

AD-A281 606


ARO Report 94-2

This document contains
blank pages that were
not filmed.

PROCEEDINGS OF THE THIRTY-NINTH
CONFERENCE ON THE DESIGN OF
EXPERIMENTS IN ARMY RESEARCH
DEVELOPMENT AND TESTING



DTIC
ELECTE
JUL 08 1994
S G D

94-20739



SBP/S

Approved for public release; distribution unlimited.
The findings in this report are not to be construed as
an official Department of the Army position, unless
so designated by other authorized documents.

DTIC QUALITY INSPECTED 8

Sponsored by

The Mathematical and Computer Sciences Division of the

ARMY RESEARCH OFFICE

94 6 183

U.S. Army Research Office

Report No. 94-2

June 1994

**PROCEEDINGS OF THE THIRTY-NINTH CONFERENCE
ON THE DESIGN OF EXPERIMENTS**

Sponsored by the Army Mathematics Coordination Group

HOST

**The Department of Statistics at Rice University
Houston, Texas**

20-22 October 1993

**Approved for public release; distribution unlimited.
The findings in this report are not to be construed
as an official Department of the Army position, unless
so designated by other authorized documents.**

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

Accession For	
NTIS	CRA&I
DTIC	TAB
Unannounced	
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
9-1	

FOREWORD

The host for the Thirty-Ninth Conference on the Design of Experiments in Army Research, Development and Testing was the Department of Statistics in Rice University. Professor James R. Thompson, Department of Mathematics, invited this conference to be held at Rice University. He was asked to be Chairperson of this conference which was held on the 20 - 22, October, 1993. Dr. Thompson was assisted in this task by Mrs. Diane J. Brown, Department Coordinator. These individuals are to be commended for their efforts in coordinating all the details required to conduct this large successful scientific meeting.

Members of the problem committee were pleased to obtain the services of the following distinguished scientists to speak on topics of interest to Army personnel:

<u>Speaker and Affiliation</u>	<u>Title of Address</u>
Professor Dennis Cox Rice University	Estimating Parameters in Complex Computer Codes: Designing the Computer Experiments
Professor Katherine B. Ensor Rice University	Properties of Simulation based Estimators of Stochastic Processes
Professor Wei-Yin Loh University of Wisconsin-Madison	Tree-Structured Statistical Methods
Professor Emanuel Parzen University Gave the Keynote Address	Beyond Classical Statistical Methods: Why and How
Professor J. Sethuraman Florida State University	Contamination of Failure Data can Change Nature of Failure Rate and Explain the Strength of Long Life Units
Professor Nozer Singurwalla and Jiangxian Chen George Washington University	On the Reliability of Emergency Diesel Generators at U.S. Nuclear Plants

This conference was preceded by a two day tutorial entitled "Multivariate Density Estimation and Visual Clustering" presented by Professor David W. Scott of Rice University. The purpose of these tutorials is to develop, in Army scientists, an interest in and an appreciation for the statistical methods that are needed to analyze experimental data.

Dr. Douglas B. Tang, Chief of the Department of Biostatistics at the Walter Reed Army Institute of Research, was selected to receive the Twelfth U.S. Army Wilks Award for contributions to statistical Methodologies in Army Reserve Development and testing. Based on his diverse research productivity, he has become widely recognized as an authority on clinical trials, medical decision making, bioassay, and laboratory data analysis.

The Program Committee has requested that the proceedings of the 1993 conference be distributed Army-wide so that the information contained therein can assist scientists with some of their statistical problems. Finally, committee members would like to thank the Program Committee for all the work it did in putting together this scientific meeting.

Program Committee

Gerald Andersen (ARO)
Kevin Beam (RAND)
Robert Burge (WRAIR)
Jock Grynovicki (ARL)
Douglas Tang (WRAIR)
Deloris Testerman (TEXCOM)
Henry Tingey (U. of DE)
Francis Dressel (ARO)

Carl Bates (CAA)
Barry Bodt (ARL)
Eugene Dutoit (AIS)
Carl Russell (TEXCOM)
Malcolm Taylor (ARL)
Jim Thompson (RICE U.)
David Cruess (USUHS)
Jerry Thomas (ARL)

TABLE OF CONTENTS*

<u>Title</u>	<u>Page</u>
Foreword	iii
Table of Contents	v
Agenda	vii
Beyond Classical Statistical Methods: Why and How Emanuel Parzen	1
Properties of Simulation Based Estimators of Stochastic Processes Katherine B. Ensor	15
A New Series of Supersaturated Designs Margaret Gelder Ehm, Marc N. Elliott and Monnie McGee	23
An Empirical Study Of The Distribution and Properties of The Slope Estimator Using Minimum Normed Distance Criterion Barbara A. Wainwright	73
Determination Of Desired Design and Operational Characteristics Of The Small Area Camouflage Cover By Ground Troops George Anitole, Ronald L. Johnson and Christopher J. Neubert	117
Tree-structured Statistical Methods Wei-Yin Loh	125
Combining Simulation Results Addressing Armor Vehicle Survivability Paul J. Deason, Ph.D.	129
A Rediscovery Of The Hedges-Lehmann Estimate Paul H. Thrasher	137

*This Table of Contents contains only the papers that are published in this manual.

<u>Title</u>	<u>Page</u>
Automatic Classification of Research Projects Based on Lexical Content Mel Brown	149
An Application of Classification with Potential Use In Reproductive Toxicology Barry A. Bodt and Ronald J. Young	157
Confidence Intervals and Tests of Hypotheses for Normal Coefficients of Variation Mark G. Vangel	173
Improved Periodogram-based Estimators of Frequency for the Cosinor Model R. John Weaver, Marshall M. Brunden and Jonathan Raz	185
Meta-Analysis of Gas Flow Resistance Measurements Through Packed Beds Malcolm S. Taylor and Csaba K. Zoltani	197
Desktop Models For Weapons Analysis John D'Errico and Eugene F. Dutoit	209
An Application of Generalized p-Values in Tank Gun Accuracy Research David W. Webb	295
Identifying The Critical Factors in An Adaptive Network Ann E.M. Brodeen, Barbara D. Broome, George W. Hartwig, Jr., and Maria C. Lopez	305
Measures of Effectiveness for Monte Carlo Sensitivity Analyses Andrew W. Harrell	309
Attendance List	327

**THIRTY-NINTH CONFERENCE ON THE DESIGN OF
EXPERIMENTS
IN ARMY RESEARCH, DEVELOPMENT, AND TESTING**

18-22 October 1993

**Host: Department of Statistics
Rice University
6100 South Main St.
Houston, Texas**

Location: Kyle Morrow Room, Fondren Library

Wednesday, 20 October 1993

0800 - 0915 REGISTRATION: (Kyle Morrow Room Lobby)

0915 - 0930 CALL TO ORDER: Jim Thompson, Rice University

**OPENING REMARKS: (Michael M. Carroll, Dean of Engineering,
Rice University)**

0930 - 1200 GENERAL SESSION I

Chairperson: Malcolm Taylor, Army Research Laboratory

**0930-1030 KEYNOTE ADDRESS: BEYOND AOV STATISTICAL
METHODS
Emanuel Parzen, Texas A&M University**

1030-1100 Break

**1100-1200 PROPERTIES OF SIMULATION BASED ESTIMATORS OF STOCHASTIC
PROCESSES
Katherine B. Ensor, Rice University**

1200 - 1330 Lunch

1330 - 1500 CONTRIBUTED SESSION I

Chairperson: Linda Moss, Army Research Laboratory

**PARTIALLY DUPLICATED FACTORIAL DESIGNS
Peter W. M. John, University of Texas at Austin**

**AN APPLICATION OF GENERALIZED P-VALUES IN
TANK GUN ACCURACY RESEARCH
David W. Webb, Army Research Laboratory**

A SERIES OF NEW SUPERSATURATED DESIGNS
Margaret G. Ehm, Marc N. Elliott, and Monnie McGee, Rice
University

1500 - 1530 Break

1530 - 1700 CONTRIBUTED SESSION II

Chairperson: Carl Russell, TEXCOM

JUDGING STATISTICAL SIGNIFIANCE GRAPHICAL METHODS VS
TRADITIONAL PARAMETRIC METHODS
Jock O. Grynovicki, Army Research Laboratory

AN EMPIRICAL STUDY OF THE DISTRIBUTION AND PROPERTIES OF
THE SLOPE ESTIMATOR USING THE MINIMUM NORMED DISTANCE
CRITERION

Barbara Wainwright, Salisbury State University and Henry B.
Tingey, University of Delaware

CHARACTERIZATION RESULTS IN PROBABILITY
Jerry Andersen, Army Research Office

DETERMINATION OF DESIRED DESIGN AND OPERATIONAL
CHARACTERISTICS OF THE SMALL AREA CAMOUFLAGE COVER
(SACC) BY GROUND TROOPS

George Anitole and Ronald L. Johnson Belvoir Research
Development and Engineering Center & Christopher J. Neubert,
Army Materiel Command

1830 - WILKS AWARD BANQUET (Cohen House/Faculty Club, Rice
University)

1830-1930 Cash Bar

1930- Dinner

Thursday, 21 October 1993

0800 - 0900 GENERAL SESSION II

Chairperson: Deloris Testerman, TEXCOM

TREE-STRUCTURED STATISTICAL METHODS
Wei-Yin Loh, University of Wisconsin-Madison

0900 - 0915 Break

0915 - 1100 CLINICAL SESSION

Chairperson: W. J. Conover, Texas Technological University
Panelists: Bernard Harris, University of Wisconsin-Madison
Wei-Yin Loh, University of Wisconsin-Madison
J. Sethuraman, Florida State University
Noser Singpurwalla, George Washington University

MOBILITY FACTOR INFERENCE

C. Denise Bullock and Nancy Renfroe Waterways Experiment Station

COMBINING SIMULATION RESULTS ADDRESSING ARMOR VEHICLE RESEARCH, DEVELOPMENT AND TESTING

Paul J. Deason, TRADOC Analysis Center-WSMR

P-VULTE (P-VALUE UPPER & LOWER TEST ESTIMATION)

Paul H. Thrasher, Material Test Directorate--WSMR

1100 - 1115 Break

1115 - 1215 CONTRIBUTED SESSION III

Chairperson: Doug Tang, Walter Reed Army Institute of Research

AUTOMATIC CLASSIFICATION OF DOCUMENTS BY LEXICAL CONTENT

Mei Brown, Army Research Office

AN APPLICATION OF CLASSIFICATION WITH POTENTIAL USE IN REPRODUCTIVE TOXICOLOGY

Barry A. Bodt, Army Research Laboratory & Ronald J. Young, Edgewood Research, Development and Engineering Center

1215 - 1330 Lunch

1330 - 1500 CONTRIBUTED SESSION III (CONTINUED)

IMPROVED PERIODGRAM ESTIMATORS FOR THE COSINOR MODEL

R. John Weaver and Marshall Brunden, The Upjohn Company & Jonathon Raz, University of Michigan

CONFIDENCE INTERVALS AND TESTS OF HYPOTHESES FOR NORMAL COEFFICIENTS OF VARIATION

Mark G. Vangel, Army Research Laboratory

ANALYSIS OF GAS FLOW RESISTANCE MEASUREMENT THROUGH PACKED BEDS.

Malcolm S. Taylor & Csaba K. Soltani, U. S. Army Research Laboratory

1500 - 1530 Break (POSTER SESSION, Kyle Morrow Room Lobby)

DESKTOP MODELS FOR WEAPONS ANALYSES

Eugene Dutoit and John D'Errico, Infantry School

1530 - 1630 GENERAL SESSION III

Chairperson: Jerry Thomas, Army Research Laboratory

**ON THE RELIABILITY OF EMERGENCY DIESEL GENERATORS AT U. S.
NUCLEAR POWER PLANTS**

Noser Singpurwalla & Jiangdan Chen, George Washington
University

Friday, 22 October 1993

0800 - 0900 GENERAL SESSION IV

Chairperson: Bob Burge, Walter Reed Army Institute of Research

**CONTAMINATION OF FAILURE DATA CAN CHANGE NATURE OF
FAILURE RATE AND EXPLAIN THE STRENGHT OF LONG LIFE UNITS**
J. Sethuraman, Florida State University

0900 - 0915 Break

0915 - 1015 CONTRIBUTED SESSION IV

Chairperson: LTC. Ronald Scotka, TEXCOM

IDENTIFYING THE CRITICAL FACTORS IN AN ADAPTIVE NETWORK
Ann E. M. Brodeen, Barbara Broome, George
Hartwig, and Maria Lopez, Army Research Laboratory

**ESTIMATES OF THE NUMBER OF MONTE CARLO TRIALS NECESSARY
FOR MOBILITY SENSITIVITY ANALYSES**
Andrew Harrell, Waterways Experiment Station

1030 - 1200 GENERAL SESSION V

Chairperson: Barry A. Bodt, Army Research
Laboratory, and Chairman of the AMSC
Subcommittee on Probability and Statistics

**OPEN MEETING OF THE PROBABILITY AND STATISTICS
SUBCOMMITTEE OF THE ARMY MATHEMATICS STEERING COMMITTEE**

**ESTIMATING PARAMETERS IN COMPLEX COMPUTER CODES:
DESIGNING THE COMPUTER EXPERIMENT**
Dennis Cox, Rice University

ADJOURN

Program Committee

Gerald Andersen (ARO)
Kevin Beam (RAND)
Robert Burge (WRAIR)
Jock Grynovicki (ARL)
Douglas Tang (WRAIR)
Deloris Testerman (TEXCOM)
Henry Tingey (U. of DE)
Francis Dressel (ARO)

Carl Bates (CAA)
Barry Bodt (ARL)
Eugene Dutoit (AIS)
Carl Russell (TEXCOM)
Malcolm Taylor (ARL)
Jim Thompson (RICE U.)
David Cruess (USUHS)
Jerry Thomas (ARL)

BEYOND CLASSICAL STATISTICAL METHODS: WHY and HOW

by Emanuel Parzen

Department of Statistics, Texas A&M University
College Station, TX 77843-3143

ABSTRACT: This is a philosophical and technical paper about future directions of statistical theory and practice. It discusses: 1. why and how components of statistical reasoning, 2. certified professional statisticians, 3. statistical computing, 4. statistical education, 5. defining the problem of statistics as probability modeling, 6. statistical education analogues to statistical modeling, 7. function representations of data and mathematical literacy, 8. the *P* value problems of statistics, 9. how to use correlation coefficients to develop beyond statistical methods

0. INTRODUCTION

This is a philosophical and technical paper about future directions of statistical theory and practice. We propose that the concept of comparing and combining classical statistical methods and modern data analysis methods should be called "Beyond Classical Statistical Methods". This name is inspired by Hirotsu (1993), "Beyond Analysis of Variance Techniques: Some Applications in Clinical Trials". Hirotsu reports that his new methods (such as max chi-squared statistics and average chi-squared statistics) are being accepted in Japanese statistical guidelines. One goal of this paper is to present a framework (in section 9) which shows how the statistics introduced by Hirotsu are related to other conventional statistics.

While combining conventional and modern methods has a history of academic development (Daniel (1959), Gnanadesikan (1980)), it may not be much practiced as yet because applied statisticians have a tendency not to use methods which have not been made readily available to them in statistical computing packages. This paper argues that unified methods can impact applied research and statistical education.

The technical content of this paper is the final section which outlines our research about HOW to combine non-parametric quantile and Comparison Change Correlation techniques with classical statistical methods. The first 8 sections discuss from various philosophical viewpoints WHY this research should be on the agenda of statisticians in a society whose health and prosperity is increasingly dependent on statistically literate engineers, scientists, managers, and public.

1. WHY AND HOW COMPONENTS OF STATISTICAL REASONING

I believe that courses and talks on statistics should be about both HOW and WHY.

Academic researchers often minimize the WHY component, because a HOW talk often emphasizes "get to the new material fast without worrying about motivating the results, since to enhance your reputation impress fellow experts in the short attention span that you have available that you've done something new and which works". We say that the HOW component of statistical reasoning is often "esoteric" in the sense that it is specialized technical in a way that appeals mainly to experts.

In contrast, the WHY component of statistical reasoning is intended to be "exoteric" in the sense that it seeks to be understandable to a more general technical audience by motivating WHY the methods are applicable and interpretable.

Presented as a Keynote Address on October 20, 1993 at the 39th Conference on the Design of Experiments for Army Research, Development, and Testing at Rice University. Research supported by the U. S. Army Research Office.

Statisticians need to be concerned with WHY in order to practice in their work the Deming inspired

Continuous Improvement Principle

which states that "every action should be judged by how well it positions you for subsequent actions". Methods should be called simple not by whether their theory is easy but whether their interpretation can be made easy to comprehend.

To enhance their quality (and competitiveness) many organizations are adopting a **Continuous Improvement Process**,

defined as a team approach to Total Quality Management to improve products or services to exceed the expectations of customers or clients. An understanding and implementation of statistical concepts of change, variation, and measurements is clearly important in this process which requires that decisions be based on the information in data, not just on opinions and guesses.

2. CERTIFIED PROFESSIONAL STATISTICIANS

A question of concern to a broad cross-section of applied statisticians, is the question of professional certification of statisticians. Professors should be interested in this question because I believe that it raises fundamental questions about
how to continuously improve statistics courses.

Several ideas that I believe deserve to be in the certification discussion are:

- (1) Is the best role model for professional certification of statisticians an exam structure (similar to that of the Society of Actuaries) which is not a single exam but a series of exams? In this way one can encourage and reward two or more levels of advanced statistical literacy. Statistical culture is understanding that there are several levels of professional statistical literacy, involving different aspects of the practice and theory of statistics.
- (2) Should certification require, in addition to passing exams, a lifelong process of Continuing Education credits? Do we not need to encourage and reward keeping up with the latest developments through short courses and attendance at professional meetings? I call this process "studying the contemporary history of statistics".
- (3) Certification of level of statistical literacy should be the goal of exams in each statistics course. Statistical educators should seek consensus about the content of the series of continually updated statistics courses that would provide excellent education in applied classical and modern methods. The courses should have both HOW and WHY components.
- (4) Statistics programs should have courses that focus on problems of communication and collaboration between statisticians and scientists (how to achieve a collaborative of statistical science).

3. STATISTICAL COMPUTING

An increasingly urgent question is the role in statistics education of
statistical computing and statistical packages,

especially

- (1) how to enable new methods to be quickly made available to applied researchers,
- (2) how to enable methods which are complicated (in theory and computation) to be made simple (in presentation).

A major issue of integrating Statistical Computing into the practice of statistics is: solving the problem that new methods are considered purely academic unless user friendly

software to use them is available.

A major issue of integrating Statistical Computing with statistical education is: how to use statistical packages to implement "alternative" (self learning) classroom cultures that stimulate students to develop statistical reasoning abilities by real life experience which expect students to search for patterns, relate contrasting ideas, and give reasons and arguments for the issues under discussion.

I believe that currently there is a danger in introductory courses that statistical computing is taught only from the HOW point of view with no discussion of WHY issues, especially

how can statistical computing make a statistical method simple,
and the impact of computing (especially graphics) on how statistics is practiced.

4. STATISTICAL EDUCATION

How to change the teaching of statistics is now being frequently discussed at statistics meetings; *Amstat News* for October 1993 (p. 13) reports the revolutionary views of David Moore that we need "new" teaching styles.

The goals of an "alternative" educational philosophy should be to emphasize both practice and theory using two teaching strategies:

1. "Never tell students what they can find out for themselves."
2. "Tell students about those things which they will find most difficult to learn by themselves."

Other goals for introductory statistics courses:

- (1) public respect for statisticians,
- (2) the recruitment of statisticians,
- (3) public statistical literacy, awareness that in every activity one should strive to compare "expectation" with "reality".

I recommend that courses discuss:

the "map" of statistics (its relations to other disciplines as the 'glue' of science);
the "contemporary history of statistics" (emphasizing that innovation in methods and applications are constantly occurring);
its culture (why statisticians are oriented to "continuous improvement" and how they keep up with new "hammers" (methods) and "nails" (applications)).

We must be pro-active in changing the current attitude among undergraduate students that statistics is a required and irrelevant course, to be remembered as little as possible.

5. DEFINING THE PROBLEM OF STATISTICS AS PROBABILITY MODELING

Defining what statistical science is about has always been regarded as a controversial act (many statisticians reject the hypothesis that one can be certain about the study of uncertainty). We should be aware of the various definitions of statistics:

- (1) help find scientific truth about probabilities and the fit of observations to theory;
- (2) make decisions in the face of uncertainty and loss functions;
- (3) model uncertain data by probability models.

I believe that to find truth (and make decisions) one must explore the widest range of alternatives (what I call "going to the edge"). I regard as most operational the following definition:

The most important concepts in statistics are the probability model and

likelihood; statistical thinking combines data analysis and concepts of probability.

Introductory statistics courses may not need to include techniques for theoretically computing probabilities but need to stress that probabilities are what statistics is computing from data.

Guided by the proverb "if your only tool is a hammer, every problem looks like a nail". I proposed (Parzen (1993)) that the practice of statistics can be regarded as combining "nails" (fields of applications) and "hammers" (general methods stated mathematically which are combined to provide a custom made method for each application, not just reducing each problem to fit the "simple" techniques the statistician knows).

Current important applications of statistics can be defined as analyzing change (observing and measuring changes taking place in society, industrial processes, medical treatments, the environment, economic indicators, etc.).

Current methods of statistics can be regarded as having a common theme: use probability models to model and comprehend populations and data, by an iterative process of model specification, parameter estimation, and model checking (eloquently described by George Box and Gwilym Jenkins in the context of Time Series Analysis). That applied statistics is best practiced by modeling is well described in an article in the September 1993 *American Scientist* by Gauch.

Bayesians (of the dogmatic type who preach that priors are not just techniques but are to be believed) imply that statisticians should never use non-Bayesian methods (one should not analyse data for which one does not have prior beliefs about the model). Modeling statisticians believe that data can yield patterns and models which provide insights which were not thought of before the data analysis. The magazine "The Economist" (October 9, 1993 issue) states that principles of "data analysis without theory" are the basis of current successful applied research on the mathematics of finance (investing).

6. STATISTICAL EDUCATION ANALOGUES TO STATISTICAL MODELING

Strategies for solving statistical problems are emphasized in the "new" teaching which aims to give students a sense of purpose and direction to their statistical learning. My major point is that reforms in statistical education and research are linked, because statistical learning and statistical investigation are analogous, because both require a cycle of model building, which one usually repeats (iterates) several times before reaching a satisfactory conclusion.

The SIET cycle of statistical model building consists of four stages:

- Stage 1 (S): *Specify* very general class of models.
- Stage 2 (I): *Identify* tentative parametric model.
- Stage 3 (E): *Estimate* parameters of tentative model.
- Stage 4 (T): *Test* goodness of fit, diagnose improved models.

(The slogan could be: "To SIET (see it) is to understand it.")

