Z)+(SAM-K)*LOG (Z)+U-V-Q)
30 IF Z>1-K/SAM THEN 40 ! To routine for large Z
32 TZ=Z*FZ/(1+SAM-K) @ SZ=TZ @ H=1
34 TZ=Z*TZ*(1+K-H)/((1+SAM-K+H)*(1-Z)) @ SZ=SZ+TZ
36 H=1+H @ IF TZ>.00000000004 THEN 34
38 CONF=1-SZ @ GOTO 48 ! End routine for small Z
40 TZ=(1-Z)*FZ/(1+K) @ SZ=TZ @ H=1
42 TZ=(1-Z)*TZ*(1+SAM-K-H)/((1+K+H)*Z) @ SZ=SZ+TZ
44 H=1+H @ IF TZ>.00000000004 THEN 42
46 CONF=SZ ! End routine for large Z
48 FCON=CONF*100000000 @ NFCO=FCON+FP (FCON)
50 CONF=IP (NFCO)/100000000 ! Rounds to 8 places
52 PRINT "For a sample size of";SAM;"items, among"
54 PRINT "which exactly";K;"defectives were observed,"
56 PRINT "a reliability exceeding";Z;"will occur with"
58 PRINT "probability ";CONF;CHR$ (10);CHR$ (10)
60 STOP
62 LNFC: IF P>2 THEN 68
64 FAC=1 @ FOR I=1 TO P @ FAC=FAC*I @ NEXT I
66 Q=LOG (FAC) @ RETURN
68 SERI=((1/(7*P^2)-.5)/(30*P^2)+.5)/(6*P)+.918938533205
70 Q=(P+.5)*LOG (P)-P+SERI @ RETURN
72 END
```

```
3 PRINT "                    Program 2.",CHR$ (10)
6 PRINT "     This program computes a minimum reliability"
9 PRINT "consistent with a specified level of confidence."
12 PRINT @ DISP "ENTER  Sample Size" @ INPUT SAM
15 DISP "Enter desired Level of Confidence" @ INPUT CONF
18 DISP "Enter number of failures observed" @ INPUT K
21 Z=1-K/SAM @ IF K>1 THEN 39
24 IF K=1 THEN 30
27 Z=(1-CONF)^(1/(1+SAM)) @ GOTO 66
30 FZ=Z^(SAM-1)*SAM*(1+SAM)*(1-Z)
33 SZ=Z^SAM*(1+SAM-Z*SAM) @ DLZ=(1-CONF-SZ)/FZ @ Z=Z+DLZ
36 IF ABS (DLZ)>.0000000004 THEN 30 ELSE 66
39 P=1+SAM @ GOSUB 87
42 U=Q @ P=SAM-K @ GOSUB 87
45 V=Q @ P=K @ GOSUB 87
48 FZ=EXP (K*LOG (1-Z)+(SAM-K)*LOG (Z)+U-V-Q)
51 TZ=Z*FZ/(1+SAM-K) @ SZ=TZ @ H=1
54 TZ=Z*TZ*(1+K-H)/((1+SAM-K+H)*(1-Z)) @ SZ=SZ+TZ
57 H=1+H @ IF TZ>.00000000004 THEN 54 ! Integral loop
60 DLZ=(1-CONF-SZ)/FZ @ Z=Z+DLZ
63 IF ABS (DLZ)>.0000000004 THEN 48 ! Argument loop
66 XZ=Z*100000000 @ YZ=XZ+FP (XZ) @ RZ=IP (YZ)/100000000
69  ! Rounds Z to 8 decimal places
72 PRINT "For a sample size of";SAM;"items, among"
75 PRINT "which exactly";K;"defectives were observed,"
78 PRINT "a reliability exceeding";RZ;"will"
81 PRINT "occur with probability ";CONF;CHR$ (10);CHR$ (10)
84 STOP
87 LNFC: IF P>9 THEN 96
90 FAC=1 @ FOR I=1 TO P @ FAC=FAC*I @ NEXT I
93 Q=LOG (FAC) @ RETURN
96 SERI=((1/(7*P^2)-.5)/(30*P^2)+.5)/(6*P)+.918938533205
99 Q=(P+.5)*LOG (P)-P+SERI @ RETURN
102 END
```

```
1 PRINT "                    Program 3.",CHR$ (10)
2 PRINT "     This program computes a ";CHR$ (34);
3 PRINT "Best Estimate";CHR$ (34)
4 PRINT "of the reliability in the sense of minimum"
5 PRINT "distance between bounds for a specified"
6 PRINT "level of confidence.";CHR$ (10)
7 DISP "ENTER  Sample Size" @ INPUT SAM
8 DISP "Enter desired confidence level" @ INPUT CONF
9 DISP "Enter number of failures observed" @ INPUT K
10 IF K>1 THEN 25
11 IF K=1 THEN 15
12 PRINT CHR$ (34);"Best Estimate";CHR$ (34);
13 PRINT " is not defined for zero"
14 PRINT "failures.  Use Program 2 for reliability." @ STOP
15 A=1-Z/SAM
16 FA=A^(SAM-1)*SAM*(1+SAM)*(1-A) ! Begin A loop
17 SA=A^SAM*(1+SAM-A*SAM) @ Z=1-FA/(SAM*(1+SAM))
18 FZ=Z^(SAM-1)*SAM*(1+SAM)*(1-Z) ! Begin Z loop
19 DLZ=(FA/FZ-1)*Z*(1-Z)/(SAM-1-Z*SAM)
20 Z=Z+DLZ @ IF ABS (DLZ)>.000000000004 THEN 18
21 SZ=Z^SAM*(1+SAM-Z*SAM) @ Y=SAM-1/(1-Z)
22 DLA=(CONF-SZ+SA)*Y/(FA*(Z/A*(SAM-1/(1-A))-Y))
23 A=A+DLA @ IF ABS (DLA)>SAM*.000000000001 THEN 16 ELSE 42
24  ! End of routine for a single failure
25 P=1+SAM @ GOSUB 52 ! Begin routine for K > 1
26 U=Q @ P=SAM-K @ GOSUB 52
27 V=Q @ P=K @ GOSUB 52
28 R=1-K/SAM @ RZ=SQR (R^2-R*(1-K/(SAM-1))) @ A=R-RZ
29 FA=EXP (K*LOG (1-A)+(SAM-K)*LOG (A)+U-V-Q) ! A loop
30 Z=R+RZ @ TA=A*FA/(1+SAM-K) @ SA=TA @ J=1
31 TA=A*TA*(1+K-J)/((1+SAM-K+J)*(1-A)) @ SA=SA+TA
32 J=1+J @ IF TA>K*.000000000002 THEN 31 ! A integral loop
33 FZ=EXP (K*LOG (1-Z)+(SAM-K)*LOG (Z)+U-V-Q)
34 DLZ=(FA/FZ-1)*Z/(SAM-K/(1-Z)) @ Z=Z+DLZ
35 IF ABS (DLZ)>K*.000000000002 THEN 33 ! FZ=FA loop
36 TZ=Z*FZ/(1+SAM-K) @ SZ=TZ @ H=1
37 TZ=Z*TZ*(1+K-H)/((1+SAM-K+H)*(1-Z)) @ SZ=SZ+TZ
38 H=1+H @ IF TZ>K*.000000000002 THEN 37 ! Z integral loop
39 W=SAM-K/(1-Z)
40 DLA=(CONF-SZ+SA)*W/(FA*(Z/A*(SAM-K/(1-A))-W)) @ A=A+DLA
41 IF ABS (DLA)>SAM*.000000000001 THEN 29 ! End of A loop
42 XA=A*1000000000 @ YA=XA+FP (XA) @ ZA=IP (YA)/100000000
43 XZ=Z*1000000000 @ YZ=XZ+FP (XZ) @ ZZ=IP (YZ)/100000000
44  ! Rounds results to 8 decimal places
45 PRINT "For a sample size of";SAM;"items, among"
46 PRINT "which exactly";K;"defectives were observed,"
47 PRINT "the ";CHR$ (34);"Best Estimate";CHR$ (34);
48 PRINT " of the reliability"
49 PRINT "falls between ";ZA;" and ";ZZ
50 PRINT "with probability ";CONF;CHR$ (10);CHR$ (10)
51 STOP
52 LNFC: IF P>9 THEN 55 ! Logarithm of the factorial
53 FAC=1 @ FOR I=1 TO P @ FAC=FAC*I @ NEXT I
54 Q=LOG (FAC) @ RETURN
55 SERI=((1/(7*P^2)-.5)/(30*P^2)+.5)/(6*P)+.918938533205
56 Q=(P+.5)*LOG (P)-P+SERI @ RETURN
57 END
```

```
2 PRINT "                    Program 4.",CHR$ (10)
4 PRINT "     This program tests for compliance with a"
6 PRINT "specified maximum number of allowable failures"
8 PRINT "in the original population.",CHR$ (10)
10 DISP "ENTER  Population Size" @ INPUT POP
12 DISP "ENTER  Sample Size" @ INPUT SAM
14 DISP "ENTER  Maximum Number of Allowable Failures" @ INPUT MXF
16 DISP "ENTER  Number of failures observed" @ INPUT K
18 IF MXF>= K THEN 22
20 DISP "**  ERROR  **   Maximum already exceeded" @ GOTO 14
22 P=1+POP @ GOSUB 60
24 R=Q @ P=1+SAM @ GOSUB 60
26 S=Q @ P=1+MXF @ GOSUB 60
28 T=Q @ P=POP-SAM @ GOSUB 60
30 U=Q @ P=POP-MXF @ GOSUB 60
32 V=Q @ P=POP-MXF-SAM+K @ GOSUB 60
34 W=Q @ P=SAM-K @ GOSUB 60
36 X=Q @ P=MXF-K @ GOSUB 60
38 Y=Q @ P=1+K @ GOSUB 60
40 TX=EXP (U+V-R-W+S+T-X-Y-Q) @ SX=TX
42 NU=TX*(1+MXF-P)*(1+SAM-P) @ P=1+P
44 TX=NU/(P*(POP-SAM-1-MXF+P)) @ SX=SX+TX
46 IF TX>K*.000000000002 THEN 42
48 VX=SX*100000000 @ YX=VX+FP (VX) @ ZX=IP (YX)/100000000
50 PRINT "     The population consists of";POP;"similar items,  A"
52 PRINT "sample of size";SAM;"is drawn which contains exactly";K
54 PRINT "defectives.  The original population contained fewer"
56 PRINT "than";1+MXF;"defectives with probability";ZX
58 DISP SX @ PRINT CHR$ (10);CHR$ (10) @ STOP
60 LNFC: FAC=1 @ IF P>1 THEN 64
62 Q=0 @ RETURN
64 IF P>9 THEN 70
66 FOR I=1 TO P @ FAC=FAC*I @ NEXT I
68 Q=LOG (FAC) @ RETURN
70 SERI=((1/(7*P^2)-.5)/(30*P^2)+.5)/(6*P)+.918938533205
72 Q=(P+.5)*LOG (P)-P+SERI @ RETURN
74 END
```

```
2 PRINT "                    Prcgram 5.",CHR$ (10)
4 PRINT "     This program has an expanded printout which"
6 PRINT "tabulates both specific and cumulative probabilities"
8 PRINT "for various numbers of failures.  The upper bound of"
10 PRINT "the listing can be established either by stating the"
12 PRINT "maximum number of allowable failures, or by specify-"
14 PRINT "ing a desired cumulative probability.";CHR$ (10)
16 DISP "ENTER  Population Size" @ INPUT POP
18 DISP "ENTER  Sample Size" @ INPUT SAM
20 DISP "ENTER  Number of failures observed" @ INPUT K@ A=0
22 DISP "ENTER  Desired upper bound.  Use integer to denote maximum"
24 DISP "number of allowable failures.  Or use decimal fraction to"
26 DISP "express appropriate confidence level."
28 INPUT MXF@ IF MXF<1 THEN 38
30 IF FP (MXF)=0 THEN 34
32 DISP "ERROR  ";MXF;"  Invalid parameter";@ GOTO 22
34 IF MXF>= K THEN 38
36 DISP "**  ERROR  **   Maximum already exceeded" @ GOTO 20
38 Y=25 @ DIM B(99),C(99)
40 P=1+POP @ GOSUB 82
42 R=Q @ P=1+SAM @ GOSUB 82
44 S=Q @ P=POP-K @ GOSUB 82
46 T=Q @ P=SAM-K @ GOSUB 82
48 B(A)=EXP (T-R+S-Q) @ C(A)=B(A)
50 PRINT "     The population consists of";POP;"similar items,  A"
52 PRINT "sample of size";SAM;"is drawn which contains exactly";K
54 PRINT "defectives.  The various probabilities are:";CHR$ (10)
56 PRINT "Number of      Probability          Cumulative"
58 PRINT "failures     of occurrence          probability";CHR$ (10)
60 PRINT USING 62 ; K+A;B(A);C(A)
62 IMAGE XXDDD,10X,Z.8D,10X,Z.8D
64 A=1+A @ IF FP (A/5)=0 THEN PRINT
66 B(A)=B(A-1)*(A+K)*(1+POP-SAM-A)/(A*(1+POP-K-A))
68 C(A)=B(A)+C(A-1) @ IF A<> Y THEN 74
70 DISP "INSERT  NEW  PAGE.  Then press  CONT" @ PAUSE
72 Y=Y+35 @ GOTO 56
74 IF MXF<1 THEN 78
76 IF MXF<A+K THEN 80 ELSE 60
78 IF MXF>C(A-1) THEN 60
80 PRINT CHR$ (10);CHR$ (10) @ STOP
82 LNFC: FAC=1 @ IF P>1 THEN 86
84 Q=0 @ RETURN
86 IF P>9 THEN 92
88 FOR I=1 TO P @ FAC=FAC*I @ NEXT I
90 Q=LOG (FAC) @ RETURN
92 SERI=(1/(7*P^2)-.5)/(30*P^2)+.5)/(6*P)+.918938533205
94 Q=(P+.5)*LOG (P)-P+SERI @ RETURN
96 END
```

```
2 PRINT "                    Program 6.",CHR$ (10)
4 PRINT "     This program computes the maximum number of"
6 PRINT "defectives in the original population consistent"
8 PRINT "with a specified confidence level and with the"
10 PRINT "observed number of sample failures.";CHR$ (10)
12 DISP "ENTER  Population Size" @ INPUT POP
14 DISP "ENTER  Sample Size" @ INPUT SAM
16 DISP "ENTER  Number of failures observed" @ INPUT K
18 DISP "ENTER  Desired Confidence Level.  (Use decimal fraction)"
20 INPUT CONF@ Z=1-K/SAM @ C=0 @ D=0 @ IF K>1 THEN 32
22 IF K=1 THEN 26
24 Z=(1-CONF)^(1/(1+SAM)) @ GOTO 52
26 FZ=Z^(SAM-1)*SAM*(1+SAM)*(1-Z)
28 SZ=Z^SAM*(1+SAM-Z*SAM) @ DLZ=(1-CONF-SZ)/FZ @ Z=Z+DLZ
30 IF ABS (DLZ)>.00000001 THEN 26 ELSE 52
32 P=1+POP @ GOSUB 94
34 R=Q @ P=1+SAM @ GOSUB 94
36 S=Q @ P=POP-SAM @ GOSUB 94
38 U=Q @ P=SAM-K @ GOSUB 94
40 X=Q @ P=K @ GOSUB 94
42 FZ=EXP (K*LOG (1-Z)+(SAM-K)*LOG (Z)+S-R-Q)
44 TZ=Z*FZ/(1+SAM-K) @ SZ=TZ @ H=1
46 TZ=Z*TZ*(1+K-H)/((1+SAM-K+H)*(1-Z)) @ SZ=SZ+TZ
48 H=1+H @ IF TZ>.00000001 THEN 46
50 DLZ=(1-CONF-SZ)/FZ @ Z=Z+DLZ @ IF ABS (DLZ)>.00000001 THEN 42
52 MXF=INT (K-.5+(1-Z)*(1+POP-SAM))
54 P=POP-MXF-SAM+K @ GOSUB 94
56 W=Q @ P=POP-MXF @ GOSUB 94
58 V=Q @ P=1+MXF @ GOSUB 94
60 T=Q @ P=MXF-K @ GOSUB 94
62 Y=Q @ P=1+K @ GOSUB 94
64 TX=EXP (U+V-R-W+S+T-X-Y-Q) @ SX=TX
66 NU=TX*(1+MXF-P)*(1+SAM-P) @ P=1+P
68 TX=NU/(P*(POP-SAM-1-MXF+P)) @ SX=SX+TX
70 IF TX>K*.000000000002 THEN 66
72 IF SX>= CONF THEN 76
74 J=SX @ P=MXF @ MXF=1+MXF @ C=99 @ IF D=88 THEN 78 ELSE 54
76 L=SX @ G=MXF @ MXF=MXF-1 @ D=88 @ IF C<> 99 THEN 54
78 XJ=J*100000000 @ YJ=XJ+FP (XJ) @ JJ=IP (YJ)/100000000
80 XL=L*100000000 @ YL=XL+FP (XL) @ LL=IP (YL)/100000000
82 PRINT "     The population consists of";POP;"similar items.  A"
84 PRINT "sample of size";SAM;"is drawn which contains exactly";K
86 PRINT "defectives.  The original population contained fewer"
88 PRINT "than";1+G;"defectives with probability";LL
90 PRINT "Or fewer than";G;"with probability";JJ
92 PRINT CHR$ (10);CHR$ (10) @ STOP
94 LNFC: FAC=1 @ IF P>1 THEN 98
96 Q=0 @ RETURN
98 IF P>9 THEN 104
100 FOR I=1 TO P @ FAC=FAC*I @ NEXT I
102 Q=LOG (FAC) @ RETURN
104 SERI=((1/(7*P^2)-.5)/(30*P^2)+.5)/(6*P)+.918938533205
106 Q=(P+.5)*LOG (P)-P+SERI @ RETURN
108 END
```

```
1 PRINT "                    Program 7.",CHR$ (10)
2 PRINT "     This program yields a ";CHR$ (34);"Best Estimate";
3 PRINT CHR$ (34);" of"
4 PRINT "the total number of defectives in the original"
5 PRINT "population, in the sense of minimum distance"
6 PRINT "between bounds for a specified confidence level."
7 PRINT CHR$ (10) @ DISP "ENTER  Population Size" @ INPUT POP
8 DISP "ENTER  Sample Size" @ INPUT SAM
9 DISP "Enter number of failures observed" @ INPUT K
10 DISP "Enter desired confidence level (use decimal fraction)"
11 INPUT CONF@ IF K>1 THEN 25
12 IF K=1 THEN 16
13 PRINT CHR$ (34);"Best Estimate";CHR$ (34);
14 PRINT " is not defined for zero"
15 PRINT "failures.  Use Program 6 for probability." @ GOTO 93
16 A=1-2/SAM
17 FA=A^(SAM-1)*SAM*(1+SAM)*(1-A) @ SA=A^SAM*(1+SAM-A*SAM)
18 Z=1-FA/(SAM*(1+SAM))
19 FZ=Z^(SAM-1)*SAM*(1+SAM)*(1-Z)
20 DLZ=(FA/FZ-1)*Z*(1-Z)/(SAM-1-Z*SAM)
21 Z=Z+DLZ @ IF ABS (DLZ)>.000000001 THEN 19
22 SZ=Z^SAM*(1+SAM-Z*SAM) @ Y=SAM-1/(1-Z)
23 DLA=(CONF-SZ+SA)*Y/(FA*(Z/A*(SAM-1/(1-A))-Y))
24 A=A+DLA @ IF ABS (DLA)>.000000001 THEN 17 ELSE 43
25 P=1+POP @ GOSUB 88
26 R=Q @ P=1+SAM @ GOSUB 88
27 S=Q @ P=POP-SAM @ GOSUB 88
28 U=Q @ P=SAM-K @ GOSUB 88
29 X=Q @ P=K @ GOSUB 88
30 F=Q @ RX=1-K/SAM @ RZ=SQR (RX^2-RX*(1-K/(SAM-1))) @ A=RX-RZ
31 FA=EXP (K*LOG (1-A)+(SAM-K)*LOG (A)+S-X-Q) @ Z=RX+RZ
32 TA=A*FA/(1+SAM-K) @ SA=TA @ J=1
33 TA=A*TA*(1+K-J)/((1+SAM-K+J)*(1-A)) @ SA=SA+TA
34 J=1+J @ IF TA>K*.0000000002 THEN 33
35 FZ=EXP (K*LOG (1-Z)+(SAM-K)*LOG (Z)+S-X-Q)
36 DLZ=(FA/FZ-1)*Z/(SAM-K/(1-Z)) @ Z=Z+DLZ
37 IF ABS (DLZ)>K*.0000000002 THEN 35
38 TZ=Z*FZ/(1+SAM-K) @ SZ=TZ @ H=1
39 TZ=Z*TZ*(1+K-H)/((1+SAM-K+H)*(1-Z)) @ SZ=SZ+TZ
40 H=1+H @ IF TZ>K*.0000000002 THEN 39
41 WX=SAM-K/(1-Z) @ DLA=(CONF-SZ+SA)*WX/(FA*(Z/A*(SAM-K/(1-A))-WX))
42 A=A+DLA @ IF ABS (DLA)>SAM*.0000000001 THEN 31
43 MNF=INT (K-.5+(1-Z)*(1+POP-SAM))
44 MXF=INT (K+.5+(1-A)*(1+POP-SAM))
45 PRINT "     The population consists of";POP;
46 PRINT "similar items."
47 PRINT "A sample of size";SAM;"is drawn which contains exactly"
```

```
47 PRINT "A sample of size";SAM;"is drawn which contains exactly"
48 P=POP-MXF-SAM+K @ GOSUB 88
49 W=Q @ P=POP-MXF @ GOSUB 88
50 V=Q @ P=MXF @ GOSUB 88
51 T=Q @ P=MXF-K @ GOSUB 88
52 Y=Q @ P=POP-MNF-SAM+K @ GOSUB 88
53 N=Q @ P=POP-MNF @ GOSUB 88
54 M=Q @ P=MNF @ GOSUB 88
55 L=Q @ P=MNF-K @ GOSUB 88
56 G=Q @ P=1+K
57 PX=EXP (U+V-R-W+S+T-X-Y-F)
58 TX=PX*(1+MXF)/P @ SX=TX
59 NU=TX*(1+MXF-P)*(1+SAM-P)/(POP-SAM-MXF+P)
60 P=1+P @ TX=NU/P @ SX=SX+TX
61 IF TX>K*.00000000002 THEN 59
62 CX=SX @ CF=MXF @ P=1+K
63 PY=EXP (U+M-R-N+S+L-X-G-F)
64 TY=PY*(1+MNF)/P @ SY=TY
65 MU=TY*(1+MNF-P)*(1+SAM-P)/(POP-SAM-MNF+P)
66 P=1+P @ TY=MU/P @ SY=SY+TY
67 IF TY>K*.00000000002 THEN 65
68 CY=SY @ CE=MNF @ P=1+K
69 DISP MNF,MXF
70 IF CX+PY-CY<CONF THEN 78
71 IF PX<PY THEN 75
72 IF CX-CY<CONF THEN 83
73 MNF=MNF+1 @ PY=PY*MNF*(POP-SAM-MNF+K+1)/((MNF-K)*(1+POP-MNF))
74 CY=CY+PY @ GOTO 69
75 IF CX+PY-PX-CY<CONF THEN 83
76 MXF=MXF-1 @ CX=CX-PX
77 PX=PX*(1-K/(1+MXF))*(POP-MXF)/(POP-MXF-SAM+K) @ GOTO 69
78 IF PX<PY THEN 81
79 MXF=MXF+1 @ PX=PX*MXF*(POP-SAM-MXF+K+1)/((MXF-K)*(1+POP-MXF))
80 CX=CX+PX @ GOTO 69
81 MNF=MNF-1 @ CY=CY-PY
82 PY=PY*(1-K/(1+MNF))*(POP-MNF)/(POP-MNF-SAM+K) @ GOTO 69
83 XX=(CX+PY-CY)*100000000 @ YY=XX+FP (XX) @ ZZ=IP (YY)/100000000
84 PRINT K;"defectives.  At the";ZZ;"level of confi-"
85 PRINT "dence, the original population contained between"
86 PRINT MNF;"and";MXF;"defectives, inclusive."
87 PRINT CHR$ (10) @ STOP
88 LNFC: IF P>9 THEN 91
89 FAC=1 @ FOR I=1 TO P @ FAC=FAC*I @ NEXT I
90 Q=LOG (FAC) @ RETURN
91 SERI=((1/(7*P^2)-.5)/(30*P^2)+.5)/(6*P)+.918938533205
92 Q=(P+.5)*LOG (P)-P+SERI @ RETURN
93 PRINT CHR$ (10);CHR$ (10) @ END
```

```
2 PRINT "                    Program 8.",CHR$ (10)
4 PRINT "     This program has an expanded printout which"
6 PRINT "tabulates both specific and cumulative probabilities"
8 PRINT "for various numbers of failures.  The upper and lower"
10 PRINT "bounds of the listing reflect the maximum and minimum"
12 PRINT "number of failures asscciated with the ";CHR$ (34);
14 PRINT "Best Estimate";CHR$ (34)
16 PRINT "consistent with a stated level of confidence.";CHR$ (10)
18 DISP "ENTER  Population Size" @ INPUT POP
20 DISP "ENTER  Sample Size" @ INPUT SAM
22 DISP "ENTER  Number of failures observed" @ INPUT K@ A=0
24 DISP "ENTER  Confidence level.  (Use decimal fraction.)"
26 INPUT CONF@ DIM P(99),C(99)
28 IF POP-SAM>99 THEN 98
30 P=1+POP @ GOSUB 84
32 R=Q @ P=1+SAM @ GOSUB 84
34 S=Q @ P=POP-K @ GOSUB 84
36 T=Q @ P=SAM-K @ GOSUB 84
38 P(A)=EXP (T-R+S-Q) @ C(A)=P(A)
40 FOR A=1 TO POP-SAM
42 P(A)=P(A-1)*(K+A)*(1+POP-A-SAM)/(A*(1+POP-K-A))
44 C(A)=P(A)+C(A-1) @ NEXT A
46 PRINT "     The population consists of";POP;"similar items."
48 PRINT "A sample of size";SAM;"is drawn which contains exactly"
50 PRINT K;"defectives.  The various probabilities are:";CHR$ (10)
52 PRINT "Number of      Probability          Cumulative"
54 PRINT "failures     of occurrence          probability";CHR$ (10)
56 B=0 @ D=POP-SAM
58 IF P(B)<P(D) THEN 64
60 IF C(D-1)-C(B)+P(B)<CONF THEN 68
62 D=D-1 @ GOTO 58
64 IF C(D)-C(B)<CONF THEN 68
66 B=B+1 @ GOTO 58
68 A=0 @ FOR H=B TO D
70 PRINT USING 72 ; K+H;P(H);C(H)
72 IMAGE XXDDD,10X,Z.8D,10X,Z.8D
74 A=1+A @ IF FP (A/5)=0 THEN PRINT
76 NEXT H @ PRINT
78 PRINT USING 80 ; "Sum",C(D)-C(B-1),"(Conf. level)"
80 IMAGE XXAAA,10X,Z.8D,7X,13A
82 PRINT CHR$ (10);CHR$ (10) @ STOP
84 LNFC: FAC=1 @ IF P>1 THEN 88
86 Q=0 @ RETURN
88 IF P>9 THEN 94
90 FOR I=1 TO P @ FAC=FAC*I @ NEXT I
92 Q=LOG (FAC) @ RETURN
94 SERI=(1/(7*P^2)-.5)/(30*P^2)+.5)/(6*P)+.918938533205
96 Q=(P+.5)*LOG (P)-P+SERI @ RETURN
98 DISP "**  ERROR  **  Population too large for array."
100 DISP "                Redimension on lines 26 and 28."
102 END
```

```
1 PRINT "                    Program 9.",CHR$ (10)
2 PRINT "     This program yields a '";CHR$ (34);
3 PRINT "Best Estimate";CHR$ (34);" of"
4 PRINT "the total number of defectives in the original"
5 PRINT "population, in the sense of minimum distance"
6 PRINT "between bounds for a specified confidence level."
7 PRINT "Also given is the maximum likelihood estimate."
8 PRINT @ DISP "ENTER  Population Size" @ INPUT POP
9 DISP "ENTER  Sample Size" @ INPUT SAM
10 DISP "Enter number of failures observed" @ INPUT K
11 DISP "Enter desired confidence level (use decimal fraction)"
12 INPUT CONF@ IF K>1 THEN 26
13 IF K=1 THEN 17
14 PRINT CHR$ (34);"Best Estimate";CHR$ (34);
15 PRINT " is not defined for zero"
16 PRINT "failures.  Use Program 6 for probability." @ GOTO 98
17 A=1-2/SAM
18 FA=A^(SAM-1)*SAM*(1+SAM)*(1-A) @ SA=A^SAM*(1+SAM-A*SAM)
19 Z=1-FA/(SAM*(1+SAM))
20 FZ=Z^(SAM-1)*SAM*(1+SAM)*(1-Z)
21 DLZ=(FA/FZ-1)*Z*(1-Z)/(SAM-1-Z*SAM)
22 Z=Z+DLZ @ IF ABS (DLZ)>.000000001 THEN 20
23 SZ=Z^SAM*(1+SAM-Z*SAM) @ Y=SAM-1/(1-Z)
24 DLA=(CONF-SZ+SA)*Y/(FA*(Z/A*(SAM-1/(1-A))-Y))
25 A=A+DLA @ IF ABS (DLA)>.000000001 THEN 18 ELSE 44
26 P=1+POP @ GOSUB 93
27 R=Q @ P=1+SAM @ GOSUB 93
28 S=Q @ P=POP-SAM @ GOSUB 93
29 U=Q @ P=SAM-K @ GOSUB 93
30 X=Q @ P=K @ GOSUB 93
31 F=Q @ RX=1-K/SAM @ RZ=SQR (RX^2-RX*(1-K/(SAM-1))) @ A=RX-RZ
32 FA=EXP (K*LOG (1-A)+(SAM-K)*LOG (A)+S-X-Q) @ Z=RX+RZ
33 TA=A*FA/(1+SAM-K) @ SA=TA @ J=1
34 TA=A*TA*(1+K-J)/((1+SAM-K+J)*(1-A)) @ SA=SA+TA
35 J=1+J @ IF TA>K*.0000000002 THEN 34
36 FZ=EXP (K*LOG (1-Z)+(SAM-K)*LOG (Z)+S-X-Q)
37 DLZ=(FA/FZ-1)*Z/(SAM-K/(1-Z)) @ Z=Z+DLZ
38 IF ABS (DLZ)>K*.0000000002 THEN 36
39 TZ=Z*FZ/(1+SAM-K) @ SZ=TZ @ H=1
40 TZ=Z*TZ*(1+K-H)/((1+SAM-K+H)*(1-Z)) @ SZ=SZ+TZ
41 H=1+H @ IF TZ>K*.0000000002 THEN 40
42 WX=SAM-K/(1-Z) @ DLA=(CONF-SZ+SA)*WX/(FA*(Z/A*(SAM-K/(1-A))-WX))
43 A=A+DLA @ IF ABS (DLA)>SAM*.0000000001 THEN 32
44 MNF=INT (K-.5+(1-Z)*(1+POP-SAM))
45 MXF=INT (K+.5+(1-A)*(1+POP-SAM))
46 PRINT "     The population consists of";POP;"similar items."
47 PRINT "A sample of size";SAM;"is drawn which contains exactly"
48 P=POP-MXF-SAM+K @ GOSUB 93
49 W=Q @ P=POP-MXF @ GOSUB 93
50 V=Q @ P=MXF @ GOSUB 93
```

```
50 V=Q @ P=MXF @ GOSUB 93
51 T=Q @ P=MXF-K @ GOSUB 93
52 Y=Q @ P=POP-MNF-SAM+K @ GOSUB 93
53 N=Q @ P=POP-MNF @ GOSUB 93
54 M=Q @ P=MNF @ GOSUB 93
55 L=Q @ P=MNF-K @ GOSUB 93
56 G=Q @ P=1+K @ PX=EXP (U+V-R-W+S+T-X-Y-F)
57 TX=PX*(1+MXF)/P @ SX=TX
58 NU=TX*(1+MXF-P)*(1+SAM-P)/(POP-SAM-MXF+P)
59 P=1+P @ TX=NU/P @ SX=SX+TX
60 IF TX>K*.00000000002 THEN 58
61 CX=SX @ CF=MXF @ P=1+K @ PY=EXP (U+M-R-N+S+L-X-G-F)
62 TY=PY*(1+MNF)/P @ SY=TY
63 MU=TY*(1+MNF-P)*(1+SAM-P)/(POP-SAM-MNF+P)
64 P=1+P @ TY=MU/P @ SY=SY+TY
65 IF TY>K*.00000000002 THEN 63
66 CY=SY @ CE=MNF @ P=1+K
67 DISP MNF,MXF
68 IF CX+PY-CY<CONF THEN 76
69 IF PX<PY THEN 73
70 IF CX-CY<CONF THEN 81
71 MNF=MNF+1 @ PY=PY*MNF*(POP-SAM-MNF+K+1)/((MNF-K)*(1+POP-MNF))
72 CY=CY+PY @ GOTO 67
73 IF CX+PY-PX-CY<CONF THEN 81
74 MXF=MXF-1 @ CX=CX-PX
75 PX=PX*(1-K/(1+MXF))*(POP-MXF)/(POP-MXF-SAM+K) @ GOTO 67
76 IF PX<PY THEN 79
77 MXF=MXF+1 @ PX=PX*MXF*(POP-SAM-MXF+K+1)/((MXF-K)*(1+POP-MXF))
78 CX=CX+PX @ GOTO 67
79 MNF=MNF-1 @ CY=CY-PY
80 PY=PY*(1-K/(1+MNF))*(POP-MNF)/(POP-MNF-SAM+K) @ GOTO 67
81 XX=(CX+PY-CY)*100000000 @ YY=XX+FP (XX) @ ZZ=IP (YY)/100000000
82 PRINT K;"defectives.  At the";ZZ;"level of confi-"
83 PRINT "dence, the original population contained between"
84 PRINT MNF;"and";MXF;"defectives, inclusive.";CHR$ (10)
85 MLX=(1+POP)*K/SAM
86 MLY=INT (MLX) @ IF MLX=MLY THEN 89
87 PRINT "The maximum likelihood estimate is";MLY;"defectives."
88 GOTO 98
89 PRINT "There are two equally likely maximum likelihood"
90 PRINT "estimates.  They are";MLX-1;"and";MLX;"defectives."
91 GOTO 98
92 STOP
93 LNFC: IF P>9 THEN 96
94 FAC=1 @ FOR I=1 TO P @ FAC=FAC*I @ NEXT I
95 Q=LOG (FAC) @ RETURN
96 SERI=((1/(7*P^2)-.5)/(30*P^2)+.5)/(6*P)+.918938533205
97 Q=(P+.5)*LOG (P)-P+SERI @ RETURN
98 PRINT CHR$ (10);CHR$ (10) @ END
```

```
3 PRINT "                    Program 10.",CHR$ (10)
6 PRINT "     This program tests for compliance with a"
9 PRINT "minimum mean life standard.",CHR$ (10)
12 DISP "ENTER Duration of Test, UNITS Employed."
15 DISP "     (Use comma between.)"
18 INPUT DUR,UNIT$
21 DISP "Enter desired mean life standard in  ";UNIT$
24 INPUT MLS@ DISP "Enter number of failures observed"
27 INPUT K@ Z=DUR/MLS @ IF K>0 THEN 33
30 CONF=1-EXP (-Z) @ GOTO 72
33 IF K>9 THEN 54
36 FAC=1 @ FOR I=2 TO K @ FAC=FAC*I @ NEXT I
39 FZ=Z^K/(FAC*EXP (Z)) @ TZ=K*FZ/Z
42 SZ=FZ+TZ @ IF K=1 THEN 51
45 FOR J=1 TO K @ TZ=TZ*(K-J)/Z
48 SZ=SZ+TZ @ NEXT J
51 CONF=1-SZ @ GOTO 72
54 SERI=((1/(7*K^2)-.5)/(30*K^2)+.5)/(6*K)
57 FZ=EXP (K*(1+LOG (Z))-(K+.5)*LOG (K)-Z-(SERI+.9189385332))
60 TZ=Z*FZ/(K+1) @ CONF=TZ @ L=2
63 TZ=Z*TZ/(K+L) @ CONF=CONF+TZ
66 IF TZ<.00000000001 THEN 72
69 L=1+L @ GOTO 63
72 XX=CONF*100000000 @ YY=XX+FP (XX) @ ZZ=IP (YY)/100000000
75 PRINT "For a test duration of  ";DUR;"  ";UNIT$
78 PRINT "during which exactly  ";K;"  failures were observed,"
81 PRINT "a mean life exceeding  ";MLS;"  ";UNIT$
84 PRINT "will occur with probability  ";ZZ;CHR$ (10)
87 END
```

```
3 PRINT "                    Program 11.",CHR$ (10)
6 PRINT "     This program computes a minimum mean life"
9 PRINT "consistent with a specified level of confidence.",CHR$ (10)
12 DISP "ENTER Duration of Test, UNITS Employed.  (Use comma between.)"
15 INPUT DUR,UNIT$
18 DISP "Enter desired Level of Confidence.    (Use decimal fraction.)"
21 INPUT CONF@ DISP "Enter number of failures observed"
24 INPUT K@ IF K>0 THEN 30
27 Z=-LOG (1-CONF) @ GOTO 75
30 Z=K+SQR (K) @ IF K>9 THEN 54
33 FAC=1 @ FOR I=2 TO K @ FAC=FAC*I @ NEXT I
36 FZ=Z^K/(FAC*EXP (Z)) @ TZ=K*FZ/Z
39 SZ=FZ+TZ @ IF K=1 THEN 48
42 FOR J=1 TO K @ TZ=TZ*(K-J)/Z
45 SZ=SZ+TZ @ NEXT J
48 DLZ=Z*(1-SQR (1-2*(Z-K)*(CONF-1+SZ)/(Z*FZ)))/(Z-K)
51 Z=Z+DLZ @ IF ABS (DLZ)<K^2*.0000000001 THEN 75 ELSE 36
54 SERI=((1/(7*K^2)-.5)/(30*K^2)+.5)/(6*K)
57 FZ=EXP (K*(1+LOG (Z))-(K+.5)*LOG (K)-Z-(SERI+.9189385332))
60 TZ=Z*FZ/(K+1) @ TCON=TZ @ L=2
63 TZ=Z*TZ/(K+L) @ TCON=TCON+TZ @ IF TZ<.00000000001 THEN 69
66 L=L+1 @ GOTO 63
69 DLZ=Z*(1-SQR (1-2*(Z-K)*(CONF-TCON)/(Z*FZ)))/(Z-K)
72 Z=Z+DLZ @ IF ABS (DLZ)<K^2*.0000000001 THEN 75 ELSE 57
75 XX=100000*DUR/Z @ YY=XX+FP (XX) @ MLS=IP (YY)/100000
78 PRINT "For a test duration of ";DUR;" ";UNIT$
81 PRINT "during which exactly ";K;" failures were observed,"
84 PRINT "a mean life exceeding ";MLS;" ";UNIT$
87 PRINT "will occur with probability ";CONF;CHR$ (10)
90 PRINT CHR$ (10)
93 END
```

```
2 PRINT "                    Program 12.",CHR$ (10)
4 PRINT "    This program computes the ";CHR$ (34);
6 PRINT "Best Estimate";CHR$ (34);" of"
8 PRINT "the Mean Life consistent with a specified Level"
10 PRINT "of Confidence.",CHR$ (10)
12 DISP "ENTER Duration of Test, UNITS employed.  (Use comma between.)"
14 INPUT DUR,UNIT$
16 DISP "ENTER desired Level of Confidence.   (Use decimal fraction.)"
18 INPUT CONF@ DISP "ENTER number of failures observed."
20 INPUT K@ IF K>0 THEN 28
22 PRINT CHR$ (34);"Best Estimate";CHR$ (34);
24 PRINT " is not defined for zero failures."
26 PRINT "Use Program 11 to compute probability." @ GOTO 110
28 Z=2 @ IF K>1 THEN 44
30 FZ=Z/EXP (Z) @ A=FZ
32 FA=A/EXP (A) @ DLA=(FZ-FA)*A/(FA*(1-A))
34 A=A+DLA @ ALB=A @ IF ABS (DLA)>K*.0000000005 THEN 32
36 TCON=FZ/A-FZ/Z
38 DLZ=(Z-(Z-1))*(1-SQR (1-2*(Z-1)*(1-A)*(CONF-TCON)/((Z-A)*FZ)))
40 Z=Z+DLZ @ FZ=Z/EXP (Z)
42 IF ABS (DLZ)>K^2*.0000000001 THEN 30 ELSE 88
44 Z=K+SQR (K) @ IF K>9 THEN 66
46 A=K-SQR (K) @ FAC=1 @ FOR I=2 TO K @ FAC=FAC*I @ NEXT I
48 FZ=Z^K/(FAC*EXP (Z)) @ TZ=K*FZ/Z @ SZ=FZ+TZ
50 FOR J=1 TO K @ TZ=TZ*(K-J)/Z @ SZ=SZ+TZ @ NEXT J
52 FA=A^K/(FAC*EXP (A)) @ DLA=A*(FZ/FA-1)/(K-A)
54 A=A+DLA @ ALB=A @ IF ABS (DLA)>K*.0000000005 THEN 52
56 TA=K*FZ/A @ SA=FZ+TA
58 FOR H=1 TO K @ TA=TA*(K-H)/A @ SA=SA+TA @ NEXT H
60 TCON=SA-SZ
62 DLZ=Z*(1-SQR (1-2*(Z-K)*(K-A)*(CONF-TCON)/(K*(Z-A)*FZ)))/(Z-K)
64 Z=Z+DLZ @ IF ABS (DLZ)>K^2*.0000000001 THEN 48 ELSE 88
66 SERI=((1/(7*K^2)-.5)/(30*K^2)+.5)/(6*K) @ A=K-SQR (K)
68 LNFC=(K+.5)*LOG (K)-K+(SERI+.918938533205)
70 FZ=EXP (K*LOG (Z)-Z-LNFC) @ TZ=K*FZ/Z @ SZ=FZ+TZ @ Q=K-1
72 TZ=TZ*Q/Z @ SZ=SZ+TZ @ Q=Q-1 @ IF TZ>.00000000004 THEN 72
74 FA=EXP (K*LOG (A)-A-LNFC) @ DLA=A*(FZ/FA-1)/(K-A) @ A=A+DLA
76 IF ABS (DLA)>K*.0000000005 THEN 74
78 SA=0 @ TA=FA @ Q=K+1
80 TA=A*TA/Q @ SA=SA+TA @ Q=Q+1 @ IF TA>.00000000004 THEN 80
82 TCON=1-SZ-SA
84 DLZ=Z*(1-SQR (1-2*(Z-K)*(K-A)*(CONF-TCON)/(K*(Z-A)*FZ)))/(Z-K)
86 Z=Z+DLZ @ IF ABS (DLZ)>K^2*.0000000001 THEN 70
88 QQ=100000*DUR @ XX=QQ/Z @ YY=XX+FP (XX) @ ZZ=IP (YY)/100000
90 CC=QQ/A @ BB=CC+FP (CC) @ AA=IP (BB)/100000 @ HH=QQ/K
92 JJ=HH+FP (HH) @ KK=IP (JJ)/100000
94 PRINT "     For a test duration of";DUR;UNIT$
96 PRINT "exactly";K;"failures were observed."
98 PRINT CHR$ (10);"At the ";CONF;" Level of Confidence, the"
100 PRINT CHR$ (34);"Best Estimate";
102 PRINT CHR$ (34);" of the Mean Life falls"
104 PRINT "between";ZZ;" and"
106 PRINT AA;UNIT$;".  The maximum like-"
108 PRINT "lihood estimate is ";KK;UNIT$;"."
110 PRINT @ END
```

APPENDIX B

The Programs Exercised

Program 1.

This program tests for compliance with a minimum reliability standard.

For a sample size of 7 items, among which exactly 2 defectives were observed, a reliability exceeding .7 will occur with probability .44822619

Program 1.

This program tests for compliance with a minimum reliability standard.

For a sample size of 21 items, among which exactly 6 defectives were observed, a reliability exceeding .7 will occur with probability .50582374

Program 1.

This program tests for compliance with a minimum reliability standard.

For a sample size of 70 items, among which exactly 20 defectives were observed, a reliability exceeding .7 will occur with probability .5754343

Program 1.

This program tests for compliance with a minimum reliability standard.

For a sample size of 700 items, among which exactly 200 defectives were observed, a reliability exceeding .7 will occur with probability .78981566

Program 2.

This program computes a minimum reliability consistent with a specified level of confidence.

For a sample size of 7 items, among which exactly 2 defectives were observed, a reliability exceeding .53790159 will occur with probability .8

Program 2.

This program computes a minimum reliability consistent with a specified level of confidence.

For a sample size of 14 items, among which exactly 4 defectives were observed, a reliability exceeding .59237177 will occur with probability .8

Program 2.

This program computes a minimum reliability consistent with a specified level of confidence.

For a sample size of 70 items, among which exactly 20 defectives were observed, a reliability exceeding .66385031 will occur with probability .8

Program 2.

This program computes a minimum reliability consistent with a specified level of confidence.

For a sample size of 700 items, among which exactly 200 defectives were observed, a reliability exceeding .69937687 will occur with probability .8

Program 3.

This program computes a "Best Estimate" of the reliability in the sense of minimum distance between bounds for a specified level of confidence.

For a sample size of 7 items, among which exactly 2 defectives were observed, the "Best Estimate" of the reliability falls between .49573206 and .87983592 with probability .8

Program 3.

This program computes a "Best Estimate" of the reliability in the sense of minimum distance between bounds for a specified level of confidence.

For a sample size of 14 items, among which exactly 4 defectives were observed, the "Best Estimate" of the reliability falls between .55515217 and .8438256 with probability .8

Program 3.

This program computes a "Best Estimate" of the reliability in the sense of minimum distance between bounds for a specified level of confidence.

For a sample size of 70 items, among which exactly 20 defectives were observed, the "Best Estimate" of the reliability falls between .6428334 and .77920786 with probability .8

Program 4.

This program tests for compliance with a specified maximum number of allowable failures in the original population.

The population consists of 26 similar items. A sample of size 7 is drawn which contains exactly 2 defectives. The original population contained fewer than 9 defectives with probability .5502503

Program 4.

This program tests for compliance with a specified maximum number of allowable failures in the original population.

The population consists of 26 similar items. A sample of size 7 is drawn which contains exactly 2 defectives. The original population contained fewer than 12 defectives with probability .8352657

Program 4.

This program tests for compliance with a specified maximum number of allowable failures in the original population.

The population consists of 48 similar items. A sample of size 7 is drawn which contains exactly 2 defectives. The original population contained fewer than 16 defectives with probability .5229367

Program 4.

This program tests for compliance with a specified maximum number of allowable failures in the original population.

The population consists of 48 similar items. A sample of size 7 is drawn which contains exactly 2 defectives. The original population contained fewer than 22 defectives with probability .8001343

Program 5.

This program has an expanded printout which tabulates both specific and cumulative probabilities for various numbers of failures. The upper bound of the listing can be established either by stating the maximum number of allowable failures, or by specifying a desired cumulative probability.

The population consists of 26 similar items, a sample of size 7 is drawn which contains exactly 2 defectives. The various probabilities are:

Number of failures	Probability of occurrence	Cumulative probability
2	0.01914530	0.01914530
3	0.04547009	0.06461538
4	0.07117057	0.13578595
5	0.09165907	0.22744502
6	0.10475322	0.33219824
7	0.10999088	0.44218912
8	0.10806121	0.55025033
9	0.10034256	0.65059289
10	0.08853755	0.73913043
11	0.07439614	0.81352657
12	0.05951691	0.87304348
13	0.04521739	0.91826087
14	0.03246377	0.95072464
15	0.02185061	0.97257525
16	0.01362116	0.98619641
17	0.00771866	0.99391507
18	0.00385933	0.99777440
19	0.00161751	0.99939191
20	0.00051350	0.99990541
21	0.00009459	1.00000000

Program 5.

This program has an expanded printout which tabulates both specific and cumulative probabilities for various numbers of failures. The upper bound of the listing can be established either by stating the maximum number of allowable failures, or by specifying a desired cumulative probability.

The population consists of 48 similar items. A sample of size 7 is drawn which contains exactly 2 defectives. The various probabilities are:

Number of failures	Probability of occurrence	Cumulative probability
2	0.00303951	0.00303951
3	0.00812740	0.01116691
4	0.01444870	0.02561561
5	0.02134467	0.04696029
6	0.02829410	0.07525439
7	0.03489600	0.11015045
8	0.04085392	0.15100438
9	0.04596067	0.19696504
10	0.05008534	0.24705038
11	0.05316076	0.30021114
12	0.05517224	0.35538338
13	0.05614751	0.41153089
14	0.05614751	0.46767840
15	0.05525830	0.52293670
16	0.05358380	0.57652050
17	0.05123951	0.62776002
18	0.04834696	0.67610698
19	0.04502903	0.72113601
20	0.04140600	0.76254201
21	0.03759229	0.80013430
22	0.03369383	0.83382814
23	0.02980608	0.86363422
24	0.02601258	0.88964680
25	0.02238402	0.91203082
26	0.01897775	0.93100857

Program 6.

This program computes the maximum number of defectives in the original population consistent with a specified confidence level and with the observed number of sample failures.

The population consists of 26 similar items. A sample of size 7 is drawn which contains exactly 2 defectives. The original population contained fewer than 12 defectives with probability .81352657
Or fewer than 11 with probability .73913043

Program 6.

This program computes the maximum number of defectives in the original population consistent with a specified confidence level and with the observed number of sample failures.

The population consists of 48 similar items. A sample of size 7 is drawn which contains exactly 2 defectives. The original population contained fewer than 22 defectives with probability .8001343
Or fewer than 21 with probability .76254201

Program 6.

This program computes the maximum number of defectives in the original population consistent with a specified confidence level and with the observed number of sample failures.

The population consists of 125 similar items. A sample of size 14 is drawn which contains exactly 4 defectives. The original population contained fewer than 51 defectives with probability .8094822
Or fewer than 50 with probability .79092954

Program 7.

This program yields a "Best Estimate" of the total number of defectives in the original population, in the sense of minimum distance between bounds for a specified confidence level.

The population consists of 26 similar items. A sample of size 7 is drawn which contains exactly 2 defectives. At the .80842809 level of confidence, the original population contained between 4 and 12 defectives, inclusive.

Program 7.

This program yields a "Best Estimate" of the total number of defectives in the original population, in the sense of minimum distance between bounds for a specified confidence level.

The population consists of 48 similar items. A sample of size 7 is drawn which contains exactly 2 defectives. At the .81667294 level of confidence, the original population contained between 6 and 23 defectives, inclusive.

Program 7.

This program yields a "Best Estimate" of the total number of defectives in the original population, in the sense of minimum distance between bounds for a specified confidence level.

The population consists of 125 similar items. A sample of size 14 is drawn which contains exactly 4 defectives. At the .80976263 level of confidence, the original population contained between 20 and 54 defectives, inclusive.

Program 8.

This program has an expanded printout which tabulates both specific and cumulative probabilities for various numbers of failures. The upper and lower bounds of the listing reflect the maximum and minimum number of failures associated with the "Best Estimate" consistent with a stated level of confidence.

The population consists of 26 similar items. A sample of size 7 is drawn which contains exactly 2 defectives. The various probabilities are:

Number of failures	Probability of occurrence	Cumulative probability
4	0.07117057	0.13578595
5	0.09165907	0.22744502
6	0.10475322	0.33219824
7	0.10999088	0.44218912
8	0.10806121	0.55025033
9	0.10034256	0.65059289
10	0.08653755	0.73913043
11	0.07439614	0.81352657
12	0.05951691	0.87304348
Sum	0.80842809	(Conf. level)

Program 8.

This program has an expanded printout which tabulates both specific and cumulative probabilities for various numbers of failures. The upper and lower bounds of the listing reflect the maximum and minimum number of failures associated with the "Best Estimate" consistent with a stated level of confidence.

The population consists of 48 similar items. A sample of size 7 is drawn which contains exactly 2 defectives. The various probabilities are:

Number of failures	Probability of occurrence	Cumulative probability
6	0.02829410	0.07525439
7	0.03489606	0.11015045
8	0.04085392	0.15100438
9	0.04596067	0.19696504
10	0.05008534	0.24705038
11	0.05316076	0.30021114
12	0.05517224	0.35538338
13	0.05614751	0.41153089
14	0.05614751	0.46767840
15	0.05525830	0.52293670
16	0.05358380	0.57652050
17	0.05123951	0.62776002
18	0.04834696	0.67610698
19	0.04502903	0.72113601
20	0.04140600	0.76254201
21	0.03759229	0.80013430
22	0.03369383	0.83382814
23	0.02980608	0.86363422
Sum	0.81667394	(Conf. level)

Program 9.

This program yields a "Best Estimate" of the total number of defectives in the original population, in the sense of minimum distance between bounds for a specified confidence level. Also given is the maximum likelihood estimate.

The population consists of 26 similar items. A sample of size 7 is drawn which contains exactly 2 defectives. At the .80842809 level of confidence, the original population contained between 4 and 12 defectives, inclusive.

The maximum likelihood estimate is 7 defectives.

Program 9.

This program yields a "Best Estimate" of the total number of defectives in the original population, in the sense of minimum distance between bounds for a specified confidence level. Also given is the maximum likelihood estimate.

The population consists of 46 similar items. A sample of size 7 is drawn which contains exactly 2 defectives. At the .81667394 level of confidence, the original population contained between 6 and 23 defectives, inclusive.

There are two equally likely maximum likelihood estimates. They are 13 and 14 defectives.

Program 9.

This program yields a "Best Estimate" of the total number of defectives in the original population, in the sense of minimum distance between bounds for a specified confidence level. Also given is the maximum likelihood estimate.

The population consists of 125 similar items. A sample of size 14 is drawn which contains exactly 4 defectives. At the .80976263 level of confidence, the original population contained between 20 and 54 defectives, inclusive.

There are two equally likely maximum likelihood estimates. They are 35 and 36 defectives.

Program 10.

This program tests for compliance with a minimum mean life standard.

For a test duration of 3200 hours
during which exactly 6 failures were observed,
a mean life exceeding 500 hours
will occur with probability .45767113

Program 10.

This program tests for compliance with a minimum mean life standard.

For a test duration of 3200 hours
during which exactly 5 failures were observed,
a mean life exceeding 500 hours
will occur with probability .61625634

Program 10.

This program tests for compliance with a minimum mean life standard.

For a test duration of 3200 hours
during which exactly 4 failures were observed,
a mean life exceeding 500 hours
will occur with probability .76492997

Program 10.

This program tests for compliance with a minimum mean life standard.

For a test duration of 3200 hours
during which exactly 3 failures were observed,
a mean life exceeding 500 hours
will occur with probability .88108124

Program 11.

This program computes a minimum mean life consistent with a specified level of confidence.

For a test duration of 3200 hours
during which exactly 6 failures were observed,
a mean life exceeding 352.60211 hours
will occur with probability .8

Program 11.

This program computes a minimum mean life consistent with a specified level of confidence.

For a test duration of 3200 hours
during which exactly 5 failures were observed,
a mean life exceeding 404.75623 hours
will occur with probability .8

Program 11.

This program computes a minimum mean life consistent with a specified level of confidence.

For a test duration of 3200 hours
during which exactly 4 failures were observed,
a mean life exceeding 476.12113 hours
will occur with probability .8

Program 11.

This program computes a minimum mean life consistent with a specified level of confidence.

For a test duration of 3200 hours
during which exactly 3 failures were observed,
a mean life exceeding 580.23091 hours
will occur with probability .8

Program 12.

This program computes the "Best Estimate" of the Mean Life consistent with a specified Level of Confidence.

For a test duration of 3200 hours exactly 6 failures were observed.

At the .8 Level of Confidence, the "Best Estimate" of the Mean Life falls between 327.58165 and 954.25284 hours. The maximum likelihood estimate is 533.33333 hours.

Program 12.

This program computes the "Best Estimate" of the Mean Life consistent with a specified Level of Confidence.

For a test duration of 3200 hours exactly 5 failures were observed.

At the .8 Level of Confidence, the "Best Estimate" of the Mean Life falls between 376.27302 and 1220.02901 hours. The maximum likelihood estimate is 640 hours.

Program 12.

This program computes the "Best Estimate" of the Mean Life consistent with a specified Level of Confidence.

For a test duration of 3200 hours exactly 4 failures were observed.

At the .8 Level of Confidence, the "Best Estimate" of the Mean Life falls between 443.51195 and 1666.43711 hours. The maximum likelihood estimate is 800 hours.

APPENDIX C

Some Notes on BASIC Programming

Like much that is utilitarian, the BASIC programming language* is in a state of constant evolution. Moreover, most vendors embellish it with additional statements and commands, intended to solve real or imagined difficulties. Some of these differences are pointed out here. But the ultimate guide is a good owner's manual.

1. Capital Letters. Older versions of BASIC require the use of capital letters. Some later ones can recognize instructions written in lower case.

2. Variable Assignment. Older versions require that the statement LET precede the variable name. With later versions, its use is optional. Variables should be assigned one at a time. In a statement such as "A = B = 4", many versions would regard the second equals sign as a logical operator.

3. Variable Names. Most versions allow names of some length so that they can act as prompts to the operator. However, as few as two characters may be used by the machine. Thus "DIVER" could not be distinguished from "DISCO". Certain reserved words -- such as FN, GO, IF, ON, OR, TO, and others -- should be avoided as variable names.

4. Subroutine Calls. Some versions allow calling a subroutine by label. Others call by line number only.

5. Multiple statements on a line, if allowed, are separated by some symbol which varies among versions. Perhaps @ \ and : are the most common.

---

* Beginners' All-purpose Symbolic Instruction Code

6. Remarks are variously preceded by ' ! or REM.

7. There are many varieties of truncation at the radix. One will encounter INT, FIX, IP, FP, FRC and others.

8. Perhaps the most confusing is the command "PRINT". In some versions it is used to address the printer, in others to address the display. To address the printer, we may find PRINT PRINT# LPRINT or PR#1. To address the display, we might encounter ? DISP PRINT or PR#0. Additionally, PRINT# sometimes is used to address a mass storage device.

9. For formatted print-out, some versions reference a numbered IMAGE line. Others reference a string variable. Thus Program 5, lines 60 and 62 might read:

```
60 Y$ = "  ###           #.########          #.########"
62 LPRINT USING Y$; K+A, B(A), C(A)
```

10. Note that punctuation is not necessarily the same from version to version.

THE APPLICATION OF EXPERIMENTAL DESIGN TO EVALUATION OF MULTI-ECHELON STOCKAGE MODELS

Carl B. Bates
US Army Concepts Analysis Agency
Bethesda, Maryland

ABSTRACT. A long-range objective of the Army is to develop a comprehensive stockage methodology which integrates the echelons and optimizes cost, weapon system availability, and transportability. As research into multi-echelon inventory models progresses, various stockage policies (models) have been proposed as ways to best satisfy requirements and constraints. The Concepts Analysis Agency conducted a study to evaluate the utility and effectiveness of two proposed multi-echelon stockage policies. A part of the study involved a sensitivity analysis on each of the simulation models and a comparative analysis on the two models. This paper discusses the experimental design employed and the statistical analysis results of the sensitivity analyses and the comparative analysis.

1. BACKGROUND. Current Army supply policies do not relate operations and maintenance, Army (OMA) funds to weapon system availability. The long-range objective of the Office of the Deputy Chief of Staff for Logistics (ODCSLOG) is to develop a comprehensive stockage policy which integrates all echelons of support activities. The Retail Inventory Management Stockage Policy (RIMSTOP) Model is the current DOD inventory stockage policy. Methodology development of multi-echelon inventory theory is ongoing. The Inventory Research Office (IRO) recently developed the Major Assemblies Stockage System (MASS) Model. MASS is a multi-echelon inventory stockage model which uses an optimization process. It is a derivative of the Multi-Echelon Technique for Recoverable Item Control (METRIC) Model developed by The Rand Corporation. Before further development or implementation of MASS, ODCSLOG wanted an independent assessment of MASS. Ultimately, the US Army Concepts Analysis Agency (CAA) was requested to evaluate MASS and compare it with RIMSTOP. The Multi-Echelon Stockage Analysis (MESA) is documented in a CAA Study Report* published in 1984. A part of the study objectives was to conduct sensitivity analyses of MASS versus RIMSTOP. This paper discusses the sensitivity analyses performed on the two models.

---

*Blake, Robert T., et al., Multi-Echelon Stockage Analysis (MESA), CAA-SR-84-18, Bethesda, MD, May 1984.

2. EXPERIMENTAL DESIGN AND MODEL. Because both models were relatively fast running, 250 to 300 runs were allowed for the sensitivity analyses. Ultimately, eight input factors common to both models were selected for the investigation. Three levels (low, nominal, and high) were selected for each of the eight input factors shown in Table 1.

Table 1. Input Factors

Direct Support (DS) Units
A - Repair cycle time (days)
B - Demand (failed major assemblies removed)
C - Percent of failed assemblies repaired locally
D - Order ship time (days)
General Support (GS) Unit
E - Repair cycle time (days)
F - Demand (failed major assemblies removed)
G - Percent of failed assemblies repaired locally
H - Order ship time (days)

This gave $3^8 = 6{,}561$ factor level combinations. A $1/27 \times 3^8$ fractional factorial design was developed which contained 243 design points. The design permitted testing of the 8 main effects and the 28 first order interaction effects. The fixed effects analysis of variance (ANOVA) model for the fractional factorial design was:

$$y = \mu + A + B + C + \ldots + H + AB + AC + AD + \ldots GH + HOI.$$

The dependent variable (y) represents the model output variable Inventory Cost; μ is a true but unknown common effect; the letters A, B, ..., H represent the eight model input factors; HOI represents the Higher Order Interactions. The ANOVA table is illustrated in Table 2. The design may appear somewhat wasteful because only 128 of the total 242 degrees of freedom were fitted. The design was employed, however, because of the desire to measure all first order interactions.

Table 2. ANOVA Table for 1/27 x 38 Fractional Factorial Design

Source	Degrees of freedom	Sum of squares	Mean squares	F-ratio
A	2	SS(A)	MS(A)	MS(A)/MS(HOI)
B	2	SS(B)	MS(B)	MS(B)/MS(HOI)
C	2	SS(C)	MS(C)	MS(C)/MS(HOI)
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
H	2	SS(H)	MS(H)	MS(H)/MS(HOI)
AB	4	SS(AB)	MS(AB)	MS(AB)/MS(HOI)
AC	4	SS(AC)	MS(AC)	MS(AC)/MS(HOI)
AD	4	SS(AD)	MS(AD)	MS(AD)/MS(HOI)
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
GH	4	SS(GH)	MS(GH)	MS(GH)/MS(HOI)
Higher order interactions	114	SS(HOI)	MS(HOI)	
Total	242	SS(T)		

3. ANALYSIS. The single output variable of interest was the cost of inventory required to achieve an anticipated level of operational availability. Analyses were performed for two end items (the M60A3 tank and the M561 truck). The data were obtained from 1982 maintenance records from the Vth and the VIIth Corp in USAREUR. The ANOVA illustrated in Table 2 above was performed on the M561 and the M60A3 data from both models. Main effects and two-factor interaction effects were tabulated, and significant two-factor interaction effects were graphically illustrated. The significant two-factor interactions for the M561 system and the M60A3 system are shown in Tables 3 and 4, respectively. Table 3 shows 11 interactions significant from MASS and 9 interactions significant for RIMSTOP. All main effects except A and F are contained in the interactions shown. The main effect A was significant for MASS, but it was not significant for RIMSTOP. Main effect F was not significant for either models; therefore, based upon the M561 data, input factor F (GS Demand) does not have a significant influence upon inventory cost from either model.

Table 3. Significant Interaction Effects for M561

Interactions	Significance	
	MASS	RIMSTOP
BC - DS assembly removals x DS percentage repaired	0.001	0.001
BD - DS assembly removals x DS order ship time	0.001	0.001
BE - DS assembly removals x GS repair cycle time	0.05	
BG - DS assembly removals x GS percentage repaired	0.001	0.001
BH - DS assembly removals x GS order ship time	0.001	0.001
CD - DS percentage repaired x DS order ship time	0.001	0.05
CG - DS percentage repaired x GS percentage repaired	0.01	0.05
CH - DS percentage repaired x GS order ship time	0.001	0.001
DE - DS order ship time x GS repair cycle time	0.05	0.05
EG - GS repair cycle time x GS percentage repaired	0.001	
GH - GS percentage repaired x GS order ship time	0.001	0.001

Table 4. Significant Interaction Effects for M60A3

Interactions	Significance	
	MASS	RIMSTOP
BC - DS assembly removals x DS percentage repaired	0.001	0.001
BD - DS assembly removals x DS order ship time	0.001	0.001
BE - DS assembly removals x GS repair cycle time	0.01	0.01
BG - DS assembly removals x GS percentage repaired	0.001	0.001
BH - DS assembly removals x GS order ship time	0.001	0.001
CD - DS percentage repaired x DS order ship time	0.001	
CE - DS percentage repaired x GS repair cycle time	0.05	
CG - DS percentage repaired x GS percentage repaired	0.001	0.001
CH - DS percentage repaired x GS order ship time	0.001	0.001
DE - DS order ship time x GS repair cycle time	0.01	0.05
EG - GS repair cycle time x GS percentage repaired	0.001	0.001
GH - GS percentage repaired x GS order ship time	0.001	0.001

Table 4 for the M60A3 data shows there were 12 significant interactions for MASS and 10 for RIMSTOP. Most of the interactions in Table 4 are the same as those in Table 3; however, the two lists are not identical. Again, all input factors are contained in the interactions in Table 4 except A and F. This time the F main effect is not significant for MASS, but it is for RIMSTOP. As with the M561 data, the three largest interactions in Table 4 are GH, BH, and BC.

Figures 1 and 2 illustrate the GH (GS percentage repaired x GS order ship time) interaction for the M561 data for MASS and RIMSTOP, respectively. The figures show the similarity of the interactions from the two models. Figures 3 and 4 show the same interaction (GH) for the M60A3 and the two models. Again, similarity of the interactions is shown. Figures 1 and 3 show the comparison of the M561 with the M60A3 data for the MASS Model. Figures 2 and 4 show the comparison of the two data sets for the RIMSTOP Model.

4. SUMMARY. The utilization of experimental design permitted an efficient evaluation of the MASS Model and a comparison of MASS with the RIMSTOP Model. Experimental design should be an integral part of the test and evaluation of computer simulation models. Analysts involved in the exercise of simulation models for the purpose of generating data for analysis should be apprised of the need for experimental design and continually reminded of its importance.

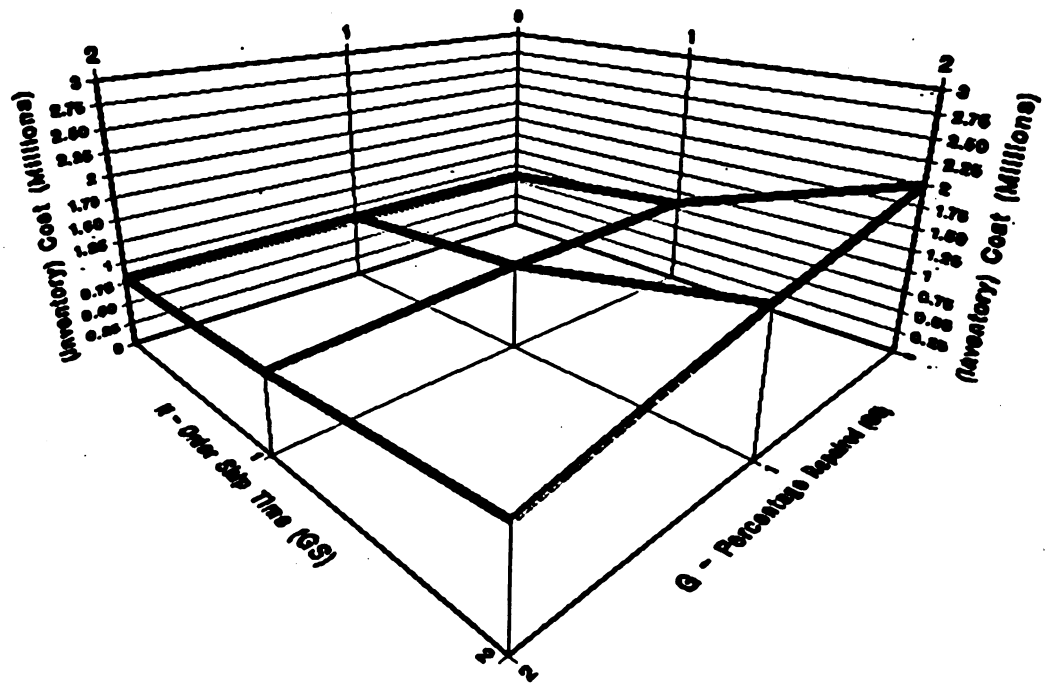


Figure 1. GH Interaction, MASS, M561

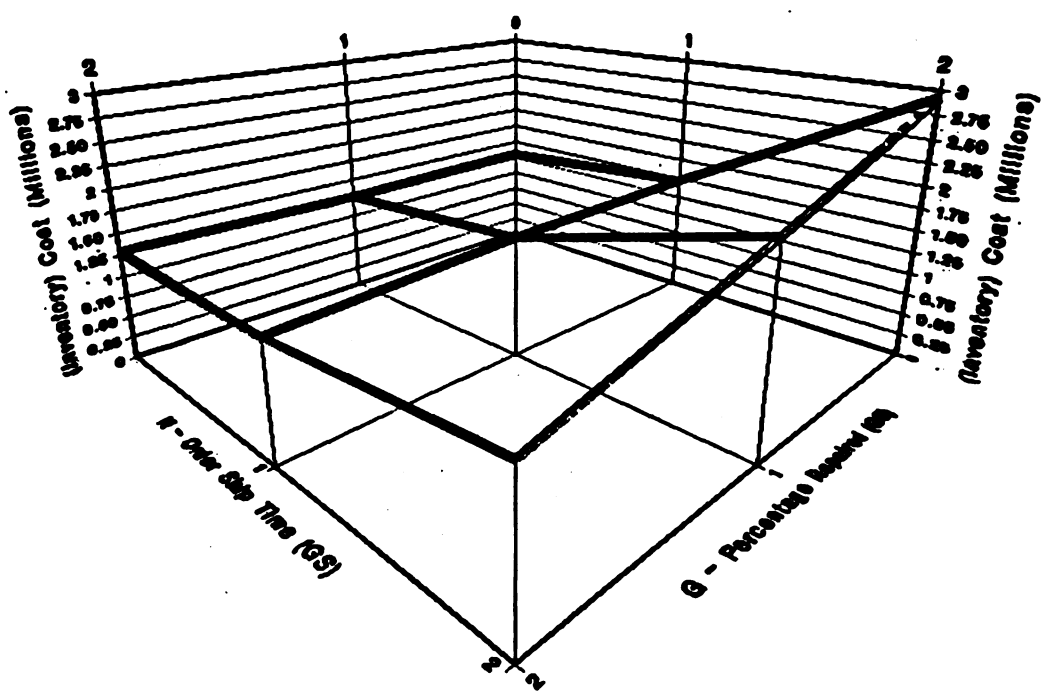


Figure 2. GH Interaction, RIMSTOP, M561

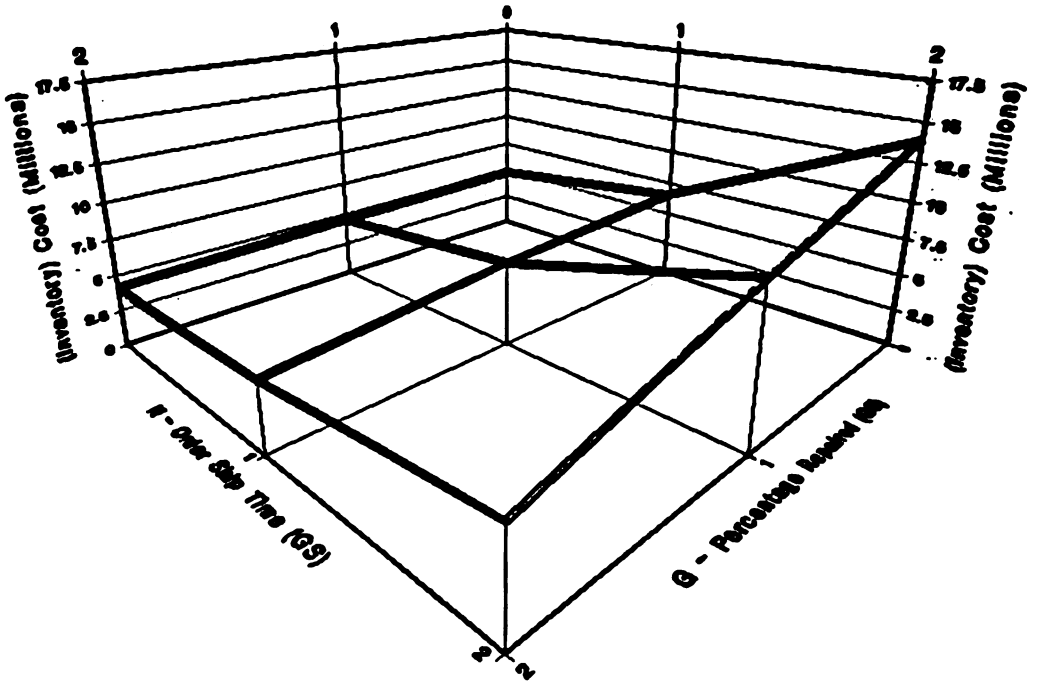


Figure 3. GH Interaction, MASS, M60A3

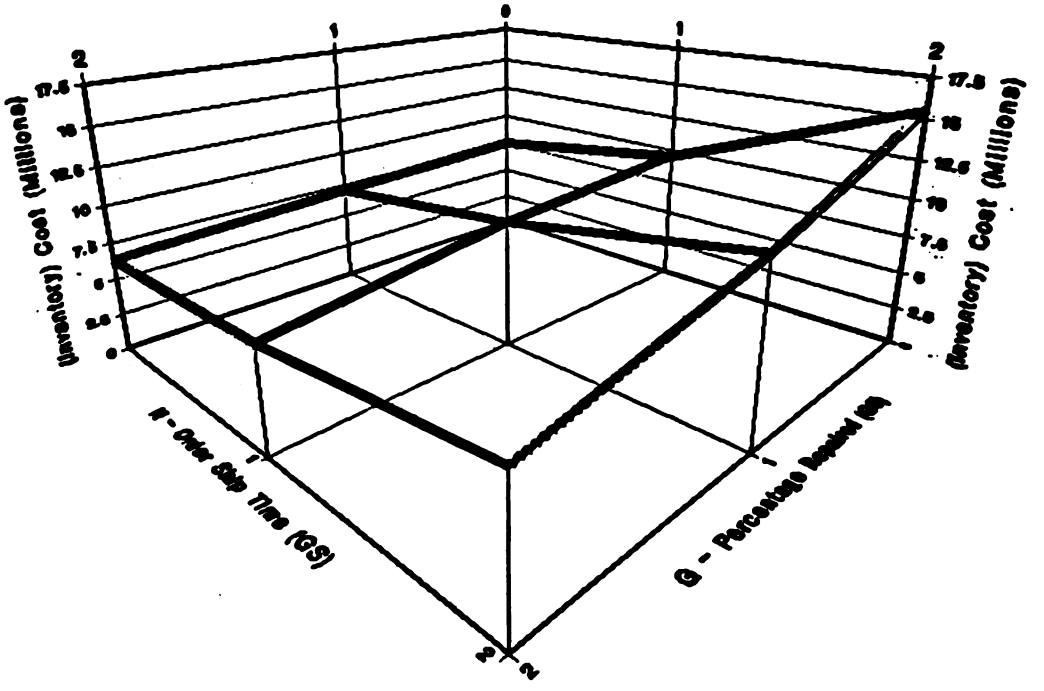


Figure 4. GH Interaction, RIMSTOP, M60A3

LEAST-TIME ANALYSIS: A METHODOLOGY FOR EXPERIMENTAL DESIGN IN LOGISTICS SYSTEMS TESTING AND EVALUATION

Clarence H. Annett
TRADOC Independent Evaluation Directorate
US Army Combined Arms Center
Fort Leavenworth, Kansas 66027-5130

Abstract: This "work in progress" paper describes the background and development of a distinctly new methodology for the design of tests and objective evaluations of logistics systems. The method is based on the physical principle known as Fermat's Principle of Least Time, and on the practical reality that "supplies delivered too late are worthless." The problem of logistical support is formulated mathematically using variational calculus. Rather than rigorously solving an abstract, general problem, an information theory approach is used to chart the flow of logistical materials and requests in time and space, segment the flow into discrete serial units, and analyze each unit separately. In this way, both the specific flow unit where a delay in delivery occurs and the reason for the delay can be identified. The optimal solution is achieved when the materials pass through all serial units in the least possible time. Tests can be designed, and criteria for evaluation of the tests can be constructed by comparing measured results to the optimal solution. In particular, the method can be used to produce specific "target values" which constitute an objective yardstick for evaluation.

I. Introduction. The design and conduct of any test of new, improved, or modified Army materiel, concepts, force structure, or doctrine are governed by a set of issues and criteria (Stevens, 1979). The issues are simply questions about the performance, reliability, safety, usability, efficiency, or other aspects of the tested item. The criteria are objective standards which the tested item must meet if it is to be judged acceptable for use (and for procurement/distribution subsequent to testing). There is a rule of thumb that says that the best test and the most objective evaluation will result when the criterion is one which can be quantified and the test can produce quantitative results for comparison to the criterion. Quantitative criteria are often referred to in the test/evaluation literature as measures of effectiveness (MOE) or measures of performance (MOP), although these terms are often used more loosely than they should be (i.e., to denote subjective or non-quantitative criteria). Together the issues and criteria form the building blocks which govern the design of the test, execution of the test plan, and analysis of the test data.

Quantitative criteria usually are obvious or follow logically from an examination of the tested item's purpose or intended use. For instance, if it is desired to have a larger-caliber howitzer because more range is needed, one of the test criterion should speak to the amount of improvement in range (either absolute, in km, or percent) which will be expected when the new howitzer is compared to the one it will replace. Indeed, for most functional areas in the Army (infantry weapons, tube artillery, ground communications, etc.), a set of "generic" issues and criteria can be developed. When the test

item is prepared, the specific performance numbers can be added to the generic criteria and the testing can proceed. Adherence to the criteria will be especially important for some issues, depending on the function of the tested item; these issues are designated as "critical" issues and the system will not be judged acceptable unless it meets all the criteria for all critical issues. Thus the importance of having numerical, objective criteria for critical issues is clear.

In some instances, however, it is not at all easy (even with logic and common sense!) to determine objective or numerical criteria for a given test item. This is particularly true for items, systems, or concepts in logistics and supply. Part of the difficulty arises because logistics cuts across all functional areas rather than having a single functional area of its own with a well-defined, limited mission. The question of logistics supportability arises in the consideration of all Army systems, but the truth of the matter is that it is never really addressed properly in testing. Typically, the only logistics issue which is usually written is:

Issue: Is the test system logistically supportable?

And its associated criterion is almost always:

Criterion: The system must be logistically supportable.

From a logical point of view, the issue and associated criterion are unassailable; the Army cannot tactically or economically afford a system that the logistics supply system cannot support, no matter how well the system works. However, it is equally obvious that this issue and criterion are badly flawed on several counts. First, the criterion begs the issue and does not readily admit any other conceivable answer, which violates the principle of objectivity in test design. Secondly, the criterion cannot be quantified, and as the issue is written no quantifiable criterion can be readily determined. These two factors alone would preclude any reasonable, objective, believable test of the system's logistic supportability. TRADOC Independent Evaluation Directorate (Ft. Leavenworth) has attempted on at least two occasions to alleviate this problem by constructing generic issues/criteria for logistic supportability, with appropriate measures of performance and/or effectiveness, but these efforts have not succeeded because of the exceptionally broad scope of logistics problems encountered in Army test and evaluation.

Hidden within the problem of testing logistic supportability is a dilemma which is neither widely recognized nor completely understood by test designers. It is true that a system should not (and in fact, often cannot) be issued to the Army in the field unless logistic supportability is demonstrated. Logistic supportability cannot be demonstrated until the total complement of systems is fielded and the total supply system is exercised. However, for testing and evaluation only a few systems (typically five or ten) are procured, so the logistics support systems and procedures are never fully exercised. This creates a chicken-and-egg dilemma: the total number of systems desired must be purchased and deployed in order to demonstrate logistic supportability, but the decision to buy more than just a few systems for testing cannot properly be made until logistic supportability is

demonstrated by a test! This dilemma would be solved if a method for assessing logistic supportability using a limited number of systems could be found.

The answer to both these problems -- non-quantifiable criteria and the dilemma of logistic supportability demonstration — appears to be in finding an alternate way of looking at the logistics supply problem. "Conventional" logistics analysis more or less relies on examining the quantity of supplies in the supply system and their availability at various supply points. Quantitative analytical tools and measures of effectiveness exist for determining how well the system works (Engel, 1980). However, the proper use of these quantitative measures and tools depends on the full utilization of the logistics supply system, which again implies full fielding of the materiel. Thus conventional analytical tools are not expected to be useful in resolving problems which fall outside of the situations for which they were designed, and the implication is that new tools will have to be developed.

II. Development of Least-Time Analysis. There is an alternate way of looking at the logistics supply problem. Instead of a conventional or "inventory" analysis, in which the functioning of the supply system is determined by monitoring the supplies on hand at certain points in the system, the alternative is to examine the delivery time for supplies and analyze the effects of the logistics supply system on the delivery time. On their face, these two analyses would appear to be quite different; however, with a little thought one can convince oneself that in fact they are really two ways of looking at the same thing and should give the same results. The advantage to avoiding any "inventory" analysis is that without the constraints of counting supplies, the method should be easily extended to any quantity of supplies, large or small, and the supportability dilemma may be overcome.

Using an analysis based on delivery time has the distinct advantage that it can embrace a "guiding principle" of logistics supply that every soldier in the field knows: Supplies and logistical materials delivered too slowly are worth no more than no supplies at all. To put this simply but graphically, the ammunition delivered to the soldier in the foxhole is of no use if the supply system is so slow that the soldier is killed before he gets it. This principle provides powerful motivation for the development of the method of Least-Time Analysis.

The first step in Least-Time Analysis is illustrated in Figure 1. Using a map model, the flow of supplies from manufacturer (1) to depot (2) to port (3) to foreign port (4) to staging area (5) to the using unit (6) is plotted and the distances and travel times (including loading, unloading, refueling, and all delays) are noted. For convenience, the map model (upper half of Figure 1) is converted to a linear delivery model (lower half of Figure 1) involving only times and locations. This idea is borrowed from information theory and is also called a Shannon model. The Shannon model is appropriate for this because the flow of supplies can be modeling as linear or serial, supplies flow in discrete units or packages which can be traced, and the locations of the supply terminals are fixed but the travel time between them is unconstrained.

FIGURE 1

Once the times T1, T2,... have been determined from the linear delivery model, a Space-Time diagram is plotted as in Figure 2. The location of each "node" in the linear delivery model is plotted on the horizontal (X) axis and is treated as a discrete variable with only certain allowed values (the discrete values come from noting that depots, ports, etc. are fixed in location and do not move frequently). Each location is represented on the graph by a vertical line. Time is treated as a continuous, cumulative variable and is plotted on the vertical (Y) axis. As shown, the time it takes for the supplies to pass between two nodes is plotted on the graph, and a curve is obtained from which the total time (called delivery time) for supplies to travel from the start point to the end point can be determined. Delays at any location are represented as vertical gaps which make the space-time diagram discontinuous.

III. Analysis using the Space-Time Diagram. After the space-time diagram is constructed for the situation being considered, it is analyzed according to three rules. These rules are based on the "guiding principle" of logistics defined two paragraphs previously.

a) The optimal path for moving logistical materiel is the path for which the space-time diagram gives the shortest delivery time.

b) Each segment of the optimal path is self-optimal; that is, each path segment by itself takes the minimum possible time to go from its beginning point to its end point.

c) Any path segment which does not take the minimum time is not optimal.

It follows from these rules that test, evaluation, and modification activities need to be directed only (and especially) toward those activities which make up the path segments which are identified as not optimal. These segments can almost always be identified as the "choke points" which hinder the proper and timely flow of logistical materials.

IV. Mathematical Foundation for Least-Time Analysis. Least-time analysis was inspired by a well-known principle of physics, Fermat's Principle of Least Time (Marion, 1970). Fermat's principle was originally applied to optical ray tracing, and later to classical kinematics and dynamics of moving bodies. It states that if a system is initially in one state or condition, and at some later time is in a different state or condition, the path taken to go from the initial state to the final state will be the path which allows the transition to occur in the shortest possible time period. A formalism called variational calculus was later developed to describe these kinds of transitions between states. Variational calculus is extensive and complex; only a few details necessary to form the basis of least-time analysis will be presented here. The notation follows the basic text on the subject (Gelfand and Fomin, 1963), and the reader is referred to this book for further details.

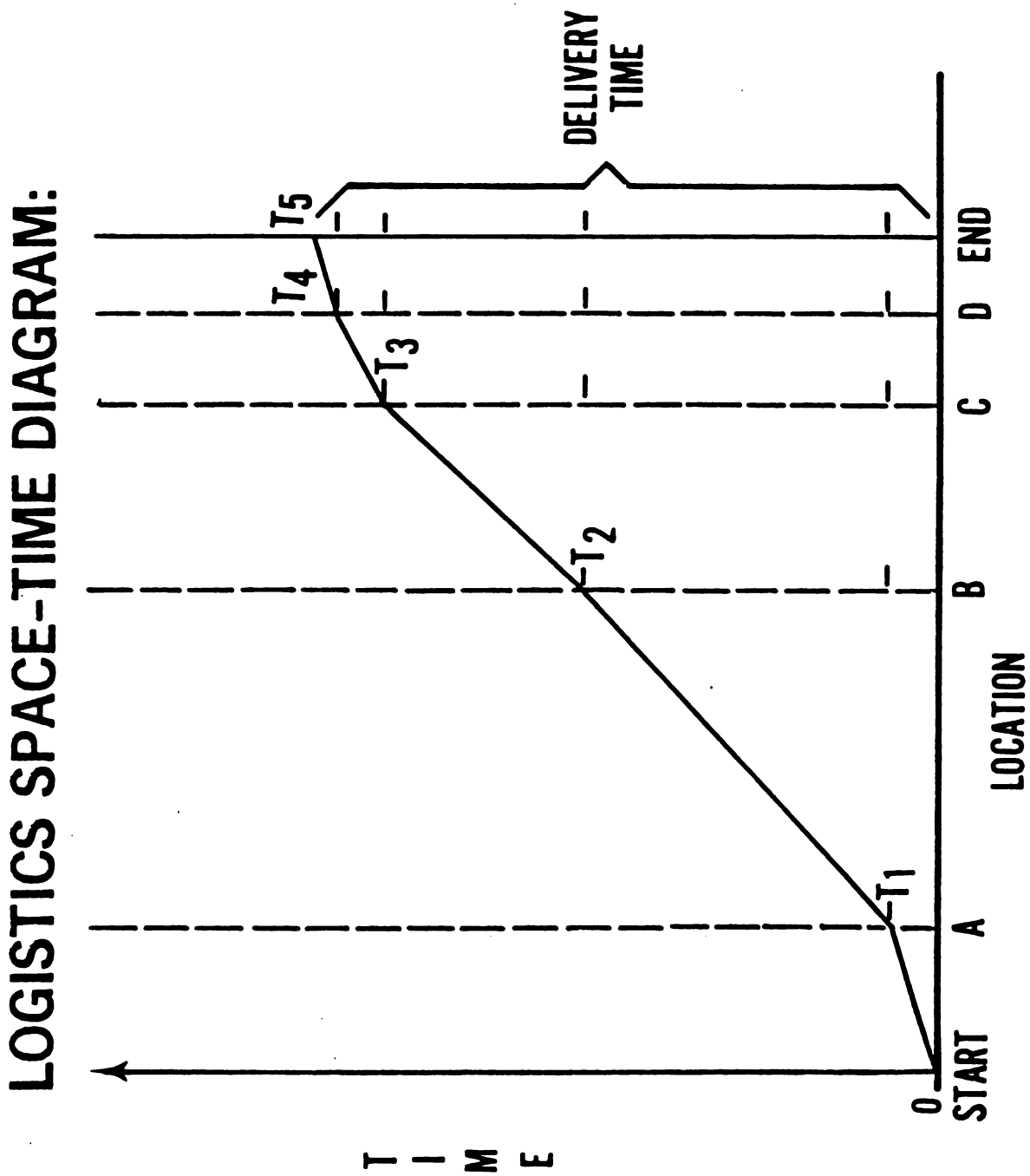
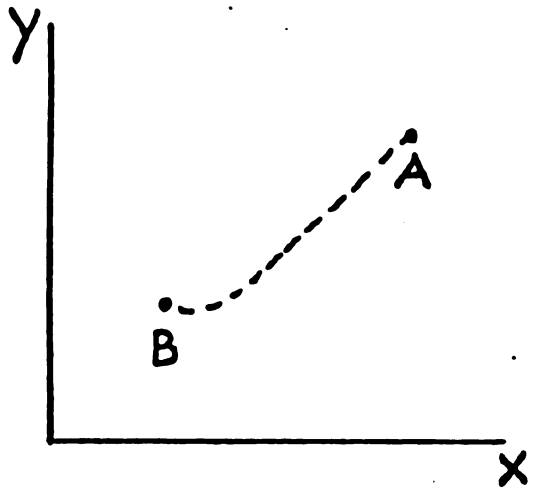


FIGURE 2

Consider the problem illustrated here: a ball is to roll frictionlessly on some path between points A and B. The path which allows the ball to complete the trip in the least possible time must be determined. This famous problem (known as the brachistochrone problem) is solved in many textbooks, and is referred to here because it illustrates the basic structure of all such problems. Both coordinates are constrained; the locations of points A, B are fixed in both x and y. In the y direction, the additional constraint of a fixed force (gravity) is imposed. These type problems are solved by defining a functional



$$F(y(x),x;y'(x))$$

where $y(x)$ defines the curve which provides the least travel time between A and B (called the optimal path), and y' is defined by

$$y'(x) = y(x) + \alpha\eta(x) \tag{1}$$

$y(x)$ is assumed to be continuous, and $y'(x)$ represents the functions which describe all paths between A and B which are <u>not</u> the optimal path $y(x)$. Thus, $\eta(x)$ represents a "difference" function between any non-ideal path and the ideal path. α is assumed to be small.

The task appears complicated because we do not know the optimal path $y(x)$, but rather are trying to solve for it. Recognizing that $\eta(x)$ can have an infinite number of forms, the equation to solve is

$$J = \int_A^B F(y(x),x;y'(x))dx \tag{2}$$

so that

$$\left.\frac{\partial J}{\partial \alpha}\right|_{\alpha\to 0} = 0. \tag{3}$$

These two equations basically say the following: Examine all possible paths $y'(x)$. For each one, look at the first derivative of J over the entire path and find the extremal path (first derivative equal to zero). If the

first derivative becomes zero when the value of α becomes zero, the function $y'(x)$ is actually $y(x)$, the optimal path we are searching for.

Gelfand and Fomin demonstrate that the problem is mathematically solvable for virtually all $y(x)$ and $\eta(x)$, and they describe exhaustively how it is done. For the purpose of least-time analysis, only 3 results need to be noted:

a) There is always an optimal path, and it can always be found.

b) The optimal path is unique; there cannot be more than one "optimal" path for a given set of constraints.

c) If $y(x)$ is not continuous but is piecewise continuous with n pieces, the optimal path is found by substituting for J:

$$J = \int_{A}^{\alpha_1} F(\,)dx + \int_{\alpha_1}^{\alpha_2} F(\,)dx + \cdots + \int_{\alpha_n}^{B} F(\,)dx \,. \qquad (4)$$

This says that if a path is not continuous, the optimal path can still be found by finding the optimal path for each segment which is continuous.

This is sufficient to establish a mathematical basis for least-time analysis. Although the methodology clearly belongs to the class of tools known as "critical path" methods, its development and motivation are quite different from most methods in this class. In fact, least-time analysis appears to be the first critical path method derived solely from a variational calculus approach.

Before leaving this section, it is appropriate to compare least-time analysis with another and perhaps more familiar critical path method, Program Evaluation Review Technique (PERT). As applied to logistics and similar scheduling problems, PERT attempts to identify the slowest among multiple paths so that coordination can be effected which will bring all the paths to a common end point at the same time. Least-time analysis, in contrast, takes a situation where a single path exists and attempts to identify the slowest segment (or, perhaps, the slowest among possible alternate paths) in order to reduce the time required for (and thus optimize) the total path. Because of this essential difference, least-time analysis appears to have at least two advantages over PERT: the data requirements are less, and the answer derived from least-time analysis is unique (there is only one optimal path).

V. Application of Least-Time Analysis. It is not necessary to use the extensive mathematical formalism behind least-time analysis in order to obtain useful results (although the existence of a mathematical background is tantalizing for logistics modelers). Application of least-time analysis requires basically four steps.

First, the level of application must be determined. Depending on how much detail is desired in the analysis, the method can be applied to parts, subassemblies, systems, supply classes, or adjuncts to systems.

Next, a linear delivery model (Shannon model; see figure 1) is constructed for the supply path at the chosen level of application. This model is used to identify each handling step or path segment in the supply process.

Next, the time required for each step is determined. This determination can be made by "desk-top" analysis, literature research, or testing/observation. When these times are known, a space-time diagram such as figure 2 can be plotted.

Finally, from the space-time diagram and the principle of analysis given in section III, all path segments which are not believed to be optimal can be identified. If the time required for each step was determined by a literature search or by desk-top analysis, appropriate issues/criteria for testing can be written. If the times required for each step were found by testing or observation, they can be analyzed to see if new issues/criteria or further testing will be necessary.

In addition to being potentially useful for attacking problems which are not easily handled by other methods, least-time analysis fits nicely within the conventional framework of logistics and thus allows quantification of some concepts which are widely considered to be only qualitative in nature. For instance, the logistician speaks of the three products of logistics: readiness in time of peace, surge at the outbreak of hostilities, and sustainability until the fighting is ended. Intuitively the meaning of these terms is obvious, but they have not previously been quantified for use by the test/evaluation community. Figure 3 illustrates how least-time analysis combines with a little common sense to make these terms quantitative. Readiness represents the "relaxed" or "natural" state of the supply system, with logistics materiel being delivered only at the rate required for normal operation of the military unit. Hence the required times for each step (and the overall delivery time) can be determined reliably and with high precision from simple analysis of records already existing. In the surge mode, the curve on the space-time diagram flattens out greatly because the delivery time for any materiel must decrease (note how this is equivalent to the conventional method of representing surge as a drastic increase in the quantity of all supplies). In the sustainability mode, the curve is not as flat because the delivery of supplies is not so critical as in the initial surge, but the delivery time must still be shorter than in the readiness phase. Using figure 3, the logistician could now quantify these - for instance, by requiring the delivery time during surge to be one-half of the delivery time during readiness. This gives the tester a quantitative way to evaluate the logistics supply system.

It is also of interest to note that some other types of information may be extracted from a least-time analysis. For example, some estimate of vulnerability of a system can be obtained by recognizing that the steps on the supply path which require more time are more vulnerable to enemy attack and disruption than steps which proceed more rapidly. The caveat to this is that vulnerability is actually more complex and involves more factors than simply the time spent on one path or in one place, so any vulnerability derived from

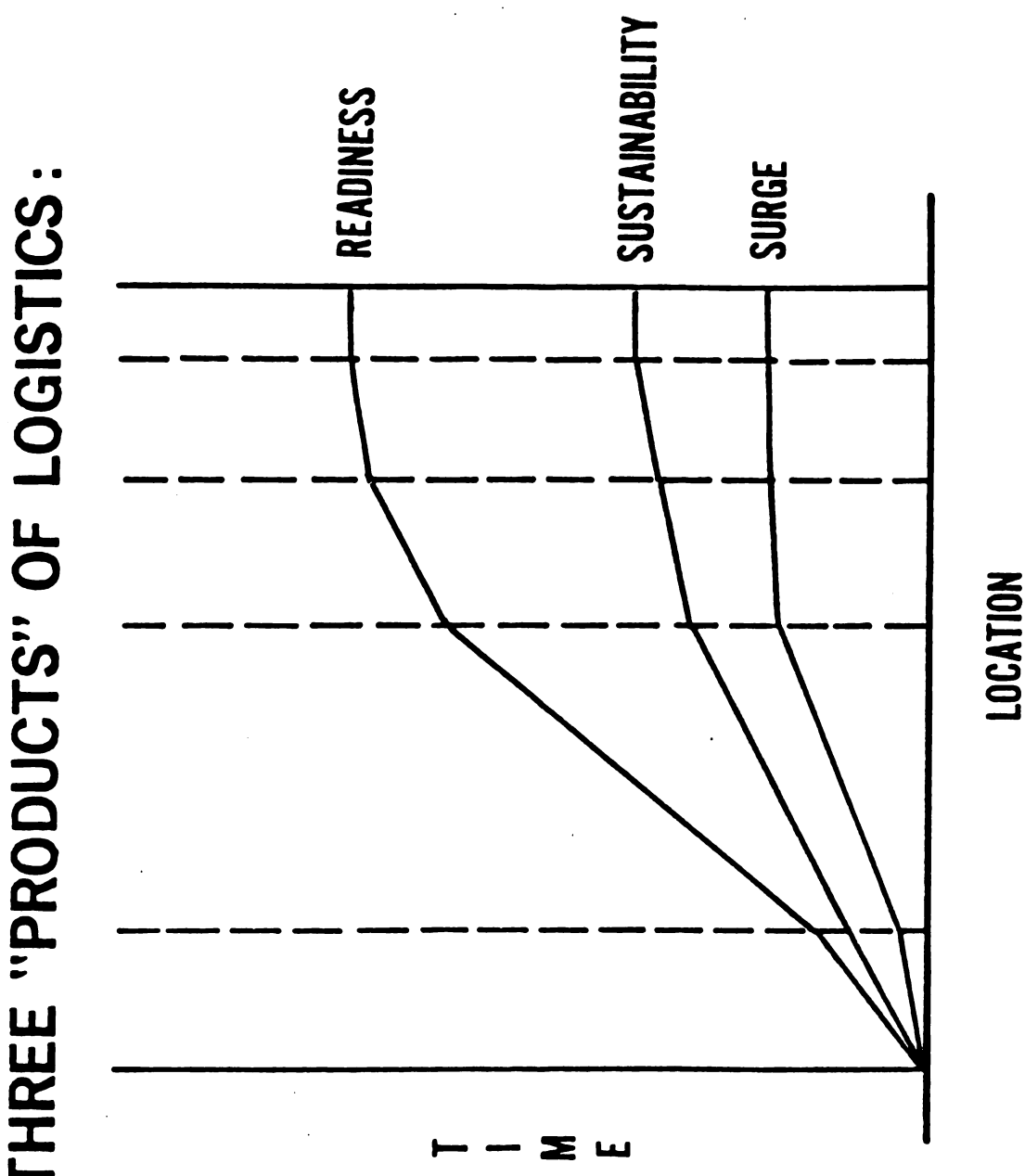


FIGURE 3

least-time analysis is strictly an estimate and should be used only with great care.

VI. An Example: Logistics over the Shore. Logistics over the shore (LOTS) was a major Army-Navy joint test which presented particular problems to the test community because it initially lacked good issues, criteria, and measures of effectiveness or performance. Some retrospective analysis of LOTS using least-time analysis will illustrate the usefulness of the method. LOTS represented the sustainment phase of logistics, with the beachhead secured and the threat from enemy fire or troops to logistics units not significant. Using several fully-loaded ships, the exercises made use of all possible methods to unload cargo of all types and move it to the beaches in the most expedient way possible. The test was to answer two general questions:

a) What is the most efficient method among all possible ways to transfer cargo from ship to shore?

b) Where is improvement needed in all methods to make them more efficient?

The results of the test were assigned special significance because they would be used to rewrite doctrine for beachhead supply and standard operating procedures (SOPs) for logistics units in ship-to-shore operations, make revisions to tables of organizational and equipment (TOEs), make possible changes in force structure, and support production decisions for amphibious lighters and other ship-to-shore watercraft. Because of space limitations, the example considered here is somewhat abbreviated and does not represent the entire LOTS exercises.

The decision was made to break down the analysis to the "type of cargo" level, and accordingly the first two classes to be considered were breakbulk cargo and containerized (non-liquid) cargo. Figure 4 shows the linear delivery model for breakbulk cargo, constructed to show all possible combinations of handling methods used to transfer the cargo from ship to shore. Analysis then shows that containerized cargo is handled in exactly the same ways, so the linear delivery model would be the same.

For breakbulk cargo, several questions can immediately be generated from figure 4. These include:

a) Which path represents the most efficient delivery method?

b) What improvements can be made to each path to decrease its delivery time?

c) What safety hazards are present on each path?

d) Is sufficient equipment and rigging available to execute all paths without delay?

BREAKBULK CARGO FLOW

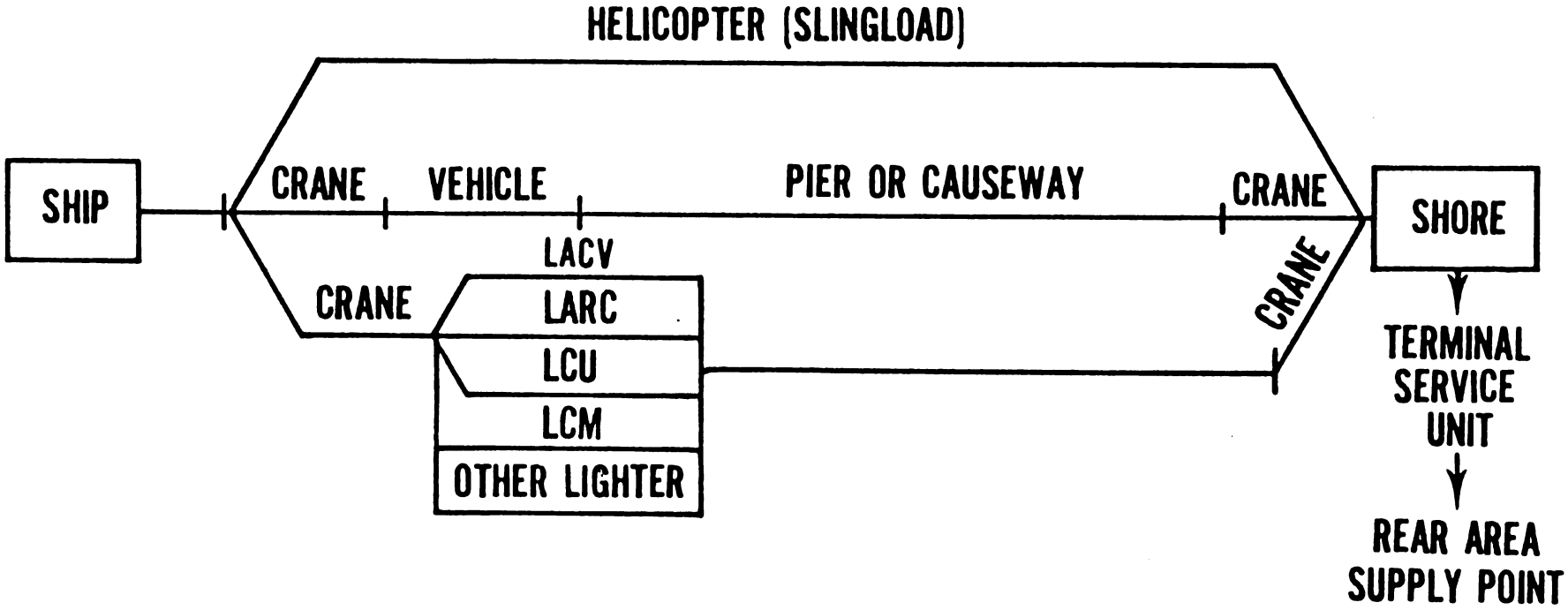


FIGURE 4

And, when containerized cargo is considered, the above questions remain and there are additional questions:

e) Is it more efficient to move containerized cargo than breakbulk cargo?

f) What improvements to breakbulk cargo handling would increase its efficiency to be competitive with containerized cargo handling?

A combination of testing and desktop analysis would be used to construct space-time diagrams for each path and for each cargo type, and from these diagrams the questions listed above could be turned into testable issues with quantitative criteria. As either a test planning aid or a data analysis aid, least-time analysis is very useful because the problem could not easily be attacked with any other method available. The analysis would be continued in the same manner for all other types of cargo (gasoline, vehicles, etc.).

VII. Conclusion. As stated in the abstract, least-time analysis is still in development. It has demonstrated its usefulness in generating issues and criteria for logistics-related testing, and it shows some promise of being useful to logistics systems designers and modelers. Much of its utility comes from the property of fitting into a conventional framework of logistics (that is, having the same point of view and not requiring new language or ways of thinking about logistics) but taking a somewhat different approach than conventional methods of analysis. Part of its strength comes from its solid mathematical foundation, and future development will involve further exploration of the underlying mathematics to identify useful results which increase the method's applicability.

Bibliography

Engel, John E., ed., 1980: Quantitative Tools for the Logistics Manager, Air Force Institute of Technology Textbook #LS-32, Wright-Patterson AFB, Ohio, DTIC #AD 121990, DLSIE # LD54498A, 645pp.

Gelfand, I.M., and S.V. Fomin, 1963: Calculus of Variations (Translated and Edited by Richard A. Silverman), Prentice-Hall, Inc., Englewood Cliffs, N.J., 232pp.

Marion, Jerry B., 1970: Classical Dynamics of Particles and Systems, 2nd edition, Academic Press, N.Y., Chapter 6, P. 178-194.

Stevens, Roger T., 1979: Operational Test and Evaluation: A Systems Process, John Wiley and Sons, New York, 275pp.

COMPARING SHOCK SENSITIVITY FOR TWO EXPLOSIVES

Gary W. Gemmill
and
Audrey E. Taub

Naval Surface Weapons Center
Dahlgren, Virginia 22448

ABSTRACT

A procedure for comparing shock sensitivity of two explosives for small sample sizes is described. In this context, shock sensitivity refers to the probability of detonation when an explosive unit is exposed to a shock stimulus. The comparison between explosive sensitivities is made by testing the hypothesis that the 100pth percentile of the population of critical shock values for a standard explosive is greater than the 100pth percentile of critical shock values for an experimental explosive. The logistic function is assumed as the distribution function for both explosives, and equality of variance (parallelism) between distributions need not exist. The procedure utilizes three fixed levels of stimulus to which explosive units are exposed and a quantal response, detonation or non-detonation, observed for each. Monotonized and truncated estimates of the detonation probabilities are used to compute minimum logit chi-square estimates of the logistic function parameters for each explosive. The test is based on the estimated difference between the 100pth percentiles of the two explosives. Monte Carlo simulation has been used to evaluate the discriminatory power of the procedure over a wide range of parametric values. Examples are given for comparisons made at the 20th and 10th percentiles.

1. INTRODUCTION

During the development of experimental explosives for gun-fired projectiles, one of the critical concerns is the sensitivity of an explosive to shock. Shock in this sense refers to the rapidly developed pressure that the explosive sees when it is loaded into a projectile and fired from a gun. If the explosive is sensitive to this type of shock, a premature detonation may occur. The explosive will detonate while the projectile is still in or just being expelled from the gun barrel. This, of course, is a very hazardous and

undesirable situation.

In order to test explosives for shock sensitivity, a premature simulator test facility was fabricated at NSWC, Dahlgren, Virginia. The simulator essentially consists of a 1000 pound drop vehicle which can accommodate a scaled down projectile loaded with explosive. The test is conducted by dropping the vehicle and projectile from a predetermined height onto a heavy steel plate. At impact a plunger is forced into the projectile thus simulating the rapidly increasing pressure that occurs in conventional gun systems. The test result is either a detonation or non-detonation. For each test unit the observed outcome is associated with a given drop height. This kind of data has been collected for standard explosives which are used in Navy projectiles. Newly developed experimental explosives are also tested for shock sensitivity in this manner. A requirement exists for a procedure which can collect and analyze this data to determine if a given experimental explosive is less sensitive than a standard explosive.

2. EXPERIMENTAL CONDITIONS AND ASSUMPTIONS

The experimental conditions and assumptions under which this comparison will be made are as follows:

a. The response is a quantal-type response; that is, there is either a detonation or no detonation. This quantal response is related to the drop height stimulus level by some probability distribution. This probability distribution governs the probability of detonation for all drop heights for a particular explosive.

b. Knowledge of the standard explosive is assumed both with respect to the type of distribution and also the value of the distribution parameters. This information will have been gained

through sufficient testing of the standard explosive.

c. The experimental explosive is an unknown. For the purposes of the present discussion we shall assume that it has the same type of probability distribution as the standard. However, the robustness of this assumption is a question which can be investigated under the proposed comparison procedure. The values of the parameters of the experimental distribution are unknown.

d. The number of test units of the experimental explosive is severely restricted. The number of units available will be between 10 and 15. This restriction arises mainly due to a high cost per test and the large number of test conditions that need to be evaluated.

e. The measure of sensitivity is defined as the 100pth percentile of the probability distribution for a particular explosive, where $0<p<.5$. That is the value of drop height (stimulus level) at which 100p percent of the population of experimental units would detonate. The smaller the value of the 100pth percentile, the more sensitive is the explosive.

f. The comparison procedure should protect against selecting an experimental explosive over the standard explosive unless the data indicates with high probability that the experimental is less sensitive than the standard.

3. INITIAL EFFORTS

A computer program was written whereby sensitivity test data could be simulated and analyzed under certain distributional assumptions and for particular data collection designs. A normal distribution was assumed as the governing probability distribution. Available data provided no evidence to seriously question this assumption. Maximum likelihood estimation was used to estimate the parameters of the

distribution for the experimental explosive.

Three data collection designs, the up-and-down, Langlie and probit designs were evaluated for sample sizes of 10 and 20. A brief description of each design is given below. More detailed information can be found in references 4, 5 and 6.

The up-and-down design consists of choosing a starting height and test increment. Then the stimulus level is raised or lowered by the amount of the test increment according to whether a non-detonation or a detonation, respectively, was obtained on the previous test.

The Langlie design requires choosing two heights initially, one at which no test units are expected to detonate and one at which all test units are expected to detonate. The first stimulus level is the average of these two heights. Subsequent heights are chosen based on the type and sequence of outcomes from previous tests. Accordingly, the next stimulus level is obtained by counting backwards until a prior stimulus level is found which has inclusively encompassed an equal number of detonations and non-detonations. The average of this stimulus level and the most recent stimulus level becomes the next level. If no such prior level can be found then the next stimulus level is obtained by averaging the most recent stimulus level with the upper bound or lower bound depending on whether the previous outcome was a non-detonation or a detonation, respectively.

The probit design is the simplest of the three and requires only that a prescribed number of stimulus levels are chosen and a designated number of items are tested at each level. Data collection designs of this type are usually associated with probit analysis which is discussed in reference 5.

The maximum likelihood estimation procedure requires that a zone

of mixed results (ZMR) exists in the data before estimates of the parameters of the distribution can be estimated. A ZMR is simply an interval on the stimulus scale defined by the maximum non-detonation and the minimum detonation where the former exceeds the latter. Table 1 shows the percentage of simulated evaluations of a total of 400 which actually yielded ZMR's or usable data under optimal design conditions.

TABLE 1

Probability of ZMR

N	Up & Down	Langlie	Probit
10	65%	75%	60%
20	90%	95%	95%

For the probit design the values of N were actually 9 (3 observations at each of 3 stimulus levels) and 21 (3 observations at each of 7 stimulus levels). As one can see the percentage of time that estimates were not obtainable ranges from 40% to 25% for a sample size of 10. A sample size of 20 performs somewhat better. However, the figures shown are for optimal choices of design parameters. The choices for these design values must be made with respect to the unknown parameters of the distribution for the experimental explosive. This makes an optimal choice unlikely. For less than optimal choices of the design parameters the percentage of time a ZMR is obtained may go as low as 40%. This is unacceptable.

4. PROPOSED PROCEDURE

Based on the preceding results, we turned to a different distribution, the logistic function, and a different estimation

technique which can be easily applied when using a probit design. The logistic function closely approximates the normal distribution and, as stated in reference 5, there is little difference between conclusions drawn from an analysis based on either distribution. The cumulative distribution function is

$$p = 1/(1 + \exp(- \alpha - \beta x)),$$

where α and β are distribution parameters, x represents the value of the stimulus level (height), and p is the probability of detonation for a test projectile dropped from height x. The function is easily linearized by taking the logit $p = \ln(p/q)$, where $q = 1-p$ to obtain

$$L = \text{logit } p = \ln(p/q) = \alpha + \beta x.$$

The parameters of this distribution can be estimated using the minimum logit chi-square technique. The probit design is particularly appropriate for this estimation procedure, especially for the small sample size restriction. The 100pth percentile can then be estimated by substitution in the linearized form and solving for x to get

$$x(p) = \text{100pth percentile} = (\ln(p/q) - \alpha)/\beta .$$

This application was discussed by J. Berkson in references 2 and 3.

The comparison hypothesis is stated as follows:

H: The 100pth percentile of the standard explosive is greater than or equal to the 100pth percentile of the experimental explosive.

Accepting H, of course, implies that one concludes the standard explosive is less sensitive than the experimental explosive. The alternative hypothesis is:

H(a): The 100pth percentile of the standard explosive is less than the 100pth percentile of the experimental explosive.

The alternative hypothesis is accepted if H is rejected and this

conclusion implies that the standard explosive is more sensitive than the experimental explosive. We assume that the distribution parameters, α_S and β_S, for the standard explosive are known. Therefore, the 100pth percentile for the standard distribution, S(p) can be calculated.

A number, n (less than or equal to 5), of experimental explosive units are tested at each of three equally spaced heights making the total sample size, N, equal to 3n. An estimate of probability of detonation, p, is obtained at each height by dividing the number of detonations at that height by n. These estimates are monotonized to be non-decreasing thus satisfying the requirement of a non-decreasing cumulative distribution function. Monotonization, in this respect, is discussed in reference 1. Probability estimates of 1 and 0 may occur because of the small sample sizes used. When this happens a substitution must be made so that the logit value can be computed. The values of .95 for 1 and .05 for 0 are recommended as truncated estimates to be substituted in such instances. These values were chosen based primarily on an empirical rule given in one of the earlier Berkson papers and the fact that they provide a workable result for the small sample sizes of interest. Estimates of α_E and β_E for the experimental explosive can then be computed using the minimum logit chi-square techniques given by Berkson in reference 3. These estimates are designated a and b, respectively.

If b is non-zero we can estimate the 100pth percentile, E(p), of the distribution for the experimental explosive by substitution as shown,

$$\hat{E(p)} = (\ln(p/q) - a)/b.$$

The estimated difference between the 100pth percentile for the

standard explosive and the 100pth percentile for the experimental explosive is given by

$$D = S(p) - \hat{E(p)}.$$

The variance of this difference is given by

$$s^2 = [1/b^2]\,[(1/\Sigma\, nw) + (G/\Sigma\, nw(x - \bar{x})^2\,],$$

where

$w = \hat{p}(1-\hat{p})$, x = drop height, $\bar{x} = \Sigma\, nwx/\Sigma\, nw$ and $G = (\bar{x} - \hat{E(p)})^2$.

The summation is taken over the three values of height. We shall reject H if D is less than or equal to c*s where c is chosen by the experimenter to give an acceptable probability of rejecting H when it is actually true. This is a Type I error and for the stated H is equivalent to rejecting the standard explosive when it is actually less sensitive than the experimental explosive. The value of c is chosen to be negative because we want to reject H with low probability if H is actually true. If D is greater than c*s, all possible test outcomes with more detonations at the test level of greatest stimulus must be considered before H can be accepted. If any lead to a rejection, H must be rejected. The rationale for this is that if any of these more sensitive outcomes yield a rejection of H then less sensitive outcomes should not contradict that result by accepting H.

Figure 1 is an illustration of the comparison procedure for the case where b is non-zero. The stimulus levels have been scaled in approximate units of the standard deviation of the distribution for the standard explosive. The experimental explosive was tested at stimulus levels of 1, 2 and 3. S(.2) and $\hat{E(.2)}$ represent the 20th percentile and the estimate of the 20th percentile for the standard and experimental explosives, respectively.

An alternative procedure is required when b = 0 since we then cannot estimate a 100pth percentile for the experimental explosive. Consequently, we assume all experimental tests are conducted at the lowest of the three test levels. The lowest test level is the level of least stimulus and, therefore, the level with the least probability of detonation. This is a conservative position in terms of the experimental explosive. If all tests had been done at the lowest test level (LTL), certainly no greater number of detonations would have occurred than the number actually observed. Next we compute an estimate, $\hat{p}$, for the probability of detonation at this stimulus level. An upper 95% confidence bound for the probability of detonation at this test level is computed and, if necessary, truncated at .95. The logit of this upper confidence bound is calculated and designated as the logit truncated bound or LTB. The LTB value is substituted into the logit function for the standard explosive to estimate an upper bound for an equivalent stimulus level, $S(\hat{p})$, required by the standard explosive. If the difference between this equivalent stimulus level and the lowest test level, $D = S(\hat{p}) - LTL$, is less then K, then H is rejected. As with c, K is chosen as a negative value in order to achieve an acceptable probability of rejecting H when it is true. The value of K must be adjusted in magnitude when the value of the lowest test level changes.

Figure 2 illustrates graphically the procedure for b = 0. This example is for a design with test stimulus levels of 1, 2 and 3 and the stimulus levels have been scaled as before. LTL represents the lowest test level and $S(\hat{p})$ represents an estimate of an equivalent stimulus level for the standard explosive.

5. EVALUATION OF PROPOSED PROCEDURE

Computer simulation was used to evaluate the proposed procedure. A computer model was written to generate sample data from an underlying logistic distribution for the experimental explosive. The simulated data are then compared to a nominal standard explosive using the proposed procedure. The comparison procedure determines an acceptance or rejection of the hypothesis H. By replicating this process many times one can obtain an estimate of the probability of accepting H under a given set of distributional assumptions.

The main objective of the evaluation was to demonstrate the feasibility of the proposed comparison procedure and determine the acceptance/rejection behavior over a wide range of distributional assumptions for the experimental explosive. The distribution for the nominal standard explosive was chosen to be a logistic distribution with $\alpha_S = 0$, $\beta_S = 1.668$. These parametric values were chosen because they matched a logistic distribution to a standard normal distribution (mean = 0, variance = 1) at minus one-sigma and plus one-sigma abscissa values. This, in effect, provided a standardization with respect to the parameters of the standard explosive. In order to pursue the evaluation systematically with respect to possible experimental distributions, two relative quantities, R and delta, were defined. R represents the ratio of β_S to β_E and is equivalent to the ratio of the standard deviation of the experimental explosive to the standard deviation of the standard explosive. Experimental distributions with values of R from .5(.5)2(1)4 were chosen to cover a wide range of relative variation. Delta represents the median of the experimental distribution minus the median of the standard distribution, that is, delta = E(.5) - S(.5). This provided a

convenient means of relating an assumed experimental distribution to the standard distribution with respect to location. For each value of R the evaluation was done for values of delta from -1(.5)2(1)5. This generated 60 experimental distributions to be compared to the standard distribution.

Two probit designs were selected for comparison. The first design used stimulus levels of 1, 2 and 3, c = -1.5 and k = -1.5. The second design used stimulus levels of 3, 4 and 5, c = -1.5 and k = -2.5. In terms of the assumed standard distribution, the stimulus levels for the first design correspond to approximately one-, two- and three-sigma units above the median value, whereas the stimulus levels for the second design correspond to approximately three-, four- and five-sigma units above the median value. These designs were chosen with the idea of "overtesting" the experimental explosive. This was an attempt to provide additional protection against rejecting the standard explosive unless the experimental explosive were clearly less sensitive. The values of c and k were determined empirically by running the simulation with arbitrary trial values. Both designs were evaluated with p = .2 for all 60 combinations of R and delta. Additionally, the first design was also evaluated with p = .1 for all combinations of R and delta. The evaluations were done with n = 5 and 10 which correspond to total sample sizes of 15 and 30, respectively. The probability of acceptance of H was estimated for a particular condition by replicating that condition 800 times. This number of replications ensures that the estimated probability is within .03 of the true probability with 95% confidence.

6. RESULTS

The results are shown in Figures 3-11. For both sample sizes each figure displays the plot of estimated probability of accepting H versus delta for a particular design, R-value and p-value. Figures 3-5 and 6-8 show the results for the first design for p-values of .2 and .1, respectively. Figures 9-11 show the results for the second design for p = .2. Only the results for R-values of .5, 1 and 4 are shown in the interest of saving space. These R-values, however, encompass the range used in the evaluation and are sufficient to show the effect of relative variation upon the estimated probability.

As stated above delta is the median of the probability distribution for the experimental explosive minus the median of the probability distribution for the standard explosive. For an experimental explosive with specified variability, a decrease in sensitivity is associated with an increase in delta. A comparison procedure based on sensitivity, as defined previously, should yield a decrease in the probability of accepting the standard explosive as delta increases. Also, a particular value of delta exists below which H is true (the experimental explosive is more sensitive) and above which H is false (the standard explosive is more sensitive). The probability of accepting the standard explosive should go from one to zero at this value of delta, of course, for an ideal comparison procedure. These critical values of delta are given in Table 2 for the indicated values of R and p.

TABLE 2

Critical Deltas

R	p=.1	p=.2
0.5	-.66	-.42
1.0	0.00	0.00
4.0	3.95	2.49

The closeness to which the procedure approaches the ideal varies for the illustrated examples. The capability to discriminate between explosives of different sensitivity is reasonably precise for the situations where the variation of the experimental explosive is less than or equal to the variation of the standard explosive, that is, when R = .5 and 1. This is shown in Figures 3, 4, 6 and 7 for the first design and Figures 9 and 10 for the second design. The probability of accepting the standard explosive drops quickly as the sensitivity of the experimental explosive lessens, that is, as delta increases. When the variation of the experimental explosive exceeds the variation of the standard explosive, that is, when R = 4, the discriminatory capability is much less precise for both designs as shown in Figures 5, 8 and 11. Here the probability of accepting the standard explosive lessens very gradually as delta increases.

Because of the severe consequences of incorrectly rejecting the standard explosive (Type I error), specific measures were incorporated to reduce the probability of this happening. However, as a result of reducing the probability of a Type I error, the probability of committing a Type II error is increased. In this context a Type II error means accepting the standard explosive when the experimental explosive is actually less sensitive. This effect is clearly demonstrated in Figure 3. In this case the standard explosive and the

experimental explosive are of equal sensitivity when delta is -.42. This is the critical value of delta from Table 2 for R = .5 and p = .2. For delta less than -.42, H is true. That is, the standard explosive is less sensitive than the experimental explosive and should be accepted. The probability of doing that is near 1 for delta smaller than -.42. As delta increases to .5 (the experimental explosive becomes less sensitive than the standard), the probability of accepting H remains high. Under these conditions a Type II error would be very likely because H is not true but would be accepted with high probability. In terms of explosives, this means the standard is more sensitive than the experimental but would be accepted with high probability. However, this kind of error is not as serious as incorrectly rejecting a less sensitive standard explosive in favor of accepting a more sensitive experimental explosive (Type I error). The price for keeping the probability of a Type I error small is accepting an increase in the probability of a Type II error. For both designs the desired small probability of a Type I error is attained for R = .5 and 1; and, as expected, this is accompanied by a relatively large probability of a Type II error. Additionally, the large probability of a Type II error extends over a greater range of delta for the second design. This is an effect of the increased stimulus values used in the second design. When the variation of the experimental explosive becomes relatively large, that is, when R = 4, neither design produces small probabilities of a Type I error. This is particularly true as delta approaches the critical value. This was noted earlier in the comments concerning precise discrimination between explosives with different sensitivities. This is caused by the relatively large variability of the experimental explosive.

Smaller values of c and k would decrease this probability of a Type I error but the precision of the discriminatory capability would remain low. Decreases in c and k may cause an intolerable increase in the probability of incorrectly accepting the standard explosive for R-values of .5 and 1.

The two values of p used with the first design provide an opportunity to observe the effect of the value of p upon the comparison procedure. The probability of accepting H for p = .2 is equal to or slightly lower than the probability of accepting H for p = .1 across all values of delta. This means that for delta less than the critical value, the probability of wrongly rejecting the standard explosive is greater for p = .2. When delta is greater than the critical value, the probability of wrongly accepting the standard explosive is greater for p = .1. A comparison using the smaller value of p would be of more interest and consistent with making the Type I error probability small. However, choosing too small a value of p may result in a less precise comparison. Additional simulations would be needed to offer more understanding of this situation.

Comparing the simulation results for the two sample sizes shows that for most values of delta the Type I and Type II error probabilities are less for n = 10 than for n = 5. This would be expected since the rejection values of c and k were held constant for both sample sizes. However, some exceptions do exist for certain regions of delta. These exceptions are also dependent upon the value of R. Ideally, the probability of accepting the standard explosive would be greater for n = 10 than for n = 5 when delta is less than or equal to the critical delta. For delta values greater than the critical value the reverse would be true. The probability of Type I

and Type II errors would then be less for the larger sample size in all cases. One can note that for R = .5 and 1 the curves for n = 10 cross below the curves for n = 5 after the critical value of delta has been reached. This means that for delta between the critical value and the cross-over point the probability of a Type II error is greater for n = 10 than for n = 5. When R = 4 and the variation of the experimental explosive is larger than the variation of the standard, the two curves cross before the critical value of delta. Here the probability of wrongly rejecting the standard explosive (Type I error) is greater for n = 10 than for n = 5 when delta lies between the cross-over point and the critical value. Adjustment could be made on the values of c and k to reconcile some of the apparent anomalies between the sample sizes. However, the effect of any adjustment is dependent upon the value of R. If adjustments are made to force a more consistent response for R = .5 and 1, the anomalies are spread over a greater range of delta for R = 4, and vice versa. This is one of the difficulties imposed by the initial ground rules which allow for differences in the variation for the standard and experimental explosives.

7. SUMMARY

The proposed procedure offers a methodology for comparing an experimental explosive to a standard explosive with respect to sensitivity under the stated ground rules. The procedure is particularly applicable to a situation where a limited number of experimental test units are available. The results of the simulated examples demonstrate the discriminatory capability of the comparison procedure. The computer model written to simulate the procedure can

be used to determine optimal design parameters with regard to probabilities of Type I and II errors. Also, the simulation model can be used to evaluate robustness of the procedure with regard to the assumed probability distribution of the experimental explosive.

In conclusion, the proposed procedure appears to provide a reasonable answer for every set of test outcomes. Additionally, by properly choosing values for the design and test parameters, the required protection against Type I and II errors can be obtained. This is a decided improvement over current comparison procedures being used in sensitivity testing of new explosives.

REFERENCES

1. Barlow, R. E., Bartholomew, D. J., Bremmer, J. M., and Brunk, M. D., (1972), STATISTICAL INFERENCE UNDER ORDER RESTRICTIONS; John Wiley & Sons, London.

2. Berkson, J., (1953), "A Statistically Precise and Relatively Simple Method of Estimating the Bio-Assay with Quantal Response, Based on the Logistic Function", Journal of the American Statistical Association, 48, 566-599.

3. Berkson, J., (1955), "Maximum Likelihood and Minimum χ^2 Estimates of the Logistic Function", Journal of the American Statistical Association, 50, 130-162.

4. Dixon, W. J., and Mood, A. M., (1948), "A Method for Obtaining and Analysing Sensitivity Data", Journal of the American Statistical Association, 43, 109-126.

5. Finney, D. J., (1952), PROBIT ANALYSIS, 2nd Edition, Cambridge University Press.

6. Langlie, H. J., (1962), "A Reliability Test Method for One-Shot Items", Aeronutronic Publication No. U-1792.

FIGURE 1

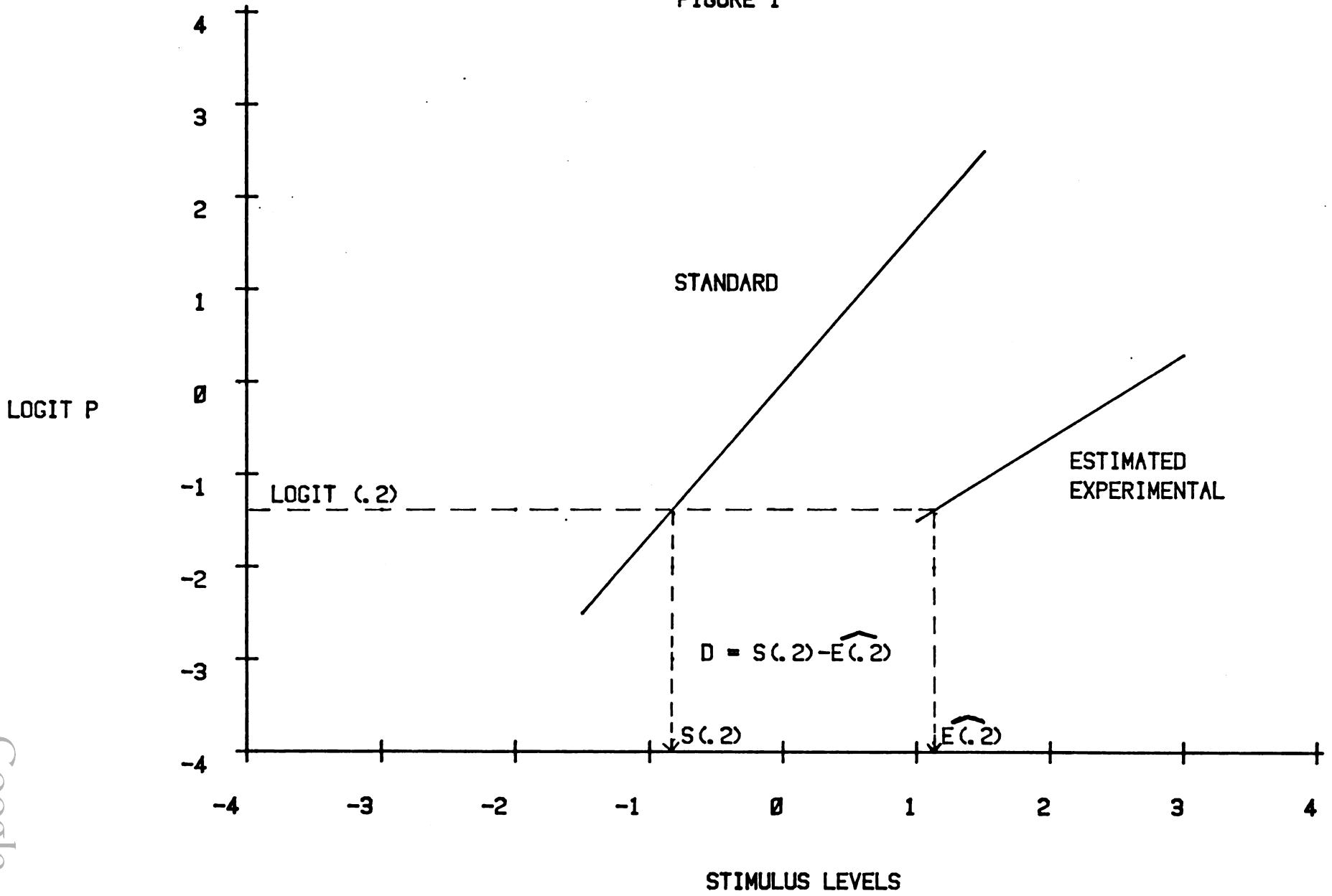


FIGURE 2

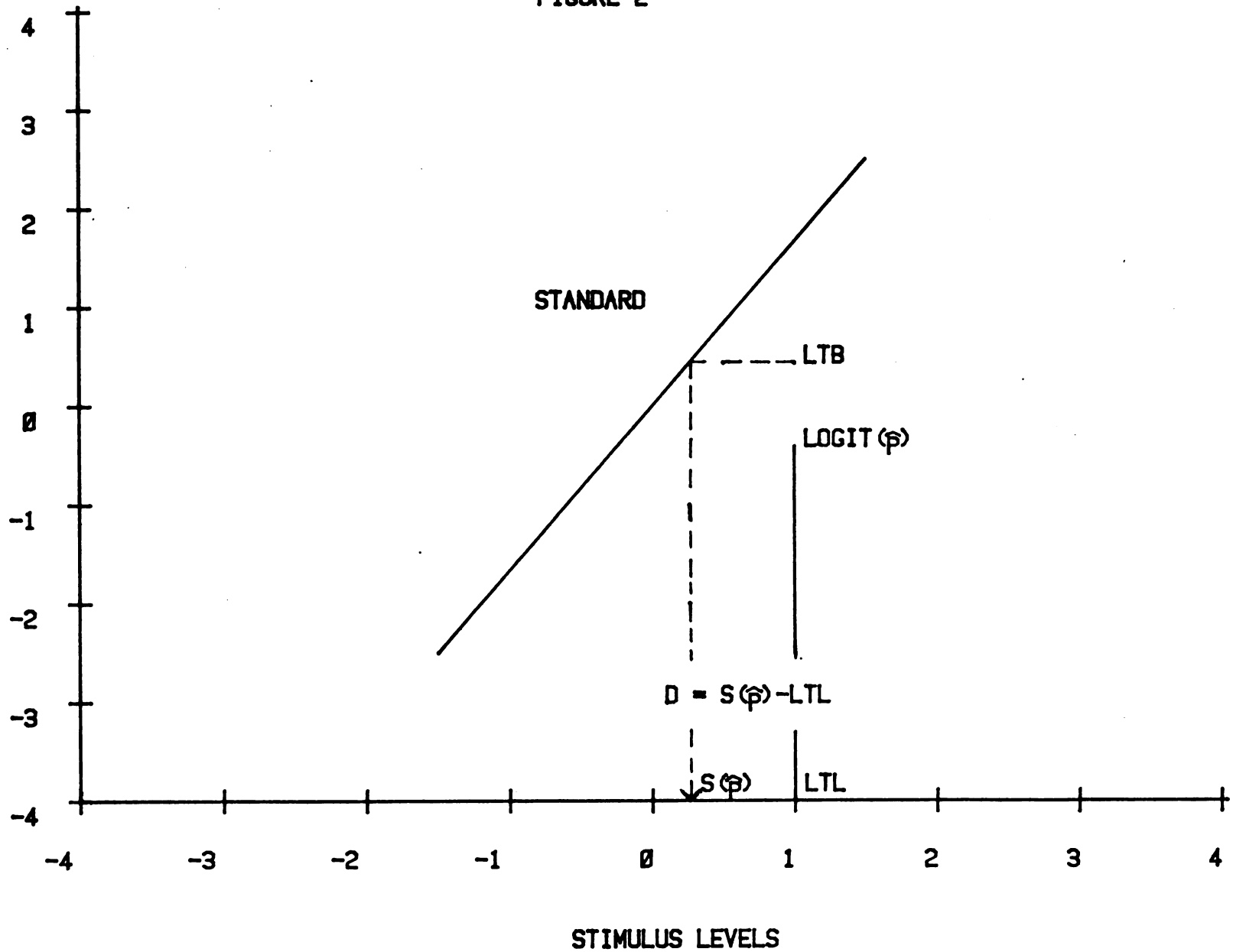


FIGURE 3

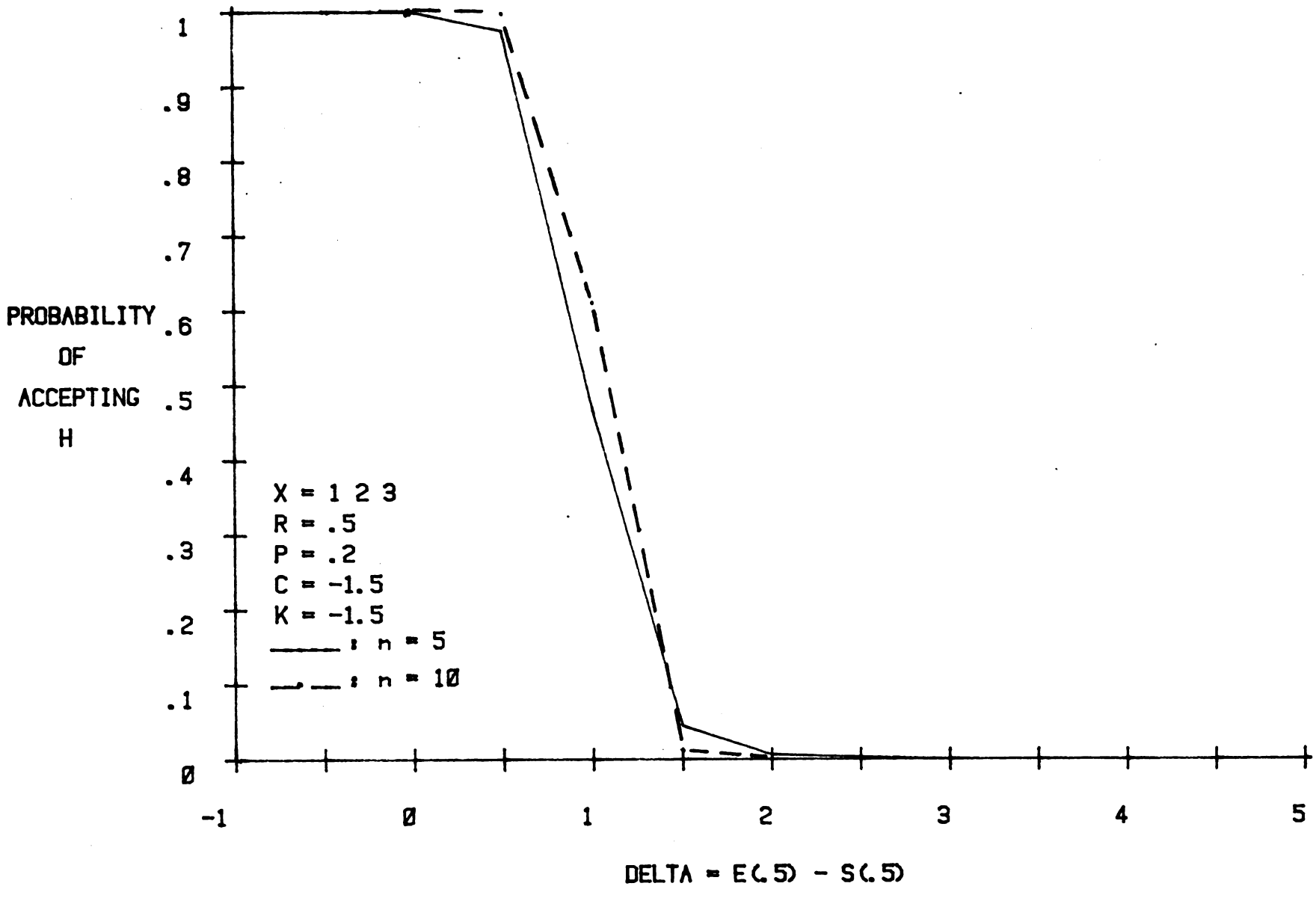


FIGURE 4

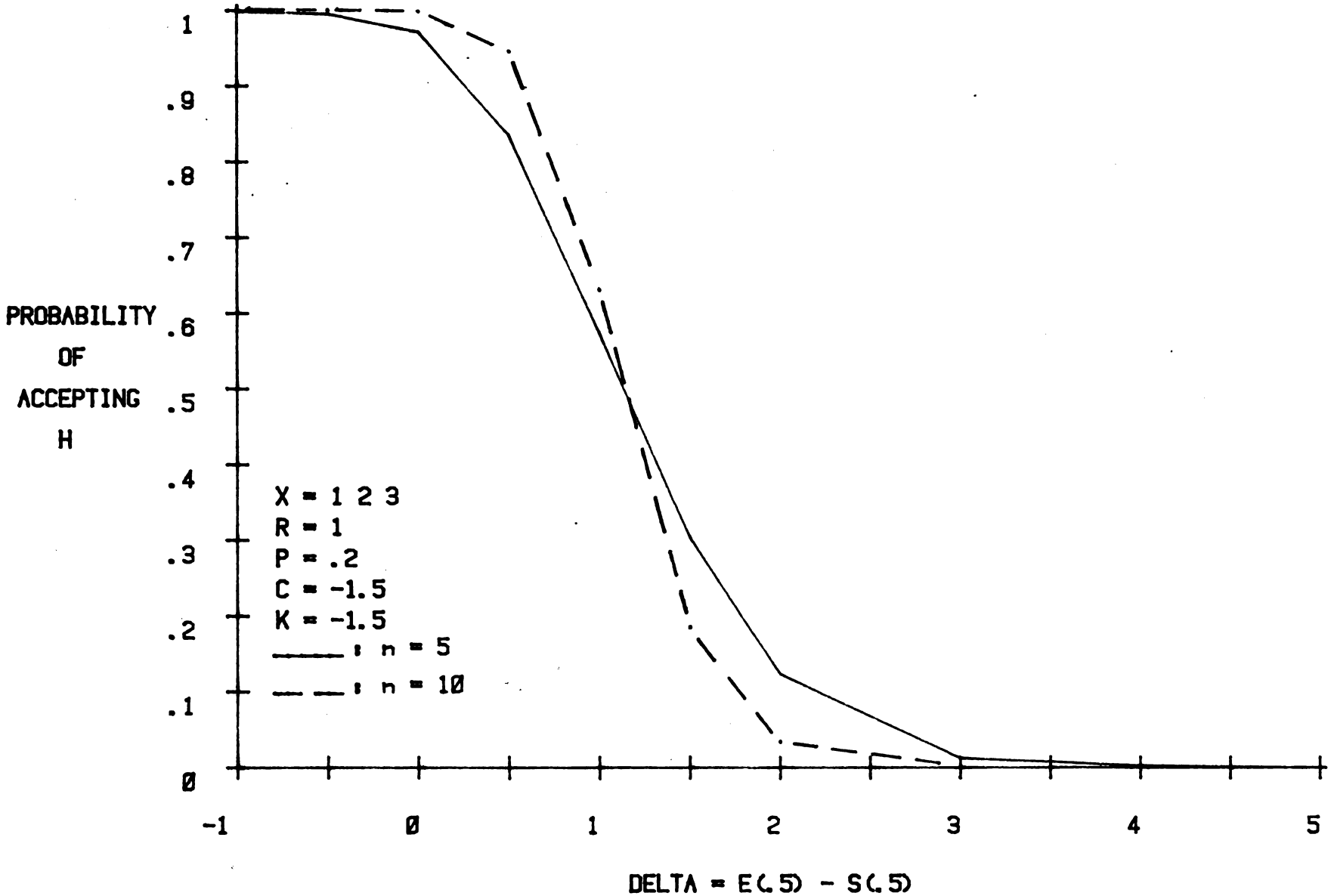


FIGURE 5

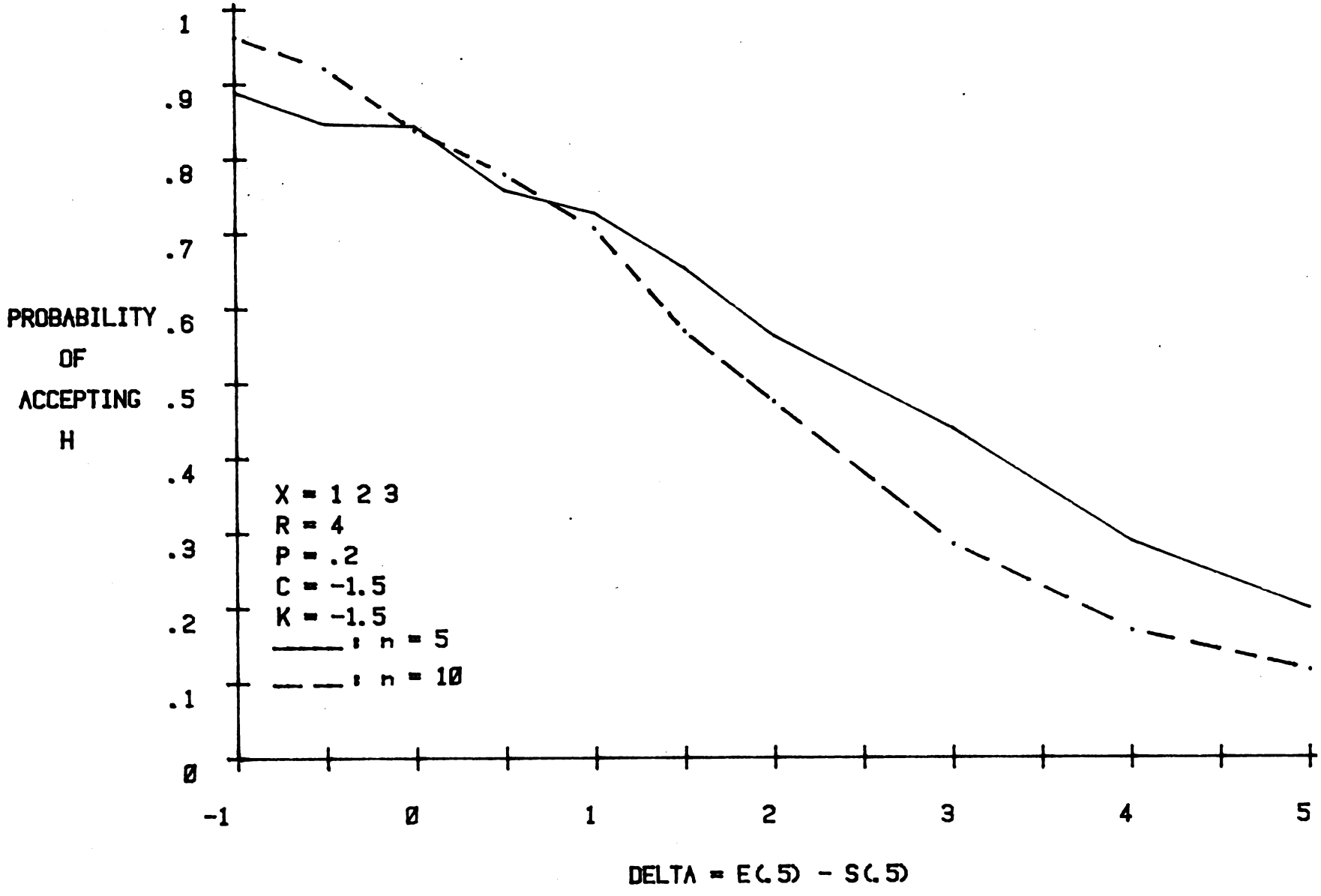


FIGURE 6

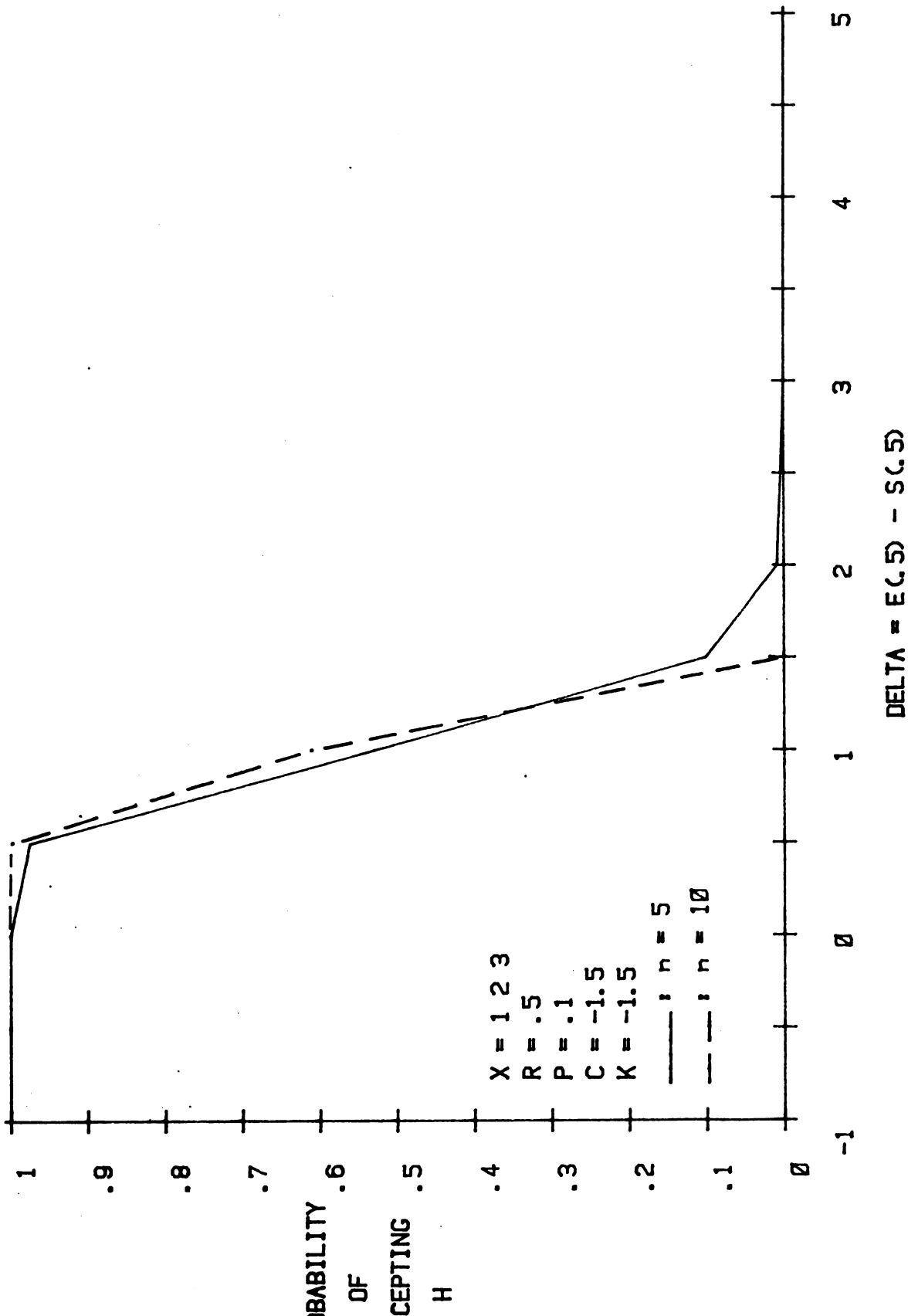


FIGURE 7

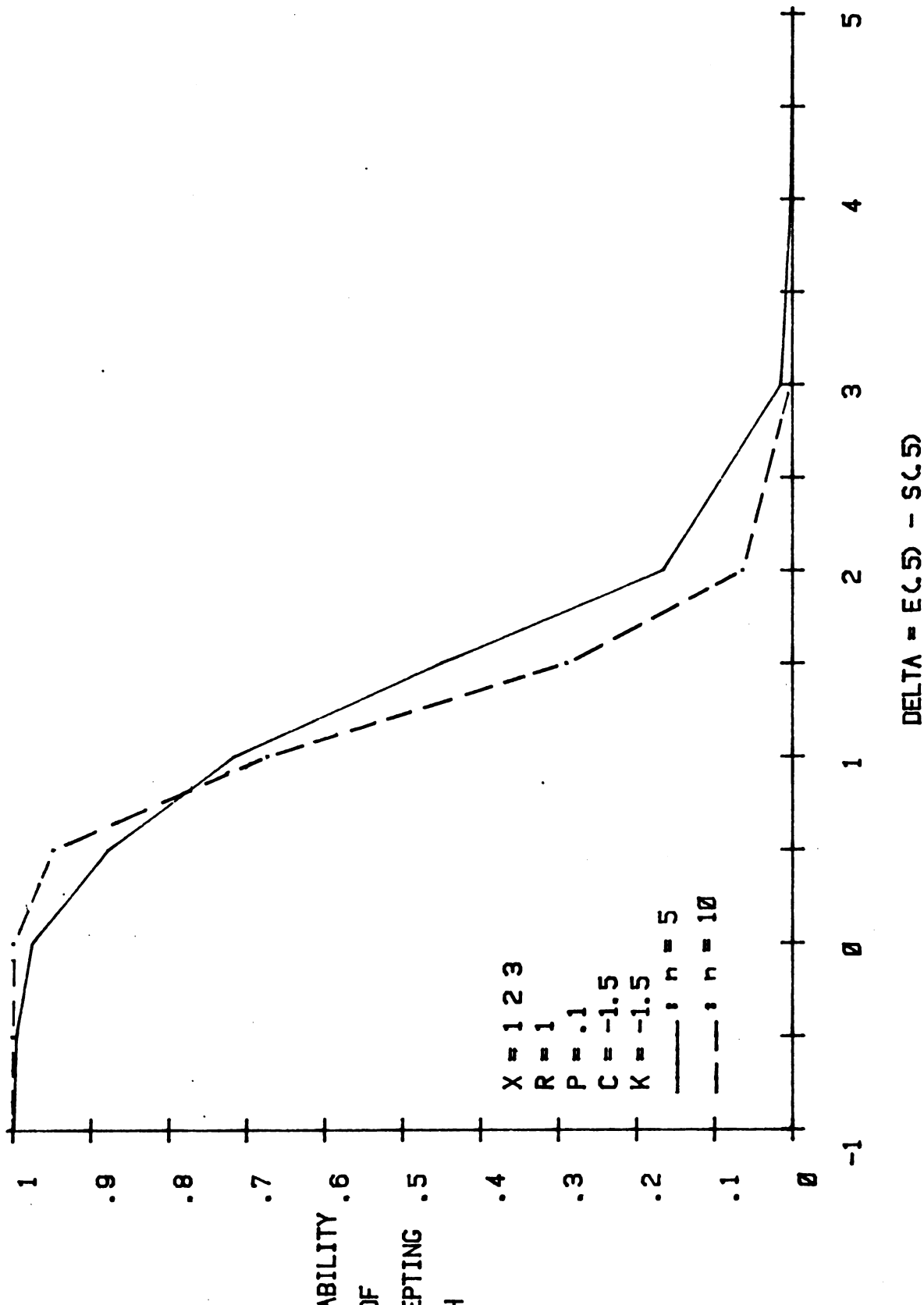


FIGURE 8

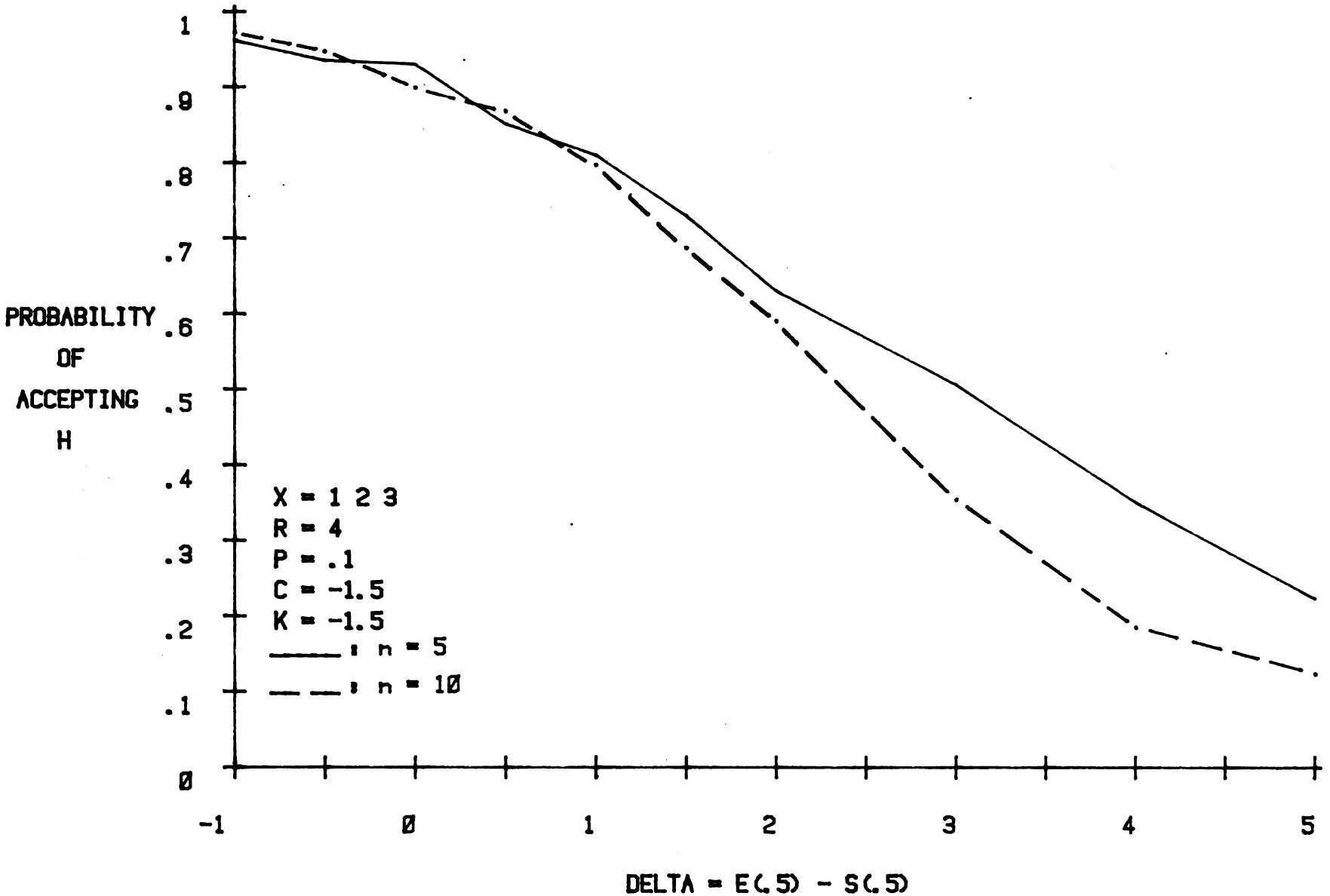


FIGURE 9

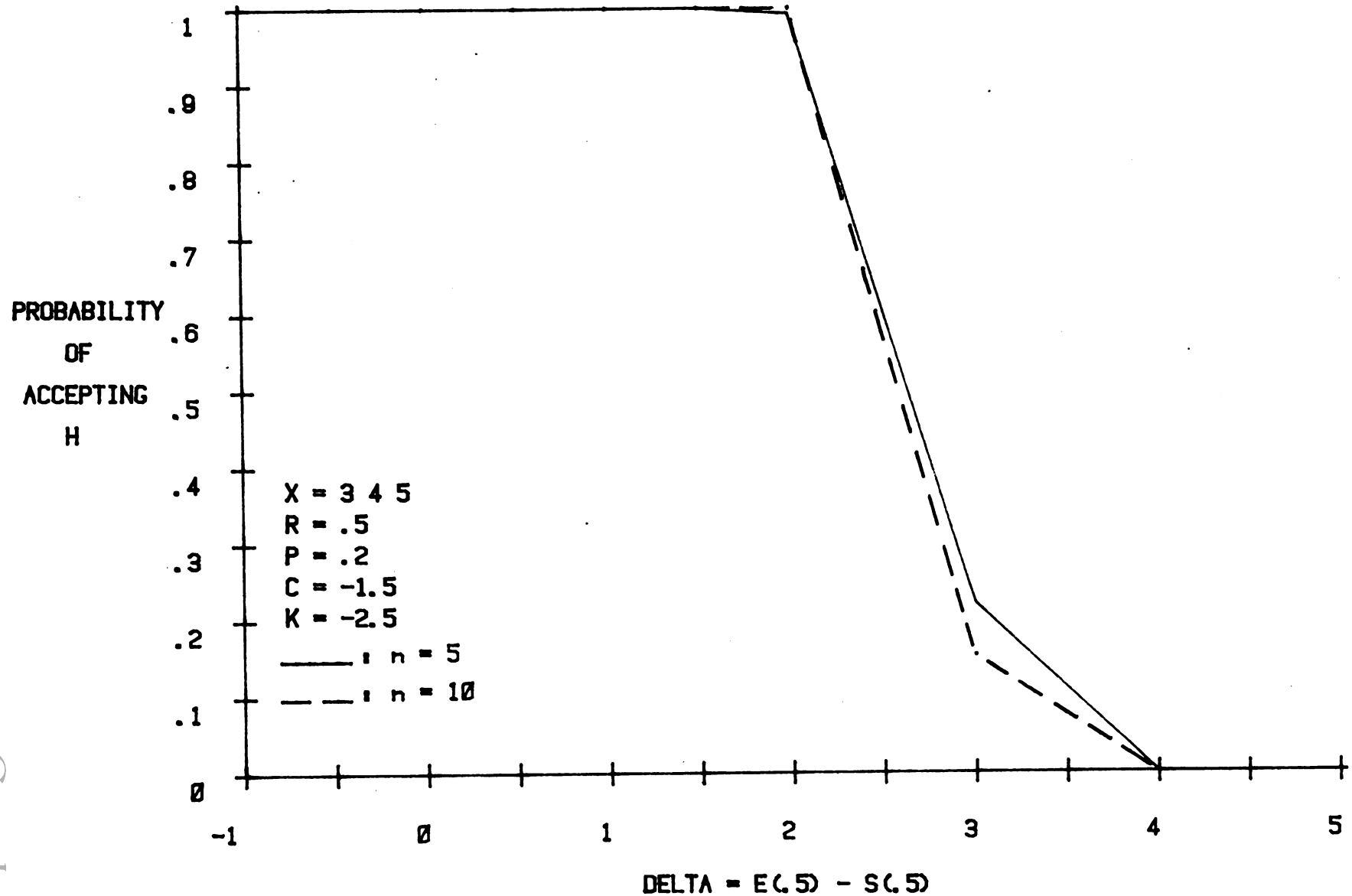


FIGURE 10

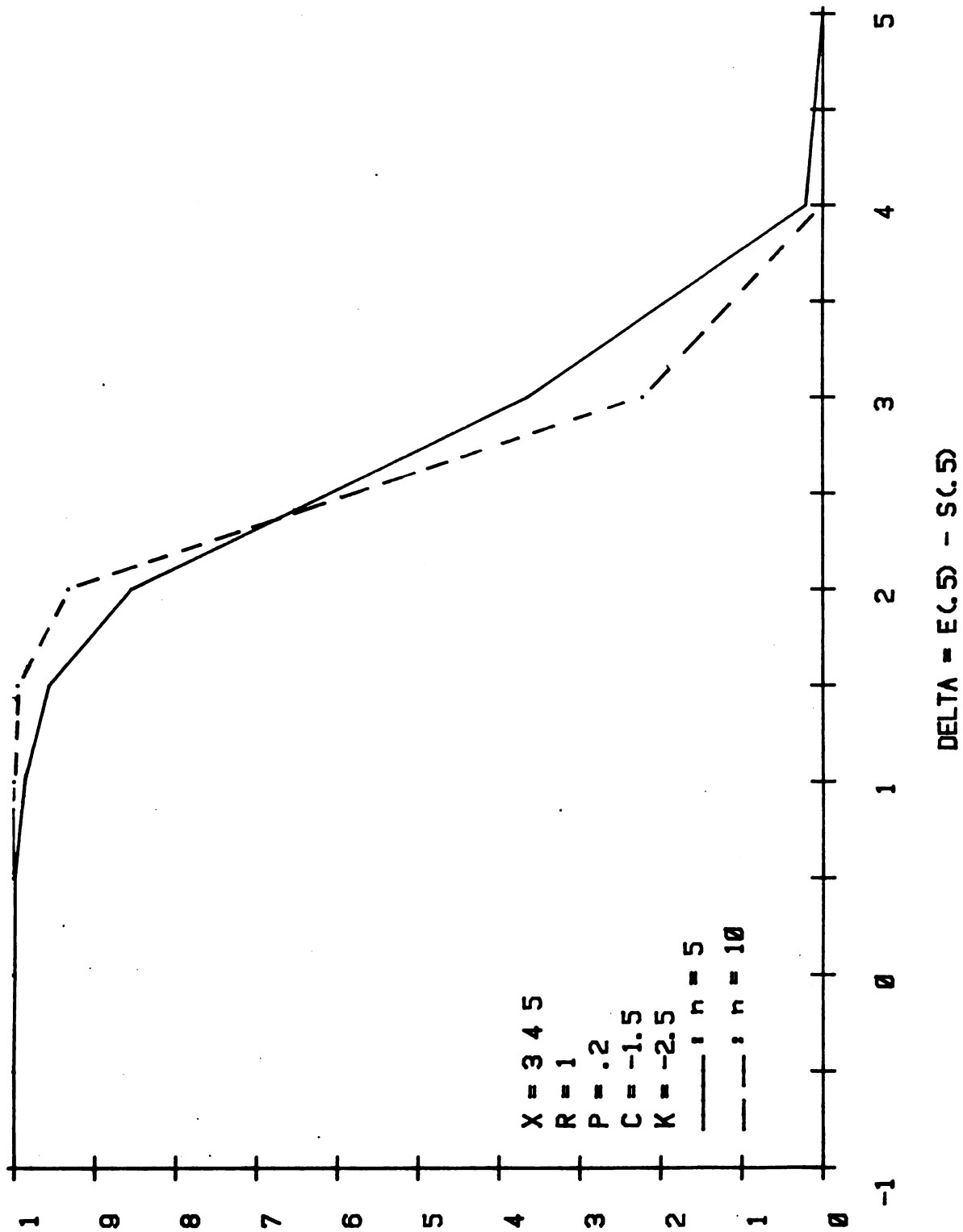
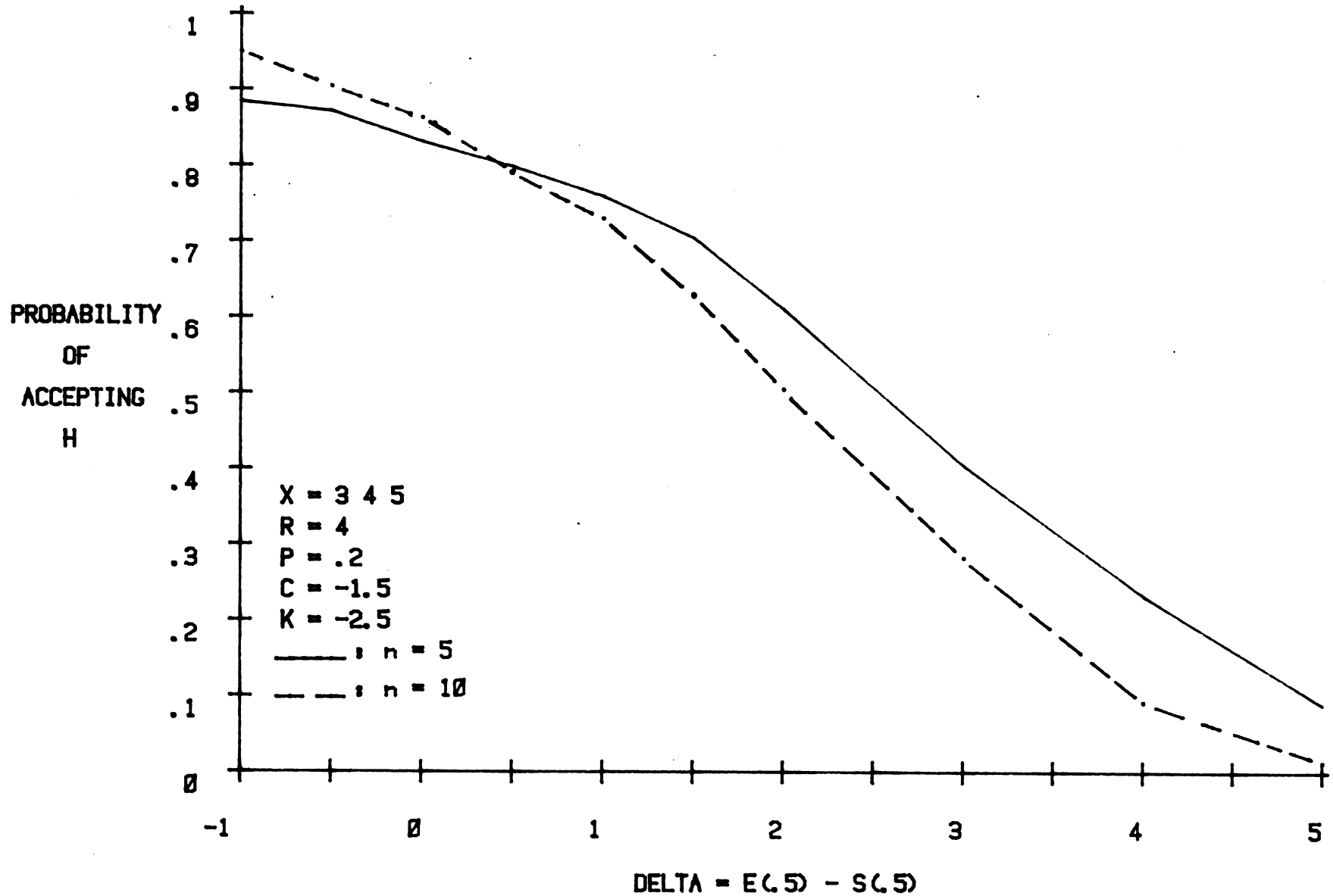


FIGURE 11



MULTIVARIATE DATA ANALYSIS APPLIED TO SKINFOLD MEASUREMENTS AND THE PERCENTAGE OF BODY FAT FOR BLACK AND WHITE MALE SOLDIERS

Karen Patricia Hobson, MAJOR, USA
Army Medical Specialist Corps
Chief Clinical Dietetics
Nutrition Care Division
William Beaumont Army Medical Center
El Paso, Texas 79920-5001

Eugene Dutoit
OR/SA Branch
Directorate of Combat Developments
US Army Infantry School
Fort Benning, Georgia 31905-5400

ABSTRACT:

The first author conducted a study to investigate racial differences in skinfold thickness measurements and the calculated percent body fat for normal weight and overweight soldiers. It was hypothesized that there was no significant differences between skinfold measurements and percent of body fat between black and white soldiers. Multivariate analysis of variance and discriminant analysis was used to analyze the data and test the experimental hypotheses. Different combinations of dependent variables were considered in the analysis to investigate the consistency of the statistical decisions between and within racial and weight groups. This study shows that the multivariate procedures used yielded consistent decisions for various combinations of the dependent variables.

INTRODUCTION:

Nutritional assessment using skinfold calipers to evaluate leanness/fatness has been used extensively since the development of calipers. For several years the United States Army has used skinfold measurements to screen its members for leanness/fatness and physical fitness. When an individual exceeds the weight for height required maximum, that person is evaluated on the basis of a percentage body-fat standard.

This research was designed to compare and evaluate the skinfold measurements and body-fat percent determination of male soldiers. It was hypothesized that (a) there is no significant difference between skinfold measurements taken from black soldiers and those taken from white soldiers, and (b) there is no significant difference in percent body fat between black soldiers versus white soldiers as determined by the established method. Skinfold measurements were made during routine physical examinations on basic trainees, and during an overweight

screening of career soldiers. Each group was divided by race and age.

PROCEDURE:

Subjects for the study were male soldiers at Fort Benning, Georgia, between 17 and 39 years of age. The purpose and procedures involved in the study were explained to each soldier before the participant signed a volunteer consent form. Four hundred fifty-two soldiers volunteered for the study. However, only 302 soldiers were used because those over 40 years of age, or members of other ethnic or racial groups were not included in the analysis.

Subjects were divided into two groups: Group I - young trainees from Regular Army, National Guard, and Reserve units who had just completed seven weeks of infantry basic training and Group II - career soldiers from an infantry brigade who had been identified as overweight by Army height-weight tables. Subjects for Group I were selected at random during a routine physical examination. The number of each race per group was kept equal for statistical comparisons. Each group is subdivided by race. Race was determined visually, by questioning the participant, and from medical history forms. The composition of each group is given in Table 1.

TABLE 1. Distribution of the Subjects by Race & Age

Age (Yrs)	Number of Subjects	(NORMAL WEIGHT) Group I Black	White	(OVERWEIGHT) Group II Black	White
17-22	208	92	92	12	12
23-29	84	21	21	21	21
30-39	10	—	—	5	5
TOTALS	302	113	113	38	38

ANTHROPOMETRIC MEASUREMENTS:

Height was measured in stocking feet with feet together, back and heels against the upright bars of a height scale, head in the Frankfort horizontal plan ("look straight ahead") and stand erect ("stand up straight"). All weights were recorded to the nearest quarter pound.

Four skinfold thickness measurements were taken with a Lange skinfold caliper calibrated to exert a pressure of 10g/sq mm of jaw surface. All skinfold measurements were taken by the first author. All measurements were taken on the right side of the body, with the subject standing. The sites selected were as follow:

1. Biceps: over the midpoint of the muscle belly with the arm hanging vertically at rest.

2. Triceps: over the midpoint of the muscle belly, midway between the olecranon and the tip of the acromion, with the arm hanging vertically at rest.

3. Subscapular: Just below the tip of the inferior angle of the scapula, at an angle of about 45^{o} to the vertical.

4. Suprailiac: Just above the iliac crest in the mid-axillary line.

A reading was taken and recorded at each of the four skinfold sites and then repeated twice more in succession. Readings were taken to the nearest 0.5mm. The three readings at each site was averaged to the nearest tenth of a millimeter and the average values totaled to give the sum of four skinfolds for obtaining percent body-fat value from the Durnin and Womersley table (Reference 2).

ANALYSIS OF DATA (GENERAL):

The data analysis was conducted using the statistical package for the Social Sciences (SPSS) Update 7-9. The normal plots for the different variables in the study all had linear trends indicating normal distributions. The test for homogeneity of the variance/covariance matricies, Wilk's Lamda and Bartlett's test of sphericity were some of the statistical tools used in the analysis. Hypothesis testing was conducted at the 5% level of significance.

RESULTS AND DISCUSSION:

1. Analysis of Group 1 (Normal Weight Group).

The results of the data analysis for the normal weight group are shown in Table 2 (next page). The differences are marked by an (*). These differences in skinfold measurements are examined more closely in Table 3 below. The dependent variables are the skinfold measures and the independent variable is race (i.e., black, white).

TABLE 3. DIFFERENCES BETWEEN BLACK/WHITE NORMAL WEIGHT GROUP

VARIABLE	UNIVARIATE F	P	MANOVA CORRELATION BETWEEN DEPENDENT AND CANONICAL VARIABLE	
TRICEP	25.6	.000	.80	RELATIVE MAGNITUDE REFLECTS UNIVARIATE RESULTS
BICEP	16.7	.000	.64	
SUBSCAPULAR	1.2	.280	.17	
SUPRAILIAC	9.5	.002	.49	

TABLE 2. MEASUREMENTS FOR NORMAL WEIGHT GROUP

RACE	N	AGE	AVERAGE HEIGHT (IN)	AVERAGE WEIGHT (LB)	← SKINFOLDS (MM) → TRICEP	BICEP	SUBSCAPULAR	SUPRAILIAC	% BODY FAT
BLACK	113	$\bar{X}$ = 19.8	69.7	156.4	10.5*	5.9*	12.6	14.7*	16.6*
		S = 3	2.7	24.1	4.1	1.7	4.0	5.3	3.5
WHITE	113	$\bar{X}$ = 19.5	70.2	156.7	13.9*	6.9*	13.4	16.8*	18.4*
		S = 3.4	2.6	21.8	6.5	2.3	4.4	6.0	3.6

* P<.05

1. HOMOGENEITY OF $\tilde{\Sigma}$ (BOX's M) M = 35.7 , P = 0.000
   NOT EQUAL BUT SAMPLE SIZES ARE EQUAL.
2. SPHERICITY (BARTLETT) = 638.95 , P = 0.000
   AT LEAST ONE SIG CORRELATION.
3. WILK'S LAMDA = .848, P = 0.000
   RACIAL DIFFERENCES

2. Analysis of Group 2 (Overweight Group).

Similar data for the overweight group are shown in Tables 4 (next page) and 5 below.

TABLE 5. DIFFERENCES BETWEEN BLACK/WHITE OVERWEIGHT GROUP

VARIABLE	UNIVARIATE F	P	MANOVA CORRELATION BETWEEN DEPENDENT AND CANONICAL VARIABLE	
TRICEP	20.4	.000	.94	RELATIVE MAGNITUDE REFLECTS UNIVARIATE RESULTS
BICEP	6.2	.015	.52	
SUBSCAPULAR	3.0	.087	.36	
SUPRAILIAC	6.8	.011	.54	

3. Summary.

For both the normal and overweight groups there is a consistent difference between the black and white soldiers with respect to skinfold measurements of the triceps, biceps, and suprailiac. The white soldiers have consistently larger measurements on these variables.

TABLE 4. MEASUREMENTS FOR OVERWEIGHT GROUP

RACE	N	AGE	AVERAGE HEIGHT (IN)	AVERAGE WEIGHT (LB)	← SKINFOLDS (MM) → TRICEP	BICEP	SUBSCAPULAR	SUPRAILIAC	% BODY FAT
BLACK	38	X = 24.4	69.8	197.1	16.7*	8.0*	20.8	24.9*	22.6*
		S = 4.1	2.4	16.1	4.7	2.6	5.0	7.8	3.3
WHITE	38	X = 24.7	69.8	199.7	21.5*	9.4*	23.0	29.6*	24.2*
		S = 4.5	2.5	20.1	4.5	2.3	5.9	7.6	4.7

* P <.05

1. HOMOGENEITY OF (BOX's M) M = 7.6 , P = .714
   EQUAL
2. SPHERICITY (BARTLETT) = 121.4 , P = .000
   AT LEAST ONE SIG CORRELATION.
3. WILK'S LAMDA = .761, P = .001
   RACIAL DIFFERENCES

4. Excursion 1. Consistency of MANOVA Results for Normal Weight Group.

The consistency of the MANOVA for detecting differences between racial groups (black, white) and identifying the variables which contributed to these differences (skinfold measurements, percent body fat) are indicated in Table 6. Note that the results are consistent for the five combinations of variables using MANOVA and univariate t-tests (combination d). Combination (e) is the MANOVA that considers all the variables examined in the previous combinations. The variables which are identified as contributing to the differences between the races for the normal weight group are completely consistent with the findings of the other combinations (a through d).

TABLE 6
EXCURSION 1: MANOVA CONSISTENCY WITHIN NORMAL WEIGHT GROUP - BETWEEN RACES

VARIABLES CONSIDERED IN MANOVA	$\tilde{\Sigma}$	SPHERITY	WILK'S λ	RELATIVE MAGNITUDE OF CORRELATION BETWEEN DEPENDENT AND CANONICAL; UNIVARIATE F
(a) Skinfold only	35.7 0.000	639 0.000	.848 0.000	TRI, BICEP, ILIAC
(b) Ht, Wt only	2.9 .41	79.1 0.000	.987 .15	No Difference
(c) a and b above	39.3 .012	881.1 0.000	.787 0.000	TRI, BICEP, ILIAC
(d) % Body fat and Age	2 Separate T-Tests			% Body Fat
(e) a, b, d above	161.5 0.000	1566 0.000	.780 0.000	TRI, BICEP, ILIAC % Body Fat

NOTE: (Statistic/P value)

5. Excursion 2: Consistency of MANOVA Results for Overweight Group Table 7

The same findings were obtained for excursion 2 (see excursion 1). The same variables were identified as contributing to differences between the racial group, and the results were consistent for all combinations of factors considered in the different analyses (a through d).

TABLE 7
EXCURSION 2: MANOVA CONSISTENCY WITHIN OVERWEIGHT GROUP - BETWEEN RACES

VARIABLES CONSIDERED IN MANOVA	$\tilde{\Sigma}$	SPHERITY	WILK'S λ	RELATIVE MAGNITUDE OF CORRELATION BETWEEN DEPENDENT AND CANONICAL; UNIVARIATE F
(a) Skinfold only	7.6 .714	121.4 0.000	.761 .001	TRI, BICEP, ILIAC
(b) Ht, Wt only	3.6 .320	74.2 0.000	.989 .674	No Difference
(c) a and b above	18.2 .737	239.8 0.000	.754 .003	TRI, BICEP, ILIAC
(d) % Body fat and Age	2 Separate T-Tests			% Body Fat
(e) a, b, d above	46 .273	450.1 0.000	.734 .006	TRI, BICEP, ILIAC % Body Fat

NOTE: (Statistic/P value)

6. Excursion 3: Factorial MANOVA.

A full factorial MANOVA was conducted on the independent variables of race (black, white) and weight group (normal weight, overweight). The dependent measures were the four skinfold measurements, height, weight, age, and percent body fat. The results were entirely consistent with the findings discussed so far.

Race Effect. Wilk's Lamda was significant. The variables that contributed to the differences between races were again identified as the tricep, bicep, suprailiac, and percent body fat.

Weight Group Effect. As expected, Wilk's Lamda was significant on all the dependent variables except for height.

Race and Weight Group Interaction. The interaction effect was not significant.

7. Excursion 4: Discriminant Analysis (All Data and a 50% Random Cut)

The results of discriminant analysis, using all data, and classifying individuals as belonging to either the normal weight or overweight group are shown below in Table 8.

TABLE 8

DISCRIMINANT ANALYSIS

GROUPS = WEIGHTS (NORMAL, OVERWEIGHT)

VARIABLES = TRICEP, BICEP, SUPRAILIAC, RACE

DATA = ALL (N = 302)

ACTUAL GROUP	N	PREDICTED MEMBERSHIP NORMAL	OVERWEIGHT
NORMAL WEIGHT	226	193 85.4%	33 14.6%
OVERWEIGHT	76	17 22.4%	59 77.6%

1. CORRECT CLASSIFICATION = (193 + 59)/302 = 83.4%.

2. REFERENCE: AAKER, C max = Max $\{\hat{p}_1$, Group I, $\hat{p}_2$ Group 2$\}$

   = Max $\{.748, .252\}$ = .748

3. SINCE .834 > .748, CLASSIFICATION IS ACCEPTABLE.

According to a criteria by Aaker (reference 1), the classification is acceptable. Note that the variable used were those determined to be significant contributors as determined from other analyses. The 50% random/ stratified cut scheme is shown in Table 9.

TABLE 9

DISCRIMINANT ANALYSIS

50% RANDOM CUT (STRATIFIED BY RACE, WEIGHT GROUP)

		ORIG	CUT
NORMAL WEIGHT	=	226	113 (56 BLACK, 57 WHITE)
OVERWEIGHT	=	76	38 (19 BLACK, 19 WHITE)

VARIABLES = TRICEP, BICEP, ILIAC, RACE

The results of the discriminant analysis based on a single 50% random cut and classifying the remaining data not used to construct the discriminant functions are shown in Table 10. Note there is only a 3% difference in the percentage of correct classification when compared to the results using all data presented in Table 8.

TABLE 10

DISCRIMINANT ANALYSIS - CLASSIFICATION BASED ON A 50% RANDOM CUT

ACTUAL GROUP	N	PREDICTED MEMBERSHIP OF REMAINING DATA NORMAL	OVERWEIGHT
NORMAL	113	91 80.5%	22 19.5%
OVERWEIGHT	38	8 21.1%	30 78.9%

1. CORRECT CLASSIFICATION = $\frac{91 + 30}{151}$ = .801

2. C max = .748 < .801

Brief Summary: This single case study indicates that multivariate procedures (MANOVA, discriminant analysis) yield consistent decisions for various combinations of data. The decisions were consistent for several excursions within and between groups. Although these results are not theoretically conclusisve, and represent only one set of data, they add confidence to drawing conclusions from data analysis that are driven to answer a series of "what if" questions.

REFERENCES

1. Aaker, David (1971), Multivariate Analysis in Marketing: Theory and Application, Wadsworth Publishing Company, California.

2. Durnin, J. and Womersley, J. (1974), Body fat assessed from total body density and its estimation from skinfold thickness; measurements on 481 men and women aged from 16-72 years. Br. J. Nutr. 32, 77-97.

3. Hull, C (1981), SPSS Update 7-9. McGraw-Hill, New York.

4. Harris, R. (1975), A Primer of Multivariate Statistics, Academic Press, New York.

DESERT CAMOUFLAGE PAINT EVALUATION FOR SAUDI ARABIAN NATIONAL GUARD

GEORGE ANITOLE AND RONALD L. JOHNSON

US Army Belvoir Research and Development Center
Fort Belvoir, Virginia 22060-5606

ABSTRACT

This paper describes the procedure for selecting candidate desert paint colors, the field test, and statistical analysis procedures which enabled the final color selection. Eleven tactical vehicles were painted in camouflage colors and viewed by ten ground observers at four different color backgrounds in Saudi Arabia. The method of testing involved a ranking technique using a direct comparison between two vehicles. This is more sensitive and discerning than measurements on a scale of values and overcomes the problem of inconsistency of judgements expressed by the same observer. The observers were presented with every possible combination of paired vehicles and forced to choose between them. The data was summarized for each observer and test location and analyzed statistically to determine preferred colors in the order of rank, establish confidence limits, and establish color groupings for each site.

1.0 INTRODUCTION

In April 1980, the Project Manager (PM), Saudi Arabian National Guard (SANG), Modernization, requested assistance to provide camouflage for SANG. A fact finding team visited Saudi Arabia in September 1980 to gather specific information required to develop a camouflage program including desert colored paints, nets, and uniforms. This paper presents the paint development effort and the selection of the final camouflage paint colors for SANG.

The color used on SANG vehicles at the time of the fact finding visit, resembled a battleship gray, high luster paint. This color contrasted greatly with the Saudi Arabian desert backgrounds, and the vehicles could be seen from great distances. It was considered an immediate requirement by PM, SANG to develop a desert colored paint suitable for the Saudi Arabian deserts. The following sections describe the procedure used in selecting candidate paint colors, the field test, and data analysis that enabled the final selection of the color(s).

2.0 COLOR DETERMINATION

During the fact finding trip, soil samples and 35mm color slides were taken at each of 32 locales visited. The color determination process was a subjective process based on an evaluation of the soil samples and slides, tempered by notes and mental recall of the terrain and background by the fact finding team. Spectrophotometric analysis of the soil samples was used to determine the proper visual and near-infrared reflectance values. Hundreds of color chips in selected color areas were procured from commercial and government sources [1/]. These chips were matched against the soil samples and the backgrounds as viewed in the slides. The evaluation of soil sample colors took into account that soil samples represented the color slightly below the surface. This color is in most cases, different from the color seen by an on-site observer, because the fine sand particles have been blown away by the desert winds, leaving the heavier particles behind. The surface layer, therefore, appears slightly different in color than a soil sample which is drawn from below the surface. Based upon the color determination process described, nine camouflage colors were selected. A sample quantity of paint was procured under paint specification TT-E-529 with a 15-18% reflectance at 60°. This paint would provide sufficient surface hardness, while still being acceptable (from

the camouflage point of view) in terms of gloss. The test vehicles consisted of eleven V-150 armored vehicles. Nine of the vehicles were painted monotone colors and two were painted in a two-color pattern specially designed for the V-150.

3.0 SITE SELECTION

At the request of SANG, all field testing was restricted to areas around Riyadh and Hofuf. A survey of the local terrain around Riyadh and Hofuf showed that there were four general terrain colors: tan, gray, buff, and tan-red. Two elevated viewing stations for the observers were selected for each of the four sites of interest. All target vehicles (V-150's) were viewed against the background of the terrain and not against the sky. Table 1 shows the four sites with the viewing ranges and site color.

TABLE 1

Site Description

SITE	NEAR RANGE	FAR RANGE
Tan	720 Meters	1,020 Meters
Gray	587 Meters	853 Meters
Buff	778 Meters	1,050 Meters
Tan-Red	738 Meters	911 Meters

The near range was selected to allow observers close study of the colors. The far range was selected to represent a quasi-realistic combat distance.

4.0 TEST METHOD

The method of testing involved a ranking technique using direct comparisons between two vehicles. This technique is more sensitive and discerning than actual measurement on a scale of values and helps overcome the problem of multidimensional judgements, i.e., inconsistency of judgements expressed by the same observer. In this paired method of ranking, an observer is presented with every possible combination of two vehicles from the set to be evaluated. The observer is forced to choose between the two vehicles, and must decide either way, even if he thinks the other is just as good. The data was recorded by each observer and was summarized for each test location. It was analyzed statistically to determine preferred colors in the order of rank, established confidence limits, and color groupings for each site. All observations were a minute in duration, and were performed between the hours of 1000 and 1400 for proper sun angle and minimum shadows. Each day

was clear and hot with temperatures between 120 and 130 degrees Fahrenheit. Table 2 summarizes the number of observers for each site and range. All observers were screened for 20/20 visual acuity and normal color vision before selection.

TABLE 2

Number of Observers for Each Site and Range

SITE	RANGE	
	NEAR	FAR
Tan	10	11
Gray	11	11
Buff	10	10
Tan-Red	10	10

15.0 RESULTS

The mean preference and 95% confidence interval was calculated for each V-150 color. These results are presented as descriptive data for each viewing range for each test site.

TABLE 3

Descriptive Data, Tan Terrain, Range 1,020 Meters

Vehicle Color Number	Mean Preference	95% Confidence Interval[2/]	
		Lower Limit	Upper Limit
1	2.90	1.86	3.94
2	3.70	2.66	4.73
3	5.00	3.96	6.04
4	2.10	1.06	3.14
5	3.10	2.06	4.14
6	8.30	7.26	9.34
7	2.30	1.26	3.34
7/6	7.70	6.66	8.74
7/10	4.10	3.06	5.14
10	7.40	6.36	8.44
SAUDI SAND	8.40	7.36	9.44

The higher the mean preference number, the greater the number of times the color was preferred by the observers. That is, Color Saudi Sand was selected as being the most preferred color with a mean preference of 8.40. The associated confidence interval states that there is 95% confidence that the true mean preference rests between 7.36 and 9.44. The remaining Tables 4 through 10 can be interpreted in the same manner as Table 3.

TABLE 4

Descriptive Data, Tan Terrain, Range 720 Meters

Vehicle Color Number	Mean Preference	95% Confidence Interval Lower Limit	Upper Limit
1	1.45	0.47	2.44
2	4.55	3.56	5.53
3	4.64	3.65	5.63
4	4.00	3.01	4.99
5	3.64	2.65	4.63
6	8.91	7.92	9.90
7	1.91	0.92	2.90
7/6	7.36	6.37	8.35
7/10	3.75	2.74	4.72
10	6.55	5.56	7.53
SAUDI SAND	8.27	7.28	9.26

TABLE 5

Descriptive Data, Gray Terrain, Range 853 Meters

Vehicle Color Number	Mean Preference	95% Confidence Interval Lower Limit	Upper Limit
1	7.91	7.07	8.75
2	6.64	5.79	7.47
3	2.36	1.52	3.21
4	7.27	6.43	8.12
5	3.55	2.70	4.39
6	1.36	0.52	2.21
7	7.82	6.97	8.66
7/6	5.36	4.52	6.21
7/10	6.91	6.07	7.75
10	4.91	4.07	5.75
SAUDI SAND	0.91	0.07	1.75

TABLE 6

Descriptive Data, Gray Terrain, Range 587 Meters

Vehicle Color Number	Mean Preference	95% Confidence Interval	
		Lower Limit	Upper Limit
1	9.18	8.34	10.03
2	8.00	7.16	8.84
3	2.45	1.61	3.29
4	7.45	6.61	8.29
5	5.18	4.33	6.03
6	1.91	1.07	2.75
7	5.91	5.07	6.75
7/6	3.73	2.88	4.57
7/10	5.91	5.07	6.75
10	4.55	3.70	5.39
SAUDI SAND	0.73	0.00	1.57

TABLE 7

Descriptive Data, Buff Terrain, Range 1,050 Meters

Vehicle Color Number	Mean Preference	95% Confidence Interval	
		Lower Limit	Upper Limit
1	2.60	1.67	3.53
2	1.60	0.67	2.53
3	4.60	3.67	5.53
4	3.40	2.47	4.33
5	1.60	0.67	2.53
6	7.70	6.77	8.63
7	4.00	3.07	4.93
7/6	8.80	7.87	9.73
7/10	6.50	5.57	4.43
10	7.60	6.67	8.53
SAUDI SAND	6.60	5.67	7.53

TABLE 8

Descriptive Data, Buff Terrain, Range 778 Meters

Vehicle Color Number	Mean Preference	95% Confidence Interval	
		Lower Limit	Upper Limit
1	1.80	0.87	2.73
2	1.10	0.17	2.03
3	4.00	3.07	4.93
4	3.40	2.47	4.33
5	3.10	2.17	4.03
6	8.50	7.57	9.43
7	3.90	2.97	4.83
7/6	7.30	6.37	8.23
7/10	6.70	5.77	7.63
10	7.20	6.27	8.13
SAUDI SAND	8.00	7.07	8.93

TABLE 9

Descriptive Data, Tan-Red Terrain, Range 911 Meters

Vehicle Color Number	Mean Preference	95% Confidence Interval	
		Lower Limit	Upper Limit
1	0.70	0.00	1.48
2	2.90	2.12	3.68
3	5.30	4.52	6.08
4	4.80	4.02	5.58
5	2.50	1.72	3.28
6	8.60	7.82	9.38
7	1.50	0.72	2.28
7/6	8.20	7.42	8.98
7/10	4.60	3.82	5.38
10	7.00	6.22	7.38
SAUDI SAND	8.90	8.12	9.68

TABLE 10

Descriptive Data, Tan-Red Terrain, Range 738 Meters

Vehicle Color Number	Mean Preference	95% Confidence Interval Lower Limit	Upper Limit
1	1.90	1.12	2.68
2	1.40	0.62	2.18
3	5.90	5.12	6.68
4	4.00	3.22	4.78
5	4.60	3.82	5.38
6	8.40	7.62	9.18
7	1.60	0.82	2.38
7/6	7.30	6.52	8.08
7/10	3.40	2.62	4.18
10	7.40	6.62	8.18
SAUDI SAND	9.10	8.32	9.88

An analysis of variance was performed upon the data to determine the effect upon the two ranges as to color preference. These results are shown in Tables 11-14.

TABLE 11

Analysis of Variance to Determine the Effect of Range and Vehicle Color Upon Color Preference - Tan Terrain

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-Ratio
Range	1	0.000**	0.000**	1.00
Vehicle Color Number	10	1,240.914	124.091	44.81*
Interaction	10	4.178	0.418	0.15
Error	209	578.800	2.769	----

* Significant at $\alpha = 0.01$ ** Less than 0.001

Conclusion: The data indicated highly significant differences in vehicle color preferences, and that these preferences are not affected by range distances. The significant interaction is only the result of the variable vehicle color.

TABLE 12

Analysis of Variance to Determine the Effect of Range and Vehicle Color Upon Color Preference - Gray Terrain

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-Ratio
Range	1	0.000**	0.000**	1.00
Vehicle Color Number	10	1,434.181	143.418	71.24*
Interaction	10	76.909	7.691	3.82*
Error	220	442.909	2.013	

* Significant at $\alpha = 0.01$ ** Less than 0.001

Conclusion: The data indicated highly significant differences in vehicle color preferences, and that these preferences are not affected by range distances. The significant interaction is only the result of the variable vehicle color.

TABLE 13

Analysis of Variance to Determine the Effect of Range and Vehicle Color Upon Color Preference - Buff Terrain

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-Ratio
Range	1	0.000***	0.000***	1.00
Vehicle Color Number	10	12,297.200	129.720	58.64**
Interaction	10	42.800	4.280	1.92*
Error	198	438.00	2.212	

* Significant at $\alpha = 0.05$ *** Less than 0.001

** Significant at $\alpha = 0.01$

Conclusion: The data indicates highly significant differences in vehicle color preferences, and that these preferences are not affected by range distances. The significant interaction is only the result of the variable vehicle color.

TABLE 14

Analysis of Variance to Determine the Effect of Range and Vehicle Color Upon Color Preference - Tan-Red Terrain

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-Ratio
Range	1	0.000**	0.000**	1.00
Vehicle Color Number	10	1,563.800	156.380	99.82*
Interaction	10	58.000	5.800	3.70*
Error	198	310.200	1.567	

* Significant at $\alpha = 0.01$ ** Less than 0.001

Conclusion: The data indicates highly significant differences in vehicle color preferences, and that these preferences are not affected by range distances. The significant interaction is only the result of the variable vehicle color.

The above Tables 11-14 have indicated that the significant differences found in this study are due solely to the variable color. For this reason, the data was collapsed for the range variable. This new data will then have the effect of doubling the sample size within each cell to enable more positive conclusions. Tables 15-18 present this data for each of the four test sites.

TABLE 15

Descriptive Data, Collapsed Across Ranges, Tan Terrain

Vehicle Color Number	Mean Preference	95% Confidence Interval Lower Limit	Upper Limit
1	2.14	1.42	2.86
2	4.14	3.43	4.86
3	4.81	4.09	5.53
4	3.10	2.38	3.81
5	3.38	2.66	4.10
6	8.62	7.90	9.34
7	2.10	1.38	2.81
7/6	7.52	6.81	8.24
7/10	3.90	3.19	4.62
10	6.95	6.24	7.67
SAUDI SAND	8.33	7.62	9.05

The higher the mean preference number the greater the number of times the color was preferred by the observers. That is, vehicle 6 was selected by the observers as being the most preferred vehicle with a mean preference of 8.62. The associated confidence interval states that there is 95% confidence that the true mean preference rests between 7.90 and 9.34. The remaining Tables 16-18 can be interpreted in the same manner as Table 15.

TABLE 16

Descriptive Data, Collapsed Across Ranges, Gray Terrain

Vehicle Color Number	Mean Preference	95% Confidence Interval Lower Limit	Upper Limit
1	8.55	7.95	9.14
2	7.32	6.72	7.91
3	2.41	1.81	3.01
4	7.36	6.77	7.96
5	4.36	3.77	4.96
6	1.64	1.04	2.23
7	6.86	6.28	7.46
7/6	4.55	3.95	5.14
7/10	6.41	5.81	7.00
10	4.73	4.13	5.32
SAUDI SAND	0.82	0.22	1.41

TABLE 17

Descriptive Data, Collapsed Across Ranges, Buff Terrain

Vehicle Color Number	Mean Preference	95% Confidence Interval Lower Limit	Upper Limit
1	2.20	1.54	2.86
2	1.35	0.69	2.01
3	4.30	3.64	4.69
4	3.40	2.74	4.06
5	2.35	1.69	3.01
6	8.10	7.44	8.76
7	3.95	3.29	4.61
7/6	8.05	7.39	8.71
7/10	6.60	5.94	7.26
10	7.40	6.74	8.06
SAUDI SAND	7.30	6.64	7.96

TABLE 18

Descriptive Data, Collapsed Across Ranges, Tan-Red Terrain

Vehicle Color Number	Mean Preference	95% Confidence Level Lower Limit	Upper Limit
1	0.70	0.11	1.29
2	2.90	1.92	3.88
3	5.30	4.23	6.37
4	4.80	4.06	5.54
5	2.50	1.53	3.47
6	8.60	7.47	9.73
7	1.50	0.99	2.01
7/6	8.20	7.26	9.14
7/10	4.60	3.33	5.87
10	7.00	6.66	7.34
SAUDI SAND	8.90	8.19	9.61

The confidence intervals for the collapsed data are displayed graphically in order to obtain a better idea of the groupings of color preference for each of the four terrains. These graphs are shown in Figures 1-4.

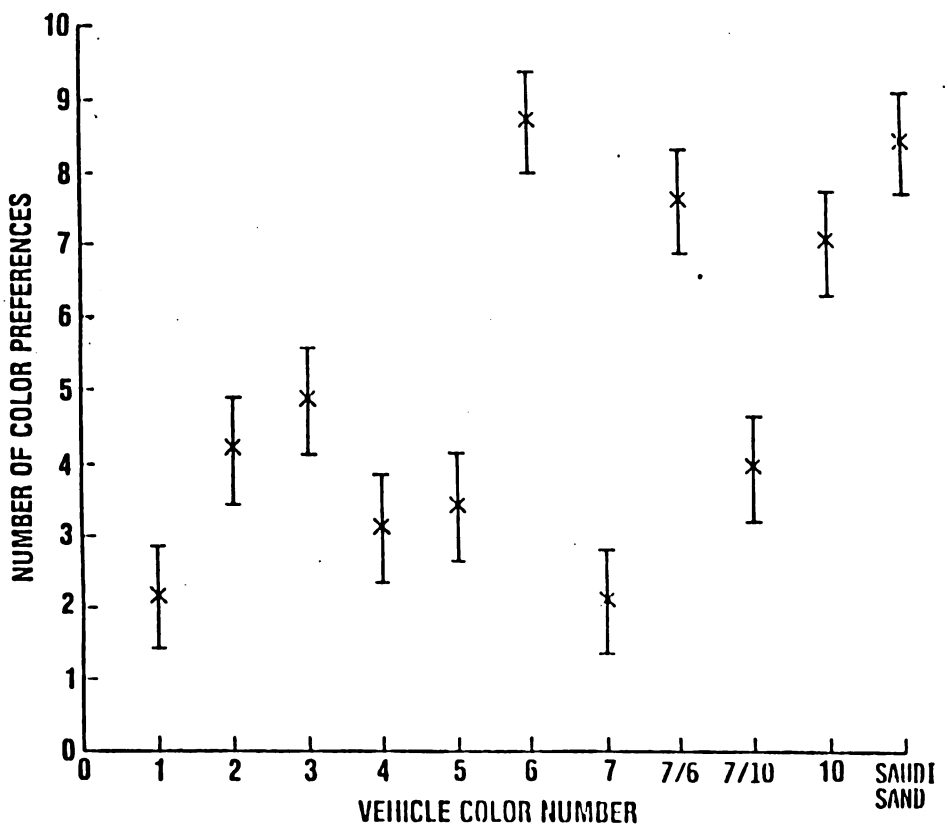


Figure 1 MEAN COLOR PREFERENCES WITH 95% CONFIDENCE INTERVALS FOR EACH VEHICLE COLLAPSED ACROSS RANGES FOR TAN TERRAIN

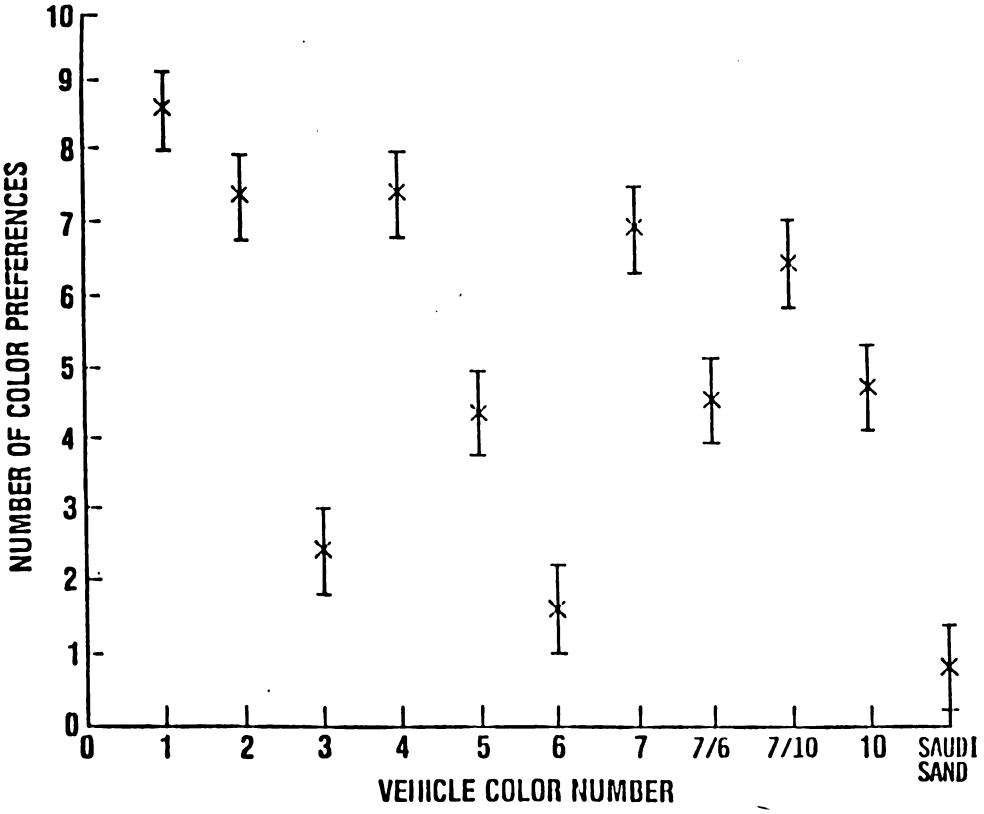


Figure 2 MEAN COLOR PREFERENCES WITH 95% CONFIDENCE INTERVALS FOR EACH VEHICLE COLLAPSED ACROSS RANGES FOR GRAY TERRAIN

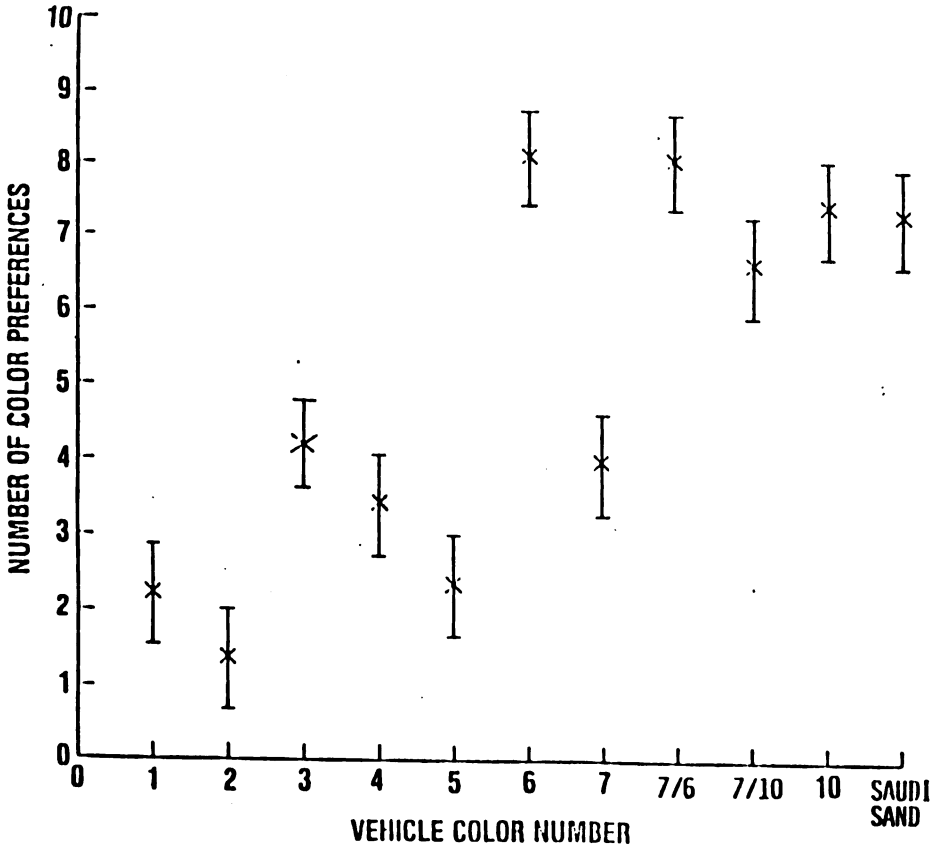


Figure 3 MEAN COLOR PREFERENCES WITH 95% CONFIDENCE INTERVALS FOR EACH VEHICLE COLLAPSED ACROSS RANGES FOR BUFF TERRAIN

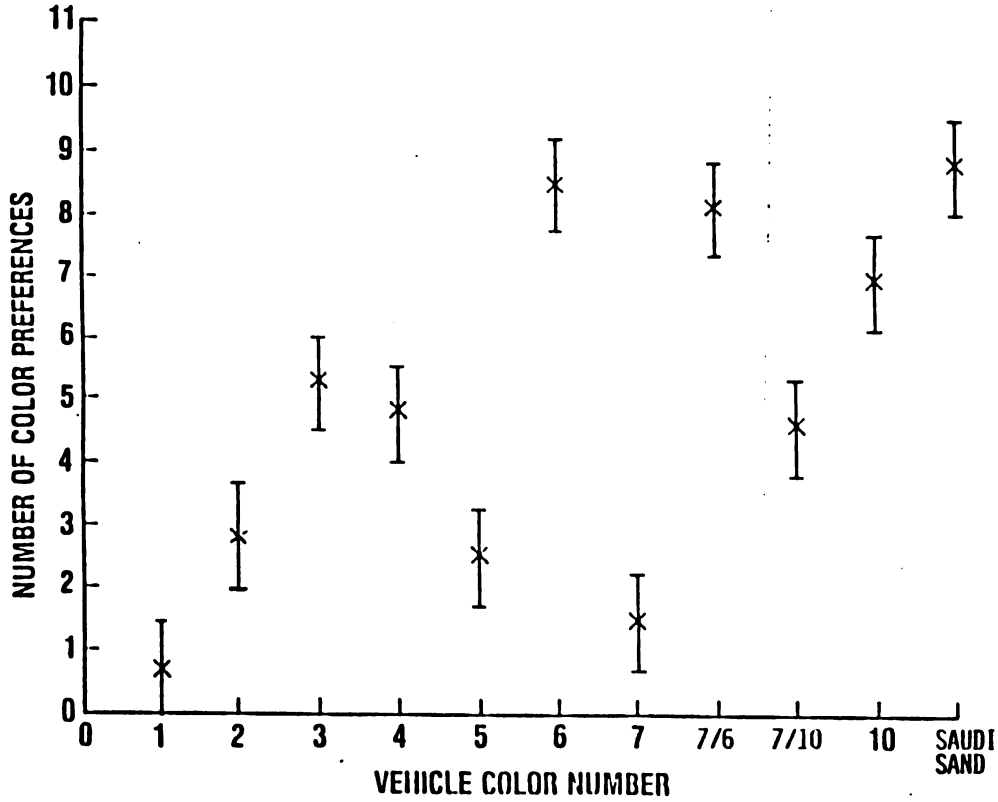


Figure 4 MEAN COLOR PREFERENCE WITH 95% CONFIDENCE INTERVALS FOR EACH VEHICLE COLLAPSED ACROSS RANGES FOR TAN-RED TERRAIN

A view of the graphs indicated the following vehicle colors were preferred for the listed sites:

- Tan Terrain - 6, 7/6, 10 and Saudi Sand
- Gray Terrain - 1, 2, 4, 7, and 7/10
- Buff Terrain - 6, 7/6, 7/10, 10 and Saudi Sand
- Tan-Red Terrain - 6, 7/6, 10 and Saudi Sand

Paired comparisons were performed for the most preferred vehicle colors for each of the sites to determine significant differences. The results are shown in Tables 19-22.

TABLE 19

Paired Comparisons of the Most Preferred Vehicle Colors on the Collapsed Data - Tan Terrain

Vehicle Color Numbers	F-Ratio	Significance Levels
6 and 7/6	4.548	$\alpha = 0.03412$
6 and 10	10.532	$\alpha = 0.00137$
6 and Saudi Sand	0.310	$\alpha = 0.57858$
7/6 and 10	1.238	$\alpha = 0.26713$
8 and Saudi Sand	2.485	$\alpha = 0.11647$
10 and Saudi Sand	7.230	$\alpha = 0.00775$

Conclusion: Vehicle color 6 differs significantly from vehicle colors 7/6 and 10, and vehicle color 10 differs significantly from vehicle color Saudi Sand.

TABLE 20

Paired Comparisons of the Most Preferred Vehicle Colors on the Collapsed Data - Gray Terrain

Vehicle Color Numbers	F-Ratio	Significance Levels
1 and 2	8.230	$\alpha = 0.00452$
1 and 4	7.631	$\alpha = 0.00622$
1 and 7	15.455	$\alpha = 0.00011$
1 and 7/10	24.937	$\alpha = 0.00000$
2 and 4	0.011	$\alpha = 1.00000$
2 and 7	1.129	$\alpha = 0.28918$
2 and 7/10	4.516	$\alpha = 0.03470$
4 and 7	1.366	$\alpha = 0.24377$
4 and 7/10	4.978	$\alpha = 0.02667$
7 and 7/10	1.129	$\alpha = 0.28918$

Conclusion: Vehicle color 1 differs significantly from vehicle colors 2, 4, 7, and 7/10; vehicle color 2 differs significantly from vehicle color 7/10; and vehicle color 4 differs significantly from vehicle color 7/10.

TABLE 21

Paired Comparisons of the Most Preferred Vehicle Colors on the Collapsed Data - Buff Terrain

Vehicle Color Numbers	F-Ratio	Significance Levels
6 and 7/6	0.011	α = 1.00000
6 and 7/10	9.734	α = 0.00207
6 and 10	2.120	α = 0.14691
6 and Saudi Sand	2.769	α = 0.09763
7/6 and 7/10	9.096	α = 0.00288
7/6 and 10	1.828	α = 0.17786
7/6 and Saudi Sand	2.433	α = 0.12029
7/10 and 10	2.769	α = 0.09763
7/10 and Saudi Sand	2.120	α = 0.14691
10 and Saudi Sand	0.043	α = 1.00000

Conclusion: Vehicle color 6 differs significantly from vehicle color 7/10, and vehicle color 7/6 differs significantly from vehicle color 7/10.

TABLE 22

Paired Comparisons of the Most Preferred Vehicle Colors on the Collapsed Data - Tan-Red Terrain

Vehicle Color Numbers	F-Ratio	Significance Levels
6 and 7/6	3.590	α = 0.05957
6 and 10	10.787	α = 0.00121
6 and Saudi Sand	1.596	α = 0.20799
7/6 and 10	1.931	α = 0.16623
7/6 and Saudi Sand	9.973	α = 0.00184
10 and Saudi Sand	20.681	α = 0.0001

Conclusion: Vehicle color 6 differs significantly from vehicle color 10; vehicle color 7/6 differs significantly from vehicle color Saudi Sand; and vehicle color 10 differs significantly from vehicle color Saudi Sand.

The analysis of the data indicated that the color preferences fell into two distinct groupings. The colors 6, 7/6, 10, and Saudi Sand were most preferred for the

tan, buff, and tan-red terrains (Figures 1, 3, and 4). The colors 1, 2, 4, 7, and 9 were the most preferred for the gray colored terrain (Figure 2). A review of Tables 19, 21, and 22 show that a ranking of the best colors for the tan, buff, and tan-red terrains would be:

- o Colors 6 and Saudi Sand - Best
- o Colors 7/6 and 10 - Next Best

A review of Table 20 for the gray color terrain shows that the ranking of the preferred colors would be as follows:

- o Color 1 - Best
- o Colors 2 and 4 - Average
- o Colors 7 and 7/10 - Worst

6.0 DISCUSSION

Observations by the test team made during this trip and during the fact finding trip in September 1980, indicated a predominance of the tan, buff, and tan-red backgrounds around the Riyadh and Hofuf vicinities. The gray test area was the exception. Although the Belvoir Research and Development Center field test personnel did not have the opportunity to determine the physical extent of the gray colored terrain, this color was not prevalent at Hofuf or anywhere else in the Riyadh area.

The overall color selections should be based on what is considered to be the predominant terrain background colors. The data analysis indicated that the best color selection for the predominant terrain background (i.e., tan, buff, tan-red) are the colors 6 and Saudi Sand. The data analysis further indicated that there is no significant statistical difference between these colors. Color 6 tends to blend in slightly better with the tan and buff backgrounds, whereas color Saudi Sand is slightly ahead in the tan-red background. A forced selection between the two colors would probably favor color 6 based upon the test results in the tan and buff sites, and the overall preponderance of tan and buff backgrounds in the Riyadh and Hofuf areas.

If for some reason the gray colored background cannot be discounted, then a color should be selected which is a compromise for all sites. Table 19 indicates that colors 5, 7/6, and 10 are in the second grouping of best colors. Colors 7/6 and 10 are also in the next

best grouping of colors for the tan, buff, and tan-red sites as indicated in the previous section. Data analysis also indicates that color 7/6 does not differ significantly from color 10. Colors 7/6 and 10 could therefore serve as overall compromise colors across all sites. Although color 7/6 has a slight edge in preference across all sites, the preferred color is color 10. This is based upon the camouflage guideline of selecting the lighter color when confronted with a choice for a desert background. The lighter color tends to counteract some of the natural on-vehicle shadows which contribute to giving the vehicle a dark appearance.

7.0 SUMMARY

A total of eleven V-150 vehicles painted in camouflage colors were viewed by a minimum of ten ground observers at tan, buff, gray, and tan-red colored backgrounds in Saudi Arabia. Each site had a near and far range. Every possible color pair combination was viewed at each site and range. In each case, the observer was forced to make a choice as to which color blended best with the background.

Analysis of the data indicated two preferred colors for the predominant background terrains found in the Riyadh and Hofuf areas, namely color 6 and color Saudi Sand. The data analysis also indicated that there is no significant statistical difference between these colors. A subjective forced choice between these colors favors the selection of color 6.

REFERENCES

1. Federal Color Standard #595

2. Natrella, Mary G., Experimental Statistics, National Bureau of Standards Handbook 91, US Department of Commerce, Washington, D.C., 1966

TETRACHORIC CORRELATION FOR MESOSCALE AREAL PERSISTENCE OF CLOUD CEILINGS

Oskar M. Essenwanger
Research Directorate
US Army Missile Laboratory
US Army Missile Command
Redstone Arsenal, Alabama 35898-5248

ABSTRACT: The correlation coefficient of attributes (tetrachoric correlation for 2X2 contingency table) in a useful tool in the study of areal persistence. This correlation permits us to evaluate persistence for subgroups (classes of attributes) while the ordinary linear correlation coefficient is a summary of the total data set. From the examination of the tetrachoric correlation it is learned that little persistence is found for low visibility and low cloud ceiling conditions in a selected mesoscale area. Clouds with a ceiling over 8000 ft, however, display an apparent wider coverage of this area and show persistence. Thus the vertical structure of persistence is confirmed in quantitative numbers by the tetrachoric correlation coefficient.

1. INTRODUCTION: The tetrachoric correlation coefficient of a 2X2 contingency table is not used very often in statistical analysis because many statisticians have pointed out its limitations, especially the problem of determining its statistical significance. In turn, the ordinary linear correlation coefficient may be of limited value if applied to non-Gaussian distributions such as cloud amounts (sky cover) and ceiling. Furthermore, areal persistence as judged by the ordinary linear correlation coefficient is not suitable to disclose vertical stratification of areal persistence. In this particular case the tetrachoric correlation is not only a useful tool to measure areal persistence in the mesoscale, but can reveal the vertical structure of areal persistence. We learn from cloud data of a mesoscale area in Central Europe that fog and low ceilings follow a joint probability which is virtually equivalent to randomness. However, ceilings above 4000 feet exhibit non-random behavior, and cloud ceilings extend over a much larger area with consistency. This interpretation is obtained in quantitative form from the use of the tetrachoric correlation coefficient.

2. THEORETICAL BACKGROUND. Let the four fields of a 2X2 contingency table be called a, b, c, d (Fig 1). Then the marginal distribution can be written:

a	b	R_1
c	d	R_2
S_1	S_2	Tot

Fig 1 Contingency Table

$$R_1 = a + b \qquad (1a)$$

$$R_2 = c + d \qquad (1b)$$

$$S_1 = a + c \qquad (1c)$$

$$S_2 = b + d \qquad (1d)$$

Let us assume that the field attributes are given as fractions of the total N, namely:

$$a + b + c + d = 1.0 \qquad (2a)$$

and

$$R_1 + R_2 = S_1 + S_2 = 1.0 \qquad (2b)$$

The tetrachoric correlation coefficient as defined by Kendall and Stuart (1958) is:

$$d = \sum_{m=0}^{\infty} \rho^i \quad \tau_i(h)\tau_j(k) \qquad (3a)$$

where $\tau_r(x)$ is the tetrachoric function:

$$\tau_r(x) = H_{r-1}(x)\,\phi(x)/(r!)^{\frac{1}{2}} \qquad (3b)$$

and $\phi(x)$ is the Gaussian integral value integrated from $-\infty$ to x.

$H_r(x)$ represents the Tchebycheff - Hermite polynomials, and $h = S_2, k = R_2$.

It is evident that ρ cannot readily be calculated and is usually determined by iterations and/or the aid of tables. Approximations have been suggested such as:

$$r_T = [(\pi/2)(\sqrt{ad} - \sqrt{bc})/(\sqrt{ad} + \sqrt{bc})] \qquad (4)$$

The form which is used here is based on the $\chi 2$ calculation and is sometimes called the phi-coefficient. The (linear) correlation coefficient of attributes in a contingency table is:

$$r_a = [(\chi^2 / N)/(m-1)]^{\frac{1}{2}} \qquad (5)$$

and for a 2X2 contingency table with m = 2, we find:

$$\chi^2/N = (ad - bc)^2 / (R_1 R_2 S_1 S_2) \qquad (6a)$$

although Yates has suggested a corrected form (adjusted to eqn 2a, b):

$$\chi^2/N = (|ad - bc| - 1/2N)^2 \qquad (6b)$$

which must be used for small N. In our case N>200 and the correction can be neglected. Hence:

$$r_a^2 = (ad - bc)^2/ (R_1 R_2 S_1 S_2) \qquad (7)$$

which can be transformed into:

$$r_t^2 = r_\phi^2 = (a - R_1S_1)^2 / (R_1 R_2 S_1 S_2) \qquad (8)$$

This has also been called the tetrachoric correlation by various authors.

The reader will notice that $r_t = 0$ for $a = S_1 R_1$ and $r_t = 1.0$ for $S_1 = R_1 = a$. The first case $a = S_1 R_1$ represents the joint probability of two events occurring at random. The second case implies a perfect match of the two events, or complete association. The size of r_t for a 2X2 contingency table of the joint probability of cloud cover expresses the degree of persistence. Caution must be exercised, however.

Similar to ρ or r_T the establishment of the statistical significance for r_t is not simple. This is supported by the fact that for $S_2 \neq R_1$ the correlation r_t has a maximum value $r_{max} < 1.0$ (see Table 1). Thus the interpretation of persistence from the use of r_t is not trivial. In order to evaluate the effect of the asymmetry $R_1 \neq S_1$ we may define a ratio:

$$\eta_t = r_t/r_{max} \qquad (9a)$$

Assume $R_1 \leqq S_1$, the field $a \leqq R_1$, then $a_{max} = R_1$; for $R_1 \geqq S_1$ we have $a_{max} = S_1$.

$$r^2_{max} = R_1 S_2/R_2 S_1 \qquad (9b)$$

or
$$\eta_t^2 = (a - R_1 S_1)^2 /[R_1^2 (1 - S_1)^2] \qquad (9c)$$

For an observed a_{obs} we finalize:

$$\eta_t = (a_{obs} - R_1 S_1)/R_1 S_2; \quad a_{obs} \leqq R_1 \qquad (9d)$$

Table 1 displays the maximum r_t for R_1 and S_1. Since the values of r_t are symmetric around the diagonal only one part is provided. Evidently the lower the ratio R_1/S_2 the lower is r_{max}. However, r_{max} is not linearly related to R_1/S_1 (see eqn. 9b).

The ratio η_t provides an evaluation whether r_t has rendered its maximum value, and can be considered an adjusted measure of persistence.

The utilization of r_t has some advantages compared with the application of the ordinary linear correlation coefficient. The calculation of r_t is based on a contingency table. Thus the correlation (and implicitly the persistence) can be determined for specific groups of cloud conditions by establishing the contingency tables for these groups and comparing r_t for them. In our case these groups are arranged by classes of the cumulative frequency of occurrence of cloud ceiling by altitude, although the contingency tables needed for comparison can be established for any grouping of attributes. The grouping of cloud cover by ceiling height permits us to study the variation of the persistence as a function of ceiling altitude, information which would not readily be available from the ordinary linear correlation coefficient which is calculated for the entire data set.

For the referenced data sample the correlation coefficient can be considered as significantly different from zero at the 95% level of significance for $r \geq 0.055$ (one-sided) or $r \geq 0.067$ (two-sided). The respective significance for the three sigma level would be $r \geq 0.10$ and 0.11, respectively. These values will be referred to in the subsequent text.

3. GROUPS OF CLOUDS. As previously outlined the question is whether persistence is the same irrespective of ceiling height and conditions designated as "adverse weather". Ceiling in this connection is defined as cloud cover of ≥ 0.5 of the sky. The following cases were of interest.

- visibility $\leq$ 1 km (fog)
- ceiling $\leq$ 500 ft
- ceiling $\leq$ 500 ft and/or visibility $\leq$ 2 mi
- ceiling $\leq$ 800 ft
- ceiling $\leq$ 800 ft and/or visibility $\leq$ 3 mi
- ceiling $\leq$ 2000 ft and/or visibility $\leq$ 4 mi
- ceiling $\leq$ 4500 ft and/or visibility $\leq$ 4 mi
- ceiling $\leq$ 8000 ft and/or visibility $\leq$ 5 mi

The frequency of occurrence and the joint (empirical) probabilities have been compiled in reports by Essenwanger and Levitt (1984) and Levitt and Essenwanger (1984). Table 2 is an example of these joint probabilities. The joint (empirical) probability of "adverse weather" is reasonable as displayed in Table 2 for the selected groups. The question remains whether this is the lowest possible joint probability. It would be if random association can be assumed but it would not be if the area persistence of adverse weather is high between the two stations. This will be investigated now.

4. TWO STATION AREAL PERSISTENCE. As a first example the condition for fog (visibility $\leq$ 1 km) is examined. Table 3a exhibits the occurrence of fog at four stations (central Europe) in fall. Table 3b lists the joint probabilities for the six possible combinations of pairs arranged by distance. This table exhibits a decrease of the joint occurrence with distance and a slight increase of the probabilities in the last two columns.

How much is this increase caused by a higher persistence? This question can be answered from Table 3c which discloses r_t for the six station pairs.

The first apparent difference between Tables 3b and c is the fact that the joint probability displays a different diurnal trend. While the highest joint probability is found at 7^h in the morning the highest persistence is found between 19^h to 01^h except for the last two columns where highest persistence and peak of joint probability coincide. These variations as depicted are caused by the differences of terrain conditions within the mesoscale area. These differences are again evident in the comparison of the station pairs as function of the distance.

It can be noticed that the correlation (and persistence) shows a tendency to decrease with distance although the last two columns disclose again a slight increase in r_t in comparison with the middle range of 70 miles. This small increase may be a terrain effect. Apparently fog conditions in the Rhine Valley show a small tendency to occur also in the Saar Valley while Hahn which is located in a mountain area does not correlate as well with Frankfurt and Heidelberg. The small value of r_t indicates very little persistence. It may surprise that even Heidelberg and Frankfurt, both in the Rhine Valley, display only a moderate persistence, i.e. a maximum association of 38% (from $r_t = 0.62$), and this only at 22^h (local time).

Table 4 is selected to show the variation of persistence with altitude and distance. Only two levels of ceiling are shown, under 800 and 8000 ft. A comparison of r_t for the top and the lower part of Table 4 reveals that for all but one column the correlations are generally higher if the ceiling height is higher. This result is not changed by the ratio n_τ (eqn. 9d).

Again, a marked decrease is evident with distance implying that persistence decreases with distance. Although this result is expected it is a quantitative formulation exhibited in Table 4. The high correlation for ceiling $\leqq 800$ ft for the station pair Heidelberg - Frankfurt contrasting the lower value for ceiling under 8000 ft (column three) may be somewhat unexpected. It has some explanation in the terrain conditions within the mesoscale area, however. It was pointed out in the discussion of Table 3c that Heidelberg and Frankfurt, both in the Rhine Valley, show a high correlation (and persistence) for fog. This high persistence is extended to include the conditions ceiling $\leqq 800$ ft. However, if ceiling $\leqq 8000$ ft is considered both stations show a more independent pattern. This conclusion could not be deduced from inspection of Table 2 where the joint probabilities are of the same magnitude and little distance dependency is shown.

5. THREE STATION AREAL PERSISTENCE. Examples of three station areal correlations are presented in Tables 5 and 6. In the case of three stations we have the option of utilizing a 2X2 or 3X3 contingency table. For the 2X2 contingency table we start with the joint probability for one station pair as R_1 or S_1 (see eqn.'s 1) and "d" denotes the joint probability of the three stations. This scheme is similar to the algorithm which was discussed before. An example is given in Table 5.

We learn from Table 5 that the correlation increases with increasing ceiling altitude. This implies that cloud layers in higher altitudes show a larger areal persistence or area coverage. This fact can be confirmed from inspection of daily weather maps and study of their frontal systems. This persistence depends, however, on the size of the mesoscale area. In Table 5 the diurnal variation was eliminated by averaging although from a point of theoretical statistic Fisher's z-function should have been used. The calculation did not alter the exhibited relative magnitude within Table 5 and is therefore omitted.

It is noted that the condition of ceiling $\leq$ 2000 ft without inclusion of visibility displays a slightly smaller correlation than with inclusion of the visibility except for the last column. This increase of the persistence is reasonable because good visibility appears to prevail over a larger area than poor visibility. The latter is a predominant effect of local terrain which is quantitatively confirmed by the correlation.

Table 6 exhibits an additional feature. While part A of the table provides the tetrachoric correlation part B shows the correlation based on eqn. 5 and the 3X3 contingency table. In the latter case the total area coverage is judged. Part A permits us to study the persistence in segments of the mesoscale area. The three station correlation in part B is only one single measure for the total area cover. Both measures reflect the increase in areal persistence from lower to higher ceiling threshold. Furthermore, Table 6A also discloses some seasonal variations. For low ceiling (under 800 ft) little areal persistence exists in summer while the winter season shows relatively strong persistence. Again, this is expected. During the summer months low ceiling is limited to small areas associated with terrain conditions. In Central Europe fog and low ceiling prevail often over even larger areas in wintertime than the mesoscale area under study here. The tendency of lower persistence in summer for low ceiling is also confirmed in Table 6B.

5. FOUR STATION AREAL PERSISTENCE. Again, several choices are available how to produce a 2X2 contingency table for four stations and examine the areal persistence. The first example is given in Table 7. The joint probability of adverse weather is compared with R_1 and S_1 being the joint probability of two stations. This permits us a study of areal persistence in the zonal and meridional direction. The correlation and ratio of correlation to the maximum possible correlation is exhibited in Table 7 for four seasons.

We learn that again low visibility and ceiling show little areal persistence in the zonal direction while in the meridional direction areal persistence is low in spring. Areal persistence improves with higher altitude of the ceiling. Apparently the persistence is about equal in the zonal and meridional direction for higher ceiling. The numerical values increase for the ratio η_t but the general trend of the areal persistence remains the same. The higher areal persistence in the meridional than the zonal direction for low visibility and ceiling altitude may be a consequence of frontal systems predominantly moving in the meridional direction in this particular mesoscale area.

Table 8 provides a summary of the diurnal trend for selected groups of adverse weather in fall. This time the correlation is calculated from R_1 being the joint probability of three stations and S_1 the probability of the fourth station. Four conditions (one for each station as S_1) can be construed. The correlations for these four conditions have been arranged for specified hours. It is evident that a definite diurnal trend exists. The minimum persistence is underlined in Table 8. The diurnal difference of the correlation appears to be reduced with increasing ceiling altitude but the minimum persistence does not display a uniform pattern.

7. CONCLUSION. It has been illustrated that the correlation coefficient of attributes for a contingency table (for a 2X2 contingency table also called tetrachoric correlation) can be useful in the study of areal persistence of cloud and visibility conditions. Since the fields of the contingency table can reflect the probability of occurrence for specified classes it is possible to study the vertical structure of the areal persistence while the ordinary linear correlation coefficient can only express the persistence for the total set of data.

It has been shown that persistence for a mesoscale area depends on ceiling altitude. It was deduced that low visibility ($\leq$ 1 km) and low ceiling ($\leq$ 1000 ft) is predominately terrain dependent with only a small areal persistence. However, clouds with ceiling $\leq$ 8000 ft tend to cover more than the mesoscale area under investigation. Thus they depict areal persistence. As Table 7 exhibits there is little seasonal fluctuation for ceiling $\leq$ 8000 ft.

The data on cloud cover over a mesoscale area in Central Europe produce asymmetric 2X2 contingency tables. In this case the maximum possible correlation is smaller than unity. A correction can be made to evaluate this bias by calculating a ratio η_t of the correlation coefficient to the maximum correlation. As it proved the result concerning the areal persistence does not change significantly by examination of this ratio.

8. REFERENCES. Essenwanger, O. M. and L. J. Levitt, 1983. Joint Probability Cloud and Visibility Thresholds around Frankfurt, Germany. Tech Rep. RR-84-1, US Army Missile Command, pp. 81.

Kendall, M. G. and A. Stuart, 1958. The Advanced Theory of Statistics Vol 2, Interence and Relationship. Hafner, New York, pp. 676.

Levitt, L. J. and O. M. Essenwanger, 1984. Joint Probability of Selected Cloud and Visibility Thresholds around Koblenz, Germany. Tech Rep. RR-84-3, US Army Missile Command, pp. 75.

9. ACKNOWLEDGEMENT. The author wishes to thank Dr. Dorathy A. Stewart for her critical review of the manuscript and Ms. Gloria McCrary for her diligent efforts in typing the manuscript.

TABLE 1 MAXIMUM TETRACHORIC CORRELATION

S_1	R_1 0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	0.999
0.01	1.0													
0.05	0.438	1.0												
0.10	0.302	0.688	1.0											
0.20	0.201	0.459	0.667	1.0										
0.30	0.153	0.350	0.510	0.764	1.0									
0.40	0.123	0.281	0.408	0.612	0.801	1.0								
0.50	0.100	0.229	0.333	0.500	0.655	0.816	1.0							
0.60	0.080	0.187	0.272	0.408	0.535	0.667	0.816	1.0						
0.70	0.066	0.150	0.218	0.327	0.429	0.535	0.655	0.802	1.0					
0.80	0.050	0.114	0.167	0.250	0.329	0.408	0.500	0.612	0.764	1.0				
0.90	0.034	0.076	0.111	0.167	0.218	0.272	0.333	0.408	0.509	0.667	1.0			
0.95	0.023	0.053	0.077	0.115	0.150	0.187	0.229	0.281	0.350	0.459	0.688	1.0		
0.99	0.010	0.023	0.034	0.050	0.066	0.082	0.101	0.123	0.154	0.201	0.302	0.438	1.0	
0.999	0.003	0.007	0.011	0.016	0.021	0.026	0.032	0.039	0.048	0.063	0.095	0.138	0.315	1.0

TABLE 2 J O I N T P R O B A B I L I T I E S (%)

B E T W E E N T W O S T A T I O N S

F A L L

MILES	40	50	70	70	80	90
HOUR	SA-HA	HEI-FR	HEI-HA	FR-HA	HEI-SA	SA-FR
1	56 36%	48 33%	49 37%	52 34%	50 32%	57 30 %
4	63 41	55 37	54 40	59 39	57 37	65 34
7	69 45	66 42	62 45	66 43	69 42	74 40
10	67 46	62 40	57 41	62 42	65 40	71 40
13	53 43	49 34	50 41	47 38	54 37	51 37
16	44 35	42 29	43 35	39 31	46 34	41 29
19	49 34	42 31	42 32	45 33	46 32	46 30
22	51 34	43 33	44 34	47 34	47 33	50 31

FIRST COLUMN: C $\leq$ 8000 FT AND/OR V $\leq$ 5 MI

SECOND COLUMN: C $\leq$ 8000 FT

SA = Saarbruecken, HA = Hahn, HEI = Heidelberg, FR = Frankfurt

C = Ceiling, V = Visibility

TABLE 3A VISIBILITY < 1 KM (PROBABILITIES)

FALL

HOUR	HEID	FRANK	SAAR	HAHN
1	5.3%	4.4%	8.2%	10.7%
4	6.2	7.9	12.0	13.8
7	9.2	10.0	16.9	19.3
10	5.9	6.9	12.4	13.7
13	2.6	3.1	4.3	6.6
16	2.6	2.4	2.2	5.6
19	2.9	2.3	3.8	6.5
22	4.2	3.2	4.1	8.1

HEIDELBERG
FRANKFURT
SAARBRUECKEN
HAHN

TABLE 3B JOINT PROBABILITIES OF VISIBILITY ≤1KM FOR STATIONS OF TABLE 3A

FALL

Miles / Hour	40(N-S) SA-HA	50(N-S) HEI-FR	70(E-W) HEI-HA	70(E-W) FR-HA	80(E-W) HEI-SA	90(SW-NE) SA-FR
1	3.7	2.9	1.9	1.4	2.3	2.2 %
4	5.5	3.3	2.2	2.1	2.8	3.1
7	7.7	4.1	3.8	3.5	5.6	5.6
10	4.8	2.8	2.1	2.0	2.9	3.3
13	1.3	1.0	0.6	0.4	0.6	0.8
16	1.2	1.1	0.3	0.6	0.3	0.4
19	2.4	1.5	0.7	0.6	0.4	0.7
22	2.4	2.3	0.9	0.9	0.9	1.0

SA = Saarbruccken

HA = Hahn

HEI = Heidelberg

FR = Frankfurt

TABLE 3C

TWO STATION CORRELATIONS

VISIBILITY ≤ 1 KM

FALL

MILES	40(N-S)	50(N-S)	70(E-W)	70(E-W)	80(E-W)	90(SW-NE)
HOUR	SA-HA	HEI-FR	HEI-HA	FR-HA	HEI-SA	SA-FR
1	.34	.57	.19	.15	.31	.33
4	.34	.43	.16	.11	.26	.24
7	.29	.36	.18	.14	.37	.35
10	.28	.39	.16	.12	.27	.29
13	.21	.33	.10	.06	.13	.18
16	.32	.42	.05	.12	.12	.17
19	.46*	.59	.12	.11	.10	.20
22	.39	.62	.10	.13	.18	.25
MEAN	.33	.46	.13	.12	.22	.25

* (MAXIMUM UNDERLINED)

SA = Saarbruecken

HA = Hahn

HEI = Heidelberg

FR = Frankfurt

TABLE 4 TWO-STATION CORRELATIONS (FALL)

C ≤ 800 FT

Tetrachoric Correlation

MILES:	40	50	70	70	80	90
HOUR	SA-HA	HEI-FR	HEI-HA	FR-HA	HEI-SA	SA-FR
1	.44	.58	.17	.22	.21	.30
7	.37	.46	.09	.21	.15	.19
13	.39	.37	.11	.25	.19	.28
19	.47	.46	.20	.35	.19	.24

RATIO

1	.53	.58	.32	.41	.34	.48
7	.46	.53	.18	.42	.25	.27
13	.44	.48	.24	.53	.37	.41
19	.56	.55	.39	.70	.32	.33

C ≤ 8000 FT

Tetrachoric Correlation

MILES:	40	50	70	70	80	90
HOUR	SA-HA	HEI-FR	HEI-HA	FR-HA	HEI-SA	SA-FR
1	.66	.23	.31	.34	.29	.31
7	.60	.20	.31	.29	.24	.23
13	.66	.23	.09	.15	.15	.19
19	.59	.22	.37	.32	.19	.25

RATIO

1	.73	.25	.36	.42	.30	.31
7	.64	.22	.33	.34	.26	.24
13	.72	.24	.10	.16	.15	.20
19	.62	.22	.42	.37	.20	.26

C = CLOUD CEILING

TABLE 5 THREE STATION MEAN CORRELATION (FALL)

FRANKFURT-HEIDELBERG (50 MILES) HAHN-SAARBRUECKEN (40 MILES)

	HAHN (60 MI)	SAAR (80 MI)	HEID (70 MI)	FRANK (80 MI)
V ≤ 1 KM	.07	.19	.15	.16
C ≤ 500 FT	.08	.17	.13	.25
C ≤ 800 FT	.11	.18	.17	.23
C ≤ 2000 FT	.20	.25	.26	.34
C ≤ 4500 FT	.36	.39	.44	.44
C ≤ 8000 FT	.44	.46	.46	.46
C ≤ 2000/V≤4 MI	.27	.36	.38	.33
C ≤ 8000/V≤5 MI	.39	.39	.44	.37

(HOURS OF DAY AVERAGED)

TABLE 6 THREE-STATION TOTAL

A) TETRACHORIC CORRELATION

	SPRING			SUMMER			FALL			WINTER		
	1	2	3	1	2	3	1	2	3	1	2	3
C ≤ 800 FT	.24	.23	.20	.08	.11	.05	.23	.29	.23	25	.38	.44
C ≤ 2000 FT	.33	.31	.28	.30	.26	.25	.34	.40	.34	.28	.39	.46
C ≤ 8000 FT	.45	.43	.43	.37	.38	.35	.47	.50	.48	.23	.33	.43

1 = Hahn-Saarbr. Versus Frankfurt (W-E)

2 = Hahn-Frankfurt Versus Saarbr. (N-S)

3 = Saarbr.-Frankfurt Versus Hahn (SE-NW)

B) TOTAL THREE-STATION CORRELATION

	SPRING	SUMMER	FALL	WINTER
C ≤ 800 FT	.39	.24	.33	.33
C ≤ 2000 FT	.40	.42	.42	.39
C ≤ 8000 FT	.47	.40	.49	.47

C = CEILING

TABLE 7 FOUR STATION PERSISTENCE

A) CORRELATION

	E-W				N-S			
	SP	SU	FA	WI	SP	SU	FA	WI
V≤1 KM	.06	(.05)	.13	.12	.09	--	.32	.25
C≤800 FT	.09	.10	.14	.15	.12	.31	.34	.27
C≤2000 FT	.25	.21	.25	.31	.32	.27	.43	.43
C≤4500 FT	.43	.36	.42	.42	.44	.36	.48	.45
C≤8000 FT	.46	.40	.50	.47	.47	.41	.50	.48
C≤500 FT/V≤2 MI	.16	.17	.24	.24	.17	.23	.41	.32
C≤2000 FT/V≤4 MI	.36	.33	.35	.34	.43	.36	.46	.43
C≤8000 FT/V≤5 MI	.45	.41	.43	.40	.47	.44	.45	.39

B) RATIO

	E-W				N-S			
	SP	SU	FA	WI	SP	SU	FA	WI
V 1 KM	.25	(.30)	.16	.27	.10	--	.39	.33
C 800 FT	.40	.30	.28	.36	.16	.51	.41	.35
C 2000 FT	.67	.61	.45	.56	.35	.28	.54	.49
C 4500 FT	.63	.62	.56	.52	.46	.39	.52	.50
C 8000 FT	.55	.47	.52	.50	.48	.44	.52	.49
C≤500 FT/V≤2 MI	.39	.47	.33	.43	.21	.30	.45	.38
C≤2000 FT/V≤4 MI	.64	.57	.48	.57	.50	.40	.49	.49
C≤8000 FT/V≤5 MI	.55	.49	.48	.50	.50	.49	.48	.42

E-W = Hahn/Saarbruecken - Frankfurt/Heidelberg

N-S = Frankfurt/Hahn - Saarbruecken/Heidelberg

TABLE 8 · FOUR STATIONS, FALL

AREA PERSISTENCE (MEAN CORRELATION)

HOUR	V ≤ 1 KM	C ≤ 500 FT	C ≤ 8000 FT	C ≤ 8000 FT/V ≤ 5 MI
1	.26	.28	.50	.41
4	.22	.23	.45	.35
7	.17	.16	.46	.32
10	.11	.10	.45	.34
13	.14	.07	.40	.38
16	.17	.17	.41	.43
19	.20	.20	.45	.44
22	.26	.23	.50	.42

(CORRELATIONS OF FOUR COMBINATIONS "THREE-STATION VERSUS ONE STATION" AVERAGED)

NONPARAMETRIC MEDIAN ESTIMATION

(With Application to Number of Simulation Replications Needed)

J. R. Knaub, Jr.

U.S. Army Logistics Center

When the variance of results from replications of a simulation is small, it is intuitively apparent that a small sample size (small number of replications) will be sufficient. If four replications are made, for instance, and the difference between the smallest and largest of the four observations is of little or no practical importance, this sample size may be sufficient.

Often, ten observations can be thought of as a minimum sample size, because, using distribution-free tolerance limits, there is a 75.6% confidence that 75.0% or more of the population lies between the extremes of the ten values observed. With four observations, this confidence is only 26.2%.

If, as in the example above, the difference between the smallest (o_1) and the largest (o_4) of four observations is of little or no practical importance, it is of interest to find the probability that the median of the population lies between the extremes of the four observations. This probability is $1-2\,(\frac{1}{2})^4 = 0.875$. If the central limit theorem is applicable*, then the median is approximately equivalent to the mean. Thus, the mean of the sample, in this case, is very likely to be very close to the mean of the population.

Suppose that seven replications are made instead of four. There is a probability of 0.875 that the population median (M) will fall between the second smallest (o_2) and second largest (o_6) observations, which is the same probability that M falls between o_1 and o_4 out of four observations. The difference is that $o_6 - o_2$ in the case of seven observations, is probably smaller than $o_4 - o_1$ for the case of four observations. See Table I for details. Column 1 in this table is the sample size. Column 2 is the probability that M will be located between the extremes of the sample of size n. Columns 3 and 4 provide the probabilities that the 10% and 5% tails, respectively, of the population will be drawn from at least once in n attempts. Columns 5, 6, and 7 provide similar information, but for the case where o_1 and o_n are ignored. Columns 8 and 9 are distribution-free tolerance limit results. They provide the probabilities that 50% or more, or 75% or more of the population is to be found between o_1 and o_n, the first and n^{th} ordered observations (ordered from smallest to largest).

Although a fixed sample size approach could be taken using the tables provided here, and some prior knowledge as to what variance will be experienced in most cases, a sequential approach would appear to often be more appropriate. Consider the following equation:

* The central limit theorem is applicable if each observation is actually the mean of a number of observations.

$P[o_j \leq M \leq o_{n+1-j}]$ = probability that the median of the population lies between the j^{th} smallest and j^{th} largest of n observations, where o_j is the j^{th} ordered observation.

$$= 1 - (1/2)^{n-1} \sum_{i=1}^{j} \binom{n}{n+1-i}$$

If one first determines the minimum acceptable value for $P[o_j \leq M \leq o_{n+1-j}]$ and the maximum acceptable value for $o_{n+1-j} - o_j$, then one may increase n and j until these conditions are satisfied.

Table II contains values for $P[o_j \leq M \leq o_{n+1-j}]$ for selected values of n and j. The rows are labelled by sample size (n) and the columns by the number of ordered observations (j - 1) eliminated from both the low and high ends of the set of ordered observations. (Therefore 2j-2 observations are eliminated.)

To use Table II in a sequential manner, first determine the value of $o_{n+1-j} - o_j$ to be used, or perhaps $(o_{n+1-j} - o_j)/\bar{X}$ and the value of $P[o_j \leq M \leq o_{n+1-j}]$ will be read from the table to determine the minimum sample size which could possibly satisfy the former criterion. If conditions are not satisfied, look in the table to see how many more observations must be taken so that j may be incremented by one, and the criterion for $P[o_j \leq M \leq o_{n+1-j}]$ met. After those additional observations are taken, the other criterion (e.g., $o_{n+1-j} - o_j$) can be checked for the new case. This process is repeated until both criteria are met.

Example:

Let the criteria be that $P[o_j \leq M \leq o_{n+1-j}]$ must be greater than or equal to 0.90, and that $o_{n+1-j} - o_j$ be less than or equal to 2.0, where the observations are number of personnel required to operate a given unit as structured by a computer model. Looking at Table II, it is seen that at least five observations (replications) are needed. Suppose the following observations are obtained: 42.6, 41.8, 41.9, 42.1, and 43.9.

Since $o_1 = 41.8$, and $o_5 = 43.9$, and $43.9 - 41.8 = 2.1 > 2.0$, more observations must be taken. From Table II, it is seen that at least three more observations are needed. Suppose that the next three observations are 42.4, 43.0, and 40.7. Now, $o_2 = 41.8$, $o_7 = 43.0$, and $43.0 - 41.8 = 1.2 < 2.0$. Therefore, both criteria are met, and sampling ceases. A sample size of eight observations (replications) has therefore been found to be adequate. Suppose further that the central limit theorem is applicable in this simulation analysis. The mean is 42.3 and the median is 42.4. Therefore, 42.3 is the mean value to be reported.

Note that in many other applications, such as values collected in surveys, the median is a far more appropriate measure to use. It would limit the influence of any outliers, and since median estimation is the focus of the theory provided here, the application is exact.

TABLE I

COL. 1 SAMPLE SIZE (n)	COL. 2 $P[O_1 \leq M \leq O_n]$	COL. 3 P[10% TAILS NOT REPRESENTED BETWEEN O_1 AND O_n]	COL. 4 P[5% TAILS NOT REPRESENTED BETWEEN O_1 AND O_n]	COL. 5 $P[O_2 \leq M \leq O_{n-1}]$	COL. 6 P[10% TAILS NOT REPRESENTED BETWEEN O_2 AND O_{n-1}]	COL. 7 P[5% TAILS NOT REPRESENTED BETWEEN O_2 AND O_{n-1}]	COL. 8 P[Y=0.500]	COL. 9 P[Y=0.750]
1	0.000	0.800	0.900	-	-	-	0.000	0.000
2	0.500	0.640	0.810	-	-	-	0.250	0.062
3	0.750	0.512	0.729	0.000	0.944	0.986	0.500	0.156
4	0.875	0.410	0.656	0.375	0.896	0.972	0.688	0.262
5	0.938	0.328	0.590	0.625	0.840	0.955	0.812	0.367
6	0.969	0.262	0.531	0.781	0.778	0.935	0.891	0.466
7	0.984	0.210	0.478	0.875	0.714	0.912	0.938	0.555
8	0.992	0.168	0.430	0.930	0.650	0.888	0.965	0.633
9	0.996	0.134	0.387	0.961	0.587	0.861	0.980	0.700
10	0.998	0.107	0.349	0.979	0.527	0.833	0.989	0.756
15	1.000	0.035	0.206	0.999	0.283	0.682	1.000	0.920

FORMULATIONS FOR TABLE I:

COL. 1: self-explanatory

COL. 2: $1 - (\frac{1}{2})^{n-1}$

COL. 3: $(0.8)^n$

COL. 4: $(0.9)^n$

COL. 5: $1 - (n+1)\,(\frac{1}{2})^{n-1}$

* COL. 6: $(0.8)^n + 0.2n\,(0.8)^{n-1} + 0.01n\,(n-1)(0.8)^{n-2}$

* COL. 7: $(0.9)^n + 0.1n\,(0.9)^{n-1} + 0.0025n\,(n-1)\,(0.9)^{n-2}$

** COL. 8: $1 - n(0.5)^{n-1} + (n-1)\,(0.5)^n$

** COL. 9: $1 - n(0.75)^{n-1} + (n-1)\,(0.75)^n$

* $(1-2p)^n + 2pn\,(1-2p)^{n-1} + p^2n(n-1)\,(1-2p)^{n-2}$
where $p = 0.10$ for column 6 and 0.05 for column 7

** $1-n\gamma^{n-1} + (n-1)\gamma^n$
where $\gamma = 0.5$ for column 8 and 0.75 for column 9

TABLE II

J-1 / N	0	1	2	3	4	5	6	7	8	9	10
1											
2	0.5000										
3	0.7500										
4	0.8750	0.3750									
5	0.9375	0.6250									
6	0.9688	0.7813	0.3125								
7	0.9844	0.8750	0.5469								
8	0.9922	0.9297	0.7109	0.2734							
9	0.9961	0.9609	0.8203	0.4922							
10	0.9980	0.9785	0.8906	0.6563	0.2461						
11	0.9990	0.9883	0.9346	0.7734	0.4512						
12	0.9995	0.9937	0.9614	0.8540	0.6123	0.2256					
13	0.9998	0.9966	0.9775	0.9077	0.7332	0.4189					
14	0.9999	0.9982	0.9871	0.9426	0.8204	0.5760	0.2095				
15	0.9999	0.9990	0.9926	0.9648	0.8815	0.6982	0.3928				
16	1.0000	0.9995	0.9958	0.9787	0.9232	0.7899	0.5455	0.1964			
17	1.0000	0.9997	0.9977	0.9873	0.9510	0.8565	0.6677	0.3709			
18	1.0000	0.9999	0.9987	0.9925	0.9691	0.9037	0.7621	0.5193	0.1855		
19	1.0000	0.9999	0.9993	0.9956	0.9808	0.9364	0.8329	0.6407	0.3524		
20	1.0000	1.0000	0.9996	0.9974	0.9882	0.9586	0.8847	0.7368	0.4966	0.1762	
21	1.0000	1.0000	0.9998	0.9985	0.9928	0.9734	0.9216	0.8108	0.6167	0.3364	
22	1.0000	1.0000	0.9999	0.9991	0.9957	0.9831	0.9475	0.8662	0.7137	0.4765	0.1682
23	1.0000	1.0000	0.9999	0.9995	0.9974	0.9894	0.9653	0.9069	0.7900	0.5951	0.3224
24	1.0000	1.0000	1.0000	0.9997	0.9985	0.9934	0.9773	0.9361	0.8484	0.6925	0.4587
25	1.0000	1.0000	1.0000	0.9998	0.9991	0.9959	0.9854	0.9567	0.8922	0.7705	0.5756
26	1.0000	1.0000	1.0000	0.9999	0.9995	0.9975	0.9906	0.9710	0.9245	0.8314	0.6731
27	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9941	0.9808	0.9478	0.8779	0.7522
28	1.0000	1.0000	1.0000	1.0000	0.9998	0.9991	0.9963	0.9875	0.9643	0.9128	0.8151
29	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9977	0.9919	0.9759	0.9386	0.8640
30	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9986	0.9948	0.9839	0.9572	0.9013

This table was compiled from the computer program of Appendix A by Mrs. Ida D. Price, et. al.

Appendix A