The cycle of statistical problem solving consists of four stages:

- Stage 1 (P): *Pose* the question, form expectations.
- Stage 2 (C): *Collect* the data, make observations.
- Stage 3 (A): *Analyze* the data, compare observations and expectation.
- Stage 4 (I): *Interpret* the results, find the best theory or decision that fits the data.

The PCAI cycle of a statistical investigation should be represented in a diagram as a circular process (rather than a linear process); see figure from p. 183 A. Graham (1993).

We prefer to call the cycle EOCl (Expect, Observe, Compare, Interpret).

Reformers of mathematics education take the view that teachers should communicate the four aspects of learning which cognitive sciences recommend for success:

1. simple recall,
2. algorithmic learning,
3. conceptual learning, and
4. problem solving strategies.

In statistical teaching we can make these cognitive concepts more concrete by teaching that statistical concepts (such as the sample mean or sample variance) have three aspects:

1. how to define it (mean of sample distribution);
2. how to compute it (average the values or the quantile function);
3. how to interpret it (estimate location parameter of sample);

The fourth aspect of statistical learning consists of ideas about combining concepts to conduct an iterative statistical investigation whose output is data models which can be applied.

When one is a discussant of a technical paper it may be helpful to use the four aspects of learning as a basis for evaluation.

One reason the definition

"the methods of statistics are modeling"

may be controversial among statisticians is because many introductory statistics courses adopt approaches which avoid the use of concepts of probability.

Statistics and probability need to be linked not only to define probability models but in order to make judgements (and simulations) about how to interpret the significance of a set of results, to explain that unusual results do sometimes occur just by chance.

Professors of education report that the dilemma of mathematics education reform is that it requires teachers to have a deeper understanding of ideas and concepts, which they are reluctant to study. Teachers prefer "ready to apply" modules rather than professional development. How can we overcome these inhibitions to mathematics and statistics educational reform?

7. FUNCTION REPRESENTATIONS OF DATA AND MATHEMATICAL LITERACY

The philosophy of "Beyond Classical Statistical Methods" proposes that to practice statistics, one must be aware of the relations between statistics and computing, between statistics and probability, and between statistics and mathematics.

Early childhood study of statistical data analysis and probability is now regarded as critical to developing mathematically literate students who can function in a society driven by technology. Current mathematics educational reform movements believe experience (with statistical data analysis) is the ideal way to teach and reinforce mathematical concepts; I propose that statistics can benefit from mathematical tools (such as representations of data by functions).

My Comparison Change Correlation statistical methods emphasize innovations in functions that can be used to describe probability relationships and the "shape" of data. These functions are defined on the unit interval (denoted $[0,1]$ or $0 < u < 1$ or $0 < t < 1$) and the unit square (denoted $0 < t, u < 1$); they can be plotted and interpreted by their shapes, as well as their numerical magnitudes, and yield functional statistics.

"Safe" (best) statistical methods provide two hypotheses between which the researcher must choose. In the Comparison Change Correlation approach, the null hypothesis of no relationship is formulated as implying

data representation on $[0,1]$ =white noise

while the alternative hypothesis implies

data representation on $[0,1] = \text{signal} + \text{white noise}$.

Test statistics are "linear detectors" of the form integral over $[0,1]$ of the product of the data representing function and the signal representing function. Quadratic detectors are sums of squares of linear detectors. Information theory detectors are entropy measures of comparison density estimators.

Factoid: The concept of null hypothesis was introduced by R. A. Fisher as a hypothesis set up for the purpose of being nullified (invalidated). Source: Fisher (1990), p. 322.

8. THE P VALUE PROBLEMS OF STATISTICS

Statistics has as its goals specification and identification of models that fit data, and assigning "p values" to models selected by multiple comparisons. If we use modeling methods to decide which of two treatments is better the client wants and expects a p value for our conclusion! Answering such distributional questions may be feasible using computer intensive re-sampling methods which can generate the distribution of the statistics that we propose to test relationships.

I would like to tell you a true story that happened to me in Israel in September 1993 on a bus to the Weizmann Institute. When a statistician meets a scientist, one often gets the reaction:

"All scientists need statisticians (good news).
But we do not need them very much (bad news).
How complicated is it to compute a p value?"

Revising this attitude requires a public relations campaign to educate the scientific public about "Beyond Classical Statistical Methods."

9. HOW TO USE CORRELATION COEFFICIENTS TO DEVELOP BEYOND STATISTICAL METHODS

This section is a technical outline, without examples, of Comparison Change Correlation statistical methods, emphasizing HOW conventional statistical methods can be expressed in terms of diverse correlation coefficients.

We start with the multi-sample problem that we reformulate as data analysis of bivariate (X, Y) . Multi-sample statistical data analysis arises when observe a variable Y in c cases or samples (corresponding to c treatments or c populations). The samples are usually regarded as the value of c variables Y_1, \dots, Y_c with respective true distribution functions $F_1(y), \dots, F_c(y)$ and quantile functions $Q_1(u), \dots, Q_c(u)$. The general problem is to model how the distribution functions F_k vary with the value of the conditioning variable $k = 1, \dots, c$, and in particular to test the hypothesis of homogeneity of distributions:

$$H_0 : F_1 = \dots = F_c = F$$

The distribution F to which all the others are equal under H_0 is considered to be the unconditional distribution of Y (which is estimated by the sample distribution of Y in the pooled sample).

For $k = 1, \dots, c$, we observe a random sample $Y_k(j), j = 1, \dots, n_k$ for $k = 1, \dots, c$. The pooled sample, of size $n = n_1 + \dots + n_c$, represents observations of the pooled (or unconditional) variable Y . The c samples are assumed to be independent of each other.

We propose that we regard the data as consisting of bivariate observations (X, Y) , where X represents the population $k = 1, \dots, c$ observed and Y the response observed. The

observation that is usually denoted $Y_k(j)$ is denoted in our notation ($X = k, Y = Y_k(j)$). While X is a deterministic variable rather than a random variable, the probability notation we use can be interpreted for both cases. The marginal (unconditional) distribution of X is specified by the probability mass function

$$p_X(k) = n_k/n.$$

The distribution function of X is defined

$$F_X(x) = \sum_{k \leq x} p_X(k).$$

Define the indicator function $I(B)$ of an event B to be 1 or 0 according as the event B did or did not occur. Thus $I(X = k)$ denotes the indicator function of the event $X = k$, which equals 1 or 0 according as $X = k$ or $X \neq k$. $I(Y \leq y)$ denotes the indicator function of the event $Y \leq y$. The distribution function of the values of Y in the k -th sample, previously denoted F_k , is now described in the notation of conditional distributions of Y given X :

$$F_k(y) = F_{Y|X=k}(y) = E[I(Y \leq y)|I(X = k)]$$

We henceforth use empirical distributions (based on the observed data) rather than theoretical distributions (based on the unobserved population). Then

$$\begin{aligned} F_{Y|X=k}(y) &= E[I(Y \leq y)|X = k] \\ &= (1/n_k) \sum_{\text{observations } (X,Y)} I(X = k)I(Y \leq y) \\ &= E[I(X = k)I(Y \leq y)]/p_X(k). \end{aligned}$$

An important general formula: for function $g(Y)$ and set B of real numbers

$$E[g(Y)|X \text{ is in } B] = E[g(Y)I(X \text{ is in } B)]/P[X \text{ is in } B]$$

An important general concept is correlation coefficient. We now show that correlation can be used to describe a statistic that is a conditional mean:

$$\begin{aligned} R(X \text{ is in } B, g(Y)) &= CORR[I(X \text{ is in } B), g(Y)] \\ &= E[I(X \text{ is in } B)(g(Y) - E[g(Y)])]/\sigma[g(Y)]/\sigma[I(X \text{ is in } B)] \\ &= (\text{odds } P[X \text{ is in } B])^{\frac{1}{2}} E[(g(Y) - E[g(Y)])/\sigma[g(Y)]|X \text{ is in } B] \end{aligned}$$

where we define $\text{odds}(p) = p/(1 - p)$. Note that $P[X \text{ is in } B]/\sigma[I(X \text{ is in } B)] = (\text{odds } P[X \text{ is in } B])^{\frac{1}{2}}$.

The pooled sample has unconditional empirical distribution

$$F_Y(y) = (1/n) \sum_{\text{observations } (X,Y)} I(Y \leq y)$$

The empirical quantile function of Y is denoted $Q_Y(u)$, $0 < u < 1$, and is piecewise constant between points u satisfying $F_Y(Q_Y(u)) = u$, called exact values of u ; exact values u are of the form $u = F_Y(y)$ for some y . The quantile function of X is denoted $Q_X(t)$,

$0 < t < 1$, and is piecewise constant between points t satisfying $F_X(Q_X(t)) = t$, called exact values of t ; exact values of t are of the form $t = F_X(x)$ for some x .

To test the null hypothesis H_0 , three main methods are proposed in introductory statistics courses, based on comparing (1) means, (2) scored ranks, (3) distribution functions. We propose to unify these methods by expressing the test statistics in terms of basic types of indicator correlations:

- (1) $R(X = Q_X(t), Y) = \text{CORR}[I(X = Q_X(t))), Y], 0 < t < 1;$
- (2) $R(X = Q_X(t), \text{scored ranks of } Y) = \text{CORR}[I(X = Q_X(t)), \text{scored ranks of } Y], 0 < t < 1;$
- (3) $R(X = Q_X(t), Y \leq Q_Y(u)) = \text{CORR}[I(X = Q_X(t)), I(Y \leq Q_Y(u))], 0 < t, u < 1.$

Additional statistics to be investigated for multi-sample problems are accumulation correlations:

- (4) $R(X \leq Q_X(t), Y) = \text{CORR}[I(X \leq Q_X(t)), Y], 0 < t < 1;$
- (5) $R(X \leq Q_X(t), \text{scored ranks of } Y) = \text{CORR}[I(X \leq Q_X(t)), \text{scored ranks of } Y], 0 < t < 1;$
- (6) $R(X \leq Q_X(t), Y \leq Q_Y(u)) = \text{CORR}[I(X \leq Q_X(t)), I(Y \leq Q_Y(u))], 0 < t, u < 1.$

To motivate how the indicator correlations (1) arise in the Analysis of Variance we introduce the following notation. The sample mean \bar{Y}_k of the k -th sample is the conditional mean of Y given $X = k$:

$$E[Y|X = k] = \bar{Y}_k = (1/n_k) \sum_{j=1}^{n_k} Y_k(j) = p_X(k) E[Y I(X = k)]$$

The pooled sample mean is the unconditional mean of Y :

$$\bar{Y} = E[Y] = \sum_{k=1}^c p_X(k) E[Y|X = k] = \sum_{k=1}^c p_X(k) \bar{Y}_k.$$

The unconditional variance of Y , and the conditional variance of Y given $X = k$, are respectively denoted

$$\begin{aligned} \text{VAR}[Y] &= \sigma^2[Y] = \sum_{k=1}^c \sum_{j=1}^{n_k} (Y_k(j) - \bar{Y})^2 / n, \\ \text{VAR}[Y|X = k] &= \sum_{j=1}^{n_k} (Y_k(j) - \bar{Y}_k)^2 / n_k. \end{aligned}$$

The common variance σ^2 of Y under H_0 is estimated by the pooled variance

$$\hat{\sigma}^2 = E[\text{VAR}[Y|X]] = \sum_{k=1}^c p_X(k) \text{VAR}[Y|X = k]$$

Define the multiple correlation

$$R^2[Y|X] = \text{VAR}[E[Y|X]] / \text{VAR}[Y]$$

What may be novel is the observation that one can write

$$R^2[Y|X] = \sum_{k=1}^c (1 - p_X(k)) (R(X = k, Y))^2.$$

From the important representation

$$\text{VAR}[Y] = E[\text{VAR}[Y|X]] + \text{VAR}[E[Y|X]],$$

infer that the pooled variance σ^2 can be shown to be related to the original variance $\text{VAR}[Y]$ by

$$\sigma^2 = \text{VAR}[Y](1 - R^2[Y|X])$$

The F statistic used in the Analysis of Variance to test H_0 can be shown to be $(n - c)T^2/(c - 1)$, defining

$$\begin{aligned} T^2 &= R^2[Y|X]/(1 - R^2[Y|X]) \\ &= \sum_{k=1}^c (1 - p_X(k)) T^2(X = k), \end{aligned}$$

defining

$$T(X = k) = R(X = k, Y)/(1 - R^2[Y|X])^{.5} = (\text{odds } p_X(k))^{.5} (Y^-_k - Y^-)/\sigma.$$

A plot of $(n - c)^{.5} T(X = Q_X(t))$, $0 < t < 1$, can help determine which sub-samples are most different from the others. Note $(n - c)^{.5} T(X = k)$ has a Student- t distribution with $n - c$ degrees of freedom under H_0 and normality, while $(n - c)T^2/(c - 1)$ has $F(c - 1, n - c)$ distribution.

The foregoing discussion has outlined how a conventional statistical method (one way Analysis of Variance) can be expressed in terms of correlations. We next state results for expressing other conventional and beyond methods in terms of correlations.

$R(X = x, Y = y)$, Contingency Table Analysis

The chi-squared statistic Chi used to test independence in a contingency table of n observations (X, Y) where X has c possible values and Y has r possible values can be expressed $\text{Chi} = nC(X, Y)$, in terms of a probability concept

$$C(X, Y) = \sum_{x=1}^c \sum_{y=1}^r (p_{X,Y}(x, y) - p_X(x)p_Y(y))^2 / p_X(x)p_Y(y),$$

expressed in terms of (empirical) probabilities. We propose to interpret this formula in terms of indicator correlations

$$R(X = x, Y = y) = (p_{X,Y}(x, y) - p_X(x)p_Y(y)) / (p_X(x)p_Y(y)(1 - p_X(x))(1 - p_Y(y)))^{.5};$$

then

$$C(X, Y) = \sum_{x=1}^c \sum_{y=1}^r (1 - p_X(x))(1 - p_Y(y)) |R(X = x, Y = y)|^2$$

To study the independence of X and Y given data on (X, Y) we propose a "Chi-square and Indicator Correlation Tableau", consisting of the r by c matrix $n^{\frac{1}{2}}R(X = x, Y = y)$ and bordering rows and columns $nC(x, .)$, $nC(., y)$, $nCav$, defining

$$C(x, .) = (r - 1)^{-1} \sum_y (1 - p_Y(y)) |R(X = x, Y = y)|^2$$

$$C(., y) = (c - 1)^{-1} \sum_x (1 - p_X(x)) |R(X = x, Y = y)|^2$$

$$Cav = (c - 1)^{-1} \sum_x (1 - p_X(x)) C(x, .) = (r - 1)^{-1} \sum_y (1 - p_Y(y)) C(., y)$$

We assign p value to these statistics as tests of the null hypothesis H_0 by using their known asymptotic or exact distributions under the null hypothesis. The use in practice of these statistics is best illustrated by examples which require their own paper to discuss.

Rather than a table of $R(X = x, Y = y)$ we prefer a graphical presentation of

$$n^{\frac{1}{2}}R(X = Q_X(t), Y = Q_Y(u))$$

as either a function on $0 < u < 1$ for each exact t fixed, or as a function on $0 < t < 1$ for each exact u fixed. We also plot $nC(Q_X(t), .)$, $nC(., Q_Y(u))$.

The chi-squared statistic C is a portmanteau or omnibus statistic. When it is significant we want to know the cause of the rejection of independence, the nature of the dependence, which can be obtained from the above plots which show which coefficients are most significant.

Comparison Analysis

The ultimate approach to modelling is to estimate and interpret comparison density $d(u|t)$ and comparison distribution $D(u|t)$:

$$d(u|t) = d(u; F_Y, F_{Y|X=Q_X(t)}), 0 < u < 1;$$

$$D(u|t) = \int_0^u d(u'|t) du' = D(u; F_Y, F_{Y|X=Q_X(t)})$$

If u and t are exact values in the sense that they satisfy $u = F_Y(y)$, $t = F_X(x)$ for some y and x , one can show that

$$D(u|t) = F_{Y|X=Q_X(t)}(Q_Y(u))$$

The joint dependence density $d(t, u)$ is defined as a comparison density

$$d(t, u) = d(u|t) = d(t|u) = d(t; F_X, F_{X|Y=Q_Y(u)}), 0 < t < 1;$$

The joint dependence distribution or copula function is defined by

$$D(t, u) = \int_0^t \int_0^u d(t', u') dt' du' = \int_0^t D(u|t') dt'.$$

The changePP process is defined on $0 < u < 1$ for fixed exact t by

$$CPP(u|t) = \int_0^u cPP(u'|t)du' = (\text{odds}(p_X Q_X(t)))^{-\delta} (D(u|t) - u)$$

The change distribution is defined on $0 < t < 1$ for fixed exact u by

$$D(t|u) = \int_0^t d(t'|u)dt' = \int_0^t d(t', u)dt'.$$

$R(X = x, Y \leq y)$, Multi-Sample Comparison, Accumulation Analysis
 $R(X \leq x, Y = y)$, Change Analysis of a Response

The chi-square statistic, based on correlations $R(X = x, Y = y)$, is most appropriate to compute when X is discrete and Y is discrete. Alternative correlations for diagnosis of the dependence of discrete X and discrete Y , and essential correlations when one variable is continuous and the other is discrete, are the accumulation correlation coefficients

$$R(X = x, Y \leq y) = (\text{odds } p_X(x)\text{odds } F_Y(y))^{-\delta} ((F_{Y|X=x}(y)/F_Y(y)) - 1)$$

At exact u

$$R(X = Q_X(t), Y \leq Q_Y(u)) = (\text{odds } p_X(Q_X(t)))^{-\delta} (D(u; F_Y, F_{Y|X=Q_X(t)}) - u)/(u(1-u))^{\delta}$$

We could plot for each exact t a change PP process $CPP(u|t)$, $0 < u < 1$, which compares conditional and unconditional distributions and is asymptotically a Brownian Bridge under H_0 :

$$CPP(u|t) = (\text{odds } p_X(Q_X(t)))^{-\delta} (D(u; F_Y, F_{Y|X=Q_X(t)}) - u)$$

We always plot for each exact t change test process which is a collection of accumulation correlation coefficients

$$CT(u|t) = CPP(u|t)/(u(1-u))^{\delta} = R(X = Q_X(t), Y \leq Q_Y(u))$$

Recall that the set of exact u values consists of $u = F_Y(y)$, $y = 1, \dots, r - 1$. The (Hirotsu) maximum chi-square statistic is defined for each treatment t (more precisely, treatment x with $t = F_X(x)$)

$$R^2 \text{accummax}(t) = n \max_{\text{exact } u} |R(X = Q_X(t), Y \leq Q_Y(u))|^2$$

$$R^2 \text{accumave}(t) = n \sum_{\text{exact } u} |R(X = Q_X(t), Y \leq Q_Y(u))|^2 / (r - 1)$$

By introducing weights $W(u)$, such as $W(u) = (u(1-u))$, one can define weighted Hirotsu statistics:

$$R^2 \text{accummax}W(t) = n \max_{\text{exact } u} W(u) |CT(u|t)|^2,$$

$$R^2 \text{accumave}W(t) = n \sum_{\text{exact } u} W(u) |CT(u|t)|^2 / (r - 1)$$

For contingency tables (X discrete, Y discrete) these statistics provide alternatives to the standard Chi-squared statistic to test for independence.

For multi-sample problems (X discrete, Y continuous) they provide goodness of fit type statistics for testing homogeneity of populations. Modeling rather than testing is provided by density estimation techniques which estimate $cPP(u|t)$, the change PP density or derivative of the Change PP process.

One of the accomplishments of our research is to relate the accumulation statistics introduced by Hirotsu (1993) to conventional statistical methods.

$R(X = x, Y)$, Two sample and Multi-sample tests of homogeneity, Analysis of Variance (One Way)

The output of the $R(X = x, Y)$ command is a plot of the change test density $R(X = Q_X(t), Y)$, $0 < t < 1$, and the values of the conventional F test statistics of the Analysis of Variance.

$R(X \leq x, Y)$, Change Analysis of Multi-Samples

Plot $R(X \leq Q_X(t), Y)$, $0 < t < 1$, and the corresponding max and ave statistics.

$R(X = x, \text{ranks})$, Non-parametric tests, Wilcoxon, Kruskal-Wallis

$R(X \leq x, \text{ranks})$, Non parametric Change analysis of multisamples

We define ranks to be a transformation of Y to $P_Y(Y)$, where $P_Y(y)$ is the mid distribution function

$$P_Y(y) = F_Y(y) - .5p_Y(y)$$

$R(X = x, \text{scored ranks})$

$R(X \leq x, \text{scored ranks})$

Scored ranks are a transformation of Y to $J(P(Y))$, where $J(u)$ is a score function, often chosen to be a Legendre orthogonal polynomial. Their correlations can be used to guide estimation of comparison densities.

$R(\text{scored ranks}, \text{transformation of } Y)$. Change analysis of data Y transformed by one of the transformations $I(Y = y)$, $I(Y \leq y)$, Y , $P_Y(Y)$, $J(P_Y(Y))$, and guide to estimation of change density (non-parametric regression).

$R((X, Y), R(P_X(X), P_Y(Y)), R(X, P_Y(Y)), R(P_X(X), Y), \text{Correlations and non-parametric Spearman correlations}$

Compute the correlations and plot the functions of which they are diagnostics. The important formula

$$R(X, Y) = \int_0^1 (Q_X(t) - X^-)/\sigma[X]) E[(Y - Y^-)/\sigma[Y]) | X = Q_X(t)] dt$$

suggests that we plot the two functions on $0 < t < 1$ that are in the integrand. Smoothing the second function, called the change density, is the problem considered in non-parametric regression.

BIBLIOGRAPHY

Daniel, Cuthbert. (1959). Use of half-normal plots in interpreting factorial two level experiments. *Technometrics*, 1, 311-341.

Fisher, R. A. (1990). *Statistical Inference and Analysis: Selected Correspondence of R. A. Fisher*, ed. by J. H. Bennett, Oxford.

- Gnanadesikan, R. (1980). Graphical Methods for Internal Comparisons in ANOVA and MANOVA, *Handbook of Statistics*, Vol. 1, 133-177.
- Graham, A. (1993). Handling Data to a Purpose, in *Innovation in Maths Education by Modeling and Applications*, ed. Jan de Lange, Christine Keital, Ian Huntley, Mogen Niss, Ellis Horwood: London.
- Hirotsu, Chihiro. (1993). Beyond Analysis of Variance Techniques: Some Applications in Clinical Trials, *International Statistical Review*, 61, pp. 183-201.
- Nair, V. N. (1987). Chi-squared-type tests for ordered alternatives in contingency tables. *J. Amer. Statist. Assoc.* 82, 283-291.
- Parzen, Emanuel. (1979). Nonparametric Statistical Data Modeling, *J. Ameri. Statistic. Assoc.*, (with discussion), 74, 105-131.
- Parzen, Emanuel. (1989). Multil-Sample Functional Statistical Data Analysis, *Statistical Data Analysis and Inference*, ed. Y. Dodge, Elsevier, Amsterdam: 71-84.
- Parzen, Emanuel. (1992). Comparison Change Analysis. *Nonparametric Statistics and Related Topics* (ed. A. K. Saleh), Elsevier: Amsterdam, 3-15.
- Parzen, Emanuel. (1993). Comparison Change Analysis Approach to Change-Point Estimation, *Applied Change-Point Analysis Symposium, Journal of Applied Statistical Science*.
- Parzen, Emanuel. (1993). History of Statistics in Real Time: Hammers and Nails, *Computer Science-Statistics, Interface Foundation*, 24, 602-608.

PROPERTIES OF SIMULATION BASED ESTIMATORS OF STOCHASTIC PROCESSES

Katherine B. Ensor
Department of Statistics
Rice University
Houston, TX 77251-1892

e-mail: ensor@stat.rice.edu

Abstract: Simulation based methods of estimation have proven a useful tool for parameter estimation of complicated stochastic processes. We examine a simulation based estimation procedure comparable to the conditional least squares estimates for parameters of a stochastic process. The simulated conditional least squares estimates are shown to be consistent and the asymptotic distribution is derived.

KEY WORDS: SIMEST, Conditional Least Squares

1. INTRODUCTION

Ensor, Bridges & Lawera (1993) illustrated the viability of estimation for stochastic processes through simulation. Their method built on the original work by Thompson, Brown and Atkinson (1987) in this area. The premise of such estimators is that a process can be simulated directly from the defining axioms. Parameter estimates are then obtained by minimizing over the parameter space some measure of error between the simulated process at a given point in the parameter space and the observed series. The acronym SIMEST, for simulation based estimation, is used to describe this general method.

This method of estimation has been successfully applied in the area of marketing by Bridges, Ensor and Thompson (1992). They model the number of "types" of personal computers in the marketplace at time t . A personal computer is considered a new type if something about the technology changed, for example the 486 chip replacing the 386 chip. The proposed stochastic model was not solvable in closed form. Using SIMEST they were able to use the proposed stochastic model rather than resulting to the simplifying assumption of a deterministic model plus random noise. Another example in marketing is presented by Bridges, Ensor and Raman (1994). They model the number of customers for a particular home inspection firm in the Los Angeles area as a birth and death process with constant death rate and a birth rate which is a function of advertising expenditures.

The proximity measure minimized determines the type of estimators found via SIMEST. In this paper, simulation based estimators comparable to conditional least squares estimators for stochastic processes will be examined.

Let $\{N(t), t \geq 0\}$ denote the stochastic process of interest which is observed at n different time points, i.e. $N(t_1), \dots, N(t_n)$. For simplicity in notation, we refer to the observed process at time points t_1, \dots, t_n as Y_1, \dots, Y_n . If one can simulate the $E[Y_i]$ or $E[Y_i | Y_{i-1}]$ for $i = 1, \dots, n$ in theory a SIMEST estimator can be obtained. As an example of the use of SIMEST in this setting consider a general birth and death process.

1.1. Simulation of Birth and Death Processes

Consider the Markov counting process $N(t)$ with parameters λ_n and μ_n which satisfies the following axioms:

- i) $P(N(t + \delta t) = n + 1 | N(t) = n) = \lambda_n \delta t + o(\delta t)$
- ii) $P(N(t + \delta t) = n - 1 | N(t) = n) = \mu_n \delta t + o(\delta t)$
- iii) The probability of more than one event in $(t, t + \delta t] = o(\delta t)$.

From the above axioms it is simple to derive the distribution of the time of the next arrival, $F_B(t)$ and the distribution of the time of the next exit from the system, $F_D(t)$ so that

$$F_B(t) = 1 - P\{0 \text{ births in } (t, t + \delta t]\} = 1 - e^{-\lambda_n t}$$

and

$$F_D(t) = 1 - P\{0 \text{ deaths in } (t, t + \delta t]\} = 1 - e^{-\mu_n t}.$$

Using the inverse c.d.f. transformation we obtain obtain the time until the next birth, t_B , or death, t_D , in our process from

$$t_B = \frac{-\log(U_1)}{\lambda_n} \quad \text{or} \quad t_D = \frac{-\log(U_2)}{\mu_n}, \quad (1)$$

where U_1 and U_2 represent independent random variables from the uniform distribution defined over the unit interval. It is then a simple matter to simulate the conditional mean of Y_i given the observed value of Y_{i-1} as the following algorithm illustrates. Let $X_{i,j}(\theta)$ denote the j^{th} simulated value of the process at time t_i given the observed value at t_{i-1} or Y_{i-1} as the starting point of the simulation. The process is simulated assuming parameter θ . Also, let $\bar{X}_{i,m}(\theta) = (1/m) \sum_{j=1}^m X_{i,j}(\theta)$. In other words, $\bar{X}_{i,m}(\theta)$ is the simulated conditional mean based on m realizations of the process at time t_i given the value of the process at time t_{i-1} .

Simple Algorithm to Simulate $\bar{X}_{i,m}(\theta)$

-
- | | |
|----|---|
| 1. | Set $k = Y_{i-1}$. |
| 2. | Simulate U_1 and U_2 from $U(0,1)$ distribution. |
| 3. | Compute t_D and t_B . |
| 4. | Set $t = t + \min(t_B, t_D)$. |
| 5. | If $t_D < t_B$ then $k = k - 1$ else $k = k + 1$. |
| 6. | If $t < t_i - t_{i-1}$ go to 2 otherwise $X_{i,j}(\theta) = k$. |
| 7. | Repeat 1-6 m times. Average |
| | $X_{i,1}(\theta), \dots, X_{i,m}(\theta)$ to obtain $\bar{X}_{i,m}(\theta)$. |
| 8. | Move to time $i + 1$, go to 1. |
-

An important consideration is that the computation of t_D and t_B depends on the parameter values θ .

2. SIMULATED CONDITIONAL LEAST SQUARES

ESTIMATES OF θ

An often used alternative estimator to the maximum likelihood estimators for stochastic processes is the conditional least squares estimators discussed by Klimko & Nelson (1978) (see also Hall & Heyde (1980)). The conditional least squares estimator, $\tilde{\theta}$, is the value of θ minimizing the conditional least squares equation

$$Q_{n,m}(\theta) = \sum_{i=1}^n (y_i - \mu_i(\theta))^2$$

over the parameter space Θ , where $\mu_i(\theta) = E_\theta[Y_i | Y_1, \dots, Y_{i-1}]$.

To obtain the simulated conditional least squares estimator, $\hat{\theta}_{nm}$, the simulated conditional mean replaces the conditional mean in the above equation. In other words, the SIMEST estimator based on the conditional least squares equation is the value $\hat{\theta}$ which minimizes $S_{n,m}(\theta)$ over the parameter space Θ where

$$S_{n,m}(\theta) = \sum_{i=1}^n (y_i - \bar{X}_{i,m}(\theta))^2$$

and $\bar{X}_{i,m}(\theta)$ is defined in the previous section.

Since $\bar{X}_{i,m}(\theta)$ is the average of m i.i.d. random variables with expectation $\mu_i(\theta)$ as m , the number of simulations, approaches infinity $\bar{X}_{i,m}(\theta) \xrightarrow{a.s.} \mu_i(\theta)$. Hence, the simulated conditional least squares estimator maintains the same properties as the conditional least squares estimator for large m .

2.1 Properties of the Simulated Conditional Least Squares Estimator

Under certain regularity conditions, Klimko and Nelson (1978) show that $\tilde{\theta}$ exists, is a strongly consistent estimator of θ , and is asymptotically normally distributed. Specifically,

$$n^{1/2}(\tilde{\theta} - \theta) \xrightarrow{d} \text{MVN}_k(0, V^{-1}WV^{-1})$$

where

$$V_{k \times k} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=2}^n \mathbf{g}_i \mathbf{g}_i^T$$

and \mathbf{g}_i is a $k \times 1$ vector representing the derivative of the conditional mean $\mu_i(\theta)$ with respect to the parameter vector θ . Also,

$$W = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=2}^n \sigma_i^2(\theta) \mathbf{g}_i \mathbf{g}_i^T.$$

As the number of simulations, m , goes to infinity, $\hat{\theta}_{nm}$ has the same asymptotic properties as $\tilde{\theta}$. It can be shown that for large fixed m ,

$$n^{1/2}(\hat{\theta}_{nm} - \theta)$$

is approximately distributed as a Multivariate Normal random vector of dimension k with 0 mean and covariance matrix

$$V^{-1}(I + \frac{1}{m^2} I)WV^{-1}.$$

The regularity conditions of Klimko and Nelson (1978) must be met for the above results on the simulated conditional least squares estimator to hold. It is important to note, however, that in the SIMEST situation often the regularity conditions can only be checked empirically through simulation since transition probabilities are never explicitly stated. For birth and death processes with a limit on the population size, the regularity conditions are met if the birth rate is greater than the death rate. If the regularity conditions are not met then multiple realizations of the process must be observed before one can estimate the parameters. If multiple realizations are observed, SIMEST estimators can still be obtained.

2.2 Weighted Simulated Conditional Least Squares Estimates

At each stage of our optimization we can easily compute a consistent estimator of $\sigma_i^2(\theta)$ by computing the sample variance of $X_{i,1}(\theta), \dots, X_{i,m}(\theta)$. Therefore, it is a simple matter to find the weighted simulated conditional least squares estimator by minimizing

$$S_{n,m}^*(\theta) = \sum_{i=1}^n (1/\hat{\sigma}_{i,m}^2(\theta))[(y_i - \bar{X}_{i,m}(\theta))]^2$$

where $\hat{\sigma}_{i,m}^2(\theta) = \frac{1}{m-1} \sum_{j=1}^m (X_{i,j}(\theta) - \bar{X}_{i,m}(\theta))^2$. For large fixed m , the resulting estimator $\hat{\theta}_w$ is approximately normally distributed with mean vector 0 and covariance matrix

$$(1 - \frac{1}{m^2})V^{-1}.$$

In practice V is obtained by estimating the gradients via central differences using a large number of simulations, then computing

$$V_n = \frac{1}{n} \sum_{i=2}^n \hat{g}_i \hat{g}_i^T.$$

However, using the method of Glynn (1990) in conjunction with a large number of simulations leads to efficient estimation of the gradients, thereby yielding the optimal variance estimate.

3. DOES THIS METHOD WORK IN PRACTICE?

A Modest Simulation Study and an Example

Extensive simulation studies were conducted by Ensor, Bridges and Lawera (1993). Their simulation studies focused on simulated least square estimates instead of conditional least square estimates but clearly indicated the utility of the SIMEST procedure. To investigate the usefulness of the simulated conditional least squares estimation procedure, this method of estimation was repeated numerous times and summary statistics of the replicated estimates obtained. The model used consisted of a linear death rate $\mu_n = \mu n$ (with $\mu = .1$) and birth rate $\lambda_n = (1000 - n)n\lambda$ (with $\lambda = .1$). The Nelder-Mead (1965) optimization routine was used. One thousand replicates of the simulated weighted conditional least squares estimation described in Section 2.2 with $m = 500$, resulted in a mean of .0978 with standard deviation .006261 for the parameter $\lambda = .1$ and a mean of .1047 with a standard deviation of .03604 for the parameter $\mu = .1$. The average of $S_{n,m}^*(\theta)$ for the 1,000 replications

was 12.81 with a standard deviation of 4.00. Comparable results were obtained for the simulated conditional least squares estimate based on $m = 500$ and $m = 2500$. For this particular model, the conditional variance at each time point is relatively constant hence the weighted conditional least squares estimate does not provide significant improvement. Often this will not be the case.

In addition to repeated replications of the various estimators, we examined the asymptotic properties for one realization. Estimating the gradient for the covariance matrix via central differences based on 10,000 simulated values we obtained a standard error of .0103 for the parameter estimate of λ which for this realization was .09068 and a standard error of .0517 for the parameter estimate of μ which was .0893 for this realization. The correlation between the two estimates was -.14. Again, we note that better estimates of the covariance matrix can be obtained using the method of Glynn (1990).

As mentioned in the introduction, Bridges, Ensor and Raman (1994) model the number of customers for a particular home inspection firm in the Los Angeles area as a birth and death process. The data consists of annual observations of the number of customers and information on both direct and indirect advertising costs for the first 13 years of the company's existence. Direct advertising consists of such costs as yellow pages, brochures, etc. Indirect advertising primarily consists of the cost of networking with the real estate agents in the area. The marketing model proposed was a birth and death model with constant death rate $\mu_n = \mu n$ and birth rate which depended on both types of advertising, namely $\lambda_n = (N - n)(\lambda_1(\sqrt{a_d}) + \lambda_2(\sqrt{a_i}))$, where a_d and a_i represent the direct and indirect, respectively, advertising expenditures at the current time. The advertising expenditures are linearly interpolated between years to yield a continuous function of time. The maximum number of potential customers N is assumed to be 50,000. Using the simulated weighted conditional least squares estimator described in Section 2.2 we obtain estimates of .0000 (standard error=.0000) for λ_1 , .1934 (standard error=.003375) for λ_2 , and 11.83 (standard error=.2205) for μ . For this example, the correlation between the estimate of the indirect advertising coefficient and the exit coefficient is very high, namely .986. Again the standard errors and correlation are found from the asymptotic covariance matrix. As hypothesized by the marketing researchers, direct advertising (coefficient λ_1) does not affect the number of customers the company obtains.

4. SUMMARY

We have presented an alternative method of estimating the parameters of a stochastic process when a closed form representation of the conditional expected value of the process is not available. This method of estimation is comparable to conditional least squares estimators of the parameters. Klimko and Nelson (1978) compare the performance of conditional least squares estimators and maximum likelihood estimators in similar scenarios.

The simulated conditional least squares estimator is preferred over the previously proposed simulated least squares estimator (Ensor, Bridges and Lawera (1993)). To obtain the simulated least squares estimator one must simulate multiple realizations from an initial starting point. This method of simulation can lead to high variability in the sample mean path, thereby leading to instability in the least squares criterion function which is minimized. However, the simulated conditional mean is very stable for a moderate number of simulations resulting in a criterion function with very little noise at a given point in the parameter space. The gain in stability is due to the fact that the estimate of the conditional mean at a particular time is independent of the estimate of the conditional mean at any other time; whereas, the estimate of the mean at a particular time is dependent on the previous history of the process. This independence also facilitates the proofs of the asymptotic properties of the simulated conditional least squares estimators.

REFERENCES

- Bridges, E., Ensor, K. B. and Thompson, J. R. (1992). Marketplace competition in the personal computer industry. *Decision Sciences*, 23, 467-477.
- Bridges, E., Ensor, K. B. and Raman, K. (1994). Sales response to optimal vs. actual promotional expenditures. Submitted to *International Journal of Forecasting*.
- Ensor, K. B., Bridges, K. and Lawera, M. (1993). Simulation based estimation for birth and death processes. *Proceedings of the 88th Conference on the Design of Experiments in Army Research Development and Testing*. ARO Report 93-2, 343-352.
- Gynn, Peter W. (1990). Likelihood ratio gradient estimation for stochastic systems. *Communications of the ACM*, 33, 75-84.
- Hall, P. and Heyde, C. C. (1980). Martingale Limit Theory and its Application, Academic Press, New York.

- Keiding, N. (1975). "Maximum Likelihood Estimation in the Birth-and-Death Process," *The Annals of Statistics*, 3, 363-372.
- Klimko, Lawrence A. and Nelson, Paul I. (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics*, 6, 629-642.
- Nelder, J. A. and Mead, R. (1965). "A simplex method for function minimization." *Computational Journal*, 7, 308-313.
- Thompson, J. R., Brown, B. W. and Atkinson, E. N. (1987). SIMEST: An algorithm for simulation-based estimation of parameters characterizing a stochastic process. Cancer Modeling, eds. James R. Thompson and Barry W. Brown. Marcel Dekker, New York, NY, 387-415.

A NEW SERIES OF SUPERSATURATED DESIGNS

Margaret Gelder Ehm² Marc N. Elliott³
Mounie McGee

Department of Statistics
Rice University
Houston, Texas 77251

February 2, 1994

¹The authors would like to acknowledge Dr. Carl Bates for proposing the problem and to thank Dr. James Thompson assisting with its solution.

²Supported by the U. S. National Library of Medicine Medical Informatics Research Training Program

³supported in part by the U. S. Army Research Office (Durham) under DAAL-01-01-0210 at Rice University

A New Series of Supersaturated Designs

Supersaturated designs are factorial designs in which the number of factors exceeds the number of observations. Such designs preclude the possibility of complete orthogonality, making near orthogonality the obvious goal. In the present paper, designs are constructed using a new and general method. These designs surpass previous designs in all cases but one. Design matrices are presented in the appendix.

1 Introduction

There are many settings in which it is desirable to examine the effects of a large number of factors simultaneously. Plackett and Burman (1946) devised optimal designs for studying $f = n - 1$ factors with n observations. These designs are completely pairwise orthogonal. A natural extension of their work involves studying a number of factors, f , greater than the number of observations, n . Such designs may be useful when it is necessary to examine the influence of many factors, and observations are expensive to collect or are otherwise limited. When $f < n$ complete orthogonality is achievable, but when $f \geq n$ the goal is to obtain a design matrix where the columns are as nearly orthogonal as possible.

One early approach to this problem was the method of group screening proposed by Watson (1961). The method involves combining f factors into g groups. Each group is then tested as a single factor in a standard design. If the effect of a grouped factor is significant its component factors are then tested individually.

Another approach to this problem is the method of supersaturated designs. A supersaturated design is a single design matrix for which f is greater than n . The first approach to supersaturated designs was that of Satterthwaite (1959), who suggested randomly selecting the design vectors. Later, Booth and Cox (1962) devised optimality criteria and a method for generating supersaturated designs. One of their criteria, *near orthogonality*, involves minimizing the maximum absolute value of the dot product of all pairs of vectors. Of the designs that achieve this criterion, one then selects the design that minimizes the number of pairs of vectors with this dot product. Booth and Cox show that a dot product of four is a lower bound for all designs with $f > n$. Note that near orthogonality is a minimax procedure and produces designs such that no pair of vectors is highly correlated. Booth and Cox proposed a second criterion, denoted $E[s^2]$, which is the mean of the squared pairwise dot products. $E[s^2]$ results in designs with a few highly correlated

vectors, but most are pairwise orthogonal. For designs in which no pairwise dot products are larger than 4, near orthogonality and $E[s^2]$ are equivalent criteria in that they yield the same design.

Booth and Cox's method begins with a Plackett and Burman (1946) orthogonal design, to which they add $f - n + 1$ randomly generated trial vectors, resulting in an $n \times f$ initial design matrix. Next, they determine the pair of vectors with the greatest dot product, and attempt to replace each of the two vectors with a new randomly generated vector. A vector is replaced if the resulting design matrix is superior in terms of the near orthogonality criterion. Booth and Cox continued this process until 30 minutes of clock time on the University of London Mercury computer passed without an improvement to the design matrix. Booth and Cox used $n = 12, 18, 24$ and f as large as $2n$.

Rosenberger and Smith (1984) focused on very small designs ($f = 4, 5, \dots, 9$ and $n \leq f$). For these designs, they were able select the best design according to the near orthogonality criterion by means of an exhaustive search. For larger designs this approach is computationally untenable. For example, the number of possible designs when $f = 24$ and $n = 12$ is on the order of 10^{47} .

Lin's designs (1993) involve selecting a half fraction of a Plackett and Burman design matrix of size $2n$. The resulting matrices have n observations and $f = 2n - 2$ factors. He examined all such half fraction designs resulting from a given Plackett and Burman design and reported the best design according to the near orthogonality criterion. To obtain designs with $f < 2n - 2$, Lin selected a subset of the columns from his $f = 2n - 2$ design. Lin used a variety of n 's between 6 and 30.

In the present paper we seek a general method that improves upon existing supersaturated designs with $n = 4k$, such that k is a positive integer. This is the class of designs for which pairwise orthogonality is possible and for which corresponding Plackett and Burman designs exist.

2 Method

We begin our method by creating a matrix of f randomly generated design vectors. The next stage involves a series of passes designed to improve the initial matrix. Each pass examines each of the f vectors in sequential order. When a vector is examined it is compared to a set of alternative vectors with respect to the resulting near orthogonality criterion for the entire matrix. If a superior alternative vector exists, the original vector is replaced by the best of the alternatives. The series of passes continues until a pass occurs that fails to improve any of the f vectors. For $n = 8, 12, 16$, and 20 the set of

alternative vectors is composed of all possible design vectors. For $n \geq 24$, this is not computationally practical. The set of alternative vectors consists of a randomly sampled subset of all possible design vectors. Note that this method is sufficiently general to be used with supersaturated designs of any dimension and with any optimality criterion based on pairwise dot products, such as $E[s^2]$.

3 Results

We generated designs for $n = 8, 12, 16, 20$, and 24 , and corresponding sets of $f : n < f \leq 2n$. A summary of the designs obtained appears in tables 1–5. The actual design matrices are given in the appendix. Note that “EEM” refers to Ehm, Elliott, and McGee and denotes our method. The column heading “0” refers to the number of pairwise dot products equal to 0, the column heading “4” refers to the number of pairwise dot products equal in absolute value to 4, and so forth.

$n = 8$ EEM		
f	0	4
12	46	20
14	63	28
16	76	44

Table 1: Designs with $n = 8$

Table 1 refers to designs with $n = 8$. In this case, Booth and Cox do not present a design and Lin’s method cannot be used.

f	$n = 12$ EEM		$n = 12$ Lin		$n = 12$ B & C	
	0	4	0	4	0	4
16	81	39	73	47	67	53
18	94	59	90	63	—	—
20	109	81	—	—	75	115
22	128	103	132	99	—	—
24	141	135	—	—	101	175

Table 2: Designs with $n = 12$

Table 2 refers to designs with $n = 12$. Here, all three methods have been applied. In Figure 1, we plot the proportion of pairwise orthogonal vectors in

Proportion of Pairwise Orthogonal Vectors

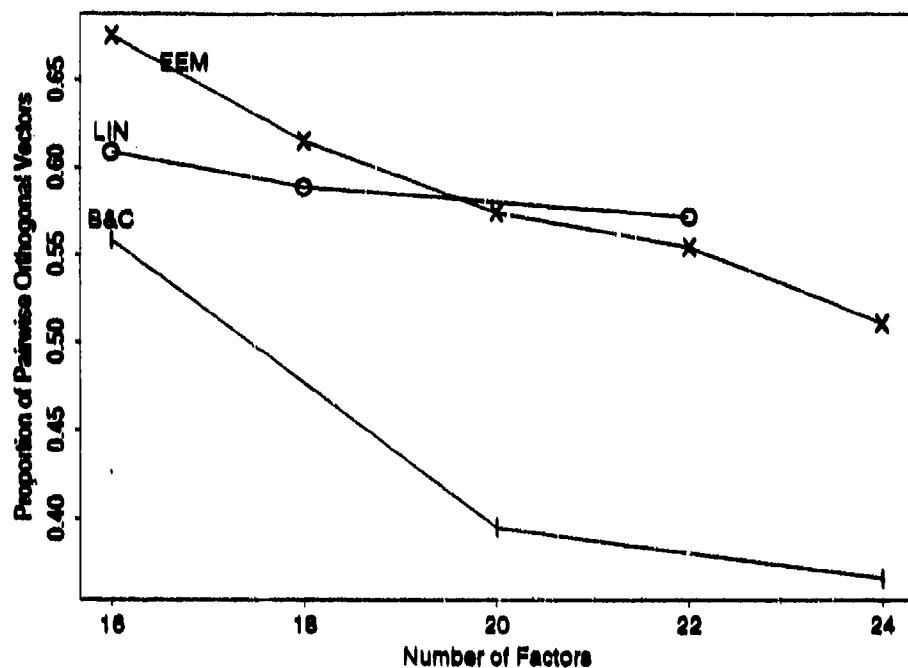


Figure 1: Proportions of Pairwise Orthogonal Vectors for the EEM, Lin, and Booth and Cox Methods

order to compare the designs graphically. Clearly, both our method and Lin's are superior to Booth and Cox's method. Our method surpasses Lin's in all cases except when $f = 2n - 2$, the case for which his method is designed.