```
      DIMENSION ARRAY(30,11)
      OPEN(8,FILE='FAC.DAT',STATUS='OLD',DISP='KEEP')
      OPEN(12,FILE='FAC.OUT',STATUS='NEW',DISP='KEEP')
100   READ(8,10,END=1000)N,J
10    FORMAT(1X,I2,1X,I2)
      SUM=0
      ANS=0
      DO 200 I=1,J
      A=N+1-I
      CALL FACT(A)
      B=I-1
      CALL FACT(B)
      C=1.0/(A*B)
      SUM=SUM+C
200   CONTINUE
      XN=N
      CALL FACT(XN)
      D=(.5)**(N-1)
      ANS=1-D*XN*SUM
      ARRAY(N,J)=ANS
      GO TO 100
1000  CONTINUE
      WRITE(12,15)
 15   FORMAT(1X,'J-1',4X,'0',6X,'1',6X,'2',6X,'3',6X,'4',6X,'5',
     & 6X,'6',6X,'7',6X,'8',6X,'9',6X,'10',/,4X,'N')
      DO 116 N=1,30
      WRITE(12,16)N,(ARRAY(N,J),J=1,11)
 16   FORMAT(3X,I2,11(1X,F6.4)/)
116   CONTINUE
      STOP
      END
C
      SUBROUTINE FACT(D)
      PROD=1
      DO 100 I=1,D
      PROD=PROD*I
100   CONTINUE
      D=PROD
      RETURN
      END
```

Mrs. Ida D. Price, et. al.

Appendix B

Finally, note the following: Suppose n_1 observations are taken, and then n_2-n_1 additional observations are taken in the event that the first n_1 have an unacceptable range. Let $P[o_1 \leq M \leq o_{n_1}] \geq P[o_2 \leq M \leq o_{n_2-1}]$. This is the procedure recommended in this paper. It is of interest to note that the probability of saying the median has been bounded is larger in this two step method than if all n_2 observations were taken at once. This is explained in the unlikely event that the first n_1 observations have an acceptable range, but the first n_2 ($n_2>n_1$) are such that $o_{n_2-1} - o_2$ is too large. However, if the first n_1 are acceptable, it is more likely that these n_1 values are closer to the true median than any other value. This is analogous to applying a chi-square goodness-of-fit test and calculating a very small chi-square statistic. There may be a very small probability of having that value under the null hypothesis; however, there is an even smaller probability of having such a value under any alternative hypothesis. Thus, the null hypothesis is the most likely candidate. Similarly, here it would have been appropriate to quit after n_1 such observations as described above. The true median is probably indicated, although it was not likely that this would happen when the underlying population probably was more disperse than indicated by such an early finish to this sequential procedure. Thus, one could stop after the first step with a savings in sampling, because the median may be located more quickly than expected.

The above considerations affect the number of observations needed; however, it does not affect table II which gives the probability that the true median is bounded by two specific values found in the sample. The considerations of this appendix deal with the probability that the specific values will be close enough at some step along the way.

Appendix C

Modifications were made to the program found in Appendix A, but extension of P values in Table II for larger sample sizes was only accomplished to N = 125. From, approximately, N = 80 to N = 125, however, a simple linear relationship can be seen which, for a given value of N, accurately predicts the j value for which P is approximately 0.90. A simple simulation was used to test the validity of this relationship for N = 200. Results were good, as shown in the computer printouts that follow. Thus the methodology discussed in this paper can be extended to more classes of problems.

The following values form an extension to Table II:

N	j	P
40	15	0.9193
50	19	0.9351
60	24	0.9075
70	28	0.9278
80	33	0.9071

For N > 80, (N,j) => P > 0.9 when j = ||(N-80)(0.46)+33||, where ||x|| is x truncated to a whole number.

```
      FAC.FOR:

      REAL*8 A,B,C,D,XN,SUM,ANS
      OPEN(8,FILE='FAC.DAT',STATUS='OLD',DISP='KEEP')
      OPEN(12,FILE='FAC.OUT',STATUS='NEW',DISP='KEEP')
  100 READ(8,10,END=1000)N,J
   10 FORMAT(1X,I3,1X,I3)
      XXN=N
      SUM=0
      ANS=0
      DO 200 I=1,J
      A=N+1-I
      CALL FACT(A)
      B=I-1
      CALL FACT(B)
            C=1.0/(A*B)
            SUM=SUM+C
  200 CONTINUE
      XN=N
      CALL FACT2(XN)
            D=2.
      ANS=1.-XN*D*SUM
      WRITE(12,16)N,J,ANS
   16 FORMAT(/,4X,2(3X,I3),5X,F6.4,/)
      GO TO 100
 1000 CONTINUE
      STOP
      END
C
      SUBROUTINE FACT(D)
      REAL*8 D,PROD
      PROD=1.
      DO 100 I=1,D
      II=D-I+1
      PROD=PROD*II
                  PROD=PROD/30.
  100 CONTINUE
      D=PROD
      RETURN
      END
C
      SUBROUTINE FACT2(D)
      REAL*8 D,PROD
      PROD=1.
      DO 100 I=1,D
      II=D-I+1
      PROD=PROD*II
                  PROD=PROD/60.
  100 CONTINUE
      D=PROD
      RETURN
      END
```

FAC.DAT:

2	1
3	1
4	1
4	2
6	3
9	3
9	4
10	1
10	2
10	3
10	4
10	5
15	2
16	7
20	10
30	11
80	32
80	33
80	34
120	51
120	52
125	53
125	54

FAC.OUT:

2	1	0.5000
3	1	0.7500
4	1	0.8750
4	2	0.3750
6	3	0.3125
9	3	0.8203
9	4	0.4922
10	1	0.9980
10	2	0.9785
10	3	0.8906
10	4	0.6563
10	5	0.2461
15	2	0.9990
16	7	0.5455
20	10	0.1762
30	11	0.9013
80	32	0.9433
80	33	0.9071
80	34	0.8544
120	51	0.9176
120	52	0.8797
125	53	0.9268
125	54	0.8930

SIMFACALL.FOR;

```
      DIMENSION Y(5000)
1     READ(5,*,END=99)N,I
      II=I
      READ(5,*)NC,J1
      IF(NC.GE.5000)GO TO 99
      J2=NC-J1
      Z=N
      DO 10   J=1,N
       NX=0
      DO 20   K=1,NC
      Y(K)=RAN(I)
      IF(Y(K).GT.0.5)GO TO 20
      NX=NX+1
20    CONTINUE
       IF((NX.GE.J1).AND.(NX.LE.J2))X=X+1.
10    CONTINUE
      P=X/Z
      WRITE(6,*)N,II
      WRITE(6,*)NC,J1,P
      WRITE(7,*)N,II
      WRITE(7,*)NC,J1,P
      X=0.
      GO TO 1
99    STOP
      END
```

SIMFACALL.COM;

```
$RUN SIMFACALL
100000,78125
2,1
100000,65557
2,1
100000,-13
2,1
100000,78125
3,1
100000,78125
4,1
100000,78125
4,2
100000,78125
6,3
250000,78125
9,3
250000,65557
9,3
250000,-13
9,3
250000,78125
9,4
250000,78125
10,1
250000,78125
10,2
250000,78125
10,3
250000,78125
10,4
250000,78125
10,5
500000,78125
15,2
500000,78125
16,7
500000,78125
20,10
500000,78125
30,11
1000000,78125
80,32
1000000,78125
80,33
1000000,78125
80,34
1000000,78125
120,51
1000000,78125
120,52
1000000,78125
125,53
1000000,78125
125,54
5000000,78125
200,88
5000000,78125
200,89
55,55
5555555,55
$EXIT
```

FOR007.DAT:

```
100000      78125
     2          1  0.5023800
100000      65557
     2          1  0.4989200
100000        -13
     2          1  0.4981800
100000      78125
     3          1  0.7520400
100000      78125
     4          1  0.8771000
100000      78125
     4          2  0.3773400
100000      78125
     6          3  0.3132800
250000      78125
     9          3  0.8190960
250000      65557
     9          3  0.8194000
250000        -13
     9          3  0.8208720
250000      78125
     9          4  0.4912520
250000      78125
    10          1  0.9982160
250000      78125
    10          2  0.9782400
250000      78125
    10          3  0.8899640
250000      78125
    10          4  0.6562600
250000      78125
    10          5  0.2466600
500000      78125
    15          2  0.9990780
500000      78125
    16          7  0.5458920
500000      78125
    20         10  0.1769040
500000      78125
    30         11  0.9012580
1000000     78125
    80         32  0.9434850
1000000     78125
    80         33  0.9071050
1000000     78125
    80         34  0.8546970
1000000     78125
   120         51  0.9176350
1000000     78125
   120         52  0.8796270
1000000     78125
   125         53  0.9265760
1000000     78125
   125         54  0.8922570
5000000     78125
   200         88  0.9232274
5000000     78125
   200         89  0.8962938
```

QUANTILE-INFORMATION-FUNCTIONAL STATISTICAL INFERENCE AND UNIFICATION OF DISCRETE AND CONTINUOUS DATA ANALYSIS

Emanuel Parzen
Department of Statistics
Texas A&M University

ABSTRACT. This paper presents results from our research program on the development of statistical methods based on quantile-information-function estimation statistical inference. Its major goal is to define, and to apply the comparison quantile function $D(u;F,G)$ of two distribution functions F and G. We outline how standard methods of fitting a parametric model $F(y,\theta)$ to one or two random univariate samples can be developed in terms of the empirical quantile function $Q^{\sim}(u)$ by suitably defining the sample comparison quantile function $D^{\sim}(u,\theta) = D(u;F^{\sim},F(\cdot,\theta)$ and the sample comparison density function $d^{\sim}(u;\theta) = D^{\sim}{}'(u;\theta)$ for both continuous and discrete data. In the continuous case we define $D^{\sim}(u);\theta) = F(Q^{\sim}(u);\theta)$. In the discrete case we define first $d^{\sim}(u;\theta)$ and then define $D^{\sim}(u;\theta)$ to be its integral. Other concepts defined include: continuous versions of discrete quantile functions, identification quantile function, information distributions, and scientific statistical science. Emphasis is placed on developing a notation which applies to both discrete and continuous data analysis.

KEY WORDS. Quantile function, quantile density function, information, entropy, function estimation, functional inference, comparison quantile function, comparison density function, identification quantile function, goodness of fit, probability model identification, scientific statistical science.

1. PROBABILITY MODEL IDENTIFICATION AND COMPARISON QUANTILE FUNCTIONS

Given data which one regards as a random sample (of size n) of a random variable Y, one would like to infer the probability law of Y.

Ensemble Probability Laws. The probability law of Y is described in general by its distribution function $F(y)=PROB[Y \leq y]$. $-\infty < y < \infty$, and/or its quantile function $Q(u) = F^{-1}(u)$ defined by

$$Q(u) = \inf \{y: F(y) \geq u\}. \quad 0 \leq u \leq 1 \quad .$$

Research supported by the U. S. Army Research Office Grant DAAG29-83-K-0051.

Sample Probability Laws (Nonparametric Estimators). The first step in statistical data analysis of a random sample is to form the sample distribution function

$$F^{\sim}(y) = \text{fraction of sample} \leq y.$$

and the sample quantile function $Q^{\sim}(u) = F^{\sim -1}(u)$. The probability theory required to develop statistical theory based on $F^{\sim}(y)$ and $Q^{\sim}(u)$ is given in Durbin (1973) and Csörgo (1983) respectively.

Parametric Probability Laws. One approach to identifying $F(y)$ and $Q(u)$ is to assume a parametric family of distributions $F(y;\theta)$ indexed by a parameter θ which is an m-dimensional vector belonging to a parameter space S_m.

This paper develops a general definition of the comparison quantile function $D(u;F,G)$ of two distribution functions $F(y)$ and $G(y)$. The ensemble or population comparison quantile function is defined $D(u;\theta) = D(u;F, F(\cdot;\theta))$.

The maximum likelihood estimator $\hat{\theta}$ can be shown to be definable as minimizing a measure of distance between $F(y;\theta)$ and $F^{\sim}(y)$ or equivalently the distance between the sample comparison quantile function $D^{\sim}(u,\theta) = D(u;F^{\sim},F(\cdot\ \theta)$ and the uniform distribution $D_0(u) = u$.

2. GOODNESS OF FIT AND COMPARISON QUANTILE FUNCTIONS

The role of sample comparison quantile functions is best introduced by considering the goodness of fit problem: test the hypothesis $H_0: F(y) = F(y,\theta^0)$ for some specified parameter value θ^0.

When $F(y;\theta^0)$ is continuous one transforms Y to $Z = F(Y,\theta^0)$, called the probability integral transform. Then H_0 is equivalent to $H_0: Z$ is uniform on the interval 0 to 1.

When the sample consists of observations $Y(1),\ldots Y(n)$ one transforms to $Z(j) = F(Y(j);\theta^0)$. Let $G^{\sim}(u)$, $0 \leq u \leq 1$, denote the sample distribution function of $Z(1),\ldots,Z(n)$. The Kolmogorov-Smirnov statistic for testing H_0 is defined by

$$D_n = \sqrt{n} \sup_{-\infty<y<\infty} \left| F^{\sim}(y) - F(y;\theta^0) \right|$$

or

$$D_n = \sqrt{n} \sup_{0<u<1} \left| G^{\sim}(u) - u \right|$$

One can show that

$$D_n = \sqrt{n} \sup_{0<u<1} \left| G^{\sim -1}(u) - u \right|$$

We propose that the sample quantile function $G^{\sim -1}(u)$ is the most useful way to express statistics to test H_0. One reason for this is the elegant formula expressing $G^{\sim -1}(u)$ in terms of the sample quantile function $Q^{\sim}(u)$ of $Y(1), \ldots Y(n)$ and the comparison quantile function.

$$G^{\sim -1}(u) = D^{\sim}(u;\theta^{o}) \quad .$$

To make this formula concrete, let us consider the realistic case of samples described by empirical probabilities, using notation such as the following:

K, number of distinct values in the sample,

$V(1) < \ldots < V(K)$, ordered distinct values in the sample;

$NV(J)$, number of sample members equal to $V(J)$, for $J=1 \ldots K$

$PV(J) = NV(J)/n$, empirical probability of $V(J)$;

$FV(J) = PV(1)+\ldots+PV(J)$, $FV(0) = 0$.

Note that $F^{\sim}(V(J)) = FV(J)$ and $FV(K) = 1$.

The sample quantile function $Q^{\sim}(u)$ has the elegant formula:

$$Q^{\sim}(u) = V(J), \quad FV(J-1) < u < FV(J) \text{ for } J=1 \ldots K.$$

The sample comparison quantile function has the formula

$$D^{\sim}(u\ \theta) = F(V(J), \theta), \quad FV(J-1) < u \leq FV(J) \text{ for } J=1 \ldots K.$$

To calculate the Kolmogorov-Smirnov statistic D_n let $Z(J)=F(V(J);\theta^{o})$. Then

$$D_n = \sqrt{n} \max_{J=1,\ldots,K} \max \{ \left| Z(J)-FV(J) \right| \ , \ \left| Z(J)-FV(J-1) \right| \}$$

To perform a goodness of fit test one needs both numerical and graphical procedures. A rejection region for H_0 at the 95% significance level for large sample sizes n is:

$D_n > 1.36$ for a simple hypothesis. and $D_n > .89$ if the null hypothesis is the composite hypothesis of a normal distribution with mean and variance estimated from the sample. A numerical procedure compares the value of D_n with these critical values. A possible graphical procedure is: plot, on the same graph, $F^{\sim}(y)$ and $F(y;\theta^o)$. More insight is obtained by plotting $D^{\sim}(u;\theta^o)$ and $D_o(u) = u$. In addition many additional numerical procedures for testing H_0 can easily be defined in terms of $D^{\sim}(u;\theta^o)-u$. For example one might use the Wilcoxon type statistic

$$WL = \sqrt{12} \int_0^1 \{D^{\sim}(u.\theta^o) - u\}\, du$$

Quick graphical procedures for identifying the probability law of a random sample are provided by $Q^{\sim}I(u)$. the sample identification quantile function defined below [section 4].

3. CONTINUOUS VERSION OF DISCRETE QUANTILE FUNCTIONS

Moments can be expressed elegantly in terms of $Q(u)$; thus

$$\text{Mean } M = \int_0^1 Q(u)\, du,$$

$$\text{Variance } VAR = \int_0^1 \{Q(u)-M\}^2\, du.$$

$$\text{Standard deviation } DS = \sqrt{VAR}\ .$$

We use a notation for mean and standard deviation which provides analogies for measures of location and scale based on the quantile function

$$\text{Median } MQ = Q(0.5).$$

$$\text{Quartile deviation } DQ = 2\,\{Q(.75) - Q(.25)\}\ .$$

Our concept of quartile deviation (which equals twice the inter-quartile range) is motivated by the concept that in the case that $Q(u)$ has a derivative $Q'(u)$. the derivative $Q'(0.5)$ is a useful universal measure of scale; DQ is a difference quotient which can be regarded as a very rough approximation to $Q'(0.5)$.

An important step in our program for unifying discrete and continuous data analysis is to associate to a discrete quantile function a continuous quantile functon which is used to define the median and quartile deviation of a discrete distribution.

The distribution function and quantile function of a discrete random variable Y are piecewise constant. For example let Y be Bernoulli with PROB[Y = 0] = q. PROB[Y = 1] = p, where p+q = 1. Then F(y) = 0, q, 1 according as $y < 0$, $0 \leq y < 1$, $1 \leq y$. Further Q(u) = 0, 1 according as $0 < u \leq q$, $q < u \leq 1$. We propose a method of associating with a discrete quantile function Q(u) a continuous quantile function denoted QC(u), which for the Bernoulli distribution has value 0 at u = q/2, value 1 at u = 1-(p/2), is a line of slope 2 between these points, and has value p at u = 0.5. Consequently the median MQ = p, in agreement with the fact that the mean M = p.

We use the following notation:

K, number of discontinuity points of F(y);

V(1)<...<V(K) values at which F(y) jumps,

FV(J) = F(V(J)); FV(0) = 0; PV(J) = FV(J)-FV(J-1);

U(J) = FV(J)-(PV(J)/2).

We call UF(J), J=1...k, the midranks of the discrete distribution F(y); they play a (two-) key role in data analysis. We call V(J), J=1,...,K, the probable values of the discrete distribution.

The discrete quantile function Q(u) is given by

$$Q(u) = V(j). \quad FV(J-1) < u \leq FV(J) \quad .$$

The continuous quantile function QC(u) associated with a discrete quantile function Q(u) is defined to be piecewise linear between its values at u = UF(J) where it is defined to satisfy

$$QC(UF(J)) = V(J), \quad J = 1...K.$$

For $0 < u < UF(1)$, define QC(u) = V(1); for $UF(K) < u < 1$, define QC(u) = V(K).

For a discrete quantile function, we define its median and quartile deviation by

$$MQ = QC(0.5). \quad DQ = 2\ \{QC(.75) - QC(.25)\}$$

The mean and variance of the midranks U are denoted MU and VARU:

$$MU = \sum_{J=1}^{K} U(J)\ PV(J), \quad VARU = \sum_{J=1}^{K} \{U(J)-MU\}^2\ PV(J).$$

Always MU = 0.5. VARU is approximately 1/12 (the variance of a uniform distribution on 0 to 1).

4. IDENTIFICATION QUANTILE FUNCTION

To a quantile function Q(u), one associates a quantile function, denoted QI(u). We call QI(u) the identification quantile function because its values near 0 and 1 and its overall shape can be shown to provide quick heuristic methods for identifying the type of distribution (normal, exponential, etc.) and its tail behavior (short tail medium tail, long tail; within medium tail, one can distinguish medium-short, medium-medium, medium-long). An excellent discussion of these concepts is given by Schuster (1984).

When Q(u) is continuous, we define QI(u) = {Q(u)-MQ}/DQ. When Q(u) is discrete, we define QI(u) = {QC(u) -MQ}/DQ. One may similarly define the sample identification quantile function Q~I(u).

The definition and applications of QI(u) was pioneered by Parzen (1984) under the name of informative-quantile function.

Transforming a random sample by subtracting its mean M, and dividing by the standard deviation DS, is equivalent to forming the Z-Quantile function QZ(u) = {Q(u)-M}/DS.

5. COMPARISON QUANTILE AND COMPARISON DENSITY FUNCTIONS CONTINUOUS CASE

Let G(y) and F(y) be continuous distribution functions. We define the comparison quantile function to be

$$D(u;F,G) = GF^{-1}(u).$$

Measures of distance between D(u;F,G) and $D_0(u) = u$ provide measures of distance between F(y) and G(y).

Although the mathematical definitions require no interpretation for G and F, we usually think of G(y) as a probability model for Y and think of F(y) as the true distribution function of Y. The quantile function of Z = G(Y) is $GF^{-1}(u)$.

The quantile density d(u;F,G) = D'(u;F,G) is called a comparison density function. Explicitly,

$$d(u;F,G) = gF^{-1}(u) \;/\; fF^{-1}(u) \;.$$

Its interpretation is more evident by writing it as a likelihood ratio:

$$d(u;F,G) = g(x)/f(x) \text{ when } F(x) = u.$$

6. INFORMATION DIVERGENCE, ENTROPY, CROSS ENTROPY

When F and G are both continuous, with respective probability density functions f(y) and g(y), information divergence is defined by

$$I(F;G) = \int_{-\infty}^{\infty} [-\log \{g(y)/f(y)\}] \, f(y) \, dy$$

Information divergence has a fundamental decomposition

$$I(F;G) = H(F;G) - H(F)$$

defining cross-entropy H(F;G) and entropy H(F) by

$$H(F;G) = \int_{-\infty}^{\infty} \{-\log g(y)\} \, f(y) \, dy$$

$$H(F) = H(F,F) = \int_{-\infty}^{\infty} \{-\log f(y)\} \, f(y) \, dy.$$

It should be noted that

$$H(F) = H(Q) = \int_0^1 \{-\log fQ(u)\} \, du = \int_0^1 \{\log q(u)\} \, du$$

where $q(u) = Q'(u)$ is the quantile-density function.

An important measure of the distance between D(u;F,G) and $D_0(u) = u$ is the negative of the entropy of d(u;F,G):

$$-H(d) = -\int_0^1 \log d(u;F,G) \, du.$$

By making a change of variable $u = F(x)$, $x = Q(u)$ one can verify that $-H(d) = I(F;G)$. This fact provides a new interpretation of information divergence as a measure of distance between D(u;F,G) and u. Minimizing information divergence I(F;G) is equivalent to maximizing entropy H(d).

7. COMPARISON DENSITY DISCRETE DISTRIBUTIONS

Let F(y) and G(y) be discrete distributions with the same probable values $V(1)<\ldots<V(K)$. Their probability mass functions are denoted PF(J) and PG(J) respectively. In terms of FV(J) = F(V(J)) and GV(J) = G(V(J)) we define

$$PF(J) = FV(J)-FV(J-1), \quad PG(J) = GV(J)-GV(J-1).$$

We define the comparison density function d(u;F,G) by

$$d(u;F,G) = PG(J)/PF(J) \text{ for } FV(J-1) < u \leq FV(J).$$

We define the comparison quantile function

$$D(u;F,G) = \int_0^u d(t;F,G)\, dt$$

As an example of how d(u;F,G) provides a unified notation for concepts that usually are defined separately for continuous and discrete random variables, consider the information divergence of discrete distributions

$$I(F;G) = \sum_{J=1}^{K} [-\log \{PG(J)/PF(J)\}]PF(J)$$

One may verify that

$$I(F;G) = \int_0^1 \{-\log d(u,F,G)\}\, du$$

The chi-squared statistic is interpreted by

$$\sum_{J=1}^{K} \frac{\{PG(J)-PF(J)\}^2}{PF(J)} = \int_0^1 [d(u;F,G)-1]^2\, du.$$

Empirical probabilities are denoted $PF\tilde{}(J)$. The asymptotic distribution theory of chi-squared statistics can be expressed as properties of the stochastic process $D(u;F\tilde{},G)-u$, assuming PG(J) is the true probability mass function.

8. WILCOXON STATISTICS

An important measure of the "distance" of D(u;F,G) from u is

$$WL(F;G) = -\int_0^1 \{D(u;F,G) - u\}\, du$$

The Wilcoxon statistic used in two-sample nonparametric statistical inference has mathematical similarities to WL; we therefore call WL a Wilcoxon statistic.

By integration by parts one may show that

$$WL(F;G) = \int_0^1 u\, d(u;F,G)\, du - 0.5$$

When F and G are both discrete, WL has an important interpretation in terms of midranks.

$$\int_0^1 u\, d(u;F,G)\, du = \sum_{J=1}^{K} \{\int_{F(J-1)}^{F(J)} u\, du\} \{PG(J)/PF(J)\}$$

$$= \sum_{J=1}^{K} PG(J)\, UF(J)$$

where we write UF(J) for the midranks of F(y) and UG(J) for the midranks of G(y).

We can interpret WL as a difference of means of midranks UF:

$$WL = E_G[UF] - E_F[UF]$$

One can also show that

$$WL = E_G[UG] - E_F[UG]$$

9. BIVARIATE AND TWO SAMPLE PROBABILITY LAW MODELING

Techniques of comparing univariate distribution functions can also be used to model the relations between several variables X and Y. Relations between Y and X can be expressed as relations between the conditional distribution functions FY:X(y:x) and the unconditional distribution function F(y). We use : rather than | to denote conditioning and we define

$$FY:X(y:x) = PROB\,[Y \leq y : X = x] .$$

The concepts and algorithms of conditional distributions can be interpreted to apply even when X and Y are not random variables. An example of this approach is the problem of comparing two samples.

The problem of the comparison of the distributions of two random variables Y_1 and Y_2 given random samples $Y_1(1),\ldots,Y_1(n_1)$ and $Y_2(1),\ldots Y_2(n_2)$ can be formulated as a comparison of empirical conditional distributions with empirical unconditional distributions. Let $n = n_1 + n_2$.

Define n observations $(X(1),Y(1)),\ldots,(X(n),Y(n))$ of a pair of hypothetical random variables X and Y by:

$$(X(j),Y(j)) = (1,Y_1(j)) \quad \text{if } j = 1,\ldots,n_1$$

$$= (2,Y_2(j-n_1)) \quad \text{if } j = n_1+1,\ldots,n \ .$$

The unconditional distribution of Y is the empirical distribution, denoted $F\tilde{}Y$, of the pooled sample $Y_1(1),\ldots.Y_1(n_1)$, $Y_2(1),\ldots Y_2(n_2)$. The conditional distribution of Y given X=1. denoted $F\tilde{}Y\ X=1$, is the empirical distribution of the first sample $Y_1(1),\ldots,Y_1(n_1)$.

In terms of the comparison distribution function

$$D\tilde{}(u) = D(u\ F\tilde{}Y, F\tilde{}Y:X = 1)$$

one can express the Kolmogorov-Smirnov two-sample statistic as essentially the supremem of $|D\tilde{}(u) - u|$, and the Wilcoxon two-sample statistic as essentially the integral of $D\tilde{}(u) - u$.

One can give explicit formulas relating various statistics which one forms in a comprehensive data analysis computer program, and which are in fact equivalent to the Wilcoxon rank sum statistic. Let

S = sum of ranks of first sample in pooled sample,

M = mean of pooled sample midranks $UF\tilde{}Y$ in first sample (M = conditional mean of $UF\tilde{}Y$ under $F\tilde{}Y:X = 1$);

RU = rank correlation of midranks $UF\tilde{}Y$ and $UF\tilde{}X$. (for truly bivariate samples, RU is the Spearman rank correlation statistic).

One can show that:

$$M = (S/n_1 n) - (1/2n)$$

$$(RU)^2 = (M-0.5)^2(12n_1/n_2) \quad .$$

10. INFORMATION DISTRIBUTIONS AND APPROXIMATIONS TO t AND F DISTRIBUTIONS

Another aspect of the unification made possible by a quantile-information-functional approach is that statisticians may be able to approximately compute sigificance levels of standard statistical tests without having to consult a large array of statistical tables.

To test H_0: Y is $N(\mu,\sigma^2)$, $\mu = 0$ against H_1:Y is $N(\mu.\sigma^2)$ one should use the log likelihood ratio test statistic log Λ for testing $\mu = 0$ versus $\mu \neq 0$, which can be shown to satisfy [see Rohatgi. (1984), p. 723, 725]

$$-2 \log \Lambda = n J^{\sim} \quad , \quad J^{\sim} = \ln \left(1 + \frac{(M^{\sim}Y)^2}{VAR^{\sim}Y}\right) \quad .$$

The distribution of $J^{\sim}$ is called an information distribution. It can be described symbollically by writing

$$J^{\sim} = \ln \left(1 + \frac{1}{n-1} t^2_{n-1}\right)$$

where t_{n-1} denotes a random variable obeying a t distribution with (n-1) degrees of freedom.

The sample correlation coefficient $R^{\sim}$ of a bivariate normal sample has the property that the log likelihood ratio test statistic for testing $R = 0$ against $R \neq 0$ satisfies [see Rohatgi, (1984), p. 724]

$$-2 \log \Lambda = - n \ln (1-R^{\sim 2}) = nJ^{\sim}$$

defining $J^{\sim} = - \ln (1-R^{\sim 2})$. The distribution of $J^{\sim}$ is an information distribution which can be described symbolically

$$J^{\sim} = \ln \left(1 + \frac{1}{n-2} t^2_{n-2}\right) \quad .$$

Asymptotic distribution of information distributions. One can show that asymptotically

$$n \ln \left(1 + \frac{1}{n} t^2_n\right) \to \chi^2_1 \quad , \text{ chi-square 1 d.f.}$$

More generally, Let $F_{m,n}$ denote a random variable obeying an F distribution with degrees of freedom (df) equal to m and n in the numerator and denominator respectively. Then

$$n \ln \left(1 + \frac{m}{n} F_{m,n}\right) \to \chi^2_m \quad . \text{ chi-square m d.f.}$$

To obtain a finite sample approximation to the distribution of the information distribution we write

$$(*) \qquad n \ln \left(1 + \frac{m}{n} F_{m,n}\right) = \chi^2_m / h(m.n)$$

where h(m,n) (which tends to 1 as $n \to \infty$) is a constant given explicitly by

$$h(m.n) = h = \frac{2}{n} \left\{1 - \left(1 + \frac{m}{n-2}\right)^{-2/m)}\right\}^{-1}$$

For n large, $h = \frac{n-2}{n}$. For m = 1, $h = \frac{(n-1)^2}{n(n-1.5)}$.

The constant h is determined by equating the means of both sides of the representation for F implied by eq. (*):

$$\frac{m}{n} F_{m,n} = e^{\chi^2_m/nh} - 1$$

which determines h to satisfy

$$\frac{m}{n-2} = (1 - \frac{2}{hn})^{-(m/2)} - 1 \quad .$$

The high degree of accuracy of formula (*) in the case m = 1 is discussed in a paper by Gaver and Kafadar (1984).

If one uses the fact that the t-distribution enjoys a close approximation (for $n \geq 7$) by the distribution of χ^2_1/nh then one need never consult tables of the t-distribution to form tests of the hypothesis H_0: $\mu = 0$. One can directly interpret without any table lookup the statistic $NJ^{\sim}$ defined by

$$NJ^{\sim} = (3.84)/h(1,n-1)J^{\sim}$$

Reject H_0 at the 5% level if $NJ^{\sim} < 1$. The interval $.5 < NJ^{\sim} < 1.5$ corresponds linearly to levels of significance $.995 > \alpha > .905$. The statistic $NJ^{\sim}$ is useful for sample size determination; the minimum sample size at which the observed value of $J^{\sim}$ would reject the hypothesis that J = 0 is approximately equal to $NJ^{\sim}$ times the observed sample size.

11. SCIENTIFIC STATISTICAL SCIENCE

This paper has presented results from our research program on the development of new statistical methods based on quantile-information-functional statistical inference (which we abbreviate FunStAT and call FUNctional STatistical Analysis Technology). The overall goal of this research program is to contribute to the development of scientific statistical science.

We propose that statistical science be called "scientific" (as opposed to "artistic") when it attempts to develop statistical methods in a unified way that can be systematically applied in many different fields of statistics. Scientific statistical science is desirable for its elegance (in order to help dispel the impression among applied scientists and engineers that statistical reasoning is at most a bag of

tricks). Scientific statistical science is desirable for its utility, since it enables statisticians to adapt statistical methods (in order to develop for different problems innovative methods customized to the unique features of each problem).

Scientific statistical science may be a necessity in the emerging era of "PC/StAT", personal portable computing statistical analysis technology. Statistical computing for personal computers can be developed to be interactive for ease of use and for effective integration of classical and currently emerging styles of statistical data analysis. The high cost of data relative to computing makes it sensible and wise to analyze one's data from as many points of view as possible. A unified framework for statistical reasoning will make it possible to more rigorously combine the results yielded by different algorithms which are applied to the same data.

Examples of identification quantile functions and comparison quantile functions generated by our IBM PC program EPSTAT are given in a separate report available from the author.

REFERENCES

Csörgo, M. (1983) "Quantile Processes with Statistical Applications," Vol. 42, Regional Conference Series in Applied Mathematics, Philadelphia: Society for Industrial and Applied Mathematics.

Durbin, J. (1973) "Distribution Theory for Tests Based on the Sample Distribution Function," Vol. 9, Regional Conference Series in Applied Mathematics, Philadelphia Society of Industrial and Applied Mathematics.

Gaver, D. P. and Kafadar, K. (1984) "A Retrievable Recipe for Inverse t", American Statistician, 38, 308-311.

Parzen, E. (1979) "Nonparametric Statistical Data Modeling" Journal of the American Statistical Association, (with discussion), 74, 105-131.

Parzen, E. (1979) "A Density-Quantile Function perspective on Robust Estimation" Robustness in Statistics, ed. R. Launer and G. and G. Wilkinson, New York: Academic Press, 237-258.

Parzen, E. (1982) "Data Modeling Using Quantile and Density-Quantile Functions," Some Recent Advances in Statistics, ed. J. Tiago de Oliveira and B. Epstein, New York: Academic Press. 23-52.

Parzen, E. (1982) "Data Modeling Using Quantile and Density-Quantile Functions," Some Recent Advances in Statistics, ed. J. Tiago de Oliveira and B. Epstein, New York: Academic Press. 23-52.

Parzen, E. (1983) "Quantiles, Parametric-Select Density Estimation, and Bi-Information Parameter Estimators," Proceedings of the 14th Annual Symposium on the Interface of Computer Science and Statistics, New York: Springer Verlag, 241-245.

Parzen, E. (1983) "Entropy Interpretation of Goodness of Fit Tests," Proceedings of the 28th Army Conference on the Design of Experiment (Monterey). Army Research Office: Durham, N.C.

Parzen, E. (1984) "Informative Quantile Functions and Identification of Probability Distribution Types," Proceedings of the 29th Army Conference on the Design of Experiments (Washington) Army Research Office: Durham, N.C. 97-107.

Rohatgi, V. K. (1983) Statistical Inference, John Wiley, New York.

Schuster, E. F. (1984) "Classification of Probability Laws by Tail Behavior," Journal of the American Statistical Assoc. 79, 936-939.

THE DISTRIBUTION OF A WEIGHTED VISIBILITY MEASURE ON A LINE SEGMENT UNDER SHADOWS CAST BY RANDOM DISKS HAVING A BIVARIATE NORMAL SCATTERING*

M. Yadin and S. Zacks
Technion, Israel Institute of Technology and
State University of New York at Binghamton

ABSTRACT

The present paper develops formulae for the computations of moments of a weighted measure of visibility along short line segments in the plane, when the shadowing objects consist of N (fixed number) of disks, which are randomly scattered around the origin, according to a bivariate normal distribution. These moments are based on visibility probabilities of points on the line segment. Formulae for the computation of these probabilities are derived. These formulae are based on the probabilities of general rectangles, half circles and triangles when the points have a bivariate normal distribution. The appendices provide formulae and FORTRAN subroutine functions for the computation of the required functions. An approximation is given to the distribution of the random measure of weighted visibility. The results of this research are applicable in various areas. In particular in the evaluation of the performance of laser range finders and other similar problems, when random objects in the field cast their shadows on a target.

Key Words: *Visibility, weighted measure of visibility, multinomial random field, shadowing process, probabilities of convex sets under bivariate normal distributions.*

*Partially supported by Contract DAAG29-83-D-0176 with the U. S. Army Research Office.

1. Introduction

The present paper is motivated by the following actual problem. A laser beam is oriented from a source at a point (x_o,y_o) to a line segment, T, whose center is located at the point (x_t,y_t). The energy in the beam is maximal at its center and is tapering off fast as the distance from the center increases. A common model for the distribution of the intensity of the energy around the center of the beam is the spherically symmetric Gaussian distribution, with very small standard deviation. Certain portions of the beam may be obstructed by obstacles which are randomly dispersed in the field. These obstacles could consist of different types of objects which are in the field, such as trees, bushes, piles of dirt, etc. The ray does not penetrate through such objects. If such an obstructing object intersects any ray from (x_o,y_o) to T, we say that the object casts a shadow on T. There may be different, sometimes overlapping shadows which are cast on T by different objects in the field. Thus, certain points on T are in the "light" and certain ones might be in the "dark". Let (x_R,y_R) denote the right hand limit point of T and (x_L,y_L) its left hand limit point. Let $\Delta = x_R - x_L$. We assume that the intensity of a ray connecting (x_o,y_o) with a point (x,y) on T is proportional to the normal (Gaussian) probability density function (PDF). The intensity function, $w(x)$, is normalized, so that its integral from $x_t-\Delta$ to $x_t+\Delta$ is equal to 1. We define an integrated measure, W, of the random amount of light (energy) that reaches T from the source. W is a random measure having a distribution which depends on the characteristics of the random field of the shadowing objects. In our previous studies (Yadin and Zacks [3,4]) we discussed

properties of random visibility measures when the shadowing objects are disks of random radii and random location. Furthermore, we assumed that the number of centers of disks falling within any specified (Borel) set, B, in the plane has a Poisson distribution with mean $\lambda(B)$. The radii of disks may or may not depend on the location of their centers. Such a model is called a Poisson random field of shadowing objects. The present study follows a different scattering model. First, we assume that the number of disks, N, is fixed (finite). Furthermore, given any partition of the plane to (Borel) sets $B_1, \ldots, B_m$, the number of disks $J_1, \ldots, J_m$, whose centers belong to $B_1, \ldots, B_m$, respectively, have a multinomial distribution, with probability vector $(\pi_1, \ldots, \pi_m)$ which depends on the specified sets and on the stochastic scattering mechanism. In addition each disk has a radius which is a realization of a random variable with a specified distribution. Such a model is called a <u>multinomial</u> field of shadowing objects.

In the present paper we further assume that the centers of disk have coordinates which are independent random vectors having a given bivariate normal distribution, and that the radii of disks are independent and identically distributed random variables, independent of the center locations. The motivation for studying such a model is due to a particular military application, in which the shadowing objects are artillery rounds. N rounds are scattered according to a bivariate normal distribution around an aim point (x^*,y^*). Each round when exploded creates a cloud of dust and debris of random size. A planar cut of such a cloud is modeled

as a disk. The model of the present paper can be further generalized to cases of several clusters of N disks, each one characterized by a different bivariate normal distribution.

In Section 2 we formally present the structure of the field and the probability models of the associated random variables. In Section 3 we discuss methods for computing the probability that any specified point on T is in the light, and the probabilities that any n specified points of T are simultaneously in the light. These are called visibility probabilities. The visibility probabilities are required for the computation of the moments of the random measure W. In Section 4 we discuss some properties of the distribution of W and the computation of its moments. A beta-mixture approximation to the distribution of W is discussed in Section 5. In Section 6 we discuss the subroutines which were programmed for computing the moments of W and the parameters of the beta-mixture approximation. Numerical examples are presented in Sections 3-5. As shown in Section 3, the computation of the visibility probability of a point requires subroutines to compute the probabilities of arbitrary rectangles and half-circles, under a standard bivariate normal distribution. The determination of the simultaneous visibility probability of n points, $n \geq 2$, requires subroutines for the computation of probabilities of arbitrary triangles under standard bivariate normal distributions. Formulae of these probabilities are derived in Appendices A and B. A.R. Didonato and R. K. Hageman [2] published a method for computing the probability of an arbitrary polygon under a bivariate normal

distribution. Their method is based primarily on the computation of probabilities of angular sectors. The leading factor in such computations is $\exp(-R^2/2)$ where R is the distance of the vertex of the angle from the origin (the center of the bivariate normal distribution). The main formula cannot be applied if R is large and some asymptotic approximations are provided. As shown in the examples, we are interested in triangles having at least one vertex which is very far away from the origin ($R \geq 100$). The direct calculations which are presented in the appendices do not require computations of angular probabilities, as in Didonato et al [2]. The subroutines provided in the present paper yield very accurate results in a matter of a few seconds of computer time.

2. The Theoretical Model

Let $P_o=(x_o,y_o)$ denote the source of the light (laser) beam. Let T be a line segment of width $w = 2\Delta$ centered at $P_t = (x_t,y_t)$ and perpendicular to the line segment $\overline{P_oP_t}$. Furthermore, let $P_R = (x_R, y_r)$ and $P_L = (x_L,y_L)$ denote the right and left end points of T respectively. Thus, if b_t denotes the slope of the line through P_o and P_t, the coordinates of P_R and P_L are given by

$$
\begin{aligned}
x_R &= x_t + \Delta|b_t|/(1+b_t^2)^{1/2} \\
x_L &= x_t - \Delta|b_t|/(1+b_t^2)^{1/2} \\
y_R &= y_t + \mathrm{sgn}(b_t)\Delta/(1+b_t^2)^{1/2} \\
y_L &= y_t - \mathrm{sgn}(b_t)\Delta/(1+b_t^2)^{1/2} \quad ,
\end{aligned}
\tag{2.1}
$$

where $\text{sgn}(b_t)$ is the sign of the slope b_t. Let $D_1, D_2, \ldots, D_N$ be N random disks. Each random disk is characterized by the triplet (X_i, Y_i, R_i), $i = 1, \ldots, N$, where (X_i, Y_i) are the coordinates of the center of the disk D_i and R_i is its radius. We assume that $(X_1, Y_1), \ldots, (X_N, Y_N)$ are independent and identically distributed (IID) random vectors having a bivariate normal distribution centered at the origin (0,0). Without loss of generality we assume that the bivariate normal distribution has zero correlation, i.e., $\rho=0$. Indeed, if $\rho\neq 0$, one can apply the orthogonal transformation $(X', Y') = (X,Y)(B)$, where (B) is a 2x2 matrix consisting of the eigenvectors of the covariance matrix of (X, Y). The orthogonal transformation B maps the random disks into random disks having the same radii. Furthermore, the distribution of (X', Y') is bivariate normal centered at (0,0) with correlation zero and variances λ_1 and λ_2, respectively where λ_1 and λ_2 are the eigenvalues of the covariance matrix

$$\sum = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}.$$

These eigenvalues are

$$\lambda_1 = \frac{\sigma_x^2 + \sigma_y^2}{2} + \frac{1}{2}\left[(\sigma_x^2 - \sigma_y^2)^2 + 4\rho^2\sigma_x^2\sigma_y^2\right]^{1/2}$$

(2.2)

$$\lambda_2 = \frac{\sigma_x^2 + \sigma_y^2}{2} - \frac{1}{2}\left[(\sigma_x^2 - \sigma_y^2)^2 + 4\rho^2\sigma_x^2\sigma_y^2\right]^{1/2}.$$

The elements of B are B_{ij}, i, j = 1,2 , where

$$b_{1j} = \frac{1}{c_j} \rho\sigma_x\sigma_y$$

$$b_{2j} = \frac{1}{c_j} (\sigma_y^2 - \lambda_j) \quad , \tag{2.3}$$

$$c_j = [(\sigma_x^2 - \lambda_j)^2 + \rho^2\sigma_x^2\sigma_y^2]^{1/2} . \quad j = 1,2.$$

The points (x_o,y_o), (x_R,y_R) and (x_L,y_L) are transformed to corresponding points (ξ_o,η_o), (ξ_R,η_R), (ξ_L,η_L). The line segment T is transformed to the line segment T', connecting (ξ_L,η_L) with (ξ_R,η_R).

The radii of the random disks, R_1, ..., R_N, are assumed to be independent random variables, having a common distribution, $G_R(r)$, concentrated over the interval [a,b], where $0 \leq a < b < \infty$. It is also assumed that $\{R_i; i=1, \ldots, N\}$ are independent of $\{(x_i,y_i);i=1,\ldots,N\}$. The randomly scattered disks may cast shadows on T. In Figure 1 we illustrate a possible realization of such a random shadowing process, in which N = 4 disks are randomly scattered. Two disks cast shadows on the line segment T. Thus, creating two dark intervals at the edges of T and one visible interval in the middle.

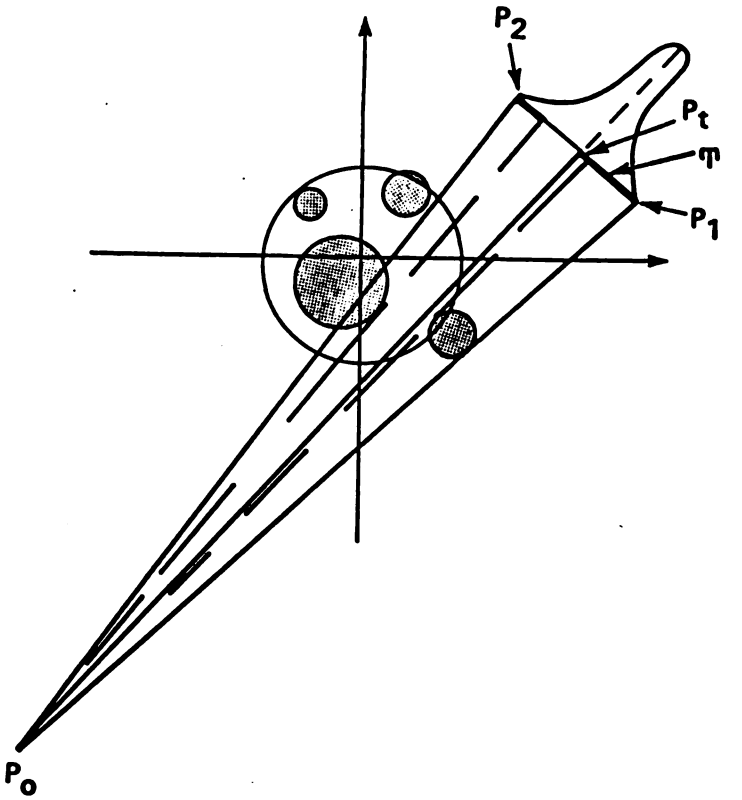


FIGURE 1. The Light (Laser) Beam and the Shadowing Disks

Let P = (x,y) be a point on T, and let I(x) = 1 or 0 if P is in the light (visible) or in the dark. The integrated measure of visibility on T is

$$(2.4) \qquad W = \frac{1}{A\tau} \int_{x_t-\Delta}^{x_t+\Delta} \phi\left(\frac{x-x_t}{\tau}\right) I(x)\,dx$$

$$= \frac{1}{A} \int_{-\Delta/\tau}^{\Delta/\tau} \phi(u)\, I(x_t+\tau u)\,du \quad ,$$

where $0<\tau<\infty$ is the standard deviation of the dispersion of the beam, and $\phi(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$, $-\infty<u<\infty$, is the probability

density function (PDF) of the standard normal distribution. Moreover, $A = 2\Phi(\frac{\Delta}{\tau})-1$, where $\Phi(z)$ is the standard normal integral.

Notice that W is a random variable assuming values in the interval [0,1]. W = 0 if T is completely in the dark and W = 1 if T is completely visible. The distribution of W has two jump points, at 0 and at 1. In the open interval (0,1) the distribution of W is absolutely continuous. Let $F_W(w)$ denote the cumulative distribution function (CDF) of W and let $p_0 = Pr\{W=0\}$, $p_1 = Pr\{W=1\}$, then

$$(2.5) \qquad F_W(w) = \begin{cases} p_0+(1-p_0-p_1)\int_0^w g_W(y)dy, & \text{for } 0\leq w<1 \\ 1, & \text{for } w\geq 1. \end{cases}$$

where $g_W(w)$ is the PDF of the absolutely continuous component of $F_W(w)$.

3. Visibility Probabilities

In the present section we develop formulae for the probabilities that rays connecting P_0 with n ($n\geq 1$) specified points on T are not intersected by any one of the N random disks. The methodology always follows the following algorithm:

S.1. Determine the set B(r) of (x,y) points such that, if a random disk of radius R=r has a center (x,y) in B(r) then at least one of the n specified rays is intersected by the disk.

S.2 Determine the probability of B(r), i.e.

$$Pr\{B(r)\} = \frac{1}{2\pi}\iint_{B(r)} \exp(-\tfrac{1}{2}(x^2+y^2))dxdy \quad .$$

S.3. Compute the total probability

$$H = \int_a^b \Pr\{B(r)\} dG_R(r).$$

S.4. Compute the simultaneous visibility probability of the n points.

$$\psi = (1-H)^N$$

3.1 The Visibility Probability of a Single Point on T.

Let $P = (x,y)$ be a point on T. Let $\overline{P_oP}$ be the line segment connecting P_o with P. The set $B(r)$ is the union of a rectangle $C_o(r)$ and two half circles $C_1(r)$ and $C_2(r)$ (see Figure 2). Two sides of $C_o(r)$ are parallel to $\overline{P_oP}$ and at distance r from it. The half circles $C_1(r)$ and $C_2(r)$ are each of radius r, and are attached to $C_o(r)$ at the sides perpendicular to $\overline{P_oP}$. In Appendix A we provide the formula for the computation of $\Pr\{B(r)\}$ in the case that $G_R(r)$ is a uniform distribution on $[0,b]$. Let $H_1(x_o,y_o,x,y)$ denote the H-function corresponding to the present case. This function provides the probability that a random disk intersects $\overline{P_oP}$. The visibility probability of the point P is

$$\psi_1(x_o,y_o,x,y) = (1-H_1(x_o,y_o,x,y))^N \quad . \tag{3.1}$$

Notice that for large values of N,

$$\psi_1(x_o,y_o,x,y) \cong \exp(-\Lambda_1(x_o,y_o,x,y)) \quad , \tag{3.2}$$

where $\Lambda_1(x_o,y_o,x,y) = NH_1(x_o,y_o,x,y)$.

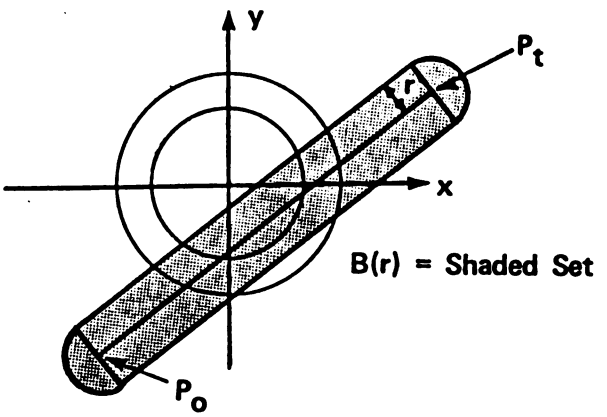


FIGURE 2. The Set B(r) for the Visibility Probability of a Single Point.

3.2 Simultaneous Visibility Probability of n=2 Points on T

Consider three points $P_o=(x_o,y_o)$, $P_1=(x_1,y_1)$ and $P_2=(x_2,y_2)$ in the plane. Let $T(x_o,y_o,x_1,y_1,x_2,y_2)$ denote the bivariate normal probability that a random vector (X,Y) belongs to the set inscribed by the triangle $\Delta P_oP_1P_2$. We adopt the convention that P_o is the vertex having the smallest y coordinate, and P_1,P_2 are reached from P_o in a counterclockwise direction. In Appendix B we provide the formula for determining $T(x_o,y_o,x_1,y_1,x_2,y_2)$. The set B(r) corresponding to the simultaneous visibility of two points, P_1 and P_2 on T is shown at Figure 3.

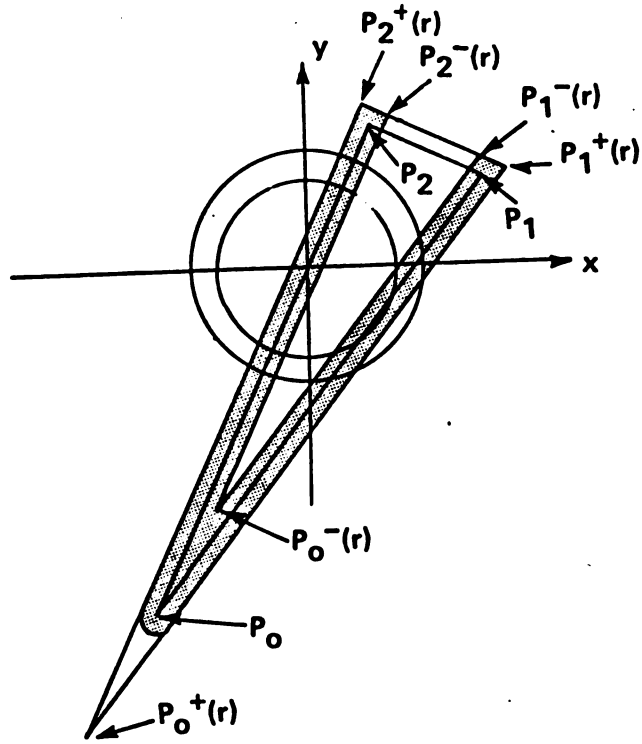


FIGURE 3. The Set B(r) for the Simultaneous Visibility of P_1 and P_2.

Let $P_i^+(r) = (x_i^+(r), y_i^+(r))$, i=0,1,2 , be the vertices of the triangle $\Delta P_o^+(r)P_1^+(r)P_2^+(r)$. The sides of this triangle are parallel to those of $\Delta P_oP_1P_2$ and at distance r from them. $\Delta P_o^+(r)P_1^+(r)P_2^+(r)$ is an outer-triangle of distance r. Similarly, we define an inner-triangle of distance r, $\Delta P_o^-(r)P_1^-(r)P_2^-(r)$. The inner triangle is contained in $\Delta P_oP_1P_2$, with sides parallel, respectively, to those of $\Delta P_oP_1P_2$ and at distance r.

The set B(r) corresponding to the points P_o, P_1 and P_2 is contained in the set $\Delta P_o^+(r)P_1^+(r)P_2^+(r) - \Delta P_o^-(r)P_1^-(r)P_2^-(r)$, which is the r-outer triangle minus the r-th inner triangle. If the distance of the point P_o from the origin is greater than 4 the probability of B(r) is given to a high degree of approximation by

$$\text{(3.3)} \qquad \Pr\{B(r)\} \doteq T(x_o^+(r), y_o^+(r), x_1^+(r), y_1^+(r), x_2^+(r), y_2^+(r)) - T(x_o^-(r), y_o^-(r), x_1^-(r), y_1^-(r), x_2^-(r), y_2^-(r)).$$

We remark here that in certain instances the inner triangle does not exist. In such cases we say that the inner triangle is an empty set, and its probability is zero. We further remark that in the many actual applications the source of light is more than 4 units away from the origin, and formula (3.3) provides excellent approximation. The conditions under which the inner triangle does not exist is given in formula (3.4) below. The H-function corresponding to (3.3) will be denoted by $H_2(x_o, y_o, x_1, y_1, x_2, y_2)$. This function is evaluated by numerical integration. For the determination of $x_i^+(r), y_i^+(r)$ $(i=0,1,2)$ and of $x_i^-(r), y_i^-(r)$ $(i=0,1,2)$, we distinguish between five cases:

Case I: $x_2 < x_o < x_1$;

Case II: $x_2 < x_1 = x_o$;

Case III: $x_2 < x_1 < x_o$

Case IV: $x_o = x_2 < x_1$;

Case V: $x_o < x_2 < x_1$.

Let $y=a_1+b_1x$ be the formula of the line passing through P_o and P_2, let $y=a_2+b_2x$ be the formula of the line passing through P_1 and P_2 and let $y=a_3+b_3x$ be the formula of the line through P_2 and P_o. The line passing through $P_o^+(r)$ and $P_1^+(r)$ has in Case I the formula $y=a_1-r(1+b_1^2)^{1/2}+b_1x$. Similarly, the line through $P_o^+(r)$ and $P_2^+(r)$ has in Case I the formula $y=a_3-r(1+b_3^2)^{1/2}+b_3x$. $x_o^+(r)$ is the x-coordinate of the point of intersection of these two lines. In this manner one can obtain explicit formulae for the coordinates $x_i^+(r)$, $y_i^+(r)$, $i=0,1,2$; and of $x_i^-(r),y_i^-(r)$. These formulae are listed in Appendix C. In all the above five cases, the condition under which the r-inner triangle is empty is

$$a_2+r(1+b_2^2)^{1/2}+b_2x_o^-(r)\leq y_o^-(r). \tag{3.4}$$

3.3 Probabilities of Simultaneous Visibility of n Points on T.

Let $P_i=(x_i,y_i)$, $i=1, \ldots, n$, be n points on T, such that $x_n<x_{n-1}< \ldots < x_1$. Generalizing the results of the previous section, the probability of B(r) is given approximately by

$$P_r\{B(r)\}\doteq T(x_o^+(r),y_o^+(r),x_1^+(r),y_1^+(r),x_n^+(r),y_n^+(r)) \tag{3.5}$$

$$- \sum_{i=1}^{n-1} T(x_{o,i}^-(r),y_{o,i}^-(r),x_i^-(r),y_i^-(r),x_{i+1}^-(r),y_{i+1}^-(r)).$$

Notice that some or all the r-inner triangles might be empty and the corresponding T-functions are then equal to zero. The condition under which the i-th inner triangle is empty is

$$(3.6) \qquad a_n + r(1+b_n^2)^{1/2} + b_n \bar{x}_{o,i}(r) \leq \bar{y}_{o,i}(r) \quad , \; i=1, \ldots, n-1.$$

a_n and b_n are the intercept and slope of the line passing through P_1 and P_n. The coordinates $\bar{x}_{o,i}(r)$ and $\bar{y}_{o,i}(r)$ can be obtained in the following manner. Consider the triangle $\Delta P_o P_1 P_{i+1}$. The coefficients a_2 and b_2 in the formulae of Appendix C should be replaced by a_n and b_n. a_3 and b_3 should be replaced by the intercept and slope of the line through P_o and P_{i+1}, i.e., a_{i+1} and b_{i+1}. Finally, the H-function corresponding to (3.5) is denoted by $H_n(x_o, y_o, x_1, y_1, \ldots, x_n, y_n)$. The probability that n points are simultaneously visible is

$$(3.7) \qquad \psi_n(x_o, y_o, x_1, y_1, \ldots, x_n, y_n) = (1 - H_n(x_o, y_o, \ldots, x_n, y_n))^N.$$

EXAMPLE 3.1

In the present example we illustrate some of the visibility probabilities. The parameters of the calculations are:

$N=20$	$x_o=0$	$x_t=0$
$\sigma_x=1$	$y_o=-33.3$	$y_t=-.333$
$\sigma_y=1$	$\Delta=.167$	$b=.333$

$G_R(r)$ is uniform on $[0,b]$.

We consider k=7 triplets on T, having coordinates $(x_{i,1},y_t)$, $(x_{i,2},y_t)$ and $(x_{i,3},y_t)$, where

$$x_{i,3} = -\Delta+i\Delta/4 \ , \ x_{i,2} = i\Delta/8 \text{ and } x_{i,1} = \Delta \ , \ i = 1,\ldots,7.$$

Let $$\psi_{1,i} = \psi_1(x_o,y_o,x_{i,3},y_t)$$

$$\psi_{2,i} = \psi_2(x_o,y_o,x_{i,1},y_t,x_{i,3},y_t)$$

$$\psi_{3,i} = \psi_3(x_o,y_o,x_{i,1},y_t,x_{i,2},y_t,x_{i,3},y_t).$$

The values of these visibility probabilities are:

i	$\psi_{1,i}$	$\psi_{2,i}$	$\psi_{3,i}$
1	.3080	.1233	.1108
2	.3064	.1360	.1262
3	.3055	.1516	.1439
4	.3052	.1692	.1642
5	.3055	.1927	.1885
6	.3064	.2201	.2186
7	.3080	.2539	.2534

Notice that $\psi_{1,i}$ is symmetric around the origin (i=4). The simultaneous visibilities of two and three points increase as the points become closer. Notice also that $\psi_{2,i}$ and $\psi_{3,i}$ are quite close, for each i, and that their difference diminishes as the distance between the three points decreases.

The probability p_1 that the line segment T is completely visible is determined in the following manner. Let

$$(3.8)\qquad H_2^*(x_o,y_o,x_R,y_R,x_L,y_L) = \int_a^b T(x_o^+(r),y_o^+(r),x_R^+(r),y_R^+(r),x_L^+(r),y_L^+(r))\,dG_R(r).$$

$H_2^*(x_o,y_o,x_R,y_R,x_L,y_L)$ is the probability that a random disk will cast a shadow on T. Hence, the probability that T is completely visible is

$$(3.9)\qquad p_1=(1-H_2^*(x_o,y_o,x_R,y_R,x_L,y_L))^N$$

4. The Moments of the Random Measure W.

The k-th moment of W is given according to (2.4) by the formula

$$(4.1)\qquad \mu_k = \frac{1}{A^k} E\{(\int_{-\Delta/\tau}^{\Delta/\tau} \phi(u) I(x_t+\tau u)\,du)^k\}$$

$$= \frac{k!}{A^k} \int\cdots\int_{S_k} \prod_{i=1}^{k} \phi(u_i) E\{\prod_{i=1}^{k} I(x_t+\tau u_i)\}\,du_i\,,$$

where $S_k=\{(u_1,\ldots,u_k);\ -\Delta/\tau \le u_k<u_{k-1}<\ldots<u_2<u_1\le\Delta/\tau\}$.

Moreover

$$(4.2)\qquad E\{\prod_{i=1}^{k} I(x_t+\tau u_i)\} = \psi_k(x_o,y_o,x_t+\tau u_1,y_1,\ldots,x_t+\tau u_k,y_k).$$

is the probability of simultaneous visibility of the k points.

Making the transformation $z_i = \frac{\tau}{\Delta} u_i$, i=1, ..., k, the k-th moment of W can be expressed as

$$(4.3) \qquad \mu_k = \frac{k!\Delta^k}{A^k\tau^k} \int_{-1}^{1} \phi(\tfrac{\Delta}{\tau} z_k) \int_{z_k}^{1} \phi(\tfrac{\Delta}{\tau} z_{k-1}) \int_{z_{k-1}}^{1} \cdots \int_{z_2}^{1} \phi(\tfrac{\Delta}{\tau} z_1) \cdot$$

$$\psi_k(x_o, y_o, x_t+\Delta z_1, y_1, \ldots, x_t+\Delta z_k, y_k)\, dz_1 \ldots dz_k$$

Notice that $y_i=\alpha+\beta\Delta z_i$ (i=1, ..., k), where α and β are the intercept and slope of the line segment T. In particular, the expected value of W is

$$(4.4) \qquad \mu_1 = \frac{\Delta}{A\tau} \int_{-1}^{1} \phi(\tfrac{\Delta}{\tau} z) \psi_1(x_o, y_o, x_t+\Delta z, y_t+\beta\Delta z) dz.$$

The second moment of W is

$$(4.5) \qquad \mu_2 = \frac{2\Delta^2}{A^2\tau^2} \int_{-1}^{1} \phi(\tfrac{\Delta}{\tau} z_2) \int_{z_2}^{1} \psi(\tfrac{\Delta}{\tau} z_1) \psi_2(x_o, y_o, x_t+\Delta z_1, y_t+\Delta\beta z_1,$$

$$x_t+\Delta z_2, y_t+\Delta\beta z_2) dz_1 dz_2 \ .$$

These integrals have to be computed numerically, as discussed in Section 6. The computation of a high order moment according to (4.3), with a high degree of accuracy, is time consuming. One can obtain, however, lower and upper bounds for $\mu_k (k \geq 3)$, in a manner which will be readily discussed. We precede this discussion with the following comment: As $k \to \infty$ μ_k approaches p_1. Indeed,

according to (2.5),

$$\mu_k = (1-p_o-p_1)\int_o^1 y^k g_W(y)dy + p_1. \tag{4.6}$$

Hence, by applying the Dominated Convergence Theorem, $\mu_k \to p_1$ as $k\to\infty$.

To obtain bounds for $\mu_k (k\geq 3)$ consider the inequality

$$\psi_2^*(x_o,y_o,x_1,y_1,x_k,y_k) \leq \tag{4.8}$$

$$\psi_k(x_o,y_o,x_1,y_1,x_2,y_2,\ldots,x_k,y_k) \leq$$

$$\psi_2(x_o,y_o,x_1,y_1,x_k,y_k) \quad ,$$

for every $k\geq 3$, where $\psi_2^*(x_o,y_o,x_1,y_1,x_k,y_k) = (1-H^*(x_o,y_o,x_1,y_1,x_k,y_k))^N$.

This follows directly from (3.5) and (3.8). We remind that $\psi_2^*(x_o,y_o,x_1,y_1,x_k,y_k)$ is the probability that the entire interval between (x_1,y_1) and (x_k,y_k) is visible. Substituting these lower and upper bounds of $\psi_k(x_o,y_o,x_1,y_1,\ldots,x_k,y_k)$ in (4.3) we obtain lower and upper bounds for μ_k. Thus, if $\mu_{k,L}$ and $\mu_{k,u}$ denote the lower and upper bound for μ_k one obtains, for every $k\geq 3$,

$$\mu_{k,L} = \frac{k(k-1)\Delta^2}{A^k \tau^2} \int_{-1}^{1} \phi(\tfrac{\Delta}{\tau}z_k) \int_{z_k}^{1} \phi(\tfrac{\Delta}{\tau}z_1) \tag{4.9}$$

$$\psi_2^*(x_o,y_o,x_t+\Delta z_1,y_t+\Delta\beta z_1,x_t+\Delta z_k,y_t+\Delta\beta z_k).$$

$$[\Phi(\tfrac{\Delta}{\tau}z_1)-\Phi(\tfrac{\Delta}{\tau}z_k)]^{k-2}dz_1dz_k \quad ,$$

where $\Phi(\cdot)$ is the standard normal integral. $\mu_{k,u}$ is obtained from (4.9) by replacing the function ψ_2^* by ψ_2. The proof that substitution of ψ_2^* in (4.5) yields (4.9) can proceed by induction. Indeed, for k=3 we obtain from (4.3)

$$(4.10) \qquad \mu_{3,L} = \frac{3!\Delta^3}{A^3\tau^3} \int_{-1}^{1} \phi(\tfrac{\Delta}{\tau}z_3) \int_{z_3}^{1} \phi(\tfrac{\Delta}{\tau}z_2) \int_{z_2}^{1} \phi(\tfrac{\Delta}{\tau}z_1)\, F(z_1,z_3)dz_1dz_2dz_3$$

where $F(z_1,z_3) = \psi_2^*(x_o,y_o,x_t+\Delta z_1,y_t+\Delta\beta z_1,x_t+\Delta z_3,y_t+\Delta\beta z_3)$

Furthermore

$$(4.11) \qquad \frac{\Delta}{\tau} \int_{z_3}^{1} \phi(\tfrac{\Delta}{\tau} z_2) \int_{z_2}^{1} \phi(\tfrac{\Delta}{\tau}z_1)F(z_1,z_3)dz_1dz_2$$

$$= \frac{\Delta}{\tau} \int_{z_3}^{1} \phi(\tfrac{\Delta}{\tau} z_1)F(z_1,z_3) \int_{z_2}^{z_1} \phi(\tfrac{\Delta}{\tau} z_2)dz_2dz_1$$

$$= \int_{z_3}^{1} \phi(\tfrac{\Delta}{\tau} z_1)F(z_1,z_3)\, [\Phi(\tfrac{\Delta}{\tau} z_1)-\Phi(\tfrac{\Delta}{\tau} z_3)]dz_1$$

Substitution of (4.11) in (4.10) yields (4.9) for k=3. For $k\geq 4$ one obtains from the induction hypothesis,

$$(4.12) \qquad \frac{\Delta^j}{\tau^j} \int_{z_{j+2}}^{1} \phi(\tfrac{\Delta}{\tau}z_{j+1}) \int_{z_j}^{1} \phi(\tfrac{\Delta}{\tau}z_j) \int_{z_{j-1}}^{1} \cdots \int_{z_2}^{1} \phi(\tfrac{\Delta}{\tau}z_1)F(z_1,z_k)dz_1\ldots dz_{j+1}$$

$$= \frac{1}{j!} \int_{z_{j+2}}^{1} \phi(\tfrac{\Delta}{\tau} z_1)F(z_1,z_k)\, [\Phi(\tfrac{\Delta}{\tau}z_1)-\Phi(\tfrac{\Delta}{\tau}z_{j+2})]^j \quad ,$$

for all $j=1,2,\ldots,\ell$, $\ell\leq k-3$, that

$$(4.13)\qquad \frac{\Delta^{\ell+1}}{\tau^{\ell+1}} \int_{z_{\ell+3}}^{1} \phi(\tfrac{\Delta}{\tau}z_{\ell+2}) \int_{z_{\ell+2}}^{1} \phi(\tfrac{\Delta}{\tau}z_{\ell+1}) \int_{z_{\ell+1}}^{1} \cdots \int_{z_2}^{1} \phi(\tfrac{\Delta}{\tau}z_1) F(z_1,z_k)\, dz_1 \cdots dz_{\ell+2}$$

$$= \frac{\Delta}{\ell!\tau} \int_{z_{\ell+3}}^{1} \phi(\tfrac{\Delta}{\tau}z_{\ell+2}) \int_{z_{\ell+2}}^{1} \phi(\tfrac{\Delta}{\tau}z_1) F(z_1,z_k) [\Phi(\tfrac{\Delta}{\tau}z_1) - \Phi(\tfrac{\Delta}{\tau}z_{\ell+2})]^{\ell} dz_1 dz_{\ell+2}$$

$$= \frac{\Delta}{\ell!\tau} \int_{z_{\ell+3}}^{1} \phi(\tfrac{\Delta}{\tau}z_1) F(z_1,z_k) \int_{z_{\ell+3}}^{z_1} \phi(\tfrac{\Delta}{\tau}z_{\ell+2}) [\Phi(\tfrac{\Delta}{\tau}z_1) - \Phi(\tfrac{\Delta}{\tau}z_{\ell+2})]^{\ell} dz_{\ell+2} dz_1$$

But,

$$(4.14)\qquad \frac{\Delta}{\tau} \int_{z_{\ell+3}}^{z_1} \phi(\tfrac{\Delta}{\tau}z_{\ell+2}) [\Phi(\tfrac{\Delta}{\tau}z_1) - \Phi(\tfrac{\Delta}{\tau}z_{\ell+2})]^{\ell} dz_{\ell+2}$$

$$= \frac{1}{\ell+1} [\Phi(\tfrac{\Delta}{\tau}z_1) - \Phi(\tfrac{\Delta}{\tau}z_{\ell+3})]^{\ell+1}$$

Substituting (4.14) in the right hand side of (4.13) we obtain

$$(4.15)\qquad \frac{\Delta^{\ell+1}}{\tau^{\ell+1}} \int_{z_{\ell+3}}^{1} \phi(\tfrac{\Delta}{\tau}z_{\ell+2}) \int_{z_{\ell+2}}^{1} \cdots \int_{z_2}^{1} \phi(\tfrac{\Delta}{\tau}z_1) F(z_1,z_k)\, dz_1 dz_2 \cdots dz_{\ell+2}$$

$$= \frac{1}{\ell+1} \int_{z_{\ell+3}}^{1} \phi(\tfrac{\Delta}{\tau}z_1) F(Z_1,Z_k) [\Phi(\tfrac{\Delta}{\tau}z_1) - \Phi(\tfrac{\Delta}{\tau}z_{\ell+3})]^{\ell+1} dz_1$$

In particular, for $\ell=k-3$, (4.3) and (4.15) yield (4.9).

Example 4.1

In the present example we illustrate numerically the bounds for the first 10 moments in a given case. The first two moments are computed exactly according to formulae (4.4) and (4.5). The

bounds to the moments of order $k \geq 3$ were computed according to (4.9). The parameters of the computation are:

$$\sigma_x = \sigma_y = 1$$

$$x_o = -.333$$

$$y_o = -100.$$

$$x_t = 0$$

$$y_t = -.333$$

$$\Delta = .167$$

$$\tau = \Delta/3$$

and $a = 0,\ b = .333.$

The first two moments are $\mu_1 = .3057$ and $\mu_2 = .2375$. The lower and upper bounds for μ_k, k=3, ..., 10 are given in the following table.

k	3	4	5	6	7	8	9	10
Lower	.2099	.1950	.1847	.1771	.1711	.1662	.1622	.1588
Upper	.2147	.2014	.1924	.1858	.1806	.1765	.1731	.1703

TABLE 2. Lower and Upper Bounds for μ_k, k=3, ..., 10.

A good estimate of μ_k can be obtained as the midpoint between the lower and upper bounds. The limit of μ_k as $k \to \infty$ is $p_1 = .0855$.

5. Approximating the Distribution of W and Determining p_o.

The CDF of W is a mixture of a discrete and absolutely continuous component, as presented in formula (2.5). The only element of that formula which we have already discussed is p_1, as specified in formula (3.9). We provide here an approximation for p_o and for the absolutely continuous component of $F_W(w)$, which is called the beta-mixture distribution.

The beta CDF is represented by the incomplete beta function ratio

$$(5.1) \qquad I_x(\nu_1,\nu_2) = \frac{1}{B(\nu_1,\nu_2)} \int_o^x u^{\nu_1-1}(1-u)^{\nu_2-1}\,du,$$

where $0<\nu_1,\nu_2<\infty$, and $B(\nu_1,\nu_2)$ is the complete beta integral. This two-parameter family of distributions can represent a variety of different shapes and forms of distributions on the interval (0,1). We, therefore, approximate $F_W(w)$ by the beta-mixture.

$$(5.2) \qquad F_W(w) \doteq p_o + (1-p_o-p_1)I_w(\nu_1,\nu_2) \quad , \ 0\le w<1$$

and, obviously $F_W(w) = 1$ for $x\ge 1$. The parameters p_o, ν_1 and ν_2 of (5.2) are determined by equating the first three moments of (5.2) to those of W. The solution of these moment equations (see [3]) is

$$\nu_1 = (2\mu_2'-\mu_3'(\mu_1'+\mu_2'))/\text{Det.}$$

$$(5.3) \qquad \nu_2 = (\mu_1'-\mu_2')(\mu_2'-\mu_3')/\text{Det.}$$

and

$$p_o = 1-p_1-\mu_1'(\nu_1+\nu_2)/\nu_1 \quad ,$$

where $\mu_i'=\mu_i-p_1$ $(i=1,2,3)$ and $\text{Det.} = \mu_1'\mu_3'-(\mu_2')^2$.

Example 5.1

In Table 3 we provide a numerical illustration of the beta-mixture approximation, under the parameters $\sigma_x=\sigma_y=1$, $x_o=-.333$, $y_o=-100.$, $x_t=0,.333,.666,\ 1.000,\ 1.333$, $y_t=-.333$, $\Delta=.167$, $\tau=\Delta/5$ and $a=0$, $b=.333$.

x_t	μ_1	μ_2	μ_3	p_o	p_1	ν_1	ν_2
.000	.3057	.2375	.2126	.6928	.0855	53.3	.357
.333	.3257	.2570	.2314	.6728	.0980	52.9	.353
.667	.3864	.3170	.2898	.6120	.1404	52.2	.340
1.000	.4854	.4190	.3914	.5131	.2261	55.1	.324
1.333	.6105	.5538	.5287	.3884	.3633	67.6	.312

TABLE 3. The First Three Moments and the Parameters of the Beta-Mixture Distributions

The beta distributions with $\nu_1>1$ and $0<\nu_2<1$ have PDF's which are monotonically increasing from 0 to ∞. They are called J-shaped distributions. In the cases of Table 3, since ν_1 is very large, most of the mass of the absolutely continuous component is concentrated near the value W=1. This means that the distribution of W is almost a 2-point distribution. For example, $x_t=0$, W=0 with probability $p_o=.693$ and with probability .307 $W\geq.95$.

The values of (5.2) in the case of $x_t=0$ are given in Table 4.

W	0	.90	.950	.955	.975	.985	.995	1.000
$F_W(w)$	.693	.693	.694	.698	.707	.719	.744	1.000

TABLE 4. The Beta-Mixture Approximation to $F_W(w)$, in the Case of $x_t=0$.

6. Numerical Computations

The numerical computations of Tables 1-4 were performed on an IBM-370 system. The programs were written in FORTRAN. They consist of the main parts and the following subroutine functions.

1. FUNCTION VPT (Xϕ,Yϕ,XT,YT,B)
2. FUNCTION TRL (Xϕ,Yϕ,X1,Y1,X2,Y2)
3. FUNCTION SVP (K,Xϕ,Yϕ,X1,Y1,X2,Y2,X3,Y3,B)
4. FUNCTION TVP(Xϕ,Yϕ,X1,Y1,X2,Y2,B)
5. FUNCTION CNDX(Y)
6. FUNCTION BETA(X,V1,V2)

and

7. FUNCTION GAMA(V).

Subroutine functions 5-7 compute the standard normal CDF, the beta CDF and the gamma function, respectively. Subroutine VPT(Xϕ,Yϕ,XT,YT,B) computes the visibility probability of the point (XT,YT) from the source (Xϕ,Yϕ), where the distribution of the disks radii, R, is uniform (rectangular) on the interval (0,B). The subroutine TRL (Xϕ,Yϕ,X1,Y1,X2,Y2) computes the standard bivariate normal probability that a point (X,Y) falls within the triangle with vertices

$P_o=(X\phi,Y\phi)$, $P_1=(X1,Y1)$, $P_2=(X_2Y_2)$. P_o is always the point with the smallest y-coordinate. The vertices P_1 and P_2 are reached from P_o in a counter clockwise direction. The subroutine SVP(K,Xϕ,Yϕ,X1,Y1,X2,Y2,X3,Y3,B) computes the probabilities H_2 or H_3 for the simultaneous visibility of k=2 or k=3 points (see formula (3.7)). K=2 or K=3 is the argument for the number of points. If K=2, the only arguments which are taken into account are Xϕ,Yϕ,X1,Y1,X2,Y2 and B. The values of X3 and Y3 can be arbitrary. The points $P_o=(X\phi,Y\phi)$, $P_1=(X1,Y1)$ and $P_2=(X2,Y2)$ are the vertices of the triangle, while $P_3=(X_3,Y_3)$ is a point in the line segment connecting P_1 and P_2. The function TVP(Xϕ,Yϕ,X1,Y1,X2,Y2,B) computes the probability $H_2^*(x_o,y_o,x_1,y_1,x_2,y_2)$. The subroutine functions VPT and TRL were programmed for the standard bivariate normal model with zero correlation, i.e., a bivariate normal distribution centered at the origin, with $\sigma_x=\sigma_y=1$ and $\rho=0$. In actual problems the variances of X and Y are generally unequal. In such cases we make the additional transformation $x\to x/\sigma_x$ and $y\to y/\sigma_y$. This reduces the model to a standard bivariate normal. The random disks are transformed to random ellipses. However, we replace each such random ellipse by a random disk having a radius equal to the large radius of the corresponding ellipse.

The moments μ_1, μ_2 and the lower and upper bounds for $\mu_k (k\geq 3)$ are computed according to formula (4.9) by employing the subroutine functions TVP and SVP and using a Gauss quadrature 12-point numerical integration (see Abramowitz and Stegun [1] pp.916). The FORTRAN programs of these subroutine functions are given in Appendix D.

The main program reads the parameters of the problem and reduces it to the standard bivariate normal model under which all the computations are performed. For example, suppose that a source of a laser beam is located at the point (-1, -3000) and a target of width 1 is centered at (0,-10). N=20 disks are scattered according to a bivariate normal distribution centered at (0,0) with parameters $\sigma_x=3$, $\sigma_y=30$ and $\rho=0$. The radii of disks are uniformly distributed in the interval (0,1). The transformation $x'=x/\sigma_x$, $y'=y/\sigma_y$, $b'=b/\sigma_x$, $w'=w/\sigma_x$ reduces the above parameters to those of Table 3.

7. References

[1] Abramowitz, M. and I. A. Stegun
Handbook of Mathematical Functions (5th Edition)
Dover Publication, New York, 1968

[2] Didonato, A. R. and R. K. Hageman (1982)
A method for computing the integral of the bivariate normal distribution over an arbitrary polygon.
SIAM J. Sci. Stat. Comput., 3:434-446.

[3] Yadin, M. and S. Zacks (1982)
Random coverage of a circle with application to a shadowing problem, J. Appl. Prob., 19:562-577

[4] Zacks, S. and Yadin, M. (1984)
The Distribution of the Random Lighted Portion of a Curve in a Plane Shadowed by a Poisson Random Field of Obstacles.
Statistical Signal Processing, E. J. Wegman and J. G. Smith Eds, Marcel Dekker, New York, pp. 273-286.

APPENDIX A. The Bivariate Normal Probability of a Rectangle with Two Half Circles and The Probability that a Random Disk Intersects a Line Segment

We develop here formulae for the computation of the standard bivariate normal probability of the set B(r) described in Figure 2. The set B(r) is the union of a rectangle $C_o(r)$ and two half circles $C_1(r)$ and $C_2(r)$, having a radius r. The rectangle $C_o(r)$ has two sides which are parallel to the line segment $\overline{P_oP_t}$, where $P_o=(x_o,y_o)$ and $P_t=(x_t,y_t)$. Moreover, these sides are of distance r from $\overline{P_oP_t}$. We provide now a formula for $P_r\{C_o(r)\}$. We distinguish between three cases:

Case I: $x_t<x_o$,

Case II: $x_t=x_o$,

Case III: $x_t>x_o$.

In Case II we have

$$(A.1) \qquad P_r\{C_o(r)\} = [\Phi(x_t+r)-\Phi(x_t-r)][\Phi(y_t)-\Phi(y_o)],$$

where $\Phi(z)$ is the standard normal integral. In Cases I and III we consider the orthogonal matrix

$$(A.2) \qquad A = \frac{1}{(1+B^2)^{1/2}} \begin{bmatrix} B & -1 \\ 1 & B \end{bmatrix} ,$$

where $B=(y_t-y_o)/(x_t-x_o)$ is the slope of the line segment $\overline{P_oP_t}$. It is easy to verify that the orthogonal transformation

$$\begin{pmatrix} y' \\ x' \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix}$$

yields points P_o' and P_t' having the same x-coordinate. Furthermore $y_t'>y_o'$ if $x_t>x_o$ and $y_t'<y_o'$ if $x_t<x_o$. Accordingly,

$$\text{(A.3)} \qquad \Pr\{C_o(r)\} = \begin{cases} [\Phi(x_t'+r)-\Phi(x_t'-r)][\Phi(y_o')-\Phi(y_t')] & \text{, in Case I} \\ [\Phi(x_t'+r)-\Phi(x_t'-r)][\Phi(y_t')-\Phi(y_o')] & \text{, in Case III.} \end{cases}$$

Indeed, the distribution of (X',Y') is a standard bivariate normal distribution, with correlation $\rho=0$. The half circle $C_1(r)$ is centered at (x_t',y_t') and $C_2(r)$ is centered at (x_o',y_o'). Consider Case III, in which $y_t'>y_o'$. The probability of $C_1(r)$ is

$$\text{(A.4)} \qquad \Pr\{C_1(r)\}=\Pr\{|X'-x_t'|\le r, y_t' \le Y' \le y_t' + (r^2-(X'-x_t')^2)^{1/2}\}$$

$$= \int_{x_t-r}^{x_t+r} \phi(x)\ \Phi(y_t'+(r^2-(x-x_t')^2)^{1/2})dx -$$

$$\Phi[y_t')[\Phi(x_t'+r)-\Phi(x_t'-r)].$$

Similarly,

$$\text{(A.5)} \qquad \Pr\{C_2(r)\} = \Phi(y_o')[\Phi(x_t'+r)-\Phi(x_t'-r)] - \int_{x_t-r}^{x_t+r} \phi(x)\Phi(y_o'-(r^2-(x-x_t')^2)^{1/2}dx \quad .$$

From (A.3)-(A.5) one obtains in Case III,

$$\text{(A.6)} \qquad \Pr\{B(r)\} = \sum_{i=0}^{2} \Pr\{C_i(r)\}$$

$$= \int_{x_t-r}^{x_t+r} \phi(x)\,[\Phi(y'_t+(r^2-(x-x'_t)^2)^{1/2})$$

$$- \Phi(y'_o-(r^2-(x-x'_t)^2)^{1/2})]dx.$$

Formula (A.6) applies also for Case II, by sutstituting x_t for x'_t and y_o, y_t for y'_o, y'_t. In Case I we interchange y'_o and y'_t in (A.6). We remark also that if $y'_o < -4$, the term $\Phi(y_o-(r^2-(x-x_t)^2)^{1/2})$ can be dropped from formula (A.6).

Finally, the probability that a random disk intersects the line segment $\overline{P_oP_t}$ is, in Cases II and III,

$$\text{(A.7)} \qquad H_1(x_o,y_o,x_t,y_t) =$$

$$\frac{1}{b}\int_o^b \int_{x_t-r}^{x_t+r} \phi(x)\,[\Phi(y'_t+(r^2-(x-x'_t)^2)^{1/2})-\Phi(y'_t-(r^2-(x-x'_t)^2)^{1/2})]dxdr$$

$$= b\int_o^1 u \int_{-1}^{1} \phi(x'_t+buz)\,[\Phi(y'_t+bu(1-z^2)^{1/2})-\Phi(y'_o-bu(1-z^2)^{1/2})]\;dzdu$$

The double integral in (A.7) was evaluated by a double 8-point Gaussian integration (see Abramowitz and Stegun [1; pp.916]. Attempts at 20-point integration yielded negligible difference.

APPENDIX B. The Bivariate Normal Probability of Triangles, $T(x_o,y_o,x_1,y_1,x_2,y_2)$.

Consider the triangle $\Delta P_o P_1 P_2$, where $P_i=(x_i,y_i)$, i=0,1,2. We make first an orthogonal transformation of the plane (x,y) so that the side $\overline{P_1'P_2'}$ of the triangle $\Delta P_o'P_1'P_2'$ is parallel to the x-axis and $y_o'=\min_{o\leq i\leq 2}(y_i')$. The orthogonal matrix applied for this purpose is:

$$A = \frac{1}{(1+B^2)^{1/2}} \begin{bmatrix} 1 & B \\ -B & 1 \end{bmatrix} , \text{ if } x_1 \neq x_2 , \tag{B.1}$$

where $B=(y_2-y_1)/(x_2-x_1)$. If $x_1=x_2$ we use the transformation

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} , \text{ if } x_o < x_1$$

(B.2)

$$A = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} , \text{ if } x_o > x_1$$

Let (X',Y') = (X,Y)A'. The distribution of (x',y') is like that of (X,Y).

Let $T(x_o',y_o',x_1',y_1',x_2',y_2')$ be the standard bivariate normal probability ($\rho=0$) of the triangle $\Delta P_o'P_1'P_2'$. Notice that $y_1'=y_2'$, and $y_o'<y_1'$

Furthermore, $T(x_o,y_o,x_1,y_1,x_2,y_2) = T(x_o',y_o',x_1',y_1',x_2',y_2')$.

We distinguish between five cases.

Case 1: $x_2'<x_o'<x_1'$:

Let $y'=a_2+b_2x'$ be the line connecting P_o' with P_2' and let $y'=a_1+b_1x'$ be the line connecting P_o' with P_1'. The probability of $\Delta P_o'P_1'P_2'$ is then

(B.3)
$$T(x_o,y_o,x_1,y_1,x_2,y_2) =$$
$$\Pr\{x_2'\le X'\le x_o' \ , \ a_2+b_2X'\le Y'\le y_1'\} +$$
$$\Pr\{x_o'\le X'\le x_1' \ , \ a_1+b_1X'\le Y'\le y_1'\}$$
$$= \Phi(y_1')[\Phi(x_1')-\Phi(x_2')] - \int_{x_2'}^{x_o'}\phi(x)\Phi(a_2+b_2x)dx - \int_{x_o'}^{x_1'}\phi(x)\Phi(a_1+b_1x)\,dx .$$

The integral in (B.3) was evaluated numerically according to the formula

(B.4)
$$\int_{\xi_o}^{\xi_N}\phi(x)\Phi(a+bx)dx \cong \sum_{i=1}^{N}\Phi(a+b\tilde{\xi}_i)[\Phi(\xi_i)-\Phi(\xi_{i-1})],$$

where $N=20$, ξ_i are the endpoint of equal subinterval partition of (ξ_o,ξ_N) and $\tilde{\xi}_i=(\xi_i+\xi_{i+1})/2$.

Case 2: $x_2' = x_o'$

As in Case 1, let $y = a_1 + b_1 x$ be the line connecting P_o with P_1. Then,

(B.5) $$T(x_o, y_o, x_1, y_1, x_2, y_2) =$$

$$\Pr\{x_o' \leq X_1' \leq x_1' \ , \ a_1 + b_1 X' \leq Y' \leq y_1'\} =$$

$$\Phi(y_1')[\Phi(x_1') - \Phi(x_o')] - \int_{x_o'}^{x_1'} \phi(x)\Phi(a_1 + b_1 x)dx.$$

Case 3: $x_2' > x_o'$.

Let $y = a_2 + b_2 x$ denote the line connecting P_o with P_2 and $y = a_1 + b_1 x$ denote the line connecting P_o with P_1. Then

(B.6) $$T(x_o, y_o, x_1, y_1, x_2, y_2) = \Phi(y_1')[\Phi(x_1') - \Phi(x_2')]$$

$$+ \int_{x_o'}^{x_2'} \phi(x)\Phi(a_2 + b_2 x)dx - \int_{x_o'}^{x_1'} \phi(x)\Phi(a_1 + b_1 x)dx.$$

Case 4: $x_1' = x_o'$

As before, let $y = a_2 + b_2 x$ designate the line connecting P_o with P_1, then

(B.7) $$T(x_o, y_o, x_1, y_1, x_2, y_2) = \Phi(y_1')[\Phi(x_o') - \Phi(x_2')]$$

$$- \int_{x_2'}^{x_o'} \phi(x)\Phi(a_2 + b_2 x)dx.$$

MORE DIRECTIONS

AFTER YOU HAVE COMPLETED THIS STUDY GUIDE THE FOLLOWING WILL TAKE PLACE.

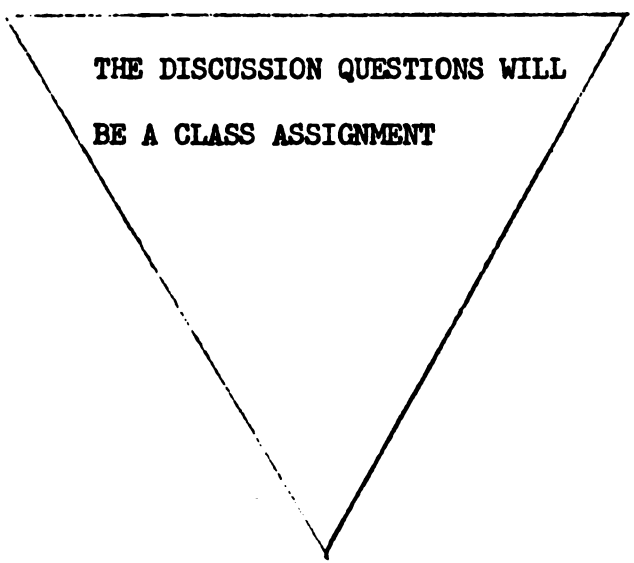
(1) YOUR TEACHER WILL GO OVER THE ANSWERS WITH YOU. MAKE SURE YOU UNDERSTAND THE QUESTIONS AND ANSWERS AFTER CLASS DISCUSSION.

(2) YOUR TEACHER WILL GIVE YOU ADDITIONAL INFORMATION THROUGH LECTURES AND OTHER WRITTEN RESOURCES.

STUDY AND LEARN

REVIEW: STUDY BY READING THE QUESTIONS. . . . .DO NOT LOOK AT THE ANSWERS. . . . .TRY TO <u>RECITE</u> THE ANSWERS. . . . .CHECK TO SEE <u>HOW</u> MUCH YOU CAN REMEMBER. . . . .

DO: THE STUDY EXERCISE AND THE VOCABULARY REVIEW TO CHECK YOUR UNDERSTANDING OF THIS GUIDE.



50

$$y_1^-(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_1^-(r) \quad ,$$

$$x_2^+(r) = \frac{a_2-a_3}{b_3-b_2} + \frac{r}{b_3-b_2}\left[(1+b_3^2)^{1/2}+(1+b_2^2)^{1/2}\right] \quad ,$$

$$y_2^+(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_2^+(r) \quad ,$$

$$x_2^-(r) = 2x_2-x_2^+(r) \quad ,$$

$$y_2^-(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_2^-(r)$$

Case 2: $x_o=x_1$

$$x_o^+(r) = x_o+r \quad ,$$

$$y_o^+(r) = a_3-r(1+b_3^2)^{1/2}+b_3x_o^+(r) \quad ,$$

$$x_o^-(r) = x_o-r \quad ,$$

$$y_o^-(r) = a_3+r(1+b_3^2)^{1/2}+b_3x_o^-(r) \quad ,$$

$$x_1^+(r) = x+r$$

$$y_1^+(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_1^+(r) \quad ,$$

$$x_1^-(r) = x-r$$

$$y_1^-(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_1^-(r) \quad ,$$

$$x_2^+(r) = \frac{a_2-a_3}{b_3-b_2} + \frac{r}{b_3-b_2}\left[(1+b_3^2)^{1/2}+(1+b_2^2)^{1/2}\right] ,$$

$$y_2^+(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_2^+(r) ,$$

$$x_2^-(r) = 2x_2-x_2^+(r) ,$$

$$y_2^-(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_2^-(r) .$$

Case 3: $x_o > x_1$

$$x_o^+(r) = \frac{a_1-a_3}{b_3-b_1} + \frac{r}{b_3-b_1}\left[(1+b_3^2)^{1/2}+(1+b_1^2)^{1/2}\right] ,$$

$$y_o^+(r) = a_1+r(1+b_1^2)^{1/2}+b_1x_o^+(r) ,$$

$$x_o^-(r) = 2x_o-x_o^+(r) ,$$

$$y_o^-(r) = a_1+r(1+b_1^2)^{1/2}+b_1x_o^-(r) ,$$

$$x_1^+(r) = \frac{a_2-a_1}{b_1-b_2} + \frac{r}{b_1-b_2}\left[(1+b_2^2)^{1/2}-(1+b_1^2)^{1/2}\right] ,$$

$$y_1^+(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_1^+(r) ,$$

$$x_1^-(r) = 2x_1-x_1^+(r) ,$$

$$y_1^-(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_1^-(r) ,$$

$$x_2^+(r) = \frac{a_3-a_2}{b_2-b_3} - \frac{r}{b_2-b_3}\left[(1+b_2^2)^{1/2}+(1+b_3^2)^{1/2}\right] ,$$

$$y_2^+(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_2^+(r) \quad ,$$

$$x_2^-(r) = 2x_2-x_2^+(r) \quad ,$$

$$y_2^-(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_2^-(r) \quad .$$

Case 4: $x_o=x_2$.

$$x_o^+(r) = x_o-r$$

$$y_o^+(r) = a_1-r(1+b_1^2)^{1/2}+b_1x_o^+(r) \quad ,$$

$$x_p^-(r) = x_o+r$$

$$y_o^-(r) = a_1-r(1+b_1^2)^{1/2}+b_1x_o^-(r) \quad ,$$

$$x_1^+(r) = \frac{a_1-a_2}{b_2-b_1} - \frac{r}{b_2-b_1}\left[(1+b_1^2)^{1/2}+(1+b_2^2)^{1/2}\right] \quad ,$$

$$y_1^+(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_1^+(r) \quad ,$$

$$x_1^-(r) = 2x_1-x_1^+(r) \quad ,$$

$$y_1^-(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_1^-(r) \quad ,$$

$$x_2^+(r) = x_2-r \quad ,$$

$$y_2^+(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_2^+(r) \quad ,$$

$$x_2^-(r) = x_2+r \quad ,$$

$$y_2^-(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_2^-(r) \quad .$$

Case 5: $x_o<x_2$

$$x_o^+(r) = \frac{a_3-a_1}{b_1-b_3} + \frac{r}{b_1-b_3}\left[(1+b_1^2)^{1/2}+(1+b_3^2)^{1/2}\right] \quad ,$$

$$y_o^+(r) = a_1-r(1+b_1^2)^{1/2}+b_1x_o^+(r) \quad ,$$

$$x_o^-(r) = 2x_o - x_o^+(r) \quad ,$$

$$y_o^-(r) = a_1-r(1+b_1^2)^{1/2}+b_1x_o^-(r) \quad ,$$

$$x_1^+(r) = \frac{a_1-a_2}{b_2-b_1} - \frac{r}{b_2-b_1}\left[(1+b_i^2)^{1/2}+(1+b_2^2)^{1/2}\right] \quad ,$$

$$y_1^+(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_1^+(r) \quad ,$$

$$x_1^-(r) = 2x_1-x_1^+(r) \quad ,$$

$$y_1^-(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_1^-(r) \quad ,$$

$$x_2^+(r) = \frac{a_3-a_2}{b_2-b_3} + \frac{r}{b_2-b_3}\left[(1+b_3^2)^{1/2}-(1+b_2^2)^{1/2}\right] \quad ,$$

$$x_2^-(r) = 2x_2-x_2^+(r) \quad ,$$

$$y_2^-(r) = a_2+r(1+b_2^2)^{1/2}+b_2x_2^-(r) \quad .$$

ON SEGMENTATION OF SIGNALS, TIME SERIES, AND IMAGES

Stanley L. Sclove

Department of Quantitative Methods
University of Illinois at Chicago

ABSTRACT

Signals and time series often are not homogeneous but rather are generated by mechanisms or processes with various phases. Similarly, images are not homogeneous but contain various objects. "Segmentation" is a process of attempting to recover automatically the phases or objects. A model for representing such signals, time series, and images is discussed. Some approaches to estimation and segmentation in this model are presented.

Key words and phrases: statistical pattern recognition, classification; temporal correlation, spatial correlation; optimization by relaxation method.

ON SEGMENTATION OF SIGNALS, TIME SERIES, AND IMAGES

Stanley L. Sclove

Department of Quantitative Methods
College of Business Administration
University of Illinois at Chicago

1. Introduction

About ten years ago, Professor James Osterburg, a colleague at the University of Illinois at Chicago and an expert on physical evidence (Osterburg and O'Hara, 1949; Osterburg, 1968, 1982), consulted me in regard to establishing probability estimates for partial fingerprints. We have had a very interesting collaboration which resulted in several papers (Osterburg, Parthasarathy, Raghavan, and Sclove, 1979; Sclove, 1979, 1980, 1981a) and may, thanks to Professor Osterburg's continuing efforts, result in changes in practices regarding the evaluation of fingerprints as evidence. For the present paper the point about fingerprints is this: When at professional meetings I would talk about the subject, certain people (usually electrical engineers or computer scientists) would tell me that what I was doing was "image processing."

The reason that our fingerprint work resembles image processing is this. Osterburg treated fingerprint analysis by placing a grid of cells over the print. One then notes the locations of any occurrences of the "Galton details," the minutiae of the ridge lines, such as ridge endings and forks. In numerical image processing a real image is divided into cells ("pixels": picture elements) and one notes numerically what occurs in each cell. The real image is expressed as a matrix -- rows by columns -- of cells, just as the TV screen has a matrix of dots which are illuminated with various colors and intensities.

2. Images

An important aspect of image processing is image "segmentation," the location of objects in the image.

Examples. (i) A picture of a house and yard is to be labeled using the labels, brick, glass, tree, grass, sky. (ii) An image is to be labeled using the labels, tank, mud, tree, sky. (iii) A medical image is to be labeled using the labels, tumor, or normal tissue.

Segmentation involves labeling each pixel according to the name of the object of which it is a part, in turn involves the grouping of neighboring pixels. Since I had earlier worked on cluster analysis, the grouping of observations (Sclove 1977), it was natural to concentrate on segmentation. The labeling process can be made explicit in a model which states that at each pixel we observe the value of a random variable X, but also along with X there is an unobservable variable, the label. (I tend to treat the labels as parameters. Others treat them as missing data, i.e., random variables.) In the context of this model, segmentation is merely estimation of the labeling parameters.

The random variable X is often a vector of several measurements.

Examples. (i) A familiar example of a vector of measurements (though it would not be measured across a two-dimensional array) is blood pressure, which is a vector of the two measurements, systolic and diastolic. (ii) In color television, X is a vector of three measurements, the red level, green level, and blue level. (iii) In Landsat satellite data, there are four spectral channels, one in the green/yellow visible range, a second in the red visible range, and the other two in the near infrared range. (iv) For a black-and-white image, the random variable X which is simply a scalar consisting of a single gray-level measurement, rather than a vector.

Images are two-dimensional; I decided first to consider a one-dimensional version of the problem. An image is a two-way series; a one-way series is a "time series." So I began this research by thinking about segmentation of time series.

3. Time Series

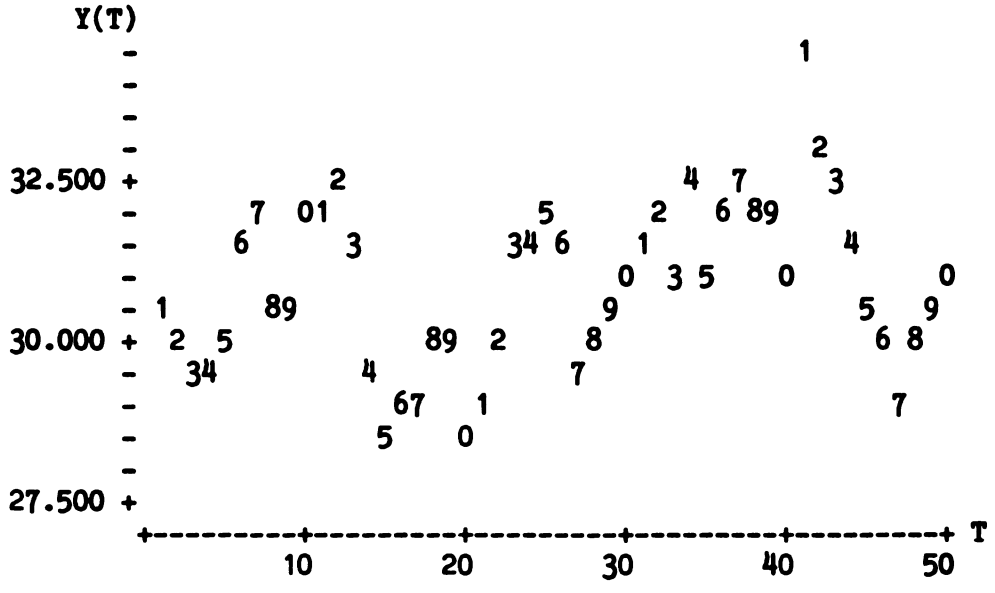
The problem of segmentation considered here is: Given a time series

$$\{x(t),\ t=1,2,\ \ldots\ ,n\},$$

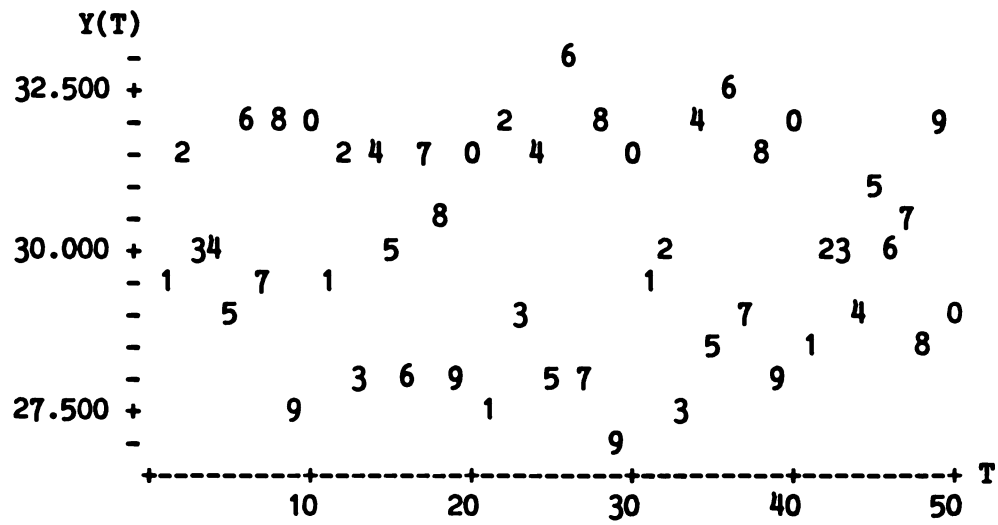
partition the set of values of t into subseries (segments, regimes) for which the values x(t) are relatively homogeneous in some sense. The segments are assumed to fall into several classes.

Examples. (i) Segment a received signal into background, target, background again, another target, etc. (ii) Segment an EEG of a sleeping persion into periods of deep sleep and restless or fitful sleep (two classes of segment). (iii) Segment an ECG into rhythmic and arhythmic periods (two classes of segment). (iv) Segment an economic time series into periods of recession, recovery, and expansion (three classes of segment).

Next, several simulated time series will be shown. The first is relatively smooth; the second, relatively rough or jumpy. (They are simulated first-order autoregressions with autoregression coefficients equal to +.8 and -.8, respectively.)

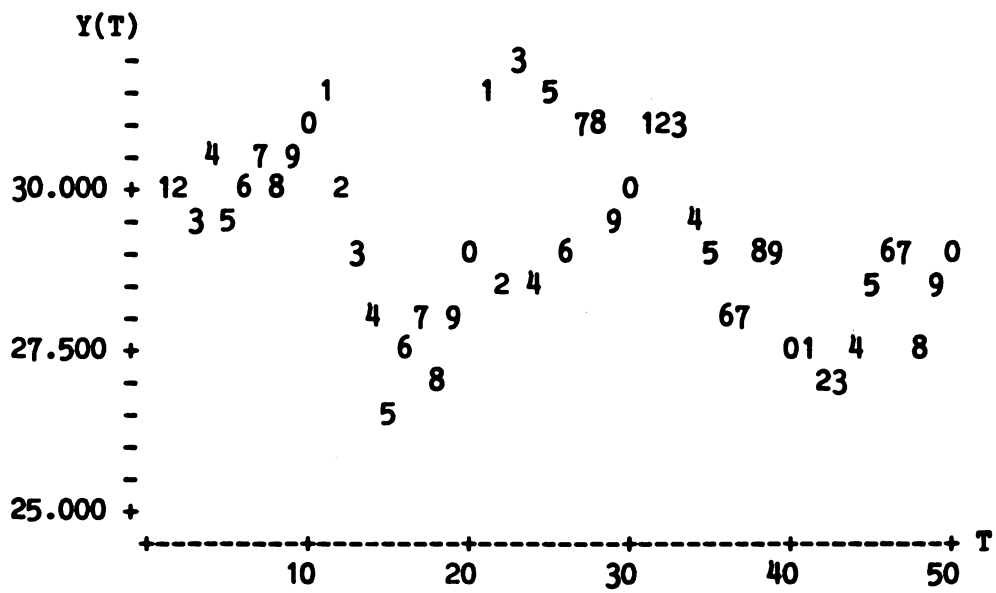


Smooth time series (simulated first-order autoregression with coefficient +.8)



Rough time series (simulated first-order autoregression with coefficient -.8)

The next series has alternating smooth and jumpy segments.

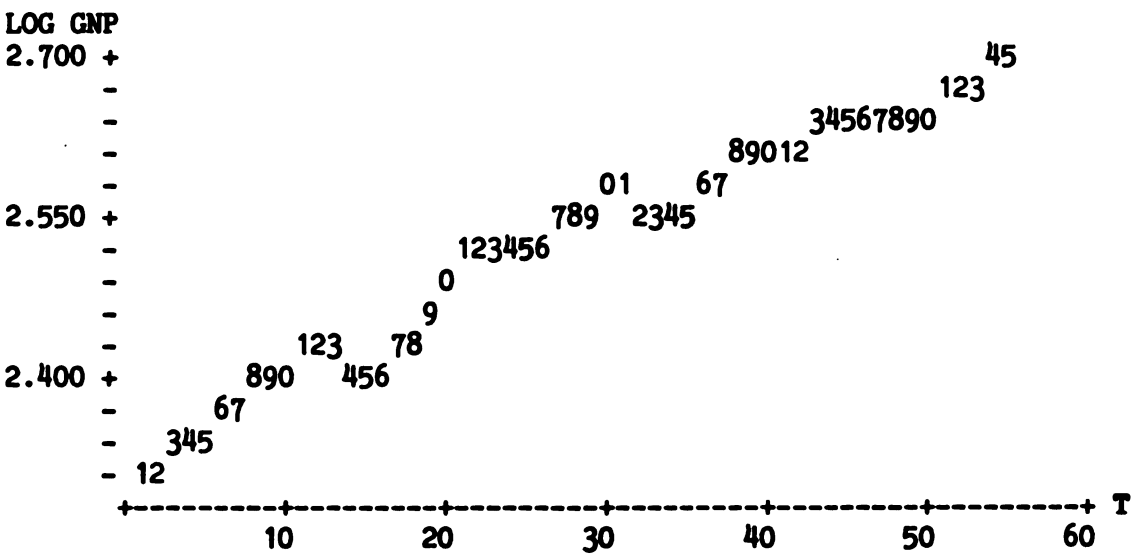


Simulated time series with three segments (smooth, then rough, then smooth)

It is a simulated series, with three segments. Each segment is a first-order autoregression. For t = 1 to 20 the coefficient is +.8; for

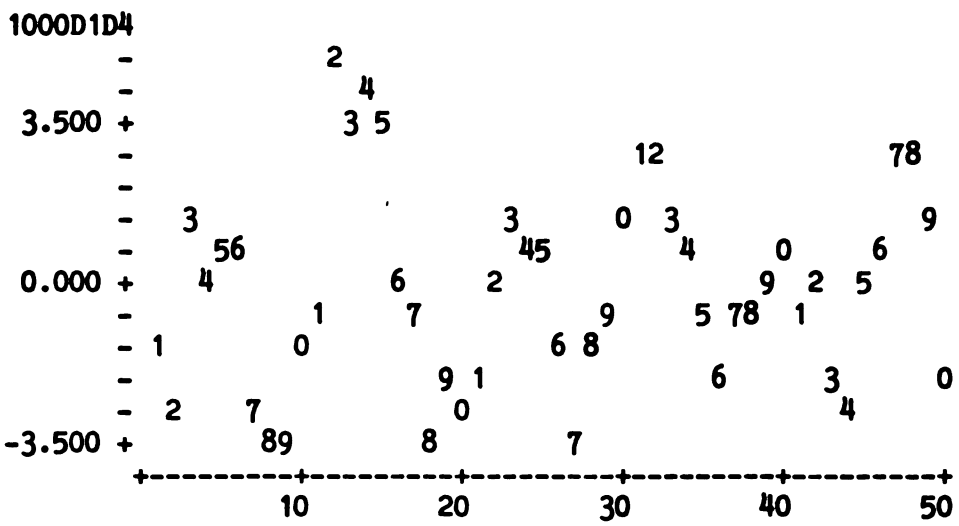
t = 21 to 30 the coefficient is -.8; then for t = 31 to 50 the coefficient is again +.8. In actual data analysis one would not know the segments but would have to find them and estimate their parameters.

Now let us consider a real example, rather than simulated data. The graph below is quarterly GNP (Gross National Product) for 55 quarters, starting with the first quarter of 1946.



Plot of Y(T), where Y(T) is log (base ten) quarterly GNP in billions of current (unadjusted) dollars. T=1 is 1946-1, T=2 is 1946-2, ..., T=5 is 1947-1, etc. (N=55)

Here is a second difference of the log quarterly GNP. Do you see anything particularly interesting?



Plot of mixed second difference of log GNP: Vertical scale, labeled '1000D1D4', is Z(T), where Z(T) = 1000[(Y(T)-Y(T-1)) - (Y(T-4)-Y(T-5))] and Y(T) = log GNP (base 10); T=1 is 1947-2, T=2 is 1947-3, ..., T=9 is 1949-2, etc. (N=50)

A first difference corresponds to a velocity. The difference

$$y(t) - y(t-1) = d(t),$$

say, is $[y(t) - y(t-1)]/[t-(t-1)]$, which is the change in y over the one time unit from time t-1 to time t. A second difference

$$d(t) - d(t-1),$$

which in terms of an original series of y's is

$$\{[y(t)-y(t-1)] - [y(t-1)-y(t-2)]\},$$

is proportional to the change in velocity across the indicated time periods and hence is essentially an acceleration. Thus a second difference is perhaps a natural transform to analyze. (Note, however, that the second difference used here is a mixed second difference, namely, the ordinary difference of the lag-four difference y(t)-y(t-4).] The second-difference series appears level. What I think is particularly interesting is that the acceleration of the economy due

to the Korean conflict of the 1950's is readily apparent. (T = 12 is 1950-1, T = 13 is 1950-2, etc.) I think that the need for segmentation is clearly indicated. One needs to give some sort of special treatment to those four exceedingly high values.

An alternative approach to such observations is to identify them as "outliers." However, "outlier" connotes spuriousness. If outlying observations are non-spurious or are associated with a recurring cause, perhaps they should not then be termed "outliers." They should be modeled.

In other cases, where, e.g., equipment failure is suspected, one truly wants to look for outliers. Here, too, segmentation can be useful.

De Alba and Zartman (1979) analyzed radiotelemetric measurements of cows' temperatures, in order to locate the periods of high estrus, with a view toward more optimally timed breeding and efficient milk production. A technique of de Alba and Van Ryzin (1979, 1980) was used. In the de Alba-Zartman report the analysis of the temperatures of one cow over 133 days is discussed in detail. Eleven observations were detected as coming from distributions with means shifted relative to the rest of the observations. Inspection of the data showed that some of these temperatures were high and some extremely low. It is concluded that the high temperature readings correspond to times of active estrus. (Perhaps the low readings correspond to instrument failure and hence are true "outliers" in that sense.)

Here, as in de Alba and Zartman (1979), the raw data were pre-processed by differencing. Model-selection criteria (see below)

estimated the number, k, of classes of segment as two, but k = 3 and k = 6 scored almost as well. Since the numbers analyzed were differences, it is perhaps especially interesting to consider three classes, particularly since the segmentation yielded one class for positive differences, one for negative differences, and one for differences close to zero. The results obtained are similar to those obtained by de Alba and Zartman. They identified eleven observations as "outliers;" 7 of these were high and 4 were low. The segmentation algorithm run with three classes of segment identified 4 observations as high; these were among the 7 identified by de Alba and Zartman. Use of six classes provided even closer agreement to the de Alba-Zartman results. The upper two of the six classes captured 6 of the 7 high observations; the 4 low observations were located by the lowest class.

(In discussion after the paper Professor Parzen mentioned that at Texas A&M University also they had dealt with the analysis of bovine estrus and found by spectral analysis that a filtered Poisson process -- see, e.g., Parzen (1962) -- provided a good fit to the data.)

Now, having discussed some examples, let me be a little more specific about the model and the algorithm. [More formal presentations are found in Sclove (1983a,c,d).] The elements of the segmentation model are the class-conditional time series or distributions, with their parameters; the labels; and the transition probabilities between the labels. Correspondingly, the algorithm alternates between estimation of the distributional parameters, estimation of the labels, and estimation of the transition probabilities. That is, given a tentative labeling,

one can obtain tentative estimates of the parameters of the class-conditional distributions and of the transition probabilities. One then relabels the observations, using these updated parameter estimates. The relabeling is done as follows: If x(t) (i.e., if time period t) is currently labeled as class c, then x(t+1) is labeled as that class d for which the product

$$p(c \text{ to } d)f(x(t+1)|d)$$

is maximal, where p(c to d) denotes the current estimate of the probability of a transition from c to d and f(x|d) denotes the tentatively estimated class-d probability density, evaluated at x. This makes sense because, under the assumptions of the model, the likelihood is the product of these terms.

To illustrate the algorithm, let us consider a short, artificial time series.

t:	1	2	3	4	5	6	7	8	9	10	11	12
x(t):	1	1	3	1	2	1	2	6	7	1	1	1

Suppose it is specified that there are two classes and that the class-conditional distributions are exponential. Suppose the initial guesses of the parameters are equal prior probabilities for the two classes and means of 2 and 3. Then initially the class-conditional densities are taken as $f(x|a) = (1/2)\exp(-x/2)$ and $f(x|b) = (1/3)\exp(-x/3)$. In the first iteration, using equal prior probabilities of .5, one labels x as having come from Class A if $.5f(x|a) > .5f(x|b)$, which simplifies to $x < 2.43$. This gives the following estimated labels at the end of the first iteration.

t:	1	2	3	4	5	6	7	8	9	10	11	12
x(t):	1	1	3	1	2	1	2	6	7	1	1	1
label:	a	a	b	a	a	a	a	b	b	a	a	a

Now the transition probabilities can be estimated from the sequence of estimated labels. The 11 transitions are: a to a, a to b, b to a, a to a, a to a, a to a, a to b, b to b, b to a, a to a, a to a. These give the following estimates of the transition probabilities.

p(a to a) = 3/4 p(a to b) = 1/4

p(b to a) = 2/3 p(b to b) = 1/3

The class means are estimated as follows (but see Section 6: Plans for Further Research, for a discussion of this point).

Mean of a's: (1 + 1 + 1 + 2 + 1 + 2 + 1 + 1 + 1)/9 = 11/9 = 1.22

Mean of b's: (3 + 6 + 7)/3 = 16/3 = 5.33

Now the condition for labeling the current x as "a", given that the preceding x has been labeled as "a", is

$$p(a \text{ to } b)f(x|b) < p(a \text{ to } a)f(x|a),$$

or

$$(1/4)(1/5.33)\exp(-x/5.33) < (3/4)(1/1.22)\exp(-x/1.22),$$

which simplifies to $x < 4.075$. Similarly, the condition for labeling x as "a", given that the preceding observation has been labeled as "b", is $p(b \text{ to } b)f(x|b) < p(b \text{ to } a)f(x|a)$, which simplifies to $x < 1.24$. These second-iteration classification rules change only the label of x(3) from "b" to "a".

At this point let me mention a possible improvement to the algorithm. At the relabeling stage, where the labels are re-estimated, based on tentative parameter estimates and transition probabilities, the

problem is one of estimating a finite sequence, given the transition probabilities. The dynamic programming approach to this problem is known as the Viterbi algorithm. [Forney (1971) is perhaps the most readable reference on this; the references Viterbi (1967, 1971) and Viterbi and Odenwalder (1969) are also given.] This approach will be applied in the future. The present approach is perhaps somewhat simpler but has its advantages. It adapts itself to operation in the purely sequential mode, and it is relatively easy to program.

An algorithm, such as the present one, which alternates between optimizing different sets of variables, is known as a "relaxation" method (Southwell 1940, 1946; Ortega and Rheinboldt 1970). In the present case the different sets of variables are different parameters, namely, the distributional parameters, the transition probabilities, and the labels.

We note in passing that the EM algorithm -- see Dempster, Laird and Rubin (1977) -- is such a relaxation method, where at the "E" step the estimation is by expectation and at the "M" step the estimation is by maximization.

An alternative approach to the presently-implemented relaxation method would involve equating partial derivatives of the likelihood function to zero and solving the resulting equations. It is clear that results analogous to those of Wolfe (1970) will be obtained by this approach, with transition probabilities replacing his mixture probabilities. In any case, the resulting equations have to be solved by an iterative, numerical method, and it is not clear whether it would be any better than the the presently-implemented relaxation method.

4. Image-segmentation experiments

Conventional approaches to segmentation (see, e.g., Ahuja and Rosenfeld (1982) for a survey of image models) include ordinary clustering; edge detectors such as the two-dimensional filters of Irwin Sobel or Judy Prewitt; and a pixel-labeling method that has (perhaps inaccurately) been called a relaxation method. This involves updating the current estimate of

$$\Pr(\text{label of } t \text{ is } c \mid \text{data}),$$

the (conditional) probability that the true label of pixel t is class c, given the data, by an updated version of these estimated probabilities, where the updating takes into account the current labels of neighboring pixels. This is done by means of "compatibility coefficients," measuring the consistency of label c for pixel t with the current labelings of neighboring pixels. (See, e.g., Eklundh, Yamamoto and Rosenfeld 1980.) More precisely, to estimate $\Pr(\text{label of } t \text{ is } c \mid \text{data})$, one moves from stage s-1 to stage s as follows. Let

$$\Pr(\text{label of } t \text{ is } c \mid \text{data};s)$$

be the s-th stage estimate. Then

$$\Pr(\text{label of } t \text{ is } c \mid \text{data};s) = \Pr(\text{label of } t \text{ is } c) \mid \text{data};s-1)\, C(t,c;s-1)/C,$$

where C is a normalizing factor equal to the sum of the denominators in this expression and C(t,c;s-1) is the compatibility coefficient; e.g., C(t,c;s-1) could be related to estimated transition probabilities for the labels or taken as the proportion of the neighbors of pixel t which are labeled c at stage s-1. A problem with this relaxation

method, noted by various researchers, is that the image gets good, then "fades," so there is a problem of knowing when to stop the iteration.

This approach is intuitively attractive. The approach I use seems to have the same intuitive attraction, plus the advantage of being embedded in a Markov model which gives a likelihood function, putting the problem in the context of parameter estimation in a probability model and making performance evaluation possible. (For instance, one can consider asking questions such as, "What if we worked with first-order neighbors and the model were really second-order?" I'm not saying that such questions are easy to answer; I'm merely stating that it is only in the presence of a model that their formulation is even possible.)

The idea of the model in the case of images is the same as for time series, but the transition matrix is more complicated. The transition probabilities are, even in the simplest, first-order, one-sided case (where one conditions only on pixels to the north and west of the given pixel, rather than those to the north, west, south, and east), functions p((c,d) to e) of three arguments, where this represents the probability of a transition to class e in pixel t, given that the pixel to the north of t is class c and the pixel to the west of t is class d. [The Markov approach used here is not unrelated to that presented by Professor Grenander at this conference (Grenander 1985; see also Grenander 1983).] The segmentation algorithm discussed in the present paper, like the "relaxation" method using compatibility coefficients, also has the property that each iteration is not necessarily better than the preceding, but there is a likelihood value associated with each

iteration, and these values can be used to pick out the best iteration.

Here are results of some experiments with the segmentation algorithm.

The Fisher iris data consist of 4 variables observed for each of 150 irises, 50 in each of three species. In order to form an experimental image, these 150 were arranged into 15 rows of 10, the first 5 rows being species 1, rows 6-10 being species 2, and rows 11-15 being species 3. Thus the true segmentation looks like this.

```
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
```

The measurements on flowers 50, 100, and 150 were used as initial estimates for the distributional parameters, namely, means of multivariate normal distributions. A common covariance matrix was assumed. (Actually, statistical tests and model-selection criteria suggest that different covariance matrices should be used; my algorithms do allow for this.) The initial estimate of the common covariance was taken to be proportional to the identity matrix. Equal prior probabilities for the three classes were assumed as initial estimates. After only three iterations, no label changed. The segmentation

obtained was as follows.

```
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
2 3 2 3 2 3 2 3 2 3
2 2 2 2 2 2 2 2 2 2
3 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
```

Experiments were run with different configurations of these data. Below is the true configuration for Experiment 1. The four-by-four block of 2's in the middle can be thought of as a target to be located.

```
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 2 2 2 2 1 1 1
1 1 1 2 2 2 2 1 1 1
3 3 3 2 2 2 2 3 3 3
3 3 3 2 2 2 2 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
```

The segmentation produced was as follows.

```
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 3 2 3 2 1 1 1
1 1 1 3 2 2 2 1 1 1
3 3 3 2 2 2 2 3 3 3
2 3 3 2 2 2 2 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
```

The following was the true configuration for Experiment 2. Note the small "target" of four 2's.

```
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 2 2 1 1 1 1
3 3 3 3 2 2 3 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
```

The segmentation obtained was as follows.