Tables 3 and 4 refer to designs where $n = 16$ and $n = 20$, respectively. In these cases, Booth and Cox do not present a design and Lin's method cannot be used.

Table 5 refers to designs with $n = 24$. Here, all three methods have been applied. With $f = 30$, our design clearly surpasses Booth and Cox's design according to the near orthogonality criterion. Also, with $f = 46$ our design is superior to Lin's in terms of near orthogonality. This case, where $f = 2n - 2$, is the case for which his was specifically designed. For designs in which some pairwise dot products are larger than 4, $E[s^2]$ and near orthogonality are no longer equivalent criteria. As discussed earlier, these criteria produce designs with different characteristics. In light of this, one could apply our method using $E[s^2]$ as the criterion if one preferred designs having a higher proportion of pairwise orthogonal vectors, but also having some pairwise dot products

equal in absolute value to 8.

As seen in Figure 2, the proportion of pairwise orthogonal vectors attainable decreases steadily as the number of factors in the design increases. We suspect this reflects the inherent geometry of the problem. Note in figure 1, the proportion of pairwise orthogonal vectors in Lin's designs does not decrease substantially as f increases. This suggests that Lin's method of selecting subsets of designs where $f = 2n - 2$ to obtain designs with smaller f is inadequate. Another observable trend is that, for a given number of factors, one obtains a slightly better design with larger n . The magnitude of the effect of f is larger than that of n .

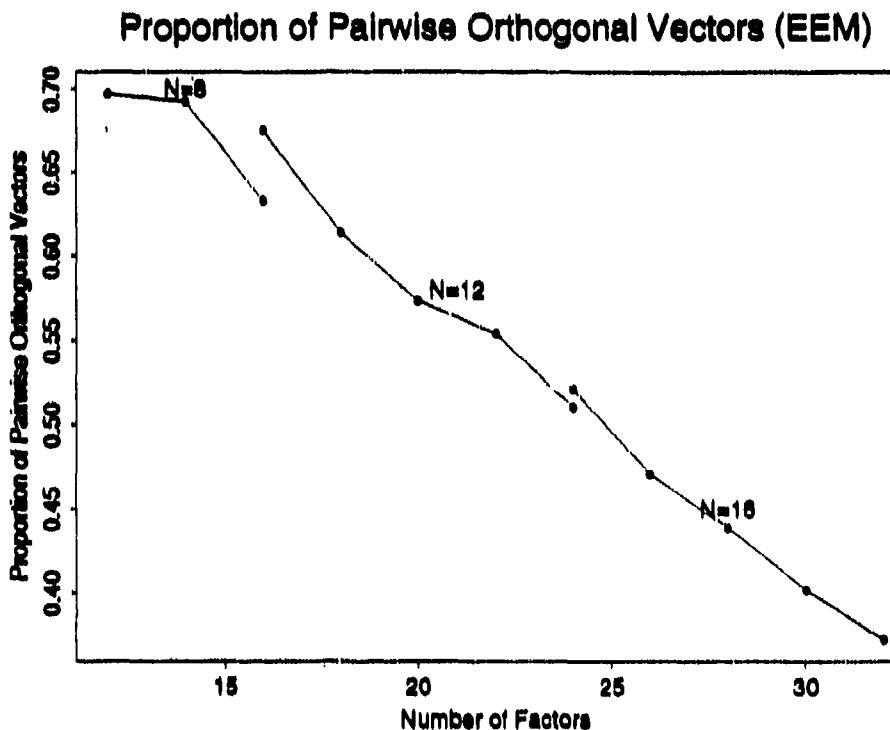


Figure 2: Decrease in the proportion of attainable pairwise orthogonal vectors as the number of factors increases

4 Discussion and Conclusions

We will consider our method in comparison to competing methods. First we will compare our approach to that of Booth and Cox. Both methods are

large scale search procedures, but ours has a natural stopping criterion where as theirs is arbitrary. Because both are general algorithms they can be used for a wide variety of n and f . For comparable designs, our method produces uniformly superior results.

We suspect the following observations may explain why we obtain better results than Booth and Cox. We found that starting with an orthogonal matrix made it difficult to add vectors with small dot products, while starting with randomly generated design vectors produced superior results. We also achieved better results when sequentially considering each of the f vectors for replacement rather than replacing the vectors at random.

We now compare our method to that of Lin. Although computationally simple, Lin's method is highly specialized and can only be applied to a limited number of values of n . For given n it only produces designs for $f \leq 2n - 2$. Furthermore, Lin's method for deriving designs for $f < 2n - 2$ gives poor results. This is consistent with our own findings. The ideal set of design vectors changes so dramatically from one level of f to another that even the best subsets of larger matrices do not yield good designs. This further suggests that methods specialized for particular values of n and f are unlikely to produce good designs for other combinations of n and f . Our method generates design matrices that exceed Lin's in all cases, but one.

Our designs for $n = 16$ and $f = 32$ is the largest supersaturated design published for which all dot products are less than or equal in absolute value to the theoretical minimum of 4.

5 Future Work

This research was inspired by a problem posed by the late Dr. Carl Bates. He needed to estimate the effects of 104 factors using 52 observations. The 104 factors were parameters in a model and the observations were the 52 Sun workstations to which he had access. In order to solve this problem we hope to thoroughly investigate designs where $n \geq 20$ with respect to the near orthogonality criterion. We also hope to produce designs in which the maximum pairwise dot product is 4 for $n \geq 20$. Finally we will investigate the class of designs generated by our method when the optimality criterion is $E[s^2]$.

References

- Booth, K.H.V. and D.R. Cox, (1962) "Some Systematic Supersaturated Designs," *Technometrics*, 4, 489-495.
- Lin, D.K.J. (1993) "A New Class of Supersaturated Designs," *Technometrics*, 35, 28-31.
- Plackett, R.L. and J.P. Burman, (1946) "The Design of Optimum Multi-factorial Experiments," *Biometrika*, 33, 303-325.
- Rosenberger, J.L. and D.E. Smith (1984) "Interruptible Supersaturated Two-Level Designs," *Communications in Statistics-Theory and Methodology*, 13(5), 599-609.
- Satterthwaite, F. (1959) "Random Balance Experimentation" (with discussion), *Technometrics*, 1, 111-137.
- Watson, G.S. (1961) "A Study of the Group Screening Method," *Technometrics*, 3, 209-217.

$n = 16$ EEM		
f	0	4
24	144	132
26	153	172
28	166	212
30	175	260
32	185	311

Table 3: Designs with $n = 16$

$n = 20$ EEM			
f	0	4	8
28	171	207	0
30	170	265	0
32	174	322	0
34	173	385	3
36	176	449	5
38	204	488	11
40	222	542	16

Table 4: Designs with $n = 20$

f	$n = 24$ EEM			$n = 24$ Lin			$n = 24$ B & C		
	0	4	8	0	4	8	0	4	8
30	178	257	0				295	81	59
32	183	311	2						
34	202	351	8						
36	202	416	12						
38	223	460	20						
40	243	507	30						
42	261	560	40						
44	273	623	50				-	-	-
46	309	662	64	414	552	69	-	-	-
48	337	711	80	-	-	-	-	-	-

Table 5: Designs with $n = 24$

6 Appendix

The design matrices generated by our method are given below. In each matrix, the rows correspond to the observations and the columns to the factors.

-	+	+	-	-	-	+	-	-	+	+	+	+
-	.	.	.	+	+	+	.	+	.	+	.	-
-	+	+	+	-	+	-	+	+	.	+	+	+
+	.	+	+	+	-	.	-	+	.	.	+	+
+	+	.	-	+	+	+	+	.	.	.	-	+
+	+	.	+	-	+	-	.	+	.	.	-	-
-	-	.	-	+	+	+	-	-	-
+	.	+	+	+	-	+	+	+	+	-	-	-

$$n = 8 \quad f = 12$$

-	+	.	.	.	+	+	.	+	.	+	.	+
+	+	+	+	.	-	+	+	.	-	+	+	+
-	+	.	.	+	-	.	-	+	.	+	.	-
+	.	+	+	+	-	+	.	-	+	.	.	-
-	.	.	+	-	+	+	+	+	-	.	-	-
+	.	.	-	+	+	+	-	+	.	+	+	+
-	.	+	+	+	-	+	+	+	+	+	+	+
+	+	+	-	-	+	-	+	+	-	.	-	-

$$n = 8 \quad f = 14$$

-	+	.	+	.	.	+	+	.	.	+	.	.
+	-	+	-	+	-	+	-	+	+	+	+	+
+	+	.	+	-	+	-	.	+	+	+	+	-
-	+	+	-	+	-	+	-	+	+	-	+	+
-	-	+	+	-	+	-	+	-	+	.	+	+
+	-	-	+	+	-	+	-	+	-	+	+	-
-	+	-	-	+	+	-	+	+	-	.	+	-
+	-	+	-	-	+	+	+	+	-	+	-	-

$$n = 8 \quad f = 16$$

A 12x20 grid of '+' symbols, representing a sparse matrix with 12 rows and 20 columns. The grid is composed of 240 '+' characters.

$$n = 12 \quad f = 20$$

$n = 12$ $f = 22$

$$n = 12 \quad f = 22$$

$$n = 12 \quad f = 24$$

A 16x24 grid of '+' and '-' symbols. The grid is composed of two sets of 8 columns each. The first set of columns contains '+' symbols at the top and '-' symbols at the bottom. The second set of columns contains '-' symbols at the top and '+' symbols at the bottom. The grid is bounded by a thick black border.

$$n = 16 \quad f = 24$$

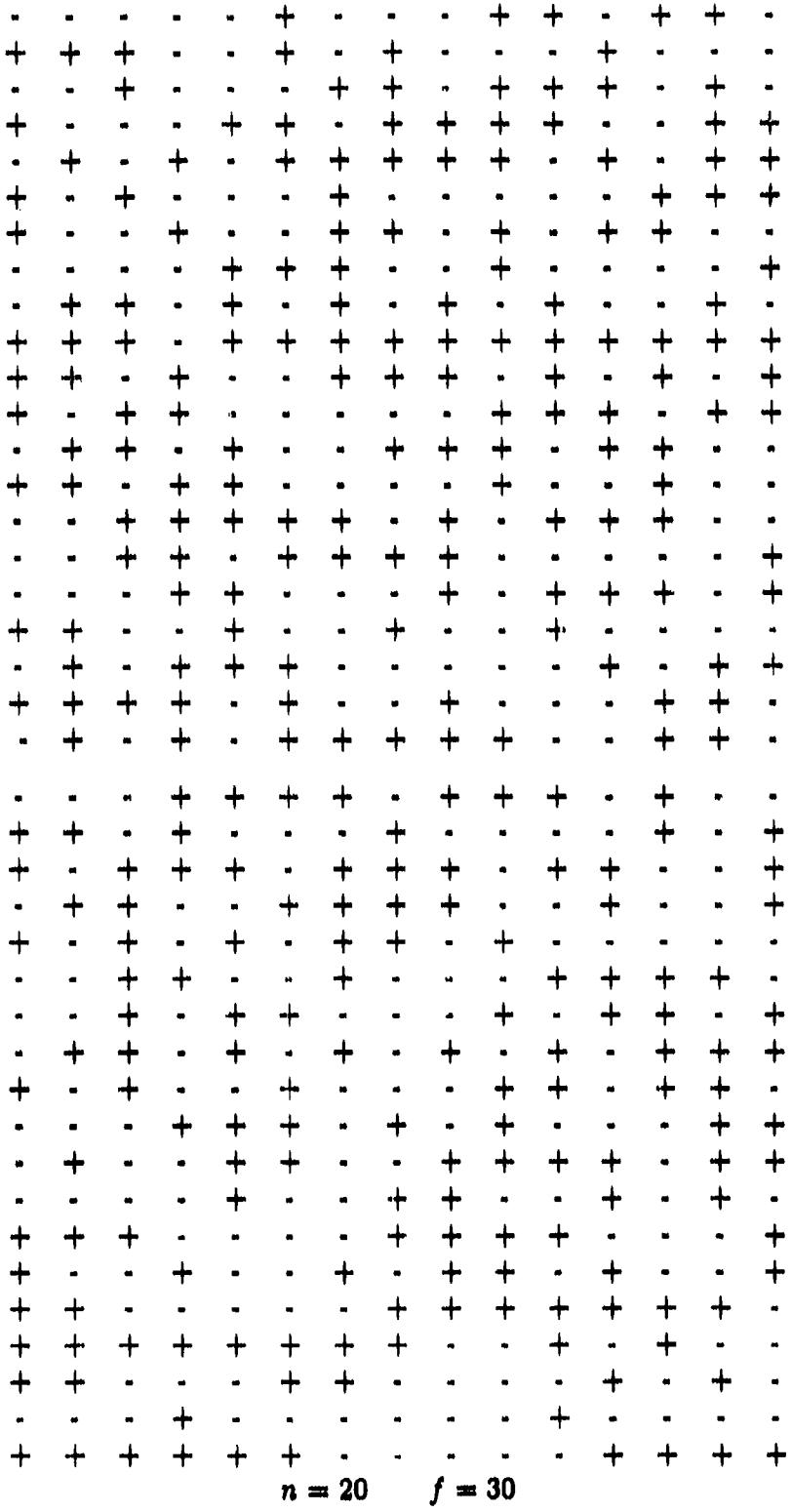
A 16x16 grid of '+' symbols, representing a sparse matrix. The grid is composed of 256 '+' symbols arranged in a square pattern.

A 16x28 grid of '+' symbols, representing a sparse matrix or a specific pattern. The grid is composed of two main parts: a top section with 16 rows and 28 columns, and a bottom section with 16 rows and 28 columns. The symbols are arranged in a repeating pattern of '+' and '-' across the entire grid.

$n = 16$ $f = 28$

A large grid of '+' symbols arranged in a rectangular pattern. The grid has 16 columns and 30 rows, starting from the top-left corner and ending at the bottom-right corner. The '+' symbols are black and have a medium size.

A 16x32 grid of '+' symbols, representing a sparse matrix or a specific pattern of data points.



A grid of black '+' symbols on a white background, representing a 20x32 matrix. The grid consists of 20 rows and 32 columns of '+' characters.

A grid of 20 columns and 34 rows of black '+' symbols on a white background. The grid is composed of two sets of parallel vertical lines and two sets of parallel horizontal lines, intersecting to form a continuous pattern of '+' symbols.

A large grid of black '+' symbols on a white background. The grid consists of approximately 20 columns and 36 rows of symbols, creating a pattern of small squares.

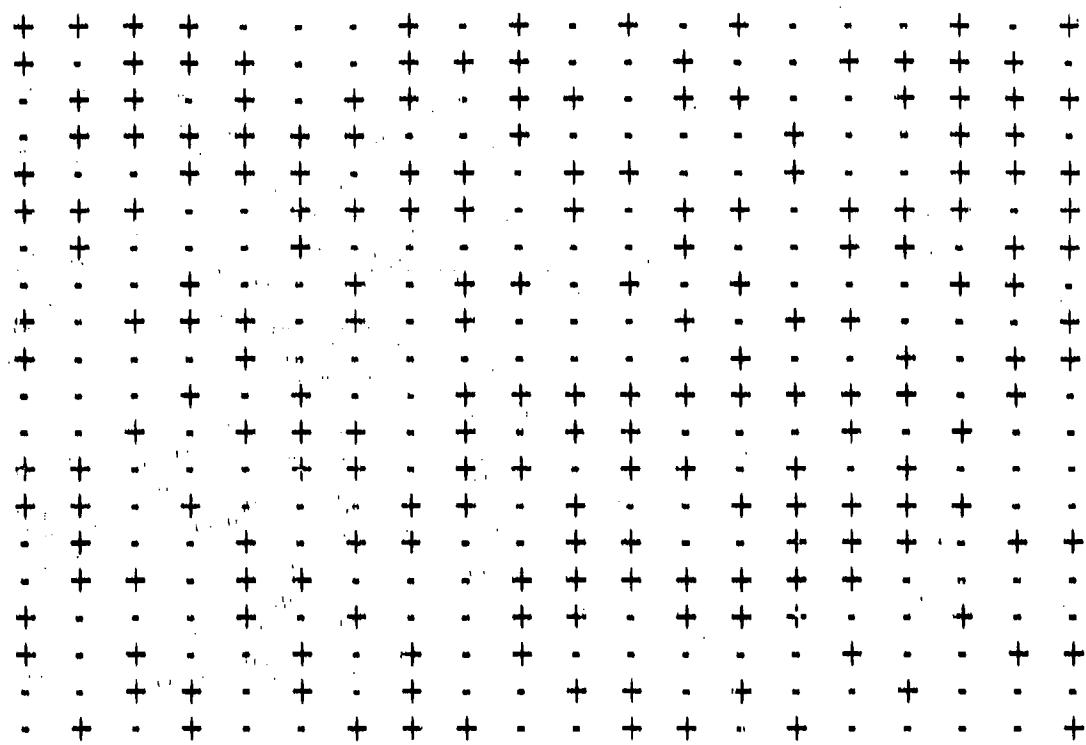
$$n = 20 \quad f = 36$$

A uniform grid of black '+' symbols arranged in approximately 20 horizontal rows and 20 vertical columns, covering most of the page.

n = 20 f = 38 continued on next page

A uniform grid of black '+' symbols arranged in approximately 25 horizontal rows and 25 vertical columns, creating a pattern of small squares across the page.

$$n = 20 \quad f = 38$$



$n = 20$ $f = 40$

A uniform grid of black '+' symbols arranged in approximately 25 horizontal rows and 25 vertical columns, creating a pattern of small squares across the page.

$n = 20$ $f = 40$ continued on next page

A uniform grid pattern consisting of small black plus signs ('+') arranged in a regular, overlapping grid across the entire page. The grid is composed of approximately 10 columns and 10 rows of symbols.

$n = 24$ $f = 30$ continued on next page

A uniform grid of black '+' symbols is arranged in 10 horizontal rows and 10 vertical columns, creating a pattern of 100 '+' symbols across the entire page.

$$n = 24 \quad f = 30$$

A uniform grid of black '+' symbols arranged in approximately 25 horizontal rows and 25 vertical columns, creating a pattern of small squares across the page.

n = 24 f = 32 continued on next page

A uniform grid of black '+' symbols arranged in approximately 25 horizontal rows and 25 vertical columns, creating a pattern of small squares across the page.

$$n = 24 \quad f = 32$$

A uniform grid of black '+' symbols arranged in a rectangular pattern. The symbols are placed at the intersections of horizontal and vertical lines, creating a dense, repeating pattern across the entire page.

$n = 24$ $f = 34$ continued on next page

A uniform grid of black '+' symbols arranged in a rectangular pattern. The symbols are placed at the intersections of horizontal and vertical lines, creating a dense, repeating pattern across the entire page.

$$n = 24 \quad f = 34$$

A uniform grid of black plus signs (+) arranged in rows and columns on a white background. The grid covers the entire page.

n = 24 f = 36 continued on next page

A uniform grid of black '+' symbols arranged in a rectangular pattern. The grid consists of approximately 20 columns and 25 rows, creating a dense, repeating texture across the entire frame.

$$n = 24 \quad f = 36$$

A uniform grid of black '+' symbols is arranged in 20 horizontal rows and 20 vertical columns, creating a pattern of small squares across the entire page.

n = 24 f = 38 continued on next page

A uniform grid of black plus signs (+) arranged in rows and columns on a white background. The grid covers the entire page.

$$n = 24 \quad f = 38$$

A uniform grid of black plus signs (+) is arranged in a rectangular pattern across the entire page. The grid consists of approximately 10 columns and 15 rows of symbols, creating a clean, mathematical, and structured visual effect.

n = 24 *f* = 40 continued on next page

A uniform grid of black plus signs (+) arranged in a rectangular pattern. The grid consists of approximately 20 columns and 25 rows, covering most of the page area.

$$n = 24 \quad f = 40$$

A uniform grid of black '+' symbols is arranged in 10 rows and 10 columns. Each symbol is positioned at the intersection of a horizontal and vertical line, creating a pattern of alternating '+' and '-' signs across the entire area.

$n = 24$ $f = 42$ continued on next page

A uniform grid of black '+' symbols is arranged in 10 horizontal rows and 10 vertical columns, creating a pattern of 100 '+' symbols across the entire page.

$$n = 24 \quad f = 42$$

A uniform grid of black '+' symbols is arranged in 16 horizontal rows and 16 vertical columns, creating a pattern of small squares across the entire page.

$n = 24$ $f = 44$ continued on next page

A uniform grid of black '+' symbols arranged in a rectangular pattern. The symbols are placed at the intersections of horizontal and vertical lines, creating a dense, repeating pattern across the entire page.

n = 24 f = 44 continued on next page

A uniform grid of black '+' symbols is arranged in 16 horizontal rows and 16 vertical columns, creating a pattern of small squares across the entire page.

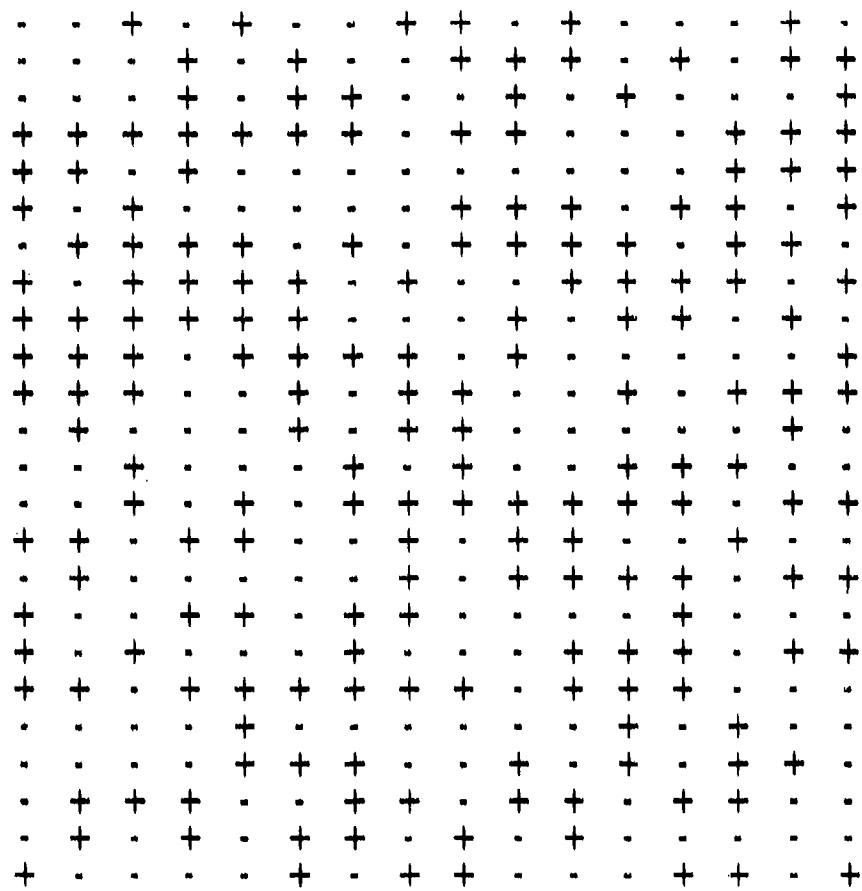
$$n = 24 \quad f = 44$$

A uniform grid of black '+' symbols arranged in a rectangular pattern. The symbols are placed at the intersections of horizontal and vertical lines, creating a dense, repeating pattern across the entire frame.

$n = 24$ $f = 46$ continued on next page

A uniform grid of black '+' symbols arranged in a rectangular pattern. The grid consists of approximately 20 columns and 25 rows, creating a dense, repeating pattern across the entire page.

n = 24 f = 46 continued on next page



$n = 24$ $f = 46$

A uniform grid of black '+' symbols arranged in a rectangular pattern. The symbols are placed at the intersections of horizontal and vertical lines, creating a dense, repeating pattern across the entire page.

n = 24 f = 48 continued on next page

A uniform grid of black '+' symbols arranged in a rectangular pattern. The grid consists of approximately 20 columns and 25 rows, covering most of the page area.

n = .24 f = 48 continued on next page

A uniform grid of black '+' symbols arranged in a rectangular pattern. The grid consists of approximately 20 columns and 25 rows, creating a dense, repeating texture across the entire page.

$$n = 24 \quad f = 48$$

AN EMPIRICAL STUDY OF THE DISTRIBUTION AND PROPERTIES OF
THE SLOPE ESTIMATOR USING MINIMUM NORMED DISTANCE CRITERION

Barbara A. Wainwright

Statement of the Problem

How should one estimate a linear relation between two variables? It is common to use a regression model and automatically apply the ordinary least squares method of estimating parameters. This is sometimes the wrong model and method and thus one should consider other types because of the variables and the assumptions in question. This leads in turn to consider various techniques of estimation. The topic of this paper is the estimation of linear structural relations when there is measurement error in both the dependent and independent variables. These problems are often referred to as Model II regression problems [Graybill, 1961] or measurement error models [Fuller, 1987]. There are several techniques for estimating the model parameters. However, the technique that will be investigated is the one that minimizes the perpendicular distance between the observed points and the estimated line. While these estimates have been derived, there is very little known about the exact distribution of the slope estimator and some of its properties other than consistency, some asymptotic properties [Fuller, 1987], and some approximate tests and confidence limits [Creasy, 1956; Kendall and Stuart, 1973]. This paper will investigate the following properties of the slope estimator:

1. the shape of the density of this estimator for small samples,
2. the expected value,
3. the bias, and
4. the probability of Type I errors for both small and large samples.

The Minimum Norm Distance Method of Estimation for The Classical Errors in Variables Case

There are many techniques for estimating the structural relation parameters but if one assumes normality and uses maximum likelihood estimation techniques, then unidentifiability is an issue. There are ways to alleviate this problem when either σ_e^2 , the measurement error variance associated with the y values, σ_d^2 , the measurement error variance associated with the x values, or the ratio of the two error variances (λ) is known [Kendall and Stuart, 1973; Lindley, 1953]. We will examine the one in which we know the ratio of the variances of the measurement errors, lambda. This

case is referred to by Fuller [1987] as the *classical errors in variables case*. The resulting maximum likelihood estimator actually minimizes the weighted sum of the squared statistical distances between the observed points and the estimated line. For the case in which $\lambda = 1$, it would minimize the perpendicular distances. This is often referred to as the "minimum norm distance." The problem of minimizing the norm distance was discussed as early as 1877 by Adcock, 1879 by Kummel, and 1901 by Karl Pearson. However, this approach is regularly attributed (especially in clinical chemistry) to W. Edwards Deming, who reintroduced it in 1943 [Cornbleet and Gochman, 1979; Goldschmidt et al., 1981; Lloyd, 1978; Mandel, 1964; Northam, 1981; Schall et al., 1980; Smith et al., 1980; Vormbrock and Helger; Wakkers et.al., 1975; Weisbrot, 1985; Westgaard and Hunt, 1973; Zucker, 1947]. According to Mandel [1964], Deming minimized the weighted sum of squares

$$S = \sum_{i=1}^n [(x_i - \hat{X}_i)^2 + (y_i - \hat{Y}_i)^2 \lambda]$$

such that

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} \hat{x}_i.$$

The resulting slope estimator may be expressed as

$$\hat{\beta} = \frac{s_{yy} - \frac{1}{\lambda} s_{xx}}{2 s_{xy}} + \sqrt{\left(\frac{s_{yy} - \frac{1}{\lambda} s_{xx}}{2 s_{xy}} \right)^2 + \frac{1}{\lambda}} \quad (1)$$

where

$$s_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$s_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$s_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Other resulting estimators are

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

and

$$\hat{\sigma}_e^2 = \frac{1}{n-2} \left[\sum_{i=1}^n (y_i - \bar{y})^2 - \hat{\beta} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) \right].$$

If lambda is not known, then assuming it is one is possibly better than ignoring it all together. However, Vormbrock and Helger suggest the use of duplicates for estimating lambda in method comparison studies analyzed with Deming's procedure. For this approach the following sums of squares are calculated

$$Q_x = \sum_{i=1}^n (x_{i1}^2 + x_{i2}^2) - \frac{\left(\sum_{i=1}^n (x_{i1} + x_{i2}) \right)^2}{2n}$$

$$Q_y = \sum_{i=1}^n (y_{i1}^2 + y_{i2}^2) - \frac{\left(\sum_{i=1}^n (y_{i1} + y_{i2}) \right)^2}{2n}$$

$$Q_{xy} = \sum_{i=1}^n (x_{ii}y_{ii} + x_{i2}y_{i2}) - \frac{\left(\sum_{i=1}^n (x_{ii} + x_{i2})\right) \left(\sum_{i=1}^n (y_{ii} + y_{i2})\right)}{2n}$$

$$\lambda = \frac{\sum_{i=1}^n (x_{ii} - x_{i2})^2}{\sum_{i=1}^n (y_{ii} - y_{i2})^2}$$

from which the following are computed

$$\hat{\beta} = \frac{\hat{\lambda} Q_y - Q_x + \sqrt{(\hat{Q}_x - \hat{\lambda} \hat{Q}_y)^2 + 4 \hat{\lambda} \hat{Q}_{xy}^2}}{2 \hat{\lambda} \hat{Q}_{xy}} \quad (2)$$

$$s_{yx} = \sqrt{\frac{\hat{Q}_y - 2 \hat{\beta} \hat{Q}_{xy} + \hat{\beta}^2 \hat{Q}_x}{(1 + \hat{\lambda} \hat{\beta}^2)(2n - 2)}}$$

According to Feldman et al. [1981] no one knows the exact sampling distribution of $\hat{\beta}$. It is important to test $H_0: \beta = 1$, particularly in method comparison studies. If this is taken as a constraint on principal components and standard principal components, it is equivalent to saying $\sigma_y^2 = \sigma_x^2$ when X and Y are dependent. Morgan [1939] transforms these measurements and then opts for a t-test. Using a similar t-test with a slightly different transformation, one can test $\beta = \beta_0$ for any value of β_0 . Confidence intervals can also be constructed. However, Morgan's test does not apply if the above constraint can not be imposed. Kendall and Stuart [1973] as well as Creasy [1956] give confidence limits and tests of hypotheses for the case $\lambda = 1$ using the fact that $\hat{\beta} = \tan \phi$, and the fact that the sample correlation coefficient, r , has a "Student's - t" distribution with $(n-2)$ degrees of freedom when y is normally distributed and the correlation coefficient, ρ , is zero.

From the literature, it does not appear that the expected value of $\hat{\beta}$ is known; especially since the sampling distribution is not known. Fuller [1987], in his exposition of the asymptotic properties, claims that $\hat{\beta} - \beta = O_p(n^{-1})$. He also gives the variances of the limiting distribution. These are

$$V(\hat{\beta}) = \frac{\hat{\sigma}^2 s + \hat{\sigma}_d^2 s - \hat{\beta}^2 \hat{\sigma}_d^4}{(n-1) \hat{\sigma}^4} \quad (3)$$

where

$$s = \frac{(n-1) \left(\frac{1}{\lambda} + \hat{\beta}^2 \right) \hat{\sigma}_d^2}{(n-2)}$$

$$\hat{\sigma}_d^2 = \frac{m_{YY} + \frac{1}{\lambda} m_{XX} - \sqrt{(m_{YY} - \frac{1}{\lambda} m_{XX})^2 + 4 \frac{1}{\lambda} m_{XY}^2}}{2 \frac{1}{\lambda}}$$

and

$$\hat{\sigma}^2 = \frac{\sqrt{(m_{YY} - \frac{1}{\lambda} m_{XX})^2 + 4 \frac{1}{\lambda} m_{XY}^2} - (m_{YY} - \frac{1}{\lambda} m_{XX})}{2 \frac{1}{\lambda}}$$

Sampling Distribution of $\hat{\beta}$

Although the exact sampling distribution of $\hat{\beta}$ is complex and there does not seem to be a general closed form solution, by examining various expressions of $\hat{\beta}$, and by using various transformations of variables and Mellin transforms, we can obtain expressions for the density of $\hat{\beta}$ for some special cases. For small samples under some of these situations, we can show that the density is far from being a t distribution.

Special Case: $n = 2$

To investigate the density of the slope estimator, let us consider the simplest case in which $n = 2$. When $n = 2$, the line is uniquely determined regardless of the intended method of estimation. The density of the estimator will, however, depend upon the conditions imposed and the assumptions made. First, we will assume that the relation between X and Y is given by

$$Y = \beta X. \quad (4)$$

That is, we will assume that the intercept, α , is zero.

Case 1: X_i Measured Without Error

For the first case we will assume that the relation above holds and that the X_i are fixed or predetermined values of the random variable x_i . These values are measured without error. This is therefore an example of a regression situation. We observe

$$y_i = Y_i + e_i$$

where

$$e \sim N(0, \sigma_e^2)$$

and thus

$$y_i \sim N(\beta X_i, \sigma_e^2)$$

The slope of the line in this case is given by

$$\hat{\beta} = \frac{y_2 - y_1}{X_2 - X_1} = \frac{Y^*}{X^*}$$

In this case

$$y^* = y_2 - y_1 \sim N(\beta X_2 - \beta X_1, 2\sigma_e^2)$$

and

$$x^* = X_2 - X_1$$

is constant. Thus

$$\hat{\beta} \sim N\left(\beta, \frac{2\sigma_e^2}{(X_2 - X_1)^2}\right)$$

Note that in the usual case of least squares,

$$\text{var}[\hat{\beta}] = \frac{\sigma_e^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

which is the same.

Case 2: X_i a Normally Distributed Random Variable

Let us now assume that the relation defined is a structural one between unobservable random variables. That is X_i is an unobservable random variable. Assume the following:

$$X_i \sim N(\mu, \sigma^2)$$

where we observe

$$x_i = X_i + d_i, \quad d_i \sim N(0, \sigma_d^2)$$

$$y_i = Y_i + \epsilon_i$$

$$= \beta X_i + \epsilon_i$$

where

$$\epsilon_i \sim N(0, \sigma_e^2).$$

Thus the following distributional theory exists.

$$x_i \sim N(\mu, \sigma^2 + \sigma_d^2)$$

$$y_i \sim N(\beta\mu, \beta^2\sigma^2 + \sigma_e^2).$$

For the case in which $n = 2$,

$$\hat{\beta} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{y^*}{x^*}$$

where

$$y^* \sim N(0, 2(\beta^2\sigma^2 + \sigma_e^2))$$

$$x^* \sim N(0, 2(\sigma^2 + \sigma_d^2))$$

In this case $\hat{\beta}$ is a ratio of normal random variables, each having mean zero. One could attempt to obtain the density of $\hat{\beta}$ by changing variables and by using the moment generating function. However, Maple and Derive could not evaluate these integrals. Through Mellin transforms, Springer [1979] derives the density for the ratio of two dependent standard normals (p. 156) and Craig [1942] derives the density for the ratio of two dependent normal random variables each with mean zero and any finite variance. Taking the bivariate normal density of (x^*, y^*) , making a change of variables, and using the Mellin transform for two dependent random variables [Craig, 1942; Springer, 1979], it is determined that the density of $\hat{\beta}$ for this case is given by

$$f(\hat{\beta}) = \frac{\sigma_x \cdot \sigma_y \cdot \sqrt{1 - \rho^2}}{\pi (\sigma_{y\cdot}^2 - 2\rho \sigma_x \cdot \sigma_y \cdot \hat{\beta} + \sigma_x \cdot \hat{\beta}^2)}$$

$$= \frac{4 \sqrt{(\sigma^2 + \sigma_d^2) (\hat{\beta}^2 \sigma^2 + \sigma_e^2) (1 - \rho)}}{\pi \left((\hat{\beta}^2 \sigma^2 + \sigma_e^2) - 2\rho \sqrt{(\sigma^2 + \sigma_d^2) (\hat{\beta}^2 \sigma^2 + \sigma_e^2)} \hat{\beta} + (\sigma^2 + \sigma_d^2) \hat{\beta}^2 \right)}.$$

It is also worth noting that if x and y obey a normal bivariate probability density function, the mean value of y/x does not exist.

Case 3: X_i an Unobservable Mathematical variable

If X_i is a mathematical variable, then the relation $y = \beta X$ is a functional relation. If we observe

$$x_i = X_i + d_i, \quad d_i \sim N(0, \sigma_d^2)$$

then

$$y_i = \beta X_i + e_i$$

where

$$e_i \sim N(0, \sigma_e^2)$$

and

$$y_i \sim N(\beta X_i, \sigma_e^2).$$

The slope estimator is computed in the usual manner, but in this case

$$y^* = y_2 - y_1 \sim N(\beta X_2 - \beta X_1, 2\sigma_e^2)$$

$$x^* = x_2 - x_1 \sim N(x_2 - x_1, 2\sigma_d^2).$$

Among the various substitutions or change of variables that Craig [1942] uses is that

$$z = \frac{y^*}{x^*}$$

$$w = z \frac{\sigma_{x^*}}{\sigma_{y^*}}.$$

For the case of a ratio of two dependent normal random variables with nonzero means, Craig [1942] provides an expression for the density of w as follows:

$$f(w) = \frac{\exp\left[\frac{1}{2(1-\rho^2)}(x_1^2 - 2x_1x_2 + x_2^2)\right]}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{au^2}{2} + bu\right) |u| du$$

where

$$a = \frac{1 - 2\rho w^2}{(1 - \rho^2)} > 0, \quad \rho^2 < 1$$

$$b = \frac{x_1 - \rho x_2 + (x_2 - \rho x_1)w}{(1 - \rho^2)}.$$

According to Craig, "this can be calculated from existing tables for particular values of w and of the parameters." No closed form solution seems to exist particularly for z.

All of this theory above still applies for the case in which $Y = \alpha + \beta X$. The means in each case will be identical since α will merely subtract out. Therefore the distributions derived for each of the three cases will be the same regardless of the value of α .

General Case: Arbitrary n

Various transformations of variables were made in an attempt to derive the density of β . Without loss of generality we can assume that the means are zero and thus we can use the sums of squares and crossproducts rather than the sums of squared deviations from the mean. It is still the case that when $\lambda = 1$,

$$\hat{\beta} = w + \sqrt{w^2 + 1} \quad (5)$$

where

$$w = \frac{a - b}{2c}$$

but now

$$a = \sum_{i=1}^n y_i^2$$

$$b = \sum_{i=1}^n x_i^2$$

$$c = \sum_{i=1}^n x_i y_i$$

We hoped to obtain the density of w through an appropriate transformation or through Mellin transforms. Once we had the density of w, then another change of variable based on Equation 5 could possibly yield the density of β . It did not.

Expected Value of $\hat{\beta}$ and Analysis of the Bias

Through Taylor series expansions an approximate expression is obtained for the asymptotic expected value of $\hat{\beta}$. For large n , under the assumption that $\lambda = 1$, we may conclude that

$$E[\hat{\beta}] \approx \left(\frac{1}{2}\beta - \frac{1}{2\beta} + \frac{\sigma_e^2 - \sigma_d^2}{2\beta\sigma^2} \right) + \sqrt{\left(\frac{1}{2}\beta - \frac{1}{2\beta} + \frac{\sigma_e^2 - \sigma_d^2}{2\beta\sigma^2} \right)^2 + 1}$$

Recall that $\lambda=1$ implies $\sigma_e^2 = \sigma_d^2$ and the above expression reduces to β which is expected of a consistent estimator. However, it can be shown that use of an incorrect value of λ can introduce an additional bias that does not approach zero in the limit as n goes to infinity. Note that when $\lambda \neq 1$, a similar expression results and the same situation exists.

Simulation Results

Large Samples

Computer simulations are performed using SPlus. Cases are considered for which X is a fixed vector with measurement error and for X a random variable. For each of these cases various values of the parameters and various sample sizes are considered. These simulations seem to support the theories. For large samples the distribution of $\hat{\beta}$ appears normal in most cases. However the use of an incorrect value of λ does introduce bias as Figures 1 through 4 indicate. Note that underestimating λ results in underestimation of β on the average and vice versa.

Table 1 provides the mean, variance, upper and lower tail probability of rejection of $H_0: \beta=1$ for various cases. It should be noted that for $n=100$ and $\beta=1$ most simulations result in a density of $\hat{\beta}$ that is very close to a normal density. However, in method comparison and bioequivalence studies, we often assume that $\lambda=1$ if the error variances are unknown. Doing so can shift the density to the left or right. The shift or biasing effect can greatly increase the chance of making a Type I error in testing $H_0: \beta=1$. It should be noted that many other simulations were performed with similar results.

Smaller Sample Simulations

In smaller samples asymptotics do not always hold and in fact, extreme values of the statistic often result. Since samples of sizes 24, 36, and 48 are commonly used for bioequivalence studies [Snikeris, 1992], these are considered along with still other sample sizes, only some of which will be addressed here.