```
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 3 2 1 1 1 1
3 3 3 3 3 2 3 3 2 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3
```

Note the algorithm worked reasonably well in detecting even a small target, even with the same initial (equal) estimates of prior probabilities, quite different from the true probabilities of .48, .04, .48 for the three classes. Thus perhaps the algorithm will not have to be well "tuned" in order to work well.

Note there are problems at the northwest edges, due to the way the algorithm is set up. In future software development I might program a procedure analogous to Box and Jenkins' back-forecasting ("backcasting") to take care of this. Then, after a first segmentation, the data would be read through in reverse order, so that northwest pixels become southeast pixels, easy to label correctly.

5. Model-selection criteria

A question of obvious importance is that of how many classes of segment to use, that is, what should be the value of the parameter k? Maximum likelihood estimation alone cannot provide an answer, because it is model-conditional; that is, maximum likelihood applies to the problem of estimating the other parameters, for a fixed value of k. It does not apply to the problem of estimating k because the likelihood itself changes as k does. An approach to the problem of estimation of k is provided by model-selection criteria, such as Akaike's, Schwarz' and Kashyap's. These criteria add a penalty for using extra parameters to the negative log-likelihood function. Thus a small score is a good score on these criteria. Akaike's criterion is based on a heuristic estimate of the cross-entropy of the true model and the model with k classes. (See Parzen, 1982.) Schwarz' criterion, perhaps more convincing, is based on a Bayesian approach. Kashyap obtains Schwarz' criterion by a relatively simple expansion, then takes this expansion a term further to obtain another criterion, which I call Kashyap's criterion. The additional term seems to be particularly meaningful. These model-selection criteria take the form

$$-2 \log L(k) + a(n)m(k) + b(k),$$

where "log" here denotes the natural (base-e) logarithm, and

$L(k)$ = likelihood under the k-th model, maximized with respect to the parameters,

$a(n)$ = 2, for all n, $b(k)$ = 0, for Akaike's criterion,

$a(n)$ = log n, $b(k)$ = 0, for Schwarz' criterion,

$a(n)$ = log n, $b(k)$ = log[det B(k), for Kashyap's criterion,

det = determinant,

and

B(k) = matrix of second partial derivatives of log L(k) with respect to the parameters.

(The mathematical expectation of B(k) is the Fisher information matrix.) The optimal k is that value which minimizes the criterion. Akaike's criterion generally chooses a higher value of k (more parameters) than do the others. Since for n greater than 8, log n is greater than 2, Schwarz' criterion will choose a value of k no larger than that chosen by Akaike's, for n greater than 8.

6. Plans for further research

Several items for future research have already been mentioned, including programming of backcasting and use of the Viterbi algorithm.

Some other plans for additional research will be mentioned now. First there is the matter of improved estimation of the distributional parameters. For purposes of discussion focus on the example of segmenting a time series into two classes, i.e., labeling each observation as an "a" or a "b." There is a truncation resulting from the present implementation of the algorithm. Namely, only those observations labeled "a" will be used in updating the current estimate of the mean of Class A. But these observations are a truncated sample from Distribution A, and the algorithm does not treat them as such. (At the Conference, Professor Tukey very kindly sought me out the day after my presentation to point this out to me, and also to remark on the limitations of the unidirectional approach of the algorithm; see below.) Rather than deal with the truncation per se, I had planned in the next stage of the work to modify the estimators by doing them Bayesianly,

e.g., estimate the mean of Group A as N/D, where

$$N = x(1)\,Pr(a|x(1)) + x(2)\,Pr(a|x(2)) + \ldots + x(n)\,Pr(a|x(n))$$

and

$$D = [Pr(a|x(1)) + Pr(a|x(2)) + \ldots + Pr(a|x(n))]$$

(In this expression, $Pr(a|x)$ can be replaced by $Pr(x|a)$ since $Pr(a|x) = f(x|a)Pr(a)/f(x)$ and $Pr(a)/f(x)$ is a common factor which will cancel out.) In this estimate, all the observations play a role, whether labeled as "a" or "b," so that at least some of the bias will be removed by allowing the larger "b" observations to enter.

I obtain the likelihood function by a one-sided approach, conditioning any given pixel on the results in the pixels to its north and west. A two-sided, full neighborhood approach seems preferable to a unidirectional one. The unidirectional approach is a device for writing down the likelihood, but this does not mean one has to be wedded to that approach in the iterative updating. That is, the parameters can be estimated with a full neighborhood approach.

Another bit of further research is to calculate Kashyap's criterion for various clustering and segmentation models. Also, so far the method, algorithm and software have been developed only for the case where the observations within a class are independent (and Gaussian). A next step will be autoregression within classes. This is of obvious importance in time series, and in the context of images, it can be used to model textures. Still another generalization is to allow some forms of time or state dependency in the transition probabilities.

REFERENCES

Ahuja, N., and A. Rosenfeld (1982). Image models. Pp. 383-397 in Handbook of Statistics 2: Classification, Pattern Recognition, and Reduction of Dimensionality (P. R. Krishnaiah and L. N. Kanal, eds.), North Holland/American Elsevier, New York.

Akaike, H. (1983). Statistical inference and measurement of entropy. Pp. 165-189 in Scientific Inference, Data Analysis, and Robustness (H. Akaike and C.-F. Wu, eds.), Academic Press, New York.

Box, G.E.P., and G.M. Jenkins (1976). Time Series Analysis: Forecasting and Control, rev. ed. Holden-Day, San Francisco.

Bozdogan, H., and S. L. Sclove (1984). Multi-sample cluster analysis using Akaike's information criterion. Annals of the Institute of Statistical Mathematics 36, 163-180.

de Alba, E., and J. Van Ryzin (1979). An empirical Bayes approach to outliers. Journal of Statistical Planning and Inference 4, 217-236.

de Alba, E., and J. Van Ryzin (1980). An empirical Bayes approach to outliers: shifted mean case. Pp. 33-44 in Recent Developments in Statistical Inference and Data Analysis (K. Matusita, ed.), North Holland/American Elsevier, Amsterdam and New York.

de Alba, E., and Zartman, D.L. (1979). Testing outliers in time series: an application to remotely sensed temperatures in cattle. Special Paper Number 130, Agricultural Experiment Station, New Mexico State University, Las Cruces, New Mexico.

Dempster, A. P., N. M. Laird and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society (B) 39, 1-38.

Eklundh, J. O., H. Yamamoto and A. Rosenfeld (1980). A relaxation method for multispectral pixel classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 2, 72-74.

Forney, G. D., Jr. (1973). The Viterbi algorithm. Proceedings of the IEEE 61, 268-278.

Grenander, U. (1983). Tutorial in pattern theory. Notes. Division of Applied Mathematics, Brown University, Providence, Rhode Island.

Grenander, U. (1985). Proceedings of this conference.

Johari, S., and S. L. Sclove (1976). Partitioning a distribution. Communications in Statistics A5, 133-147.

Kashyap, R. L. (1982). Optimal choice of AR and MA parts in autoregressive moving average models. IEEE Transactions on Pattern Analysis and Machine Intelligence 4, 99-104.

Ortega, J., and W. Rheinboldt (1970). Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York.

Osterburg, J. W., and O'Hara, C. E. (1949). An Introduction to Criminalistics. Macmillan, New York.

Osterburg, J. W. (1968). The Crime Laboratory: Case Studies of Scientific Criminal Investigation. Indiana University Press, Bloomington, Indiana.

Osterburg, J. W. (1982). The Crime Laboratory: Case Studies of Scientific Criminal Investigation, 2nd edition. Clark Boardman Co., Ltd., New York.

Osterburg, J. W., T. Parthasarathy, T.E.S. Raghavan, and S. L. Sclove (1977). Development of a mathematical formula for the calculation of fingerprint probabilities based on individual characteristics. Journal of the American Statistical Association 72, 772-778.

Parzen, E. (1962). Stochastic Processes. Holden-Day, San Francisco.

Parzen, E. (1982). Maximum entropy interpretation of autoregressive spectral densities. Statistics & Probability Letters 1, 7-11.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics 6, 461-464.

Sclove, S. L. (1977). Population mixture models and clustering algorithms. Communications in Statistics A6, 417-434.

Sclove, S. L. (1979). The occurrence of fingerprint characteristics as a two-dimensional process. Journal of the American Statistical Association 74, 588-595.

Sclove, S. L. (1980). The occurrence of fingerprint characteristics as a two-dimensional Poisson process. Communications in Statistics A9, 675-695.

Sclove, S. L. (1981a). Modeling the distribution of fingerprint characteristics. Pp. 111-130 in Statistical Distributions in Scientific Work 6 (C. Taillie et al., eds.). NATO Advanced Study Institutes Series. D. Reidel, Dordrecht, Holland

Sclove, S. L. (1981b). Pattern recognition in image processing using interpixel correlation. IEEE Transactions on Pattern Analysis and Machine Intelligence 3, 206-208.

Sclove, S. L. (1983a). Time-series segmentation: a model and a method. Information Sciences 29, 7-25.

Sclove, S. L. (1983b). Application of the conditional population-mixture model to image segmentation. IEEE Transations on Pattern Analysis and Machine Intelligence 5, 428-433.

Sclove, S. L. (1983c). On segmentation of time series. Pp. 311-330 in Studies in Econometrics, Time Series, and Multivariate Statistics (S. Karlin, T. Amemiya, and L. Goodman, eds.), Academic Press, New York.

Sclove, S. L. (1983d). On segmentation of time series and images in the signal detection and remote sensing contexts. Pp. 421-434 in Statistical Signal Processing (E. J. Wegman and J. G. Smith, eds.), Marcel Dekker, Inc., New York.

Southwell, R. (1940). Relaxation Methods in Engineering Science: a Treatise on Approximate Computation. Oxford University Press, London.

Southwell, R. (1946). Relaxation Methods in Theoretical Physics. Oxford University Press (Clarendon), London and New York.

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory IT-13, 260-269.

Viterbi, A. J., and J. P. Odenwalder (1969). Further results on optimal decoding of convolutional codes, IEEE Transactions on Information Theory (Corresp.) IT-15, 732-734.

Viterbi, A. J. (1971). Convolutional codes and their performance in communication systems. IEEE Transactions on Communications Technology COM-19, 751-772.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. Multivariate Behavioral Research 5, 329-350.

QUANTITATIVE METHODS DEPARTMENT
COLLEGE OF BUSINESS ADMINISTRATION
UNIVERSITY OF ILLINOIS AT CHICAGO
BOX 4348
CHICAGO, IL 60680

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

RECENT RESEARCH IN EXPERIMENTAL DESIGN FOR QUALITY IMPROVEMENT WITH APPLICATIONS TO LOGISTICS

George E. P. Box

Technical Summary Report #2774
November 1984

ABSTRACT

The success of Japanese industry in producing high quality products at low cost is cited. Consideration of certain aspects of scientific method leads to discussion of recent research on the role of screening designs in the improvement of quality. A projective rationale for the use of these designs in the circumstances of factor sparsity is advanced. In this circumstance the possibility of identification of sparse dispersion effects as well as sparse location effect is considered. A new method for the analysis of fractional factorial designs is advanced.

AMS (MOS) Subject Classifications: 62K15

Key Words: Logistics, Quality Control, Scientific Method, Screening Designs, Factor Sparsity, Fractional Factorials, Dispersion Effects, Location Effects, Normal Plots, Bayesian Analysis

Work Unit Number 4 (Statistics and Probability)

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

RECENT RESEARCH IN EXPERIMENTAL DESIGN FOR QUALITY IMPROVEMENT WITH APPLICATIONS TO LOGISTICS

George E. P. Box

1. LOGISTICS AND QUALITY CONTROL

A traditional philosophy of quality control has been to "inspect bad quality out" and indeed there are famous military standards that employ this philosophy. W. Edwards Deming (1982) has likened this to making toast according to the recipe "you burn it and I'll scrape it", and has urged the alternative philosophy of assuring that good quality has been built in to the product in the first place. In particular he attributes to the latter philosophy the success of Japanese industry in producing high quality products at low cost. A typical example of the dramatic consequences that have been attributed to these differences of approach are the air-conditioner defect rates shown in Table 1 and quoted by David Garvin (1983).

(In the factory: Assembly line defects per 100 units)

	American	Japanese
Total	63.5	0.95
Leaks	3.1	0.12
Electrical	3.3	0.12

(In the field: Service call rate per 100 units under first year warranty coverage)

	American	Japanese
Total	10.5	0.6
Compressors	1.0	0.05
Thermostats	1.4	0.002
Fan motors	0.5	0.028

TABLE 1. Defect rates in US and Japanese air conditioners

Similar comparisons have been made between defect rates in American and Japanese automobiles.

The same United States industry that makes air conditioners and motor vehicles also makes military hardware. It seems clear therefore that a major change in quality philosophy could produce a major improvement in the reliability of the Army's equipment. The philosophy of "building quality in" employs a policy of never ending quality improvement which may be typified in terms of the traditional statistical model

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

$$y = f(\underset{\sim}{x}_1) + e$$

where y is a quality characteristic believed to depend on a set of variables denoted by $\underset{\sim}{x}_1$ whose identity is known, and e is the difference $y - f(\underset{\sim}{x}_1)$ usually referred to as error. (Such "errors" are often somewhat arbitrarily imbued by the theoretician with properties of randomness, normality independence and homoscedasticity). In reality e is a function $e(\underset{\sim}{x}_2)$ of a number of additional variables, $\underset{\sim}{x}_2$ say, which affect the system but whose identity is usually unknown. In general, quality improvement is achieved by transferring elements of the unknown factor vector $\underset{\sim}{x}_2$ into the known factor vector $\underset{\sim}{x}_1$ as indicated below

$$y = f(\underset{\sim}{x}_1) + e(\underset{\sim}{x}_2) \quad .$$

known unknown

The effect of such transfer is two-fold

(i) to reveal effects of previously unknown factors which may then be adjusted to levels yielding higher quality and/or used to control the process.

(ii) to remove variation previously caused by haphazard changes in these factors.

Some of the statistical techniques which contribute to this transfer are quality control charting (including Shewhart, Cusum, Pareto and Fishbone charts) and designed experimentation on line and off line (employing in different and appropriate contexts factorial, fractional factorial and orthogonal array designs, evolutionary operation and response surface methods).

2. SCIENTIFIC METHOD AND QUALITY

Charting and experimentation are examples respectively of passive surveillance and active intervention both of which are important elements in scientific method which it is desirable to consider further.

Humans differ from other animals most remarkably in their ability to learn. It is clear that although throughout the history of mankind technological learning has taken place, until three or four hundred years ago change occurred very slowly. One reason for this was that in order to learn something - for example, how to make fire or champagne - two rare events needed to coincide: (a) an informative event had to occur, and (b) a person able to draw logical conclusions and to act on them had to be aware of that informative event.

Passive surveillance is a way of increasing the probability that the rare informative event will be constructively taken note of and is exemplified by quality charting methods. Thus a Shewhart chart is a means to ensure that

possibly informative events are brought to the attention of those who may be able to discover in them an "assignable cause" (Shewhart 1931) and act appropriately.

Active intervention by experimentation aims, in addition, to increase the probability of an informative event actually occurring. A designed experiment conducted by a qualified experimenter can dramatically increase the probability of learning because it increases simultaneously the probability of an informative event occurring and also the probability of the event being constructively witnessed. Recently there has been much use of experimental design in Japanese industry particularly by Genichi Taguchi (Taguchi and Wu (1980)) and his followers. In off-line experimentation he has in particular emphasized the use of highly fractionated designs and orthogonal arrays and the minimization of variance.

In the remainder of this paper we briefly outline some recent research on the use of experimental design in the improvement of quality.

3. USE OF SCREENING DESIGNS TO IMPROVE QUALITY

Table 2 shows in summary a highly fractionated two-level factorial design employed* as a screening design in an off-line welding experiment performed by the National Railway Corporation of Japan (Taguchi and Wu, 1980). In the column to the right of the table is shown the observed tensile strength of the weld, one of several quality characteristics measured.

The design was chosen on the assumption that in addition to main effects only the two-factor interactions AC, AG, AH, and GH were expected to be present. On that supposition, all nine main effects and the four selected two-factor interactions can be separately estimated by appropriate orthogonal contrasts, the two remaining contrasts corresponding to the columns labelled e_1 and e_2 measure only experimental error. Below the table are shown the grand average, the fifteen effect contrasts, and the effects plotted on a dot diagram. When the effects are plotted on normal probability paper, thirteen of them plot roughly as a straight line but the remaining two, corresponding to the main effects for factors B and C, fall markedly off the line, suggesting that over the ranges studied, only factors B and C affect tensile location by amounts not readily attributed to noise.

If this conjecture is true, then, at least approximately, the sixteen runs could be regarded as four replications of a 2^2 factorial design in factors B and C only. However, when the results are plotted in Figure 1 so as to reflect this, inspection suggests the existence of a dramatic effect of a different kind - when factor C is at its plus level the spread of the

*To facilitate later discussion we have set out the design and labelled the levels somewhat differently from Taguchi.

A: Kind of Welding Rods
B: Period of Drying
C: Welded Material
D: Thickness
E: Angle
F: Opening
G: Current
H: Welding Method
J: Preheating

Factor			D	H	e_1	G	F	GH	AC	A	E	AH	e_1	AG	J	B	C	Tensile strength
Column	Number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	kg/mm^2
	1	+	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+	43.7
	2	+	+	−	−	−	−	+	+	−	−	+	+	+	+	−	−	40.2
	3	+	−	+	−	−	+	−	+	−	+	−	+	+	−	+	−	42.4
	4	+	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+	44.7
	5	+	−	−	+	+	−	−	+	−	+	+	−	−	+	+	−	42.4
	6	+	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+	45.9
	7	+	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+	42.2
Run	8	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	40.6
	9	+	−	−	+	−	+	+	−	+	−	−	+	−	+	+	−	42.4
	10	+	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+	45.5
	11	+	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+	43.6
	12	+	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−	40.6
	13	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	44.0
	14	+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	40.2
	15	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	42.5
	16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	46.5
Effect		43.0	.13	−.15	−.30	−.15	.40	−.03	.38	.40	−.05	.43	.13	.13	−.38	2.15	3.10	

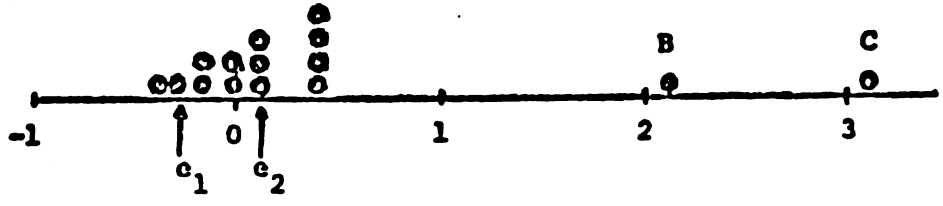


TABLE 2. A fractional two-level design used in a welding experiment showing observed tensile strength and effects. Below the estimated effects are plotted as a dot diagram.

data appears much larger* than when it is at its minus level. Thus, in addition to detecting shifts in location due to B and C, the experiment may also have detected what we will call a dispersion effect due to C. The example raises the general possibility of analyzing unreplicated designs for dispersion effects as well as for the more usual location effects.

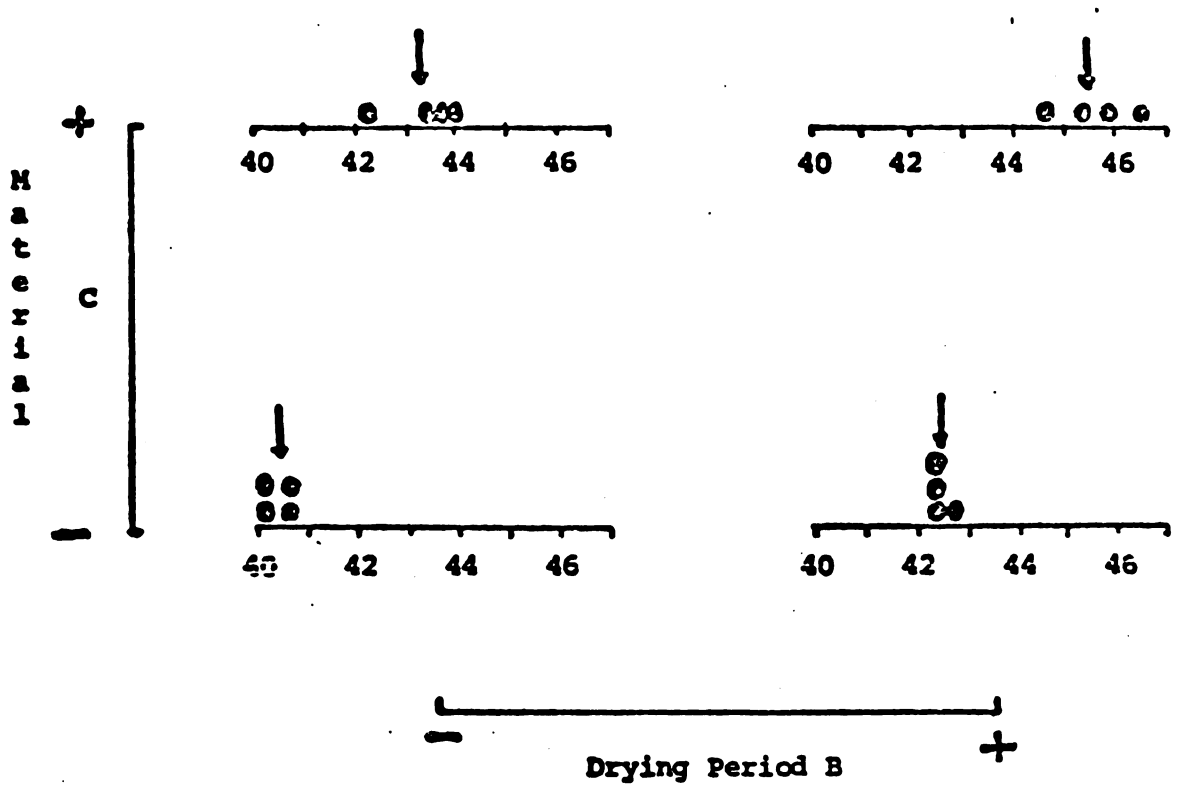


Figure 1. Tensile data as four replicates of a 2^2 factorial design in factors B and C only.

*Data of this kind might be accounted for by the effect of one or more variables other than B that affected tensile strength only at the "plus level" of C (only when the alternative material was used). Analysis of the eight runs made at the plus level of C does not support this possibility, however.

4. RATIONALES FOR USING SCREENING DESIGNS

Before proceeding we need to consider the question, "In what situations are screening designs, such as highly fractionated factorials, useful?"

4.1. Effect Sparsity

A common industrial problem is to find from a rather large number of factors those few that are responsible for large effects. The idea is comparable to that which motivates the use in quality control studies of the "Pareto diagram." (See, for example, Ishikawa (1976)). The situation is approximated by postulating that only a small proportion of effects will be "active" and the rest "inert". We call this the postulate of effect sparsity. For studying such situations, higly fractionated designs and other orthogonal arrays (Finney (1945), Plackett and Burman (1946), Rao (1947), Taguchi and Wu (1980)) which can screen moderately large numbers of variables in rather few runs are of great interest. Two main rationalizations have been suggested for the use of these designs; both ideas rely on the postulate of effect sparsity but in somewhat different ways.

4.2. Rationale Based on Prior Selection of Important Interactions

It is argued (see for example Davies (1954)) that in some circumstances physical knowledge of the process will make only a few interactions likely and that the remainder may be assumed negligible. For example, in the welding experiment described above there were 36 possible two-factor interactions between the nine factors, but only four were regarded as likely, leaving 32 such interactions assumed negligible. The difficulty with this idea is that in many applications the picking out of a few "likely" interactions is difficult if not impossible. Indeed the investigator might justifiably protest that, in the circumstance where an experiment is needed to determine which first order (main) effects are important, it is illogical that he be expected to guess in advance which effects of second order (interactions) are important.

4.3. Projective Rationale Factor Sparsity

A somewhat different notion is that of factor sparsity. Thus suppose that, of the k factors considered, only a small subset of unknown size d, whose identity is however unknown, will be active in providing main effects and interactions within that subset. Arguing as in Box and Hunter (1961) a two-level design enabling us to study such a system is a fraction of resolution $R = d + 1$ (or in the terminology of Rao (1947) an array of strength d) which produces complete factorials (possibly replicated) in every one of the $\binom{k}{d}$ spaces of $d = R - 1$ dimensions. For example, we have seen that on the assumption that only factors B and C are important, the welding design could be regarded as four replicates of a 2^2 factorial in just those two factors. But because the design is of resolution $R = 3$ the same would have been true for any of the 36 choices of two out of the nine factors tested. Thus the design would be appropriate if it were believed that not more than two of the factors were likely to be "active".

	Columns	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
(a)	2^{15-11}_{III}	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
(b)	2^{8-4}_{IV}	•	•		•			•	•			•		•	•	
(c)	2^{5-1}_{V}	•	•		•				•							•
(d)	2^4	•	•		•				•							

TABLE 3. Some alternative uses of the orthogonal array of Table 2.

For further illustration we consider again the sixteen-run orthogonal array of Table 2 and, adopting a roman subscript to denote the resolution R of the design, we indicate in Table 3 various ways in which that array might be used. It may be shown that

(a) If we associated the fifteen contrast columns of the design with fifteen factors, we would generate a 2^{15-11}_{III} design providing four-fold replication of 2^2 factorials in every one of the 105 two-dimensional projections.

(b) If we associated only columns 1, 2, 4, 7, 8, 11, 13, and 14 with eight factors we would agenerate a 2^{8-4}_{IV} design providing two-fold replication of 2^3 factorials in every one of the 56 three-dimensional projections.

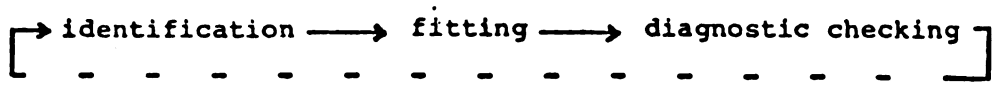
(c) If we associated only columns 1, 2, 4, 8, and 15 with five factors we would generate a 2^{5-1}_{V} design providing a 2^4 factorial in every one of the four-dimensional projections.

(d) If we associated only columns 1, 2, 4, and 8 with four factors we would obtain the complete 2^4 design from which this orthogonal array was in fact generated.

Designs (a), (b) and (c) would thus be appropriate for situations where we believed respectively that not more than 2, 3, or 4 factors would be active*. Notice that intermediate values of k could be accommodated by suitably omitting certain columns. Thus the welding design is a 2^{9-5}_{III} arrangement which can be obtained by omitting 6 columns from the complete 2^{15-11}_{III}. Notice finally that for intermediate designs we can take advantage of both rationales by arranging, as was done for the welding design, that particular interactions are isolated.

*The designs give partial coverage for a larger number of factors, for example (Box and Hunter (1961)) 56 of the 70 four-dimensional projections of the 2^{8-4}_{IV} yield a full factorial in four variables.

A discussion of the iterative model building process by Box and Jenkins (1970) characterized three steps in the iterative data analysis cycle indicated below



Most of the present paper is concerned with model identification - the search for a model worthy to be formally entertained and fitted by an efficient procedure such as maximum likelihood. The situation we now address concerns the analysis of fractional designs such as the welding design in the above context when only a few of the factors are likely to have effects but these may include dispersion effects as well as location effects.

5. DISPERSION EFFECTS

We again use the design of Table 2 for illustration. There are 16 runs from which 16 quantities -- the average and 15 effect contrasts -- have been calculated. Now if we were also interested in possible dispersion effects we could also calculate 15 variance ratios. For example, in column 1 we can compute the sample variance s_{1-}^2 for those observations associated with a minus sign and compare it with the sample variance s_{1+}^2 for observations associated with a plus sign to provide the ratio $F_1 = s_{1-}^2/s_{1+}^2$. If this is done for the welding data we obtain values for $\ln F_i$* given in Figure 2(a). It will be recalled that in the earlier analysis a large dispersion effect associated with factor C (column 15) was found, but in Figure 2(a) the effect for factor C is not especially extreme, instead the dispersion effect for factor D (column 1) stands out from all the rest. This misleading indication occurs because we have not so far taken account of the aliasing of location and dispersion effects. Since sixteen linearly independent location effects have already been calculated for the original data, calculated dispersion effects must be functions of these. Recently (Box and Meyer 1984a) a general theory of location-dispersion aliasing has been obtained for factorials and fractional factorials at two levels. For illustration, in this particular example it turns out that the following identity exists for the dispersion effect F_1, that is the F ratio associated with factor D and hence for column 1 of the design.

$$F_1 = \frac{(\hat{2}-\hat{3})^2+(\hat{4}-\hat{5})^2+(\hat{6}-\hat{7})^2+(\hat{8}-\hat{9})^2+(\hat{10}-\hat{11})^2+(\hat{12}-\hat{13})^2+(\hat{14}-\hat{15})^2}{(\hat{2}+\hat{3})^2+(\hat{4}+\hat{5})^2+(\hat{6}+\hat{7})^2+(\hat{8}+\hat{9})^2+(\hat{10}+\hat{11})^2+(\hat{12}+\hat{13})^2+(\hat{14}+\hat{15})^2} \qquad (1)$$

Now (see Table 2) $\hat{14} = \hat{B} = 2.15$ and $\hat{15} = \hat{C} = 3.10$ are the two largest location effects, standing out from all the others. The extreme value of F_1 associated with an apparent dispersion effect of factor D(1) is largely

---

*In this figure familiar normal theory significance levels are also shown. Obviously the necessary assumptions are not satisfied in this case, but these percentages provide a rough indication of magnitude.

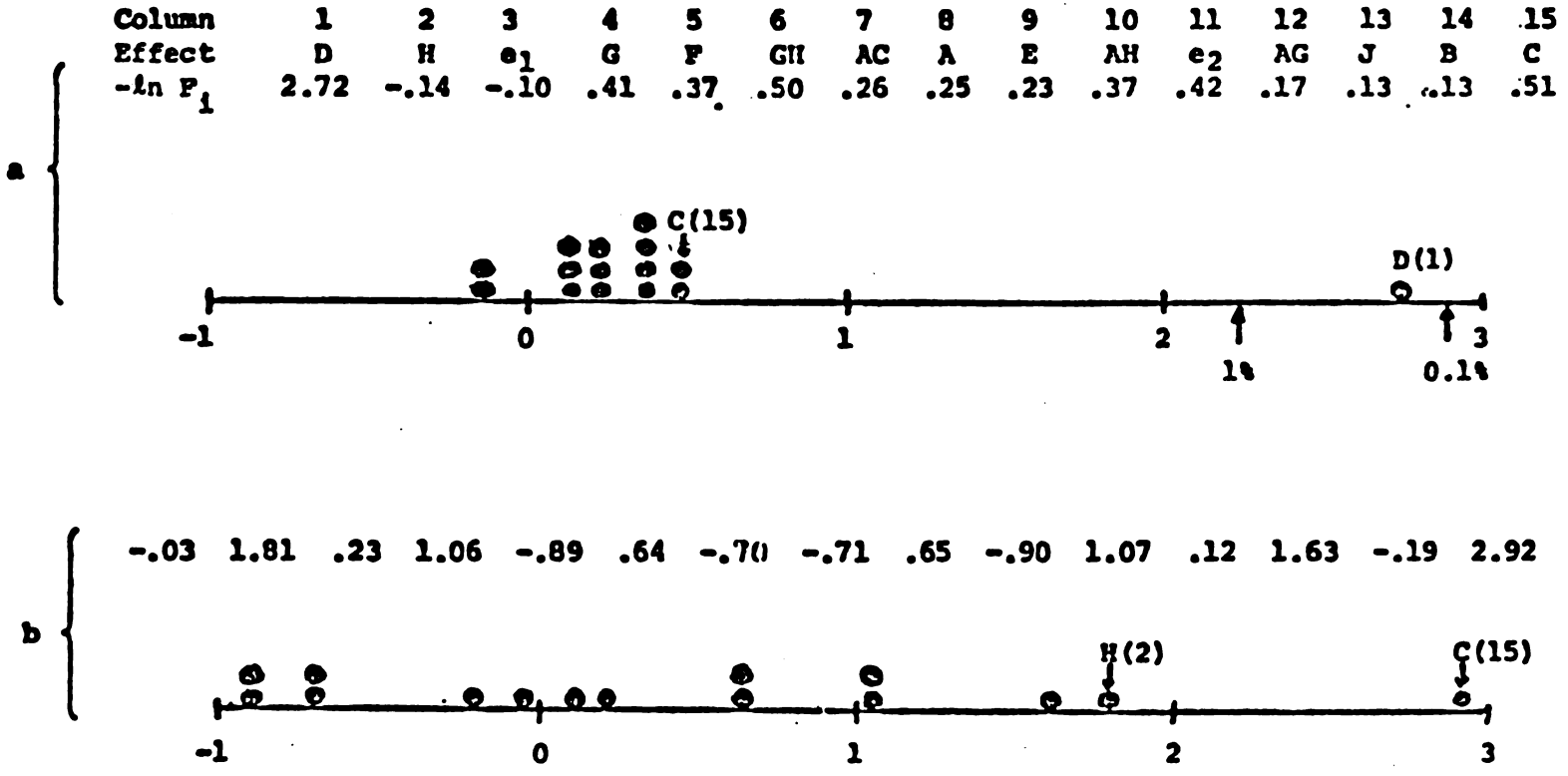


Figure 2. Welding experiment log dispersion effects (a) before, and (b) after elimination of location effects for B and C.

accounted for by the squared sum and squared difference of the location effects $\hat{B}$ and $\hat{C}$ which appear respectively as the last terms in the denominator and numerator of equation (1). A natural way to proceed is to compute variances from the residuals obtained after eliminating large location effects. After such elimination the alias relations of equation (1) remain the same except that location effects from eliminated variables drop out. That is zeros are substituted for eliminated variables. Variance analysis for the residuals after eliminating effects of B and C are shown in Figure 2(b). The dispersion effect associated with C (factor 15) is now correctly indicated as extreme. It is shown in the paper referenced above how, more generally, under circumstances of effect sparsity a location-dispersion model may be correctly identified when a few effects of both kinds are present.

6. ANALYSIS OF UNREPLICATED FRACTIONAL DESIGNS

Another important problem in the analysis of unreplicated fractional designs and other orthogonal arrays concerns the picking out of "active" factors. A serious difficulty is that with unreplicated fractional designs no simple estimate of the experimental error variance against which to judge the effects is available.

In one valuable procedure due to Cuthbert Daniel (1959, 1976) effects are plotted on Normal probability paper. For illustration Table 4 shows the calculated effects from a 2_{IV}^{8-4} design used in an experiment on injection molding (Box, Hunter and Hunter, 1978, p. 399). These effects are plotted on normal probability paper in Figure 3.

$T_1 = -0.7 + 1$ mold temp.
$T_2 = -0.1 + 2$ moisture content
$T_3 = 5.5 + 3$ holding pressure
$T_4 = -0.3 + 4$ cavity thickness
$T_5 = -3.8 + 5$ booster pressure
$T_6 = -0.1 + 6$ cycle time
$T_7 = 0.6 + 7$ gate size
$T_8 = 1.2 + 8$ screw speed

$T_9 = T_{1.2} = -0.6 + 1.2 + 3.7 + 4.8 + 5.6$
$T_{10} = T_{1.3} = 0.9 + 1.3 + 2.7 + 4.6 + 5.8$
$T_{11} = T_{1.4} = -0.4 + 1.4 + 2.8 + 3.6 + 5.7$
$T_{12} = T_{1.5} = 4.6 + 1.5 + 2.6 + 3.8 + 4.7$
$T_{13} = T_{1.6} = -0.3 + 1.6 + 2.5 + 3.4 + 7.8$
$T_{14} = T_{1.7} = -0.2 + 1.7 + 2.3 + 6.8 + 4.5$
$T_{15} = T_{1.8} = -0.6 + 1.8 + 2.4 + 3.5 + 6.7$

TABLE 4. Calculated effects from a 2_{IV}^{8-4} design showing alias structure assuming three factor and higher order interactions negligible. Injection molding experiment.

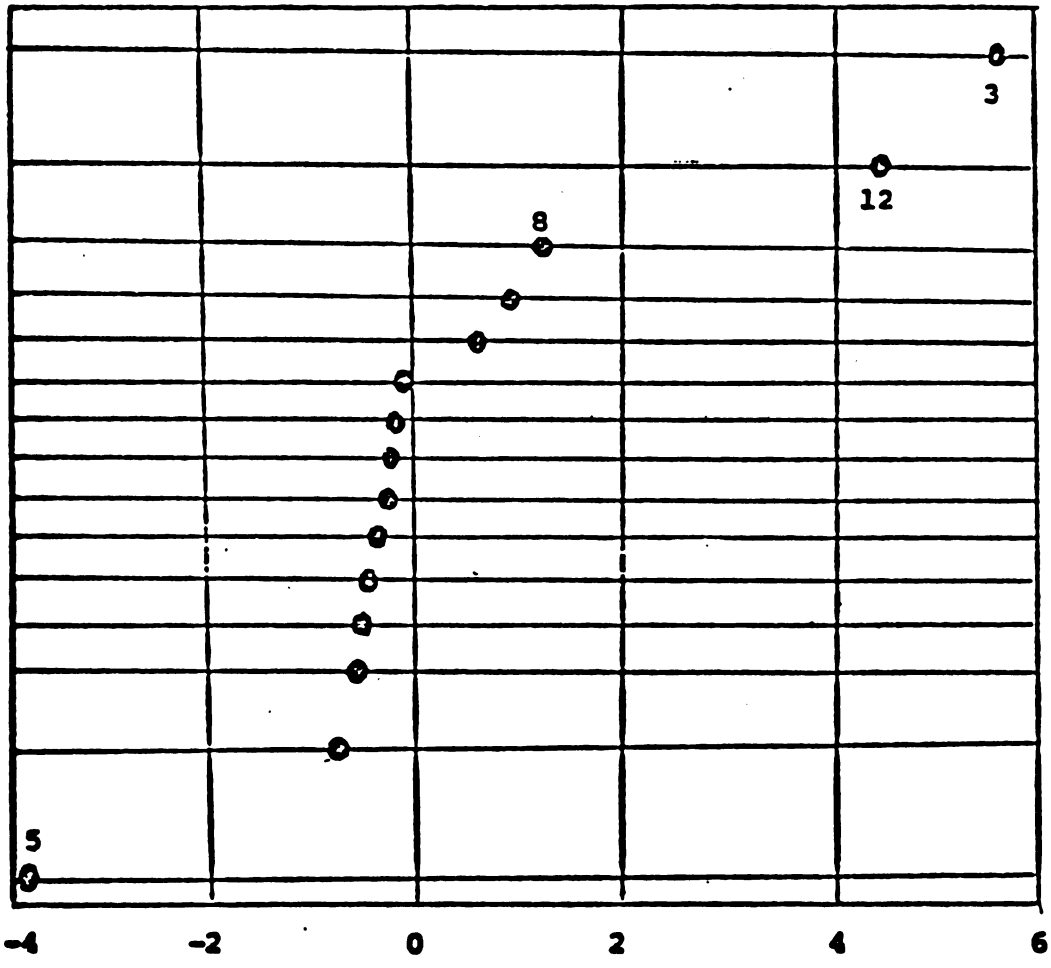


Figure 3. Normal plot of effects. Injection molding experiment.

An alternative Bayesian approach (Box and Meyer, 1984b) is as follows: Let $T_1, T_2, \ldots, T_\nu$ be standardized* effects with

$$T_i = e_i \quad \text{if effect inert}$$

$$T_i = e_i + \tau_i \quad \text{if effect active}$$

$$e_i \sim N(0,\sigma^2),\ \tau_i \sim N(0,\sigma_\tau^2)\ k^2 = \frac{\sigma^2 + \sigma_\tau^2}{\sigma^2}.$$

Suppose the probability that an effect is active is α.

Let $a_{(r)}$ be the event that a particular set of r of the ν factors are active, and let $\underset{\sim}{T}_{(r)}$ be the vector of estimated effects corresponding to active factors of $a_{(r)}$. Then, (Box and Tiao, 1968) with $p(\sigma) \propto \frac{1}{\sigma}$ the posterior probability that $\underset{\sim}{T}_{(r)}$ are the only active effects is:

$$P[a_{(r)} | \underset{\sim}{T}, \alpha, k] \propto \left[\frac{\alpha k^{-1}}{1-\alpha}\right]^r \left\{1 - \left(1 - \frac{1}{k^2}\right)\frac{S_{(r)}}{S}\right\}^{-\frac{\nu}{2}},$$

where $S_{(r)} = \underset{\sim}{T}'_{(r)}\underset{\sim}{T}_{(r)}$ and $S = \underset{\sim}{T}'\underset{\sim}{T}$. In particular the marginal probability that an effect i is active give T, α and k is proportional to

$$\sum_{\substack{a_{(r)} \\ i \text{ active}}} \left[\frac{\alpha k^{-1}}{1-\alpha}\right]^r \left\{1 - \left(1 - \frac{1}{k^2}\right)\frac{S_{(r)}}{S}\right\}^{-\frac{\nu}{2}}.$$

A study of the fractional factorials appearing in Davies (1954), Daniel (1976) and Box, Hunter and Hunter (1978) suggested that α might range from 0.15-0.45 while k might range from 5 to 15. The posterior probabilities computed with the (roughly average) values. $\alpha = 0.30$ and $k = 10$ are shown in Figure 4(a) in which N denotes the probability (negligible for this example) that there are no active effects. The results from a sensitivity analysis in which α and k were altered to vary over the ranges mentioned above is shown in Figure 4(b).

It will be seen that Figure 4(a) points to the conclusion that active effects are associated with columns 3, 5 and 12 of the design and that column 8 might possibly also be associated with an active factor. Figure 4(b) suggests that this conclusion is very little affected by widely different choices for α and k. Further research with different choices of prior, with marginization with respect to k, and with different choices of the distribution assumptions is being conducted.

---

*For three-level and mixed two and three level designs for example, this analysis is carried out after the effects are scaled so that they all have equal variances.

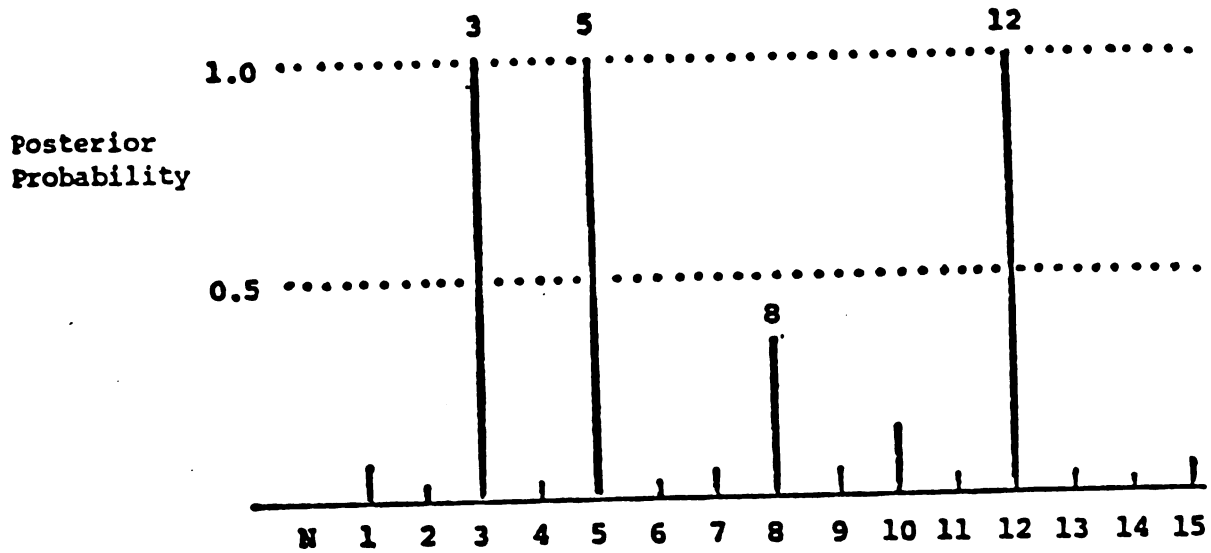


Figure 4(a) Welding experiment. Posterior probability that factor i is active ($\alpha = 0.30$, $k = 10$).

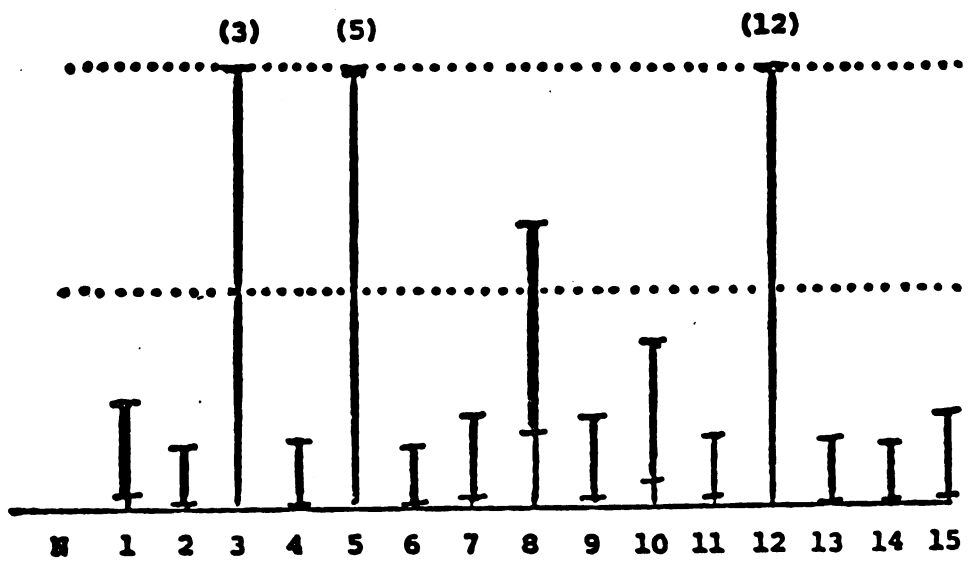


Figure 4(b) Sensitivity analysis for posterior probability $\alpha = .15 - .45$, $k = 5 - 15$.

7. OTHER RESEARCH

Topics which are emphasized in Taguchi's approach to "off line quality control" are (a) reduction of variation by error transmission studies and (b) the choosing of a product design so that it is robust with respect to environmental variables.

These topics are also receiving attention in further research.

REFERENCES

Deming, W. Edwards (1982). Quality, Productivity and Competitive Position. M.I.T., Center for Advanced Engineering Study, Cambridge, MA 02139.

Garvin, David A. (1983). Quality on the Line. Harvard Business Review, Vol. 61, No. 5, p. 64-75.

Shewhart, W. A. (1931). THe Economic Control of Quality of Manufactured Product. Van Nostrand (Reprinted in 1981 by American Society for Quality Control.)

Taguchi, G. and Wu, Y. (1980). Introduction to Off-Line Quality Control. Central Japan Quality Control Association, Nagoya, Japan.

Ishikawa, K. (1976). Guide to Quality Control. Asian Productivity Organization, Tokyo.

Finney, D. J. (1945). The Fractional Replication of Factorial Arrangements. Annals of Eugenics, 12, 4, 291-301.

Plackett, R: L. and Burman, J. P. (1946). Design of Optimal Multifactorial Experiments. Biometrika, 23, 305-325.

Rao, C. R. (1947). Factorial Experiments Derivable from Combinatorial Arrangements of Arrays. J. Roy. Statist. Soc., B9, 128-140.

Davies, O. L. Editor (1954). The Design and Analysis of Industrial Experiments. London: Oliver and Boyd.

Box, G. E. P. and Hunter, J. S. (1961). The 2^{k-p} Fractional Factorial Designs. Technometrics, 3, 311, 449.

Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis, Forecasting and Control. San Francisco: Holden-Day.

Box, G. E. P. and Meyer, R. D. (1984a). Analyzing Two-Level Fractional Factorial Designs for Possible Dispersion Effects. Mathematics Resesarch Center Technical Summary Report #2746, Madison, WI 53705.

Daniel, C. (1959). Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments. Technometrics, 1, 4, 149.

Daniel, C. (1976). Applications of Statistics to Industrial Experimentation. New York: Wiley.

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978). Statistics for Experimenters. New York: Wiley.

Box, G. E. P. and Meyer, R. D. (1984b). An Alternative Approach for the Analysis of Fractional Designs. Mathematics Research Center Technical Summary Report, Madison, WI 53705.

Box, G. E. P. and Tiao, G. C. (1968). A Bayesian Approach to Some Outlier Problems. Biometrika, 55, 119.

DETERMINING COMPLIANCE WITH WEAPON SYSTEM AND EQUIPMENT PERFORMANCE GUARANTEES

Perry C. Stewart
Army Chair
Executive Institute
Defense Systems Management College
Fort Belvoir, Virginia 22060-5426

ABSTRACT. Section 794 of the 1984 Defense Appropriations Act requires written guarantees from prime contractors that weapon systems were designed and manufactured to conform to the government's performance requirements. While warranties covering defects in material and workmanship and even product performance agreements with contingent liabilities extending up to 5 years are not new on defense equipment, the scope of this law far exceeds past contracting regulations. Since compliance must be determined using operational experience rather than controlled test data, the experimental design, data collection, screening and analysis techniques are all critical aspects in negotiating agreements. Technical problems also arise on certain types of equipment where the data collected must be from a surrogate population or some subset of the population whose performance may be affected by non-uniform environmental or operational conditions.

I. INTRODUCTION AND CLINICAL FOCUS. The aspect that was most disturbing to the Department of Defense in Section 794 was not the requirement to cover defects in material and workmanship but the ambiguous aspect of the law requiring contractors to provide a written guarantee on weapon system performance. Much of the initial activity responding to this requirement has centered around the potential legal ramifications of contractors accepting long-term contingent liabilities for the operational performance of their products. The underlying technical issue that lies at the heart of any commitment to guarantee weapon system performance has been largely ignored to date. The basic issue that needs to be understood is how does the requirement dictated by legislation change the way in which the Government determines that a contractor's product meets specified performance requirements in satisfying a military need. Therefore, the focus should really be on developing sound experimental designs when planning contract strategies so that an adequate evaluation can be made of both proposed warranty price as well as contingent risk and liabilities.

II. BASIS OF THE REQUIREMENT. The Department of Defense issued guidance on 14 March 1984 to attempt to clarify the nature of the requirement. The guidance specified that there would be two types of weapon system guarantees, both contained within a single contract. The first of these written guarantees is for conformance to specified performance requirements. Failure to meet the guarantee as evident through either test or demonstration or in operational use over some specified period of time, would require the contractor to design and manufacture the system to satisfy performance shortfalls or repair or replace parts at no increase in price to the Government. The second written guarantee covers freedom from all defects

in material and workmanship applicable to all end items for a specified period of time. Failure to meet this guarantee would require repair or replacement of components at no cost to the Government. The duration of each guarantee may be different within a given contract. There is a requirement for guarantee costs to be evaluated considering both the potential benefits and total cost to the Government. Contracting officers may include clauses to limit the contractor's total liability for guarantees and require guarantees at the component level where it is neither cost effective or feasible at the weapon system level.

III. CURRENT ACCEPTANCE PROCEDURES. In order to understand the potential impact of this requirement it is necessary to briefly review how the Government determines acceptability of weapon system performance requirements currently. In most cases the Government establishes performance requirements through incorporation into a system-level specification. Upon award of a development contract, the system design evolves to achieve the performance requirements stated by the Government. The systems engineering process is the overall mechanism that translates Government specified requirements into a delivered product useable by the Armed Forces of the United States. During the development program there are a series of configuration audits in which the contractor's design approach is reviewed by the Government. The Government implicitly accepts the contractor's design approach through an evolution on ongoing efforts during development such as conduct of paper studies, prototype evaluations, and developmental test results provided at scheduled design reviews. At the functional configuration audit, the contractor provides "design-to" documentation including allocated specifications, drawings, test results, and proposed acceptance test procedures for approval by the Government. Government approval of the results of the functional configuration audit is an indication that the contractor's design approach will meet the requirements of the contract including the performance specifications. It should be noted that there is an implicit acceptance of the technical risk associated with meeting the performance requirements at this point in the development process. In similar manner the Government determines that the contractor is ready for production through what is termed the physical configuration audit. Information supporting manufacturing readiness includes manufacturing plans, product assurance plans, initial operational test results and product readiness reviews. Upon acceptance of the product baseline, again documentation consisting of specifications, test results, and acceptance test procedures, the Government determines that a system or component will meet the requirements of the contract for production. As items are produced, the Government determines compliance with the approved "design-to" and product baseline through acceptance testing and inspection. In summary, except for latent defects, fraud or misrepresentation, as well as short-term liability for defects in materials and workmanship covered under a correction of deficiencies the Government becomes the self-insurer with respect to the use of the product in the operational environment. That is, the Government accepts the risk of the adequacy of design to meet specified performance requirements.

IV. WHAT'S DIFFERENT? A written guarantee or a warranty survives acceptance of the product by the Government. That is, the acceptability of the product to meet specified performance requirements and be free from

defects in material and workmanship may extend for a period of time during the operational use of the delivered system. Instead of determining the acceptability of the product based on developmental test results in a relatively well controlled environment, the full acceptance that performance as evident through design and manufacture is now delayed for verification in the operational environment. The basic design and manufacturing process is the same but there could be some fundamental changes that may affect the development effort including (1) Government performance goals may be lower to recognize increased risk in determining compliance in the operational environment; (2) the contractor's design approach may be more conservative using proven technologies to reduce the contingent liability; (3) additional development and operational testing may be proposed during the full-scale development effort to reduce the technical risk of failed products that do not meet the performance requirements; and (4) development contract costs may increase because of the increased testing efforts noted and increased emphasis on product assurance activities (minimizing the consequence of significant performance shortfalls of delivered systems in the operational environment).

An end result, however, the contractor's post acceptance liability increases because the technical risk that a product meets the performance requirements of the contract has been shifted to some extent from the development phase into the post delivery period for both specified performance requirements and defects in material and workmanship. A comparison of key characteristics that drive this technical risk include (1) an evaluation in the operational vs controlled environment; (2) operation by "green suiter" and maintainers compared to highly skilled engineers and technicians; (3) extended commitments of up to five years vs as low as 90 days; (4) the need to use operational weapon system related diagnostic equipment vs test instrumentation; and (5) dependence on standard service data reporting information to determine compliance in comparison with special data collection procedures used in a controlled test environment.

V. CLINICAL CHALLENGE. As described above, procedures used to determine the acceptability of products in the development process using the results of relatively controlled testing and inspection is well defined. The requirements generated by Section 794 of the 1984 Defense Appropriations Act shift the determination of compliance from the development process into the operational environment. The problem is to determine experimental design guidelines for evaluation of products in the operational environment. Some issues relative to the development of large-scale experimental designs and anticipated results should consider: (1) using the total delivered population to collect results or a selected sample of items that could be more closely controlled; (2) the desirability of sequential testing for risk reduction throughout the extended production delivery period of a program; (3) methods used to relate the reported performance of "pseudo" products to the warranted population. This occurs mainly in areas where training devices are used during peacetime operations and war reserve material may reside in deep storage; and, (5) experimental designs considering the economy in determining compliance with full consideration of the risk of accepting poor performance and rejecting poor quality products.

SUMMARY. On 3 January 1985, the Department of Defense issued Defense Acquisition Circular 84-9 as a temporary regulation on the requirements for warranties on equipment systems in compliance with the 1985 DOD Authorization Act. Extensive warranties are now a mandatory requirement for DOD programs. This continuing requirement creates a challenge and opportunity for mathematicians, statisticians and operations researchers in support of the material acquisition process to provide experimental designs using data from the operational environment to determine compliance with weapon system performance requirements.

A Prioritization Methodology for Materiel Programs

LTC F. Thomas Balzer
and
Richard T. Maruyama

Methodology Branch
Studies and Analysis Directorate
Deputy Chief of Staff for Combat Developments
US Army Training and Doctrine Command
Fort Monroe, Virginia

Abstract

The United States Army Training and Doctrine Command (TRADOC) performs combat developments activities for the Army. The basis for establishing future Army requirements is the Mission Area Analysis (MAA) process. MAA evaluate the capabilities of the programmed force against a threat projected about five years beyond the end of the current program. The deficiencies uncovered are integrated and prioritized in the Battlefield Development Plan (BDP) to put them in the perspective of the total battlefield. HQ TRADOC has now developed a methodology to link materiel programs in the Army's Long Range RDA Plan (LRRDAP) to BDP deficiencies and prioritize these programs based on their contribution to resolving BDP deficiencies. The methodology employs set theory to estimate the total relative worth of each program and establish its relative priority. The methodology contributes to year-to-year consistency in priorities that should yield greater efficiency in the Army's allocation of resources to materiel projects.

I. Introduction.

The Army's Training and Doctrine Command (TRADOC) in partnership with the Army's Materiel Command (AMC) and HQDA are forging a coordinated effort to improve the Army's materiel development and acquisition cycle. TRADOC's Mission Area Analysis (MAA), Battlefield Development Plan (BDP) and the Army's Long Range Research Development and Acquisition Plan (LRRDAP) provides the roadmap for structuring and equipping the Army of the future. These processes provide the means by which the Army can consider future requirements into the Planning, Programming, Budgeting and Execution System for resource allocation.

In response to OMB Circular A-109, AR 1000-1 tasks TRADOC with conducting Mission Area Analysis to analytically support requirements for future Army doctrine, training, organizational structure, and materiel. As part of the Concept Based Requirements System (CBRS), TRADOC has divided the wartime Battlefield mission into 13 mission areas. These mission areas serve as building blocks for measuring the capabilities of the programed force in the current Program Objectives Memorandum (POM) to fight a successful battle against a threat projected about 10 years into the future. These MAA and their resulting deficiencies establish the battlefield needs/requirements. HQ TRADOC in conjunction with its schools and centers and other major Army commands integrates and prioritizes these deficiencies across all mission areas into a single ordered list representing the broader perspective of the total battlefield. In prior years, the BDP was an end product. While indirectly influencing the priorities of doctrinal literature, training programs, force structure modifications, or materiel developments there was no direct linkage which would provide an analytically based rationale for the relative priorities of any of the Army's future developmental activities. In 1984, TRADOC set out to correct this anomaly by developing a methodology to establish the materiel priorities for battlefield systems in the Army's Long Range Research, Development, and Acquisition Plan (LRRDAP) based on their contribution toward resolving one or more of the deficiencies in the BDP. The establishment of such a methodology will lend stability and crediability to the priorities of the associated materiel programs while, at the same time provide the audit trail back to the supporting analysis against the projected threat. This methodology will therefore extend the total concept based requirement system into the PPBES system (see figure 1). The quest for an appropriate linkage methodology was triggered by TRADOC Regulation 11-9 which requires that the Studies and Analysis Directorate (S&AD), Deputy Chief of Staff for Combat Developments (DCSCD) develop such a methodology for prioritizing the LRRDAP based on BDP deficiencies. This paper highlights the theoritical underpinnings and general procedures which resulted from this methodology development.

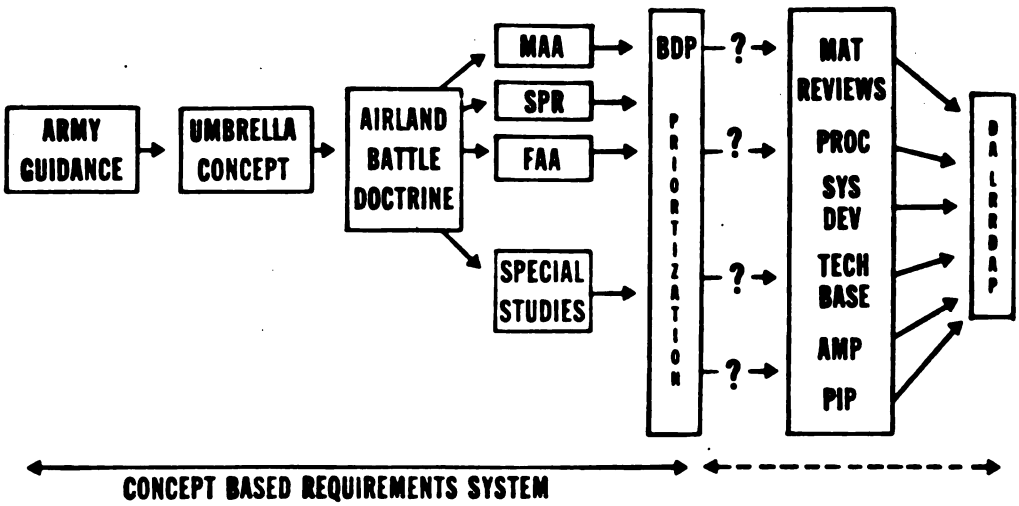


figure 1

II. Methodology.

The priority of materiel programs should be directly related to the priority of the BDP deficiencies they correct. Materiel programs that correct an important battlefield deficiency should be higher on the materiel priority list. This methodology attempts to directly establish this link. Because of the broad and general nature of existing BDP deficiencies it is likely that multiple materiel programs are associated with each deficiency. On the other hand, the same materiel program could contribute to the solution of more than one deficiency. This complication is considered and this methodology will make the necessary adjustments for those cases. An example is show in figure 2.

BDP Def #

Materiel Programs	1	2	3	4	5	6	. . . 76	. . . 227
A			X				X	
B	X					X		
C		X			X		X	X
D		X				X		
E				X				
F	X				X			X

figure 2

In this case, program A contributes to the resolution of BDP deficiencies number 3 and 76 while program C contributes to four deficiencies and program E to only one. As we will establish later in the paper this does not in itself justify that program C has more benefit to the Army. In the example, program E is the only program which contributes to the resolution of deficiency 4 while each of the deficiencies affected by program C also has other programs which make a contribution. As an additional factor in understanding the mathematical foundation of the methodology, one must also consider the nature of the BDP deficiencies which drive the prioritization effort.

BDP deficiencies are aggregations of specific MA deficiencies and as such are sets of commonly related battlefield problems. While there are approximately 1400 specific MA deficiencies in the MAA process, these are aggregated into 227 BDP deficiencies which are then prioritized by a large group of general officers. This group is representative of all combat missions of the Army. Figure 3 shows the relationship of the 1400 specific MA deficiencies which are aggregated into the 227 BDP deficiencies for prioritization by the general officers.

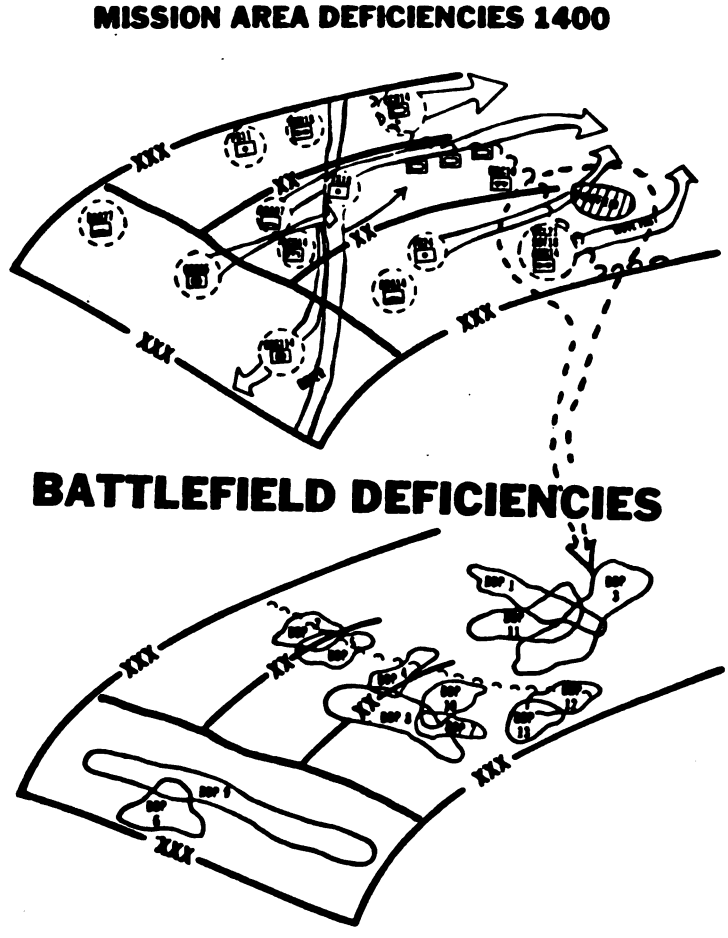
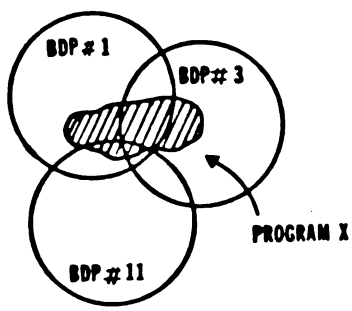


figure 3

Because BDP deficiencies are aggregations of specific mission area deficiencies and each deficiency could require multiple programs for its resolution, the mathematical concept of sets is appropriate. Having established a set of materiel programs for the resolution of each BDP deficiency (which itself is a set of the more specific MAA deficiencies) the methodology uses an underlying assumption that to the extent a materiel program contributes to multiple BDP deficiencies, these deficiencies have overlapping characteristics and there is some degree of commonality among them.

The example (figure 4) illustrates this idea. That is, if a materiel program contributes to the solution of three BDP deficiencies (BDP#1, BDP#3 and BDP#11), then by fielding this program we are simultaneously solving a fraction of all three deficiencies. The methodology must not over credit a program's battlefield contribution by assuming that the set is disjoint.

PROGRAM CONTRIBUTION



PROBLEM:

WHAT IS THE VALUE OF A MATERIEL PROGRAM THAT CONTRIBUTES DIFFERENT AMOUNTS TO DIFFERENT DEFICIENCIES?

E.G. PROGRAM X CONTRIBUTES:
0.6 TO BDP # 1
0.4 TO BDP # 3
0.1 TO BDP #11

figure 4

The question then arises about the degree of overlap. Since there is no feasible method of quantitatively measuring this overlap, existing procedures assume that it is propotional to the degree of contribution of the program to each deficiency. This assumption appears to be not only reasonable but may be also quite accurate. If a program makes a very large contribution to two different deficiencies, they should show a significant degree of overlap in the deficient battlefield functions addressed. On the other hand, a program with a very minor contribution to two (or many) deficiencies would establish only a small degree of overlap among the deficiencies in question. The mathematics to handle these phenomena utilize the well know union operations applied to sets. In the example of figure 4 program X contributes 0.6, 0.4, and 0.1 to the three deficiencies. Therefore, the combined solution (correction) to the three deficiencies is 0.78 not 1.10 (.6 +.4 +.1 = 1.1).

combined contribution = (.6)+(.4)+(.1)-(.6)(.4)-(.6)(.1)-(.4)(.1)
+(.6)(.4)(.1)
= 0.784

Another desired feature of the methodology would account for the fact that some materiel programs contribute to higher priority deficiencies (or groups of deficiencies) than others. For example, if two materiel programs contribute an equal amount to different sets of deficiencies, the one contributing to the higher priority set of deficiencies should have more battlefield worth. This is illustrated for programs X and Y in figure 5. Both programs have the same amount of contribution (.6 and .1) but the set of deficiencies affected by program X is prioritized higher than the set affected by program Y. In this case program X would be prioritized higher.

Program X		Program Y	
BDP #	Contribution	BDP #	Contribution
16	.6	62	.6
36	.1	84	.1

figure 5

In order to resolve this complication a deficiency weight must be assigned to the BDP deficiency that reflects its relative priority position. For the BDP 84 prioritization a uniform distribution of weights was used ranging from 0 to 1. This mathematical assignment was made according to the following formula.

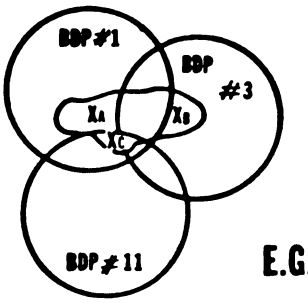
$$Q_i = 1 - \frac{i - 0.5}{n} \qquad (1)$$

In the case where n = 227 and where i = 1, 2, 3, ..., 227. Q_1, Q_3, and Q_{11} take on weighted values of 0.997, 0.988, and 0.953. These deficiency weights when applied to the contribution estimate results in a priority-adjusted contribution value of 0.780 (figure 6) for program X. Therefore, materiel programs that effect BDP deficiencies of high priority will be perferred over materiel programs that effect lower priority deficiencies.

Assigning a numerical value to the degree of contribution of a program for a deficiency is another subjective area open to debate and future research. After some preliminary research and trial applications into different scales, the authors arrived at a five point scale using the narrative descriptors shown in table 1 and values of 0.8, 0.6, 0.4, 0.2, and 0.1 for categories A through E respectively.

These five categories will discriminate enough to ensure that materiel programs making a significant contribution to high priority deficiencies will receive favorable treatment in the competition for funding.

SET THEORY



$$X_A \cup X_B \cup X_C = X_A + X_B + X_C - (X_A \cap X_B) - (X_A \cap X_C) - (X_B \cap X_C) + (X_A \cap X_B \cap X_C)$$

$$\begin{aligned} \text{E.G. } V(P)x &= (.997)(.6) + (.988)(.4) + (.953)(.1) \\ &\quad - (.997)(.6)(.988)(.4) - (.997)(.6)(.953)(.1) \\ &\quad - (.998)(.4)(.953)(.1) \\ &\quad + (.997)(.6)(.988)(.4)(.953)(.1) \\ &= .780 \end{aligned}$$

figure 6

Table I

CONTRIBUTION CATEGORIES

A. ABSOLUTELY ESSENTIAL TO SOLUTION OF DEFICIENCY

B. MAJOR CONTRIBUTOR TO SOLUTION OF DEFICIENCY

C. SUBSTANTIAL CONTRIBUTION TO SOLUTION OF DEFICIENCY

D. SMALL, YET STILL DIRECT CONTRIBUTION TO SOLUTION OF DEFICIENCY

E. LIMITED OR INDIRECT CONTRIBUTION TO SOLUTION OF DEFICIENCY

The mathematics of set theory described above, the BDP priority weight and the contribution scale, establish the materiel program relative worth. This formulation can be generalized in the following form:

GENERALIZED FORMULA

$$V_{(P)i} = \sum_{r=1}^{n} (-1)^{r-1} \sum_{i=1}^{\binom{n}{i}} \prod_{j}^{r} A_{ji}$$

WHERE:

$V_{(P)i}$ = TOTAL RELATIVE WORTH OF PROGRAM i

n = NUMBER OF PROGRAMS CONTRIBUTING TO DEFICIENCY j

A_{ji} = PRODUCT OF THE POSITION VALUE FOR DEFICIENCY j TIMES THE CONTRIBUTION OF PROGRAM i TO DEFICIENCY j

The following example is an application of the methodology applied to three notional materiel programs.

EXAMPLES

PROGRAM X			PROGRAM Y			PROGRAM Z		
PSN	BDP VALUE	CONT VALUE	PSN	BDP VALUE	CONT VALUE	PSN	BDP VALUE	CONT VALUE
.1	.998	.8	6	.977	.8	66	.721	.1
			14	.943	.4	142	.389	.4
			37	.845	.1	202	.143	.8

$$Vp = A + B + C - (AB) - (AC) - (BC) + (ABC)$$

$Vp = .798$ $Vp = .875$ $Vp = .306$

figure 7

This example demonstrates that program Y is prefered over program X and program Z because program Y contributes to three deficiencies in the top 20 percent of the overall list. It should be pointed out at this point that the numbers generated by this methodology to prioritize programs represent ordinal rather than cardinal rankings. Because the deficiency values are scaled uniformly from 0 to 1 and the contribution weights are subjectively determined, the resulting calculations produce numbers without scalar magnitude. That is a program with a value of 0.8 is more important than one with a value of 0.4 one cannot conclude that it is twice as important. In the prioritization of materiel programs it is sufficient to determine only the relative order of the candidates; it is not necessary to measure the interval between them. Research is currently underway to investigate any benefits that may accrue by trying to produce a list with more scalar properities. The steps in the prioritization methodology used in 1984 to prioritize the LRRDAP are summarized in Table II. While the first three steps have been described in some detail above, the last three need a few notes of clarification. Having established an initial prioritization of programs using the stated approach, the methodology attempts to account for the differences in time (near term, mid term, far term), R&D and procurement costs (high, medium, low) and developmental risk (high, medium, low). While mathematics have been developed in the program to consider these factors, time and the availability of input data did not permit their use in last years implementation of the priorities program. Further research is underway to evaluate different methods of considering these factors and also the manpower impacts and the amount of investment already sunk into each program.

Procurement program prioritization had to also consider what is referred to as "base case" programs because the mission area analysis studies took as a given or base case all those programs currently contained in the Program Objectives Memorandum (the Army's 5 year planning and programming document), new deficiencies for the BDP were not generated to support continued

TABLE II

1. DISTRIBUTE DEFICIENCIES UNIFORMLY BETWEEN 0 and 1
2. DETERMINE CONSTRIBUTION OF PROGRAMS TO EACH DEFICIENCY
3. DETERMINE CUMULATIVE CONTRIBUTION FOR EACH PROGRAM (CONSIDER INTERACTIVE EFFECTS)
4. ADJUST CONTRIBUTION BY TIME, COST, RISK, MANPOWER, AND SUNK INVESTMENT
5. PRIORITIZE PROGRAMS
6. PIN TO "BASE CASE" LIST (PROCUREMENT PROGRAMS ONLY)

procurement of these programs. Since they would have no calculated program value, they maintained their original LRRDAP priority and the prioritized list of newer programs was integrated or "pinned" into this list. Non-battlefield programs outside of TRADOC responsibility were also handled like base case programs.

III. RESULTS.

Although developed in only a short period of time to be used in last year's LRRDAP prioritization process, this methodology demonstrated the feasibility of linking materiel program priorities to the deficiencies contained in the battlefield development plan. Manual or "common sense" adjustments had to be made for example to align the 6.3B and 6.4 components of the same program to place them in the same funding band (funded, at risk, unfunded). In the end, of course, the list was subjected to expert review at the O-6, O-8, O-9. and O-10 levels. Of the over 300 RDTE programs prioritized with this hasty methodology, more than 65% retained their relative position and rank in the final approved DA Long Range Plan (figure 8).

OVER 300 RDT&E PROGRAMS WHERE PRIORITIZED THIS YEAR....

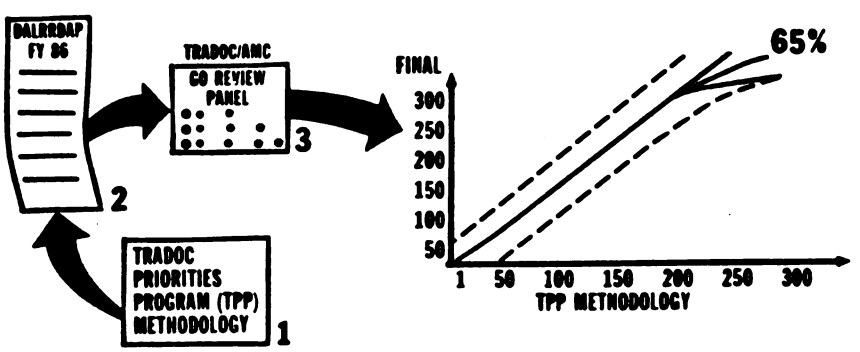


figure 8

Research continues in the areas described throughout this paper and on various means to maintain a relative consistency of program priorities from year to year. Contributors desiring additional information or wishing to comment on proposed improvements to this process are encouraged to contract the authors by phoning (804) 727-3004 or by writing HQ TRADOC, ATTN: ATCD-AM, Ft Monroe, VA 23651-5000.

PROPOSED ADDITIONAL INFERENTIAL INFORMATION DURING AND AFTER HYPOTHESIS TESTING

Paul H. Thrasher

Plans and Quality Assurance Directorate
White Sands Missile Range, New Mexico

ABSTRACT

Two statistics, denoted by β_T and a q-value, are proposed to provide flexibility and inferential information during and after hypothesis testing. These two statistics supplement the well-established statistics normally denoted by beta and the p-value. The purpose of β_T is to emphasize the true value of beta after the design of the hypothesis test but before or during the obtaining of data. A q-value, in conjuction with the p-value and associated power curves, has potential uses in (1) overruling the decision of the hypothesis test, (2) ordering additional testing, and (3) procurement cost analysis.

INTRODUCTION

During the development of statistical inference, there were two distinct needs and resulting philosophies. A statistical technique was needed to estimate the bounds of capabilities of existing equipment or processess. A different statistical technique was needed to test the conformity of newly procured equipment and/or new processes to required specifications. The first of these needs led naturally to flexible ideas such as confidence intervals. The second was met by the rigid procedure of hypothesis testing.

STANDARD HYPOTHESIS TESTING

The standard philosophy used in hypothesis testing is to (1) assume that some parameter of a proposed piece of equipment or process conforms to some standard and (2) abandon that assumption only if there is sufficient evidence that the assumption is highly unlikely. The initial assumption, which is called the null hypothesis and denoted by H_0, is expressed in an equation relating the physical parameter being tested and the specification value of that parameter. The significance level, denoted by alpha, is the minimum acceptable probability that the actual test data could have come from a population for which the null hypothesis is true. Physical measurements give a standard for evaluating the reasonableness of H_0 as a description of the equipment or process under test.

Alpha is one of two risks in hypothesis testing. It is the risk of rejecting the null hypothesis in favor the alternate hypothesis, denoted by H_a, when the null hypothesis is true. The companion risk to alpha, which is denoted by beta, is the probability of not rejecting the null hypothesis when the null hypothesis is not true. Beta depends on four parameters which must be established jointly by the statistician and the manager of the equipment or process being tested:

(1) Alpha.

(2) The specification value of the parameter being tested as stated in the null hypothesis.

(3) A specific value of the parameter satisfying the alternate hypothesis.

(4) The number of proposed physical measurements of the parameter being tested.

Beta is found from the statistics function which describes the distribution of the test statistic under various alternatives. Since there are many values of the parameter being tested that do not meet the specification, there are many betas. The power curve is defined as the graph of one minus beta versus values of the parameter being tested. Since one minus beta is the probability of rejecting the null hypothesis when the null hypothesis is not true, the power curve is defined more generally as the graph of the probability of rejecting versus values of the tested parameter. Alpha and beta have several commonly used names including type I and type II risks, the producer's and consumer's risks, and the government's and contractor's risks. The second and third sets of names are obviously interpretative and explanatory when hypothesis testing is applied in procurement actions. Alpha and beta are competing risks; as either of them increases, the other decreases.

In the operating procedure traditionally used in hypothesis testing, there are three rigidly ordered steps:

(1) Formulate a statistical test.

(2) Obtain data from a physical test.

(3) Use the data and the statistical test to either reject or not reject the null hypothesis.

The first step is called the design of the experiment; this statistical design procedure consists of six sub-steps which are performed by the statistician in consultation with the manager:

(1) State the null hypothesis.

(2) State the alternate hypothesis.

(3) Select a value for alpha.

(4) Determine a rejection rule form the appropriate statistical distribution function which describes the parameter being tested.

(5) Select a value for beta from some particular non-specification value of the parameter being tested.

(6) Determine the number of measurements that must be made on the parameter being tested.

In actual practice, there are two common modifications during the formulation of the statistical test. A range of non-specification values of the parameter being tested is normally considered in the determination of the number of measurements necessary to achieve the selected beta. This procedure requires the use of engineering judgment to determine the range of values on the power curve. A modification of both alpha and beta is often necessary to formulate the statistical test with a reasonable and achievable number of measurements in the physical test.

In classical hypothesis testing, the physical test is performed independently of the formulated statistical test; no analysis of physical test data is made until all measurements are made. As hypothesis testing is actually practiced, the data is often modified before the final analysis; that is, outliers in the data are often discarded by a process which may be systematic or logical.

In traditional hypothesis testing, the decision to reject or not reject the null hypothesis is based on the retained data and the decision rule from the statistical test. Once the decision is made to reject or not reject, the process is complete. No consideration is given to the margin of passing or failing and no questions are considered concerning possible modifications of the statistical test.

MODIFICATION OF ALPHA

The application of hypothesis testing using the traditional and rigid method outlined in the previous section has a major practical obstacle. It provides managers with no guidance other than to reject or fail to reject the null hypothesis. While this might seem sufficient to a statistician who is lead to believe that the statistical test properly considers all aspects of the proposed equipment or process, the manager may well find this "go - no go" guidance sorely lacking. The manager may have to consider the effects of more than one parameter, the political influence of pressure groups, and the economic impact of implementing the new equipment or process. A manager facing these problems needs all the help that the statistician can provide. While the statistician cannot solve the manager's problems, he/she can at least provide more information about one parameter than a bare "go - no go" recommendation. This is done by a modification of alpha based on the data from the physical test.

The p-value of a hypothesis test is defined as the value that alpha would have been necessary, in the original statistical test, to make the result of the decision rule indecisive. It is the value of alpha which would act as a pivot between rejecting and not rejecting the null hypothesis; it is calculated using the data from the physical test. It provides the manager, who must consider all factors and make the final decision, with a measure of the degree of certainty of the bare "go - no go" recommendation from the hypothesis test. If a manager wishes to make a final decision which is the opposite of the result of using data from the physical test in the decision rule of the statistical test, the p-value provides him/her with a measure of the risk that must be taken to overrule the judgment of the hypothesis test.

The existence of the p-value is certainly not new; however, its usage has increased in recent years because managers have needed more flexibility. Reporting the p-value to managers, who make final decisions, has the effect of loosening the rigor of the formal hypothesis testing. The p-value provides the final decision maker with more flexibility; it also tells him/her how flexible the result of the hypothesis test is. Very low values of the p-value imply that the null hypothesis can be rejected only if the producer is allowed a very low risk. Very high values of the p-value imply that the null hypothesis cannot be rejected unless the producer is forced to take a very high risk. Values of the p-value that are near the alpha of the original statistical test imply that the null hypothesis can be rejected if the producer takes a risk near the original alpha.

PROPOSED MODIFICATION OF BETA

The modification of alpha as outlined in the previous section is well established. In this section, modifications to beta and the power curve are proposed to further increase the flexibility and reported information of hypothesis testing. These proposals are motivated by the desire to aid decision makers as much as possible.

There are two critical parameters which may logically initiate a modification of beta after the standard hypothesis test is completed. One is the p-value; it will differ greatly from the alpha of the original statistical test if the data from the physical test leads to an extermly strong recommendation to either reject or not reject the null hypothesis. The other critical parameter is the number of measurements actually made on the physical parameter being tested. It may be larger than the number proposed by the

statistical test, but the more common change is for it to be smaller. In actual practice, the number of measurements in the physical test is often lowered by constraints of cost, time, and/or limitations of personnel, facilities, and/or equipment.

The two critical parameters result in two logical modifications of beta. These modifications are denoted by β_T and a q-value in this section. They are obtained from the same basic mathematical algorithm that yielded the original beta of the statistical test, but different inputs are used in the algorithm.

The statistic denoted by β_T is defined as the result of the beta calculation in the design of the statistical test when the number of measurements actually made is used instead of the number that was planned. The same alpha, specification value of the parameter being tested, and actual value of the parameter being tested are used in the β_T calculation as were used in the original beta calculation; only the number of measurements is changed. The numerical value of β_T will be higher than the value of the original beta if the number of actual measurements is lower than planned by the statistical test.

The statistic denoted by β_T is the consumer's true risk when the data actually available from the physical test is analyzed. The "sub T" notation is used to emphasize that β_T is based on the true number of measurements in the actual physical test. A manager can use β_T as a measure of how badly a hypothesis test would be damaged if he/she is pressured into changing the number of measurements in the planned physical test. Pressure for such a change can occur between planning and testing or during testing. Naturally the manager is concerned about more than one possible true value of the tested parameter. This means that he/she needs the entire power curve based on β_T.

Of course, this might have been one curve in a family of curves that was used to design the statistical test. Unfortunately, the manager might not have access to or knowledge of this previously calculated curve. The statistic β_T is proposed to rectify the managers' lack of information. The decision maker needs a power curve based on β_T at the time that pressure exists to change the number of measurements. Statistical advice may be needed after the statistical test is formulated but before the data from the completed physical test is analyzed. The statistician should provide the decision maker with relevant values of β_T between the formulation of the statistical test and the analysis of data from the physical test.

A q-value is defined as the result of the beta calculation in the design of the statistical test when two changes are made. A q-value is calculated from the p-value instead of the original alpha and from the number of measurements actually made whether or not this is the number that was planned in the original statistical test. The same specification value and actual value of the parameter being tested are used in a q-value calculation as in the original beta calculation. Since there are many possible values of the parameter being tested, there are many q-values. Use of the same possible parameter value that was used in the beta calculation allows direct comparison between a q-value and beta. A q-value tends to be higher than the original beta if the number of actual measurements is lower that planned or if the p-value is lower than the original alpha. Similarily, more measurements than planned or data making the p-value higher than the original alpha tends to make a q-value lower than the original beta.

A q-value integrates information about both the number of measurements actually made in the physical test and the results of these measurements.

A q-value has three potential uses:

(1) Provide information about beta if alpha is changed to the p-value after the hypothesis test in order to make the test indecisive.

(2) Indicate the need for additional testing to make the test more decisive.

(3) Consider as a rating factor for the payment in procurement cost analysis if the consumer accepts equipment or services that fail the hypothesis test.

If a decision maker has influences which contradict the result of a completed hypothesis test, there are two contrasting situations which are most intelligently considered by using both the p-value and a q-value. If the decision rule results in a recommendation to reject the null hypothesis, alpha must be lowered to the p-value if the hypothesis test is to be viewed as indecisive and the other influences are to have no opposition. In this situation, beta for a particular non-specification value of the parameter being tested must be raised to a q-value. In the opposite situation, the decision rule results in a recommendation to not reject the null hypothesis. In this case, alpha must be raised to the p-value and beta, for a given value of the tested parameter, must be lowered to the q-value in order to make the hypothesis test indecisive so the other influences have no opposition to rejecting the null hypothesis. In both of these situations, the decision maker should be aware of the changes in both alpha and beta. Thus the statistician should report both the p-value and a q-value. Of course, the decision maker is often interested in more than one possible given value of the tested parameter; thus the statistician should report the entire power curve based on q-values.

Increasing the amount of physical testing always reduces the risks in hypothesis testing; the amount of additional testing after a hypothesis test is completed may be determined from the power curve based on q-values and other power curves based on the p-value and proposed increases in the number of measurements of the parameter being tested. If a decision maker is willing to accept the lowering of alpha to the p-value but is hesitant about raising beta to a q-value, more testing is necessary. The amount of additional testing is determined by the consumer's risk that the manager is willing to take. Once that consumer's risk is determined for relevant non-specification values of the parameter being tested, power curves may be used to determine the amount of additional testing. The additional number of measurements which the manager is ordering is the difference between the total number determined by the power curves and the number of measurements in the initial hypothesis test. The statistician should report power curves based on q-values calculated from the p-value and both (1) the number of measurements actually made and (2) proposed increases in the number of measurements. This information allows the manager to evaluate the return from additional testing.

When a piece of equipment or a service is procured, the procuring agency may accept delivery of a product which produces an unexpected result in a hypothesis test; the p-value and a q-value might be used to adjust the payment to the supplier. One standard procedure in procurement actions is based on a rigid hypothesis test. The product is rejected if the decision rule yields a recommendation to reject the null hypothesis that the specification is met. There are two problems with this standard approach. The supplier has no reward if the product is good enough to pass the hypothesis test with a p-value higher than alpha. Also, the procurer must use additional

justifications to accept a product that can be made to satisfy an urgent need even though it is bad enough to fail the hypothesis test with a q-value lower than beta. A proposed solution to these two problems is for the procuring agency to execute the following cost analysis procedure:

(1) Set a specification, for the null hypothesis, which is a value of the tested parameter that allows acceptable but not exceptional operation with the equiment or service.

(2) Set an alternate hypothesis value of the tested parameter, as close to the specification as feasible, which is not acceptable.

(3) Set alpha, beta, and the number of measurements of the tested parameter after balancing the risks and the cost of testing.

(4) Call for bids to set the payment if the hypothesis test yields a p-value equal to alpha and a q-value equal to beta.

(5) Establish a continuous scale of price increases for each p-value that is higher than alpha.

(6) Establish a continuous scale of price decreases for each q-value that is higher than beta.

(7) Obtain a random sample of the equipment or service, perform the hypothesis test, and implement the result of the procedure.

An optional step might be exercised if the supplier or the procuring agency doesn't accept the result:

(8) Do more testing if either the supplier or procuring agency is willing to pay for it in an attempt to raise the p-value or lower a q-value.

Naturally, these steps must be described at the time of the invitation for bids so potential bidders can decide how to respond. Steps 5, 6, and 8 might be topics for negotiation between the procuring agency and potential

suppliers. This proposed procedure is only one possible solution to procurement confrontations. Another procedure might be desired if sufficient testing can establish that the tested parameter is within a narrow interval with a high degree of confidence. This 7 or 8 step procedure should be considered for procurements in which limitation of testing is expected to make clear-cut procurements decisions impractical or impossible.

EXAMPLE

The example presented in Figures One through Eleven illustrates modification of alpha and beta in a classical hypothesis test on the variance of a normally distributed random variable. The null hypothesis H_o for this test is the assumption that the variance is less than or equal to a standard. The chi-square distribution is appropriate for the random variable $(n-1)s^2/\sigma^2$ when n is the number of measurements used to find the sample variance s^2 which estimates the variance σ^2. The number of degrees of freedom ν is given by n-1 for this χ^2 statistic. The numbers in this example have been chosen to depict a hypothetical process-improvement development in the manufacture of glass with a low variance in its index of refraction. Existing manufacturing processes are assumed capable of yielding $\sigma_u^2 = (10)^{-8}$; but this variance is considered unacceptable for the prototype process-improvement which is intended to reach a standard of $\sigma_o^2 = 4\ (10)^{-10}$.

Figures One through Four show information that is useful before and during data taking. Figures One and Two contain the information that a statistician should present to enable a manager to complete the design of the hypothesis test. Figure Three shows the designed hypothesis test after the manager has selected a planned sample size. Figure Four reports the information that the

statistician should present to the manager if pressures exist to change the planned sample size.

Figures Five through Eleven show possible results for different sets of measurements. In Figures Five through Eight, four different values of s^2 are used to find the p-value and a q-value for σ_u^2. For any measurement, the p-value is obtained by setting the calculated value of the distributed statistic equal to the statistic that yields a percentile given by the p-value; that is, alpha is replaced numerically by the p-value. The calculation of a q-value is similar to the calculation of beta in that a non-specification value of the tested parameter must be used; the difference is that beta is a function of alpha while each q-value is a function of the p-value. Individual addition shows that the four sets of p-values and q-values in Figures Five through Eight have sums less than one. Figure Nine, which is not drawn to scale, illustrates graphically that the sum of these paired statistics is always less than one. This occurs, for alpha and beta as well as for the p-value and each associated q-value, because the area under the probability density function $f(\chi^2;\nu)$ is equal to one. Figure Nine also shows that an increase in one of these statistics is always accompanied by a decrease in the other. Measurements leading to a strong recommendation to reject H_0 will yield a low p-value and high q-values; opposite extremes in the p-value and q-values result when measurements strongly imply that H_0 should not be rejected.

Use of the p-value and associated q-values after the hypothesis test, for purposes other than gaging the intensity of the decision to reject or not reject H_0, requires analysis curves. Some of these curves are presented and interpreted in Figures Five, Six, Ten, and Eleven. In Figures Five and Six

where s^2 is close to the critical sample variance s_c^2 which separates rejection and non-rejection of the null hypothesis, power curves are presented for the manager to use in considering additional testing. Also in Figures Five and Six, hypothetical algorithms are presented for adjusting the cost and the results of these algorithms, the p-value, and a q-value for σ_u^2 are given. It must be emphasized that the manager, in consultation with the statistician, should establish these cost adjustment algorithms. The manager may design an algorithm other than a straight line; he/she may also use a maximum p-value p_m and a maximum q-value q_m other than the simple ones used in Figures Five and Six to limit the change in cost caused by results of the hypothesis test. Figure Ten summarizes the numbers of Figures Five through Eight which are all based on a sample size equal to that planned in Figure Three. Figure Eleven presents a similar summary for slightly different actual samples sizes.

CONCLUSION

Modification of alpha and beta makes hypothesis testing less rigorous and allows managers more flexibility. Modification of beta in the interim between the design of the statistical test and the completion of physical measurements allows managers to make informed decisions concerning changes to the proposed number of physical measurements. Modification of alpha and beta after the hypothesis test is completed allows managers to make informed decisions about (1) consolidating the result of the decision rule with factors not considered by the hypothesis test, (2) the value of additional testing, and (3) compensation when specifications can be considered as met with risks different from the original alpha and beta.

Figure One --- Partially Designed Hypothesis Test:

Part 1 --- H_0: $\sigma^2 \leq \sigma_0^2 = 4(10)^{-10}$ $\quad \alpha = P[\text{Reject } H_0 \mid \sigma_0^2] = .01$

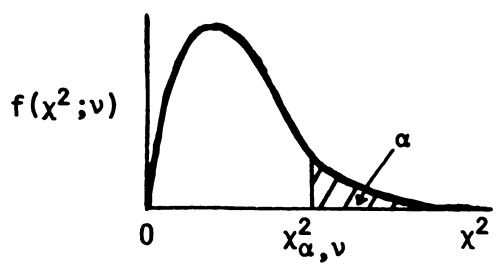
H_a: $\sigma^2 > \sigma_0^2$ $\quad$ Test is One Tailed

Part 2 --- Rejection Region: $\quad \chi^2 > \chi^2_{\alpha,\nu} = \chi_{\alpha,n-1}$

where $\chi^2 = (n-1)s^2/\sigma_0^2$

with n = number of measurements

and s^2 = sample variance



Rejection Measurement: $\chi^2_{\alpha,\nu} = (n-1)s_c^2/\sigma_0^2 \implies s_c^2 = \sigma_0^2\, \chi^2_{\alpha;n-1} / (n-1)$

Figure Two --- Power Curves for Test Design:

π = Power = $P[\text{Reject } H_0 \mid \sigma^2] = P[s^2 > s_c^2]$

$= P[(n-1)s^2/\sigma^2 > \{(n-1)/\sigma^2\}\{\sigma_0^2\, \chi^2_{\alpha,n-1}/(n-1)\}]$

$= P[\chi^2 > \{\sigma_0^2/\sigma^2\}\, \chi^2_{\alpha,n-1}]$

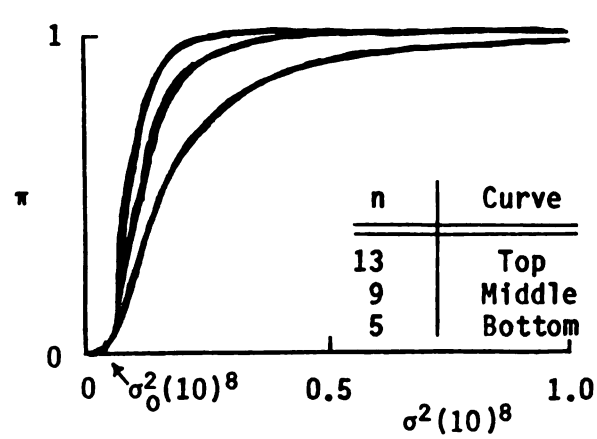


Figure Three --- Hypothesis Designed with n=7:

Part 1 --- H_0: $\sigma^2 \leq \sigma_0^2 = 4(10)^{-10}$ $\quad \alpha = P[\text{Reject } H_0 \mid \sigma_0^2] = .01$

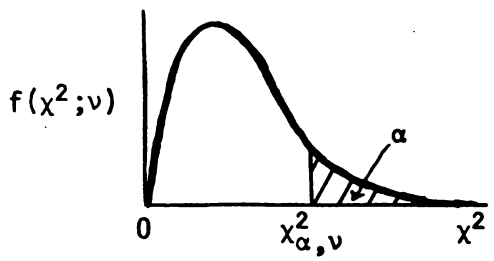
H_a: $\sigma^2 > \sigma_0^2$ $\quad$ Test is One Tailed

Part 2 --- Rejection Region: $\quad \chi^2 > \chi^2_{\alpha,\nu} = \chi^2_{\alpha,n-1} = \chi^2_{.01,6} = 16.8$

where $\chi^2 = (n-1)s^2/\sigma_0^2 = 6s^2/\sigma_0^2$

with n = number of measurements

and s^2 = sample variance

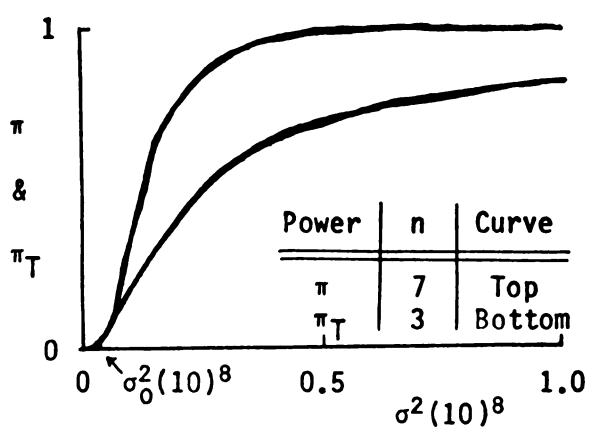


Rejection Measurement: $\chi^2 = \chi^2_{\alpha,\nu} \implies s_c^2 = \sigma_0^2 \chi^2_{.01,6}/6 = 11.2(10)^{-10}$

β Part --- If $\sigma_u^2=(10)^{-8}$; $\beta = P[s^2<s_c^2]=P[(n-1)s^2/\sigma_u^2<\{(n-1)\}/\sigma_u^2\}s_c^2]=P[\chi^2<.672] = .005$

Figure Four --- β_T and Power = $\pi_T = 1 - \beta_T$ Curves if n is Changed to 3:

β_T Part --- If $\sigma_u^2=(10)^{-8}$; $\beta_T = P[s^2<s_c^2] = P[\chi^2<\{\sigma_0^2/\sigma_u^2\}\chi^2_{\alpha,n-1}] = P[\chi^2<.368] = .168$



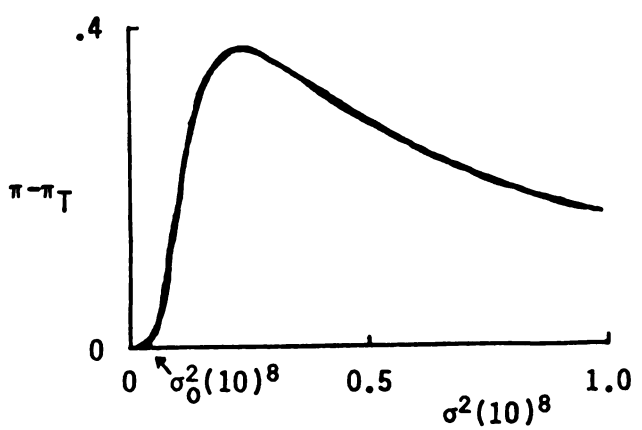


Figure Five --- Inferential Information from Design and Hypothesis Test if $s^2>\sigma_0^2$ and s^2 Implies Rejection of H_0: $\sigma^2 \leq \sigma_0^2 = 4(10)^{-10}$ with $\alpha=.01$:

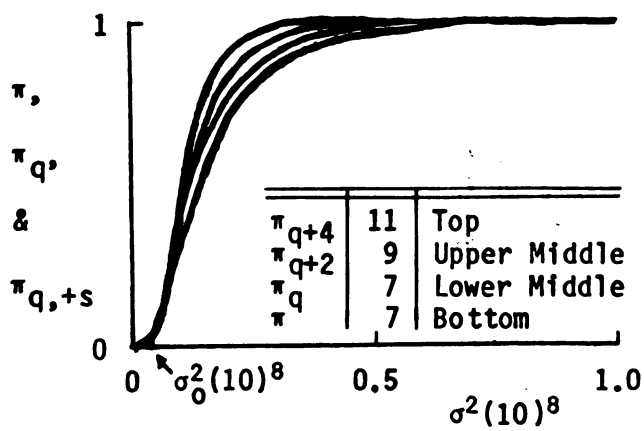
Part 3 --- If $s^2 = 12.3(10)^{-10}$ from n=7 measurements; $\chi^2 = (n-1)s^2/\sigma_0^2 = 18.45$

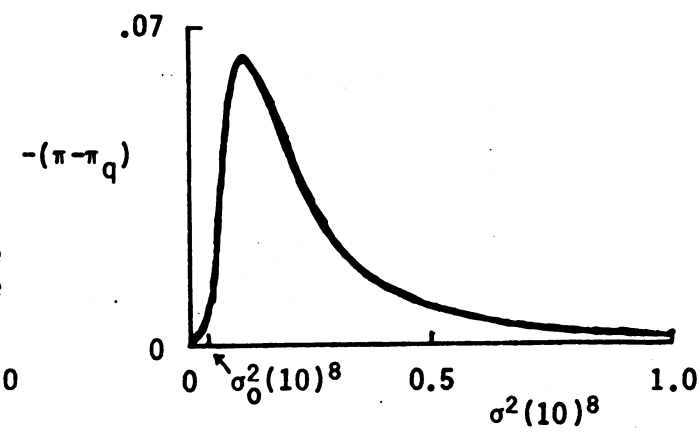
Part 4 --- Since $\chi^2 = 18.45 > 16.81 = \chi^2_{.01,6} = \chi^2_{\alpha,n-1} = \chi^2_{\alpha,\nu}$; Reject H_0

p Part --- To Barely Not Reject H_0; $\chi^2_{p,\nu} = \chi^2_{p,6} = 18.45$ ===> p-value = .005

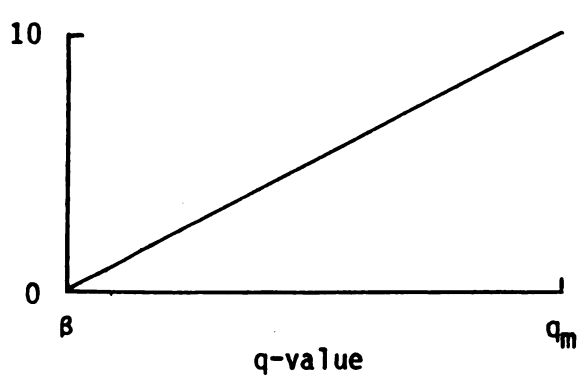
q Part --- If $\sigma_u^2 = (10)^{-8}$; q-value = $1-\pi_q = P[\chi^2<\{\sigma_0^2/\sigma_u^2\}\chi^2_{p,\nu}] = P[\chi^2<.74] = .006$

Power Curves:





Cost Analysis:



Price Decrease in Percentage of Target Cost

Use of the q-value of .006 and the graph to the left yields a price reduction equal to .017% of the target cost

β = .005 from design

$q_m = P[\chi^2<\{\sigma_0^2/\sigma_u^2\}\chi^2_{p,\nu} \mid \chi^2_{p,\nu}=(n-1)s^2/\sigma_0^2$ and $s^2=\sigma_u^2] = P[\chi^2<(n-1)] = P[\chi^2<6] = .58$

Figure Six --- Inferential Information from Designed Hypothesis Test if $s^2 > \sigma_0^2$ and s^2 Does Not Imply Rejection of H_o: $\sigma^2 \leq \sigma_0^2 = 4(10)^{-10}$ with $\alpha = .01$:

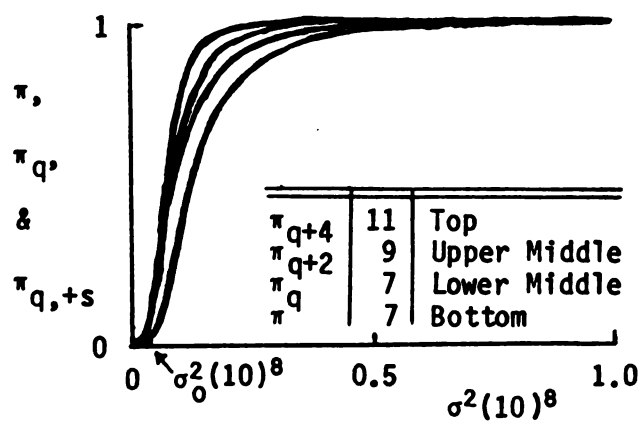
Part 3 --- If $s^2 = 8.4(10)^{-10}$ from n=7 measurements; $\chi^2 = (n-1)s^2/\sigma_0^2 = 12.6$

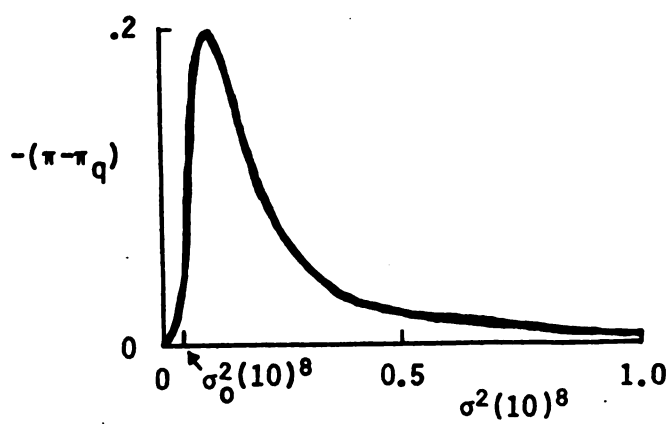
Part 4 --- Since $\chi^2 = 12.6 < 16.81 = \chi^2_{.01,6} = \chi^2_{\alpha,n-1} = \chi^2_{\alpha,\nu}$; Do Not Reject H_o

p Part --- To Barely Reject H_o; $\chi^2_{p,\nu} = \chi^2_{p,6} = 12.6$ ===> p-value = .05

q Part --- If $\sigma_u^2 = (10)^{-8}$; q-value = $1-\pi_q = P[\chi^2 < \{\sigma_0^2/\sigma_u^2\}\chi^2_{p,\nu}] = P[\chi^2 < .50] = .002$

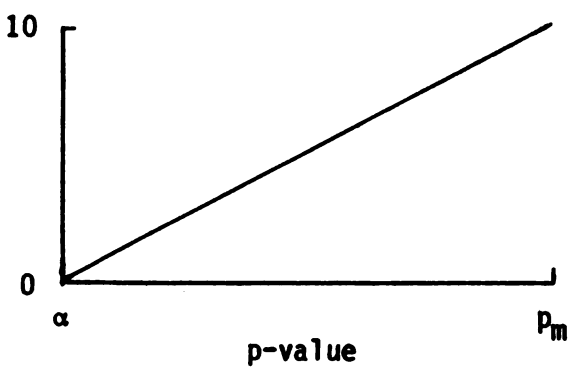
Power Curves:





Cost Analysis:

Price Increase in Percentage of Target Cost



Use of the p-value of .05 and the graph to the left yields a price enhancement equal to .17% of the target cost

$\alpha = .01$ from design

$p_m = P[\chi^2 > \chi^2_{p_m,\nu}] = P[\chi^2 > (n-1)s^2/\sigma_0^2] = P[\chi^2 > (n-1)] = P[\chi^2 > 6] = .42$

Figure Seven --- Inferential Information from Designed Hypothesis Test if $s^2 \geq \sigma_u^2$ So s^2 Forces Management to Use a Quite Low p-value to View the Test of H_o: $\sigma^2 \leq \sigma_o^2 = 4(10)^{-10}$ with $\alpha = .01$ as Indecisive:

Part 3 --- If $s^2 = 22(10)^{-10}$ from n=7 measurements; $\chi^2 = (n-1)s^2/\sigma_o^2 = 33$

Part 4 --- Since $\chi^2 = 33 > 16.81 = \chi^2_{.01,6} = \chi^2_{\alpha,n-1} = \chi^2_{\alpha,\nu}$; Reject H_o

p Part --- To Barely Not Reject H_o; $\chi^2_{p,\nu} = \chi^2_{p,6} = 33$ ===> p-value = .00001

q Part --- If $\sigma_u^2 = (10)^{-8}$; q-value = $1-\pi_q\, P[\chi^2 < \{\sigma_o^2/\sigma_u^2\}\chi^2_{p,\nu}] = P[\chi^2 < 1.32] = .03$

Figure Eight --- Inferential Information from Designed Hypothesis Test if $s^2 \leq \sigma_o^2$ So s^2 Forces Management to Use a Quite High p-value to View the Test of H_o: $\sigma^2 \leq \sigma_o^2 = 4(10)^{-10}$ with $\alpha = .01$ as Indecisive:

Part 3 --- If $s^2 = 2.55\ (10)^{-10}$ from n=7 measurements; $\chi^2 = (n-1)s^2/\sigma_o^2 = 3.825$

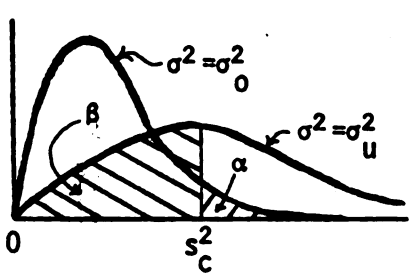
Part 4 --- Since $\chi^2 = 3.825 < 16.81 = \chi^2_{.01,6} = \chi^2_{\alpha,n-1} = \chi^2_{\alpha,\nu}$; Do Not Reject H_o

p Part --- To Barely Reject H_o; $\chi^2_{p,\nu} = \chi^2_{p,6} = 3.825$ ===> p-value = .700

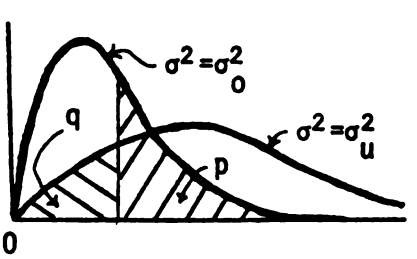
q Part --- If $\sigma_u^2 = (10)^{-8}$; q-value = $P[\chi^2 < \{\sigma_o^2/\sigma_u^2\}\chi^2_{p,\nu}] = P[\chi^2 < .153] = .00007$

Figure Nine --- α & β and p & q on Graphs of s^2 vs $f(\chi^2;\nu)$ with $\nu=n-1$ & $\chi^2=\nu s^2/\sigma^2$:

Pre-Test:



Post-Test with $s^2 < s_c^2$:



Post-Test with $s^2 > s_c^2$:

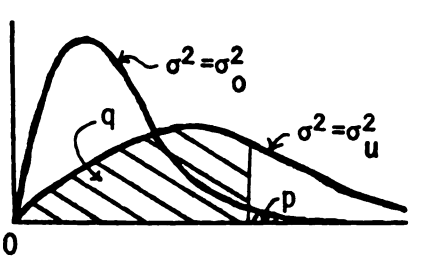


Figure Ten --- Summary of Results for Representative Values of s^2 from $n = 7$ Measurements to Test H_o: $\sigma^2 \leq \sigma_o^2 = 4(10)^{-10}$ with $\alpha = .01$, $\chi^2_{\alpha,n-1} = 16.81$, and $\beta = .005$ for $\sigma^2 = (10)^{-8}$:

s^2:	$22(10)^{-10}$	$12.3(10)^{-10}$	$8.4(10)^{-10}$	$2.55(10)^{-10}$
$\chi^2 = (n-1)\, s^2/\sigma_o^2$:	33 > 16.81	18.45 > 16.81	12.6 > 16.81	3.825 > 16.81
Decision on H_o:	Reject	Reject	Do Not Reject	Do Not Reject
p-value:	.00001<.01=α	.005<.01=α	.05>.01=α	.7>.01=α
q-value for σ_u^2:	.03>.005=β	.006>.005=β	.002<.005=β	.00007<.005=β
One Resultant Change in Cost:	Minus .4%	Minus .017%	Plus 1%	Plus 10% = Max

Figure Eleven --- Summary of Results for Representative Values of s^2 from $n = \begin{cases} 5 \\ 9 \end{cases}$ Measurements to Test H_o: $\sigma^2 \leq \sigma_o^2 = 4(10)^{-10}$ with $\alpha = .01$, $\chi^2_{\alpha,n-1} = \begin{cases} 13.28 \\ 20.09 \end{cases}$, and $\beta = \begin{cases} .03 \\ .0008 \end{cases}$ for $\sigma_u^2 = (10)^{-8}$:

s^2:	$22(10)^{-10}$	$12.3(10)^{-10}$	$8.4(10)^{-10}$	$2.55(10)^{-10}$
$\chi^2 = (n-1)s^2/\sigma_o^2$:	22 > 13.28 44 > 20.09	12.3 < 13.28 24.6 > 20.09	8.4 < 13.28 16.8 < 20.09	2.55 < 13.28 5.1 < 20.09
Decision on H_o:	Reject Reject	Do Not Reject Reject	Do Not Reject Do Not Reject	Do Not Reject Do Not Reject
p-value:	.0002 <.01 .000001<.01	.02 >.01 .002<.01	.08>.01 .03>.01	.6>.01 .8>.01
q-value for σ_u^2:	.07>.03 .01>.0008	.026<.03 .002>.0008	.01 <.03 .0004<.0008	.001<.03 .00001<.0008
One Resultant Change in Cost:	Minus .7% Minus .2%	Plus .02% Minus .02%	Plus 2% Plus .5%	Plus 10% = Max Plus 10% = Max

FORCE DEVELOPMENT TESTING AND EXPERIMENTATION OF THE FIRE SUPPORT TEAM

Jock O. Grynovicki
Ann E. McKaig
Jill H. Smith
Virginia A. Kaste
Experimental Design & Analysis Branch
System Engineering & Concepts Analysis Division
U.S. Army Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland

ABSTRACT In April and May, 1984, The Field Artillery Board, Ft. Sill, OK conducted a Force Development Test and Experimentation (FDT&E) of the Fire Support Team (FIST) concept at Ft. Riley, KS. The purpose of the FDT&E was to test and evaluate the effectiveness of the FIST HQ equipped with FIST vehicles and digital communications equipment under various tactical configurations, selected modes of operation and personnel shortages. Although traditional manual data collection methods employing human observers was used to record test data, a new automatic data recording technique based on the Artillery Control Environment (ACE) technology was used for the first time in the field. Personnel from the Ballistic Research Laboratory (BRL) assisted in the experimental design, and were responsible for the designing, coding and testing the computer software for the data collection and reduction system.

The discussions will focus on the experimental design, data reduction methodology, the methods of analysis employed and a brief reporting of the results.

I. INTRODUCTION

A. Background

During April and May 1984, The Field Artillery Board, Ft. Sill, OK conducted a Force Development Testing and Experimentation (FDT&E) of the Fire Support Team Headquarters (FIST HQ) concept at Ft. Riley, KS. The test consisted of three iterations of a 120-hour Scenario Oriented Recurring Evaluation System (SCORES) field exercise that was based upon and included the mechanized infantry and armor defensive maneuvers. The task force was confronted by an opposing force (OPFOR) of various strengths and a jamming team. All elements were strictly controlled by the test directorate during the first two exercises. The third exercise was a freeplay, uncontrolled force on force exercise.

Personnel from the Ballistic Research Laboratory (BRL) designed the experiment and assisted in the implementation of experimental design methodology in the controlled segment of the test to address a subset of the overall objectives. In addition, personnel from the BRL were responsible for designing, coding and testing a new automatic data

recording and reduction system based on the Artillery Control Environment (ACE) technology. This report will focus on the experimental design, data reduction and recording methodology, the methods of analysis employed and a brief discussion of the results. The analysis for this report is based upon data obtained from the Field Artillery Board, Ft. Sill, OK.

B. Purpose

The overall purpose of the FDT&E was to evaluate the operational effectiveness of the FIST HQ equipped with a Fire Support Team vehicle (FIST V) and digital communications. Test results will be used by the United States Army Field Artillery School (USAFAS) to further develop FIST operational and organizational concepts.

To demonstrate this effectiveness, a study of the FIST HQ ability to perform fire support coordination under two modes of Forward Observer (FO) message control and four types of FIST HQ configuration, while under various workloads, was conducted.

II. TEST CONCEPT

A. Objectives

1) To determine whether or not FIST DMD message control of the FO's, in the *review* and *automatic* communication modes of operation, has an effect on the FIST HQ ability to perform fire support coordination.

2) To determine if the FIST HQ can perform fire support coordination: a) with the Ground/Vehicle Laser Location Designator (G/VLLD) mounted on the FIST V with all FIST personnel present, b) with the G/VLLD mounted without the FIST Chief, c) inside the FIST V in a Buttoned-up environment and d) with the G/VLLD dismounted from the FIST V. To dismount the G/VLLD from the FIST V, two FIST HQ personnel must dismount both the G/VLLD and its associated equipment.

3) To determine if mission workload affects the FIST HQ performance of fire support coordination. Mission workload was defined as the number and types of missions the FIST HQ was required to process simultaneously. There were four fire mission types: missions initiated from the mechanized infantry FO, missions initiated from the armor platoon leader (Armor) by voice, FIST HQ shooting COPPERHEAD munitions, and missions from the FIST HQ shooting conventional munitions.

B. Measure of Performance

A measure of performance (MOP) is a response that is used to quantify the effects of the factors to be evaluated. For FIST initiated missions, it was defined as the elapsed time from target acquisition until the (fire request) message is transmitted from the FIST Digital Message Device (DMD). Service time for armor missions was the time

from receipt of a fire request message by the FIST HQ until transmission. The time to service FO missions was the elapsed time from when the acknowledgment (ACK) is sent from the FIST DMD indicating receipt of a fire request message until the message is transmitted. This measure indicates the combined time a message spends in the FIST DMD message queue, and the processing and decision time of the FIST HQ.

C. Scope

The first two field exercises (FEX 1, FEX 2), which were a combination of Live Fire and Force on Force, utilized three FIST HQ and one combat observation lasing team (COLT) attached to a mechanized infantry task force that consisted of two mechanized infantry companies and one armor company. (See Figure 1).

The FIST HQ consisted of:

1. The Fire Support Team Chief

2. The Fire Support Team Sergeant

3. Two radio telephone operators

All members of the FIST HQ were trained in the operation of the FIST DMD. Nine weeks of individual training was conducted and validated by the USAFAS. This individual training was followed by two weeks of collective training.

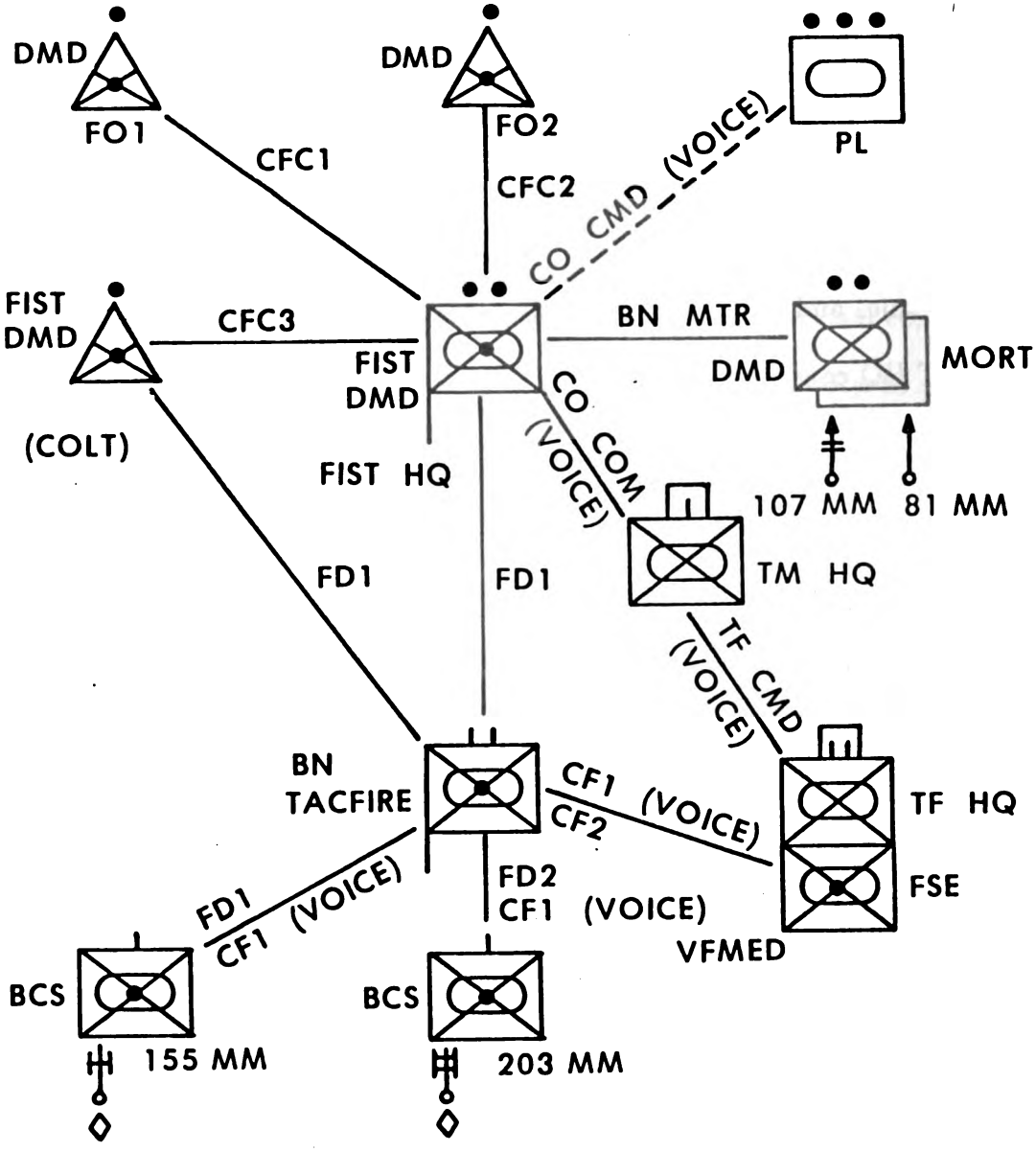
D. Limitations

After receiving the initial fire request message from a FO and deciding how the fire request should be handled, the FIST HQ routed all subsequent messages for that fire mission through the FIST DMD in the automatic "mission mode." That is, all subsequent messages for that fire mission were automatically routed through the FIST DMD. Operator intervention was needed only if a message did not get acknowledged in four transmissions.

2) Electronic Warfare was prohibited during the controlled portion of the FDT&E.

3) Range regulations at Fort Riley prevented the G/VLLD from being employed in a totally realistic environment. Laser designation and range finding were allowed in only two locations and even then had to be restricted.

4) The control cells that contained the Buttoned-up configuration were run at night.



COMPANY FIRE CONTROL NET (CFC)
BATTALION MORTAR FIRE DIRECTION NET (BN MTR)
FIELD ARTILLERY COMMAND FIRE NETS (CF1, CF2)
FIELD ARTILLERY FIRE DIRECTION NETS (FD1 FD2)
COMPANY COMMAND NET (CO CMD)
TASK FORCE COMMAND NET (TF CMD)

NOTE: THE TF WAS COMPOSED OF THREE SIMILARLY ORGANIZED COMPANY TEAMS

Figure 1. Fire Support Structure.

E. Data Collection

In addition to manual data collection methods employing human observers to record test data, a new automatic data recording technique based on the ACE technology was used for the first time in a field exercise. The procedure consisted of recording digital radio traffic time coded on analog magnetic tape. Every 24 hours the tapes were shipped to Aberdeen Proving Ground (APG), MD where HEL personnel received the tapes and played them back into the computer controlled message collection and reduction system. The resulting sorted list of messages was then written to digital magnetic tapes and shipped to Ft. Sill for analysis.

III. MESSAGE COLLECTION AND REDUCTION SYSTEM The major components of the message collection and reduction system were

1) Bit Boxes (Tactical Communication Modems, TCM)

2) VAX 11/750 Computer

3) BRL VAX Unix Operating System

4) Message collection and reduction software

A. Hardware

Bit Boxes are microprocessor based modems which enable Tactical Fire Direction System (TACFIRE) hardware to communicate with commercial computers. The Bit Boxes convert Frequency Shift Keyed (FSK) variable format and fixed format TACFIRE messages (from wire line or radio) to RS232 ASCII character format which commercial computers can accept, and visa versa.

A DEC VAX 11/750 computer was available for use as the main computer to support the message collection and reduction software. The computer operating system was a BRL enhanced version of 4.2 BSD (Berkley System Distribution) Unix.

B. Software

The application software, which was written in the C programming language, had two primary tasks: 1) message collection, and 2) message reduction.

The message collection program receives streams of characters from the Bit Boxes, separates the streams into complete messages, records the start and end time of each message, and stores this information in a computer file.

The data reduction program reads the data files created by the message collection program. The purpose of this program is to sort the messages into fire missions. The result is 3 other files that contain (1) a list of messages categorized by fire mission target number, (2) a list of messages believed to be associated with a fire mission but for some

reason that mission could not be identified, (3) and a list of messages that are known but not part of a fire mission. These lists of messages were shipped to Ft. Sill and combined with manual data to create a comprehensive data base for analysis. For an indepth description of the message collection and reduction system see "Field Artillery Digital Message Collection and Reduction Software," BRL-IMR-822, June 1984.

IV. EXPERIMENTAL DESIGN

A. Factors

The three factors that were tested during the controlled portion of the FDT&E were FIST Configuration, Mode of FIST DMD Control and Mission Workload.

1) FIST Employment Configuration alternatives were:

a) G/VLLD mounted - all hatches on the FIST V were open and the G/VLLD was mounted with the entire FIST HQ present.

b) G/VLLD mounted without FIST Chief- all hatches on the FIST V were open and the G/VLLD was mounted with the FIST Chief not available.

c) G/VLLD dismounted - the G/VLLD was placed away from the vehicle along with two of the four FIST HQ members.

d) Buttoned-Up - all hatches on the FIST V were closed and the G/VLLD was mounted with the entire FIST HQ present.

2) Mode of FIST DMD Control

a) Review - FIST DMD stops all initial fire request messages from platoon FO's for the FIST HQ to review.

b) Automatic - FIST DMD immediately forwards all initial fire request messages with out action by the FIST HQ.

3) Mission Workload

Mission workload was defined as the number and types of fire missions the FIST HQ were required to process simultaneously. The four types of fire missions were:

1) CONV - FIST HQ shooting a conventional munition

2) CPH - FIST HQ shooting a COPPERHEAD munition

3) ARMOR - Missions initiated by the armor platoon leader and

received by voice at the FIST HQ.

4) FO - Missions initiated by the mechanized infantry FO and transmitted digitally to the FIST HQ.

Based on seven combinations of mission types, thirteen categories of mission workloads were defined. They are:

TABLE 1. MISSION WORKLOAD

CATEGORY	FIRE MISSIONS PROCESSED SIMULTANEOUSLY	MISSION TYPE
a	1 CPH	CPH
b	1 FO + 1 CONV	CONV
c	1 FO + 1 CONV	FO
d	1 ARMOR + 1 CONV	CONV
e	1 ARMOR + 1 CONV	ARMOR
f	1 FO + 1 CPH	CPH
g	1 FO + 1 CPH	FO
h	1 FO + 1 ARMOR + 1 CPH	CPH
i	1 FO + 1 ARMOR + 1 CPH	ARMOR
j	1 FO + 1 ARMOR + 1 CPH	FO
k	2 FOs	FO
l	1 ARMOR + 2 FOs	FO
m	1 ARMOR + 2 FOs	ARMOR

B. Design Matrix

It was decided that the smallest period of time reasonable to test any one of the treatment combinations was two hours. A factorial design was constructed with each experimental combination being tested in a random order. This scheme assured that the effect of each of the experimental combinations on the FIST HQ ability to perform fire support coordination could be measured. The FIST HQ were tested under all of the experimental combinations and the design was repeated for each of the two controlled iterations of the FDT&E. The design matrix is presented in Table 2.

V. STATISTICAL ANALYSIS The analysis for this section is based on data reduced by the Field Artillery Board, Ft. Sill, OK, which was a combination of manual data collected by human observers and digital data that was sorted by the message collection and reduction system. This section is intended to be a supplement to the data analysis conducted by the Field Artillery Board and only focuses on several key factors and their associated levels. Unfortunately, the Buttoned-up level of the FIST Employment Configuration factor was not available in this subset of the FDT&E data base, but will be analyzed in a future BRL report.

TABLE 2. DESIGN MATRIX

MISSION WORKLOAD CATEGORIES		REVIEW MODE				AUTOMATIC MODE			
		WO CHIEF MNTD	100% FIST MNTD	100% FIST BUTTONED UP	100% FIST DMNTD	WO CHIEF MNTD	100% FIST MNTD	100% FIST BUTTONED UP	100% FIST DMNTD
1 CHD	CHD								
1 FO + 1 CONV	CONV								
1 FO + 1 CONV	FO								
1 ARMOR + 1 CONV	CONV								
1 ARMOR + 1 CONV	ARMOR								
1 FO + 1 CHD	CHD								
1 FO + 1 CHD	FO								
1 FO + 1 ARMOR + 1 CHD	CHD								
1 FO + 1 ARMOR + 1 CHD	ARMOR								
1 FO + ARMOR + CHD	FO								
2 FOs	FO								
1 ARMOR + 2FOs	FO								
1 ARMOR + 2 FOs	ARMOR								

A. Transformation

As the data was being checked for completeness, it was noted that the distribution of service time was skewed and that the variances of the observations under various experimental conditions were different. Further investigation of the data revealed a positive correlation between the cell standard deviations and the cell means. Correlation between the standard deviations and cell means is often accompanied by marked non-normality and non-homogeneity of variance, and indicates that the particular form of the original observations is unsuitable for Analysis of Variance (ANOVA) procedures.

However, a transformation can be determined which makes the standard deviation independent of the mean, corrects non-homogeneity and also results in the observations being distributed more normally. In general, if a significant functional relationship between the standard deviations and the group means can be determined, then the transformation is the integral of the reciprocal of this functional relationship. Using this procedure, the following transformation was developed:

$$1.7 \ln (18.9 + .56 \text{ (service time)})$$

The transformed data became more normal and the homogeneity of variance among the experimental conditions was improved.

B. Analysis Of Variance

An analysis of variance procedure was performed on the transformed data with one slight modification to this procedure due to unequal experimental group sizes. The sum of squares for all terms in the model, except the error term, was weighted by the harmonic mean. The ANOVA is presented in Table 3. A star next to the F-ratio indicates the factor is significant at the alpha level of .05. Since this analysis assumes a fixed effects model, the denominator for all F-ratios is the pooled error term.

Since the ANOVA was performed on the transformed data, it was decided that comparisons of medians, calculated on observed service times, would be more meaningful than comparing transformed means.

C. Results

The most significant term in the analysis was mission workload. One reason for this significance is that it took substantially less time to service fire request messages from mechanized infantry FO missions than either the FIST HQ missions (COPPERHEAD or Conventional) or Armor missions. In both FIST DMD control modes, the FIST HQ initiated fire request messages require data input, review, and transmittal. Armor messages, which are received by voice, must be reviewed and input as digital messages by the FIST HQ; whereas the digital FO fire requests require only review and transmittal in the review mode of FO control and no processing at all in the automatic mode.

TABLE 3. ANALYSIS OF VARIANCE
(SERVICE TIME)

SOURCE	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F RATIO
Mission Workload	12	101.00	8.42	10.60*
Mode	1	5.65	5.65	7.14*
Configuration	2	0.025	0.01	< 1
Mission Workload x Mode	12	9.21	0.77	< 1
Mission Workload x Configuration	24	13.43	0.56	< 1
Configuration x Mode	2	0.08	0.04	< 1
Mission Workload x Mode x Configuration	24	12.33	0.51	< 1
Pooled Error	461	365.90	0.79	

Another interesting result observed was that in mission combinations in which Armor missions were processed, Armor missions had a longer service time than any other mission type. This trend seems to indicate that it takes the FIST HQ longer to process voice initiated fire request messages than to initiate his own or service FO missions. This result is not surprising since it takes longer to input a message manually than to receive one digitally. These trends were consistent in both the Automatic and Review modes as shown in Table 4.

The number of missions processed simultaneously also affected FIST HQ service time. In plotting the median service time for the mechanized infantry FO fire missions in review mode (See Figure 2), one can see that it takes the FIST HQ longer to service FO missions when the FIST HQ are also initiating COPPERHEAD missions and receiving a armor message than when the FIST HQ are just servicing FO missions and shooting COPPERHEAD. In addition, the FIST HQ service time for FO fire request messages is shorter when they are also initiating a conventional mission as opposed to also shooting COPPERHEAD. This result is not surprising. When the FIST HQ is initiating a COPPERHEAD mission while in the review mode, the FIST DMD operator functions are disabled after sending a FO Command (Fire) or a Fire Request Quick Message and no action can be taken by the FIST DMD operator until the X button is

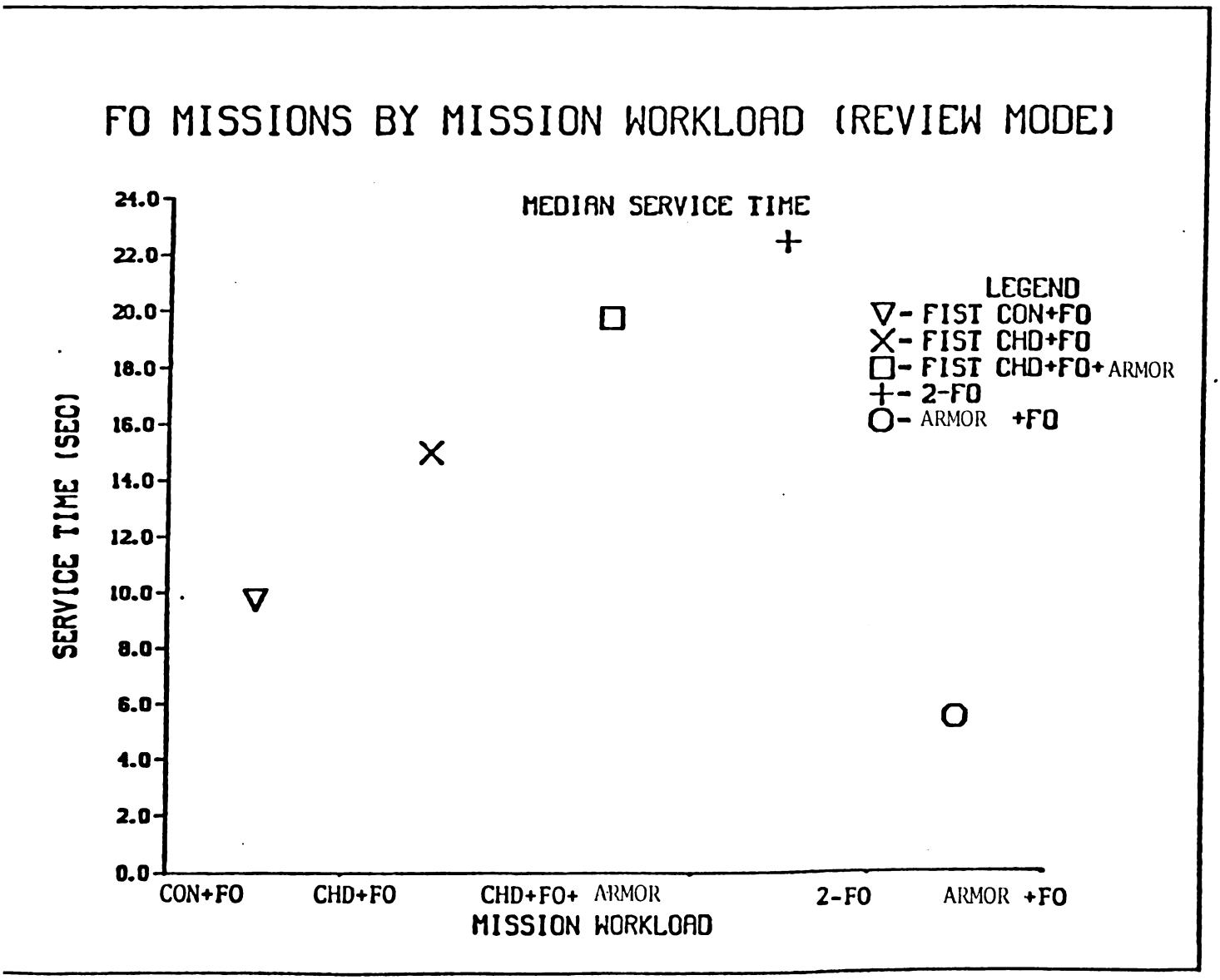


Figure 2. Mechanized Infantry FO Initiated Messages By Workload.

TABLE 4. MISSION WORKLOAD BY MODE
(MEDIAN SERVICE TIME)
(SECONDS)

FIRE MISSION TYPE	MISSION WORKLOAD	Mode	
		Review	Auto
CPH	CPH	55	35
CPH	FO + CPH	11	44
CPH	FO + ARMOR + CPH	6	34
CONV	FO + CONV	58	76
CONV	ARMOR + CONV	77	74
FO	FO + CONV	10	2
FO	FO + CPH	15	2
FO	FO + ARMOR + CPH	20	2
FO	2 FOs	22	2
FO	2 FOs + ARMOR	5	2
ARMOR	ARMOR + CONV	79	98
ARMOR	FO + ARMOR + CPH	81	69
ARMOR	ARMOR + 2 FOs	58	37

pressed to end the COPPERHEAD mission. Surprisingly, the FIST HQ spent the longest time servicing mechanized infantry FO missions in review mode, when they were not initiating or reviewing any other mission types. In this mission workload (2-FOs), the FIST HQ only responsibility was to review the two messages received from his mechanized infantry FOs. FIST personnel spent a lot of time reviewing, changing, and deciding if the initial fire request message should be sent to TACFIRE or to one of their local resources, such as the battalion mortar platoon.

From Figure 3, one can see that it took more time for the FIST HQ to service armor messages when they were also shooting a COPPERHEAD or conventional mission than when they were only reviewing a mechanized infantry FO messages and serving armor missions. This trend is consistent with both the automatic and review FIST DMD control mode.

In the automatic control mode, all initial fire request messages received by the FIST HQ are automatically forwarded to their destination. Initiating messages in the review mode must be passed by the FIST DMD operator before they can be transmitted. Therefore, one could expect the FIST DMD mode of control to significantly affect the time it takes to service digital fire request messages. The ANOVA table revealed that the Mode of FIST DMD Control was significant.

The percent of all messages processed by service time in the automatic and review modes are shown in Figures 4 and 5, respectively. The median service time for the automatic mode was small, 7.0 seconds, when compared to the median service time of 29.0 seconds to service messages in the review mode. For mechanized infantry FO missions, the median service time in the review mode ranged between 5.0 and 22.0 seconds over all workloads. However, in the automatic mode, the median service time for all workloads was 2.0 seconds. This trend was not as prevalent for messages initiated by the FIST HQ or messages received by voice from the armor as depicted in Table 4.

It is worth noting that FIST Employment Configuration was not statistically significant. The FIST's ability to service FO and ARMOR missions as well as initiate his own missions was not affected by the various configurations. This implies that the FIST can perform efficient fire support coordination when the FIST Chief is not present or when two members of the FIST are not available (due to the G/VLLD being dismounted). However, this infers nothing about the quality of the decision being made.

One puzzling result was that the median service time for a FIST HQ to service COPPERHEAD missions while in review mode and for mission workload (FO + ARMOR + CPH) was only 6.0 seconds and only 11.0 seconds for workload (FIST FO + CPH). Looking at the service time distribution for these two categories, one notes a bimodal distribution which may indicate the presence of a lurking variable.

Cluster analysis was used to try to categorize the COPPERHEAD missions into two groups: This is a multivariate statistical technique in which COPPERHEAD missions were separated into groups based on the minimization of variance within groups and the minimization of the distance between groups. A difference in values among groups from different COPPERHEAD missions is said to exist if the hypothesis of equality of means among groups is rejected by an F-test with a significance level of .05. The number of groups in which to categorize the COPPERHEAD missions was not specified.

Using cluster analysis on the COPPERHEAD mission service time, two populations were identified. One group had a median service time of 8.0 seconds and a range between 1.0 and 32.0 seconds. The other group centered at 56.0 seconds and ranged

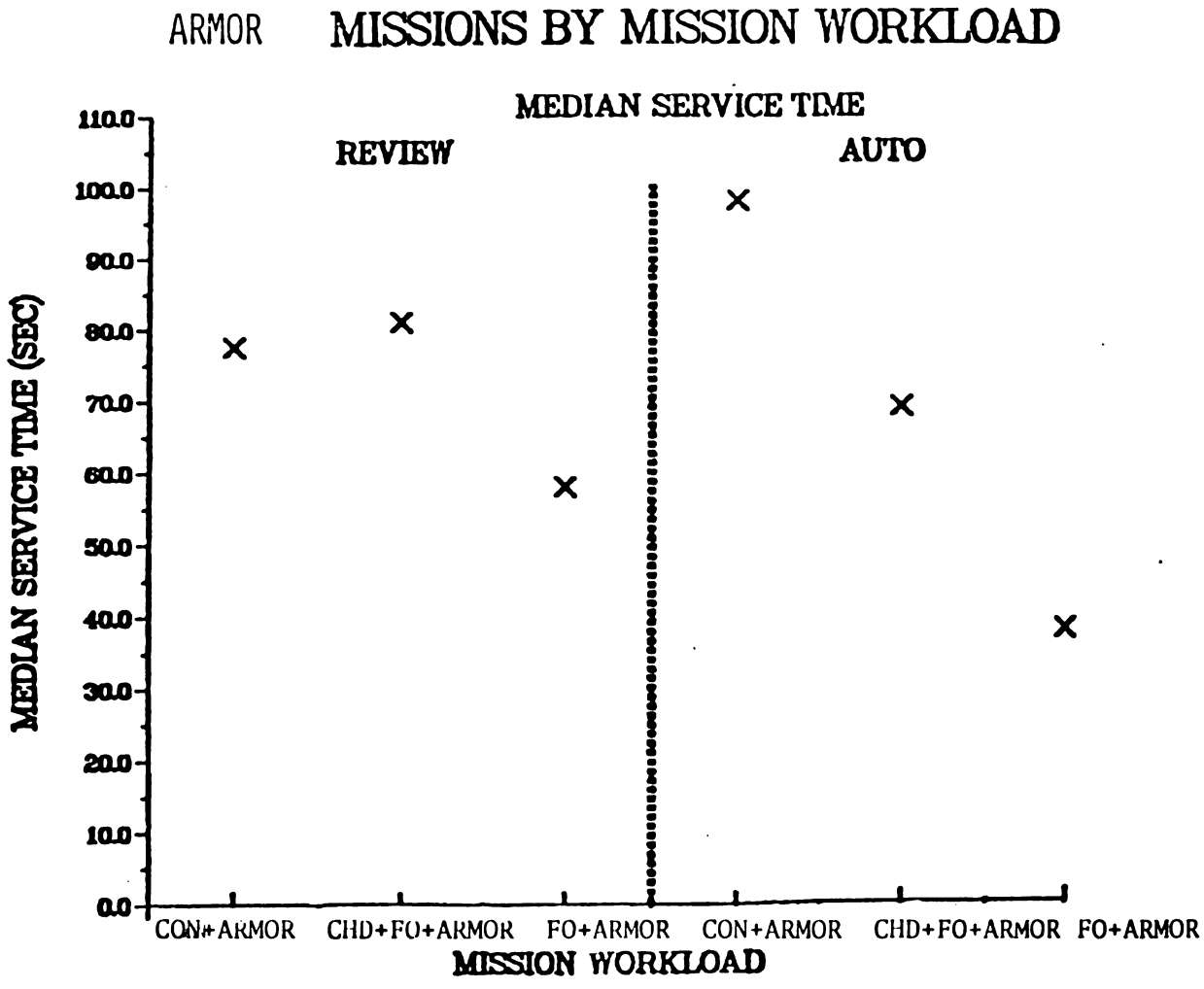


Figure 3. Armor Initiated Messages by Workload.

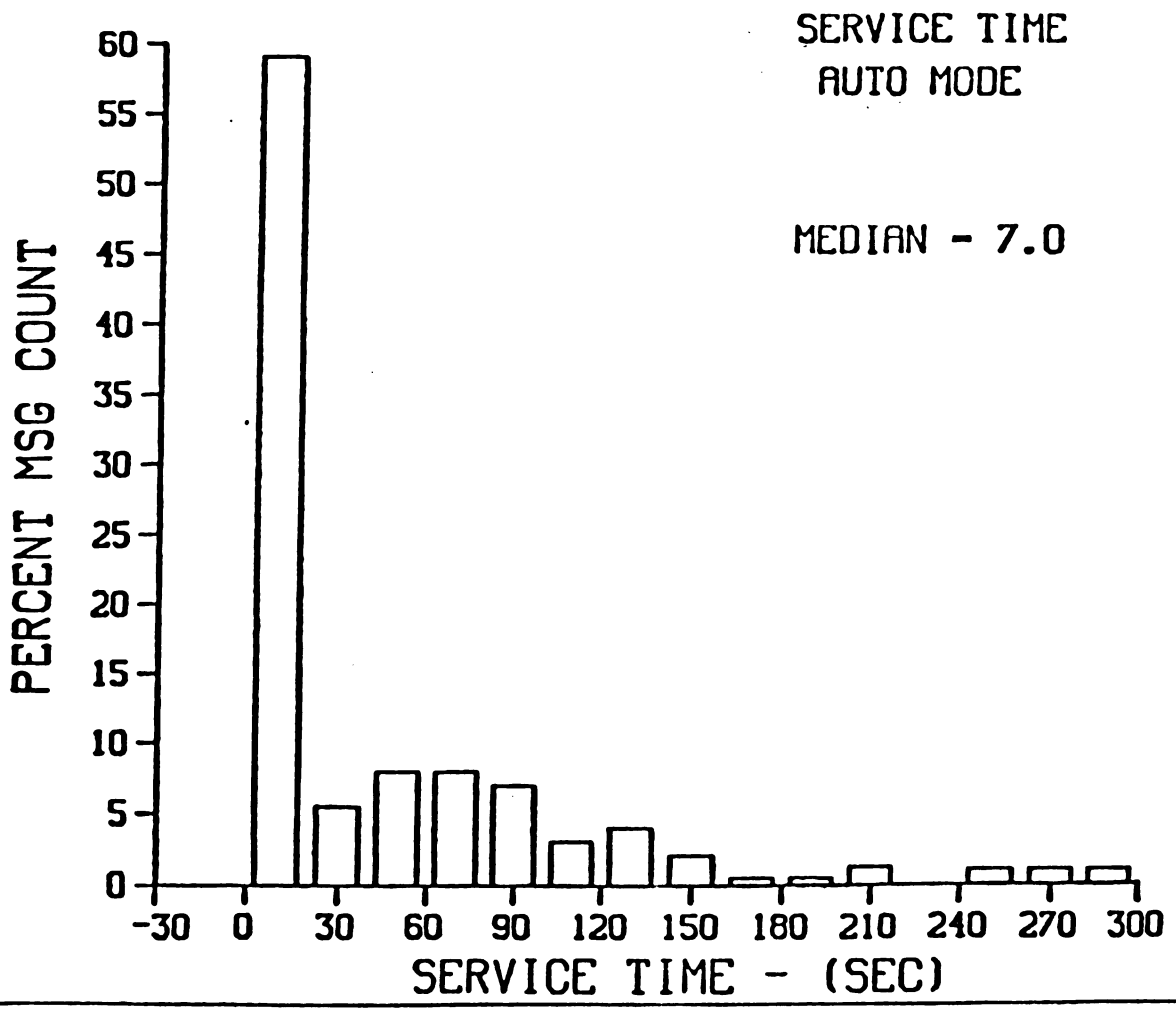


Figure 4. Percent of Messages Processed By Service Time in Automatic Mode.

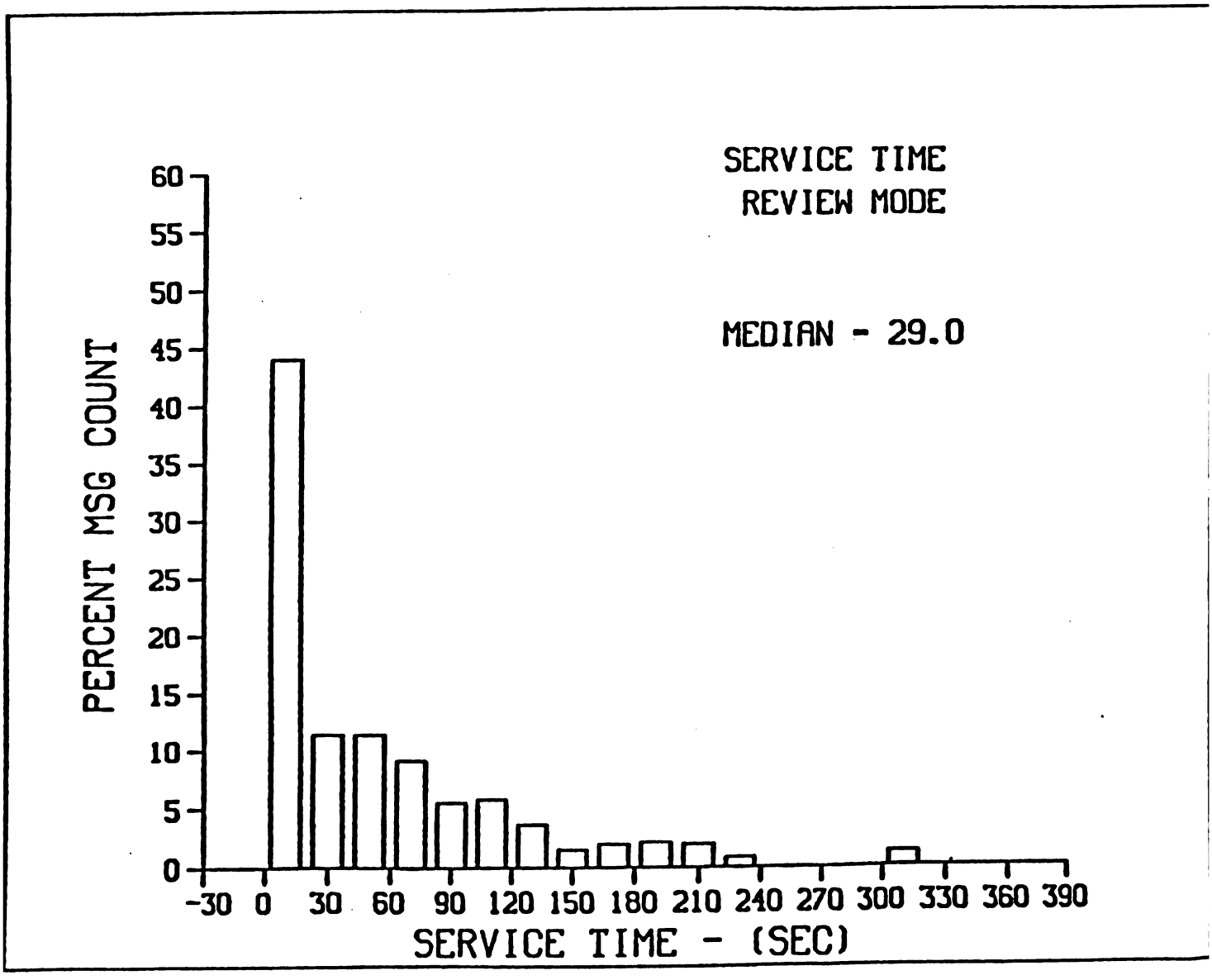


Figure 5. Percent of Messages Processed By Service Time in Review Mode.

between 34.0 and 190.0 seconds. The groups were statistically different at a significance level of .05. Based on the above categorization scheme, the analysis was redone with the two types of COPPERHEAD missions. This resulted in workload having sixteen categories. The conclusions remained unchanged from the original analysis. Mission Workload and Mode of FIST DMD Control were the only factors determined to significantly affect FIST service time especially in regard to FO missions. The median service time for the two groups of COPPERHEAD missions by Mode of FIST DMD Control and Mission Workload are given in Table 5. The median service time for the group with the smaller median service time ranged between 6.0 and 9.0 seconds while the second group ranged between 45.0 and 92.0 seconds. No statistical differences were found between the review and automatic modes of FIST DMD Control for either group. Similarly, Mission Workload had no effect on either category of COPPERHEAD missions.

There are several possible reasons as to why there are two categories of COPPERHEAD mission service time. One reason is that terrain conditions will strongly influence COPPERHEAD service time. Another reason is that there are two types of COPPERHEAD missions (priority and target of opportunity) and the data from Ft. Sill did not categorize these two types. Priority COPPERHEAD missions are preplanned missions with preassigned targets. The mission data is stored until the target appears; the mission is then reactivated and carried to its conclusion. A target of opportunity mission is not a planned mission but occurs when a target appears at an opportune time and place. Target of opportunity missions require a longer processing time by the FIST than priority COPPERHEAD missions once the target is acquired.

TABLE 5. COPPERHEAD MISSIONS
(REVIEW & AUTOMATIC)
(MEDIAN SERVICE TIME)

MODE	MISSION WORKLOAD	GROUP 1	GROUP 2
REVIEW	CPH	7.0	73.0
	FO + CPH	7.0	92.0
	FO + ARMOR +CPH	6.0	45.0
AUTOMATIC	CPH	7.0	68.5
	FO + CPH	8.0	70.0
	FO + ARMOR + CPH	9.0	66.0

VI. CONCLUSIONS Based on the results of the analysis of the limited database obtained from the Field Artillery Board, the FIST HQ demonstrated its ability to perform fire support coordination. The FIST HQ ability to service fire missions was not affected by different FIST HQ configurations. The FIST did perform efficient fire support coordinations when the FIST Chief was not present and when two members of the FIST were not available because the G/VLLD was dismounted. Although Mission Workload and Mode of FO Control were significant, the largest median service time observed was only 98.0 seconds. This occurred when the FIST HQ had to input the voice messages from the Armor.

The number and types of missions processed simultaneously influenced the FIST HQ ability to service FO and Armor missions. However, Mission Workload did not affect the two types of COPPERHEAD missions that were categorized using cluster analysis. Based on this statistical technique, COPPERHEAD missions were shown to not be affected by the FIST DMD mode of control. In fact, FIST DMD mode of control only affected the mechanized infantry FO missions.

Finally, the automatic reduction system proved to be a useful tool for data collection and reduction of field data and the ability to perform a controlled experiment during a field test was demonstrated with overwhelming success. However, it demonstrates the need for more sophisticated MOP's than simply speed of service.

VII. ACKNOWLEDGEMENTS The authors wish to express their appreciation to The Fire Support and Target Acquisition Directorate of the HEL, The Artillery Systems Concepts Branch of the BRL and to Annette Wiseman for formatting this report.

A METHOD FOR ESTIMATING DETERMINISTIC WATER WAVES CONTAMINATED WITH RANDOM BACKGROUND NOISE

Michael E. Andrew

U.S. Army Engineer Waterways Experiment Station
Coastal Engineering Research Center
P.O. Box 631
Vicksburg, MS 39180

PREFACE

This study was funded by the Defense Nuclear Agency (DNA) as part of an ongoing effort by the Prototype Measurement and Analysis Branch of the U. S. Army Coastal Engineering Research Center (CERC) at WES. The study was conducted at CERC, WES, under the direction of Dr. R. W. Whalin, Chief, and Dr. F. E. Camfield, Acting Chief, Engineering Development Division. The study was performed under the direct supervision of Dr. Dennis R. Smith, Chief, Prototype Measurement and Analysis Branch and Mr. Gary L. Howell, Acting Chief.

Directors of WES during the conduction of the study and preparation of this report were COL Tilford C. Creel, CE, and COL Robert C. Lee, CE. Techincal Director was Mr. F. R. Brown.

INTRODUCTION

1. The purpose of this study was to explore methods for predicting and reducing the error in measured explosion wave data due to the presence of a wind generated random sea suface. In this analysis, the model to be used for representing the mixture of random and deterministic waves is the linear superposition, or adding together, of explosion wave and random wave amplitude functions. An analysis method based on the linear superposition model requires the assumption that frequency components of interest in the explosion generated waves are contained in a certain frequency band. Errors resulting from the random wave frequencies outside the specified band of interest are then eliminated by means of band pass filters. Resulting estimates for the explosion wave-form will be biased by smoothing inherent in the filtering process. This bias effect is reduced by observing the affect that the filtering process has on theoretical models for explosion waves. Theoretical models have been shown to correspond well with measured data (Le Mehaute 1971) making them useful as a means of estimating bias due to the filtering process. Correction factors for a specific filter and set of explosion wave parameters are computed and used to remove filter bias from estimates of maximum explosion wave amplitude.

2. Expected error due to random background noise can be predicted prior to a test using the energy spectrum of the sea surface. This makes it possible to make GO or NO GO decisions prior to testing based on the roughness of the sea surface. The GO/NO GO model developed here is based on the parameterized sea surface spectrum developed by Ochi and Hubble (1976). An interactive computer program for predicting expected background error is developed and examples of its usage are presented in a later section.

Wave mixture model

3. Let the discrete time series associated with the random sea surface be denoted by x_n and the discrete version of the explosion wave time series by p_n for $n = 0, 1, \ldots, N-1$ with a time increment of Δt between successive numbers of the two time series. Linear superposition of the two time series produces

$$p_{on} = p_n + x_n \ , \ n = 0, 1, \ldots , N-1 \qquad 1.1$$

It is assumed that each variable in the time series x_n is distributed as a Gaussian random variable with mean zero and variance σ^2. The time series p_{on} represents the observed mixture of noise and explosion waves. This variable p_{on} is also distributed Gaussian with mean p_n and variance σ^2. The discrete fourier transform (DFT) for p_{on} is

$$P_{om} = \Delta t \sum_{n=0}^{N-1} p_{on} e^{-i2\pi mn/N} \qquad 1.2$$

or

$$P_{om} = P_m + X_m \qquad 1.3$$

Where

$$P_{om} = \Delta t \sum_{n=0}^{N-1} p_{on} e^{-i2\pi mn/N} \qquad 1.4$$

$$X_m = \Delta t \sum_{n=0}^{N-1} x_n e^{-i2\pi mn/N} \qquad 1.5$$

are the DFTs for p_n and x_n respectively. Subscript m refers to frequency values $f = m\Delta f$ for $\Delta f = 1/(N\Delta t)$.

4. Define the estimate of the DFT for P_m to be

$$\hat{P}_m = P_{om} D_m \text{ , } m = 0, 1, \text{---} \text{ , } N - 1 \qquad 1.6$$

for

$$D_m = \begin{cases} 1.0, & M_1 \leq m \leq M_2 \\ 1.0, & N - M_2 \leq m \leq N - M_1 \\ 0.0, & \text{all other values of } m \end{cases} \qquad 1.7$$

Values $f_1 = M_1 \Delta f$ and $f_2 = M_2 \Delta f$ define the frequency interval of interest for the data to be analyzed. Then, the estimate for the time series p_n is the inverse DFT of $\hat{P}_m$, given by

$$\hat{p}_n = \Delta f \sum_{m=0}^{N-1} P_{om} D_m e^{i2\pi mn/N} \qquad 1.8$$

Note that since $\hat{p}_n$ is a linear combination of linear combinations of Gaussian random variables p_{on}, $n = 0, 1, \text{---}, N-1$ then $\hat{p}_n$ is also distributed Gaussian. Substituting equation 1.3 into 1.8, the estimated time series can be rewritten as:

$$\hat{p}_n = p_n \beta_n + \varepsilon_n \tag{1.9}$$

where

$$\beta_n = \Delta f \sum_{m=0}^{N-1} P_m (D_m - 1.0) e^{i2\pi mn/N} \tag{1.10}$$

$$\varepsilon_n = \Delta f \sum_{m=0}^{N-1} X_m D_m e^{i2\pi mn/N} \tag{1.11}$$

5. The term β_n is referred to as the estimate bias term and ε_n is called the random term. Statistical expectation of $\hat{p}_n$ is

$$E[\hat{p}_n] = p_n + \beta_n \tag{1.12}$$

since $E[\varepsilon_n] = 0.0$ and since the other terms are deterministic. Variability of the estimate $\hat{p}_n$ is given by its variance, which is

$$\sigma_n^2 = \text{Var}\,[\hat{p}_n] = \Delta f \sum_{m=0}^{N-1} S_m D_m \tag{1.13}$$

where S_m is the discrete version of the power spectral density of the random sea surface. Standard deviation of $\hat{p}_n$ is

$$\sigma_n = \sqrt{\sigma_n^2} \tag{1.14}$$

6. Assuming that σ_n and β_n are known, then the $(1-\alpha/2) \times 100$ percent confidence interval for p_n is

$$(\hat{p}_n - \beta_n - \sigma_n Z_{1-\alpha/2}\,,\ \hat{p}_n - \beta_n + \sigma_n Z_{1-\alpha/2}) \tag{1.15}$$

where $Z_{1-\alpha/2}$ is the $1-\alpha/2$ percentage point for a standard normal distribution.

7. In an applied setting, σ_n is not a known quantity; however, it can be estimated by replacing S_m with $\hat{S}_m$, the estimated energy spectral density in equation 1.13.

8. The bias term β_n will be estimated by observing the affect that filtering has on functions that are used to model explosion wave time series for given explosive yields, water depths, and measurement location.

9. To simplify the discussion of the random error in later parts of this report, the bias term β_n will be included in estimates of the explosion wave time series and the unbiased estimate of p_n will then be

$$\hat{p}_{un} = \hat{p}_n - \beta_n \tag{1.16}$$

with confidence interval

$$\hat{p}_{un} \pm \sigma_n Z_{1-\alpha/2} \tag{1.17}$$

Explosion wave model

10. There are several functions that are used to model explosion wave amplitude. The function used here for testing and bias estimation is the most suitable of the available models (Le Mehaute 1971). Explosion wave amplitude as a function of time and distance from the explosion is written (dimensionless form)

$$\eta(r,t) = A \cos (Kr - t(\tanh K))^{1/2} \tag{1.18}$$

where

$$A = \frac{\varepsilon R}{rK} J_3 (KR) \left[\frac{KV (KV)}{-V' (K)}\right]^{1/2} \tag{1.19}$$

and

R = radius of initial cavitation

η = amplitude of initial cavitation

K = wave number

r = distance at which amplitude is derived

t = time

$V(K)$ = wave group velocity

$J_3(KR)$ = Bessel function of the first kind of order 3

11. The parameters listed above are dimensionless resulting in a dimensionless value for equation 1.18. In order to obtain the dimensional version, equation 1.18 is multiplied by the water depth at which the explosion occurs. Equation 1.18 is evaluated numerically for given values of the parameters to provide test data for the analysis techniques of this study.

Random noise simulation

12. Frequency domain simulation of Gaussian random time series makes use of the statistical identities between linear combinations of Gaussian variables. That is, any linear combination of independent Gaussian random variables is also Gaussian with mean and variance that are linear combinations of the means and variances of the original random variables. Discrete Fourier Transform (DFT) pairs are linear combinations and, therefore, have identities described above. A Gaussian time series with a given theoretical power spectrum can be shown to have a DFT that is Gaussian with a variance that is related to the power spectrum of the time series. That is, if X_n, $n = 0, 1, \ldots, N-1$ is Gaussian with mean zero and variance σ^2 and results from a random process with power spectral density $S(f)$, then the DFT of X_n given by $U_m - i\,V_m$ has real and imaginary parts U_m, V_m that are also mean zero Gaussian, with variances

$$\mathrm{Var}(U_m) = \begin{cases} N\Delta t\, S_m & m = 0, N/2 \\ \dfrac{N\Delta t\, S_m}{2} & 0 < m < N/2 \end{cases} \qquad 2.1$$

$$\mathrm{Var}(V_m) = \begin{cases} 0 & m = 0, N/2 \\ \dfrac{N\Delta t\, S_m}{2} & 0 < m < N/2 \end{cases} \qquad 2.2$$

for S_m the discrete version of the power spectrum $S(f)$. It is also known that (U_m, V_m) and $(U_{m'}, V_{m'})$ are statistically independent for $m \neq m'$ $0 < m, m' < N/2$, and U_m is independent of $V_{m'}$ for all m, m'. The values for $m > N/2$ can be found by the identity $U_m = U_{N-m}$ and $V_m = V_{N-m}$.

13. The steps for simulating a Gaussian random time series with power spectral density S(f) are:

(1) Generate two statistically independent sequences of independent mean zero Gaussian random numbers, say, XR_m and XI_m for $0 \leq m \leq N/2$

(2) Let $U_m = XR_m \left(N\Delta t \frac{S_m}{2}\right)^{1/2}$

$V_m = XI_m \left(N\Delta t \frac{S_m}{2}\right)^{1/2}$ $\qquad 0 < m < N/2$

$U_m = XR_m \left(N\Delta t\, S_m\right)^{1/2}$

$V_m = 0$ $\qquad m = 0, N/2$

(3) Apply DFT to

$$X_m = U_m - i\, V_m$$

to obtain the resulting times series X_n , n = 0, 1, ... , N - 1 of equation 1.5. Finally, add the time series from equation 1.18 to X_n to obtain the desired mixture of a Gaussian random sea surface with explosion generated water waves.

GO/NO GO analysis

14. Equation 1.13 gives the expected error due to the presence of random background noise in the measured explosion wave time history. The associated confidence interval of equation 1.15 represents the interval that will capture the true explosion wave time history, P_n , $(1 - \alpha/2) \times 100$ percent of the time. If the power spectral density of the random sea surface is known or specified by one of the commonly used parametric forms, then equations 1.13 through 1.17 can be computed to provide an estimate of the statistical error to be expected in filtered explosion wave.

15. A parameterized spectrum known as the Ochi Hubble (1976) six parameter spectrum was chosen for use in this analysis. This six parameter spectrum includes a spectral peakedness parameter that determines the narrow

bandedness of the spectrum. A peakedness parameter of 1.0 is used here resulting in a spectrum that is equivalent to Bretschnieder and Pierson Moskowitz type spectral densities. Parameters needed for input to the parameterized spectrum are the mean period for long period waves generated by distant storms (swell) given by T_s, mean period for locally generated waves T_w, significant wave height for swell H_s, and significant wave height for sea H_w. Figure 1 is a plot of the quantity $\sigma_n Z_{1-\alpha/2}$ from equation 1.17 for varying values of spectral parameters. Examples and instructions for the GO/NO GO device program are listed in Appendix A.

Synthetic data analysis

16. Simulations were computed for varying measurement distances from the explosion source. Distance values were 500, 800, 1,000, and 2,000 ft. There were five simulations for each location, each representing a different random background wave sequence. Random background waves were generated according to the method of section 4 using an Ochi Hubble type bimodal spectrum with significant height for swell equal to 1.0 ft, significant height for local seas equal to .75 ft, mean period for swell at 12.5 sec, and mean period for seas at 5 sec.

17. Two methods were used for estimating explosion wave time histories. The first being the boxcar type filter described in section 2; the second method was an 8 pole Butterworth digital filter with maximum flatness in the passband and stop band.

18. Plots of theoretical, observed, and estimated time series along with random noise x_n from equation 1.1, random error ε_n from equation 1.11, and bias β_n from equation 1.10 are presented in Appendix B. Of particular interest is the maximum wave amplitude for each simulation. Values for theoretical, observed and estimated maximum wave amplitudes are given in Table 1 and Table II. The tables are also include the associated values for σ^2, sea surface variance or mean square error, and σ_ε^2, error variance or mean square of the random error term ε_n from equation 1.11. The theoretical value for σ_n^2 of equation 1.13 from the GO/NO GO example of Appendix A is $\sigma_n^2 = 0.039$. Since σ_ε^2 is the sample variance and σ_n^2 is the theoretical variance of the explosion wave estimate time series, their values should be close. Note that values for σ_ε^2 generally scatter around 0.039. Average values for σ^2 and σ_ε^2 over each set of twenty simulation demonstrate that

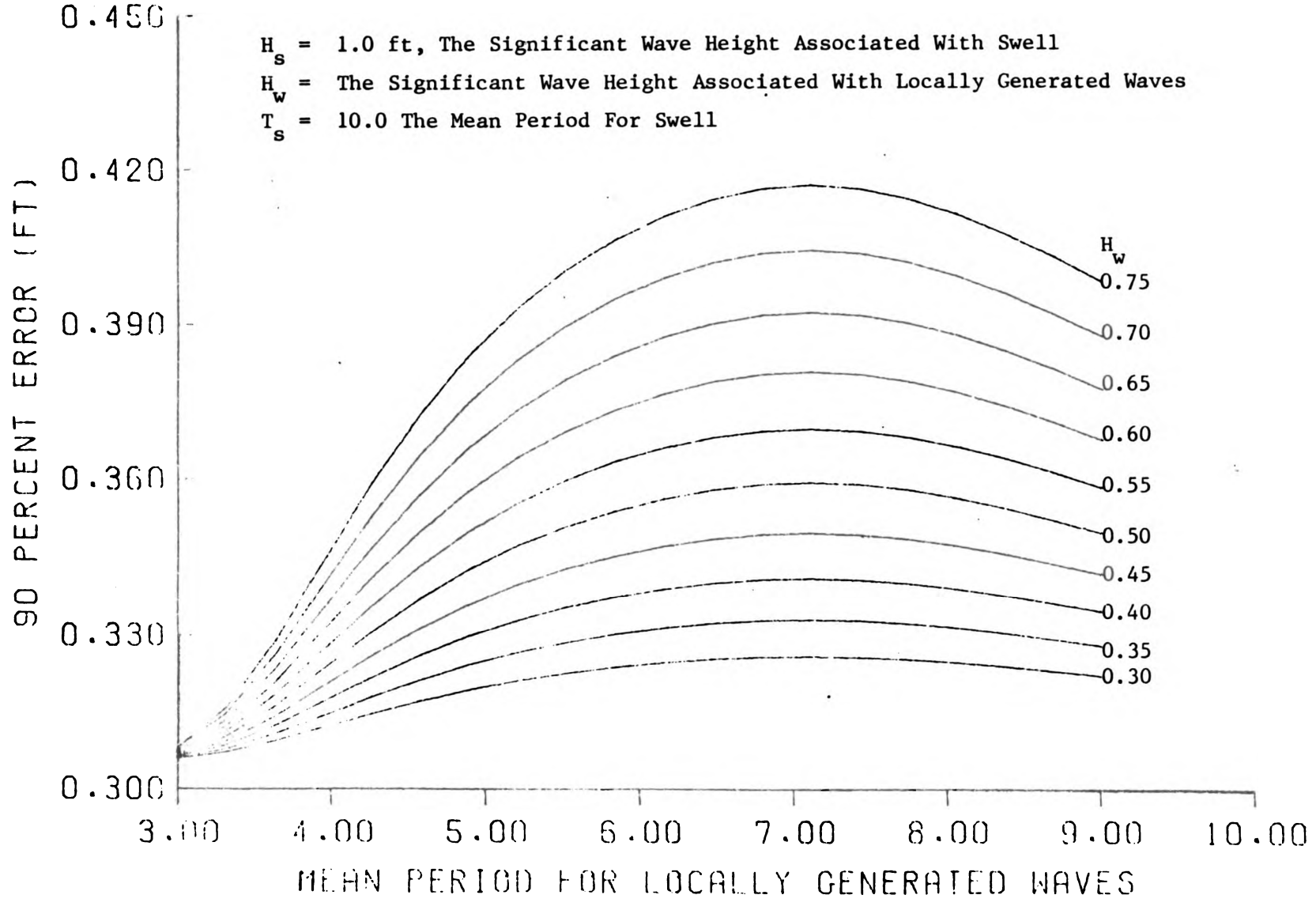


Figure 1. Confidence Limit for the Estimated Explosion Wave Amplitude

TABLE I

Boxcar Filter

R	P_{max}	$\hat{P}_{max}$	P_{obsmax}	σ^2	σ_ϵ^2
500	9.07	8.87	8.74	.096	.040
500	9.07	9.06	8.85	.100	.042
500	9.07	9.60	9.72	.103	.044
500	9.07	8.99	8.96	.102	.053
500	9.07	8.74	8.43	.092	.042
800	5.59	5.37	5.58*	.095	.039
800	5.59	5.55	5.74	.102	.043
800	5.59	5.50	5.36	.084	.043
800	5.59	5.66	5.32	.097	.040
800	5.59	5.29	5.55*	.088	.040
1000	4.15	4.25	4.27	.094	.039
1000	4.15	4.55	4.67	.103	.031
1000	4.15	3.96	4.12*	.103	.044
1000	4.15	3.87	3.88*	.095	.044
1000	4.15	3.85	4.25*	.096	.040
2000	2.25	2.33	2.18*	.102	.046
2000	2.25	2.07	1.68	.106	.048
2000	2.25	2.47	1.96	.091	.039
2000	2.25	2.26	2.38	.103	.038
2000	2.25	2.29	2.37	.109	.050

P_{max} = Theoretical maximum explosion wave amplitude.

$\hat{P}_{max}$ = Estimated maximum explosion wave amplitude.

P_{obsmax} = Observed maximum explosion wave amplitude.

* Estimated wave amplitude is worse than observed.

TABLE II

Butterworth Filter

R	P_{max}	$\hat{P}_{max}$	P_{obsmax}	σ^2	σ_ε^2
500	9.07	8.97	8.63	.100	.027
500	9.07	9.04	8.74	.099	.028
500	9.07	9.07	8.93	.096	.029
500	9.07	9.00	8.76	.092	.030
500	9.07	9.00	8.58	.090	.030
800	5.59	5.55	5.71	.091	.030
800	5.59	5.35	5.16	.102	.030
800	5.59	5.66	5.70	.105	.029
800	5.59	5.78	6.10	.090	.028
800	5.59	5.62	5.67	.094	.030
1000	4.15	4.55	4.66	.103	.031
1000	4.15	4.39	4.45	.087	.027
1000	4.15	3.76	3.76*	.106	.033
1000	4.15	3.94	3.81	.100	.032
1000	4.15	4.39	4.44	.090	.025
2000	2.25	2.16	2.06	.092	.028
2000	2.25	2.42	2.10*	.095	.028
2000	2.25	1.96	1.89	.100	.031
2000	2.25	2.06	1.61	.095	.025
2000	2.25	2.30	2.11	.095	.028

P_{max} = Theoretical maximum explosion wave amplitude.

$\hat{P}_{max}$ = Estimated maximum explosion wave amplitude.

P_{obsmax} = Observed maximum explosion wave amplitude.

* Estimated wave amplitude is worse than observed.

the Butterworth filter does better than the Boxcar in reducing the overall estimate error.

(1) Boxcar

Average (σ^2) = 0.098

Average σ_ε^2 = 0.042

(2) Butterworth

Average (σ^2) = 0.096

Average σ_ε^2 = 0.029

19. Boxcar filter estimates are closer to the theoretical value of the maximum wave amplitude than the observed maximum about 70 percent of the time. The Butterworth does better than no filter 90 percent of the time.

Results and conclusions

20. Random background noise effects on measured explosion wave time series can be reduced by means of band pass filter techniques. The Butterworth type filter tends to do a better job at reducing error than a Boxcar filter in frequency space.

21. The estimate error for the filtered explosion wave estimates can be predicted using the power spectral density of the background sea surface. In this way, GO or NO/GO decisions can be made during testing based on sea surface conditions at the time.

22. A GO or NO/GO computer program is given along with self explanatory documentation in Appendix A. The program is designed to provide quick easy error predictions using any computer that uses FORTRAN.

REFERENCES

Le Mehaute B. 1971. "Theory of Explosion-Generated Water Waves," Advances in Hydrosciences, Vol 7, Academic.

Ochi, M. K., and Hubble, N. E. 1976. "Six-Parameter Wave Spectra", Coastal Engineering, pp 301.

APPENDIX A

Appendix A consists of the listings for the GO/NO GO computer program and an example of its usage.

GO NO GO COMPUTER PROGRAM

The GO NO GO program was developed on the Honeywell DPS-1 system at the Waterways Experiment Station. The program is written in standard FORTRAN that is easily adapted to any system that supports standard FORTRAN IV. There may be some minor changes in the input and output conventions and line numbering. For DEC VAX systems the PRINT commands should be replaced by PRINT * commands and the line numbers removed.