Very skewed densities often result, while in other cases the densities are fairly symmetric. Table 2 shows the results of

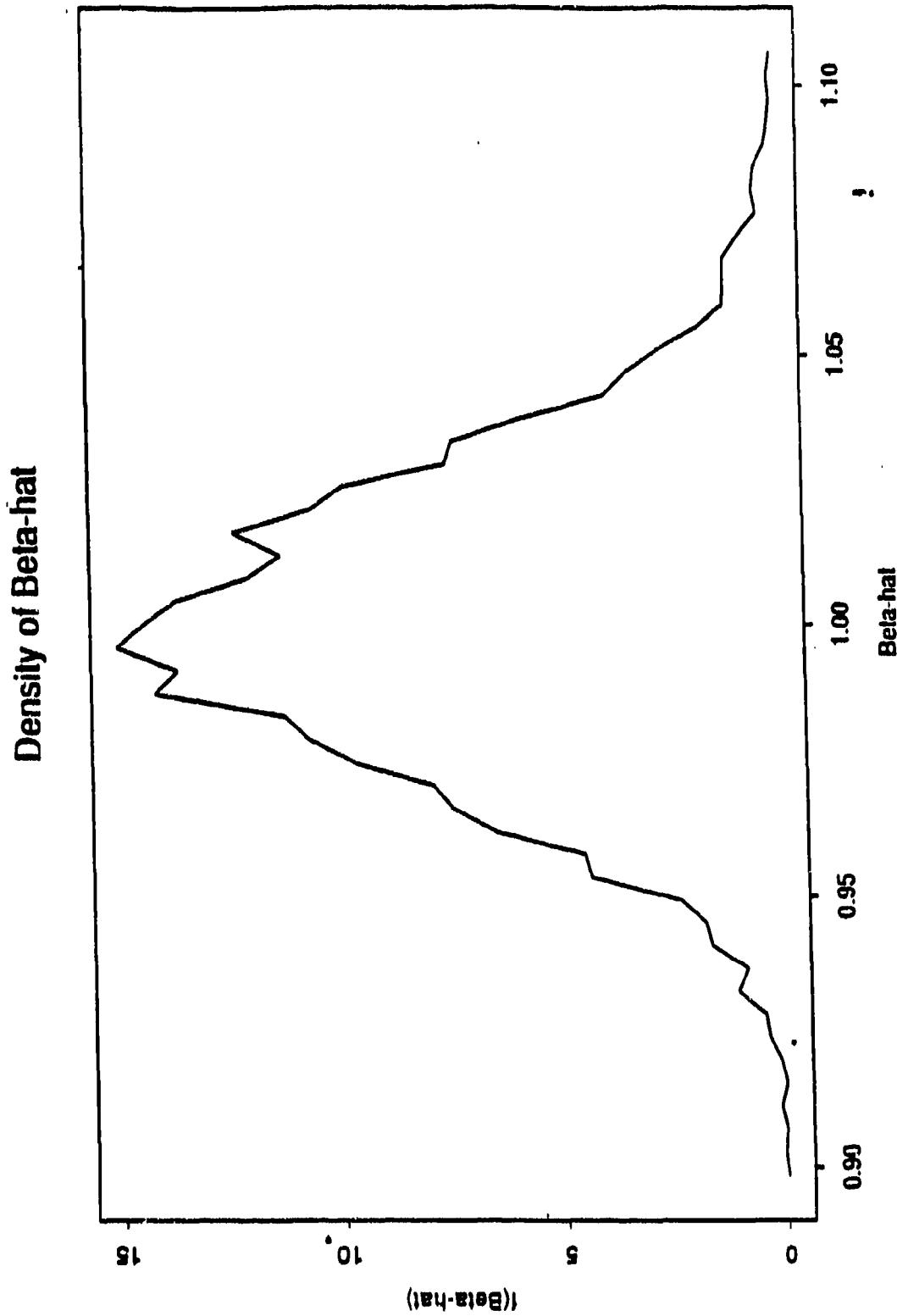


Figure 1: $n=100$, $\beta=1$, $X \sim N(5,4)$, $d \sim N(0,.25)$, $\varepsilon \sim N(0,.0625)$, mean=.9999, variance=.0008, $\lambda=4$

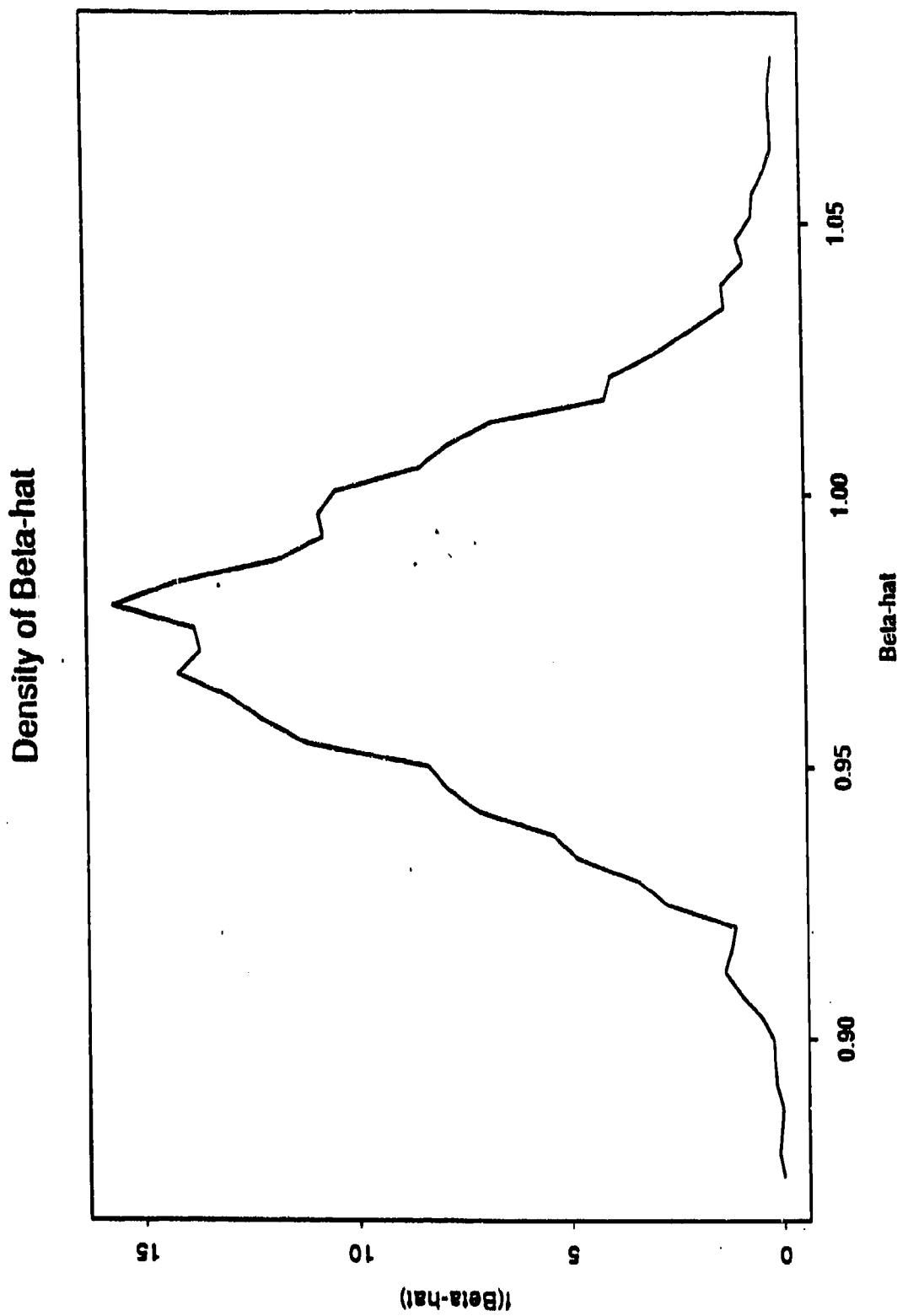


Figure 2: $n=100$, $\beta=1$, $X \sim N(5,4)$, $d \sim N(0,25)$, $c \sim N(0,0625)$, mean=.9764, variance=.0008, $\lambda=1$

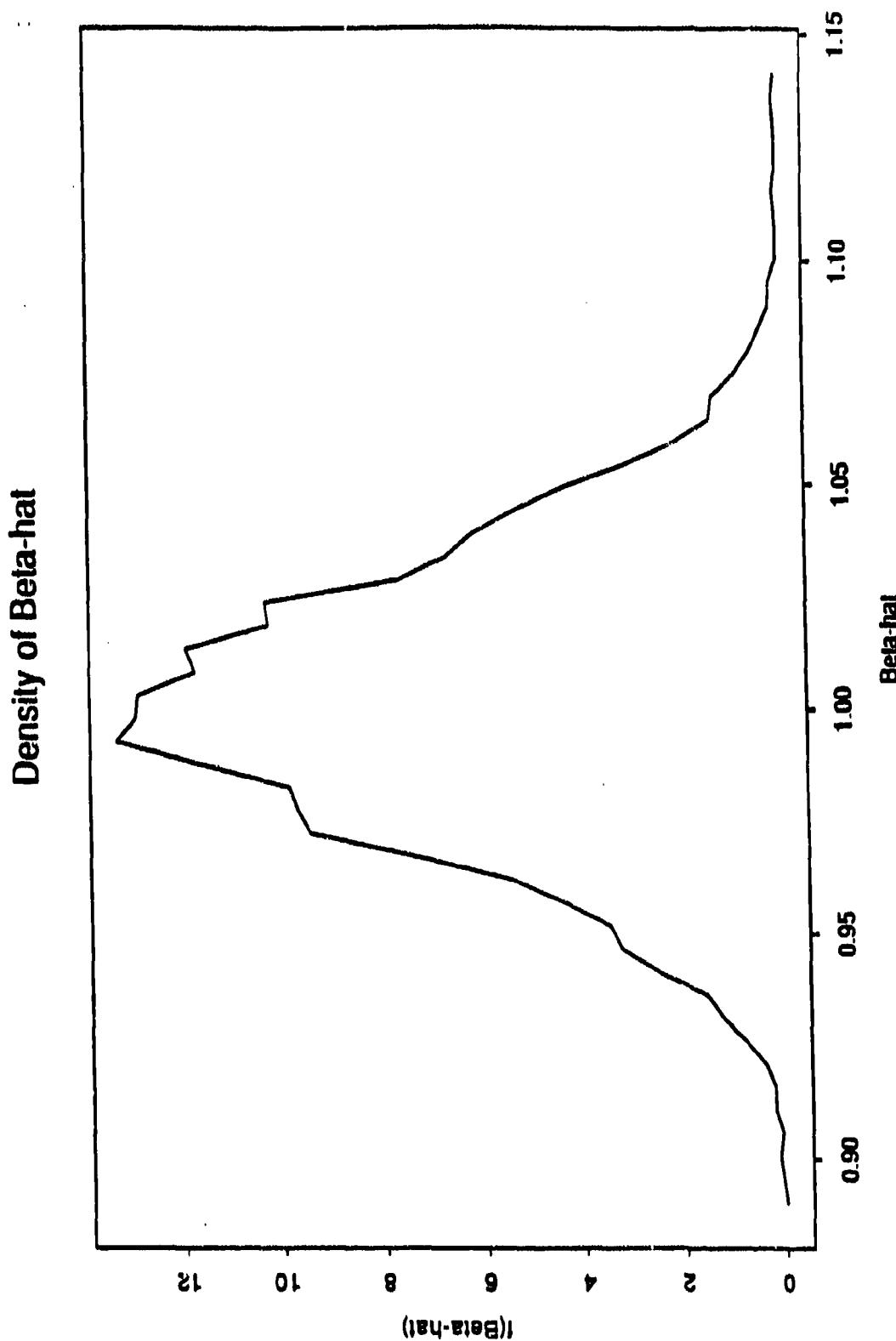


Figure 3: n=100, $\beta=1$, $X=(1:10,10)$, $d \sim N(0,25)$, $e \sim N(0,5)$, mean=1.0007, variance=.0010, $\lambda=.5$

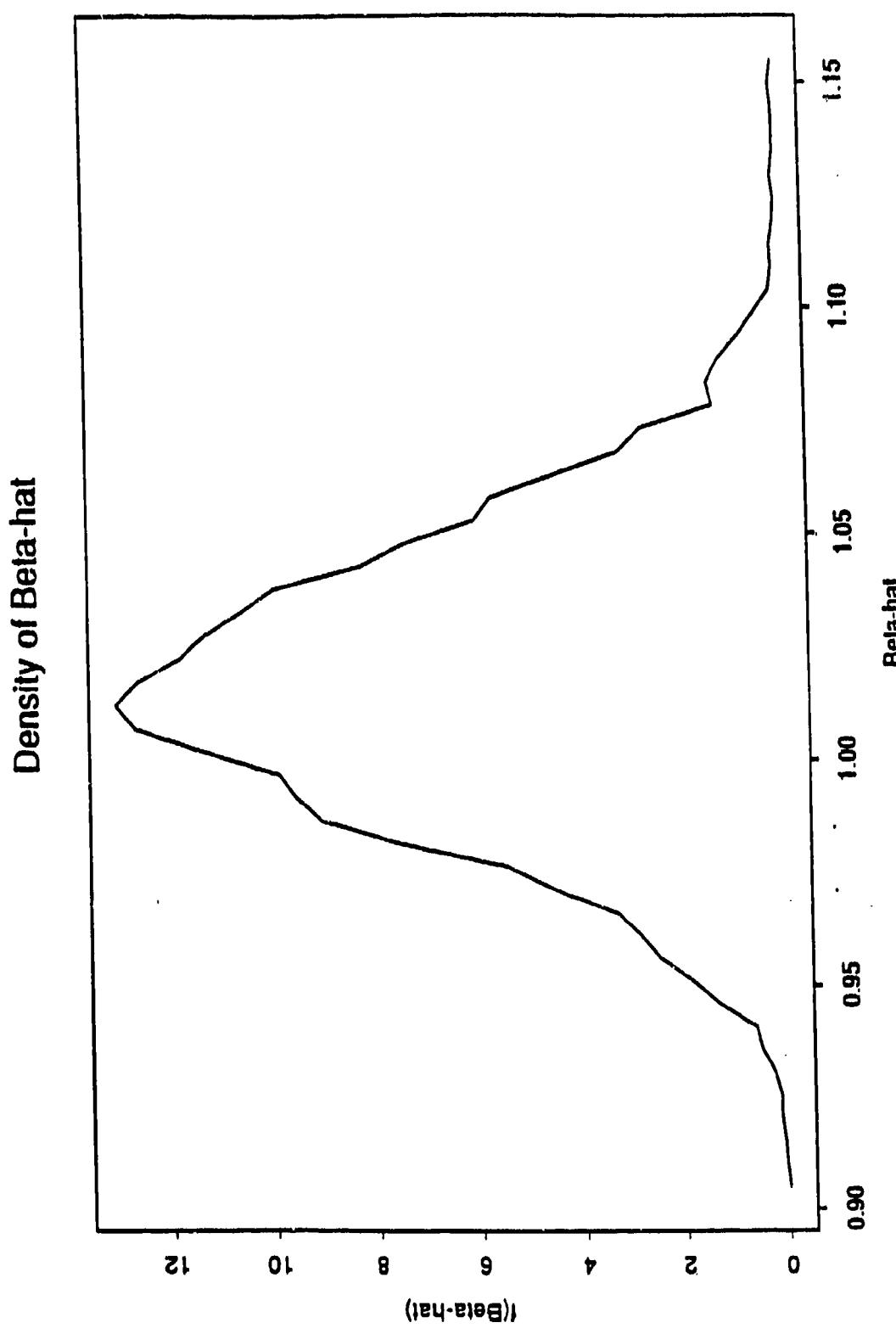


Figure 4: $n=100$, $\beta=1$, $X=(1:10,10)$, $d \sim N(0,25)$, $c \sim N(0,5)$, mean=1.0157, variance=.0010, $\lambda=1$

Table 1: Simulation Results: $n = 100, H_0 : \beta = 1, H_A : \beta \neq 1$

σ_e^2	σ_d^2	X	λ	mean $\hat{\beta}_p$	var $\hat{\beta}_p$	lower prob	upper prob	total	shape
.5	.25	1:10,10	.5	1.0007	.0010	.0190	.0260	.0450	symmetric
.5	.25	1:10,10	1	1.0157	.0010	.0057	.0750	.0807	symmetric
1	1	N(5.1)	1	1.0130	.0356	.0080	.0397	.0047	slight skew
.25	1	N(5.4)	4	1.0010	.0034	.0160	.0320	.0480	symmetric
.25	1	N(5.4)	1	.9105	.0029	.4003	.0003	.4006	symmetric
.0625	.25	N(5.4)	4	.9999	.0008	.0230	.0280	.0510	symmetric
.0625	.25	N(5.4)	1	.9764	.0008	.1383	.0037	.1420	symmetric
.25	.25	N(5.4)	1	.9995	.0013	.0230	.0270	.0500	symmetric
.25	.25	N(5,16)	1	.9996	.0003	.0250	.0230	.0480	symmetric
1	.25	1:10,10	.25	1.0008	.0016	.0230	.0270	.0500	symmetric
1	.25	1:10,10	1	1.0450	.0017	.0013	.0647	.0660	symmetric
4	.25	1:10,10	1	1.2400	.0067	.0000	.8600	.8600	symmetric
4	.25	1:10,10	.0625	.9982	.0054	.0230	.0270	.0500	symmetric
4	.25	1:10,10	16	1.4585	.0071	.0000	1.0000	1.0000	symmetric
4	.25	N(3.1)	16	5.2657	2.8979	.0000	.7273	.7273	skewed
4	.25	N(3.1)	1	4.2460	1.6350	.0000	.7237	.7237	skewed
.25	4	N(3.1)	1	.2505	.0034	1.0000	.0000	1.0000	symmetric
4	4	1:10,10	1	1.0080	.0128	.0150	.0350	.0500	symmetric
.25	.5	1:10,10	1	.9910	.0009	.0520	.0110	.0630	symmetric
.25	.5	1:10,10	2	1.0040	.0009	.0210	.0320	.0530	symmetric

several simulations when $n = 24$. Although one hopes that the error variances are not larger than the variance of X here, a few simulations indicate the general trend of how these variances greatly affect the density of $\hat{\beta}$. From the table we can see that whenever σ^2 is smaller than one or both of the error variances (in the case $\beta = 1$ it does not matter) very extreme estimates result and thus the variance of $\hat{\beta}$ is very large (larger than it should be according to Fuller). In fact the sampling distribution of $\hat{\beta}$ is very skewed to the left or right, as Figures 5 and 6 show. For smaller error variances the sample correlation tends to be larger, and therefore we do not observe as many large values for the estimator. Figure 7 displays a more stable density. This difference is also evident in the probabilities of Table 2. Simulation results for $n = 36$ and $n = 48$ appear in Tables 3 and 4 respectively. Comparing these tables, it is plain to see that for any given situation, the mean is closer to the true value as the sample size increases from 24 to 36 to 48. Notice that the variances become smaller as we would expect. In many cases densities skewed for $n = 24$ become less skewed for $n = 36$, and still less for $n = 48$.

Table 2 : Simulation Results: $n = 24$, $H_0: \beta = 1$, $H_A: \beta \neq 1$

σ_a^2	σ_d^2	X	λ used	mean $\hat{\beta}$	var $\hat{\beta}$	lower prob	upper prob	total	shape
.25	.25	N (3,.25)	1	1.2360	11.6920				skew left
.25	.25	N (3,.5)	1	1.0390	.0833	.0013	.0367	.0380	skew rt
.25	.25	N (3,1)	1	1.0156	.0303	.0053	.0340	.0393	symmetric
.25	.25	N (3,1)	1	1.0570	.0296	.0050	.0430	.0480	symmetric
.25	.25	N (3,2)	1	1.0270	.0123	.0100	.0360	.0460	symmetric
1	1	N (3,1)	1	-33.9430	3668933	.0003	.0000	.0003	skew left
1	.25	N (3,1)	.25	1.0142	.0774	.0117	.0303	.0420	symmetric
1	.25	N (3,1)	1	1.5532	.6466	.0007	.0343	.0350	sym peak
1	.25	N (3,4)	.25	1.0045	.0160	.0150	.0240	.0300	symmetric
1	.25	N (3,4)	1	1.1080	.0198	.0000	.0960	.0960	skew left
.25	1	N (3,1)	1	.7085	.0428	.2367	.0040	.2407	skew left
.25	1	N (3,1)	4	1.0552	1.1092	.0003	.0030	.0033	skew left
.25	1	N (3,4)	1	.9186	.0145	.0637	.0093	.0730	symmetric
.25	1	N (3,4)	4	1.0146	.0191	.0030	.0387	.0417	symmetric
.25	.5	N (3,1)	2	1.0344	.0566	.0003	.0370	.0373	skew rt.
.25	.5	N (3,1)	1	.9020	.0395	.0367	.0163	.0530	symmetric
.25	.5	N (3,4)	2	1.0061	.0099	.0113	.0300	.0413	symmetric
.25	.5	N (3,4)	1	.9734	.0098	.0293	.0150	.0443	symmetric
.5	.25	N (3,.25)	1	1.6814	2428.0910	.0001	.0030	.0031	skew left
.5	.25	N (3,.25)	.5	1.1049	7.6676	.0027	.0060	.0087	sym peak
.5	.25	N (3,1)	1	1.1057	.0676	.0003	.0803	.0806	symmetric
.5	.25	N (3,1)	.5	1.0162	.0468	.0083	.0310	.0393	symmetric
.5	.25	N (3,4)	.5	1.0016	.0094	.0147	.0273	.0420	symmetric
.5	.25	N (3,4)	1	1.0362	.0102	.0037	.0517	.0554	symmetric
.5	1	N (3,1)	1	.8106	.1220	.0090	.0117	.0207	sym peak
.5	1	N (3,1)	2	1.1207	2.9510	.0007	.0053	.0060	skew rt.

Density of Beta-hat

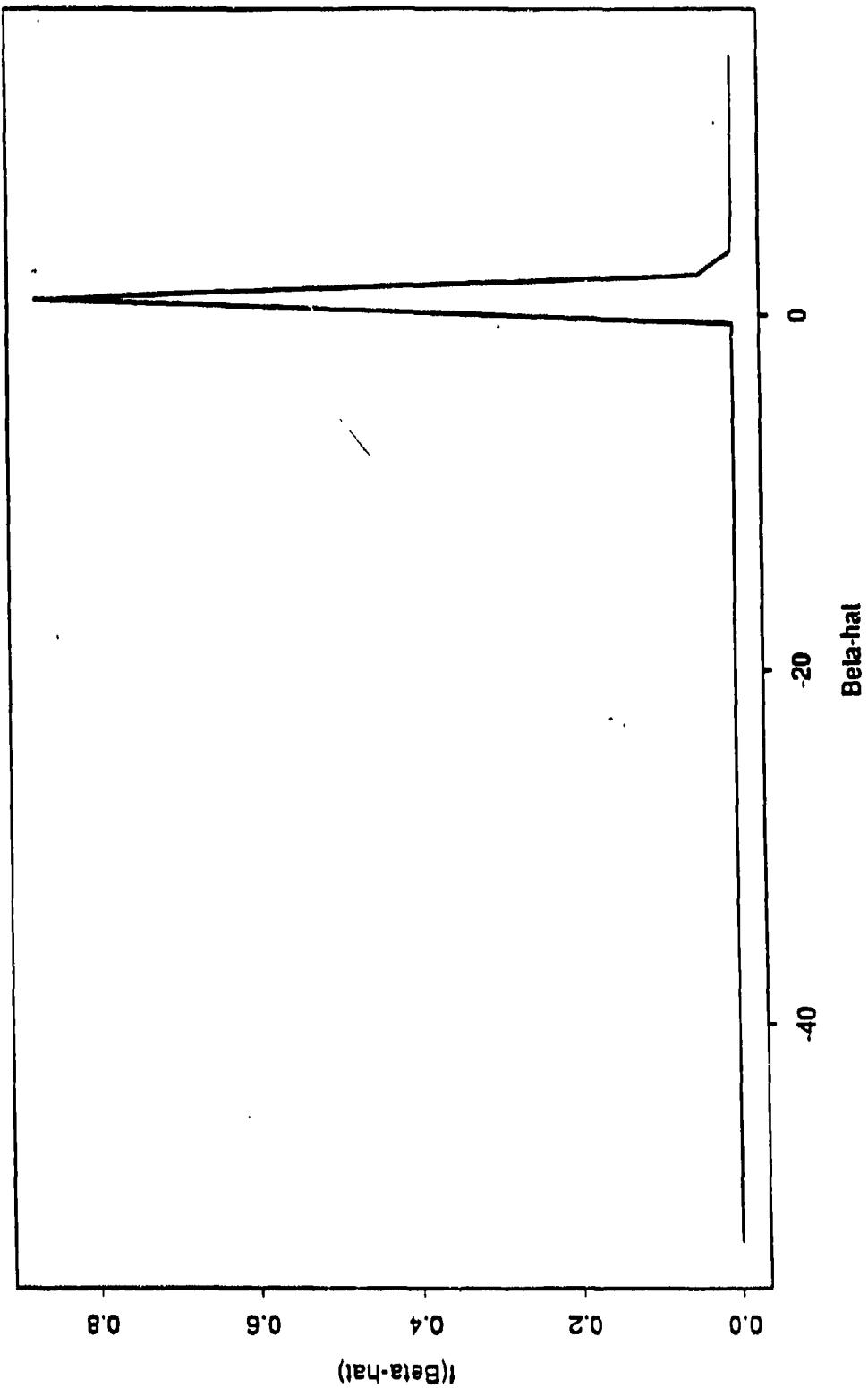


Figure 5 : n=24, $\beta=1$, $X \sim N(3,1)$, $d \sim N(0,1)$, $e \sim N(0,25)$, mean = 1.0552, variance = 1.1092, $\lambda=4$