```
10C     PROGRAM GO NO GO
20      F1=.08
30      F2=.4
40      PRINT,"THIS PROGRAM COMPUTES THE EXPECTED ERROR FOR A"
50      PRINT,"DETERMINISTIC WAVEFORM THAT IS CONTAMINATED WITH"
60      PRINT,"GAUSSIAN RANDOM NOISE"
70      PRINT,"THE INPUT PARAMETERS ARE THE FOUR PARAMETERS"
80      PRINT,"ASSOCIATED WITH THE WAVE SPECTRUM FOR A MIXTURE"
90      PRINT,"OF LONG PERIOD SWELL AND LOCALLY GENERATED SEAS"
100      PRINT,"THE INPUT UNITS DETERMINE THE UNITS OF THE OUTPUT"
110      PRINT,"THE PARAMETERS ARE:"
120      PRINT,"TS = MEAN PERIOD FOR SWELL"
130      PRINT,"TW = MEAN PERIOD FOR SEAS"
140      PRINT,"HS = SIGNIFICANT WAVE HEIGHT FOR SWELL"
150      PRINT,"HW = SIGNIFICANT WAVE HEIGHT FOR SEA"
160      PRINT,"IF YOU WISH TO RUN THIS PROGRAM TYPE 1 "
170      PRINT,"AT THE PROMPT FOLLOWED BY A CARRIAGE RETURN"
180      PRINT,"OTHERWISE TYPE 0 FOLLOWED BY A CARRIAGE RETURN"
190      READ,IRUN
200      IF(IRUN.NE.1)STOP
210   5  PRINT,"TYPE THE WAVE SPECTRUM PARAMETERS ONE TO A LINE"
220      PRINT,"ACCORDING TO THE PROMPT MESSAGES"
230    6 PRINT,"TYPE THE MEAN PERIOD FOR SWELL"
240      READ,TS
250      IF(TS.GE.5.0)GO TO 7
260      PRINT,"ILLEGAL VALUE FOR TS, RETYPE"
270      GO TO 6
280    7 PRINT,"TYPE THE MEAN PERIOD FOR SEAS"
290      READ,TW
300      IF(TW.GT.0.0.AND.TW.LT.10.0) GO TO 8
310      PRINT,"ILLEGAL VALUE FOR TW, RETYPE"
320      GO TO 7
330    8 PRINT,"TYPE THE SIGNIFICANT HEIGHT FOR SWELL"
340      READ,HS
350      IF(HS.GT.0.0) GO TO 9
360      PRINT,"ILLEGAL VALUE FOR HS, RETYPE"
370      GO TO 8
380   9 PRINT,"TYPE THE SIGNIFICANT HEIGHT FOR SEA"
390      READ,HW
400      IF(HW.GT.0.0)GO TO 10
410      PRINT,"ILLEGAL VALUE FOR HW, RETYPE"
420      GO TO 9
430   10 PRINT,"THE DEFAULT VALUES FOR THE FREQUENCY BAND OF"
```

```
      PRINT,"THE DETERMINISTIC WAVEFORM ARE :"
      PRINT,"F1 = ",F1," F2 = ",F2
      PRINT,"IF YOU WISH TO RESET THE VALUES FOR F1 AND F2"
      PRINT,"TYPE 1 FOLLOWED BY A CARRIAGE RETURN OTHERWISE"
      PRINT,"TYPE 0 FOLLOWED BY A CARRIAGE RETURN"
      READ,ICHNG
      IF(ICHNG.NE.1)GO TO 15
      PRINT,"TYPE THE NEW VALUE FOR F1"
      READ,F1
      PRINT,"TYPE THE NEW VALUE FOR F2"
      READ,F2
   15 CONTINUE
      PRINT,"PROGRAM IS RUNNING PLEASE WAIT"
      PRINT,".-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-."
C     COMPUTE THE ERROR THEN ADJUST BY THE STANDARD
C     NORMAL PERCENTAGE POINT TO OBTAIN A 90 PERCENT
C     CONFIDENCE LIMIT.
      CALL ERRP(F1,F2,TS,TW,HS,HW,E)
      ERROR=SQRT(E)
      CON=1.645*ERROR
      PRINT,"THE EXPECTED STANDARD DEVIATION FOR THE"
      PRINT,"ESTIMATE OF THE DETERMINISTIC WAVEFORM IS  "
      PRINT,"APPROXIMATELY S = ",ERROR
      PRINT,"************************************"
      PRINT,"THE 90 PERCENT CONFIDENCE INTERVAL FOR THE ESTIMATE"
      PRINT,"IS PLUS OR MINUS THE VALUE C = ",CON
      PRINT,"************************************************"
      PRINT,"IF YOU WISH TO OBTAIN ANOTHER SET OF VALUES TYPE"
      PRINT,"1 FOLLOWED BY A CARRIAGE RETURN OTHERWISE TYPE 0"
      PRINT,"THEN A CARRIAGE RETURN"
      READ,IRUN
      IF(IRUN.NE.1) STOP
      GO TO 5
      END
```

```
      SUBROUTINE ERRP(F1,F2,TS,TW,HS,HW,E)
C*****************************************************************
C     THIS SUBROUTINE COMPUTES THE AREA UNDER AN OCHI-
C     HUBBLE TYPE SPECTRUM BETWEEN THE FREQUENCY VALUES
C     F1 AND F2 FOR INPUT PARAMETERS,
C
C
C           TS = MEAN PERIOD FOR SWELL
C           TW = MEAN PERIOD FOR LOCAL SEAS
C           HS = SIGNIFICANT WAVE HEIGHT FOR SWELL
C           HW = SIGNIFICANT WAVE HEIGHT FOR LOCAL SEAS
C
C
C*****************************************************************
      DELF=(F2-F1)/100.
      SUM=0.0
      DO 10 I=1,100
      F=F1+(I-1)*DELF
      SUM=SUM+SPEC(TS,HS,F)+SPEC(TW,HW,F)
   10 CONTINUE
      E=SUM*DELF
      RETURN
      END
      FUNCTION SPEC(T,H,F)
C*****************************************************
C     COMPUTES THE OCHI-HUBBLE TYPE SPECTRUM FOR
C     PARAMETERS,
C
C           T = MEAN PERIOD
C           H = SIGNIFICANT WAVE HEIGHT
C
C*****************************************************
      A=(5./(4.*T**4))
      B=A*((H/4.)**2)/(F**5.)
      ARG=-A/(F**4.)
      IF(ARG.GE.-50.0) GO TO 5
      SPEC=0.0
      RETURN
    5 SPEC=4.0*B*EXP(ARG)
      RETURN
      END
```

GO/NO GO EXAMPLE

The GO/NO GO example begins with an FRN command. This command is specific to the Honeywell DPS-1 system and means get the program GO NO, compile it, and run it. The rest of the example is typical interactive session with the GO/NO GO package.

```
-FRN GONO
THIS PROGRAM COMPUTES THE EXPECTED ERROR FOR A
DETERMINISTIC WAVEFORM THAT IS CONTAMINATED WITH
GAUSSIAN RANDOM NOISE
THE INPUT PARAMETERS ARE THE FOUR PARAMETERS
ASSOCIATED WITH THE WAVE SPECTRUM FOR A MIXTURE
OF LONG PERIOD SWELL AND LOCALLY GENERATED SEAS
THE INPUT UNITS DETERMINE THE UNITS OF THE OUTPUT
THE PARAMETERS ARE:
TS = MEAN PERIOD FOR SWELL
TW = MEAN PERIOD FOR SEAS
HS = SIGNIFICANT WAVE HEIGHT FOR SWELL
HW = SIGNIFICANT WAVE HEIGHT FOR SEA
IF YOU WISH TO RUN THIS PROGRAM TYPE 1
AT THE PROMPT FOLLOWED BY A CARRIAGE RETURN
OTHERWISE TYPE 0 FOLLOWED BY A CARRIAGE RETURN
=1
TYPE THE WAVE SPECTRUM PARAMETERS ONE TO A LINE
ACCORDING TO THE PROMPT MESSAGES
TYPE THE MEAN PERIOD FOR SWELL
=12.5
TYPE THE MEAN PERIOD FOR SEAS
=5.0
TYPE THE SIGNIFICANT HEIGHT FOR SWELL
=1.0
TYPE THE SIGNIFICANT HEIGHT FOR SEA
=.75
THE DEFAULT VALUES FOR THE FREQUENCY BAND OF
THE DETERMINISTIC WAVEFORM ARE :
F1 =   0.80000000E-01 F2 =   0.40000000E 00
IF YOU WISH TO RESET THE VALUES FOR F1 AND F2
TYPE 1 FOLLOWED BY A CARRIAGE RETURN OTHERWISE
TYPE 0 FOLLOWED BY A CARRIAGE RETURN
=1
TYPE THE NEW VALUE FOR F1
=.11
TYPE THE NEW VALUE FOR F2
=.25
```

```
PROGRAM IS RUNNING PLEASE WAIT
.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.
THE EXPECTED STANDARD DEVIATION FOR THE
ESTIMATE OF THE DETERMINISTIC WAVEFORM IS
APPROXIMATELY S =   0.19739283E 00
**********************************
THE 90 PERCENT CONFIDENCE INTERVAL FOR THE ESTIMATE
IS PLUS OR MINUS THE VALUE C =   0.32471121E 00
**************************************************
IF YOU WISH TO OBTAIN ANOTHER SET OF VALUES TYPE
1 FOLLOWED BY A CARRIAGE RETURN OTHERWISE TYPE 0
THEN A CARRIAGE RETURN
=1
TYPE THE WAVE SPECTRUM PARAMETERS ONE TO A LINE
ACCORDING TO THE PROMPT MESSAGES
TYPE THE MEAN PERIOD FOR SWELL
=10.0
TYPE THE MEAN PERIOD FOR SEAS
=6.0
TYPE THE SIGNIFICANT HEIGHT FOR SWELL
=1.
TYPE THE SIGNIFICANT HEIGHT FOR SEA
=.5
THE DEFAULT VALUES FOR THE FREQUENCY BAND OF
THE DETERMINISTIC WAVEFORM ARE :
F1 =   0.11000000E 00 F2 =   0.25000000E 00
IF YOU WISH TO RESET THE VALUES FOR F1 AND F2
TYPE 1 FOLLOWED BY A CARRIAGE RETURN OTHERWISE
TYPE 0 FOLLOWED BY A CARRIAGE RETURN
=0
PROGRAM IS RUNNING PLEASE WAIT
.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.
THE EXPECTED STANDARD DEVIATION FOR THE
ESTIMATE OF THE DETERMINISTIC WAVEFORM IS
APPROXIMATELY S =   0.21594351E 00
**********************************
THE 90 PERCENT CONFIDENCE INTERVAL FOR THE ESTIMATE
IS PLUS OR MINUS THE VALUE C =   0.35522708E 00
**************************************************
IF YOU WISH TO OBTAIN ANOTHER SET OF VALUES TYPE
1 FOLLOWED BY A CARRIAGE RETURN OTHERWISE TYPE 0
THEN A CARRIAGE RETURN
=0
```

APPENDIX B

Time Series Plots for Synthetic Wave Data Analysis

Time series plots include examples of the theoretical explosion waveform versus the explosion waveform mixed with random noise, random noise, filtered estimate versus theoretical waveform, random error or random noise after filtering, and bias due to filtering. The plots represent a time series that would result if a device began measuring the background noise exactly 50 sec before the first explosion wave reaches the measurement location.

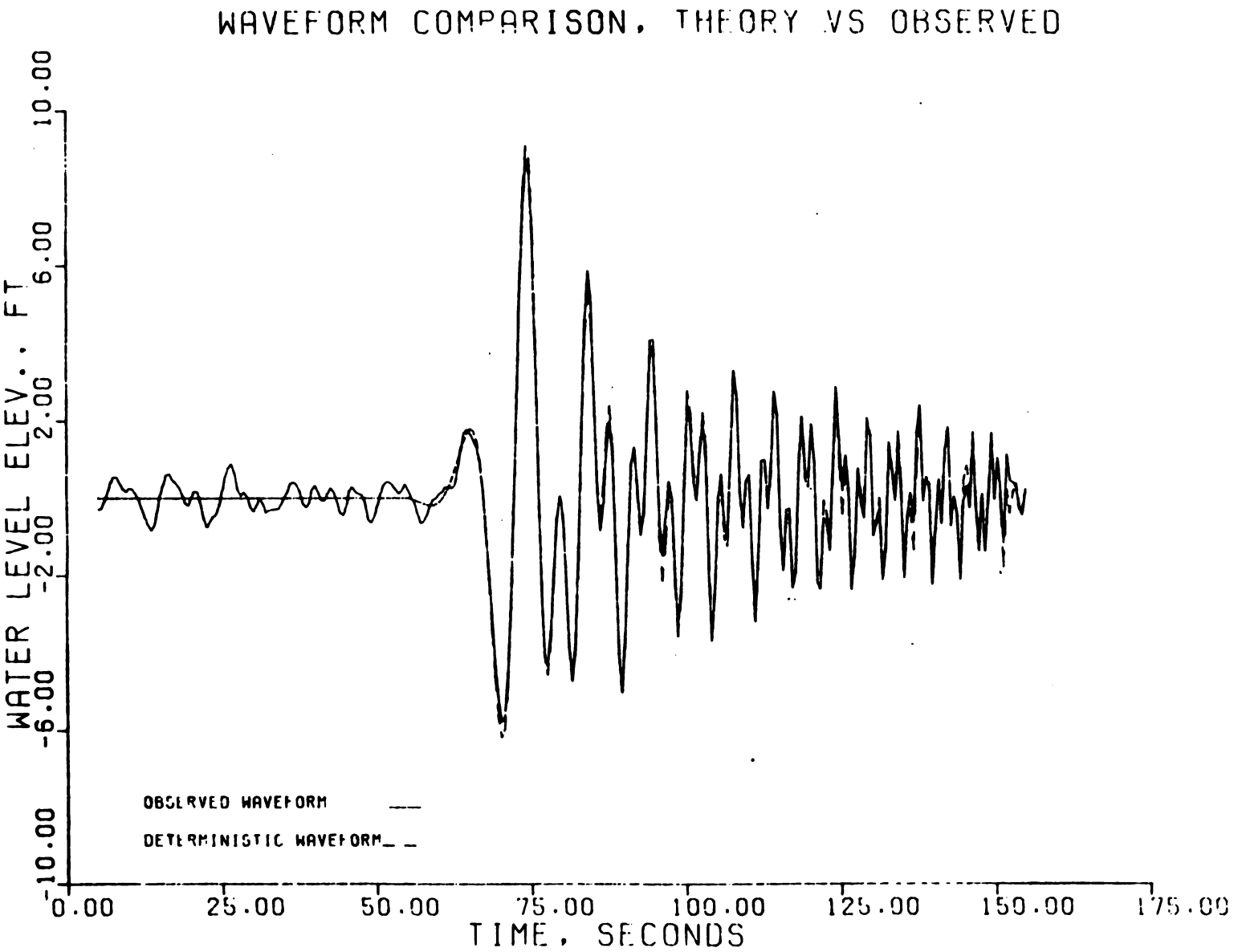


Figure B1. Time Series Plot For 500 ft from the Explosion Source

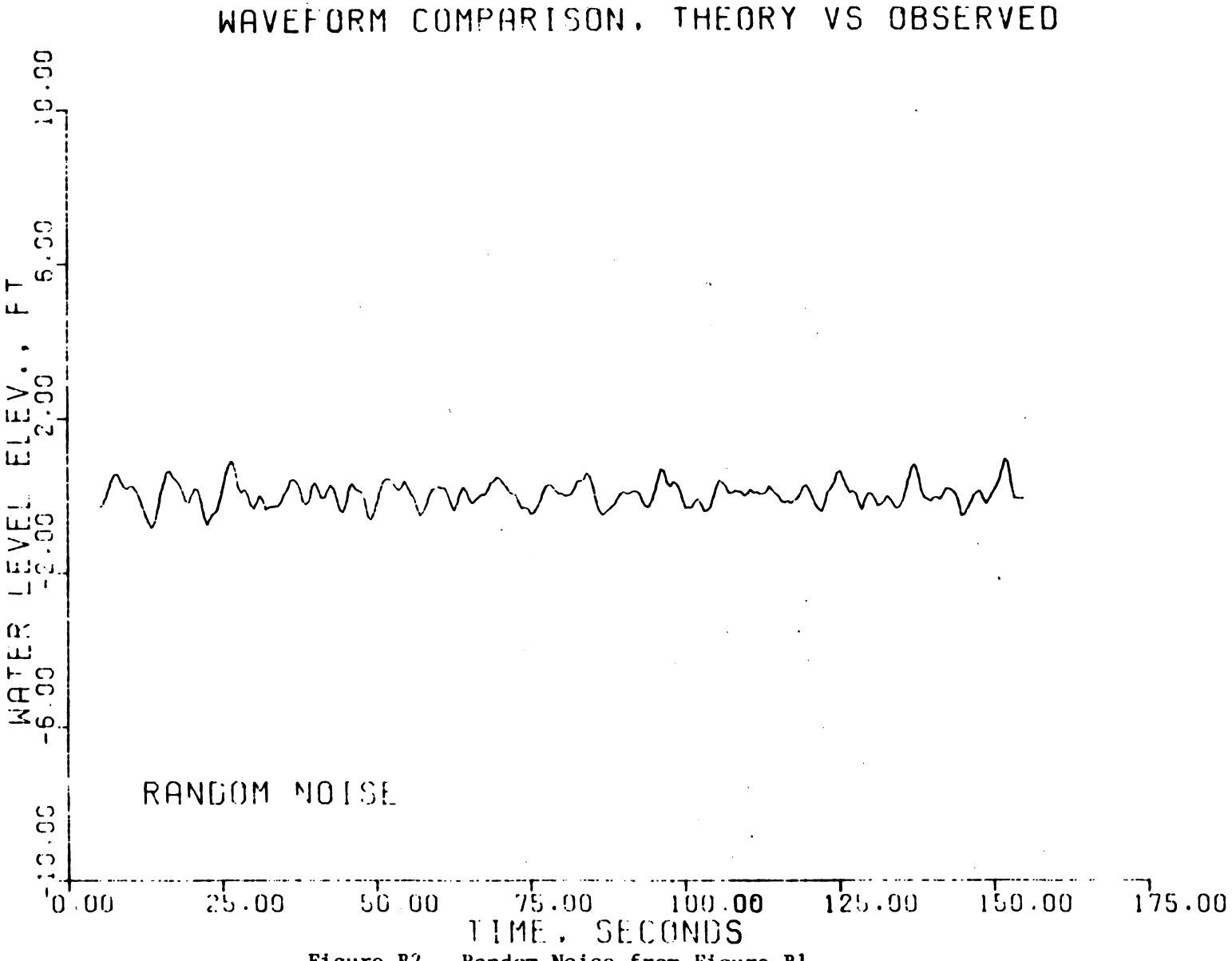


Figure B2. Random Noise from Figure B1

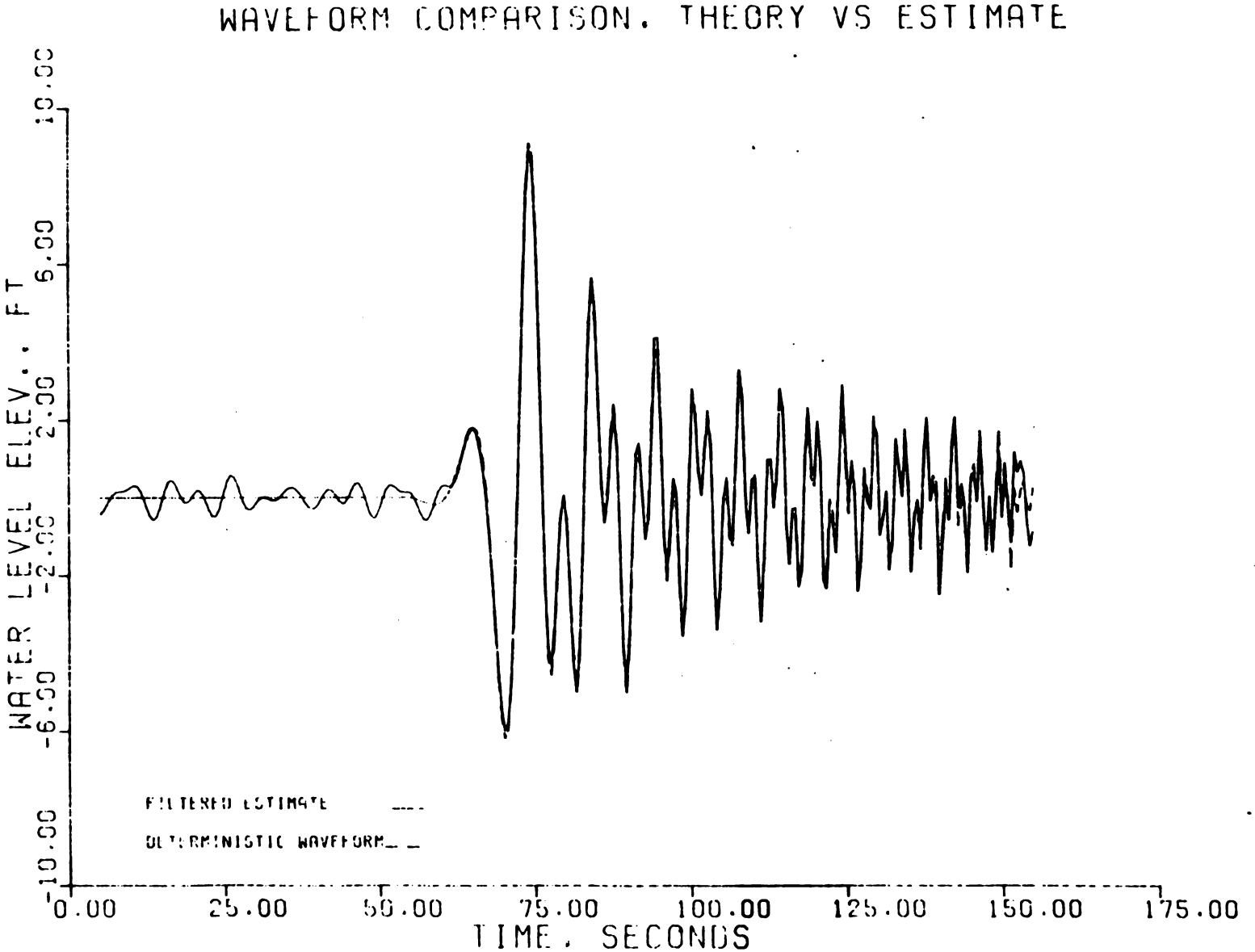


Figure B3. Time series plot for the Boxcar Estimate at 500 ft from the explosion source

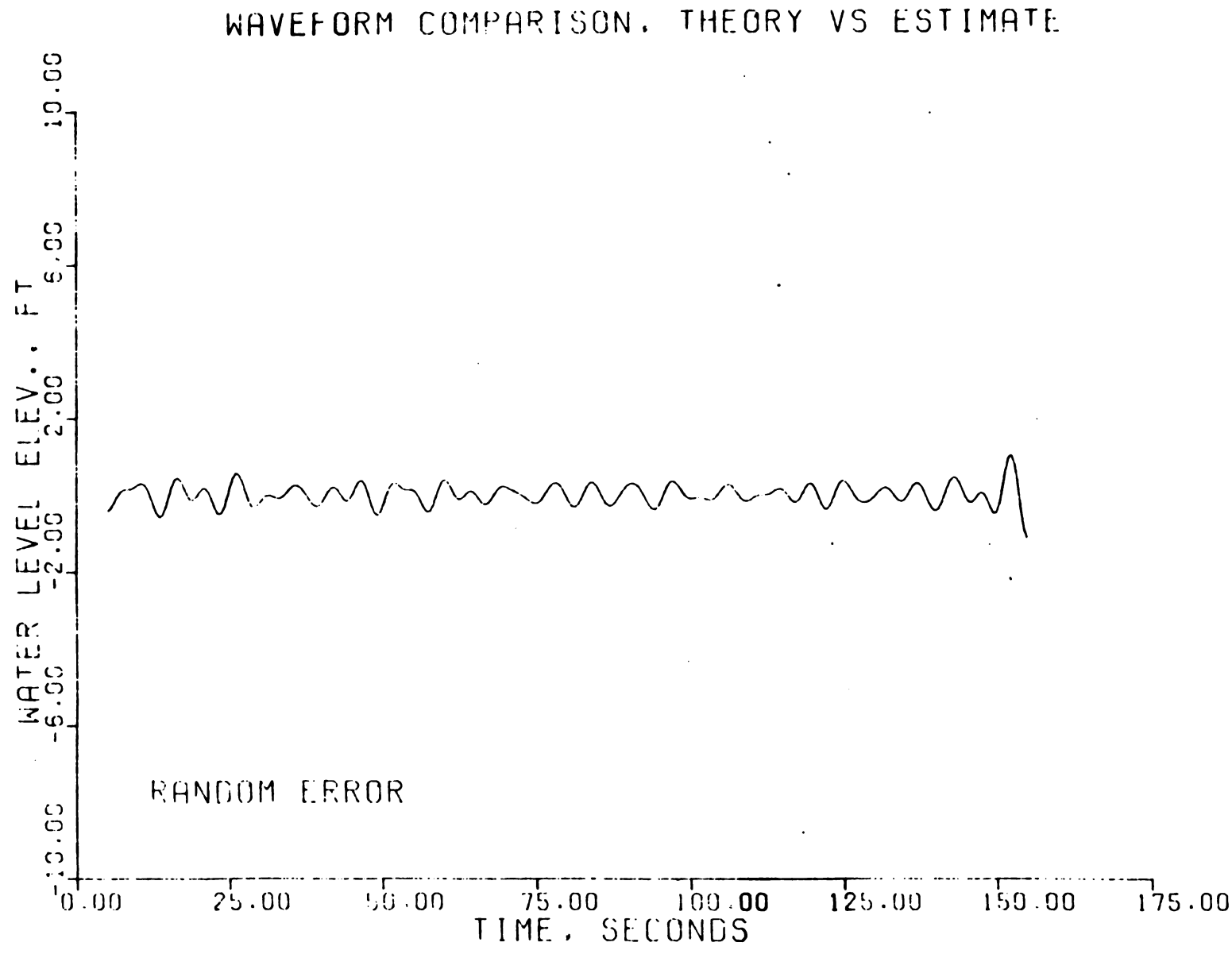


Figure B5. Random error from the estimated time series in Figure B4

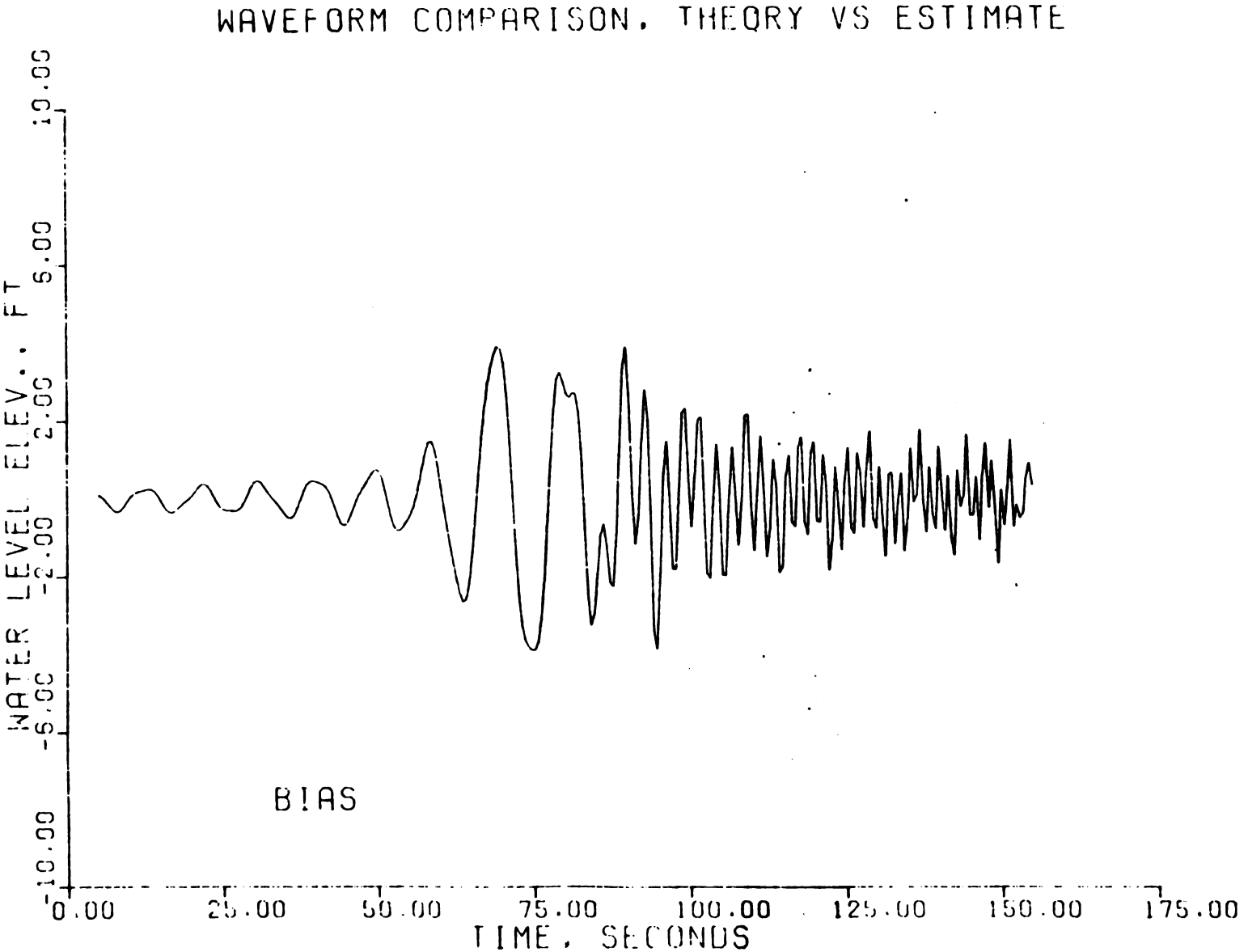


Figure B6. Bias time series for the Boxcar Estimate at 500 ft from the explosion source

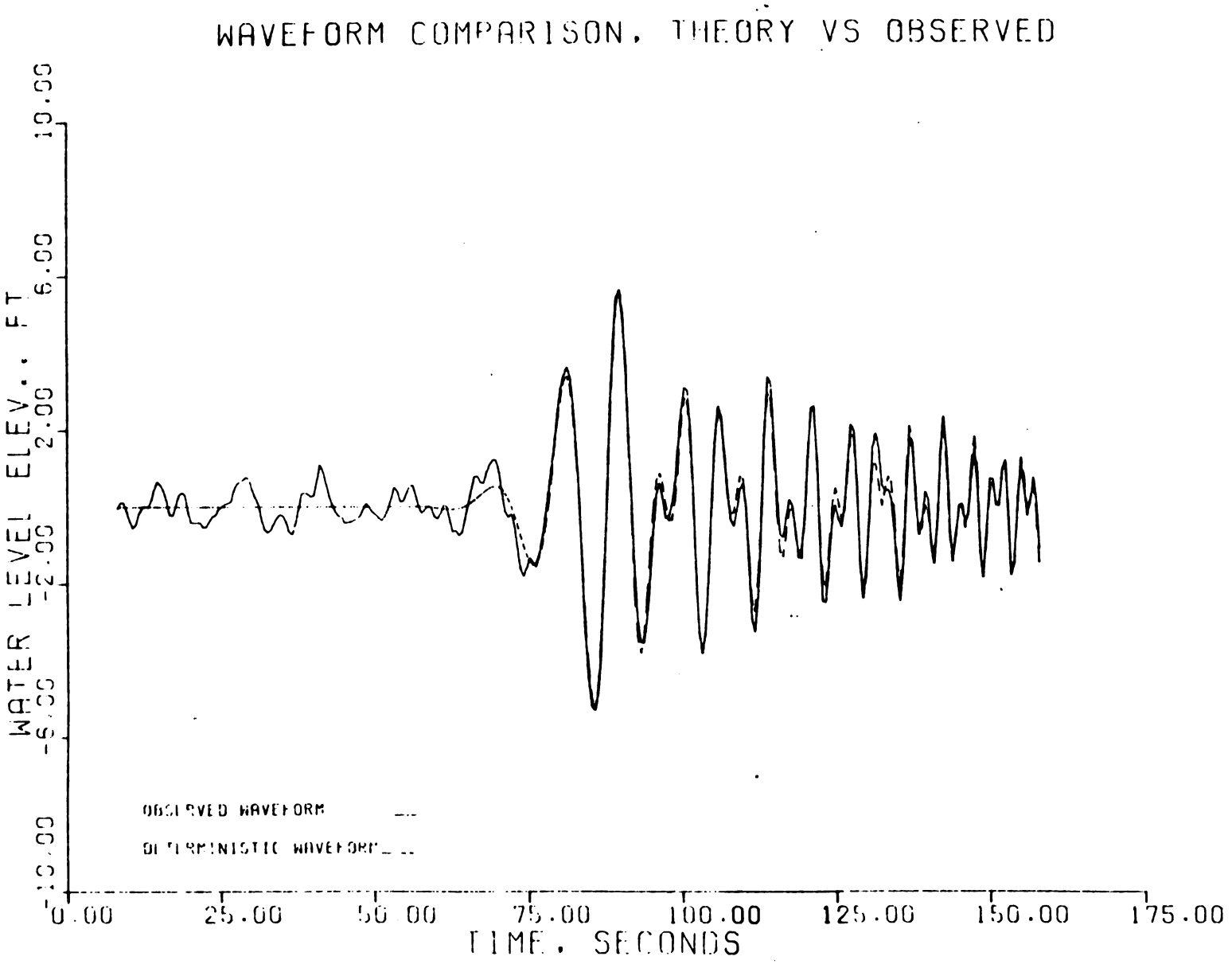


Figure B7. Time series plot for 800ft from the explosion source

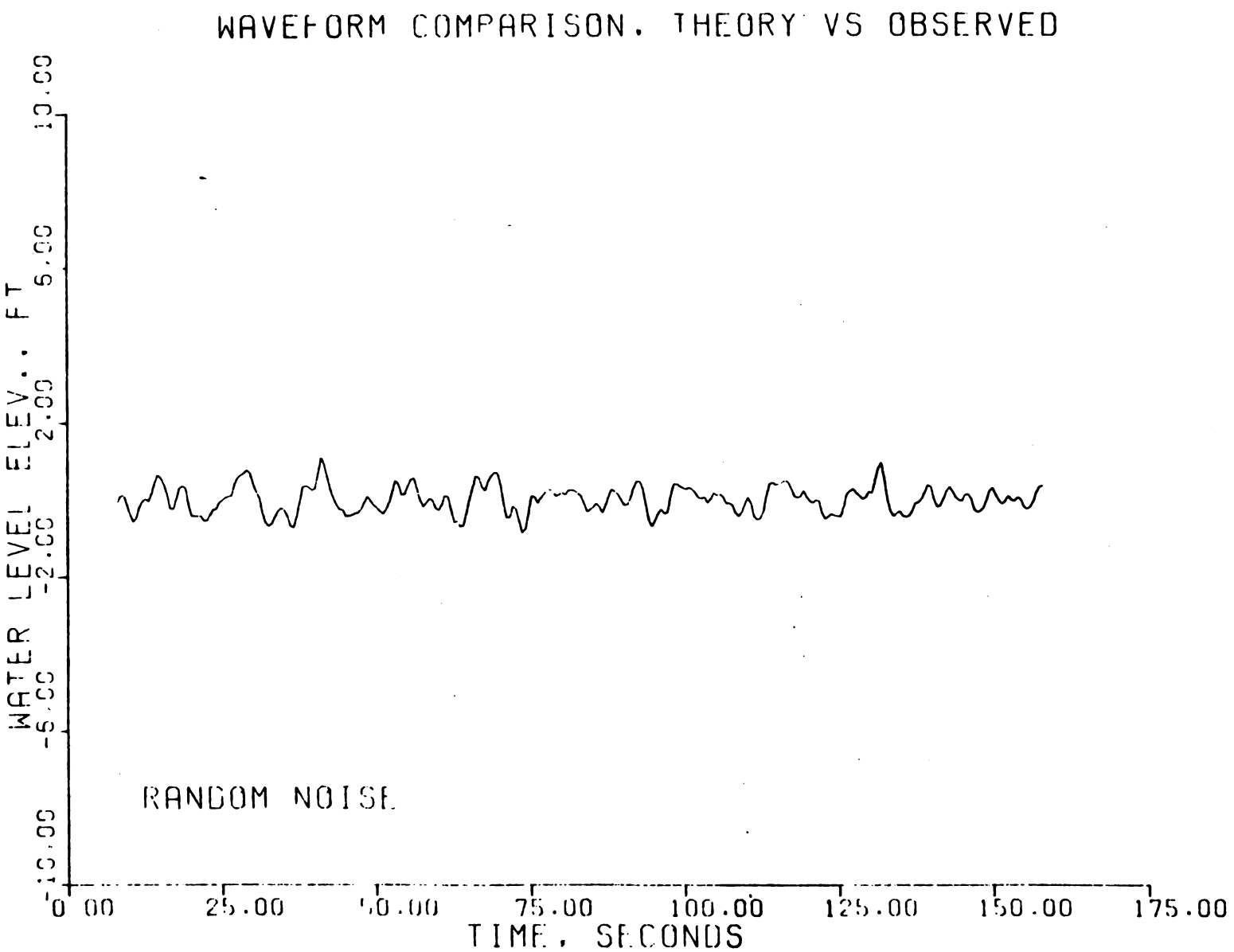


Figure B8. Random noise from Figure B7

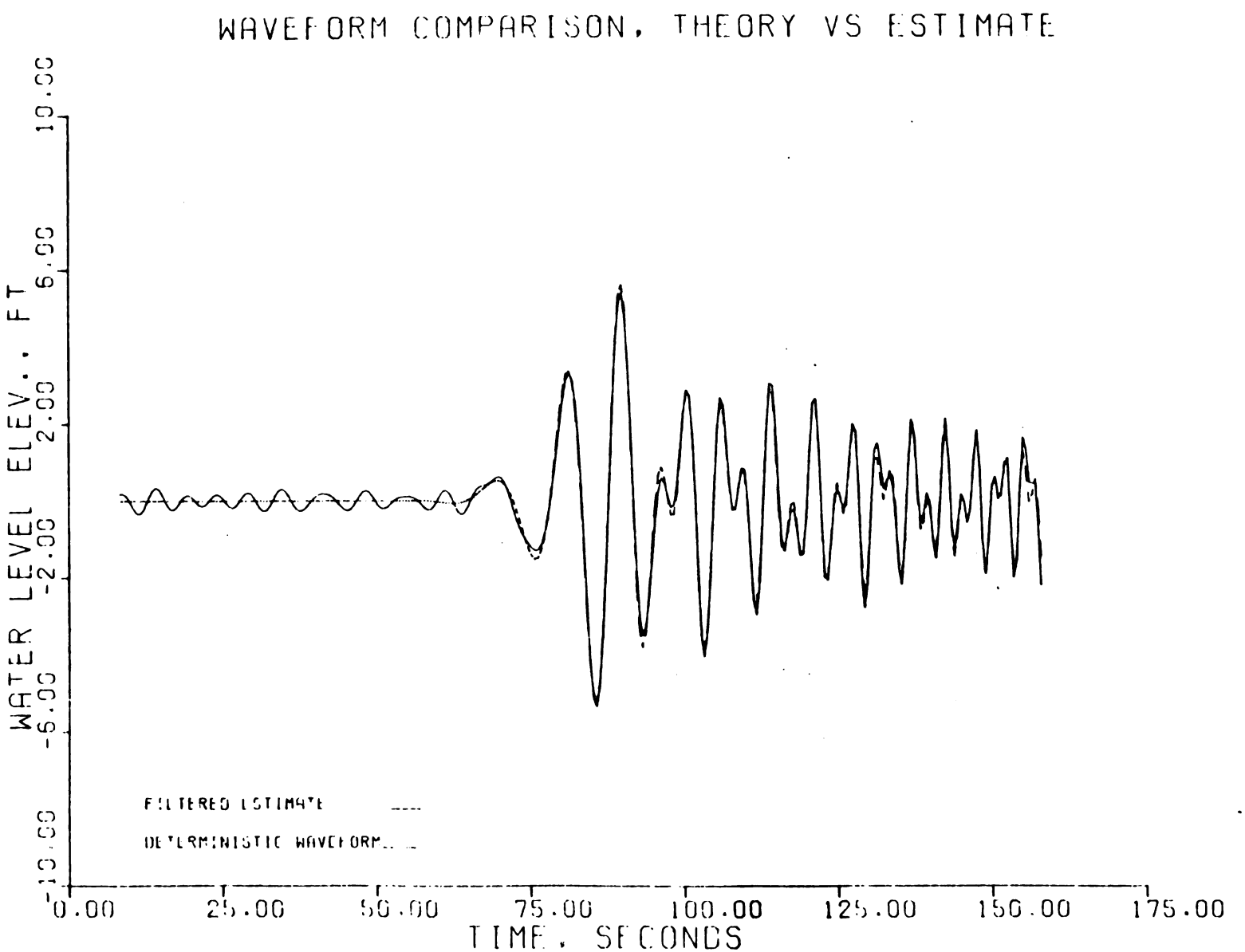


Figure B9. Time series plot for Boxcar Estimate at 800 ft from the explosion source

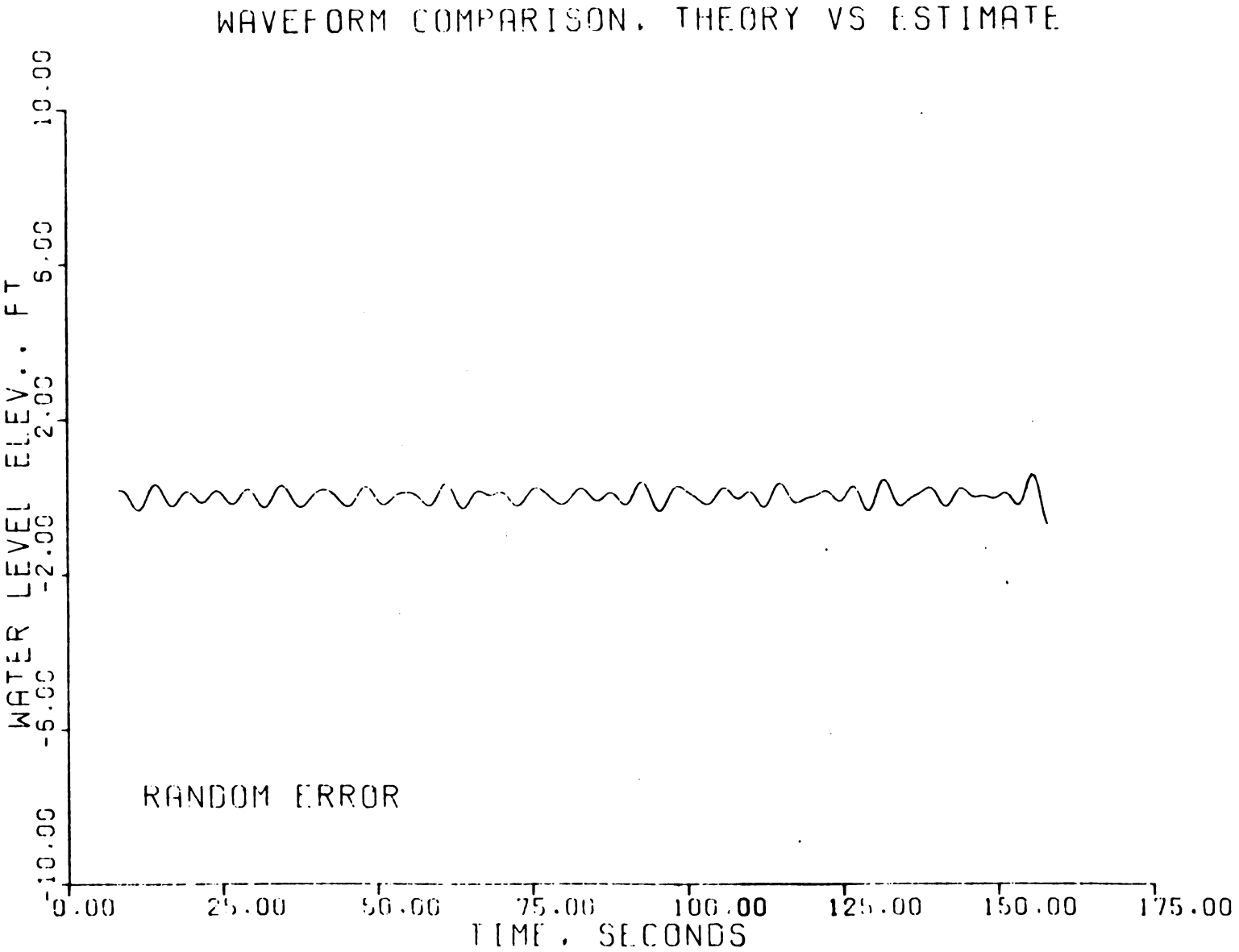


Figure B10. Random error for the estimated time series in Figure B9

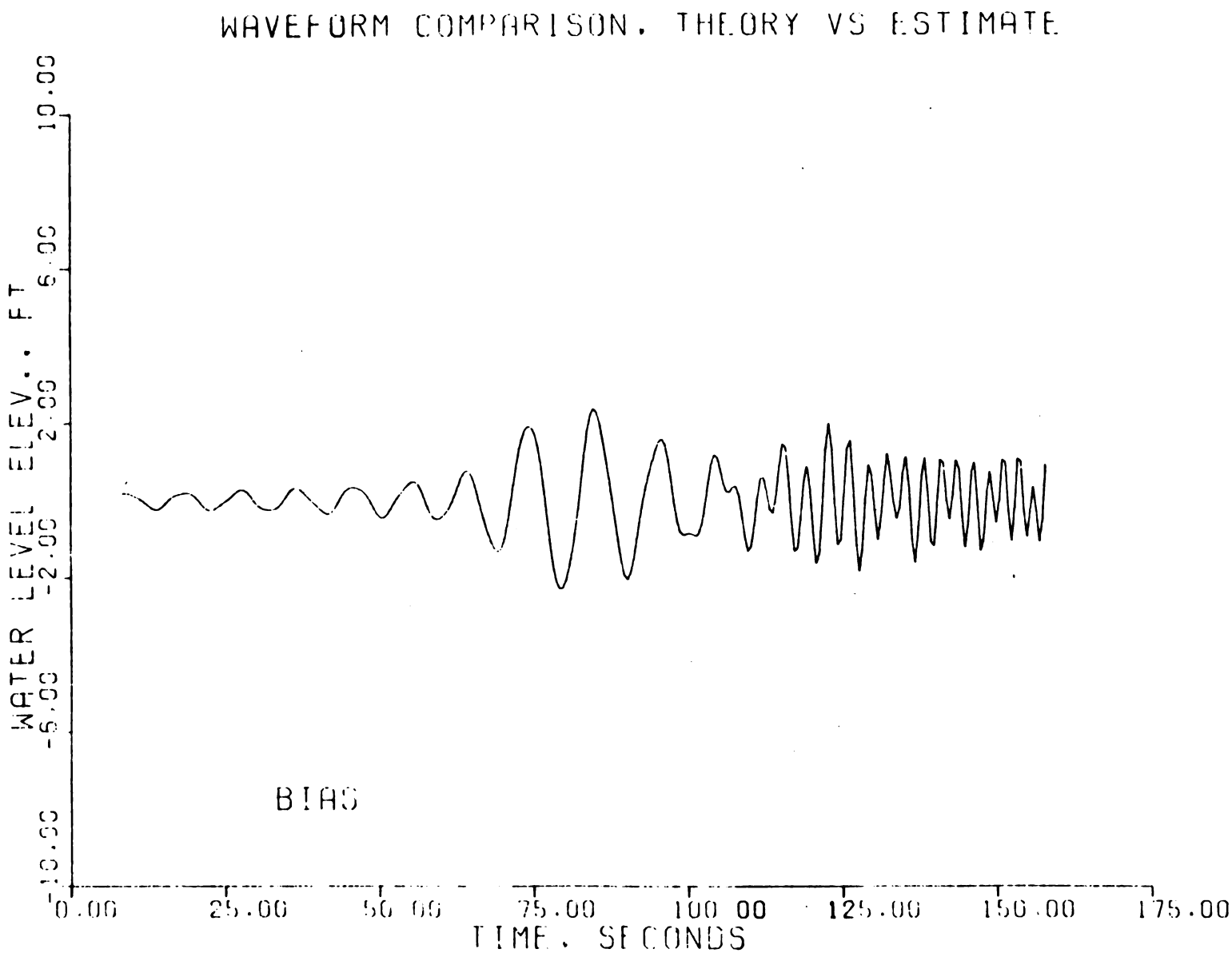


Figure B11. Bias time series for the Boxcar Estimate at 800 ft from the explosion source

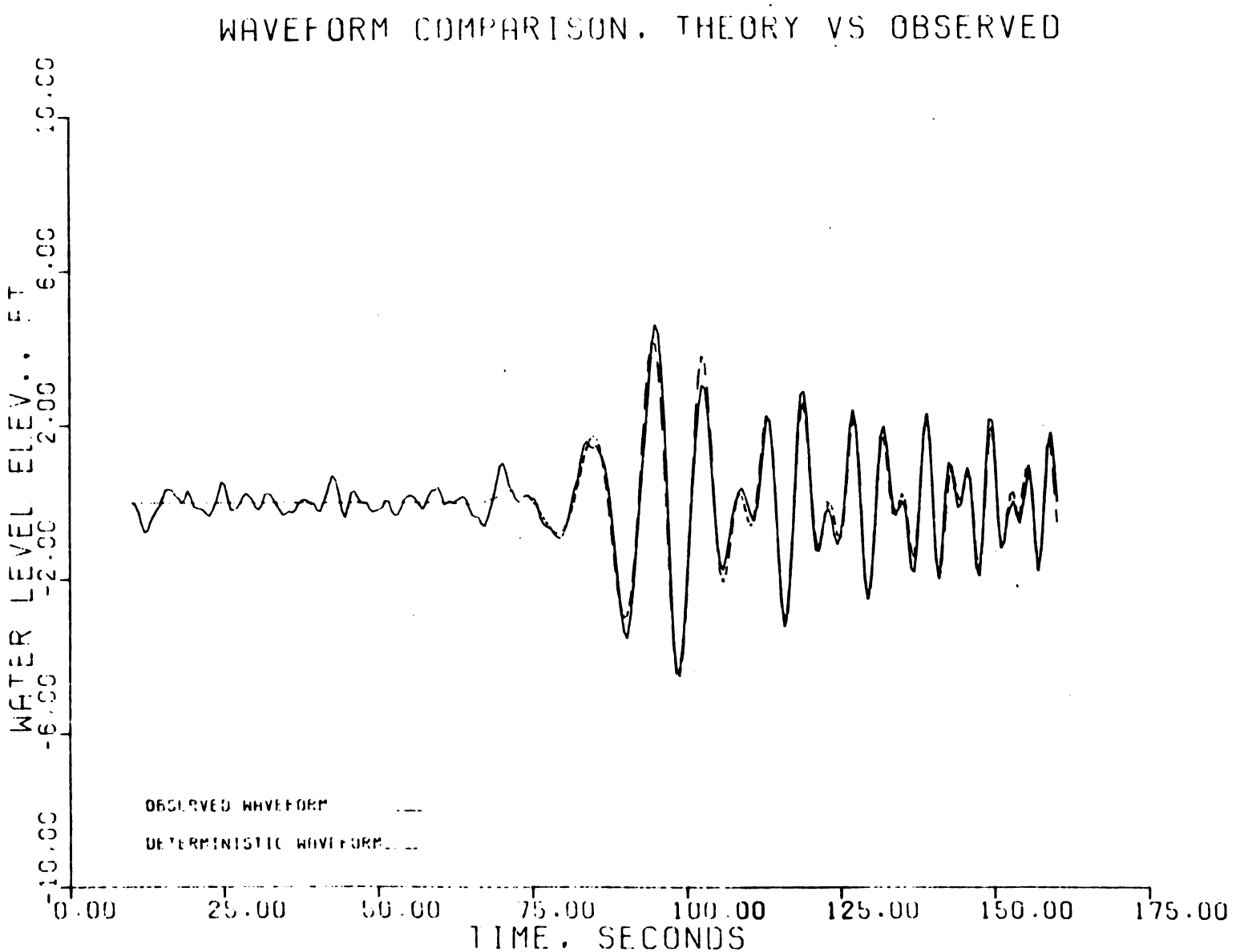


Figure B12. Time series plot for 1000 ft from the explosion source

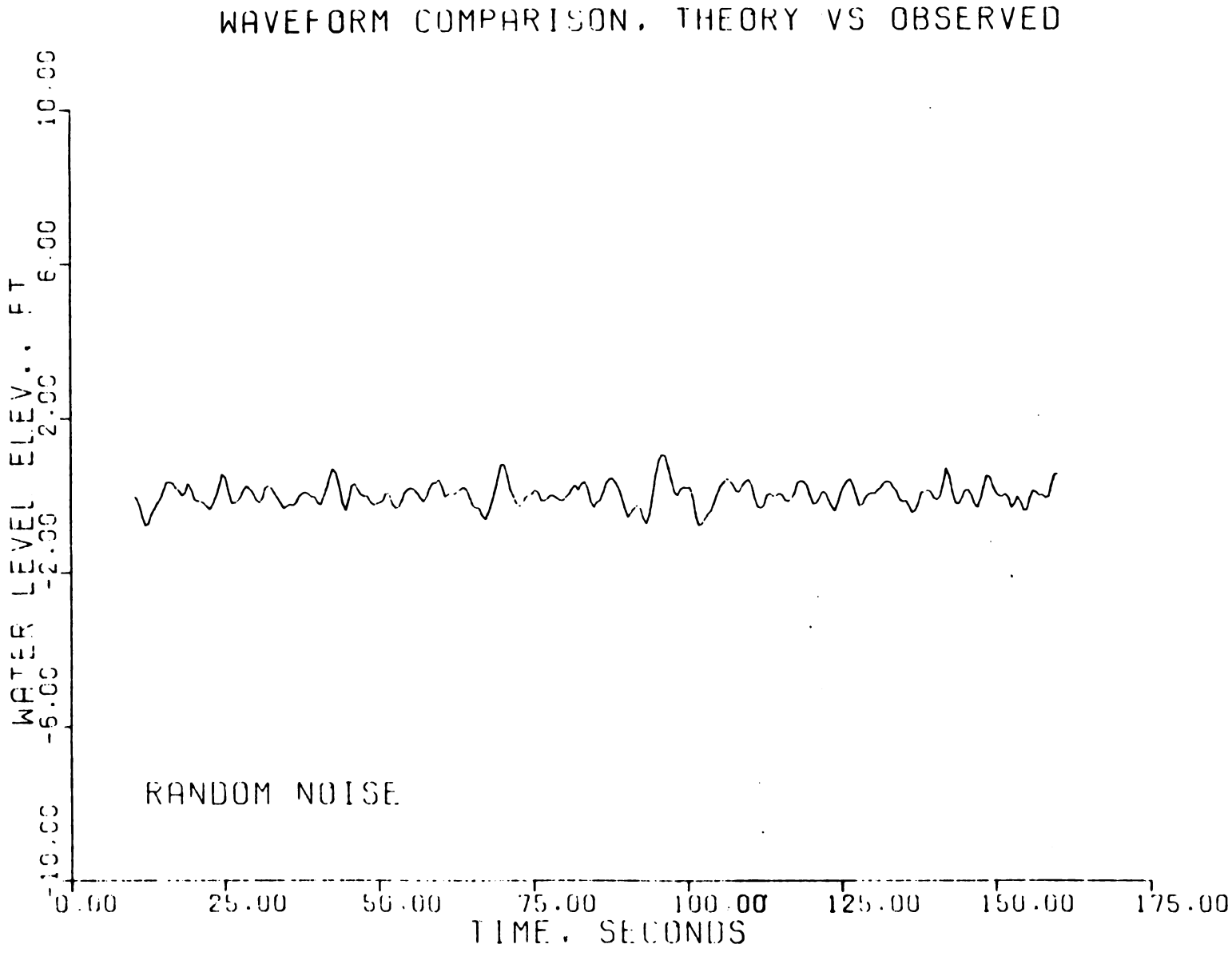


Figure B13. Random noise from Figure B12

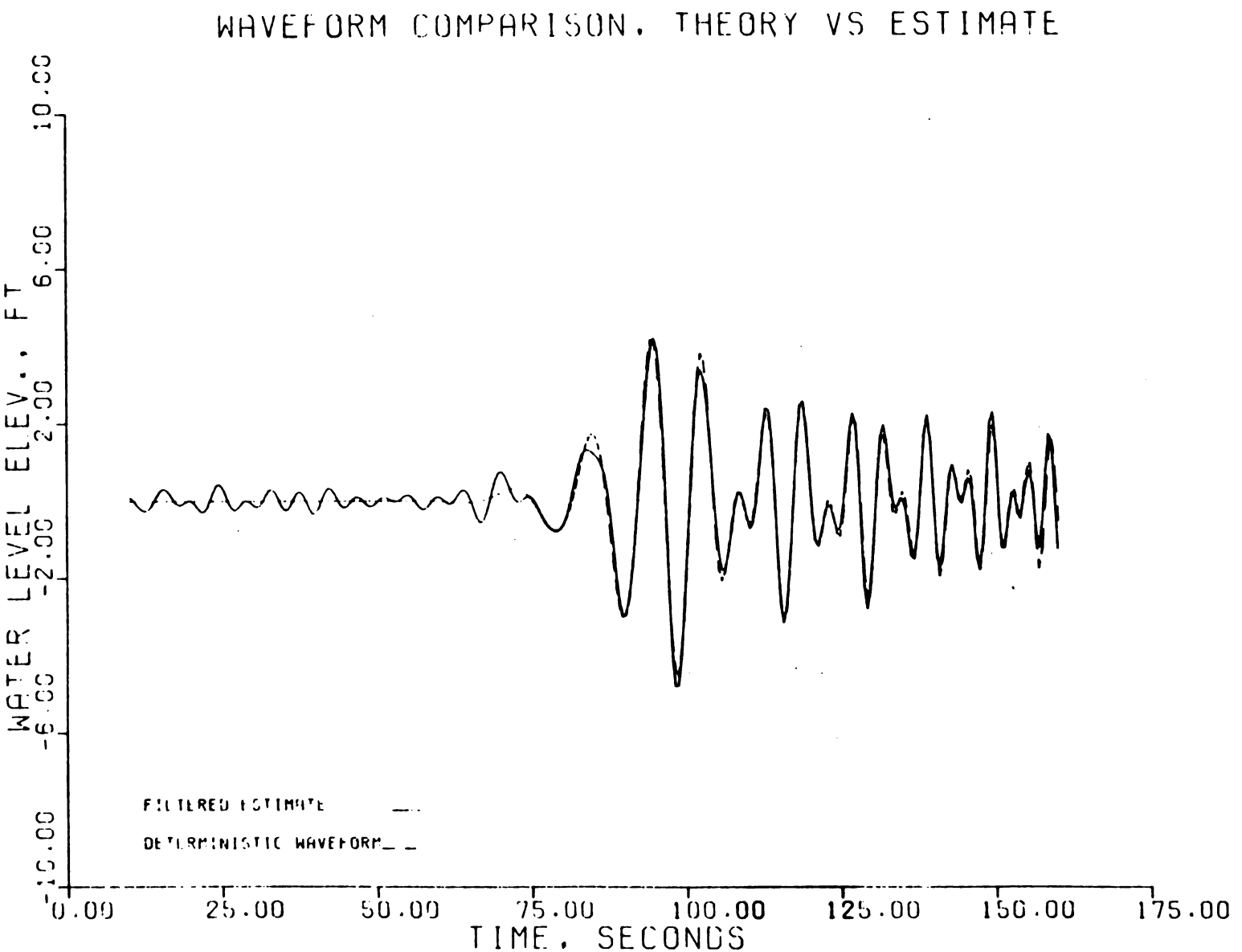


Figure B14. Time series plot for the Boxcar Estimate at 1000 ft fom the explosion source

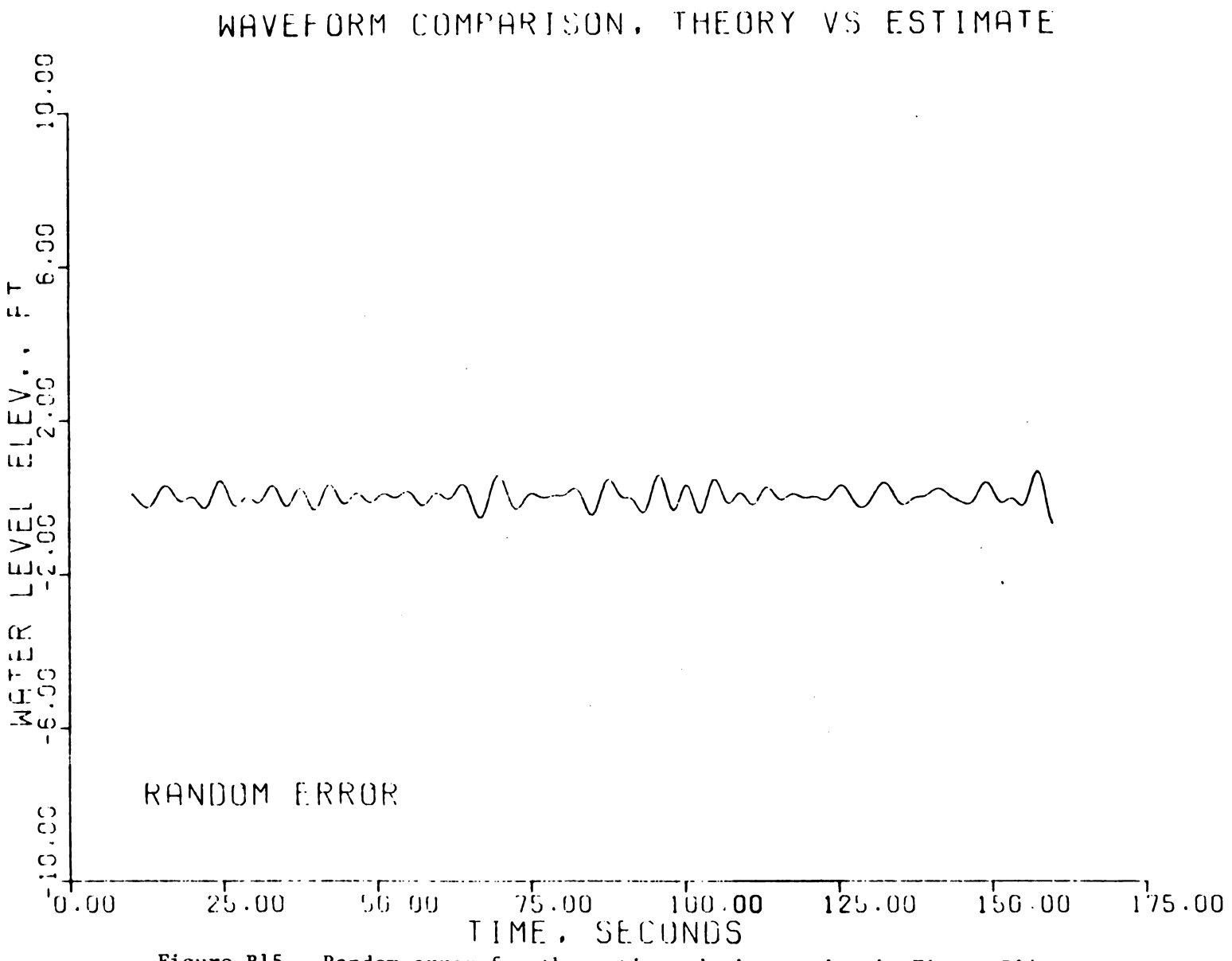


Figure B15. Random error for the estimated time series in Figure B14

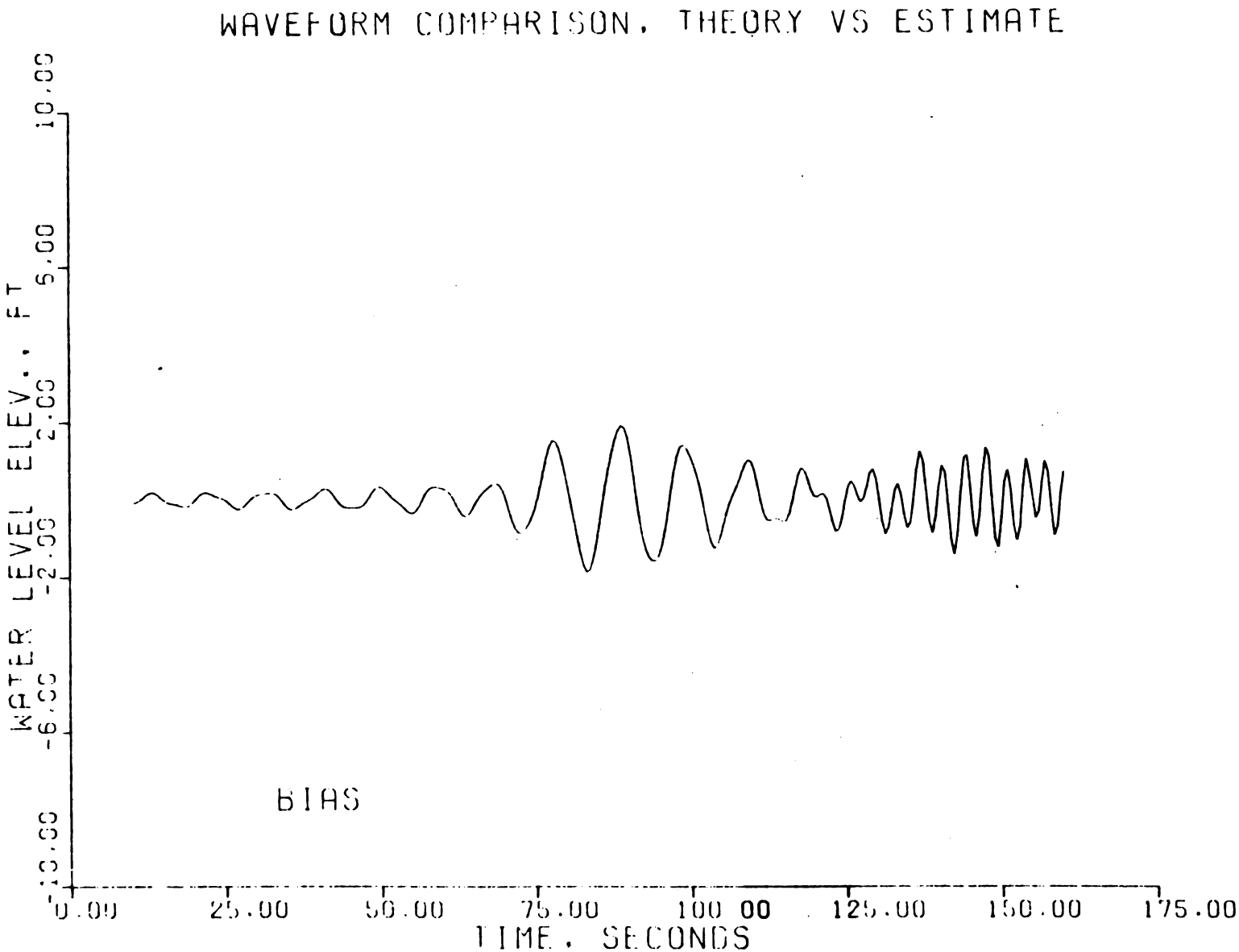


Figure B16. Bias time series for the Boxcar Estimate at 1000 ft from the explosion source

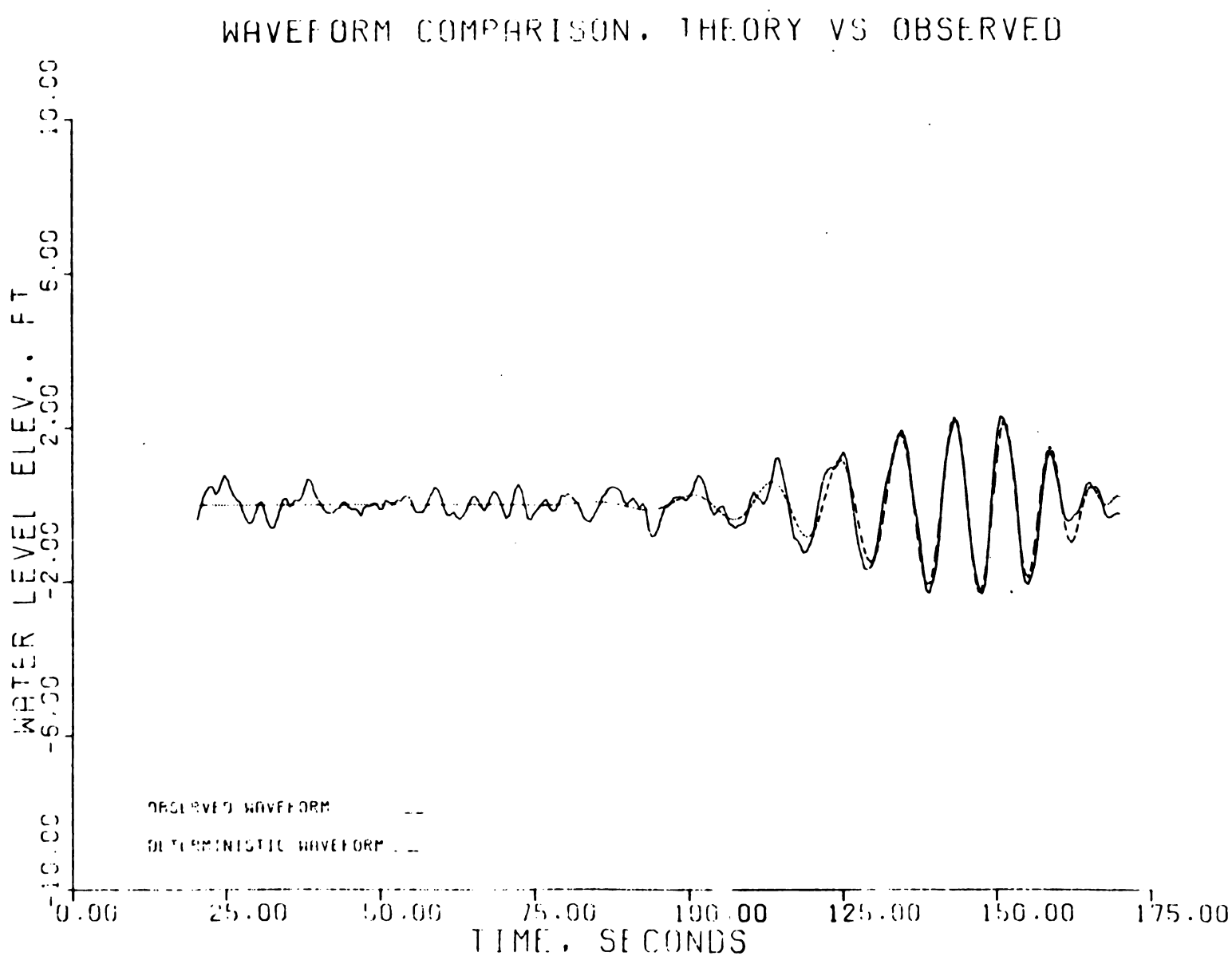


Figure B17. Time series plot for 2000 ft from the explosion source

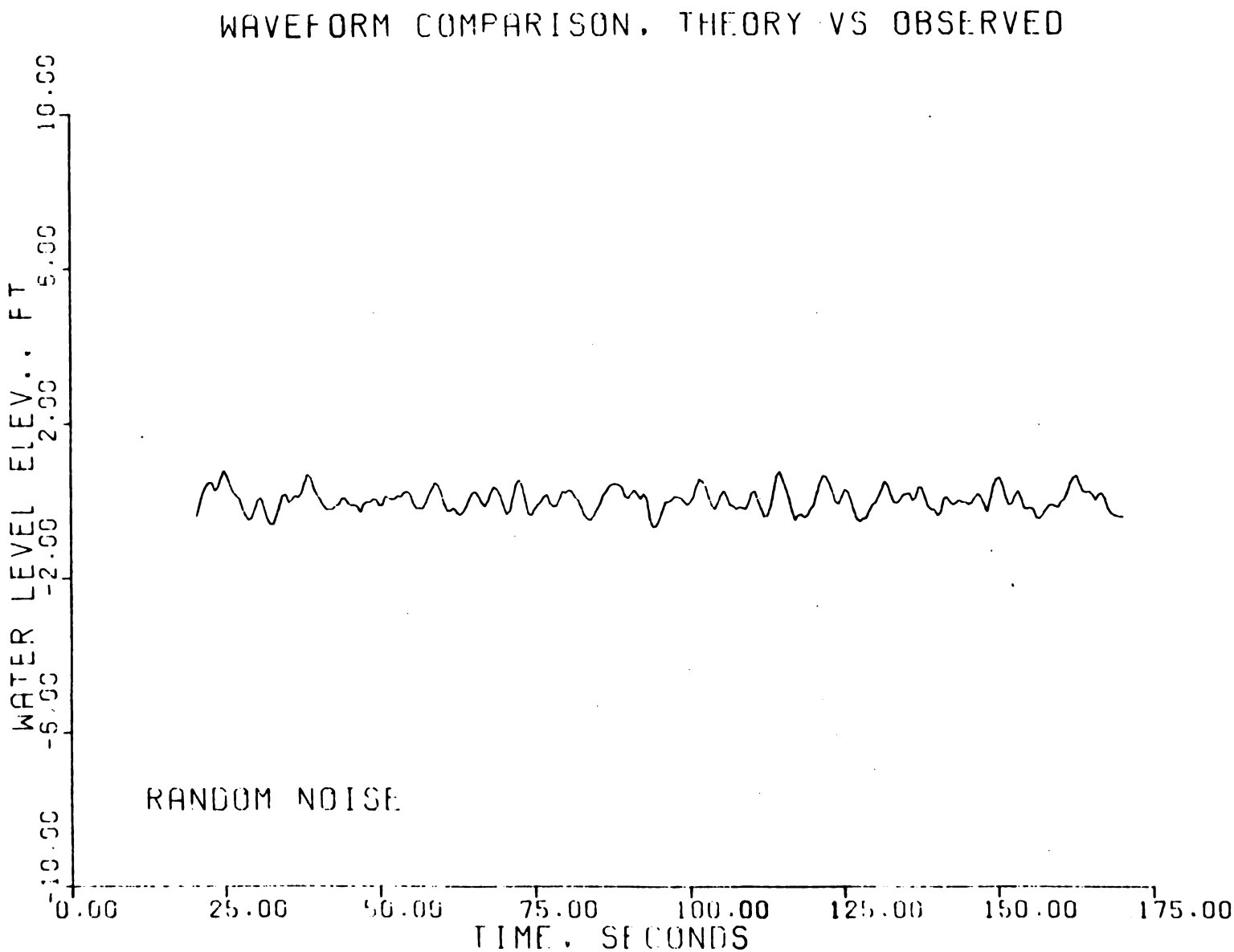


Figure B18. Random noise from Figure B17

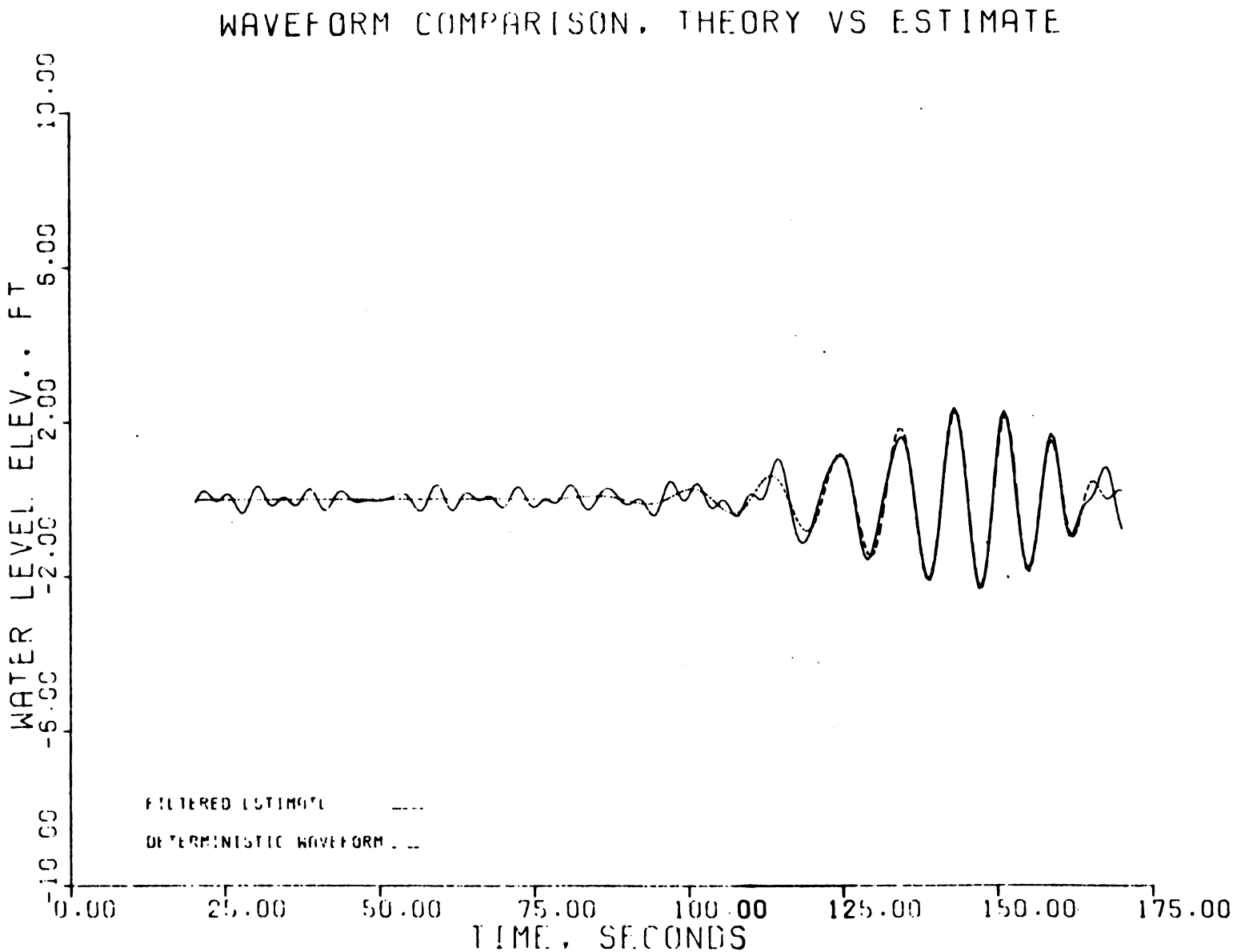


Figure B19. Time series plot for the Boxcar Estimate at 2000 ft from the explosion source

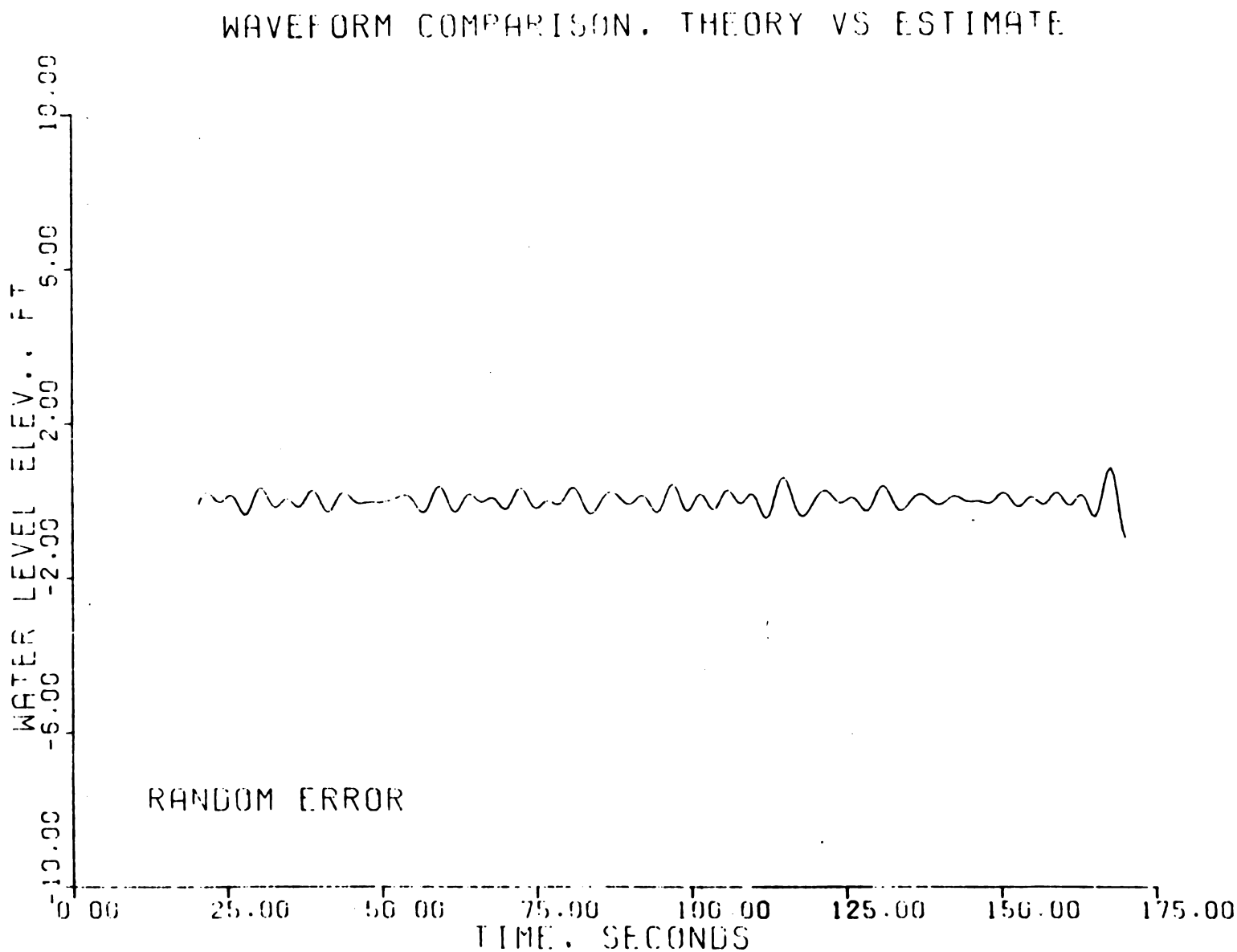


Figure B20. Random error for the estimated time series in Figure B19

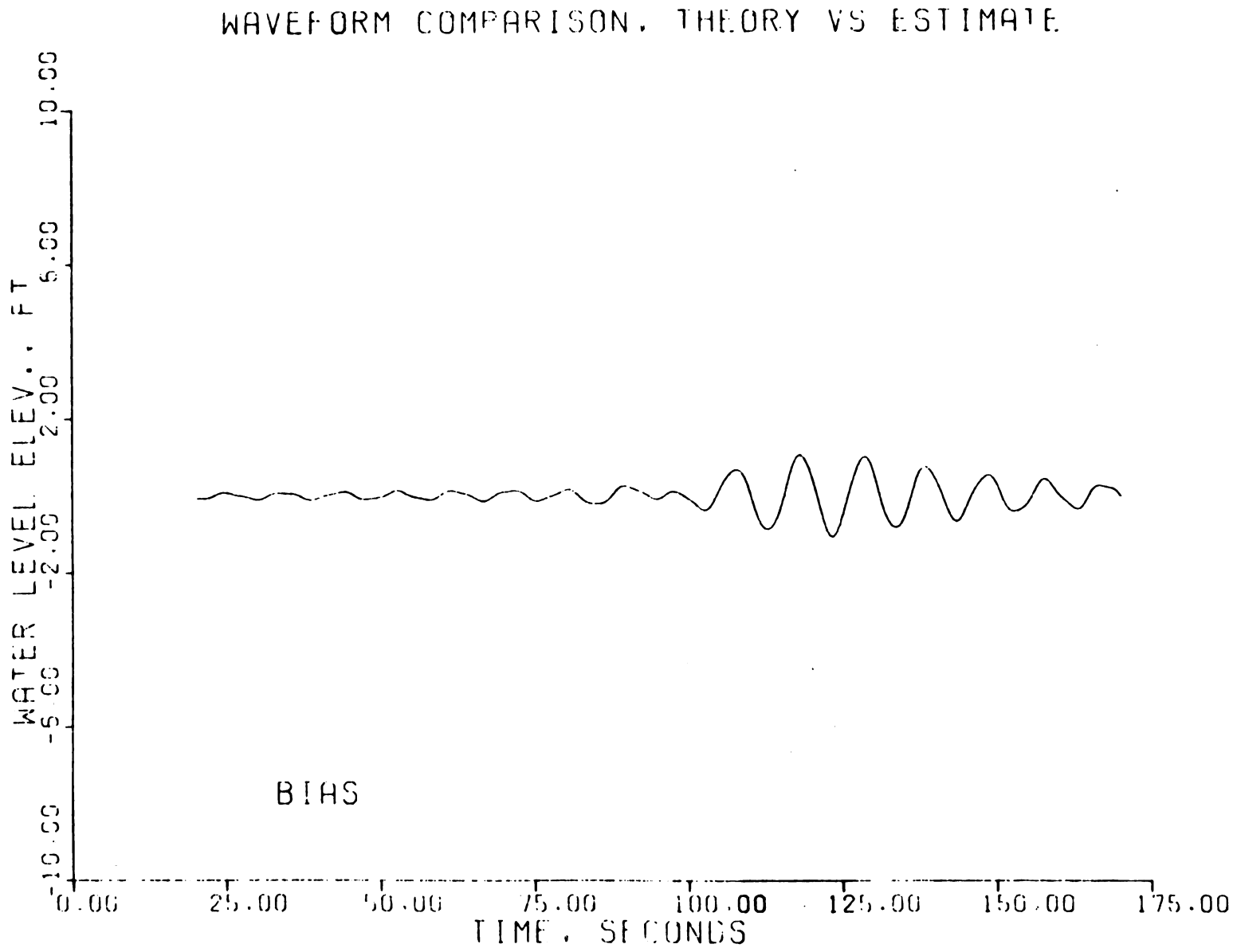


Figure B21. Bias time series for the Boxcar Estimate at 2000 ft from the explosion source

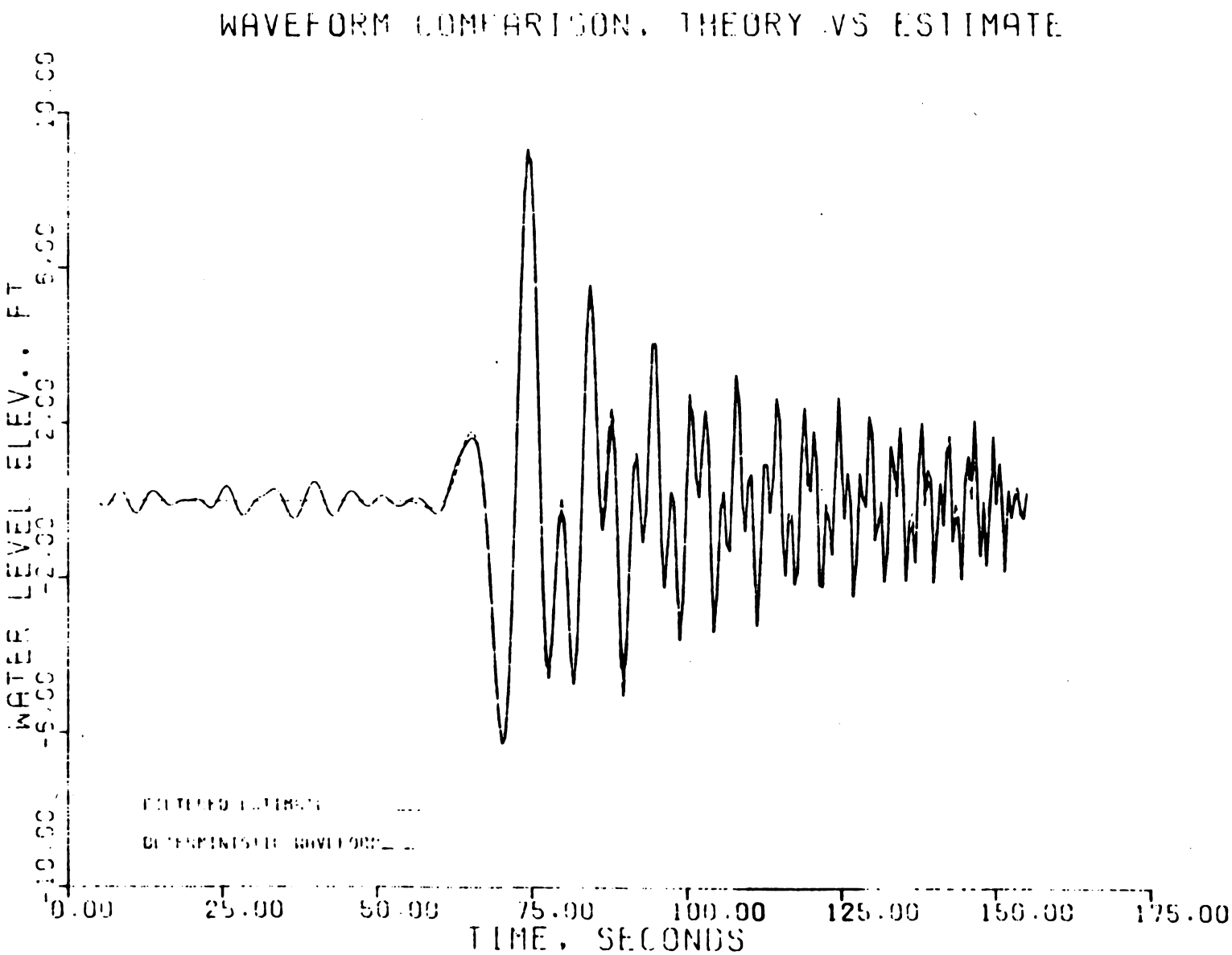


Figure B22. Time series plot for the Butterworth Estimate at 500 ft from the explosion source

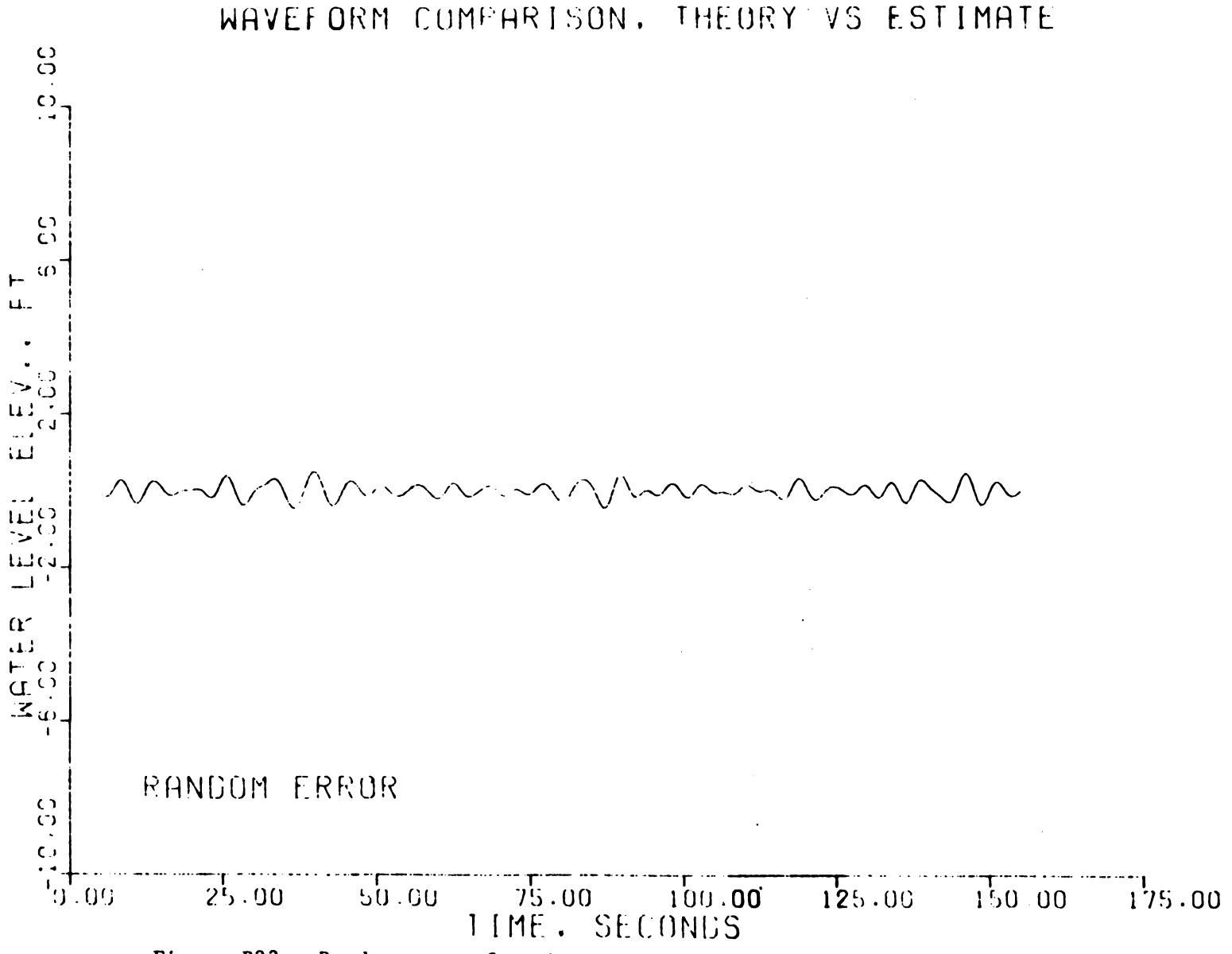


Figure B23. Random error for the estimated time series of Figure B22.

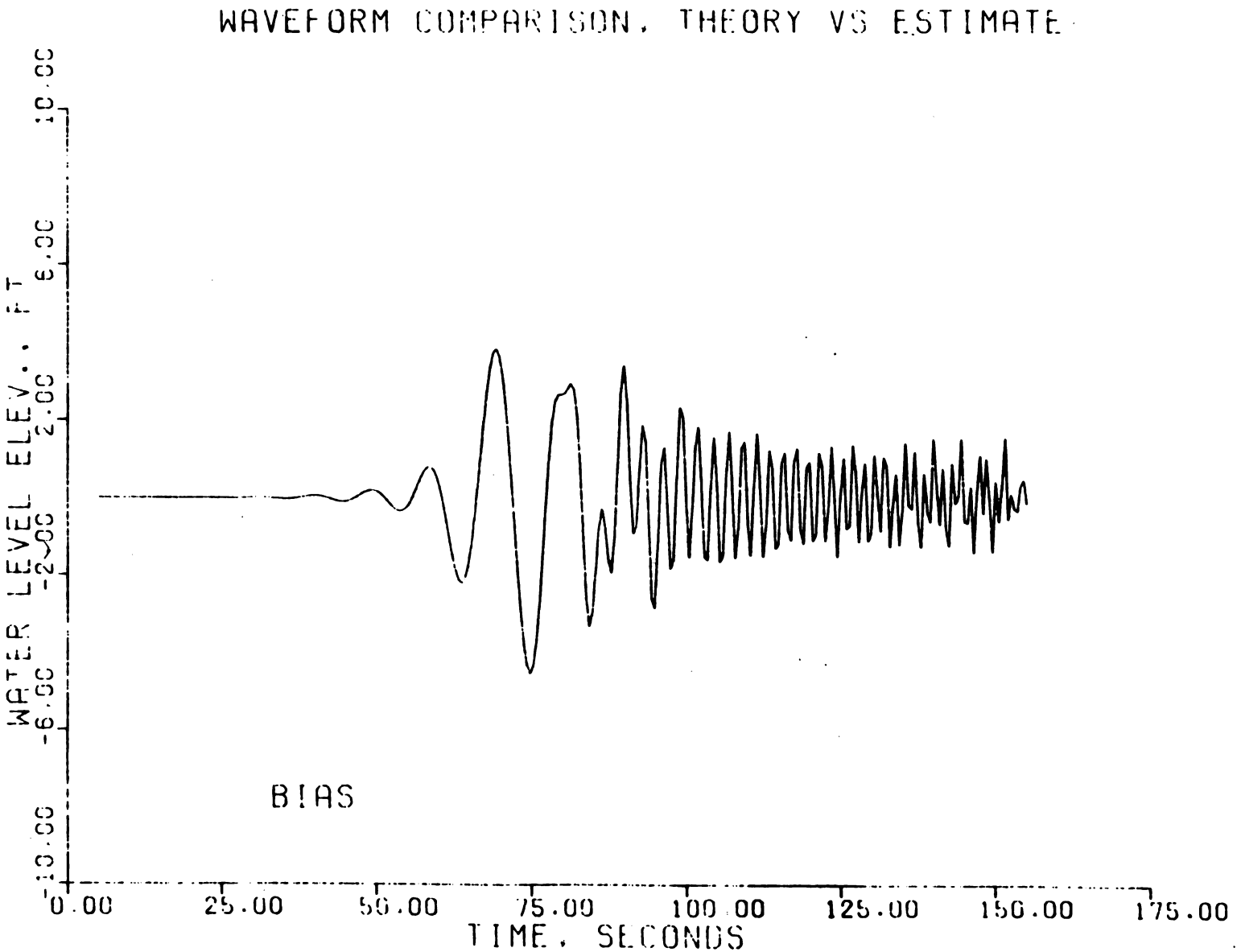


Figure B24. Bias time series for the Butterworth Estimate at 500 ft from the explosion source

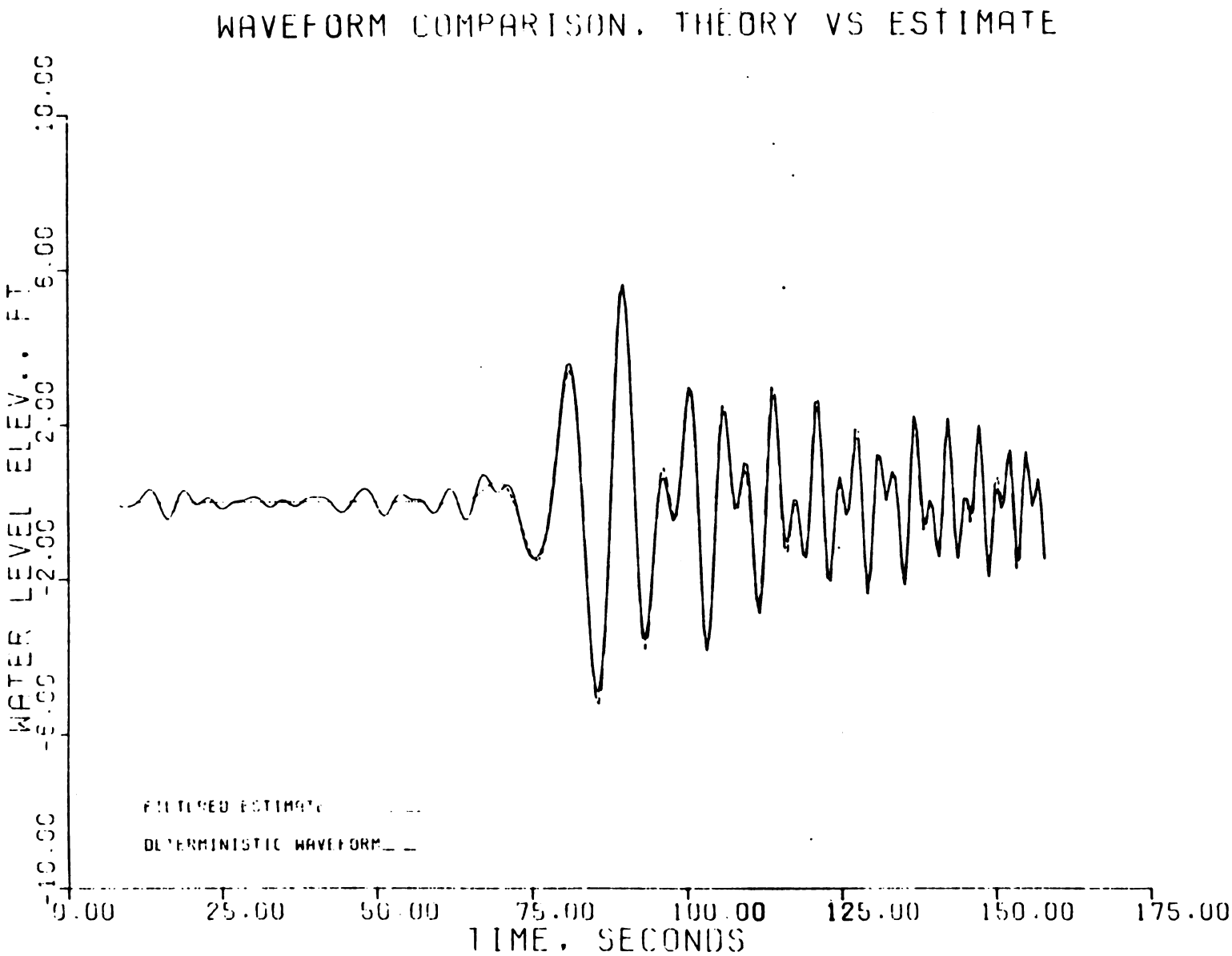


Figure B25. Time series plot for the Butterworth Estimate at 300 ft from the explosion source

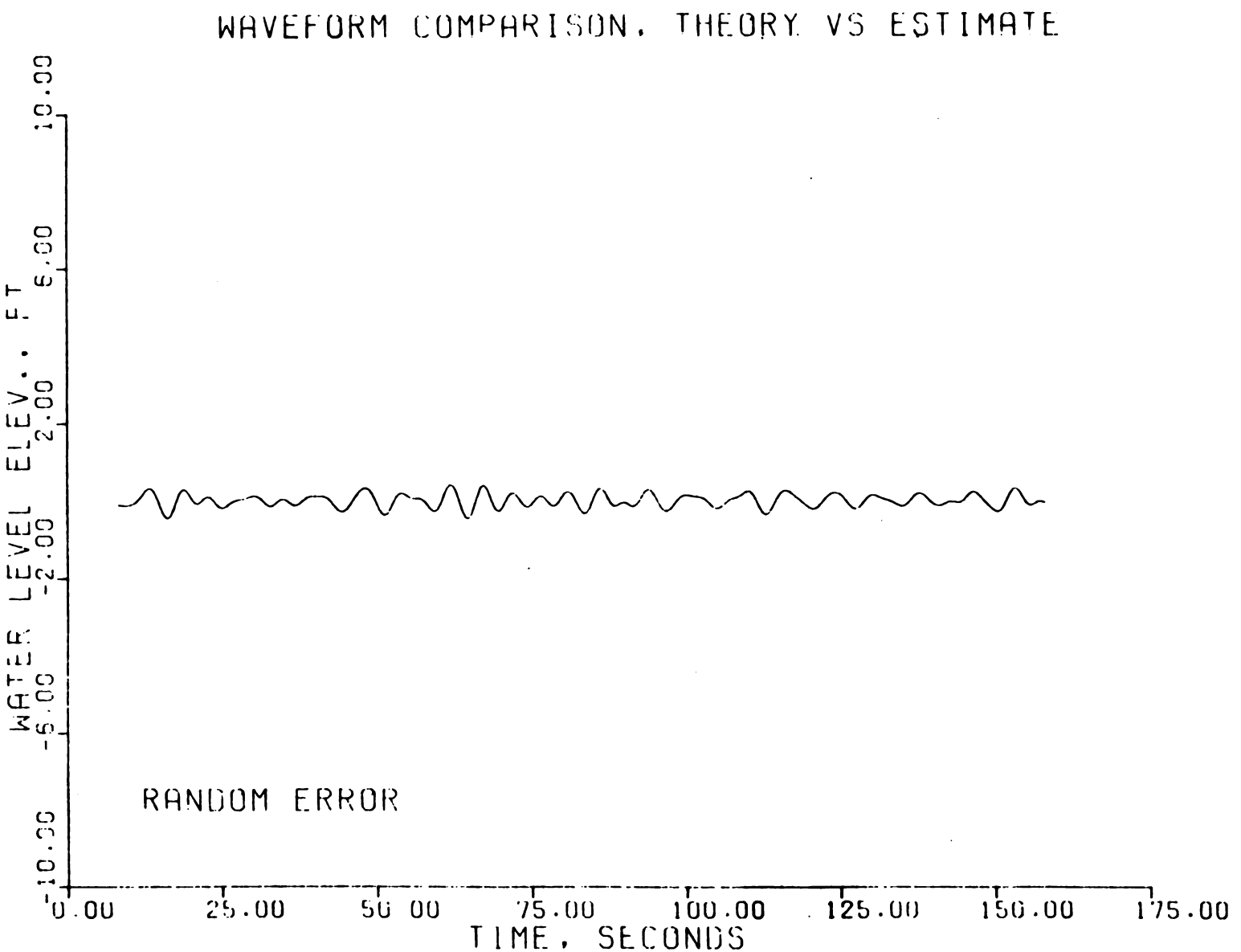


Figure B26. Random error from the estimated time series of Figure B25

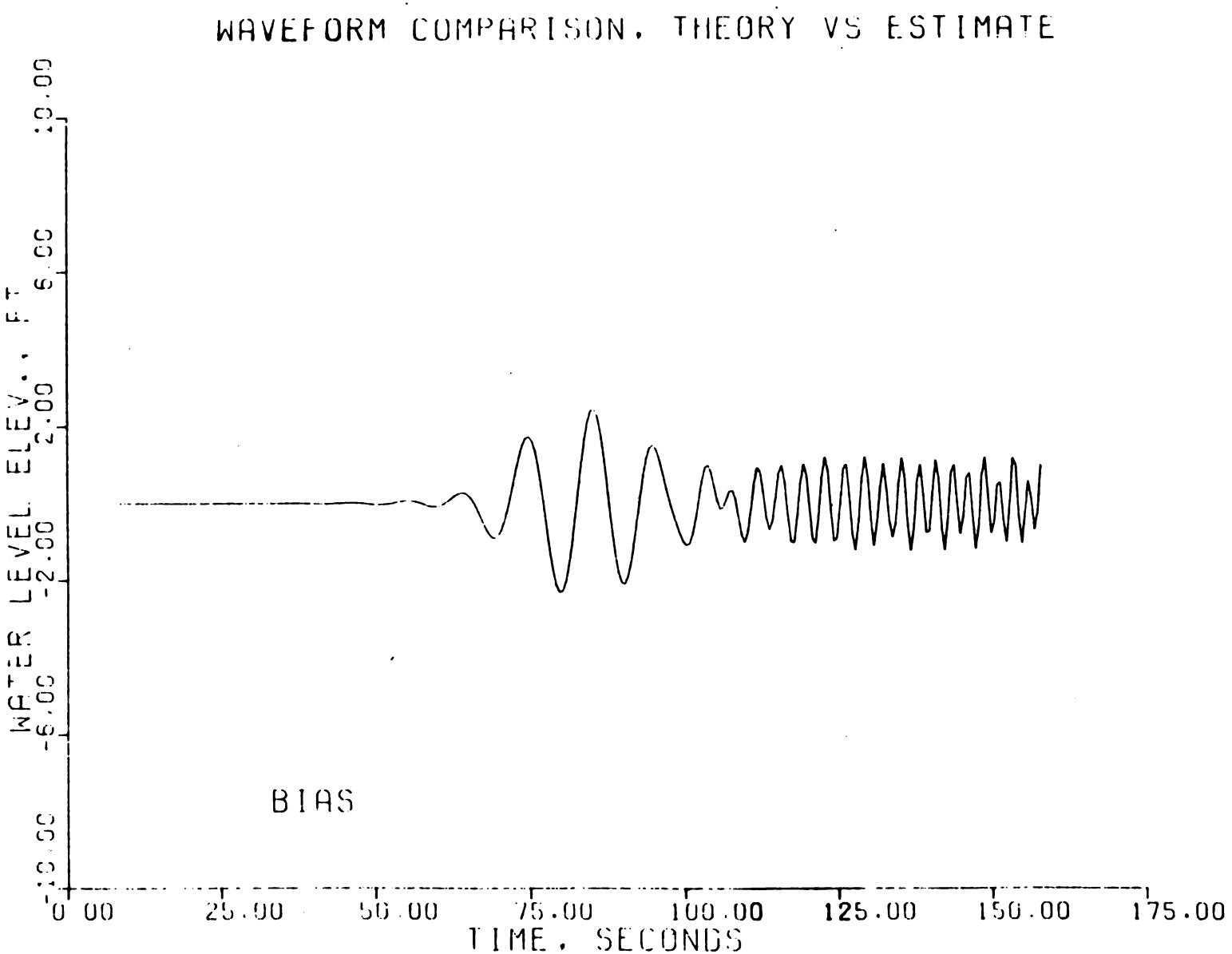


Figure B27. Bias time series for the Butterworth Estimate at 800 ft from the explosion source

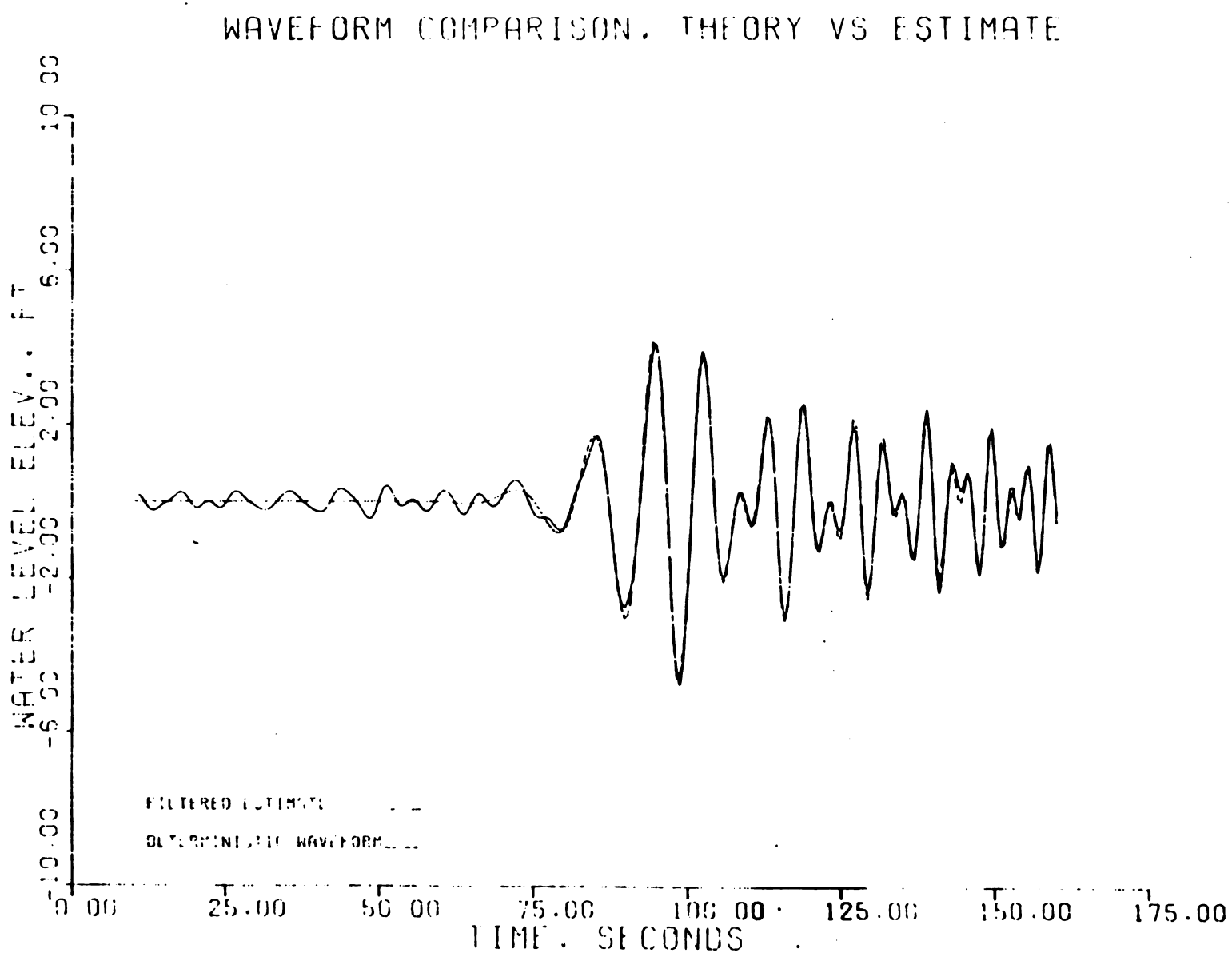


Figure B28. Time series plot for the Butterworth Estimate at 1000 ft from the explosion source

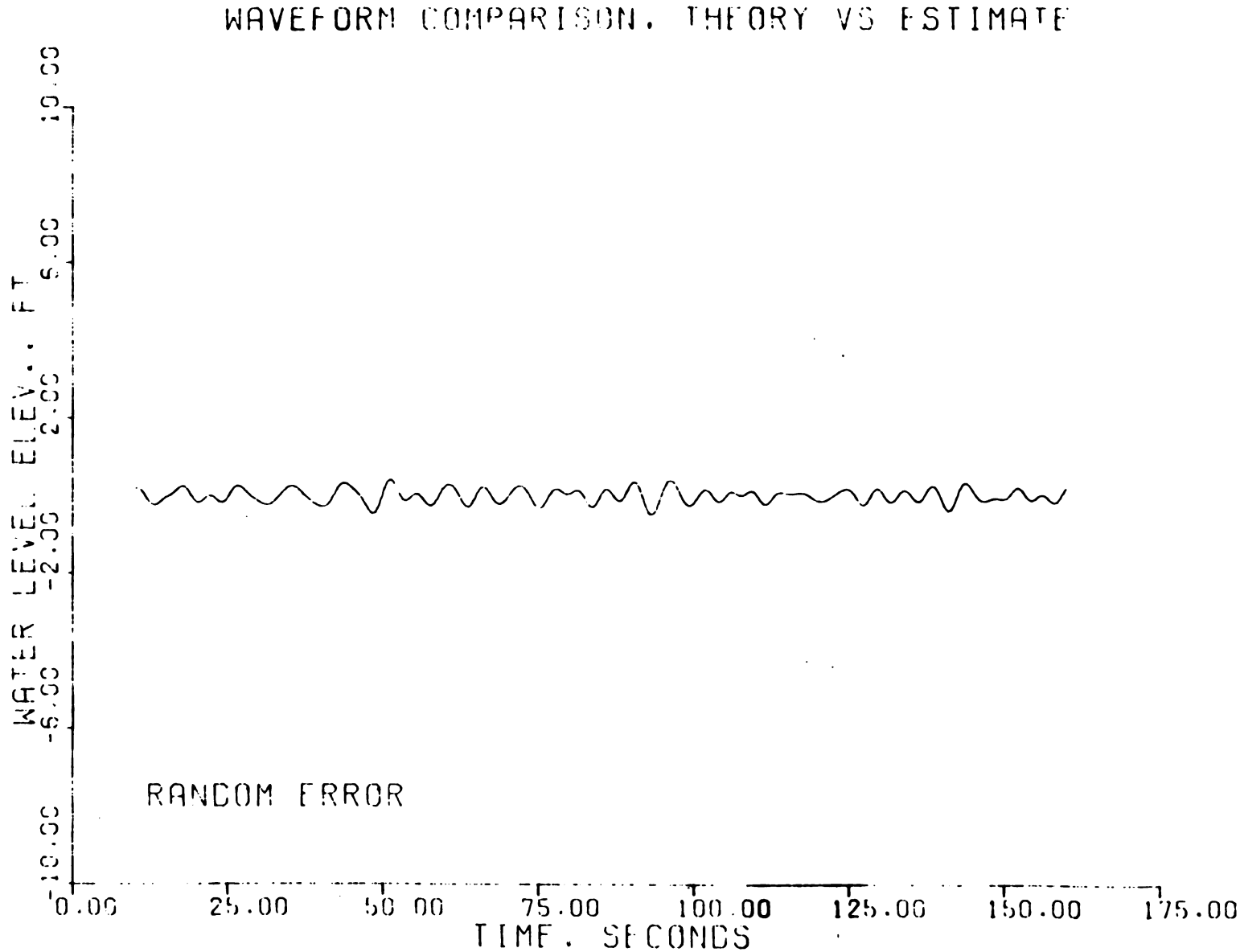


Figure B29. Random error from the estimated time series of Figure B28

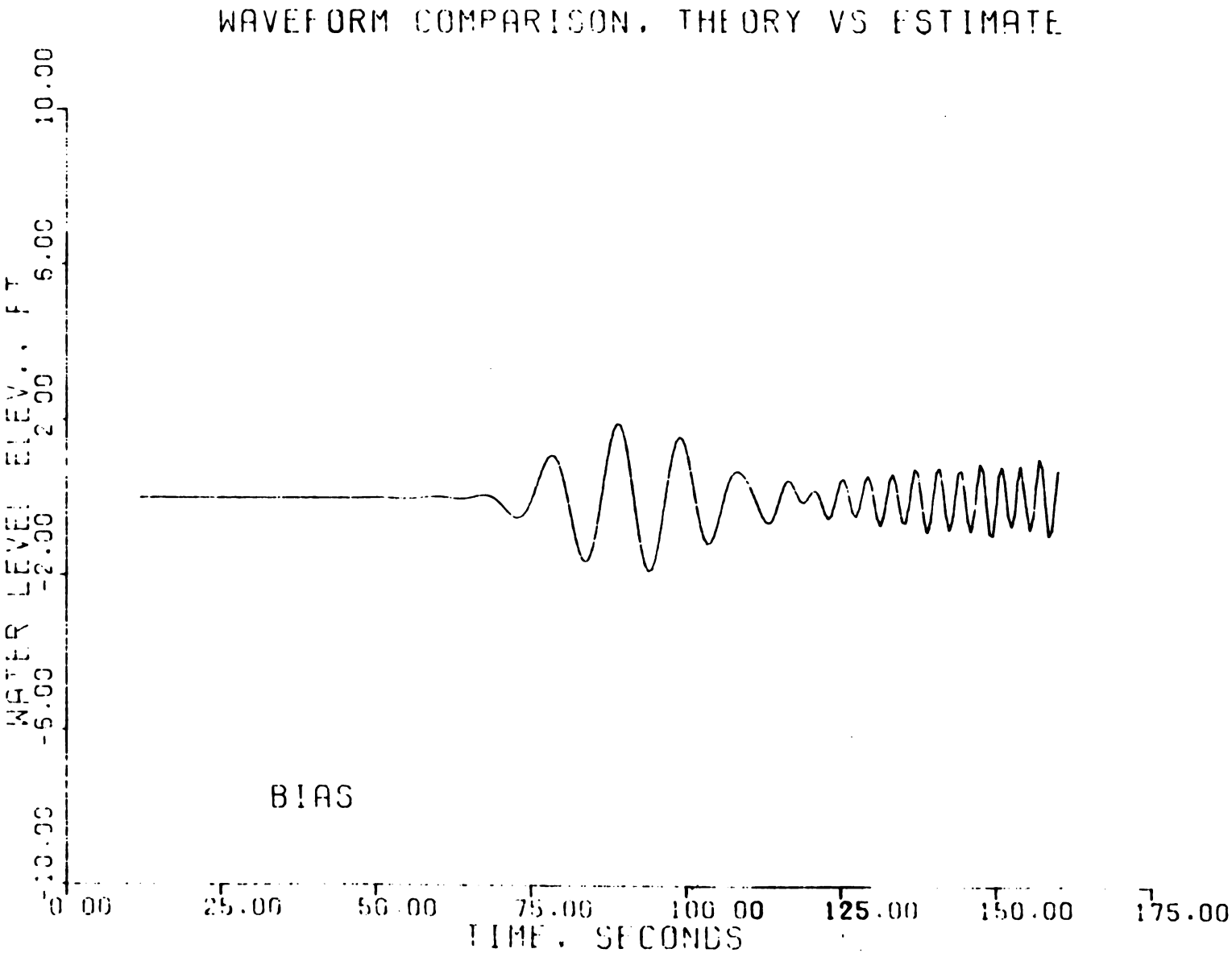


Figure B30. Bias time series for the Butterworth Estimate at 1000 ft from the explosion source

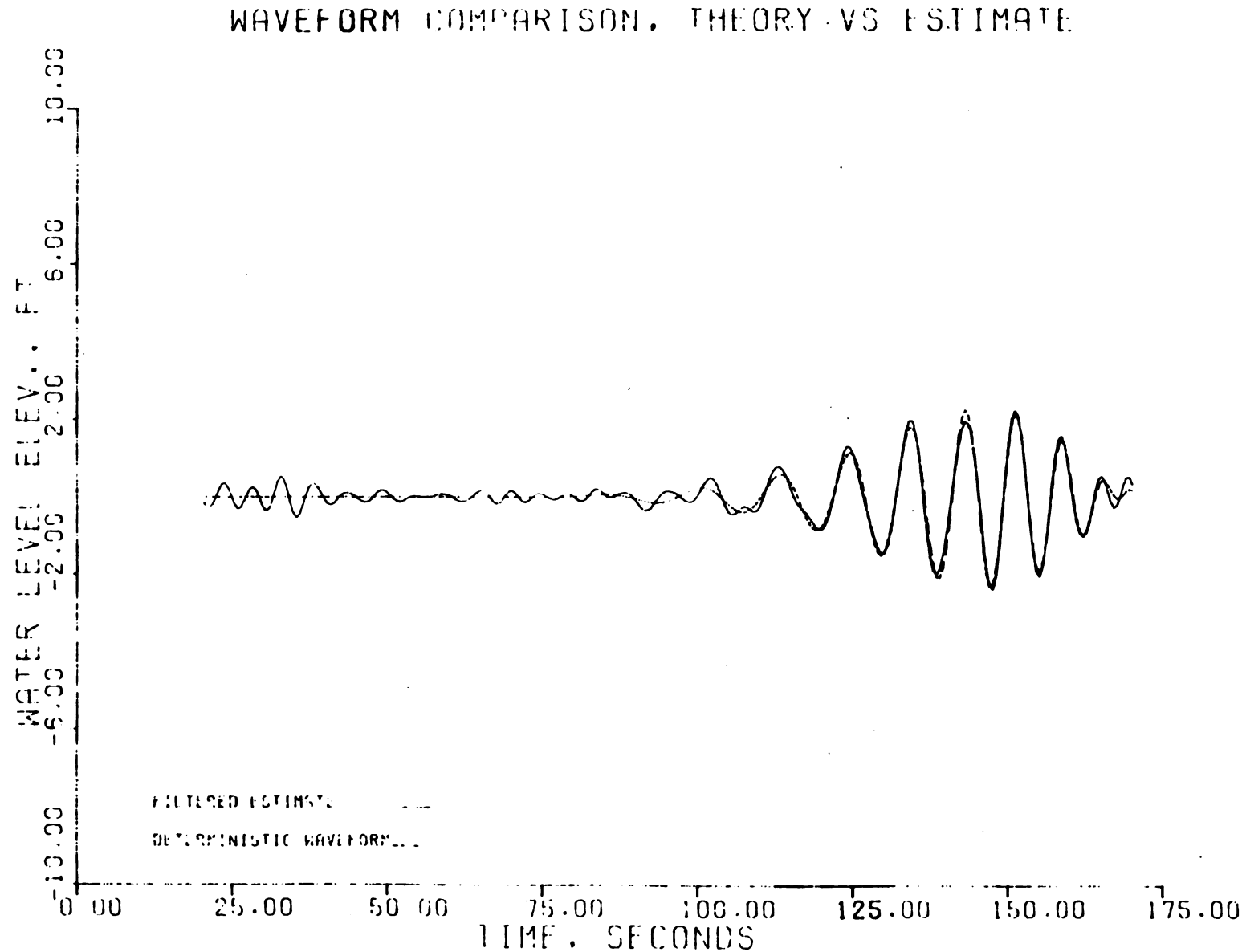


Figure B31. Time series plot for the Butterworth Estimate at 2000 ft from the explosion source

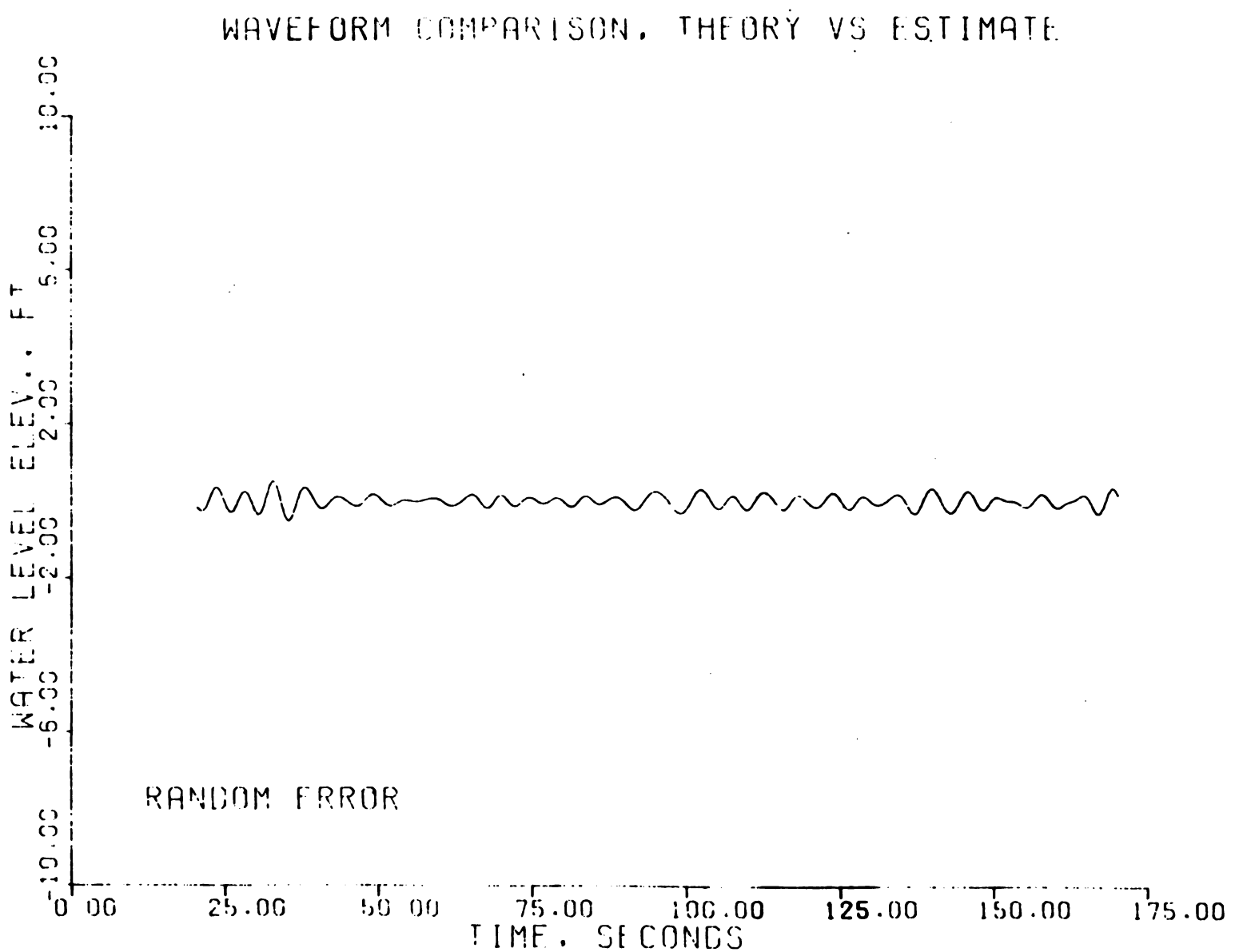


Figure B32. Random error from the estimated time series of Figure B31

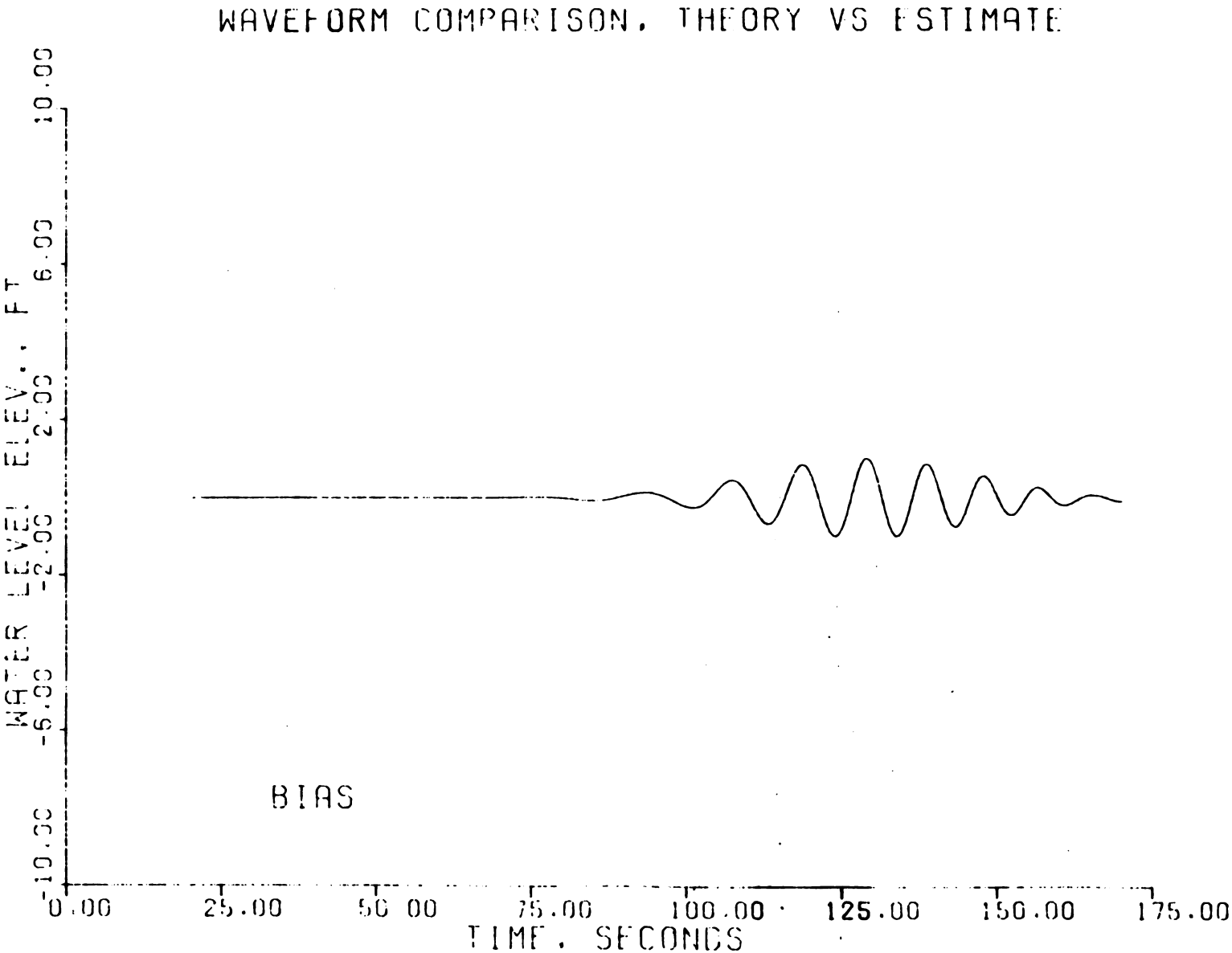


Figure B33. Bias time series for the Butterworth Estimate at 2000 ft from the explosion source

A VARIABLE SELECTION MODEL BUILDING TECHNIQUE FOR RADAR MEASUREMENT BIAS ESTIMATION

WILLIAM S. AGEE
ANDREW C. ELLINGSON, JR.
ROBERT H. TURNER

National Range Operations Directorate
White Sands Missile Range, New Mexico

ABSTRACT. Measurements of range, azimuth, and elevation from several different radars are combined to estimate the cartesian coordinates of a vehicle trajectory. Since the sequence of times t_i, i=1, 2,---,N which cover the entire trajectory. Since the measurements are subject to systematic errors as well as random measurement errors, the systematic error parameters (biases) are also estimated. The resulting estimation problem is a combined linear and nonlinear problem in which the trajectory coordinates appear nonlinearly in the measurements and the biases appear as linear parameters in the measurements. Applications of the above estimation very often result in ill-conditioned linear equations for estimating the radar biases, producing erroneous bias estimates. The problem of ill-conditioning caused by multicollinearity among the terms included in the bias model, is treated by using a backward elimination method for the selection of independent variables to be included in the radar measurement bias model. The method is illustrated with examples from WSMR radar tracking missions.

KEY WORDS. Multicollinearity, trajectory, estimation, regression, variable selection, stepwise regression

A VARIABLE SELECTION MODEL BUILDING TECHNIQUE FOR RADAR MEASUREMENT BIAS ESTIMATION

INTRODUCTION.

The WSMR radar Best Estimate of Trajectory (BET) program combines the range, azimuth, and elevation measurements from all radars tracking an object to optimally estimate a set of smoothed cartesian positions, velocities, and accelerations of the object at each measurement time t_i. The radar measurements are subject to systematic error as well as a random error component. The radar systematic errors are usually referred to as measurement biases. When an object is being tracked by multiple radars and the relative geometry between the object trajectory and the radars is sufficiently good, a measurement bias for each range, azimuth, and elevation measurement can be estimated. It is important, sometimes absolutely necessary, for the success of the multiple radar BET reduction that estimates of the radar measurement bias be obtained. When these measurement biases are estimated by least squares, the resulting bias estimates are often erroneous due to numerical ill-conditioning of the least squares estimation problem and multicollinearity between the terms included in the bias model caused by overfitting.

We have attempted to treat this problem of erroneous bias estimates by application of ridge regression techniques [1] and by application of the method of principal components [2]. We have had partial success with each of these methods but neither has proved satisfactory for automatic trajectory data reduction. Techniques are

developed in this report which successfully treat the problems of ill-conditioning and overfitting in the radar measurement bias estimation problem. These methods are being incorporated into the WSMR radar BET. These techniques are based on the use of modern, reliable numerical linear algebra algorithms and software and on the development of a statistical model building technique which estimates only the radar bias terms which make a significant reduction in the error sum of squares.

The difficulties with the radar bias estimates are illustrated with some actual WSMR radar data sets. These data sets are treated by conventional least squares estimation methods and by the model building technique developed herein.

RADAR MEASUREMENT EQUATIONS

Let x, y, z be the coordinates of an object in a local cartesian coordinate system at the radar site. The ideal mathematical model of the radar measurements in terms of these cartesian coordinates is:

$$r(\bar{x}) = (x^2 + y^2 + z^2)^{\frac{1}{2}} \qquad \text{- range}$$

$$a(\bar{x}) = \tan^{-1} \frac{x}{y} \qquad \text{- azimuth}$$

$$e(\bar{x}) = \tan^{-1} \frac{z}{(x^2 + y^2)^{\frac{1}{2}}} \qquad \text{- elevation}$$

where $\bar{x}$ is the position vector with components x, y, z. The radar does not measure these ideal values of range, azimuth, and elevation, but is subject

to a measurement error which we assume to be additive and is composed of a systematic component and a random component. Let b_R, b_A, b_E be the systematic or bias components of error for range, azimuth, and elevation, respectively. Also, let e_R, e_A, e_E be the random components of the measurement error. e_R, e_A, e_E are assumed to be independent, zero mean, and variance σ_R^2, σ_A^2, σ_E^2, respectively. Then the measured values of range, azimuth, and elevation are modelled as:

$$R = r(\bar{x}) + b_R + e_R$$

$$A = a(\bar{x}) + b_A + e_A$$

$$E = e(\bar{x}) + b_E + e_E$$

LEAST SQUARE ESTIMATION OF OBJECT POSITION AND RADAR BIASES

If measurements are available from only one radar there is no potential for estimating the radar measurement biases, b_R, b_A, b_E from radar measurements alone. Thus, estimates of the object cartesian position obtained from the single radars measurements will be biased if b_R, b_A, b_E are not zero.

Suppose, however, that we have measurements from several radars, say M. Let the radars be indexed by α, α=1,M. The radar measurement model is now written as

$$\left.\begin{aligned} R_\alpha &= r_\alpha(\bar{x}) + b_{R_\alpha} + e_{R_\alpha} \quad (1)\\ A_\alpha &= a_\alpha(\bar{x}) + b_{A_\alpha} + e_{A_\alpha} \quad (2)\\ E_\alpha &= e_\alpha(\bar{x}) + b_{E_\alpha} + e_{E_\alpha} \quad (3) \end{aligned}\right\} \alpha=1,M$$

where $\bar{x}$ is no longer the position vector of the object in a local radar cartesian system but has cartesian components in a common reference coordinate system. The functions $r_\alpha(\bar{x})$, $a_\alpha(\bar{x})$, $e_\alpha(\bar{x})$ now involve the position of the radar site with respect to the origin of the reference coordinate system,

$$r_\alpha(\bar{x}) = \sqrt{x_{T/\alpha}^2 + y_{T/\alpha}^2 + z_{T/\alpha}^2} \tag{4}$$

$$a_\alpha(\bar{x}) = \tan^{-1} \frac{x_{T/\alpha}}{y_{T/\alpha}} \tag{5}$$

$$e_\alpha(\bar{x}) = \tan^{-1} \frac{z_{T/\alpha}}{\sqrt{x_{T/\alpha}^2 + y_{T/\alpha}^2}} \tag{6}$$

where $[x_{T/\alpha} \quad y_{T/\alpha} \quad z_{T/\alpha}] = [(x-x_\alpha) \ (y-y_\alpha) \ (z-z_\alpha)]M_{w/\alpha}$

$M_{w/\alpha}$ is a rotation matrix from the local radar cartesian system to the common reference system.

Let $h_\alpha(\bar{x})$ denote the vector with components $r_\alpha(\bar{x})$, $a_\alpha(\bar{x})$, $e_\alpha(\bar{x})$, b_α the vector with components b_{R_α}, b_{A_α}, b_{E_α}, e_α the vector with components e_{R_α}, e_{A_α}, e_{E_α} and m_α the vector with components, R_α, A_α, E_α. Then the radar measurement model is:

$$m_\alpha = h_\alpha(\bar{x}) + b_\alpha + e_\alpha, \ \alpha=1,M \tag{7}$$

Suppose we have radar measurements at a sequence of times t_i, i=1,N. We index the position vectors and measurements with t_i so that the radar measurement model is $m_\alpha(t_i) = h_\alpha(\bar{x}(t_i)) + b_\alpha + e_\alpha(t_i)$, i=1,N.

For processing the radar measurements at time t_i further package the measurements by introducing the vectors

$$m^T(t_i) = (m_1(t_i) \quad m_2(t_i) \text{ --- } m_M(t_i)),$$

$$h^T(\bar{x}(t_i)) = (h_1(\bar{x}(t_i)) \quad h_2(\bar{x}(t_i)) \text{ --- } h_M(\bar{x}(t_i))),$$

$$b^T = (b_1 \; b_2 \text{ --- } b_M), \text{ and } e(t_i) = (e_1(t_i) \; e_2(t_i) \text{ --- } e_M(t_i)).$$

Then the model of the m radar measurements at time t_i is

$$m(t_i) = h(\bar{x}(t_i)) + b + e(t_i) \tag{8}$$

Using the measurements from M radars over a sequence of N times t_i the potential exists, provided the relative geometry between the object trajectory and radar sites is sufficiently good, to estimate both the position vectors $\bar{x}(t_i)$, i=1,N and the measurement bias vector b. Suppose we have a trial or guess position $\bar{x}_0(t_i)$, i=1,N. Such a trial solution is easily computed. For example, let $\bar{x}_\alpha(t_i)=(x_\alpha(t_i)$ $y_\alpha(t_i)$ $z_\alpha(t_i))$ be the cartesian components of the object position which can be computed from the α^{th} radar. Then a good guess solution is obtained from

$$\bar{x}_0(t_i) = \underset{\alpha=1,M}{\text{median}} \begin{bmatrix} x_\alpha(t_i) \\ y_\alpha(t_i) \\ z_\alpha(t_i) \end{bmatrix}.$$

Now linearize the measurement equation (8) about $\bar{x}_0(t_i)$,

$$m(t_i) \approx h(\bar{x}_0(t_i)) + H(\bar{x}_0(t_i))\delta\bar{x}(t_i) + b + e\,(t_i) \tag{9}$$

where $H(\bar{x}_o(t_i))$ is the 3Mx3 matrix of partial derivatives

$$H(\bar{x}(t_i)) = \frac{\partial h(\bar{x}(t_i))}{\partial \bar{x}(t_i)} \tag{10}$$

Suppose we have estimates $\hat{\sigma}^2_{R_\alpha}$, $\hat{\sigma}^2_{A_\alpha}$, $\hat{\sigma}^2_{E_\alpha}$ of the measurement error variances for each of the radars. Let $\hat{\sum}$ be a diagonal matrix of these variance estimates. Then a weighted least squares estimate of the $\delta\bar{x}(t_i)$ and the bias vector b minimizes,

$$\sum_{i=1}^{N} \left\| r(t_i) - H(\bar{x}_o(t_i))\delta\bar{x}(t_i) - b \right\|^2_{\hat{\sum}^{-1}} \tag{11}$$

where $r(t_i)$ is the residual vector, $m(t_i) - h(\bar{x}_o(t_i))$.

Rather than forming the least squares normal equations by differentiating (11), it is more accurate and convenient to solve the weighted least squares problem using the QR algorithm.

APPLICATION OF THE QR ALGORITHM TO THE WEIGHTED LEAST SQUARES PROBLEM

In order to cast the weighted least squares problem into an ordinary least squares problem to which the QR algorithm can be applied, the following replacements are made

$$r(t_i) \triangleq \hat{\Sigma}^{-\frac{1}{2}} r_\alpha(t_i) \tag{12}$$

$$H(\bar{x}_o(t_i)) \triangleq \hat{\Sigma}^{-\frac{1}{2}} H(\bar{x}_o(t_i)) \tag{13}$$

$$e(t_i) \triangleq \hat{\Sigma}^{-\frac{1}{2}} e(t_i) \tag{14}$$

Then the weighted least squares problem posed by (11) becomes the ordinary least squares problem to minimize

$$\sum_{i=1}^{N} \left|\left| r(t_i) - H(\bar{x}_o(t_i))\delta\bar{x}(t_i) - \hat{\Sigma}^{-\frac{1}{2}} b \right|\right|^2 \tag{15}$$

Thus, at each time t_i we have the modified measurement equation

$$r(t_i) = H(\bar{x}_o(t_i))\delta\bar{x}(t_i) - \hat{\Sigma}^{-\frac{1}{2}} b + e(t_i) \tag{16}$$

Suppose that at each time t_i, an orthogonal matrix Q_i is constructed such that

$$Q_i^T H(\bar{x}_o(t_i)) = \begin{bmatrix} R_i \\ 0 \end{bmatrix} \tag{17}$$

where R_i is 3x3 upper triangular. Let

$Q_i^T r(t_i) = y(t_i)$ and $Q_i^T \hat{\Sigma}^{-\frac{1}{2}} = P(t_i)$. Then

$$y(t_i) = \begin{bmatrix} R_i \\ 0 \end{bmatrix} \delta x(t_i) + P(t_i) b + Q_i^T e(t_i) \tag{18}$$

If $y(t_i)$ and $P(t_i)$ are partitioned as,

$$y(t_i) = \begin{bmatrix} y_x(t_i) \\ 3x1 \\ y_b(t_i) \\ (3M-3)x_1 \end{bmatrix} \tag{19}$$

$$P(t_i) = \begin{bmatrix} P_x(t_i) \\ 3x3M \\ P_b(t_i) \\ (3M-3)x3M \end{bmatrix} \tag{20}$$

then

$$\begin{bmatrix} y_x(t_i) \\ y_b(t_i) \end{bmatrix} = \begin{bmatrix} R_i \\ 0 \end{bmatrix} \delta \bar{x}(t_i) + \begin{bmatrix} P_x(t_i) \\ P_b(t_i) \end{bmatrix} b + \begin{bmatrix} e_x'(t_i) \\ e_b'(t_i) \end{bmatrix} \tag{21}$$

If the bias vector b is known, then the first of equations (21) provides the least squares solution for the incremental position vector, $\delta\bar{x}(t_i)$. Thus, $\delta\bar{x}(t_i)$ is obtained by solving the upper triangular set of equations

$$R_i \delta\bar{x}(t_i) = y_x(t_i) - P_x(t_i)b, \quad i=1,N \tag{22}$$

The second of equations (21),

$$P_b(t_i)b = y_b(t_i) \qquad , \quad i=1,N$$

provides an overdetermined set of equations to be solved for the bias vector b. Let $P_b^T=[P_b(t_1)\ P_b(t_2)---P_b(t_N)]$, $y_b^T=[y_b(t_1)\ y_b(t_2)---y_b(t_N)]$, and $e_b=[e_b(t_1)\ e_b(t_2)---e_b(t_N)]$. Then

$$P_b b = y_b + e_b \tag{23}$$

Suppose that an orthogonal matrix Q_b is constructed such that

$$Q_b^T P_b = \begin{bmatrix} R_b \\ 0 \end{bmatrix} \tag{24}$$

where R_b is a 3Mx3M upper triangular matrix. Then the least squares solution for the bias vector is the solution to,

$$\begin{bmatrix} R_b \\ 0 \end{bmatrix} b = Q_b^T y_b = u = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \tag{25}$$

where the dimension of the vector U_1 is 3M and the dimension of U_2 is N(3M-3)-3M. Thus, the bias estimate is given by the solution to the upper triangular system of equations,

$$R_b b = U_1 \tag{26}$$

Having computed the linearized least squares estimate of the bias vector from (26), the incremental position solution $\delta\bar{x}(t_i)$ is computed from (22) for each t_i. In order to conserve on core storage requirements, the QR decomposition in (24) is not done directly, since P_b is often a very large matrix. Instead of a direct QR decomposition on P_b, the QR decomposition is done sequentially in time as each $P_b(t_i)$ arrives.

Since the original observation equations and thus the original least squares estimation problem is nonlinear, the above QR decomposition process for the least squares problem must be iterated with the replacement,

$$\bar{x}_o(t_i) \doteq \bar{x}_o(t_i) + \delta\bar{x}(t_i) \tag{27}$$

After the iteration process has converged, new estimates of the measurement variances, $\sigma_{R_\alpha}^2$, $\sigma_{A_\alpha}^2$, $\sigma_{E_\alpha}^2$ are computed for each radar from the measurement residuals,

$$r(t_i) = m(t_i) - h(\bar{x}_f(t_i)) - b_f \tag{28}$$

where $\bar{x}_f(t_i)$ is the estimate of the position vector at convergence, and b_f is the estimate of the bias vector at convergence. These new estimates of measurement error variance are inserted into the diagonal variance matrix, $\hat{\sum}$ and the QR iteration process again is iterated until convergence. This outer iteration loop which reestimates the measurement error variance is repeated for a fixed number of iterations.

Since the use of the full set of measurement times for the estimation of the measurement biases would consume considerable computer time, a highly thinned set of measurement times is selected for use in bias estimation. This thinned set of measurement times is selected to cover the entire object trajectory. This allows a reliable estimate of the measurement biases while not requiring a great amount of extra computer time.

SOME EXAMPLES

Although there have not been any convergence problems in iteratively solving the nonlinear least squares problem for the object position and the radar measurement biases, another problem which sometimes arises in linear least squares estimation problems occurs frequently in the radar bias estimation problem. Very often, the estimate of the bias vector, b, converges to a value for which several of the components are too large and may have the wrong sign. Sometimes the bias estimate is obviously erroneous. One obviously erroneous case which arises frequently is that the elevation bias components will be large and of the same sign. In linear least squares estimation the problem of the vector of regression coefficients being too long is often attributed to multicollinearity among the predictor variables. This problem in the linear estimation case is often successfully treated by some method of biased estimation such as ridge regression or principal components. The problem has not been properly recognized or successfully treated when it arises in trajectory estimation. Although the existence of these erroneous bias estimates has been recognized in trajectory estimation, the source of the difficulty was not properly recognized. Some workers in trajectory estimation have stated that the existence of this problem demonstrates the need to specify a prior distribution for the biases in order to "tie down" or statistically constrain the bias estimates. It does not take much experience in using these priors for trajectory estimation to realize that the problem of inflated bias estimates is as much present with the prior as without the prior.

We have attempted to treat the problem of inflated measurement bias estimates with both ridge regression [1] and with the method of principal components [2]. We have had partial success with each of these methods but neither method is satisfactory for an automatic data reduction program.

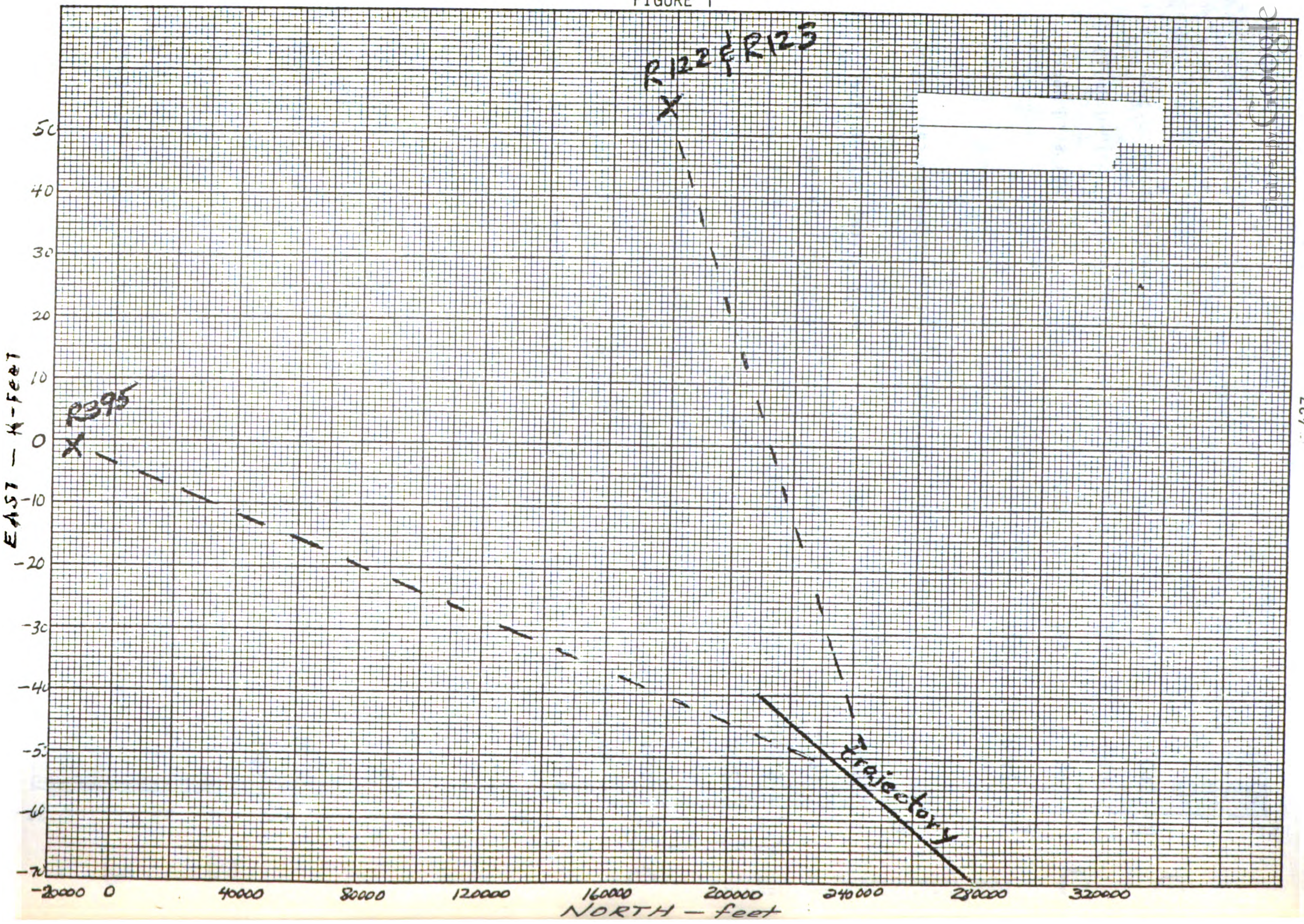
EXAMPLE 1:

Consider the following example from WSMR tracking data. This example has three radars, R122, R123, R395 tracking a level flying drone, flying at an altitude of about 30,000 ft. The graph of Figure 1 shows the relative geometry between the target trajectory and radars. The least squares estimates of the radar measurement biases for this example are,

	R122	R123	R395
Range bias (ft)	118.0	115.1	72.4
Azimuth bias (miliradians)	.186	.143	.179
Elevation bias (miliradians)	-.705	-.911	-.526

The values of the elevation bias estimates, which are all large and negative, illustrate a common type of erroneous solution occurring in radar trajectory estimation. In this example we are able to confirm that the radar bias estimates given above are greatly in error. Using measurements from tracking cameras we are able to obtain position estimates of the target trajectory which are considerably more accurate than the position estimates obtained from radar measurements. By using the optically derived positions to compute what the radar measurements

FIGURE 1



should have been, we can compute the actual values of the radar measurement biases. The following biases were computed using the optically derived positions.

	R122	R123	R395
Range bias (ft)	157.3	152.9	80.3
Azimuth bias (miliradians)	.05	.02	.09
Elevation bias (miliradians)	.11	-.08	-.09

The large errors in the radar bias estimates are readily apparent.

EXAMPLE 2:

This example has three radars, R124, R125, R442 beacon tracking a high performance missile. The graph of Figure 2 shows the relative geometry between the radars and object trajectory. The least squares estimates of the radar biases for this example are

	R124	R125	R442
Range bias (ft)	257.8	304.8	166.9
Azimuth bias (miliradians)	.056	.049	-.373
Elevation bias (miliradians)	-.170	-.368	-.479

Radar bias estimates derived from optical measurements are

	R124	R125	R442
Range bias (ft)	254	297	189
Azimuth bias (miliradians)	0	-.04	-.16
Elevation bias (miliradians)	0	-.05	-.09

FIGURE 2

Again the measurement biases estimated from least squares are seriously in error.

VARIABLE SELECTION PROCEDURE

The stepwise regression procedure in the BMDP software package was tried as a method of variable selection for the radar measurement bias estimation problem. Several examples were tried using the BMDP stepwise regression. Although the BMDP procedure proved to be unsatisfactory for many of the examples, the results suggested changes necessary to develop a satisfactory variable selection procedure for the bias estimation problem.

The computer output from the BMDP stepwise regression routine applied to the first of the previous two examples is presented on pages 433-436. Variables x(1), x(2), and x(3) correspond to the range biases for R122, R123, R395 respectively. Variables labeled x(4), x(5), and x(6) correspond to the azimuth biases for R122, R123, and R395. Variables labeled x(7), x(8), and x(9) correspond to the elevation biases for R122, R123, and R395. The stepwise regression for this example uses an $F_{IN} = 6$. Thus, at each step the variable with the largest F-to-Enter (provided it is greater than F_{IN}) is entered into the radar bias model. Thus, in step #1 the range bias for R395 is entered into the model. In step #2 the azimuth bias for R395 is entered. In step #3 R122 elevation bias is entered into the bias model and in step #4 R122 range bias is entered. In step #5 a difficulty occurs in the BMDP program. At this point variable x(2) (R122 range bias) with an F-to-Enter of 2102 should have been entered into the bias model. The BMDP output indicates on page 435 that it was unable to enter x(2) into the model because to do so would lower the tolerance

of variable x(1) below its limit of .01. This is equivalent to saying that this bias estimation problem is so ill-conditioned that, for the numerical method being used by BMDP, the minimum conditions for satisfactory operation cannot be met. Finally, the variables x(9) corresponding to R395 elevation bias and x(4) corresponding to R122 azimuth bias are entered into the radar bias model. The final radar bias estimates in the model are given on page 436. Comparing these bias estimates with the radar biases derived from optical measurements indicates that the bias model obtained from BMDP stepwise regression is erroneous. The reason for this is the failure of BMDP to enter the variable x(2) corresponding to R123 range bias. The value of the R123 range bias obtained from the optical measurements is 152.9 ft which indicates that it must be a significant influence in the radar bias model. Thus, this example suggests that a better numerical method needs to be used for the current application.

The computer output from the BMDP stepwise regression routine applied to the second of the previous two examples is presented on pages 437-441. As in the previous examples variables x(1), x(2), x(3) correspond to radar range biases for R124, R125, and R442. Variables x(4), x(5), x(6) correspond to azimuth measurement biases for R124, R125, and R442. Variables x(7), x(8), x(9) correspond to elevation measurement biases for R124, R125, and R442. In this example of stepwise example an $F_{IN} = 1$ is being used. This is done so that all variables may eventually be entered into the model. It is informative in this example to observe the sequence of multiple R^2 values as variables are entered into the measurement bias model. After having entered the three range bias

variables, x(1), x(2), x(3) into the model the multiple R^2 value is .9944 indicating that 99.44% of the sum of squares has been explained by these three variables. In step #4 x(6) corresponding to R442 azimuth bias is entered and results in a multiple R^2 of .9959, an increase of only .0015 over the previous value. As additional variables are entered into the model the sequence of increases of multiple R^2 is .0003, .0004, .0007, .0001, .0000. Also, as these additional variables are entered into the model, the magnitudes of these additional biases are large and erroneous, and the magnitude of x(6) which was entered in a previous step also becomes inflated and erroneous. Thus, variables x(4), x(5), x(7), x(8), and x(9) are erroneous and are ineffective in explaining the sum of squares. This example suggests that, in order to combat the problem of erroneous bias estimates, an effective model building procedure might be constructed by placing a lower limit on the amount of change in the multiple R^2 that is acceptable as a variable is entered into the model.

BMDP OUTPUT FOR EXAMPLE 1

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
X(3)	3	-30.84901	1.2951	-.755	1.00000	567.41	1

	VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
	X(1)	1	.18716	.86622	15.39	1
	X(2)	2	-.07397	.95372	2.33	1
	X(4)	4	.07365	.99503	2.30	1
	X(5)	5	-.24999	.17615	28.25	1
	X(6)	6	.51531	.99990	259.69	1
	X(7)	7	.37863	.99372	70.37	1
	X(8)	8	-.39301	.99961	77.65	1
	X(9)	9	-.22111	.99999	21.80	1

STEP NO. 2

VARIABLE ENTERED 6 X(6)

MULTIPLE R .8570
MULTIPLE R-SQUARE .7344
STD. ERROR OF EST. 3.5507

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	14855.237	2	7432.619	589.24
RESIDUAL	5375.7034	424	12.67854	

VARIABLES IN EQUATION

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
X(3)	3	-30.53693	1.0211	-.752	.99990	903.13	1
X(6)	6	.45145-003	.2854-004	.603	.99990	259.69	1

VARIABLES NOT IN EQUATION

	VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
	X(1)	1	.19351	.86203	17.35	1
	X(2)	2	-.11725	.93789	5.93	1
	X(4)	4	.10111	.99493	4.37	1
	X(5)	5	-.08355	.15765	2.75	1
	X(7)	7	.51675	.98364	155.75	1
	X(8)	8	-.49373	.99977	135.37	1
	X(9)	9	-.27981	.99999	35.95	1

STEP NO. 3

VARIABLE ENTERED 7 X(7)

MULTIPLE R .8977
MULTIPLE R-SQUARE .8059
STD. ERROR OF EST. 3.0477

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	16311.837	3	5437.395	585.38
RESIDUAL	3929.0530	423	9.286536	

VARIABLES IN EQUATION

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL

VARIABLES NOT IN EQUATION

	VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL

X(3)	3	-31.75835	.8781	-.778	.99070	1305.52	1	• X(1)	1	.23297	.84208	24.22	1
X(5)	5	.47016-003	.2454-004	.415	.99759	375.57	1	• X(2)	2	-.13736	.93739	5.11	1
X(7)	7	.20755-003	.1557-004	.259	.98564	155.75	1	• X(4)	4	.11653	.99693	5.99	1
								• X(5)	5	-.08831	.15765	3.29	1
								• X(8)	8	.00415	.09315	.01	1
								• X(9)	9	-.17577	.93255	13.65	1

STEP NO. 4

VARIABLE ENTERED 1 X(1)

MULTIPLE R .9036
MULTIPLE R-SQUARE .8154
STD. ERROR OF EST. 2.9574

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	15525.156	4	4131.264	459.19
RESIDUAL	3715.8011	422	8.8035215	

VARIABLES IN EQUATION

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
X(1)	1	3.94729	.8021	.112	.84208	24.22	1
X(3)	3	-29.93982	.9298	-.734	.83763	1035.95	1
X(5)	5	.47027-003	.2397-004	.411	.99508	395.65	1
X(7)	7	.20766-003	.1619-004	.259	.98343	164.65	1

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(2)	2	.91277	.01059	2102.15	1
X(4)	4	.12152	.99493	5.52	1
X(5)	5	-.09241	.15752	3.53	1
X(8)	8	.00567	.09315	.01	1
X(9)	9	-.17835	.93247	13.33	1

STEP NO. 5

VARIABLE ENTERED 9 X(9)

MULTIPLE R .9058
MULTIPLE R-SQUARE .8203
STD. ERROR OF EST. 2.9232

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	15543.341	5	3328.568	389.53
RESIDUAL	3597.5952	421	8.545359	

VARIABLES IN EQUATION

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
X(1)	1	3.91795	.7902	.111	.84200	24.59	1
X(3)	3	-29.85087	.9162	-.732	.83718	1062.30	1
X(5)	5	.46907-003	.2357-004	.410	.99489	396.11	1
X(7)	7	.19174-003	.1552-004	.238	.93173	136.68	1
X(9)	9	-.19011-004	.1234-004	-.079	.93247	2.37	1

• VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(2)	2	.91095	.01055	2043.37	1
X(4)	4	.12376	.99493	5.53	1
X(5)	5	-.08885	.15753	3.56	1
X(8)	8	-.04705	.77811	1.00	1

NOTE THAT VARIABLE 2 X(2) WAS NOT ENTERED BECAUSE ITS ENTRY
WOULD LOWER THE TOLERANCE OF VARIABLE 1 X(1) BELOW THE TOLERANCE LIMIT.

STEP NO. 6

VARIABLE ENTERED 4 X(4)

MULTIPLE R .9085
MULTIPLE R-SQUARE .9250
STD. ERROR OF EST. 2.9042

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	15695.459	6	2733.077	329.96
RESIDUAL	3542.4779	423	8.4344471	

VARIABLES IN EQUATION

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
X(1)	1	5.91773	.7851	.111	.84203	24.90	1
X(5)	5	-29.71038	.9121	-.728	.83369	1060.99	1
X(4)	4	.52197-004	.2062-004	.052	.99493	6.53	1
X(6)	6	.45967-003	.2362-004	.411	.99479	602.30	1
X(7)	7	.19171-003	.1641-004	.248	.92178	136.44	1
X(9)	9	-.10925-003	.2917-004	-.079	.93247	14.05	1

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(2)	2	.91520	.01355	2251.33	1
X(5)	5	.68101	.03353	125.13	1
X(8)	8	-.89463	.03360	1675.99	1

NOTE THAT VARIABLE 2 X(2) WAS NOT ENTERED BECAUSE ITS ENTRY
WOULD LOWER THE TOLERANCE OF VARIABLE 1 X(1) BELOW THE TOLERANCE LIMIT.

* * * * * F-LEVELS(6.000, 5.900) OR TOLERANCE INSUFFICIENT FOR FURTHER STEPPING

STEPWISE REGRESSION COEFFICIENTS

VARIABLES STEP	1 X(1)	2 X(2)	3 X(3)	4 X(4)	5 X(5)	6 X(6)	7 X(7)	8 X(8)	9 X(9)
0	1*.5028*	5.2310	-30.8499	.0001	.0002	.0005	.0001	-.0002	-.0002
1	4.7037	-1.6532	-30.8499*	.0000	-.0002	.0005	.0002	-.0002	-.0002
2	3.9341	-2.3267	-30.6959*	.0001	.0000	.0005*	.0002	-.0002	-.0002
3	3.9473	-2.4301	-31.7356*	.0001	.0000	.0005*	.0002*	.0000	-.0001
4	3.9473*	161.7443	-29.9379*	.0001	.0000	.0005*	.0002*	.0000	-.0001
5	3.9150*	160.6938	-29.8603*	.0001	.0000	.0005*	.0002*	-.0002	-.0001*
6	3.9177*	160.7215	-29.7105*	.0001*	.0000	.0005*	.0002*	-.0002	-.0001*

NOTE - 1) REGRESSION COEFFICIENTS FOR VARIABLE IN THE EQUATION ARE INDICATED BY AN ASTERISK
2) THE REMAINING COEFFICIENTS ARE THOSE WHICH WOULD BE OBTAINED IF THAT VARIABLE WERE TO ENTER IN THE NEXT STEP

PAGE 1 BMDP 10/04/84 12:54:16

REGRESSION TITLE. . . . . .
10/04/84 12:54:16

STEPPING ALGORITHM.	F
MAXIMUM NUMBER OF STEPS	20
DEPENDENT VARIABLE.	10 X(10)
MINIMUM ACCEPTABLE F TO ENTER	1.000, 20.000
MAXIMUM ACCEPTABLE F TO REMOVE.	.000, 19.000
MINIMUM ACCEPTABLE TOLERANCE.	.01000
FORCED LEVELS	1
SUBSCRIPTS OF THE INDEPENDENT VARIABLES	1 2 3 4 5 6 7 8 9

STEP NO. 0

STD. ERROR OF EST. 8.2731

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE
RESIDUAL	18206.223	266	68.44445

VARIABLES IN EQUATION

VARIABLE	COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(1)	1	.57042	1.00000	127.81	1
X(2)	2	.22728	1.00000	14.44	1
X(3)	3	.61656	1.00000	162.52	1
X(4)	4	-.20127	1.00000	11.19	2
X(5)	5	.19614	1.00000	10.60	2
X(6)	6	-.32361	1.00000	31.00	2
X(7)	7	.16865	1.00000	7.76	2
X(8)	8	.21519	1.00000	12.87	2
X(9)	9	-.39533	1.00000	49.09	2

STEP NO. 1

VARIABLE ENTERED 3 X(3)

MULTIPLE R	.6166
MULTIPLE R-SQUARE	.3801
STD. ERROR OF EST.	6.5258

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	6921.0413	1	6921.041	162.52
RESIDUAL	11285.182	265	42.58550	

VARIABLES IN EQUATION

VARIABLES NOT IN EQUATION

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
X(7)	3	425.96373	73.4057	.697	1.00000	162.52	1

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(1)	1	.54855	.93934	113.63	1
X(2)	2	.20790	.98906	11.93	1
X(4)	4	-.34349	.98800	35.32	2
X(5)	5	.16580	.98842	7.46	2
X(6)	6	-.34692	.99304	36.12	2
X(7)	7	.26386	.99606	19.76	2
X(8)	8	.31592	.99710	29.27	2
X(9)	9	-.44086	.99359	63.69	2

STEP NO. 2

VARIABLE ENTERED 1 X(1)

MULTIPLE R .7528
MULTIPLE R-SQUARE .5667
STD. ERROR OF EST. 5.4666

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	10316.820	2	5158.410	172.61
RESIDUAL	7880.4034	244	29.88410	

VARIABLES IN EQUATION

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
X(1)	1	109.54795	10.2767	.446	.93934	113.63	1
X(2)	3	350.04227	28.8730	.507	.93934	147.00	1

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(2)	2	.97355	.51530	4774.54	1
X(4)	4	-.42257	.98769	57.17	2
X(5)	5	-.01952	.88056	.10	2
X(6)	6	-.08731	.72593	2.02	2
X(7)	7	.21597	.97198	12.87	2
X(8)	8	.27677	.97175	21.82	2
X(9)	9	-.40836	.95661	52.63	2

STEP NO. 3

VARIABLE ENTERED 2 X(2)

MULTIPLE R .9886
MULTIPLE R-SQUARE .9774
STD. ERROR OF EST. 1.2514

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	17794.333	3	5931.444	3787.34
RESIDUAL	411.89049	243	1.566124	

VARIABLES IN EQUATION

VARIABLES NOT IN EQUATION

	STD. ERROR	STD REG	F TO		PARTIAL	F TO

VARIABLE		COEFFICIENT	OF COEFF	COEFF	TOLERANCE	REMOVE	LEVEL
X(1)	1	245.41793	3.2593	1.080	.48940	6691.36	1
X(2)	2	305.71358	4.4243	.893	.54530	4774.54	1
X(3)	3	177.69654	7.0648	.257	.82223	632.64	1

	VARIABLE		CORR.	TOLERANCE	ENTER	LEVEL
•	X(4)	4	-.04509	.80973	.53	2
•	X(5)	5	-.11091	.88053	3.26	2
•	X(6)	6	-.22076	.72489	13.42	2
•	X(7)	7	.02905	.92687	.22	2
•	X(8)	8	-.04151	.88799	.45	2
•	X(9)	9	-.17559	.81748	8.34	2

STEP NO. 4

VARIABLE ENTERED 6 X(6)

MULTIPLE R .9892
MULTIPLE R-SQUARE .9785
STD. ERROR OF EST. 1.2229

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	17814.606	4	4453.402	2978.03
RESIDUAL	391.81732	262	1.495486	

VARIABLES IN EQUATION

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
X(1)	1	259.99879	3.5117	1.058	.40257	5481.62	1
X(2)	2	305.11264	4.3265	.891	.51456	4973.26	1
X(3)	3	179.32611	6.9179	.260	.81884	671.94	1
X(6)	6	-.27570-003	.7525-004	-.079	.72489	13.42	2

VARIABLES NOT IN EQUATION

	VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
•	X(4)	4	.01400	.75333	.05	2
•	X(5)	5	-.02421	.73574	.15	2
•	X(7)	7	.03632	.92610	.34	2
•	X(8)	8	-.03902	.88786	.42	2
•	X(9)	9	-.20400	.80911	11.33	2

STEP NO. 5

VARIABLE ENTERED 9 X(9)

MULTIPLE R .9896
MULTIPLE R-SQUARE .9794
STD. ERROR OF EST. 1.1995

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	17830.711	5	3566.142	2478.65
RESIDUAL	375.51216	261	1.438744	

VARIABLES IN EQUATION

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
X(1)	1	254.86068	3.7674	1.037	.37650	4576.28	1
X(2)	2	299.13796	4.5098	.874	.43795	4229.19	1
X(3)	3	192.09665	6.8352	.264	.80696	709.75	1
X(6)	6	-.30099-003	.7419-004	-.043	.71746	16.46	2
X(9)	9	-.14959-003	.4443-004	-.033	.80911	11.33	2

VARIABLES NOT IN EQUATION

	VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
•	X(4)	4	-.02298	.72969	.14	2
•	X(5)	5	-.01324	.73449	.05	2
•	X(7)	7	-.10039	.62211	2.65	2
•	X(8)	8	-.15571	.69537	6.46	2

STEP NO. 6

VARIABLE ENTERED 8 X(8)

MULTIPLE R .9800
MULTIPLE R-SQUARE .9700
STD. ERROR OF EST. 1.1871

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	17870.816	6	2977.703	2109.94
RESIDUAL	346.40713	240	1.409259	

VARIABLES IN EQUATION

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
X(1)	1	255.85052	3.7497	1.041	.33290	4656.99	1
X(2)	2	300.49670	4.5937	.878	.43200	4297.72	1
X(3)	3	179.33776	6.8517	.260	.78671	685.17	1
X(6)	6	-.30848-003	.7749-004	-.044	.71670	17.62	2
X(8)	8	-.14350-003	.5446-004	-.027	.69577	6.46	2
X(9)	9	-.20840-003	.4069-004	-.046	.63368	17.59	2

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(4)	4	.00624	.70432	.01	2
X(5)	5	-.00649	.73307	.01	2
X(7)	7	-.25756	.37717	18.40	2

STEP NO. 7

VARIABLE ENTERED 7 X(7)

MULTIPLE R .9806
MULTIPLE R-SQUARE .9612
STD. ERROR OF EST. 1.1403

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	17864.122	7	2552.017	1932.10
RESIDUAL	342.10099	239	1.320853	

VARIABLES IN EQUATION

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
X(1)	1	257.22894	3.6439	1.046	.37025	4993.47	1
X(2)	2	301.78728	4.4479	.891	.47002	4603.59	1
X(3)	3	171.06450	6.9074	.248	.73538	613.29	1
X(6)	6	-.33236-003	.7736-004	-.047	.71105	21.69	2
X(7)	7	-.15287-003	.3564-004	-.059	.37717	18.40	2
X(8)	8	-.37245-003	.7020-004	-.062	.42158	22.43	2
X(9)	9	-.44990-003	.7404-004	-.100	.26757	36.91	2

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(4)	4	.07002	.66663	1.27	2
X(5)	5	.00549	.73154	.01	2

VARIABLE ENTERED 4 X(4)

MULTIPLE R .9906
MULTIPLE R-SQUARE .9813
STD. ERROR OF EST. 1.1487

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	17865.800	8	2233.225	1692.51
RESIDUAL	340.42380	258	1.319472	

VARIABLES IN EQUATION

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
X(1)	1	259.40606	3.7986	1.051	.30516	4652.05	1
X(2)	2	304.65759	5.1229	.890	.32382	3536.67	1
X(3)	3	148.09978	7.3979	.243	.67348	517.77	1
X(4)	4	.47925-004	.4251-004	.012	.66663	1.27	2
X(6)	6	-.35613-003	.7439-004	-.050	.65473	22.93	2
X(7)	7	-.14242-003	.3661-004	-.063	.35609	19.68	2
X(8)	8	-.35627-003	.7727-004	-.067	.38654	23.64	2
X(9)	9	-.44077-003	.7463-004	-.102	.26303	38.12	2

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(5)	5	.03272	.63830	.28	2

* * * * * F-LEVELS(1.000, .000) OR TOLERANCE INSUFFICIENT FOR FURTHER STEPPING

* * * * * BACKWARD STEPPING BEGINS NOW * * * * *

STEP NO. 9

VARIABLE REMOVED 4 X(4)

MULTIPLE R .9906
MULTIPLE R-SQUARE .9812
STD. ERROR OF EST. 1.1493

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	17864.122	7	2552.017	1932.10
RESIDUAL	342.10099	259	1.320853	

VARIABLES IN EQUATION

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
X(1)	1	257.22886	3.6438	1.046	.33025	4993.47	1
X(2)	2	301.78728	4.4479	.891	.43002	4603.69	1
X(3)	3	171.06450	6.9076	.248	.73538	613.29	1
X(6)	6	-.33236-003	.7134-004	-.047	.71195	21.69	2
X(7)	7	-.15287-003	.3564-004	-.059	.37717	18.40	2
X(8)	8	-.33245-003	.7020-004	-.062	.42158	22.43	2
X(9)	9	-.44980-003	.7404-004	-.100	.26757	36.91	2

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(4)	4	.07002	.66663	1.27	2
X(5)	5	.00549	.73154	.01	2

RADAR MEASUREMENT BIAS MODEL BUILDER

Both a forward selection procedure and a backward elimination procedure were developed and tested as radar measurement bias model builders. At each step of the model building process the forward selection procedure enters the variable into the model which produces the largest increase in the multiple R^2. The selection process is stopped when the largest increase in the multiple R^2 falls below a specified threshold, δ. Starting from the model which includes all of the radar bias parameters, at each step of the model building process the backward elimination procedure deletes the variable from the model which produces the smallest decrease in the multiple R^2. The elimination process is stopped when the smallest decrease in the multiple R^2 is greater than a specified threshold, δ or when the error sum of squares has increased by more than (1+f) the error sum of squares of the full model. The two model building procedures have been tested on numerous data sets. Sometimes the two selection methods result in the same model, but more often they result in two slightly different but reasonable measurement bias models. The backward elimination model builder was selected for implementation in the WSMR radar reduction program because it allows one to obtain some idea of the effects of the parameters which are not in the final model.

Starting with the full bias model given in equation (23),

$$P_b b = y_b,$$

we construct the orthogonal Q_b such that $Q_b^T P_b = \begin{bmatrix} R_b \\ o \end{bmatrix}$, i.e., equation (24). Then the measurement bias estimate for the full model is given by the solution to equation (25),

$$\begin{bmatrix} R_b \\ o \end{bmatrix} b = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$

The sum of squares due to regression for the full model is

$$SSR(B) = \left\| U_1 \right\|^2 \tag{29}$$

where $B = \{b_j, j=1, 3M\}$ denotes the full model

and the sum of squares due to error is

$$SSE(B) = \left\| U_2 \right\|^2 \tag{30}$$

thus, the multiple R^2 for the full model is

$$R^2(B) = \frac{SSR(B)}{SSR(B)+SSE(B)} = \frac{\|U_1\|^2}{(\|U_1\|^2 + \|U_2\|^2)} \tag{31}$$

Now suppose variable b_j is deleted from the full model. Let B_j denote the model with b_j deleted, $B_j = \{b_k, k \neq j\}$. Then the regression sum of squares with b_j deleted is,

$$SSR(B_j) = SSR(B) - SSR(b_j|B_j) \tag{32}$$

where $SSR(b_j|B_j)$ is the extra sum of squares due to entering b_j into the model given that all other variables are already in the model. We want to delete from the full model the variable, b_{j_1}, such that $SSR(b_{j_1}|B_{j_1})$ is a minimum provided $SSR(b_{j_1}|B_{j_1})/(SSR(B)+SSE(B) \leq \delta$. Thus, we choose b_{j_1} such that $SSR(b_{j_1}|B_{j_1}) \leq SSR(b_j|B_j)$, j=1,3M and $SSR(b_{j_1}|B_{j_1})/SSR(B)+SSE(B) \leq \delta$. Choosing the minimum of $SSR(b_j|B_j)$ is equivalent to selecting the variable b_j for which the partial F-coefficient,

$$F_j = \frac{SSR(b_j|B_j)}{SSE(B)} \tag{33}$$

is a minimum, since SSE(B) is independent of b_j. The partial F is easily computed as,

$$F_j = \frac{\hat{b}_j^2}{C_{jj}} \tag{34}$$

where $\hat{b}_j$ is the estimated value of b_j in the full model and $\sqrt{C_{jj}}$ is the standard error of $\hat{b}_j$. Thus we delete the variable b_{j_1} from the full model, where $F_{j_1} = \min \{F_j, j=1, 3M\}$. The $j_1^{\underline{st}}$ column of the matrix is then deleted from P_b and columns $j_1 + 1$ thru 3M are moved down one column, i.e., we do the replacements $p_j \doteq p_{j+1}$, $j=j_1, 3M-1$ where p_j is the $j^{\underline{th}}$ column of P_b. The backward elimination is then repeated using the new matrix P_b which has only 3M-1 columns. The algorithm for the backward elimination model building process is summarized:

$N_s = 3M - STEP$

1. Construct an orthogonal Q_b such that $Q_b^T P_b = \begin{bmatrix} R_b \\ o \end{bmatrix}$, where R_b is 

2. Solve $R_b \hat{b} = Q_b^T y_b = U_1$ dimension $\{b\} = N_s$

3. Compute $SSR = \|U_1\|^2$, $SST = \|U\|^2$, $R^2 = \frac{SSR}{SST}$

4. Compute $C = (R_b^T R_b)^{-1}$

5. Compute partial F's, $F_j = \frac{\hat{b}_j^2}{C_{jj}}$, $j=1,N_s$

6. $F_{j_1} = \min \{F_j, j=1, N_s\}$

   If $F_{j_1}/SST > \delta$, stop and use $\hat{b}$ as the final model,

   else make the column replacements, $p_b(i) \leftarrow p_b(i+1), i=j_1, N_s-1$.

} STEP=0, 3M-1

An alternative for stopping the elimination process is to stop at step j_{k-1} when $SSE(B_{j_1 j_2 \cdots j_k}) > (1+f)SSE(B)$ where $B_{j_1 j_2 \cdots j_k} = \{b_i, i \neq j_1, j_2, \ldots, j_k\}$ and f is a small fraction, say f = .1.

The deleted variables, b_{j_1}, b_{j_2} ---$b_{j_{k-1}}$ are considered as dependent variables and the set of estimated variables $B_{j_1 j_2 \cdots j_{k-1}}$ are considered as independent variables. Having identified and estimated the independent variables, values for the dependent variables are computed as averages of residuals. Let $\hat{b}_i$, $i \neq j_1, j_2, \cdots j_{k-1}$ be the estimated values of the independent variables. Let $\hat{b}_I$ be the vector with components $\hat{b}_i$ for $i \neq j_1, j_2, \cdots j_k$ and 0 for the components corresponding to the dependent variables. Let $m'(t_i)$ be the observation vector at time t_i with the vector $\hat{b}_I$ removed,

$$m'(t_i) = m(t_i) - \hat{b}_I = h(\bar{x}(t_i)) + (b - \hat{b}_I) + e(t_i) \tag{35}$$

Linearization and the QR algorithm are again applied to computing the position vectors $\bar{x}_i$, i=1,N which minimize

$$\sum_{i=1}^{N} \left\| m'(t_i) - h(\bar{x}(t_i)) \right\|^2_{\hat{\Sigma}_f^{-1}} \tag{36}$$

where $\hat{\Sigma}_f$ is the final estimated covariance matrix computed from the measurement residuals, (28). The values of the dependent variables $\hat{b}_i$, $i=j_1, j_2, \cdots j_{k-1}$ are now computed as

$$\hat{b}_i = \frac{1}{N} \sum_{\ell=1}^{N} (m'(t_\ell) - h(\hat{\bar{x}}(t_\ell)) \quad , \; i=j_1, j_2 \cdots j_{k-1} \tag{37}$$

where $\hat{\bar{x}}(t_i)$ is the estimated position vector at time t_i obtained by minimizing (36).

EXAMPLES REVISITED

EXAMPLE 1:

When the backward elimination variable selection technique described in the last section is applied to the first of the two examples previously described, the following results are obtained.

Full Model

Estimated biases

122-R	122-A	122-E	123-R	123-A	123-E	395-R	395-A	395-E
118.0	.196	-.705	115.1	.143	-.911	72.4	.179	-.526

SSR = 18972, SSE = 547 R^2 = .9810

Partial F's

161.8	3.0	3.98	154.4	[1.60]	6.67	39.0	13.0	7.27

123-A deleted

Step 1 bias estimates

122-R	122-A	122-E	123-R	123-E	395-R	395-A	395-E
118.0	.053	-.779	115.1	-.985	59.1	.118	.559

SSR = 18971, SSE = 549 R^2 = .9810

Partial F's

162.0	57.3	[5.0]	154.5	8.0	149.5	91.0	8.9

122-E deleted

Step 2 bias estimates

122-R	122-A	123-R	123-E	395-R	395-A	395-E
137.9	.053	135.0	-.207	69.0	.134	-.140

SSR = 18966, SSE = 554 R^2 = .9807

Partial F's

122-R	122-A	123-R	123-E	395-R	395-A	395-E
2756.1	[57.1]	2632.5	1291.2	1278.4	173.3	204.9

122-A deleted

Step 3 bias estimates

122-R	123-R	123-E	395-R	395-A	395-E
137.9	134.9	-.207	68.8	.134	-.140

SSR = 18908, SSE = 610 R^2 = .9777

Partial F's

122-R	123-R	123-E	395-R	395-A	395-E
2753.5	2629.9	1291.2	1271.2	[172.1]	204.8

395-A deleted

Step 4 bias estimates

122-R	123-R	123-E	395-R	395-E
159.0	156.1	-.207	84.2	-.129

SSR = 18736, SSE = 782 R^2 = .9688

Partial F's

122-R	123-R	123-E	395-R	395-E
5869.0	5628.5	1291.8	2998.7	[175.2]

395-E deleted

122-R	123-R	123-E	395-R
162.1	159.1	-.205	86.4

- independent bias variables

SSR = 18561, SSE = 958 R^2 = .9509

Partial F's

6176.1	5927.1	1264.3	3197.9

Further deletion of variables would cause more significant decreases in R^2. The dependent variables are the 122-A, 122-E, 123-A, 395-A, and 395-E biases. These biases are computed from the corresponding residuals. The results are

122-A	122-E	123-A	395-A	395-E
.057	.001	-.002	.075	-.127

- dependent bias variables

These values of the radar bias estimates compare favorably with the values computed from optics,

122-R	122-A	122-E	123-R	123-A	123-E	395-R	395-A	395-E
157.3	.05	.11	152.9	.02	-.08	80.3	.09	-.09

EXAMPLE 2:

The following results are obtained when the backward elimination variable selection technique is applied to the second of the two examples previously described.

Full Model
Estimated Biases

124-R	124-A	124-E	125-R	125-A	125-E	442-R	442-A	442-E
257.7	.056	-.170	304.8	.049	-.368	166.9	-.373	-.479

SSR = 17867, SSE = 339 R^2 = .9813

Partial F's

5537.0	2.0	28.1	4579.0	[0.6]	32.3	661.9	25.0	53.4

125-A deleted

Step 1 bias estimates

124-R	124-A	124-E	125-R	125-E	442-R	442-A	442-E
258.6	.046	-.167	304.2	-.360	167.6	-.344	-.472

SSR = 17866, SSE = 339 R^2 = .9813

Partial F's

6159.0	[1.5]	27.5	4672.1	31.7	679.9	28.2	52.8

124-A deleted

Step 2 bias estimates

124-R	124-E	125-R	125-E	442-R	442-A	442-E
257.5	-.158	301.5	-.337	170.4	-.321	-.461

SSR = 17865, SSE = 341 R^2 = .9813

Partial F's

6602	[26.0]	6084.2	30.4	804.2	26.8	51.4

124-E deleted

Step 3 bias estimates

124-R	125-R	125-E	442-R	442-A	442-E
256.1	300.2	-.141	178.9	-.296	-.212

SSR = 17839, SSE = 367 R^2 = .9798

Partial F's

6580.2	6058.3	[8.8]	961.0	22.9	25.7

125-E deleted

Step 4 bias estimates

124-R	125-R	442-R	442-A	442-E
255.1	298.8	181.7	-.289	-.154

SSR = 17830, SSE = 376 R^2 = .9793

Partial F's

6604.3	6090.2	1018.2	21.8	[17.3]

442-E deleted

Step 5 bias estimates

124-R	125-R	442-R	442-A
260.4	305.0	178.8	-.263

SSR = 17813, SSE = 393 R^2 = .9784

Partial F's

8230.9	7456.9	1001.0	[18.3]

442-A deleted

Step 6 bias estimates

124-R	125-R	442-R
265.5	305.5	177.3

- independent variables

SSR = 17794, SSE = 411 R^2 = .9773

Partial F's

10411.9 7496.6 988.0

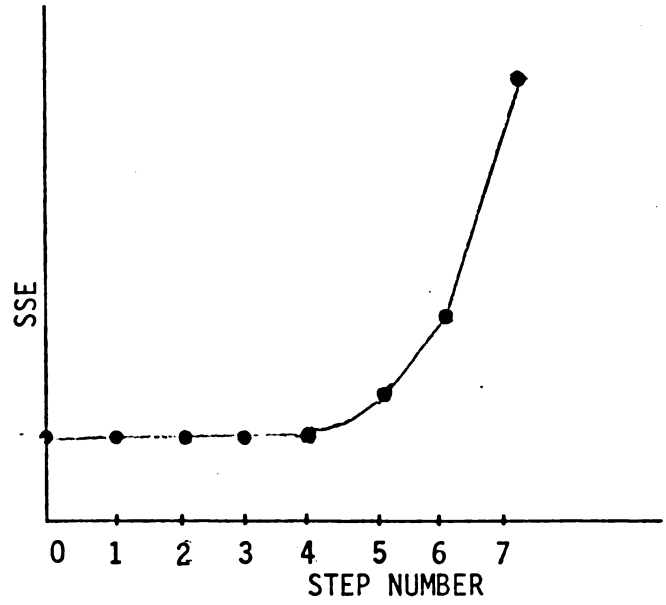
Additional deletion of variables would produce large decreases in R^2. The dependent variables are identified as the 124-A, 124-E, 125-A, 125-E, 442-A, 442-E bias. The values of these biases computed from the residuals are,

124-A	124-E	125-A	125-E	442-A	442-E
.020	.002	-.105	-.048	-.180	-.016

These values of the radar bias estimates again compare favorably with the values computed from optics,

124-R	124-A	124-E	125-R	125-A	125-E	442-R	442-A	442-E
254	0	0	297	-.04	-.05	189	-.16	-.09

The backward elimination variable selection technique described above has been successfully applied to numerous other examples of radar bias estimation. A question in the application of this method which has not been totally answered is when to stop the elimination process. From all applications tried, it appears that it is best to stop the elimination when the fractional change in the SSE exceeds a given threshold. For these applications a plot of the SSE versus the elimination step number yields a curve whose general characteristics are shown in the following graph,



One would expect that a good stopping point is close to the knee of this curve, probably just past the knee. In some applications it has been advantageous to go two points past the knee. A good answer about when to stop will probably come with additional application experience.

REFERENCES

1. Agee, W. S. and Turner, R. H., "The Use of Ridge Regression in Trajectory Estimation," Proceedings of the Twenty-Sixth Conference on the Design of Experiments in Army Research, Development and Testing, ARO Report 81-2.

2. Agee, W. S. and Turner, R. H., "Application of the Principal Component Method to Trajectory Estimation," Transactions of the First Army Conference on Applied Mathematics and Computing, ARO Report 84-1.

This paper appeard in the Proceedings of the 29th Conference on the Design of Experiments. I is reprinted here to correct several errors that occurred in typesetting the authors formulas.

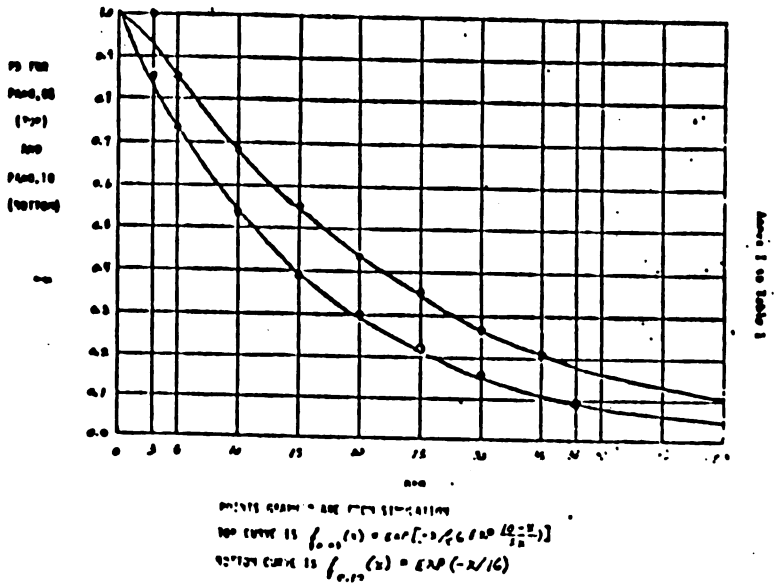
ON THE LEHMANN POWER ANALYSIS FOR THE WILCOXON RANK SUM TEST

James R. Knaub, Jr.

US Army Logistics Center

ABSTRACT

The Wilcoxon Rank Sum (or Mann-Whitney) Test is among the most useful and powerful of the non-parametric hypothesis tests. However, as with many hypothesis tests, when a clear alternative hypothesis and corresponding power analysis is not present, the practical interpretation of results using this test suffers greatly. This paper presents and clarifies an alternative suggestd by E. L. Lehmann in 1953 and provides tables of practical use which have not prviously been calculated due to computational difficulties.



On the Lehmann Power Analysis for the Wilcoxon Rank Sum Test

The Wilcoxon Rank Sum (or Mann-Whitney) Test is among the most useful and powerful of the non-parametric hypothesis tests. However, as with many hypothesis tests, when a clear alternative hypothesis and corresponding power analysis is not present, the practical interpretation of results using this test suffers greatly. This paper presents and clarifies an alternative suggested by E. L. Lehmann in 1953 (Annals of Mathematical Statistics [7]) and provides tables of practical use which have not previously been calculated due to computational difficulties. This work has recently been applied to survey data gathered for the US Army Logistics Center. (See reference [5].)

When sample sizes are small, and a power analysis is not available, one may fail to reject the null hypothesis when the true state of nature is very different from what is stated in the null hypothesis. With a small sample size and small α, it may be impossible to reject H_0. Further, when sample sizes are very large, the null hypothesis may be rejected at a very small significance level when actually the null hypothesis is so nearly true, that it is close enough for all practical purposes. Taken to the extreme, with infinite sample sizes, the attained significance level will be zero, even when there is only a very small, but finite difference between H_0 and the true state of nature. Thus significance level can be very misleading if used alone.

When a null and a definitive alternative hypothesis can both be stated, and probability distributions found under each, the results of an hypothesis test can be stated similarly to a confidence interval if the "point estimate" from the observed values falls between the two hypotheses. In the case of the Wilcoxon Rank Sum Test, only one alternative hypothesis has been well developed and will be presented here. Due to the nature of this test, however, even if the evidence may strongly indicate that the true state of nature is not bounded between this alternative and the null hypothesis, this power analysis can still be used to obtain a reasonable estimate of what the actual state of nature happens to be. (In the case of the Multiple-sample Westenberg-type tests of reference [4], an alternative must be picked such that the true state of nature is indicated to be bounded by the null and alternative hypotheses. Fortunately, that is not the case here, nor was it the case in reference [6], which is a multi-sample test.)

Consider that the null hypothesis, H_0, of the Wilcoxon Rank Sum Test indicates that $P(X<Y) = 1/2$. That is, under H_0, any value picked at random from the Y population, is larger than any value picked at random from the X population, with probability of 1/2. Here an alternative hypothesis, H_1, is used such that $P(X<Y) = 2/3$. (The exact form of H_1 is discussed in [7].) Graph 1 illustrates a possible configuration for this alternative hypothesis. For this example, consider that under H_0, all observations are taken from a N(r,s) distribution such as the N(5,1) shown on the left in graph 1, but under H_1, the Y sample comes from the N(r+0.61s, s) distribution, while the X sample comes from the N(r,s) distribution.

Another example of a possible situation satisfying the alternative hypothesis, H_1, given approximately by comparing a gamma (4,1) with a gamma (3,1), is illustrated by graph 2.

Note that the Wilcoxon Rank Sum Test is most sensitive to location, a little sensitive to shape, but not to dispersion (except as it relates proportionately to differences in location). Therefore, it is the differences in location that are of primary importance in graphs 1 and 2.

In order to determine the probability of drawing a value from distribution A which is larger than a simultaneously drawn value from distribution B, the following may be used:

$$P = \int_{x=-\infty}^{\infty} f_B(x) \int_{t=x}^{\infty} f_A(t)\,dt\,dx$$

where f_A and f_B represent density functions.

For the case where A and B are both gamma distributions,

$$P = 1 - \beta_A^{-\alpha_A}\,\beta_B^{-\alpha_B}\,\frac{1}{(\alpha_A - 1)!}\sum_{r=0}^{\alpha_B - 1}\frac{\beta_B^{r+1}}{(\alpha_B - 1 - r)!}\,\frac{(\alpha_A + \alpha_B - 2 - r)!}{([\beta_A + \beta_B]/\beta_A\beta_B)^{\alpha_A + \alpha_B - 1 - r}}$$

For gamma (4,1) and gamma (3,1), $P = 21/32 \approx 0.656$.

For normal distributions, use $\Phi[(\mu_A - \mu_B)/\sqrt{\sigma_A^2 + \sigma_B^2}]$, as in the Church-Harris-Downton (C-H-D) method of missile motor safety testing [2]. (Note: This reference to the C-H-D method should not be construed as the author's endorsement of this method for the purpose of missile motor safety testing.)

The calculation of power under this alternative involves a summation over a typically large number of products. Calculation of this value can become extremely time consuming, even for a high speed computer. A program was written for the author at White Sands Missile Range which will calculate these exact values, however, in general, the sample sizes must be very small. Recently, however, the author constructed a simulation which provides estimates of the power for much larger sample sizes. A number of the "products" mentioned earlier are calculated and the mean is computed. The number of products involved in the exact calculation can be determined, and it is multiplied by this mean. Comparison to values calculated exactly (when practical), and a study of the sensitivity of the results to increased replications, as well as comparison to other simulated values bounding the results in the tables, led to the use of from 1 to 20 million replications to simulate values for the tables found in this paper. (Work has been done, reference [3], to determine the number of simulation replications needed under less radical circumstances. Here, however, a larger number of replications appears necessary.) (For n = m

= 50, up to 35 million replications were used. It appeared, however, that fewer replications using a number of different seeds yielded mean answers which more quickly converged to reasonable results, especially when using antithetic seeds.)

In the tables, n is the sample size of the X sample, m is the sample size of the Y sample, RS is the rank sum for which type I and type II error probabilities are calculated, PA is the former of those probabilities, and PB is the later. Specifically, PA is the attained probability of making an error if H_0 is rejected, and PB is the attained probability of error if H_1 is rejected, both corresponding to the same RS value. RS is always calculated by adding the ranks of the Y elements in the combined sample. Note that for smaller sample sizes, power +PB is noticeably larger than unity due to the discrete nature of this test. That is, the probability of obtaining exactly the event observed (and no other) is non-zero.

Three significant digits are given for PA and only two for power and PB simply because it takes fewer replications of the simulation to satisfactorily obtain a value for PA than for the others.

From the annex to table 1, it is found empirically that if x is the size of each of the two samples, and $f_\alpha(x)$ is the probability of a type II error under the alternative used here, adjusted to correspond to a specific significance level, then, as a continuous representation of actually a discrete process,

$$f_{0.10}(x) \approx \exp(-x/16)$$

for at least $3 \leq x \leq 40$, and perhaps this approximation could be trusted for x = 45 or larger. However, extrapolations are always more dangerous than interpolations, so caution is advised for further extensions.

For $\alpha = 0.05$,

$$f_{0.05}(x) \approx \exp(-x/[26\exp\tfrac{10-x}{5x}])$$

for at least $4 \leq x \leq 40$, and perhaps for x substantially larger. Using this approximation, it is conjectured that for n = m = 66, when PA is approximately 0.05 (RS = 4751), then PB for this alternative is also approximately 0.05 and the true state of nature would then quite safely be said to (probably) lie between the null and alternative hypotheses. (At the 0.1 probability level for PA and PB, this could be said when n = m = 37, and RS = 1507.) An extrapolation to n = m = 66 is questionable, however, and further extrapolation is not advised. Computer simulation for n = m = 50 indicates that for the top curve (PA ≈ 0.05) in Annex I to table 1, true values in this area for PB may be somewhat smaller than this curve predicts. For PA ≈ 0.10, PB values for large n and m may be somewhat larger than predicted.

In Conover's book [1], an approximation is given to find RS for a given PA value. ($RS \gtrsim m(m + n + 1)/2 + x_{1-\alpha}\sqrt{mn(n + m + 1)/12}$, where $x_{1-\alpha}$ is from the table of the cumulative normal distribution.) The two functions given earlier can be used to estimate PB values when PA $\simeq$ 0.10 or 0.05.

The final graphs, 3-7, are taken from work the author directed at White Sands Missile Range in order to study this alternative for the Wilcoxon Rank Sum Test with emphasis on simulation validation for missile flight simulations. When comparing a very few live firings to a substantially larger number of simulations for each scenario, it can be seen from these graphs that once one sample is substantially larger than the other, increasing the larger sample size further does very little to improve the power. These graphs are continuous representations of what are actually discrete points. The values for those points were calculated analytically as noted in the acknowledgements.

Finally, when $n \neq m$, PB can be bounded using the exponential formulations found earlier in this paper. If, for example, RS is such that PA = 0.1, and x_1, is the smaller of n and m, and x_2 is the larger, then one has that approximately $\exp(-x_2/16) < PB < \exp(-x_1)$, with PB somewhat closer to $\exp(-x_1/16)$, especially when $x_1 << x_2$.

For larger sample sizes than are handled here, parametric methods may be used. However, in addition to the probability of error associated with any conclusion drawn from a parametric test, there is the additional risk involved in assuming the distributional forms used in such a test. Hypothesis tests should also be used to study these distributional assumptions to provide a more complete risk analysis.

EXAMPLE:

Consider two sources of data, X and Y, where it is suspected that Y may represent a population of larger location than X, but this is not clear. If 11 observations are taken from the X population, and 19 observations taken from Y, then the critical value of the rank sum (RS) of the Y sample observations within the combined sample which represents the point at which rejection of the null hypothesis would occur using α = 0.10, is approximately

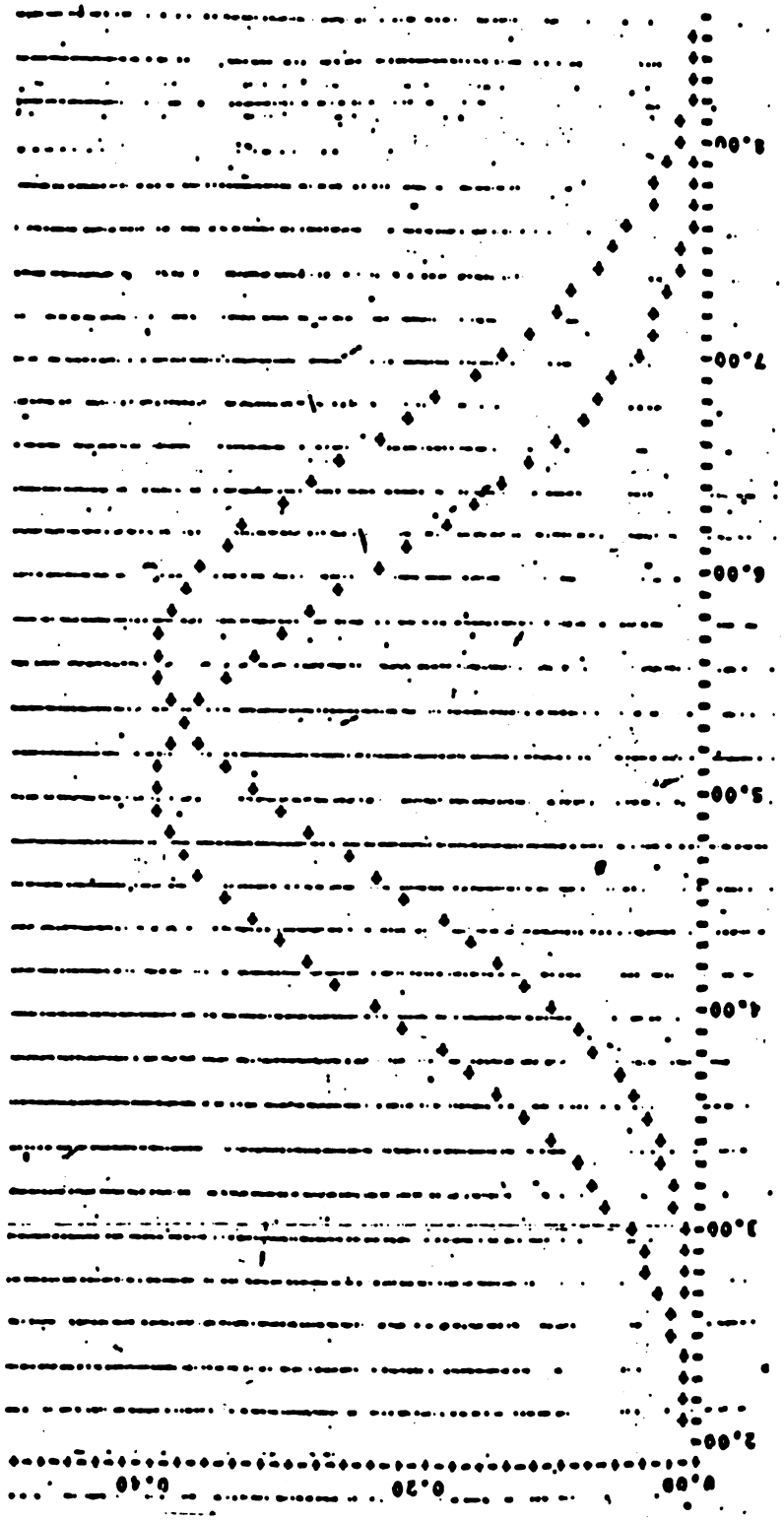
$$RS \gtrsim m(m + n + 1)/2 + 1.2816\sqrt{mn(m + n + 1)/12}$$

$$= (19)(31)/2 + 1.2816\sqrt{(19)(11)(31)/12}$$

$$\simeq 324.3$$

Therefore, if $RS \geq 325$, H_0 would be rejected at the α = 0.10 level. However, should RS = 325, and H_0 not be rejected, then the probability of making a type II error with respect to the alternative hypothesis illustrated in graphs 1 and

2 is approximately bounded by exp (-19/16) and exp (-11/16), so 0.30<PB<0.50. Note that, from table 2 , when PA = 0.099, PB (10,20) ≈ 0.43. Using 4,000,000 replications in the program given in Appendix A, for m = 19, n = 11, and RS = 325, resulted in PA = 0.100 and PB = 0.42.



Alternative Hypothesis (H_1) Using a
N(5,1) and an N(5.61,1)

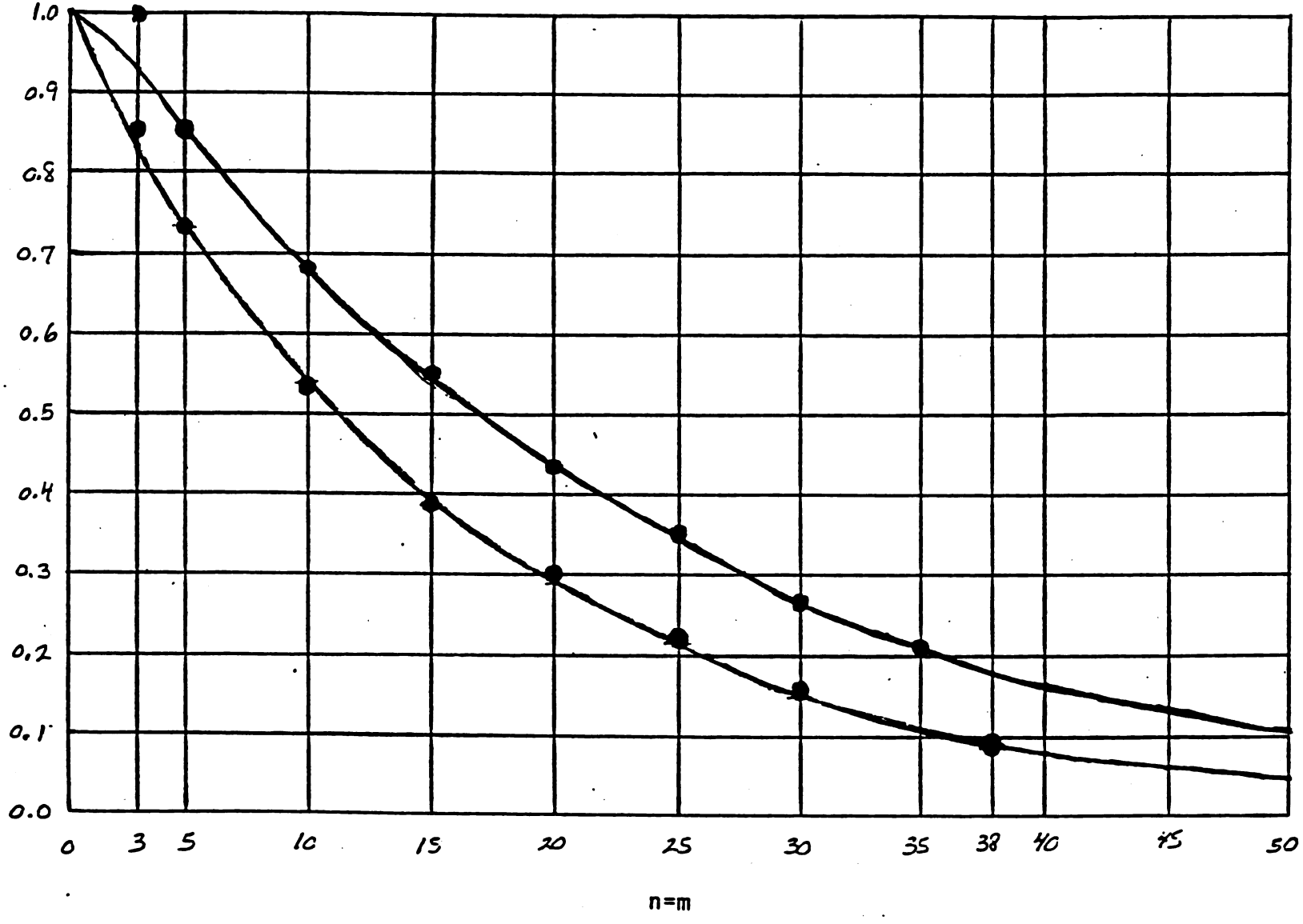
Graph 1

Graph 2

Table 1

n = m	RS	PA	power	PB
3	12	0.350	0.62	0.56
3	14	0.100	0.27	0.85
3	15	0.050	0.15	1.00
5	32	0.210	0.54	0.55
5	34	0.111	0.37	0.71
5	35	0.075	0.29	0.79
5	36	0.048	0.21	0.86
5	39	0.008	0.05	0.97
10	122	0.108	0.52	0.51
10	123	0.095	0.49	0.54
10	127	0.052	0.36	0.67
10	128	0.045	0.34	0.69
10	136	0.009	0.13	0.89
15	264	0.101	0.63	0.39
15	265	0.094	0.61	0.41
15	273	0.049	0.47	0.55
15	289	0.009	0.21	0.80
20	458	0.101	0.71	0.30
20	459	0.096	0.70	0.30
20	471	0.051	0.58	0.43
20	472	0.048	0.57	0.44
20	496	0.010	0.30	0.71
25	704	0.101	0.79	0.22
25	705	0.098	0.78	0.23
25	723	0.050	0.66	0.35
25	758	0.009	0.38	0.63
30	1002	0.101	0.85	0.16
30	1003	0.099	0.84	0.16
30	1027	0.050	0.74	0.27
30	1073	0.010	0.47	0.54
35	1383	0.050	0.79	0.21
38	1587	0.100	0.91	0.09

PB FOR PA≈0.05 (TOP) AND PA≈0.10 (BOTTOM)



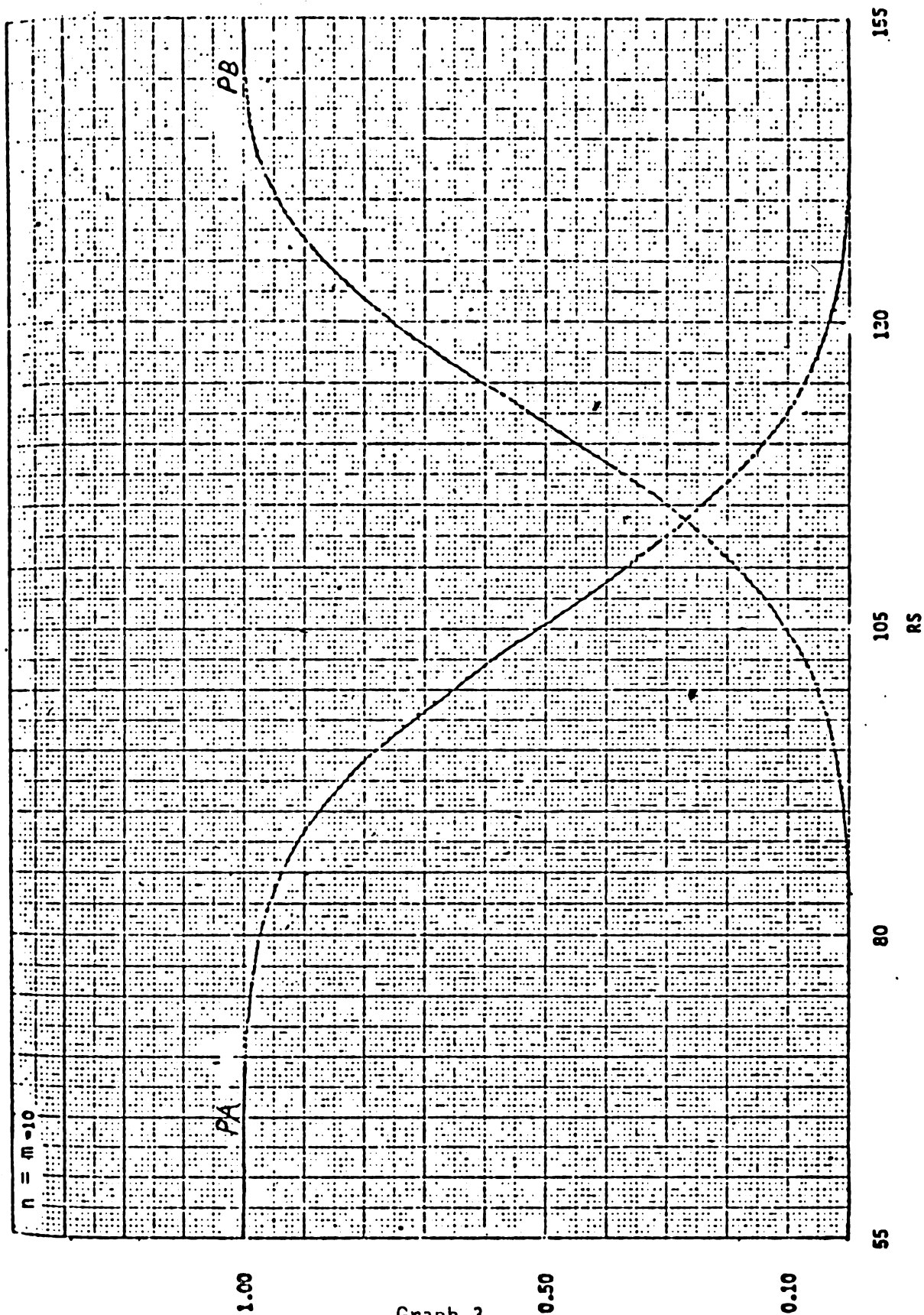
POINTS GRAPHED ARE FROM SIMULATION

TOP CURVE IS $f_{0.05}(x) = \exp[-x/(26\exp\frac{10-x}{5x})]$

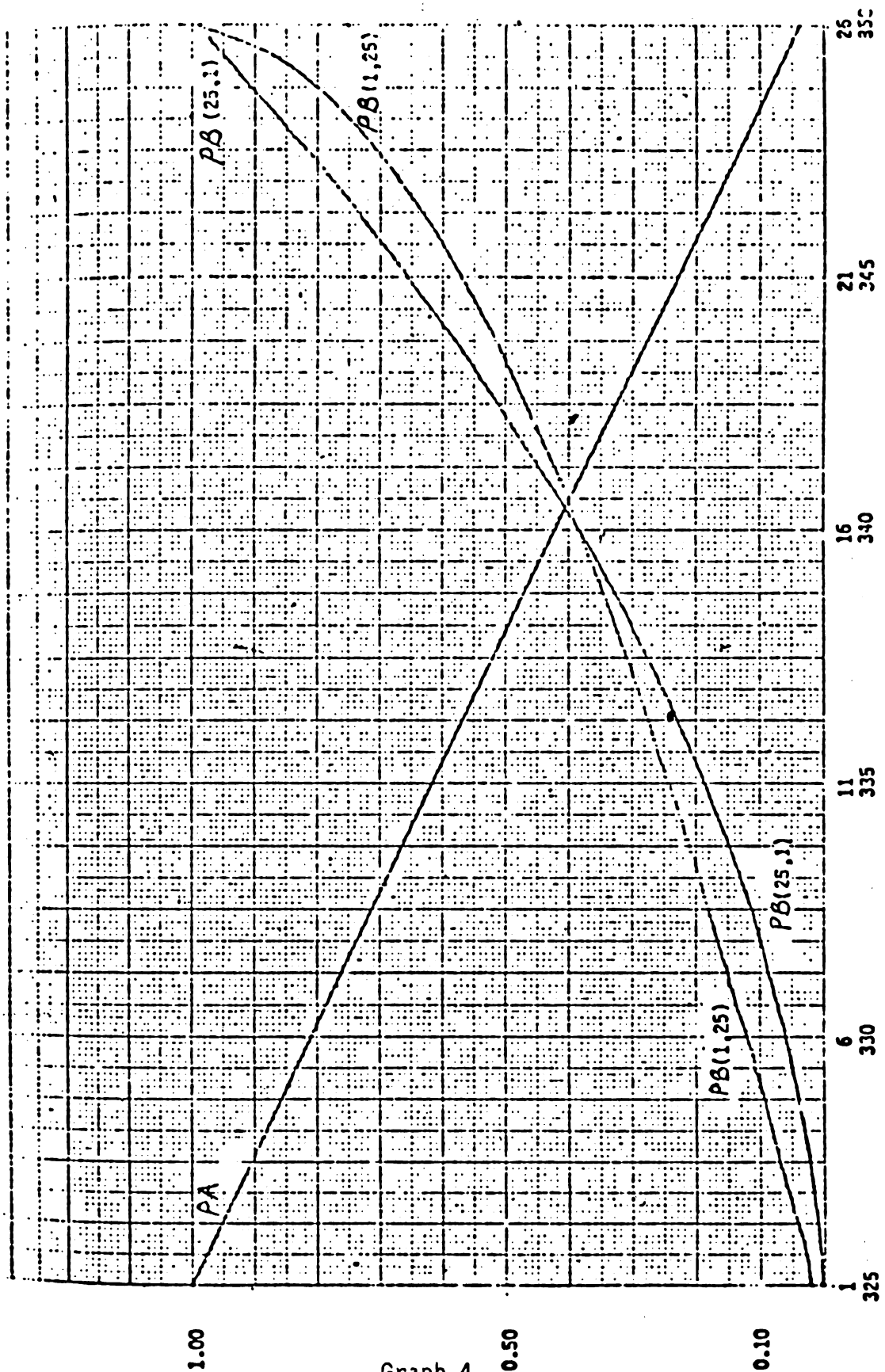
BOTTOM CURVE IS $f_{0.10}(x) = \exp(-x/16)$

Table 2

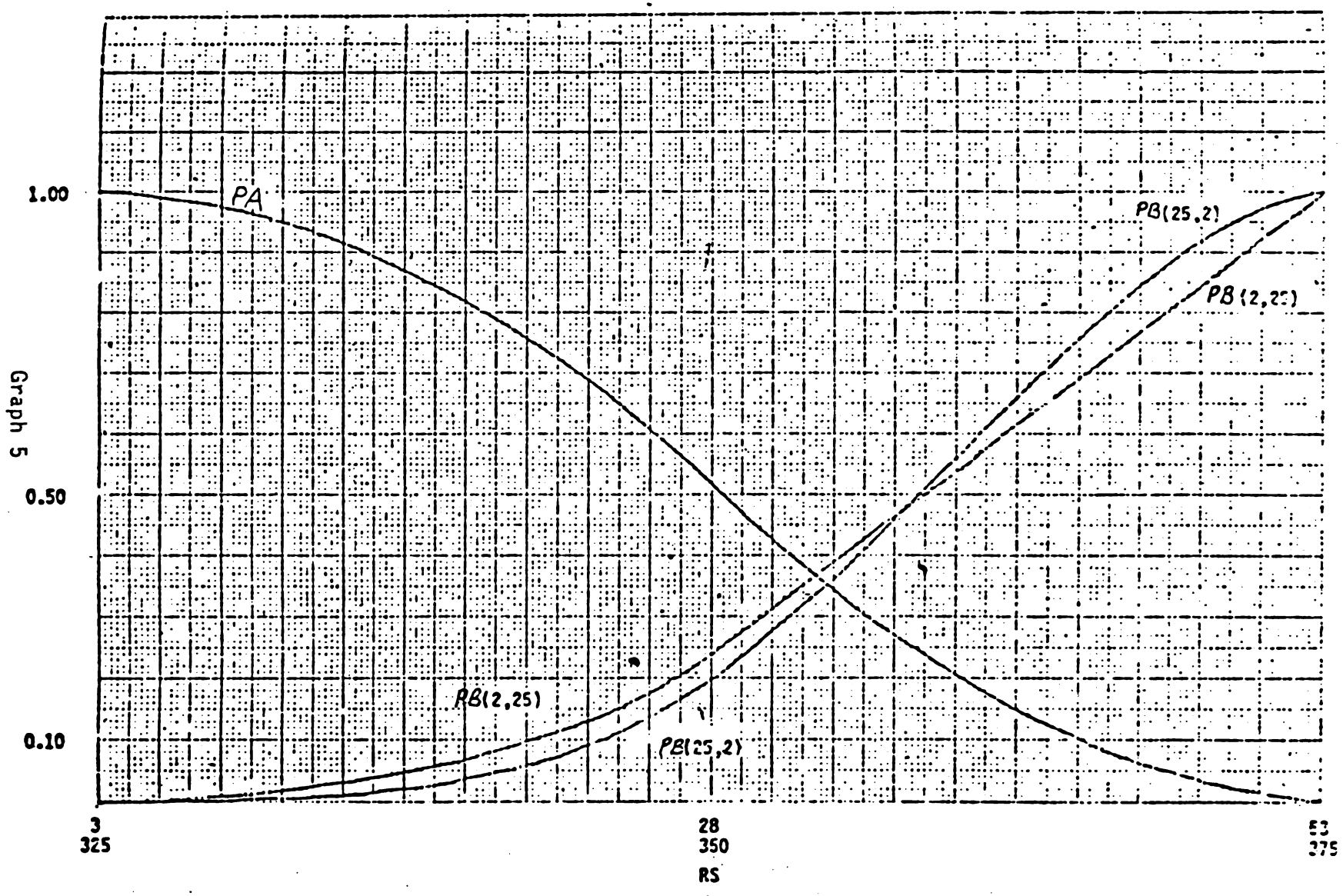
n, m n ≠ m	RS	PA	power	PB
5,10	85	0.297	0.70	0.35
5,10	91	0.103	0.42	0.63
5,10	92	0.082	0.37	0.68
5,10	94	0.050	0.27	0.77
5,10	99	0.010	0.10	0.93
10,5	45	0.297	0.71	0.34
10,5	51	0.103	0.41	0.65
10,5	52	0.082	0.35	0.70
10,5	54	0.050	0.26	0.79
10,5	59	0.010	0.08	0.95
5,25	412	0.094	0.45	0.57
5,25	418	0.048	0.33	0.69
5,25	429	0.009	0.13	0.88
5,25	430	0.008	0.12	0.89
25,5	102	0.094	0.44	0.59
25,5	108	0.048	0.29	0.73
25,5	119	0.009	0.09	0.92
25,5	120	0.008	0.08	0.93
10,20	340	0.099	0.58	0.43
10,20	348	0.050	0.44	0.58
10,20	363	0.009	0.20	0.81
20,10	185	0.099	0.59	0.43
20,10	193	0.050	0.44	0.58
20,10	208	0.010	0.18	0.84
5,50	1444	0.105	0.50	0.51
5,50	1457	0.050	0.36	0.65
5,50	1480	0.008	0.14	0.87
50,5	184	0.105	0.50	0.52
50,5	197	0.050	0.32	0.69
50,5	220	0.008	0.09	0.92
10,50	1590	0.101	0.65	0.36
10,50	1608	0.051	0.52	0.49
10,50	1643	0.009	0.26	0.75
50,10	370	0.102	0.68	0.33
50,10	388	0.051	0.52	0.49
50,10	423	0.009	0.22	0.79



Graph 3

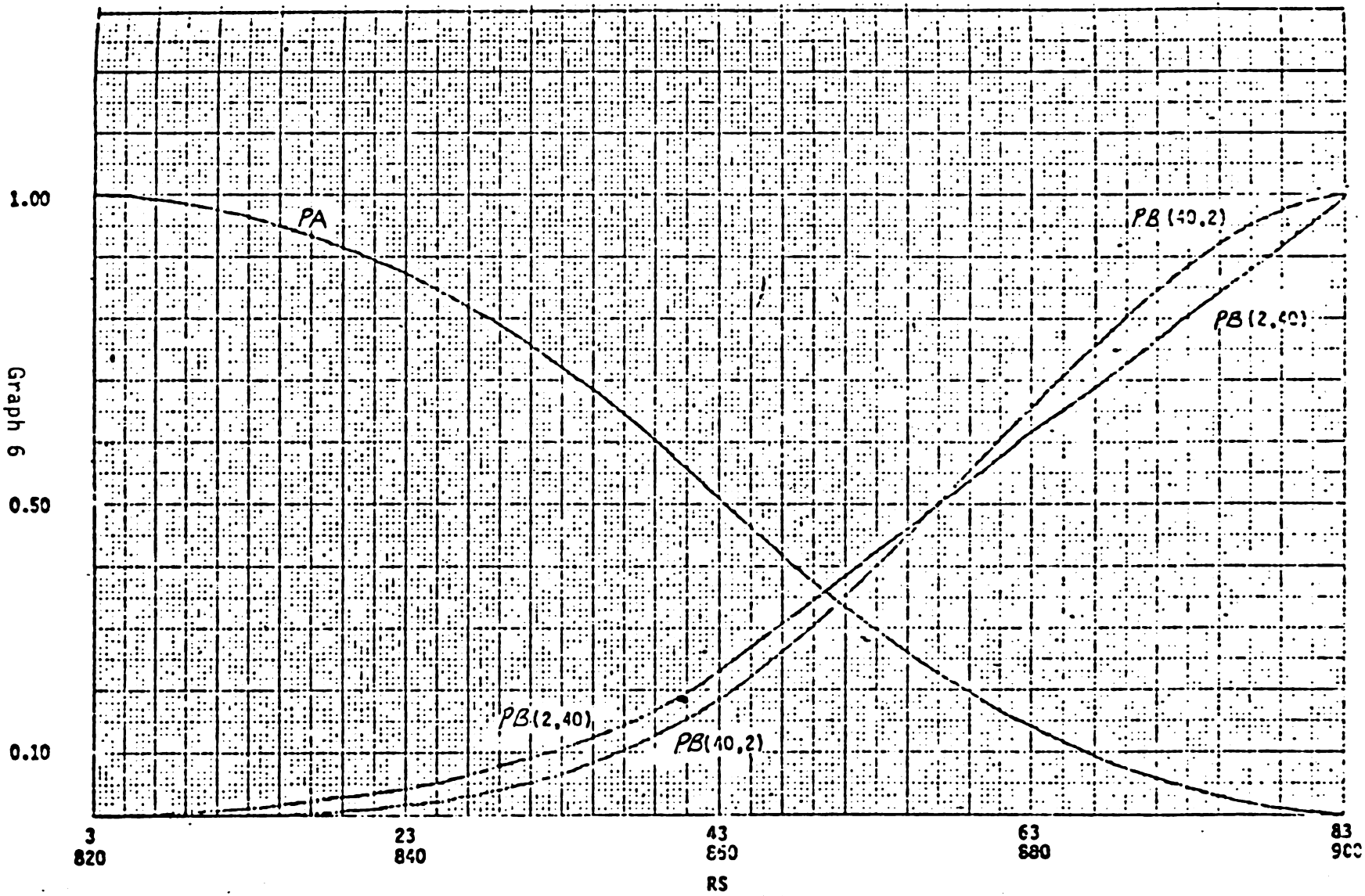


Graph 4



Graph 5

Graph 6



Graph 7

APPENDIX A

FORTRAN CODE FOR

SIMULATION:

"LEHMANN POWER ANALYSIS

FOR THE

WILCOXON RANK SUM TEST"

(LPAWRST)