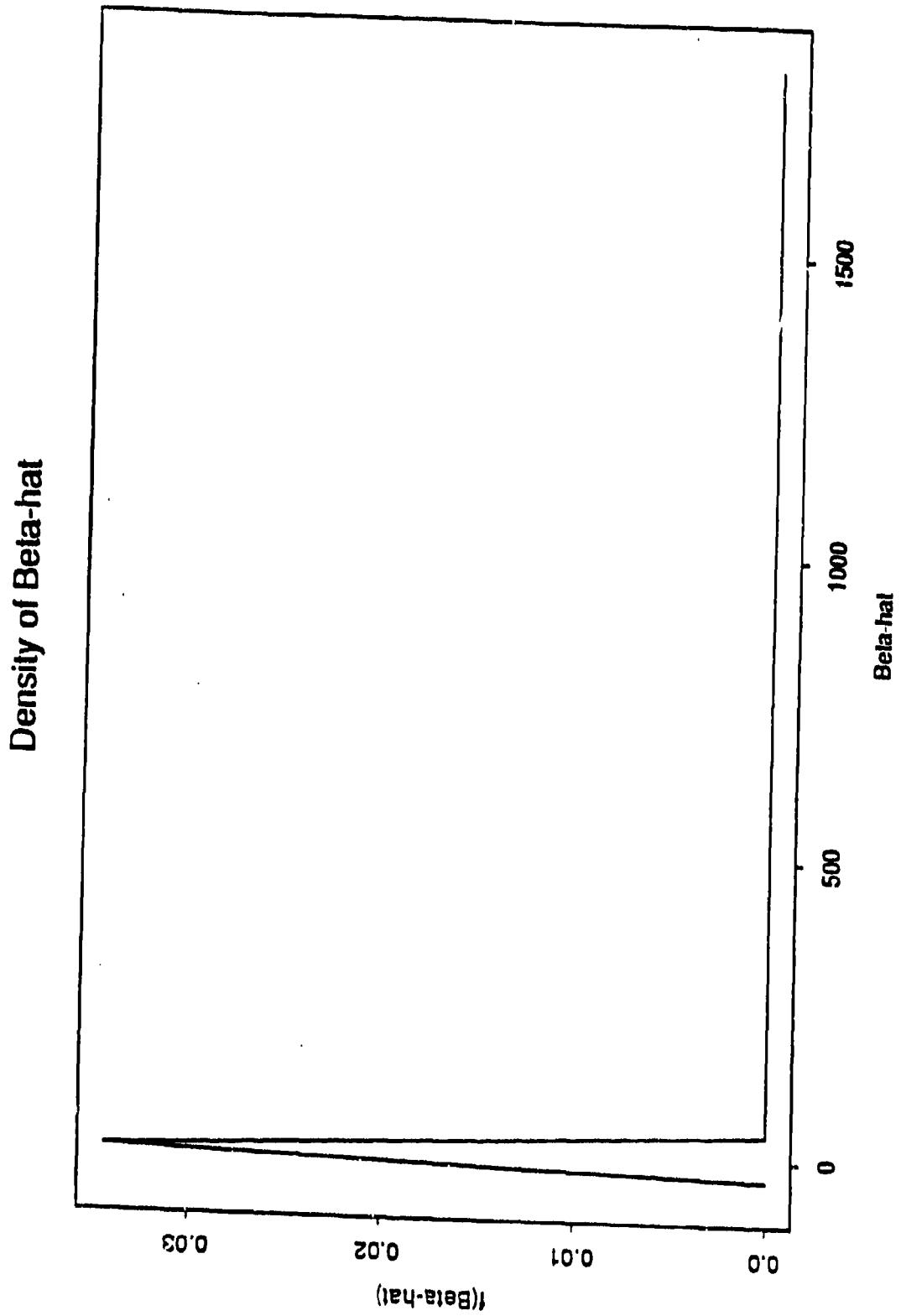


Figure 6: $n=24$, $\beta=1$, $X \sim N(3,1)$, $d \sim N(0,1)$, $e \sim N(0,1)$ mean=2.9114, variance=3153.962, $\lambda=1$

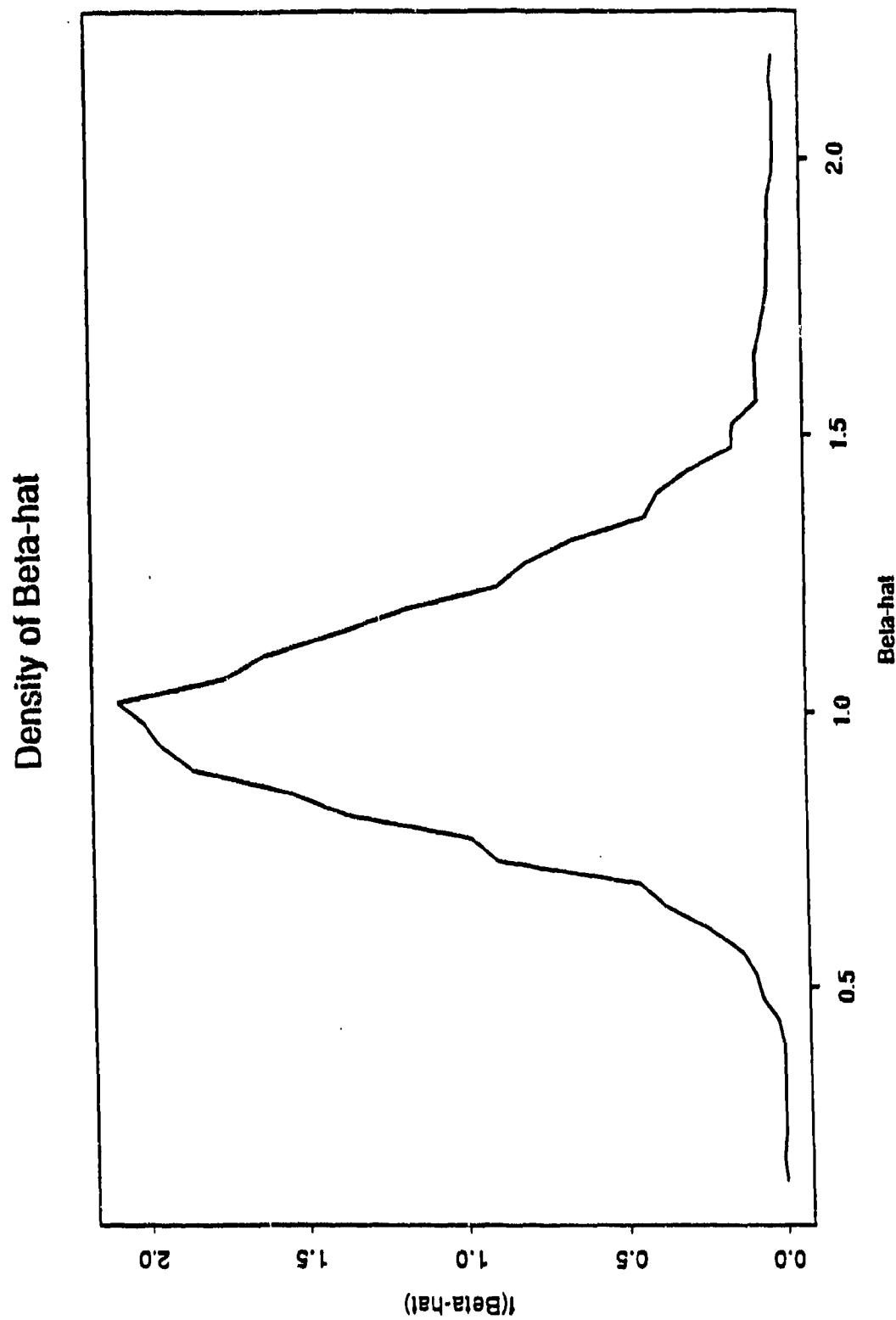


Figure 7 : $n=24$, $\beta=1$, $X \sim N(3,1)$, $d \sim N(0,25)$, $\epsilon \sim N(0,5)$ mean=1.0162, variance=.0468, $\lambda=.5$

Table 3 : Simulation Results: $n = 36$, $H_0: \beta = 1$, $H_A: \beta \neq 1$

σ_e^2	σ_d^2	X	λ used	mean $\hat{\beta}$	var $\hat{\beta}$	lower prob	upper prob	total	shape
.25	.25	N (3,.25)	1	1.0690	.2047	.0010	.0290	.0300	sym peak
.25	.25	N (3.1)	1	1.0110	.0175	.0100	.0370	.0470	symmetric
1	1	N (3.1)	1	1.0690	.2047	.0010	.0290	.0300	sym peak
1	1	N (3,.25)	1	.4436	3318.9860	.0030	.0020	.0050	peak
1	.25	N (3.1)	.25	1.0090	.0483	.0160	.0310	.0470	symmetric
1	.25	N (3.1)	1	1.4900	.1295	0	.2000	.2000	slight skew
1	.25	N (3.4)	1	1.1046	.0124	.0003	.1313	.1316	symmetric
1	.25	N (3.4)	.25	1.0040	.0106	.0190	.0290	.0480	symmetric
.25	1	N (3.1)	1	.7063	.0237	.4690	.0010	.4700	slight skew
.25	1	N (3.1)	4	1.0470	.0063	0	.0460	.0460	skewed
.25	1	N (3.4)	1	.9140	.0082	.1203	.0053	.1346	symmetric
.25	1	N (3.4)	4	1.0084	.0106	.0080	.0383	.0463	symmetric
.5	.25	N (3,.25)	1	1.8740	21.0020	.0060	.0070	.0130	sym peak
.5	.25	N (3,.25)	5	1.0970	1.4640	.0013	.0073	.0080	skew left
.5	.25	N (3.4)	.5	1.0027	.0060	.0147	.0290	.0437	symmetric
.5	.25	N (3.4)	1	1.0341	.0066	.0037	.0560	.0597	symmetric
.5	.25	N (3.1)	.5	1.0120	.0271	.0103	.0363	.0466	symmetric
.5	.25	N (3.1)	1	1.1550	.0377	.0007	.1000	.1007	symmetric
.25	.5	N (3.4)	2	1.0018	.0065	.0110	.0337	.0443	symmetric
.25	.5	N (3.4)	1	.9720	.0060	.0370	.0133	.0513	symmetric
1	.25	N (3,.25)	.25	1.1001	1.9390	.0030	.0083	.0113	sym peak
1	.25	N (3,.25)	1	2.6050	3979.8870	.0030	.0020	.0050	peak skew left

Table 4 : Simulation Results: $n = 48$ $H_0: \beta = 1$, $H_A: \beta \neq 1$

σ_e^2	σ_d^2	X	λ used	mean $\hat{\beta}$	var $\hat{\beta}$	lower prob	upper prob	total	shape
.25	.25	N (3,.25)	1	1.0406	.1014	.0020	.0330	.0350	skew left
.25	.25	N (3,1)	1	1.0060	.0125	.0140	.0300	.0440	symmetric
1	1	N (3,.25)	1	.7950	205.8400	.0047	.0027	.0074	skew left
1	1	N (3,1)	1	1.0406	.1014	.0020	.0330	.0350	skew right
1	1	N (3,4)	1	1.0056	.0125	.0140	.0300	.0440	symmetric
1	.25	N (3,.25)	1	2.9860	1735.5600	.0020	.0020	.0040	skew left
1	.25	N (3,.25)	.25	1.0470	.8793	.0017	.0110	.0217	skew left
1	.25	N (3,1)	.25	1.0070	.0340	.0143	.0257	.0400	symmetric
1	.25	N (3,1)	1	1.4800	.0875	.0000	.2000	.2000	skew right
1	.25	N (3,4)	.25	1.0010	.0072	.0180	.0210	.0390	symmetric
1	.25	N (3,4)	1	1.1028	.0090	0	.1670	.1670	symmetric
1	.25	1.8,6	1	1.0720	.0056	.0013	.1530	.0543	symmetric
1	.25	1.8,6	.25	1.0016	.0051	.0210	.0290	.0500	symmetric
.25	1	N (3,1)	1	.6970	.0171	.6503	.0003	.6506	symmetric
.25	1	N (3,1)	4	1.0270	.0307	.0010	.0450	.0460	skew left
.25	1	N (3,4)	1	.9132	.0062	.1700	.0030	.1730	symmetric
.25	1	N (3,4)	4	1.0883	.0078	.0057	.0340	.0397	symmetric
.5	.25	N (3,.25)	1	1.7510	134.8598	.0007	.0047	.0054	skew left
.5	.25	N (3,.25)	.5	1.0344	.1717	.0033	.0367	.0400	skew right
.5	.25	N (3,4)	.5	.9993	.0042	.0153	.0263	.0416	symmetric
.5	.25	N (3,4)	1	1.0348	.0048	.0033	.0657	.0690	symmetric
.25	.5	N (3,4)	2	1.0033	.0044	.0100	.0313	.0413	symmetric
.25	.5	N (3,4)	1	.9697	.0040	.0553	.0093	.0646	skew right

We can see from the simulation results that large values of $\hat{\beta}$ often result even when the true value of the parameter is one. Recalling the expression for $\hat{\beta}$, we can see that a low covariance (or correlation coefficient in the case of normality) will result in a large slope estimate. This low sample correlation often results from introducing large error variances. Another objective becomes developing a rule based on checking the sample correlation, and determining what to do with these extremes. An empirical rule was developed to detect and test for extremes. Once an estimate is detected as extreme, the correlation coefficient is tested for significance. If the correlation coefficient is not significantly different from zero, then the low correlation is the cause for the extreme estimate and therefore it is an unreliable estimate of β . Now the problem becomes what to consider extreme. The objective is to screen those large estimates that are due to low correlation. Therefore we do not want to detect as extreme an estimate that is large because β is large. Since we would probably not know the actual variances in a single sample problem, a conservative estimate of variance is needed to screen an estimate for extremeness. It is generally the case that the measurement error variances will be smaller than the variance of X . When testing $H_0: \beta=1$, the maximum variance is attained when $\sigma_e^2=\sigma_d^2=\sigma^2$. When this is the case a good approximation of the variance is $3/n$. After much screening using various estimates, it seems that what works best for detecting most of the extremes that resulted from low correlation is $\pm 5/(3/n)$. Thus any estimate that is more than five standard deviations from the hypothesized value of β will be tested for significant correlation.

In general, if it is believed that

$$\sigma_e^2 = \sigma_d^2 = k\sigma^2$$

then the estimate of variance is given by

$$\frac{2k + k^2}{n}$$

If $\beta \neq 1$, then the estimate of variance is

$$\frac{(\beta^2 + 1)k + k^2}{n}.$$

Table 5 provides various simulations when $n = 8$ and the screening rule is used. One can see that in practically all cases 100% of those detected as extreme were due to low correlation.

Table 5 : Simulation Results: $n = 8$, Truncation Results Using $\pm 5\sqrt{\frac{3}{8}}$

σ_e^2	σ_d^2	X	λ used	Fraction deleted due to low correlation
.25	.25	N (3,1)	1	$\frac{16}{18} = 1.0000$
1	1	N (3,1)	1	$\frac{16}{18} = 1.0000$
1	.25	N (3,1)	.25	$\frac{16}{18} = 1.0000$
1	.25	N (3,1)	1	$\frac{16}{20} = .9864$
1	.25	N (3,4)	.25	$\frac{1}{1} = 1.0000$
1	.25	N (3,4)	1	$\frac{1}{1} = 1.0000$
.25	1	N (3,1)	1	$\frac{16}{16} = 1.0000$
.25	1	N (3,1)	4	$\frac{16}{16} = 1.0000$
.25	1	N (3,4)	1	$\frac{1}{1} = 1.0000$
.25	1	N (3,4)	4	$\frac{1}{1} = 1.0000$
.25	.5	N (3,1)	1	$\frac{15}{18} = 1.0000$
.25	.5	N (3,1)	2	$\frac{15}{18} = 1.0000$
.25	.5	N (3,4)	2	$\frac{1}{1} = 1.0000$
.5	.25	N (3,1)	1	$\frac{10}{18} = 1.0000$
.5	.25	N (3,1)	.5	$\frac{10}{18} = 1.0000$
.5	.25	N (3,4)	1	$\frac{1}{1} = 1.0000$
.5	.25	N (3,2)	1	$\frac{16}{18} = .9474$
.5	.25	N (3,2)	.5	$\frac{1}{1} = 1.0000$

Application

We have the opportunity to analyze some real data using this estimation technique. The data come from two systems, called A and B, being compared for equivalence in handling specimens. Each system analyzed one hundred specimens, not once but twice consecutively. Therefore we have replicates for estimating lambda. Figure 8 gives a plot of system A measurements versus system B measurements. It shows a strong linear trend. We are interested, however, in whether the slope is significantly different from one, suggesting that the systems are not equivalent.

Analysis of Original Data

Normal plots for x and y indicate that the densities of x and y do not differ drastically from a normal one, and since $n=200$, we can assume asymptotics hold. We will analyze the two hundred pairs assuming that the ratio of error variances is one. The analysis results in

Scatterplot

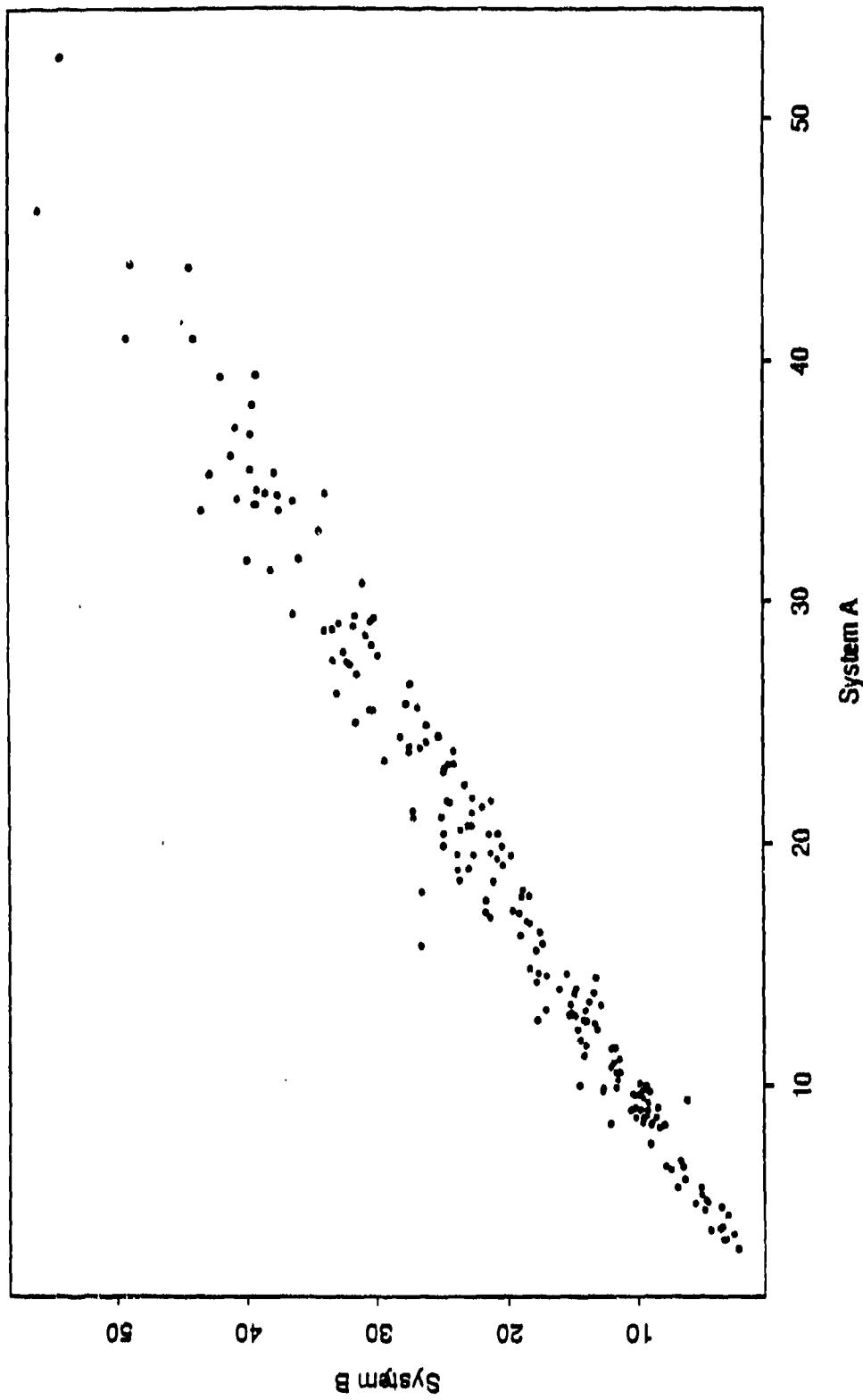


Figure 8 : Scatterplot of Specimen Measurements from System A vs System B, n=200

$$\hat{\beta} = 1.140371$$

$$\hat{V}_{\hat{\beta}} = .00627$$

so

$$z = 1.7724.$$

This does not suggest that we should reject the hypothesis $\beta = 1$. As we saw from the simulations, we will have large probabilities of Type I errors if we assume that $\lambda = 1$ when in fact it is not. Here we do not have to worry about a Type I error, but if underestimation occurs because we let $\lambda = 1$, we could be making a Type II error. Therefore, lambda will be estimated using Vormbrock and Helger's method of duplicates and we will use this estimate in computing $\hat{\beta}$. Using the duplicates to estimate the error variances, we find that

$$\sigma_d^2 = 1.840312$$

$$\sigma_e^2 = .463623$$

so

$$\hat{\lambda} = 3.9694.$$

Using this estimated ratio in the computation of $\hat{\beta}$ yields

$$\hat{\beta} = 1.149295$$

and

$$z = 1.88$$

These results do not differ much from the previous ones, but when $n = 200$ we would expect the density of $\hat{\beta}$ to be approximately normal and rather stable, particularly when the variance of X is large relative to the measurement error variances. The sample variances of x and y are 106.5541 and 138.0568, respectively. Recall that $\sigma_x^2 = \sigma_e^2 + \sigma_d^2$ and so $\sigma_e^2 \approx 106.5541 - 1.840312 = 104.71379$. This is very large relative to the measurement error variance. As we saw from the theory, the expected value of $\hat{\beta}$ will shift when $\lambda = 1$ is incorrectly assumed. We have only one estimate, yet we still can see that this second estimate is slightly larger and this suggests that when $\sigma_e^2 < \sigma_d^2$ and $\lambda = 1$ is used, there is underestimation on average. This may suggest in turn, that the true value of β is closer to the second estimate. However, we do not have a probabilistic statement of this fact since we are using only an estimate of λ and we are comparing only one estimate of β obtained by each approach. Although we have not studied the effects of λ on the density of $\hat{\beta}$ in this paper, it seems better to estimate lambda than to assume it is one. Therefore, when we select only one large sample and estimate β , we should use an estimate of λ rather than an assumption that $\lambda = 1$.