```
VAX/VMS  KNAUB        LPA0:RST  14-JUL-1983 15:35   LPA0:  14-JUL-1983 15:35   USER:[KNAUB]LPA0.PST
VAX/VMS  KNAUB        LPA0:RST  14-JUL-1983 15:35   LPA0:  14-JUL-1983 15:35   USER:[KNAUB]LPA0.PST
VAX/VMS  KNAUB        LPA0:RST  14-JUL-1983 15:35   LPA0:  14-JUL-1983 15:35   USER:[KNAUB]LPA0.PST

                      K    K  N    N   AAA   U    U  BBBB
                      K   K   N    N  A   A  U    U  B   B
                      K  K    NN   N  A   A  U    U  B   B
                      KKK     N N  N  A   A  U    U  BBBB
                      K  K    N  N N  AAAAA  U    U  B   B
                      K   K   N   NN  A   A  U    U  B   B
                      K    K  N    N  A   A  UUUUU  BBBB

     LL          PPPPPPPP      AAAAAA    WW      WW  RRRRRRRR      SSSSSSSS  TTTTTTTTTT
     LL          PPPPPPPPP     AAAAAA    WW      WW  RRRRRRRRR     SSSSSSSS  TTTTTTTTTT
     LL          PP      PP    AA  AA    WW      WW  RR      RR    SS            TT
     LL          PP      PP   AA    AA   WW      WW  RR      RR    SS            TT
     LL          PP      PP   AA    AA   WW      WW  RR      RR    SS            TT
     LL          PP      PP   AA    AA   WW      WW  RR      RR    SS            TT
     LL          PPPPPPPPP    AA    AA   WW      WW  RRRRRRRRR      SSSSSS        TT
     LL          PPPPPPPPP    AA    AA   WW      WW  RRRRRRRRR      SSSSSS        TT
     LL          PP           AAAAAAAAAA   WW  WW  WW  RR   RR             SS    TT
     LL          PP           AAAAAAAAAA   WW  WW  WW  RR   RR             SS    TT
     LL          PP           AA    AA     WWWW  WWWW  RR    RR            SS    TT
     LL          PP           AA    AA     WWWW  WWWW  RR    RR            SS    TT
     LLLLLLLLLLL PP           AA    AA      WW    WW   RR    RR    SSSSSSSS    TT
     LLLLLLLLLLL PP           AA    AA      WW    WW   RR    RR    SSSSSSSS    TT

      FFFFFFFFF    000000     RRRRRRRR        1111      222222
     FFFFFFFFFF   00000000    RRRRRRRRR       1111      222222
     FF          00      00   RR      RR      1111     22    22
     FF          00      00   RR      RR      1111     22    22
     FF          00      00   RR      RR                     22
     FF          00      00   RR      RR                     22
     FFFFFFFF    00      00   RRRRRRRRR       1111          22
     FFFFFFFF    00      00   RRRRRRRRR       1111          22
     FF          00      00   RR    RR        1111        22
     FF          00      00   RR    RR        1111        22
     FF          00      00   RR    RR         11       22
     FF          00      00   RR    RR         11       22
     FF           000000      RR      RR      11      2222222222
     FF           000000      RR      RR      11      2222222222

                      K    K  N    N   AAA   U    U  BBBB
                      K   K   N    N  A   A  U    U  B   B
                      K  K    NN   N  A   A  U    U  B   B
                      KKK     N N  N  A   A  U    U  BBBB
                      K  K    N  N N  AAAAA  U    U  B   B
                      K   K   N   NN  A   A  U    U  B   B
                      K    K  N    N  A   A  UUUUU  BBBB

VAX/VMS  KNAUB        LPA0:RST  14-JUL-1983 15:35   LPA0:  14-JUL-1983 15:35   USER:[KNAUB]LPA0.PST
VAX/VMS  KNAUB        LPA0:RST  14-JUL-1983 15:35   LPA0:  14-JUL-1983 15:35   USER:[KNAUB]LPA0.PST
VAX/VMS  KNAUB        LPA0:RST  14-JUL-1983 15:35   LPA0:  14-JUL-1983 15:35   USER:[KNAUB]LPA0.PST
```

```
      INTEGER PREV(317)
      DIMENSION I(1000),NEXT(1000),II(3),MIN(317),MAX(317)
      LOGICAL*1 FLAG(317)
      REAL*4 I,II,MINVAL,MAXVAL
      REAL*8 BOUND(316),XX,PS,TPS,PTPS,PRC,PRP,C2,PRC1,PRC2,
     /   PRC3,PRC4
      DATA ISEED/78125/,IDIR/1/
      M=0.
      WRITE(19,111)
111   FORMAT(1X,'LEHMANN POWER ANALYSIS FOR THE WILCOXON
     /RANK SUM TEST, LPAWRST')
      WRITE(19,1)
      WRITE(6,1)
1     FORMAT(1X,'ENTER NO. OF OBSERVATIONS, NO. Y')
      READ(5,*)NOBS,NY
      WRITE(19,*)NOBS,NY
      WRITE(19,101)
 7     PRINT 101
101   FORMAT(1X,'INPUT TEST STATISTIC')
      READ(5,*)IX
      WRITE(19,*)IX
      WRITE(19,102)
      PRINT 102
102   FORMAT(1X,'INPUT NO. OF REPLICATIONS')
      READ(5,*)IREPS
      WRITE(19,*)IREPS
100   I(1)=RAN(ISEED)
      IYRNK=0
      DO 105 J=1,317
      FLAG(J)=.FALSE.
105   CONTINUE
      MINVAL=I(1)*IDIR
      MAXVAL=0
      DO 10 J=2,NOBS
      I(J)=RAN(ISEED)
      II(1)=I(J)*IDIR
      IF(II(1).GT.MAXVAL)MAXVAL=II(1)
      IF(II(1).LT.MINVAL)MINVAL=II(1)
10    CONTINUE
      J=J-1
      NUM=J
      X=NUM
      NCELLS=(X/(SQRT(X)))+.5
      RANGE=MAXVAL-MINVAL
      BOUND(1)=MINVAL
      XX=RANGE/NCELLS
      DO 60 J=2,NCELLS
      BOUND(J)=BOUND(J-1)+XX
60    CONTINUE
      BOUND(NCELLS+1)=MAXVAL
      DO 50 J=1,NUM
      XX=I(J)*IDIR
      DO 70 JJ=1,NCELLS
      IF(XX.GE.BOUND(JJ).AND.XX.LT.BOUND(JJ+1))GO TO 75
70    CONTINUE
      JJ=JJ-1
75    IF(FLAG(JJ).EQ..FALSE.)THEN
         MIN(JJ)=J
         MAX(JJ)=J
         FLAG(JJ)=.TRUE.
         GO TO 50
```

```
      END IF
      II(1)=XX
      II(2)=I(MIN(JJ))*IDIR
      II(3)=I(MAX(JJ))*IDIR
      IF(II(1).LE.II(2))THEN
         NEXT(J)=MIN(JJ)
         MIN(JJ)=J
      ELSE IF(II(1).GT.II(3))THEN
         NEXT(MAX(JJ))=J
         MAX(JJ)=J
      ELSE
         PREV(JJ)=MIN(JJ)
         K=NEXT(MIN(JJ))
20       II(1)=I(J)*IDIR
         II(2)=I(K)*IDIR
         IF(II(1).GE.II(2))GO TO 30
         PREV(JJ)=K
         K=NEXT(K)
         GO TO 20
30       NEXT(PREV(JJ))=J
         NEXT(J)=K
      END IF
50    CONTINUE
      L=0
      PS=1.
      IIY=0
      DO 80 JJ=1,NCELLS
      IF(FLAG(JJ).EQ..FALSE.)GO TO 80
      K=MIN(JJ)
40    IIY=IIY+1
      IF(K.LE.NY)THEN
         L=L+1
         IYRNK=IYRNK+IIY
         PS=PS*(IIY+L-1)/100.
         IALPH='Y'
      ELSE
         IALPH='X'
      END IF
C     WRITE(6,2)I(K),IALPH
2     FORMAT(1X,E15.7,5X,A1)
      IF(K.EQ.MAX(JJ))GO TO 80
      K=NEXT(K)
      GO TO 40
80    CONTINUE
C     WRITE(6,3)IYRNK
C3    FORMAT(1X,'SUM OF Y - RANKS = ',I6)
      IF(IYRNK.GE.IX)M=M+1
       IF(IYRNK.EQ.IX)THEN
                  PTPS=PTPS+PS
                  NP1PS=NPTPS+1
            END IF
      IF(IYRNK.LE.IX)GO TO 200
      TPS=TPS+PS
      NP1PS=NPTPS+1
      NTPS=NTPS+1
      PTPS=PTPS+PS
200   ITRACK=ITRACK+1
      XM=M
      IF(ITPACK-IREPS)100,201,201
201   XNOBS=NOBS
      XNY=NY
```

```
      C1=IKEPS
      PA=XM/C1
      XNIPS=NIPS
       WRITE(19,*)XNIPS
      XNPTPS=NPTPS
        IF(TPS.EQ.0.)THEN
            ATPS=0.
       ELSE
        ATPS=IPS/XNTPS
              END IF
      IF(PIPS.EQ.0.)THEN
           APTPS=0.
       ELSE
        APTPS=PIPS/XNPTPS
              END IF
      C2=(XMOBS+XNY)*(10.**20)
      DO 203 IL=1,NY-1
      XI=IL
      C2=C2*(XMOBS+XNY-XI)/100.
203   CONTINUE
C       WRITE(6,*)C2
      XNYF=1.
        PB=ATPS*PA
        PBP=APTPS*PA
      PBC1=(2.**30)*(10.**20)
      PBC2=(2.**30)*PBP
      PBC3=(10.**2)*(2.**(NY-60))
      PBC4=PBC3/C2
        POWER=PBC1*PBC2*PBC4
      WRITE(19,*)ITY,L,PB,C2
      IF(ATPS.EQ.0.)THEN
              PB=0.
            ELSE
        PB=POWER*(PB/PBP)*(XNIPS/XM)
          END IF
900     PB=1.0-PB
      WRITE(6,*)PA,POWER,PB
      WRITE(19,*)PA,POWER,PB
      STOP
      END
```

APPENDIX B

ACKNOWLEDGEMENTS

Thanks to Lynne Grile and Gerard Petet for developing the programming necessary to generate graphs 1 and 2. Keith Haycock (White Sands) constructed graphs 3-7, and Dr. Larry Armijo (White Sands - KENTRON) wrote the computer program to provide analytical solutions used in graphs 3-7. Also, thanks to Jeffrey Greenhill for providing a customized sorting routine for the author's simulation. Carlyle Comer and Frank Lawrence were helpful in obtaining massive quantities of needed CPU time on the USALOGC's VAX-11/780. Finally, thanks to other analysts for helpful conversations and/or influence.

APPENDIX C

REFERENCES

1. Conover, W. J., Practical Nonparametric Statistics, 2 ed, John Wiley & Sons, 1980.

2. Downs, R. S., P. C. Cox, "The Probability of Motor Case Rupture," ARO Report 75-2.

3. Juritz, J. M., J. W. F. Juritz and M. A. Stephens, "On the Accuracy of Simulated Percentage Points," Journal of the American Statistical Association, 78 (June 1983).

4. Knaub, J. R., Jr., "Design of a Multiple Sample Westenberg Type Test for Small Sample Sizes," ARO Report 82-2.

5. Knaub, J. R., Jr., Appendix D, US Army Manpower Nonavailability and Indirect (Unit Related) Productive Factors - Final Report, US Army Logistics Center, August 1983 - not released

6. Knaub, J. R., Jr., L. M. Grile and G. Petet, "Analyzing n Samples of 2 Observations Each," ARO Report 83-2.

7. Lehmann, E. L., "The Power of Rank Tests," Annals of Mathematical Statistics, 24 (1953), 23-43.

ADDENDUM

Multiple applications of this test can be used to compare two levels of a factor under a number of conditions. If, for example, manufacturer A produces a machine which is suspected to have higher reliability under most scenarios than a similar machine made by manufacturer B, then under each of the γ scenarios, m_i is the sample size of A's machines and n_i is the sample size of B's machines, for $i = 1$ to γ. PA_i and PB_i can be calculated for each of the scenarios. Consider $0 \leq a \leq \gamma$ and $0 \leq b \leq \gamma$.

PA is the probability of a or more PA_i's being less than p_A $(i = 1, \gamma)$, when H_0 is true.

PB is the probability of b or more PB_i's being less than p_B $(i = 1, \gamma)$, when H_1 is true.

Therefore,

$$PA = \sum_{x=a}^{\gamma} \binom{\gamma}{x} p_A^x (1 - p_A)^{\gamma - x}$$

and

$$PB = \sum_{x=b}^{\gamma} \binom{\gamma}{x} p_B^x (1 - p_B)^{\gamma - x}$$

p_A and p_B are chosen to be reasonable considering sample sizes for each of the γ cases.

If $\frac{PA}{PB} = 1$ then the evidence shows that, in general, the true state of nature is just as likely to be equivalent to H_1 as H_0.

If $\frac{PA}{PB} = 2$ then the evidence indicates that, in general, the true state of nature is twice as likely to be equivalent to H_0 as H_1. If PA and PB are small, then the indication is only that the true state of nature is closer to H_0 than H_1, although possibly not very close to either.

(Note that another paper in this conference, "Numerical Validation of Tukey's Criteria for Clinical Trials and Sequential Testing," by C. R. Leake, also deals with this type of problem, and was of interest to this author.)

At this time, this methodology is being used to determine whether survey data from a presumably less reliable source is compatible with a presumably superior data source. Difficult to obtain data on U.S. Army warehousing activities have, as one obvious characteristic, a very flat "peak." Therefore, a sample median value can be changed drastically by the addition or deletion of one data point. If the secondary data source proves to provide values distributed closely enough to that of the primary source, the advantage of including this source may outweigh the disadvantage. The current situation is more complex than this. However, some results employing the methodology of this addendum have been realized.

ADDENDUM 2

Two approximations for the power of this test which apparently are good for a wide range of normal alternative hypotheses are to be found in E. L. Lehmann, Nonparametrics: Statistical Methods Based on Ranks, Holden-Day, 1975. Although restricted to normal alternatives in the format in which they are written, these approximations can be used to extend the tables given here to larger n and m. The easier of the two approximations to apply, in its simplest form, is found on page 73 of the above reference and is essentially as follows:

$$\text{power} \approx \Phi\left[\sqrt{\frac{3mn}{(m+n+1)\pi}}\,\frac{\mu_A - \mu_B}{\sigma} - x_{1-\alpha}\right]$$

where in our case we have $(\mu_A - \mu_B)/\sigma \approx 0.610$.

Note that in the example in the main body of this paper ($m = 19, n = 11$), that this approximation gives power ≈ 0,60, which is consistent with what was shown earlier.

Limited randomization with detailed reassignment as the key to taking advantage of modern summaries

John W. Tukey

Princeton University* and AT&T Bell Laboratories
Princeton, New Jersey 08544 and Murray Hill, New Jersey 07974

1. Uncertainty, Validity, and Stringency.

Some experimentation needs to be done with the least uncertainty possible. Some of this is devoted toward direction of change, and can thus be analyzed in terms of a simple significance test or a simple confidence interval -- perhaps obtained by trial and error bisection among displaced significance tests.

In such cases we need to reduce ***all*** kinds of uncertainty, ***including*** uncertainty about any statistical hypotheses that matter for our analysis.

For a simple significance test to be ***valid,*** we must control the chance of significance in the null situation -- allowing it to be at most inappreciably greater than the announced P-value.

We would like the null situation to be as broad as possible, since the broader it is, the more usefully valid is our analysis. The broadest null situation that we know how to handle well , when we have treatment and control in blocks of two, is:

> Any set of observed results is as likely to occur for any interchanges, each within a block, of treatment with control -- "all reassignments are equally null-probable".

Notice that no assumption is made about relative likelihoods of ***"different"*** sets of observed results. There is no "statistical model" in the ordinary sense, and there is thus no need to worry how well -- or how poorly -- such a model approximates the real world. When we are being careful, we should ***know*** why our analysis is valid! We should not depend on unverifiable assumptions for validity.

Validity is of course not enough. We surely do not want to be wasteful. So we must also seek ***stringency*** as expressed in such technical terms as

$$\frac{\textit{minimized variance}}{\textit{actual variance}} = \text{efficiency}$$

or

$$\left\{ \frac{\textit{minimized typical length of confidence interval}}{\textit{actual typical length of confidence interval}} \right\}$$

Here "variance" or "typical length", while hopefully imitating what goes on in the real

*Prepared in part in connection with research at Princeton, supported by the Army Research Office (Durham).

Presented as Keynote Address, first annual Frontiers of Industrial Experimentation Conference, Mohonk, New York, April 23, 1984; at the Northern New Jersey Chapter of the American Statistical Association, October 9, 1984; and as Keynote Address, 30th Conference on Design of Experiments, Army Research, Development and Testing, Las Cruces, New Mexico, October 17, 1984

world, can only be assessed in terms of (i) a statistical model or (ii) sampling from a larger body (probably bodies, since one body is unlikely to suffice) of real data. Assessment in terms of a statistical model can be either by *formula-manipulation* or by *numerical calculation*, which will usually require some form of experimental sampling (direct, Monte-Carlo, configural, etc.).

If we could gain stringency without fixing a statistical model – or class of statistical models – our desire for certainty would drive us to do so. But we do know that pairs of statistical models can be invented to be so antithetic to one another that stringency for either rules out stringency for the other. The best certainty we can reach, therefore, comes by using a fairly flexible *class* of statistical models – one which we hope parallels the real world in whatever ways are most important.

This OFTEN means that we must seek *robustness of stringency.*

During the last decade or so much progress has been made in identifying summaries of small and moderate-sized samples that do provide robustness of stringency, but little attention has been paid to how such summaries should be built into experimentation.

We can have *both* high validity *and* high robustness of stringency. When we are being careful we SHOULD arrange for BOTH.

And if, in special circumstances, some other attainable property is more important than robustness, we can usually combine it with guaranteed validity by selection of a different kind of summary. When we wish to be careful, we should arrange for both validity and preferred specially-useful-property.

2. The Fisher-Babbage "legacy of a legacy".

Sir Ronald Fisher and Charles Babbage left us two things, which together were not then a legacy, but could – and have – grown up to become one. Namely randomized experiments and computers.

3. The economics of computation has changed; explicit reassignment is feasible and needed!

It was almost 50 years ago that we heard – with respect and perhaps a touch of awe – that L.J. Comrie had *half a million multiplications* to do, and had rented a set of punched-cards machines (what are now dignified as "unit record equipment") for *six months* to do just this.

This year, one can buy – for no more than a statistician's annual salary – a workstation with a few megaflops of capacity. One which, since it can do millions of multiplications in a second, could (I/0 aside) do Comrie's 6-month chore in a fraction of a second.

The important conclusions are these:

- If a million arithmetic operations can save a second's time for a statistician – and the opportunity recurs often encough – we *cannot afford NOT to do the arithmetic.*
- If reanalyzing the observed data according to a *few thousand* alternative assignments (of treatment and control) let us increase the certainty of our analysis enough to save either the statistician or his client *one second* of worry, we *cannot afford not* to do the few thousand reanalyses.
- If reanalyzing the observed data according to a *few thousand* alternative assignments let us provide more robustness of stringency for our analysis, we cannot afford not to do the few thousand analyses.

Randomization analysis, as urged by R. A. Fisher, and founded by E. J. G. Pitman (and B. L. Welch) grew up in an era of expensive computing. As a result even Pitman's combinatorially obtained higher moments were forgotten in practice. (No one wanted to sum the cubes and 4th powers of the observed differences; summing squares was hard enough.) Moreover, our attention was focused on *means*, as the summary to be used, in large part

because it was only for them that the necessary formula manipulation seemed feasible.

In the interim, while computing has become so many powers of ten cheaper, we have learned a lot about when using means will give high stringency -- ***and when it will not.*** The way to provide ***both*** the ***validity*** that randomization has always provided, ***and*** the ***robust stringency*** that it can provide is to plan to actually do many alternative analyses, using whatever summary -- often one providing robust stringency -- seemed most desirable when the experiment was planned! By doing MANY reassignments explicitly, we can be quite sure of what we are doing.

For experiments of certain sizes, we can use complete randomization, which means making calculations for all possible assignments (within the general framework of the experiment), but we cannot do this for all sizes of experiments! 2^t possible assignments means 2^t reanalyses, and we can only afford this for modest t (The bound will go up year after year). So we will have to consider combine explicit reassignment with limited randomizations -- something we can approach in different ways.

4. The details of explicit reassignment with complete randomization (2^t case)

If we have t blocks, each composed of 2 units, and if treatment and control are each to appear once in each block, there will be 2^t possible assignments.

If we have chosen a summary, for simplicity based on the t within-block treatment-MINUS-control differences (more general summaries cause no difficulties), it is straightforward to introduce all possible sequences of t $\pm$ signs, and thus, for each such sequence, t values, one per block, of

possible difference = $\pm$ actual difference

whose 2^t summaries correspond to all possible 2^t reanalyses. For each such "randomization" there is a value of the corresponding summary -- which might be a mean, a midmean, some other trimmed mean, or a biweight of the n possible differences.

We can take these 2^t summary values, rank them, and then take, as our P-value:

$$\text{P-value} = \frac{\textit{rank of actual assignment}}{2^t}$$

If we want a one-tailed P-value, we will rank from one end -- the top or the bottom, as appropriate -- starting with the relevant extreme.

If we have a two-tailed P-value, we will rank from both ends and combine the results into a single ranking. Here

Rank above	*Rank below*	*Combined rank*
1	2^n	2
2^n	1	2
2	2^n-1	4
2^n-1	2	4
3	2^n-2	6
2^n-2	3	6
4	2^n-3	8
2^n-3	4	8

(and so on)

is conventional. If, however, we want to give the rank from above a shade of preference, it is both legitimate and proper to use (instead):

Rank above	*Rank below*	*Combined rank*
1	2^n	1
2^n	1	2
2^n-1	2	4
3	2^n-2	5
2^n-2	3	6
4	2^n-3	7
2^n-3	4	8

(and so on)

Both of these rerankings satisfy the essential inequality:

$$\#\text{ of assignments with combined rank} \leq k \text{ IS ITSELF } \leq k$$

The first of these rerankings alternates $<$ and $\leq$, while the second always has $\leq$. The first is ***symmetrical,*** the second is ***not wasteful.***

If we proceed in this way, choosing our summary (and our combined ranking procedure, if we are to use one) in advance of our data gathering, and doing a trustworthy job of selecting one assignment out of 2^t sufficiently nearly at random, the ***validity*** of our P-value is guaranteed (by the quality of our randomization, which we know all about, and which others can inspect, if we keep good records).

It may be worthwhile to explain how to guarantee that 1 out of 2^t -- say 1 out of 1024 -- will be chosen very, very nearly at random. A simple approach calls for 3 people who

- can be trusted to work independently of one another, AND
- of whom at least one is both honest and skilled in randomization.

We then proceed as follows:

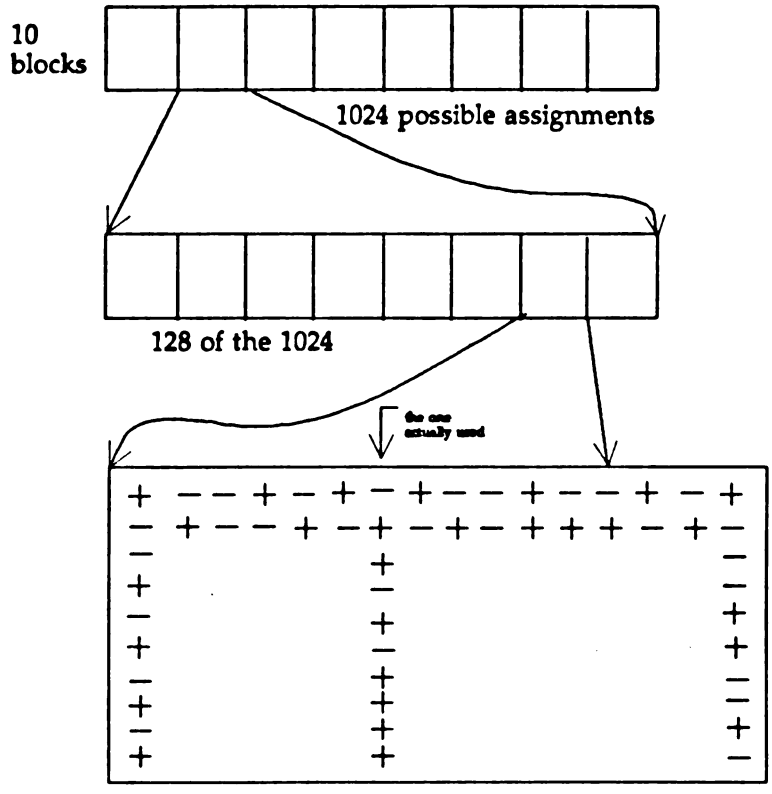
- number all possible 1024 assignments from 0 to 1023,
- request each of the three persons to produce a random integer between 0 and 1023, AND
- add up the three integers, reduce the result modulo 1024, and adopt the corresponding assignment.

No two people could combine to keep the third from making the overall choice random.

How well we do about stringency (or other desideratum) depends on how well we choose our summary. We do know which summary we chose. We may know how this summary performs for a more or less flexible statistical model -- today almost certain to involve an infinite population. The n absolute differences we actually have may collectively throw some light on how the summary performed in our specific case.

We are just about as well off about stringency -- about robustness of stringency -- as we ever are, and we are as well off about validity as we can be -- since we depend only on a process that can be monitored, recorded and inspected.

A picture may help us to get a feel:



16 of the 128 of the 1024

We have a ***stock*** of 1024 assignments and chose one at random. We summarize sets of 10 values of the form {± an observed difference} 1024 times, to get 1024 values of a given summary.

5. Some kinds of limited randomization.

The complete-randomization design and analysis of experiment just described is probably most reasonable for t blocks, where $9 \leq t \leq 12$. At $2^9 = 512$, we still have enough assignments so that comparison with both 5% and 1% points is not seriously plagued by granularity. At $2^{12} = 4096$, we can still face doing one analysis for each of four-thousand-odd assignments. But what of more than 12 blocks of 2.

We shall discuss three approaches, briefly.

* *stacking up* *

We can stack our n blocks into k stacks, locking together the assignment of treatment and control for blocks in the same stack. We then need one random ± sign per stack.

If, for example, we have 24 blocks, we may stack them in 12 stacks, say #24 with #1, #23 with #2, ..., #13 with #12, and in each stack we may choose a locking together, perhaps like this

#1	y	z	#2	y		z		#12	y	z
#24	z	y	#23	y		z		#13	z	y

where we plan to randomly assign treatment and control to y, and z or to z, and y in each

stack separately.

All our techniques, discussed today or not, WORK just as WELL for stacks -- locked-together groupings of one or more blocks -- as for single blocks. So, as we go on, we will usually talk about ***stacks*** since this covers all cases.

With k stacks stacked and locked, there are 2^k possible assignments, and we have the full validity that we had before, and the same preference for $9 \leq k$ (≤ 12). We may have to be careful to use a summary that is appropriate for blocks in stacks. If we do this carefully, there is no reason why the stacks should all contain the same number of blocks!

Used alone, this is already a very flexible form of limited randomization -- one that seems particularly appropriate for treatment-control differences whose distribution seems vaguely Gaussian in shape (because we are likely* to add up the differences within each stack).

Used together with those to be described next, "stacking up" contributes to an even greater degree of flexibility.

* *orthogonal arrays* *

There are a wide variety of ways to use orthogonal arrays -- I am about to tell you of a rather new one. (A convenient reference is Raktoe, Hedayat, and Federer 1981, ***Factorial Designs,*** specifically pages 168-189 and certain of the references at pages 188-193.)

A simple example is

1	0	0	0	0	1	1	1
0	1	0	0	1	0	1	1
0	0	1	0	1	1	0	1
0	0	0	1	1	1	1	0

whose main virtue is that, if we pick out any three rows, say the first, third, and fourth

1	0	0	0	0	1	1	1
0	0	1	0	1	1	0	1
0	0	0	1	1	1	1	0

each of the 2^3 possible columns appears an equal number of times -- namely, here, one.

Officially, this is an orthogonal array of size, n=8; t=4 constraints; s=2 levels (or alphabet size 2); and strength d=3 (meaning that ***any three*** rows are completely balanced.

In conventional notation it is an

$$OA(8, 4, 2, 3)$$

More generally, an orthogonal array is an

$$OA(n, t, s, d)$$

or an

$$OA(n, t, s, d, \lambda)$$

where λ is the common number of appearances of all subcolumns for any chosen set of d rows. Our example is thus also an OA(8, 4, 2, 3; 1).

Lest you think that every OA is a fractional factorial, let us look at an OA (18, 7, 3, 2), namely:

---

*But do not have to!

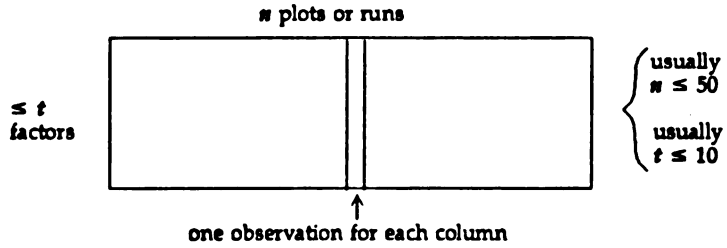
0 0 0 1 1 1 2 2 2 0 0 0 1 1 1 2 2 2
0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2
0 1 2 1 2 0 2 0 1 2 0 1 0 1 2 1 2 0
0 2 1 1 0 2 2 1 0 1 0 2 2 1 0 0 2 1
0 1 2 1 2 0 1 2 0 0 1 2 2 0 1 2 0 1
0 1 2 2 0 1 0 1 2 1 2 0 2 0 1 1 2 0
0 1 2 0 1 2 2 0 1 1 2 0 1 2 0 2 0 1

For each of the $\binom{7}{2} = 21$ choices of *two* rows from this array, we will find

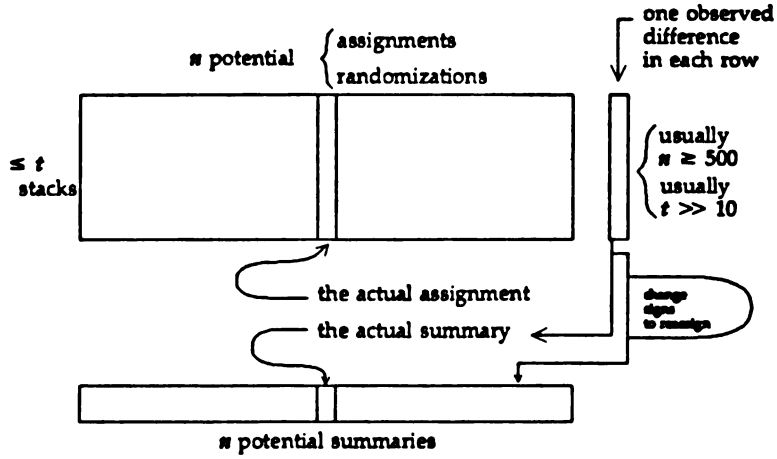
0	0	0	1	1	1	2	2	2
0	1	2	0	1	2	0	1	2

each occurring twice. Hence the orthogonality of ***strength two.***

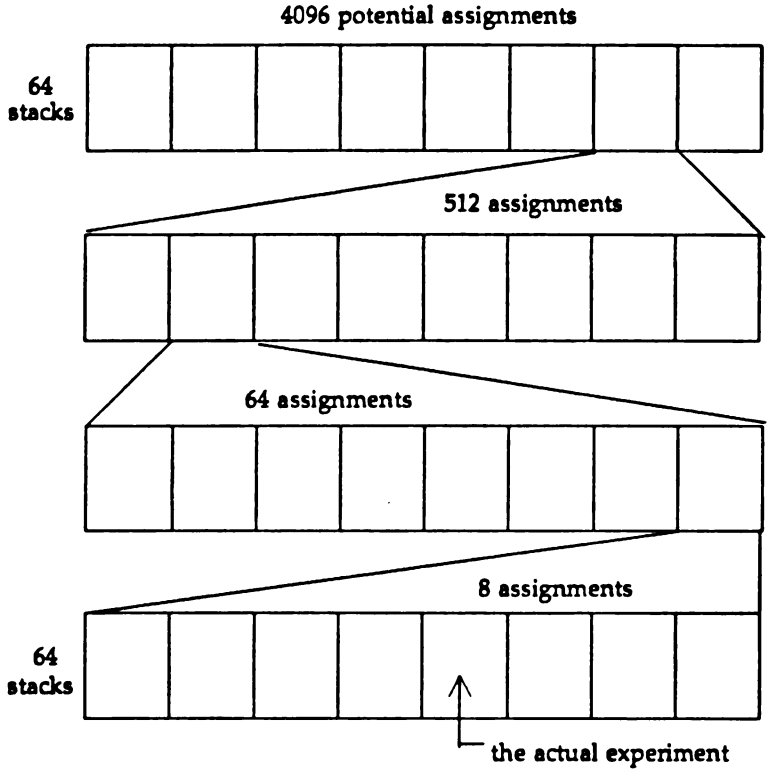
The "classical" employment of orthogonal arrays is as follows:



We are going to use them quite differently, namely:



A rather extreme example – one that will probably prove useful now and then – is shown below:



(Also bear in mind that taking all 2^t assignments as potential assignments – complete randomization for t stacks – is also using an OA!) Each of our assignments, then specifies, for each of k stacks (each stack having one or more blocks, how the treatment and control are to be assigned to the halves of that stack – and hence to the units of each of its blocks.

The ***design*** of the experiment can usefully be thought of as including the orthogonal array from which the actual assignment is to be drawn at random.

But where do we get such medium-sized orthogonal arrays?

When this question first arose, I was fortunate enough to ask Neil Sloane, who reached in his supply kit and produced a small array of such OA's, each one corresponding to a fairly well-known code (in the sense of signal-coding theory) These included (the parentheses give their labels as fractional factorial designs):

OA(512, 23, 2, 4)	(2^{23-14})
OA(2048, 23, 2, 6)	(2^{23-12})
OA(1024, 24, 2, 5)	(2^{24-14})
OA(1024, 32, 2, 4)	(2^{32-22})
OA(2048, 33, 2, 5)	(2^{33-22})
OA(2048, 63, 2, 4)	(2^{63-52})
OA(4096, 64, 2, 5)	(2^{64-52})

where we recall that, for our uses, the notations mean OA(# of assignments, up to this many blocks, 2, strength). Clearly we do quite well up to 60 blocks (or even 60 stacks of blocks).

A matter of convenience and ease of use arises because

a) these arrays are cosets of certain codes,

b) if one coset is an OA, so is any other coset of the same code, with the same parameters,

c) no two cosets overlap, while taken together they include all 2^k assignments.

As a result, we can proceed as follows:

1) choose a code, and thus a family of cosets,

2) choose, randomly, any assignment as the actual assignment – and choose the coset that includes this assignment as the set of possible assignments.

This convenient, one-step process is logically and philosophically equivalent to the more obviously sound two-step process:

(i) choose a coset at random,

(ii) choose an assignment in the coset at random

Besides convenience, and making all assignments candidates for possible assignments, this use of OA's has a particular advantage in what we might describe as *ANTI-ROBUST* situations, where there are likely to be a few very large differences – for a few stacks, or for a few blocks – whose effects we want to study and focus upon, rather than to set aside – as the use of a robust summary would do. Most weather-modification experiments, for instance, whether concerned with rainfall increase or hail-damage reduction, require antirobust analysis, since the occasional large storms yield most of the water and the occasional violent storms cause most of the damage.

* *Gaussian combination of subexperiments* *

The third approach to many blocks need not take us long. If we are willing to

- divide the experiment into subexperiments
- summarize these subexperiments separately, and then
- combine these summaries

we have only to convert the summary of each subexperiment into a reasonable (discrete) facsimile of a unit Gaussian deviate, add up the facsimiles, and divide by (# of facsimiles)$^{1/2}$, referring the result to the unit Gaussian distribution.

So how are we to get our nearly unit Gaussians from our subexperiment summaries? We may as well plan to begin by ranking our 2^j possible summaries for each subexperiment.

A little calculation of cumulants then leads us to suggest taking

$$\textit{unit Gaussian score} = Gau^{-1}\left(1 - \frac{3i-2}{3m-1}\right)$$

where

$$i = \textit{given rank}, \quad m = 2^j$$

and

$$Gau(\,) \textit{ is the unit Gaussian cumulative.}$$

This seems fairly certain to work well for 3 or more subexperiments with $2^j \geq 256$. (The case of 2 subexperiments can be bypassed treated directly, without Gaussianizing, by a different calculation.)

* *combination* *

We can clearly use these devices of combination in sequence as well as alone.

* *protective limitations* *

There have been recurrent suggestions for limiting randomization to avoid unusual assignments – principally very unbalanced assignments, occasionally too regular ones. While this is clearly more important for unblocked experiments than for blocked ones, there is no reason for not superposing such deletions on any of the randomization-reassignment patterns, complete or limited, we have considered so long as: so long as, before choosing the actual assignment:

- we decide on which deletions are to be made, AND
- we implement these deletions UNIFORMLY, to potential and actual assignments alike.

6. Compromising by rank combination.

All procedures that are robust of stringency are compromises, at least implicitly! When we want the greatest care, we are likely to want an especially thorough compromise as our chosen summary. If we can find a sufficiently broadly oriented compromise summary, so much the better. If not, then we may need to start with two or more (compromise) summaries, and, further, compromise them with one another ***explicitly***.

How should we do this? To make separate significance tests with each, and then double the smaller of the two P-values is the counsel of the Bonferroni inequality. This would almost surely be wasteful – probably substantially so. So what is surely better?

- To rank the 2^k assignments according to each summary separately – and then to combine the two sets of ranks in some appropriate way.

A simple and natural way is to put the two ranks for each assignment in lexicographical order within each pair, and then sort the ordered pairs in lexicographical order of pairs, and use this latter order to define a combined rank. The resulting correspondence

$$\begin{matrix}\text{observed numbers} \\ \text{possible assignment}\end{matrix} \longrightarrow \text{combined rank}$$

does not at all depend on which of the ***possible*** assignments was the ***actual*** assignment. So we know that

$$\frac{\textit{combined rank}}{2^k}$$

is a legitimate P-value, so that we are making a proper significance test. If the two summaries were secretly the same, so too would be the individual ranks – and the combined ranks – so that the compromise P-value would reduce to the common P-value for the individual summaries. If the two summaries are very, very different, the result of this sort of rank combination would be very close to the Bonferroni result.

A hypothetical example may help to fix the ideas.

assignment #	Ranks according to 1st summary	2nd summary	ordered pair	combined rank
351	5	1	(1,5)	1
679	1	29	(1,29)	2
224	2	42	(2,42)	3
534	97	2	(2,97)	4
976	3	4	(3,4)	5
214	14	3	(3,14)	6
723	4	53	(4,53)	7
431	8	5	(5,8)	8
92	6	6	(6,6)	9
127	7	8	(7,8)	10
923	11	7	(7,11)	11
23	9	11	(9,11)	12

Notice that, except for tie-breaking, the first 7 lines follow the Bonferroni bound to within a unit, but, that beyond 7, the similarities in the two separate-summary rankings are so great as to produce deviations from the bound. Indeed, for (5,x), (6,y), (7,z) and (8,w) together we use up only 4 ranks instead of 8.

This sort of indirect explicit compromise is certainly interesting, and probably promising.

7. Introduction to 3 or more treatments.

What if we do 3 treatments in n blocks of 3? How do we proceed?

Unless the treatments are associated with the "levels" of a quantitative x, our most likely purpose is to assess the significance of (to say how sure we seem to be of the direction of) the pair-wise comparisons. So how should we compare any two treatments?

Again there is more than one approach.

Let the three treatments be F, G, H. There are $6=3!$ ways of assigning one each to the 3 units of any block. With n blocks; and all assignments possible, there are 6^n possible assignments.

If we want to compare F with G, and do this without interference from treatment H, which may produce very different values, a very simple thing is to only consider those 2^n assignments as "possible" in which the *UNordered* pair F, G are assigned to the same two units in each block.

This means that we look at one family of 2^n assignments to compare F with G and another family of 2^n to compare F with H. These two families will, in fact, have only one assignment in common -- namely the actual one. (This means that we do not have alternative rankings for different pair-wise comparisons, so that we seem to be driven back on a Bonferroni-combined P-value.)

Like the cosets in our code-derived OA's, the families of 2^n -- out of 6^n -- assignments are non-overlapping. The selection of any one of the 2^n members of such a family as the actual assignment forces that family on us as the basis for a particular paired comparison. Although there is not, in this case, a clear two-step equivalence, the one-step procedure is just as valid as the one-step procedure in the earlier instance.

* *second approach* *

To avoid both

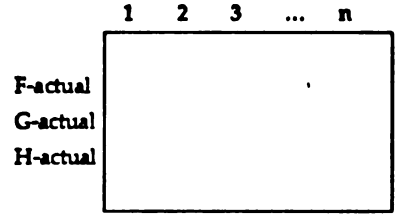
i) the potential effects of the other treatment(s) -- regarded as well-behaved though possibly large, and

ii) the full Bonferroni penalty, in those circumstances where it is not all justified,

we have to be somewhat trickier – though the resulting procedure is almost surely nearly enough legitimate and valid. (Any doubts that we may have of practical performance must then be devoted to its stringency.)

What appears to be the simplest such approach is the following:

- use only a cyclic three of the six assignments to the units of any block – e.g. use FGH, HFG or GHF but not HGF, FGH, or GFH (or *vice versa*)
- bend the rules by analyzing the actual two-way table



additively -- perhaps by lomedian polish, and

- adjust every H-actual value by subtracting

$$\textit{H-fit MINUS } \frac{\textit{(F-fit)} + \textit{(G-fit)}}{2}$$

from it

- summarize the comparisons of F-possible and G-possible for all 3^n possible assignments, using whichever two of F-actual, G-actual, and ***H- adjusted*** (notice that only one of the three is adjusted) correspond to the possible assignment at each of the blocks in question.
- repeat the last two steps with first F, and then G, replacing H.

The result of all this is a ranking for each combination of one of the $3 = \binom{3}{2}$ paired comparisons with each of the possible assignments. We can go on to combine the 3 ranks for a given possible assignments in each of a number of ways, leading to P-values for a number of kinds of multiple comparisons – ranging from multiple comparisons focusing on finding at least one difference significant to multiple comparisons focusing on finding as many differences as possible significant.

8. Closing comment.

Not only does limited randomization with explicit reassignment provide us with guaranteed validity, but, as the previous sentence about multiple comparisons suggest, it provides us with a basis for better understanding of a variety of statistical devices.

REFERENCES

Brillinger, D.R., Jones, L.V., Tukey, J.W. (1978). "The role of statistics in weather resources management," II *The Management of Weather Resources,* Stock No. 003-018-00091-1, Superintendent of Documents, U. S. Government Printing Office, Washington, D. C. 20402

Tukey, J. W. (1984). "Improving crucial randomized experiments -- by double randomization and rank combination," in *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer,* Volumes I and II, Lucien M. Le Cam and Richard A. Olshen, eds., Wadsworth Publishing Company, Monterey, California.

ATTENDANCE LIST
FOR THE
TUTORIAL SEMINAR ON "THE BOOTSTRAP"
15-16 OCTOBER 1984
AND THE
THIRTIETH CONFERENCE ON THE DESIGN OF EXPERIMENTS
IN ARMY RESEARCH, DEVELOPMENT AND TESTING
17-19 OCTOBER 1984

CONFERENCE (C)
SEMINAR (S)
BANQUET (B)

C — ACCHIONE-NOEL, SYLVIA
USA TRADOC OPERATIONS RESEARCH ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-2038 (AV 258-2038)

S,C — AGEE, WILLIAM S.
NATIONAL RANGE OPERATIONS DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2361 (AV 258-2361)

C,B — ANDREW, MIKE
COASTAL ENGINEER RESEARCH CENTER
601-634-2077

C — ANITOLE, GEORGE
US ARMY BELVOIR R&D CENTER
703-664-6741 (AV 354-6741)

C,B — ANNETT, CLARENCE
US ARMY COMBINED ARMS CENTER
(AV 552-3635)

S — ARAGON, FILEMON
NATIONAL RANGE OPERATIONS DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-1628 (AV 258-1628)

S,C — AVARA, ELTON P.
ATMOSPHERIC SCIENCES LABORATORY
WHITE SANDS MISSILE RANGE
505-678-4501 (AV 258-4501)

S BAAS, ELWOOD
COMPUTER SYSTEMS DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-5603 (AV 258-5603)

S BANKS, THOMAS
DOD AREA FREQUENCY COORDINATOR
WHITE SANDS MISSILE RANGE
505-678-5417 (AV 258-5417)

S,C BARBER, TED
ATMOSPHERIC SCIENCES LABORATORY
WHITE SANDS MISSILE RANGE
505-678-2542 (AV 258-2542)

B BARELA, CHARLENE
INSTRUMENTATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2969 (AV 258-2969)

S BARKER, RICHARD
UNIVERSITY OF TEXAS AT EL PASO
915-747-5761

S,C,B BATES, CARL B.
US ARMY CONCEPTS ANALYSIS AGENCY
(AV 295-0163)

C BINGHAM, PAUL
PLANS AND QUALITY ASSURANCE DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-678-3026 (AV 258-3026)

S,C,B BISSINGER, BARNEY
PENN STATE UNIVERSITY
717-948-6091

S BLANCO, ABEL J.
ATMOSPHERIC SCIENCES LABORATORY
WHITE SANDS MISSILE RANGE
505-678-4207 (AV 258-4207)

S BOLING, CLIFFORD
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-1070 (AV 258-1070)

C,B BOX, GEORGE E. P.
UNIV OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER
608-263-2658

C,B
BUMGARNER, THOMAS C.
PLANS AND QUALITY ASSURANCE DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2171 (AV 258-2171)

S,C,B
BURGE, J. ROBERT
WALTER REED ARMY INSTITUTE OF RESEARCH
202-576-3151

C
BURLBAW, EDWARD
SCIENCE & TECHNOLOGY CORPORATION
LAS CRUCES, NEW MEXICO
505-523-8541

S,C
CASSADY, PATRICK D.
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-5386 (AV 258-5386)

S,C
CASTILLO, CESAR
NATIONAL RANGE OPERATIONS DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2311 (AV 258-2311)

S,C
CASTRO, OSCAR J.
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2184 (AV 258-2184)

S,C
CHANDRA, JAGDISH
ARMY RESEARCH OFFICE
919-549-0641 (AV 935-0641)

C
COL CHERNAULT, JAMES A.
DEPUTY COMMANDER
WHITE SANDS MISSILE RANGE
505-678-1959 (AV 258-1959)

C
CHIEFFO, ANGELO
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-2038 (AV 258-2038)

C
CLEVENGER, TOM
NEW MEXICO STATE UNIV
DEPT OF AG ECON & AG BUS
505-646-2524

S
COLTHARP, BOB
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-9334 (AV 258-9334)

S,C CONOVER, W. J.
TEXAS TECH UNIVERSITY
COLLEGE OF BUSINESS ADMINISTRATION
806-742-1546

C,B CONROY, ROBERT
ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY
ABERDEEN PROVING GROUND
301-278-3675 (AV 283-3675)

C CORPENING, LOUIS
PLANS AND QUALITY ASSURANCE DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2174 (AV 258-2174)

S,C COUCHMAN, DIANE
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-3496 (AV 258-3496)

S,C COUNCIL, KONRAD K.
NATIONAL RANGE OPERATIONS DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2311 (AV 258-2311)

C,B COX, PAUL
RETIRED CIVIL SERVICE
505-522-1756

C,B CROW, LARRY H.
ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY
ABERDEEN PROVING GROUND
301-278-6623 (AV 283-6623)

S,C,B CRUESS, DAVID F.
U.S.U.H.S. SCHOOL OF MEDICINE
DEPT OF PREVENTIVE MED
202-295-3709

C,B CULPEPPER, GIDEON
RETIRED CIVIL SERVICE
505-524-4247

S,C DALTON, OREN N.
NATIONAL RANGE OPERATIONS DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2311 (AV 258-2311)

C DANNHAUS, DALE
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-2034 (AV 258-2034)

C
DARUWALLA, A. N.
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-2043 (AV 258-2043)

S,C
MAJ DAVIS, DAVID F.
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-4046 (AV 258-4046)

C
DAVIS, ROGER
SCIENCE & TECHNOLOGY CORPORATION
LAS CRUCES, NEW MEXICO
505-523-8541

B
DAW, HAROLD A.
ASSOC ACADEMIC V. P.
NEW MEXICO STATE UNIVERSITY
505-646-2022

S,C,B
DEASON, PAUL J.
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-3289 (AV 258-3289)

C,B
DEMPSTER, ARTHUR
HARVARD UNIVERSITY
617-495-5498

S,C
DIAZ, ALFONSO A.
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-1373 (AV 258-1373)

S
DIAZ, SYLVIA
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-5934 (AV 258-5934)

S,C
DONAHE, LEONARD D.
NATIONAL RANGE OPERATIONS DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2311 (AV 258-2311)

S,C
DONOVAN, KAYE
PROG ANAL & EVAL
AMSEL-POD-SA
FT MONMOUTH, NJ
(AV 992-2926)

C
DRAGON, ROBERT A.
PLANS AND QUALITY ASSURANCE DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2173 (AV 258-2173)

S,C
DUTOIT, EUGENE
DIRECTORATE OF
COMBAT DEVELOPMENTS
FT BENNING, GA
404-545-3165 (AV 835-3165)

S
DYLLA, FRANK
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-3496 (AV 258-3496)

B
COL EBERT, VERNON E.
CHIEF OF STAFF
WHITE SANDS MISSILE RANGE
505-678-2228 (AV 258-2228)

C
ELLINGSON, ANDREW C.
NATIONAL RANGE OPERATIONS DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2311 (AV 258-2311)

S,C
ELLIS, JIM
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-3717 (AV 258-3717)

C
ELMORE, ROBERT E.
INSTRUMENTATION PLANS & ANALYSIS DIVISION
YUMA PROVING GROUND
602-328-3111 (AV 899-3111)

S
ELOWITZ, BILL
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-9135 (AV 258-9135)

C
MAJ ENGSTROM, GREG
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-2043 (AV 258-2043)

C,B
ESSENWANGER, OSKAR
RESEARCH DIRECTORATE
US ARMY MISSILE COMMAND
205-876-4872

S,C
EVANS, TROY
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-9227 (AV 258-9227)

S,C,B
FARRELL, RUSSELL
ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY
ABERDEEN PROVING GROUND
(AV 283-6628)

C
FINCH, PAUL
IE DEPARTMENT
NEW MEXICO STATE UNIVERSITY
505-646-2950

C,B
FREITAS, STEPHEN
US ARMY NATICK
RESEARCH & DEVELOPMENT CENTER
(AV 256-5071)

C,B
FREYER, GUSTAV J.
PHYSICAL SCIENCE LAB
NEW MEXICO STATE UNIVERSITY
505-522-9409

C,B
GEMMILL GARY W.
NAVAL SURFACE WEAPONS CENTER
703-663-7420

C
GLAZE, RICHARD
STATISTICS CENTER
NEW MEXICO STATE UNIVERSITY
505-646-2936

B
GONZALEZ, ALBERTO
INSTRUMENTATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-5214 (AV 258-5214)

C
GONZALEZ, FRANK
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-1070 (AV 258-1070)

C
GOODALL, DON
PLANS AND QUALITY ASSURANCE DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-5683 (AV 258-5683)

C
GOODWILL, GORDON
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-2043 (AV 258-2043)

S,C,B
GREEN, ROBERT E.
INSTRUMENTATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2291 (AV 258-2291)

C
GREGORY, GAVIN
UNIVERSITY OF TEXAS AT EL PASO
915-747-5332

S GRIGORY, WALTER
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-9135 (AV 258-9135)

C,B GRENANDER, ULF
BROWN UNIVERSITY
401-863-2463

C,B GRYNOVICKI, JOCK
US ARMY BALLISTIC
RESEARCH LABORATORY
301-278-6659 (AV 283-6659)

C HALL, CARROLL
EXPERIMENTAL STATISTICS DEPARTMENT
NEW MEXICO STATE UNIVERSITY
505-646-2936

C HALLIGAN, JAMES E.
PRESIDENT
NEW MEXICO STATE UNIVERSITY
505-646-2035

S,C,B HARRIS, BERNARD
UNIVERSITY OF WISCONSIN
MATHEMATICS RESEARCH CENTER
608-263-2656

S HEATH, ROY L.
NATIONAL RANGE OPERATIONS DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2311 (AV 258-2311)

S,C,B HOFMANN, JAMES D.
US ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY
ABERDEEN PROVING GROUND
301-278-6629 (AV 283-6629)

S,C,B HOLMAN, HERBERT H.
NATIONAL SECURITY AGENCY
(AV 859-6323)

C HULL, JAN C.
DEPARTMENT OF MATHEMATICAL SCIENCES
NEW MEXICO STATE UNIVERSITY
505-646-2936

S,C,B HUNTER, CHARLES J.
DMS/ORAE
DND (CANADA)
613-992-3684

C,B
HUSSAIN, MOHAMMED Y.
NEW MEXICO STATE UNIVERSITY
EXPERIMENTAL STAT DEPT
505-646-2937

C
IMAN, RON
SANDIA NATIONAL LABORATORIES
505-844-8834

C
JACOBS, PAMELA M.
BELVOIR R&D CENTER
703-664-5086

S
JAMES, SHELDON
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-3717 (AV 258-3717)

C
JELIHOUSEHI, ENIO
UNIVERSITY OF TEXAS AT EL PASO
915-747-5761

C
JOHNSON, RONALD
US ARMY BELVOIR R&D CENTER
(AV 354-6771)

C,B
KAPUR, KAILASH
US ARMY TANK AUTOMOTIVE COMMAND
313-574-6864

S
KILLAN, WILLIAM R.
DIRECTED ENERGY DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-1930 (AV 258-1930)

S,C,B
KNAUB, JIM
US ARMY LOGISTICS CENTER
804-734-5662 (AV 687-5662)

C
MAJ KOLB, RICKEY A.
US MILITARY ACADEMY
914-938-2719

S,C,B
LAUNER, ROBERT, L.
ARMY RESEARCH OFFICE
MATH SCIENCES DIVISION
(AV 936-3331)

C LINLBACK, WILBUR
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-2043 (AV 258-2043)

S,C LYSNE, DALE
PROG ANAL & EVAL
AMSEL
(AV 995-2590)

S,C,B MAAR, JAMES R.
NATIONAL SECURITY AGENCY
(AV 859-6461)

S,C MADDEN, DALE A.
FOOD & DRUG ADMINISTRATION
301-443-3240

C MAESE, GLORIA
NEW MEXICO STATE UNIVERSITY
EXPERIMENTAL STAT DEPT
505-646-2936

S,C MAHER, MARY ANNE
NEW MEXICO STATE UNIVERSITY
505-646-2957

S,C MARDO, JOHN G.
US ARMY AMCCOM
201-724-3008 (AV 880-3008)

S MARTINEZ, GIL
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-679-3496 (AV 258-3496)

S,C MARUYAMA, RICHARD T.
HQ TRADOC
(AV 680-3004)

S,C,B MCAFEE, WALTER S.
AMDEL-SA
(AV 992-4749)

C,B MCINTOSH,WILLIAM B.
TECOM HQ
(AV 283-3026)

S,C MCLAUGHLIN, DALE R.
NATIONAL RANGE OPERATIONS DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2311 (AV 258-2311)

S MCMANUS, LAWRENCE
USA ARMY TRADOC OPERATIONS RESEARCH ACTIVITY
505-678-4631 (AV 258-4631)

C MEASON, JOHN
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-679-9484 (AV 258-9484)

S,C,B MOORE, J. RICHARD
US ARMY BALLISTIC RESEARCH LABORATORY
301-278-6666 (AV 283-6666)

C MURRAY, LEIGH
NEW MEXICO STATE UNIVERSITY
EXPERIMENTAL STAT DEPT
505-646-5186

S MYERS, BRENDA
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-9330 (AV 258-9330)

C NANCE, ANN
UNIV STATISTICS CENTER
NEW MEXICO STATE UNIVERSITY
505-646-2936

C,B NEAL, DONALD M.
AMMRC
617-923-5165

C NGUYEN, HUNG T.
NEW MEXICO STATE UNIVERSITY
DEPT OF MATH SCIENCES
505-646-2105

C NOBLE, JOHN L.
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-4672 (AV 258-4672)

S ODOM, ROGER
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-3496 (AV 258-3496)

S,C,B
ORTIZ, MELCHOR
NEW MEXICO STATE UNIVERSITY
EXPERIMENTAL STAT DEPT
505-646-2936

C
PARRA, MARIO Z.
PLANS AND QUALITY ASSURANCE DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2173 (AV 258-2173)

C,B
PARZEN, EMANUEL
TEXAS A&M UNIVERSITY
DEPT OF STATISTICS
409-845-3188

S
PETERSON, EMERY
COMPUTER SYSTEMS DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2947 (AV 258-2947)

C,B
PETTIBONE, TIMOTHY
NEW MEXICO STATE UNIVERSITY
ED. RES. CENTER
505-646-1500

C,B
PHILLIPS, KEITH
NEW MEXICO STATE UNIVERSITY
DEPT OF MATH SCIENCES
505-646-1223

S
PRICE, RAYMON
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-1763 (AV 258-1763)

C,B
PROSCHAN, FRANK
FLORIDA STATE UNIVERSITY
904-644-3218

C,B
COL RANDALL, JACK
DEPUTY FOR AIR FORCE
WHITE SANDS MISSILE RANGE
505-678-1251 (AV 258-1251)

C
RANKIN, DONALD
RETIRED CIVIL SERVICE
915-755-5346

C
RICHARDS, WINSTON
NEW MEXICO STATE UNIVERSITY
505-646-2936

S
RICKLIN, JENNIFER
ATMOSPHERIC SCIENCES LABORATORY
WHITE SANDS MISSILE RANGE
505-678-4207 (AV 258-4207)

C
ROBERTS, PAMELA
USA TRADOC OPERATIONS RESEARCH ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-2034 (AV 258-2034)

S
ROENFELDT, ROGER D.
ARMY MATERIEL TEST AND EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2184 (AV 258-2184)

S,C
ROGERS, GERALD
NEW MEXICO STATE UNIVERSITY
505-646-2217

C
ROJO, JAVIER
UNIVERSITY OF TEXAS AT EL PASO
915-747-5761

S,C,B
RUSSELL, CARL T.
OTEA
(AV 289-2389)

C
SAIA, ED
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-2043 (AV 258-2043)

C,B
SAMPSON, JAMES B.
US ARMY NATICK
RESEARCH & DEVELOPMENT CENTER
(AV 256-4698)

C
SANCHEZ, JAIME
EXP STAT DEPT
NEW MEXICO STATE UNIVERSITY
505-646-2936

S
SANCHEZ, SAM
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-9135 (AV 258-9135)

S
SCHOEN, O. W.
DIRECTED ENERGY DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-3736 (AV 258-3736)

C SCHUSTER, EUGENE
UNIVERSITY OF TEXAS AT EL PASO
915-747-5761

S,C,B SCLOVE, STANLEY L.
UNIVERSITY OF ILLINOIS AT CHICAGO
312-996-2601

C SCOTT, DAVID
NEW MEXICO STATE UNIVERSITY
505-646-2936

C,B SETHURAMAN, J.
FLORIDA STATE UNIVERSITY
DEPT OF STATISTICS
904-644-3218

C SHOOK, GEORGE W.
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-4300 (AV 258-4300)

C,B SINGPURWALLA, NOZER D.
GEORGE WASHINGTON UNIVERSITY
DEPT OF OPERATIONS RESEARCH
202-676-7515

C,B SMITH, DAVID W.
NEW MEXICO STATE UNIVERSITY
EXPERIMENTAL STATISTICS DEPARTMENT
505-646-2936

S,C,B SOUTHWARD, G. MORRIS
NEW MEXICO STATE UNIVERSITY
EXPERIMENTAL STATISTICS DEPARTMENT
505-646-2936

C STEWART, PERRY
EXECUTIVE INSTITUTE
DEFENSE SYS MANAGEMENT COLLEGE
(AV 354-5785)

S SWEANY, TOM
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-3496 (AV 258-3496)

S,C SWINGLE, DONALD
CONSULTANT
505-522-5197

S,C,B TANG, DOUGLAS
WALTER REED
ARMY INSTITUTE OF RESEARCH
202-576-2212

S,C TESIERMAN, DELORIS
ABERDEEN PROVING GROUND
COMBAT SYSTEMS TEST ACTIVITY
301-278-3943

C THORNE, WALTER
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-2038 (AV 258-2038)

S,C THRASHER, PAUL H.
PLANS AND QUALITY ASSURANCE DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-2172 (AV 258-2172)

S TIBSHIRANI, ROBERT J.
STANFORD UNIVERSITY
415-497-2627

S TOFSTED, DAVID
ATMOSPHERIC SCIENCES LABORATORY
WHITE SANDS MISSILE RANGE
505-678-3039 (AV 258-3039)

C,B TUKEY, JOHN W.
PRINCETON UNIVERSITY
BELL LABS
609-452-4219

S,C TURNER, ROBERT H.
NATIONAL RANGE OPERATIONS DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-5833 (AV 258-5833)

C URQUHART, N. SCOTT
DEPARTMENT OF EXPERIMENTAL STATISTICS
NEW MEXICO STATE UNIVERSITY
505-646-2936

S VALVERDE-WARD, DEBORAH
ARMY MATERIEL TEST & EVALUATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-1285 (AV 258-1285)

S,C,B VANGEL, MARK
ARMY MATERIALS
AND MECHANICS RESEARCH CENTER
617-923-5414

B VOSS, ROBERT A.
ADVANCED TECHNOLOGY OFFICE
INSTRUMENTATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-5702 (AV 258-5702)

C WALKER, CAROL L.
MATHEMATICS DEPARTMENT
NEW MEXICO STATE UNIVERSITY
505-646-3901

C WALKER, ELBERT
NEW MEXICO STATE UNIVERSITY
505-646-2707

C,B WEEKS, DAVID L.
OKLAHOMA STATE UNIVERSITY
DEPT OF STATISTICS
405-624-5684

C,B WELSCH, ROY E.
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
617-253-6601

C WILLEY, DARRELL
NEW MEXICO STATE UNIVERSITY
505-646-1629

S WOEHL, WALTER
DIRECTED ENERGY DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-3736 (AV 258-3736)

S WOOD, ROBERT
USA TRADOC SYSTEMS ANALYSIS ACTIVITY
WHITE SANDS MISSILE RANGE
505-678-4949 (AV 258-4949)

C YAKOWITZ, SID
UNIVERSITY OF ARIZONA
602-621-6559

C YOUNG, BARRY H.
INSTRUMENTATION DIRECTORATE
WHITE SANDS MISSILE RANGE
505-678-3712 (AV 258-3712)

C,B ZACKS, SHELEMYAHU
STATE UNIVERSITY OF NEW YORK-BINGHAMTON CTR
607-798-2619

★ U.S. GOVERNMENT PRINTING OFFICE: 1985 539-921

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: ARO Report 85-2	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): PROCEEDINGS OF THE THIRTIETH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING	5. TYPE OF REPORT & PERIOD COVERED: Interim Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Army Mathematics Steering Committee on Behalf of the Chief of Research, Development and Acquisition	12. REPORT DATE: July 1985
	13. NUMBER OF PAGES: 506
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): US Army Research Office, PO Box 12211, Research Triangle Park, NC 27709-2211	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. The findings in this report are to be construed as Official Department of the Army position, unless so designated by other authorized documents.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This is a technical report from the Thirtieth Conference on the Design of Experiments in Army Research, Development and Testing. It contains most of the papers presented at that meeting. These articles treat various Army statistical and design problems.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

limited randomization
regression diagnostics
object correlation
reliability growth potential
allowable value
basic programs
experimental designs
least-time analysis
shock sensitivity
cloud ceilings
multivariate data analysis
desert camouflage
nonparametric median estimation
quantile functional statistical inference
intensity of energy penetration
segmentation of signals
experimental designs
equipment performance guarantees
prioritization methodology
hypothesis testing
fire support team
water waves
radar measurements
Wilcoxon rank sum test

DD FORM 1473