If we carry out a least squares analysis on the entire set of observations, the results are as follows:

$$\hat{\beta} = 1.12217$$

$$s_b = .013555$$

and

$$z = 9.01.$$

The standard error of this estimate is smaller than the standard error of the estimate obtained from the norm distance technique. The smaller variance (.00018) for least squares results from ignoring measurement error variability. Mandel takes the relationship between Deming's estimator and the least squares estimator to be

$$\hat{\beta} = LSb_{yx} \left(1 + \frac{s_d^2}{s_x^2 - s_d^2} \right).$$

This allows us to compare the variance of Deming's estimator with the least squares estimator by

$$\text{Var}[\hat{\beta}] = \text{Var}\left[LSb_{yx} \left(1 + \frac{s_d^2}{s_x^2 - s_d^2} \right)\right].$$

Analysis of Transformed Data

In scanning these data, it seems that the variances tend to increase as the measurement values increase. This can be seen slightly from the scatterplot. While this may not be large enough to worry about, we can deal with it by splitting the ranked data into two equal groups and testing for homogeneity of variance. While Bartlett's Test suggests a significant difference between the two variances, this test is often considered too sensitive. Cochran's test also suggests heterogeneity of variances. In light of this, we can try to achieve homogeneity of variance with a data transformation. According to Bartlett [1947], if $\sigma^2 = k^2m$ where m is the mean, then \sqrt{x} is a possible transformation. If this works to correct the variance problem, the variance of the transformed data will be $.25k^2$. For x , System A, the variance is 2.38 times the mean. The variance of the transformed data is 1.45 and $.25(2.38)^2 = 1.42$. Along with the scatterplot in Figure 9, this suggests that we now have homogeneity of variance for the transformed data. Analyzing the transformed data under the assumption that $\lambda = 1$ we have the following results:

$$\hat{\beta} = 1.1011$$

$$\text{Var}[\hat{\beta}] = .0054$$

so that

$$z = 1.38.$$

Transformed Data

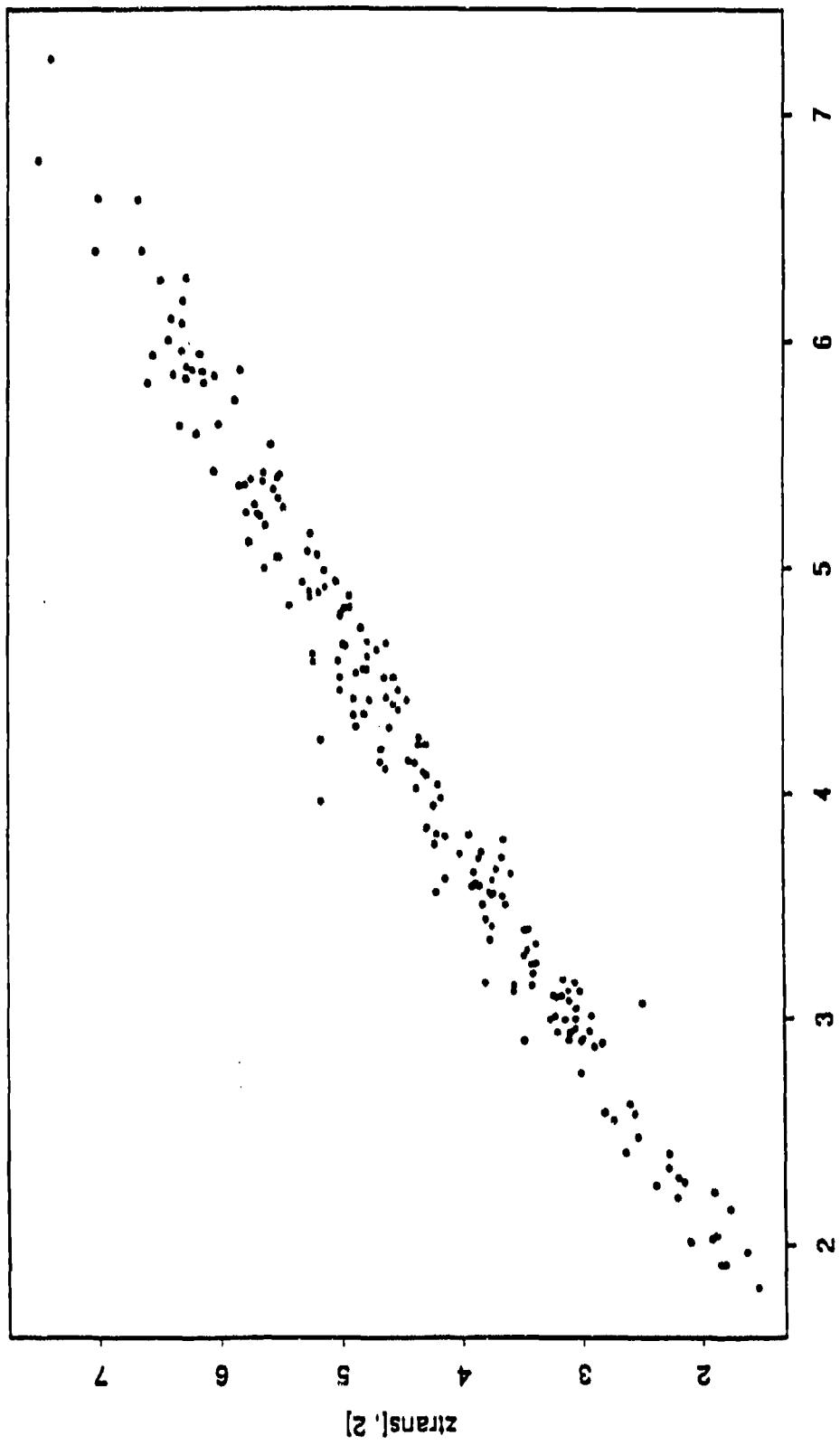


Figure 9 : Scatterplot of Specimen Measurements from System A vs System B, n=200

Using the previous estimate of $\hat{\lambda} = 4$, the results are:

$$\hat{\beta} = 1.1093$$

$$v[\hat{\beta}] = .0054$$

and hence

$$z = 1.49.$$

Next we can estimate lambda again for the transformed data. This time $\hat{\lambda} = 3.2096$. If we analyze the transformed data using $\hat{\lambda} = 3$, the results are:

$$\hat{\beta} = 1.1080$$

$$v[\hat{\beta}] = .0054$$

and

$$z = 1.47.$$

In none of these analyses do we find a significant difference between the two systems.

If we carry out another least squares analysis on the transformed data, the results are:

$$\hat{\beta} = 1.0855$$

$$\hat{V}[\hat{\beta}] = .00016$$

and thus

$$z = 6.81.$$

Once again we observe that the variance of the least squares estimator is smaller than that of the norm distance estimator. As a result we find significance when in fact there is no significant difference between the systems and we are led astray by the least squares method because it does not account for measurement error.

Recommendations

If measurement errors exist, then it is best to use an appropriate estimation technique making sure to account for these errors in both variables. As we have seen with the application, ignoring measurement error may well result in inaccurate conclusions.

If this technique of estimation is to be used, then it is best to select a large sample whenever possible. We have seen that when $n = 100$ the density of $\hat{\beta}$ is approximately normal for practically all typical situations, and even some less than typical. If smaller samples are necessary, then it is best to select values of X such that the spread of X is large relative to σ^2 and σ_d^2 . For samples of size 36, 48, and larger we have seen how the density reasonably resembles a normal one for cases when σ^2 was larger than σ^2 and σ_d^2 . However, for smaller samples there are many cases in which the density is far from resembling a normal or even a t distribution even when σ^2 is quite larger than σ^2 and σ_d^2 . For these cases we have no reliable test statistic.

It is best to sample replicates (repeated measures) whenever possible in order to estimate lambda. As we have seen, the often recommended use of $\lambda = 1$ adds additional bias to the estimate if $\lambda \neq 1$. However, using $\lambda = 1$ is better than completely disregarding the measurement error.

If one is testing a hypothesis and obtains an estimate that seems extremely different from that specified in the hypothesis, then one should test for a significant correlation. If the correlation coefficient is not significantly different from zero, then the estimate of β is an unreliable one. Hence, one should resample if possible.

REFERENCES

- [1] Acton, F.S. *Analysis of Straight Line Data*, 1959, Wiley and Sons, New York.
- [2] Adcock, R.J. 1877. "Note on the Method of Least Squares." *Analyst*, 4: 183.
- [3] Allen, H.V. 1938. *Stat. Research Memoirs*, 2: 60.
- [4] Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*, 2nd ed. 1984, John Wiley & Sons, New York, NY.
- [5] Barnett, V.D. 1967. "A Note on Linear Structural Relationships When Both Residual Variances Are Known." *Biometrika*, 54: 670-671.
- [6] Barnett, V.D. 1969. "Simultaneous Pairwise Linear Structural Relations." *Biometrics*, 25: 129-142.
- [7] Barnett, V. and Lewis, T. *Outliers in Statistical Data*, 2nd ed. 1984, John Wiley & Sons, New York, NY.
- [8] Bartlett, M.S. 1947. "The Use of Transformations." *Biometrics*, 3: 39-52.
- [9] Bartlett, M.S. 1949. "Fitting A Straight Line When Both Variables Are Subject to Error." *Biometrics*, 5: 207-212.
- [10] Berkson, J. 1950. "Are There Two Regressions?" *Journal of the American Statistical Association*, 45: 164-180.
- [11] Boggs, P.T., Donaldson, J.R., Byrd, R.H., and Schnabel, R.B. 1989. "ALGORITHM 676 ORDPACK: Software for Weighted Orthogonal Distance Regression." *ACM Transactions on Mathematical Software*, 15 (4): 348- 364.

- [12] Boggs, P.T., Spiegelman, C.H., Donaldson, J.R., Schnabel, R.B. 1988. "A Computational Examination of Orthogonal Distance Regression." *Journal of Econometrics*, 38: 169- 201.
- [13] Brace, R.A. 1977. "Fitting Straight Lines to Experimental Data." *American Journal of Physiology*, 233: R94-R99.
- [14] Cabut, S., and Springer, M.D. 1990. "Distribution of the Quotient of Noncentral Normal Random Variables." *Communications in Statistics: Theory and Methods*, 19: 1157-1168.
- [15] Cornbleet, P.J., and Gochman, N. 1979. "Incorrect Least Squares Regression Coefficients in Method Comparison Analysis." *Clinical Chemistry*, 25 (3): 432-438.
- [16] Craig, C.C, 1929. "The Frequency Function of y/x ." *Annals of Mathematics*, 30: 471-486.
- [17] Craig, C.C. 1942. "On Frequency Distributions of the Quaotient and of the Product of Two Statistical Variables." *American Mathematical Monthly*, 49: 26-32.
- [18] Cramer, H. *Mathematical Methods of Statistics*, 1957, Princeton University Press, Princeton.
- [19] Creasy, M.A. 1956. "Confidence Limits for the Gradient in the Linear Functional Relationship. *Journal of the Royal Statistical Society, B*, 18: 65-69.
- [20] Curtiss, J.H. 1941. "On the Distribution of the Quotient of Two Chance Variables." *Annals of Mathematical Statistics*, 12: 409-421.
- [21] Davies, R.B., and Hutton, B. 1975. "The Effect of Errors in the Independent Variables in Linear Regression." *Biometrika*, 62 (2): 383-391.
- [22] Deming, W.E. 1931. "The Application of Least Squares." *Philosophical Magazine (Series 7)*, 11: 146-158.

- [23] Deming, W.E. *Statistical Adjustment of Data*, 1943, John Wiley & Sons, New York, NY.
- [24] Dixon, W.J. and Massey, F.J. *Introduction to Statistical Analysis*, 1983, McGraw-Hill Book Company, New York, NY.
- [25] Dorff, M., and Gurland, J. 1961. "Estimation of the Parameters of a Linear Functional Relation." *Journal of the Royal Statistical Society, Series B*, 23: 160-170.
- [26] Draper, N., and Smith, H. *Applied Regression Analysis*, 2nd Edition, 1981, John Wiley & Sons, New York, NY.
- [27] Eisenhart, C. 1939. "The Interpretation of Certain Regression Methods and Their Use in Biological and Industrial Research." *Annals of Mathematical Statistics*, 10: 162-185.
- [28] Epstein, B. 1948. "Some Applications of the Mellin Transform in Statistics." *Annals of Mathematical Statistics*, 20: 370-379.
- [29] Erdelyi, A. *Tables of Integral Transforms*, Vol. 1, 1954, McGraw-Hill, New York, NY.
- [30] Feldman, U., Schneider, B., Klinkers, H., and Haekel, R. 1981. "A Multivariate Approach for the Biometric Comparison of Analytical Methods in Clinical Chemistry." *J. Clin. Chem. Clin. Biochem.*, 19: 121-137.
- [31] Fieller, E.C. 1932. "The Distribution of the Index in a Normal Bivariate Population." *Biometrika*, 24: 428-440.
- [32] Fieller, E.C. 1940. "The Biological Standardization of Insulin." *Journal of the Royal Statistical Society, Series B*, 7: 1-54.
- [33] Finney, D.J. 1938. "The Distribution of the Ratio of Estimates of the Two Variances in a Sample from a Normal Bi-variate Population." *Biometrika*, 30: 190- 192.

- [34] Finney, D.J. *Probit Analysis*, 3rd Edition, 1971, Cambridge University Press, New York, NY.
- [35] Fox, C. 1957. "Some Applications of Mellin Transforms to the Theory of Bi-variate Statistical Distributions." *Proceedings of the Cambridge Philosophy Society*, 53: 620- 628.
- [36] Fulks, W. *Advanced Calculus*, 2nd Edition, 1969, John Wiley & Sons, Inc., New York, NY.
- [37] Fuller, W.A. *Measurement Error Models*, 1987, John Wiley & Sons, New York, NY.
- [38] Fuller, W.A. 1980. "Properties of Some Estimators for the Errors-In-Variables Model." *The Annals of Statistics*, 8 (2): 407-422.
- [39] Geary, R.C. 1930. "The Frequency Distribution of the Quotient of Two Normal Variates." *Journal of the Royal Statistical Society*, 93: 442-447.
- [40] Geary, R.C. 1953. "Nonlinear Functional Relationship Between Two Variables When One Variable Is Controlled." *American Statistical Association Journal*, 94-103.
- [41] Goldschmidt, H.M.J., Griep, P.A.M., and de Jong, E.B.M. 1981. "The Quantitative Evaluation of Linearity in Method Comparison Studies Using a Simplified Method." *J. Clin. Chem. Clin. Biochem.*, 19: 681.
- [42] Graybill, F.A. *Linear Statistical Models*, 1961, McGraw-Hill Book Company, Inc., New York, NY.
- [43] Grubbs, F. E. 1950. "Sample Criteria for Testing Outlying Observations." *Annals of Mathematical Statistics*, 21: 27-58.
- [44] Gruber, M.H.J. *Regression Estimators A Comparative Study*, 1990, Academic Press, San Diego, CA.

- [45] Grynovicki, J.O. *Variance Component Estimation Diagnostics and Associated Distribution Theory for All Random and Mixed Designs: Applications to Repeated Measures*, 1989, Doctoral Dissertation, University of Delaware.
- [46] Gurland, J. 1953. "Distribution of Quadratic Forms and Ratios of Quadratic Forms." *Annals of Mathematical Statistics*, 24: 416-420.
- [47] Gurland, J. 1948. "Inversion Formulae for the Distribution of Ratios." *Annals of Mathematical Statistics*, 19: 228-237.
- [48] Halperin, M. 1961. "Fitting of Straight Lines and Prediction When Both Variables Are Subject to Error." *Journal of the American Statistical Association*, 56: 657-669.
- [49] Hastings, N.A.J. and Peacock, J.B. *Statistical Distributions*, 1975, John Wiley & Sons, New York, NY.
- [50] Hathaway, N.R. 1980. "Some Comments of Linear Regression Analysis by the Method of W.E. Deming." *Clinical Chemistry*, 26 (10): 1511.
- [51] Hocking, R.R. *The Analysis of Linear Models*, 1985, Brooks/Cole Publishing Company, Monterey, CA.
- [52] Hodges, S.D., and Moore, P.G. 1972. "Data Uncertainties and Least Squares Regression." *Applied Statistics*, 21: 185-195.
- [53] Housner, G.W., and Brennan, J.F. 1948. "The Estimation of Linear Trends." *Annals of Mathematical Statistics*, 19: 380-388.
- [54] Huntington, E.V. 1939. "Frequency Distribution of Product and Quotient." *Annals of Mathematical Statistics*, 10: 195-198.
- [55] Jessop, W.N. 1952. "One Line or Two?" *Applied Statistics*, 2: 131-137.
- [56] Keeping, E.S. 1956. "Note on Wald's Method of Fitting a Straight Line When Both Variables Are Subject to Error." *Biometrics*, 12: 445-448.

- [57] Kendall, M.G. 1951. "Regression, Structure, and Functional Relationship. Part I" *Biometrika*, 38: 11-25.
- [58] Kendall, M.G. 1952. "Regression, Structure, and Functional Relationship. Part II." *Biometrika*, 39: 96-108.
- [59] Kendall, M.G. and Stuart, A. *The Advanced Theory of Statistics*, Vol.II, 3rd Edition, 1973, Hafner Publishing Company, New York, NY.
- [60] Klefer, J. 1964. "Book Review: Kendall and Stuart, The Advanced Theory of Statistics, Vol. 2, 1961." *Annals of Mathematical Statistics*, 35: 1371-1380.
- [61] Kinnison, R.R. *Applied Extreme Value Statistics*, 1985, Macmillan Publishing Company, New York, NY.
- [62] Klepper, S. 1988. "Regressor Diagnostics for the Classical Errors-in-Variables Model." *Journal of Econometrics*, 37: 225-250.
- [63] Klepper, S. and Leamer, E.E. 1984. "Consistent Sets of Estimates for Regressions with Errors in All Variables." *Econometrica*, 52(1): 163-183.
- [64] Klepper, S. and Stapleton, D.C. 1987. "Consistent Sets of Estimates for Restricted Regressions with Errors in All Variables." *International Economic Review*, 28 (2): 445-457.
- [65] Kummel, C.H. 1879. "Reduction of Observation Equations Which Contain More Than One Observed Quantity." *The Analyst*, VI (4): 97-105.
- [66] Laurer, G.N. and Han, C.P. 1972. "Joint Distribution Function of Certain Ratios of Chi-Square Variates." *Journal of Statistical Computation and Simulation*, 1: 255-260.
- [67] Levine, D. 1985. "The Sensitivity of MLE to Measurement Error." *Journal of Econometrics*, 28: 223-230.
- [68] Lindley, D.V. 1947. "Regression Lines and the Linear Functional Relationship." *Journal of the Royal Statistical Society*, 9: 218-244.

- [69] Lindley, D.V. 1953. "Estimation of a Functional Relationship." *Biometrika*, 40: 47-49.
- [70] Lloyd, P.H. 1978. "A Scheme for the Evaluation of Diagnostic Kits." *Annals of Clinical Biochemistry*, 15: 136-145.
- [71] Lybanon, M. 1984. "Comment on 'Least Squares When Both Variables Have Uncertainties'." *American Journal of Physics*, 52 (3): 276-278.
- [72] Madansky, A. 1959. "The Fitting of Straight Lines When Both Variables Are Subject to Error." *Journal of the American Statistical Association*, 54: 173-205.
- [73] Mandel, J. *The Statistical Analysis of Experimental Data*, 1964, Interscience Publishers Division of Wiley and Sons, New York, NY.
- [74] McNulty, F. 1972. "Quadratic Form Distributions Associated with Special Functions." *Sankhya: Series B*, 34: 21-26.
- [75] McNulty, F., and Tomsky, J. 1972. "Some Properties of Special-Function, Bivariate Distributions." *Sankhya: Series B*, 34: 251-264.
- [76] Merrill, A.S. 1928. "Frequency Distribution of an Index When Both the Components Follow the Normal Law." *Biometrika*, 20A: 53-63.
- [77] Michotte, Y. "Evaluation of Precision and Accuracy - Comparison of Two Procedures." In: *Evaluation and Optimization of Laboratory Methods and Analytical Procedures* (Massart, D.L., Dijkstra, A., and Kaufmann, L. eds.), 1978, New York, NY.
- [78] Mood, A.M., Graybill, F.A., and Boes, D.C. *Introduction to the Theory of Statistics*, 3rd Edition, 1974, McGraw-Hill, Inc., New York, NY.
- [79] Moran, P.A.P. 1971. "Estimating Structural and Functional Relationships." *Journal of Multivariate Analysis*, 1: 232-255.

- [80] Morgan, W.A. 1939. "A Test for the Significance of the Difference Between the Two Variances in a Sample from a Normal Bivariate Population." *Biometrika*, 31: 13-19.
- [81] Mowchan, W.L. *Monte Carlo Investigation of the Probability Distributions of Dixon's Criteria for Testing Outlying Observations*, 1967, Master's Thesis, University of Delaware.
- [82] Myers, R.H., and Milton, J.S. *A First Course in the Theory of Linear Statistical Models*, 1991, PWS-KENT Publishing Company, Boston, MA
- [83] Nair, K.R. 1984. "On the Behavior of Some Estimators from Probability Plots." *Journal of the American Statistical Association*, 79: 823-831.
- [84] Nair, K.R., and Banerjee, K.S. 1942. "A Note on Fitting of Straight Lines If Both Variables Are Subject to Error." *Sankhya*, 6: 331.
- [85] Nair, K.R., and Shrivastava, M.P. 1942. "On a Simple Method of Curve Fitting." *Sankhya*, 6: 121-132.
- [86] Neyman, J. 1951. "Existence of Consistent Estimates of the Directional Parameter in a Linear Structural Relation Between Two Variables." *Annals of Mathematical Statistics*, 22: 496-512.
- [87] Northam, B.E. 1981. "Whither Automation?" *Ann. Clin. Biochem.*, 18: 189-199.
- [88] Orear, J. 1982. "Least Squares When Both Variables Have Uncertainties." *American Journal of Physics*, 50 (10): 912-916.
- [89] Orear, J. 1984. "Erratum for Least Squares When Both Variables Have Uncertainties." *American Journal of Physics*, 52 (3): 278.
- [90] Papoulis, A. *Probability, Random Variables, and Stochastic Processes*, 1965, McGraw Hill, Inc., New York, NY.

- [91] Passing, H. and Bablok, W. 1983. "A New Biometrical Procedure for Testing the Equality of Measurements from Two Different Analytical Methods." *J. Clin. Chem. Clin. Biochem.*, 21: 709-720.
- [92] Pearson, K. 1901. "On Lines and Planes of Closest Fit to Systems of Points in Space." *Philosophical Magazine*, 2: 559-573.
- [93] Pierce, D.A., Stram, D.O., Vaeth, M., and Schafer, D.W. 1992. "The Errors-in-Variables Problem: Considerations Provided by Radiation Dose-Response Analyses of the A-Bomb Survivor Data." *Journal of the American Statistical Association*, 87: 351-359.
- [94] Reed, B.C. 1989. "Linear Least-squares Fits with Errors in Both Coordinates." *American Journal of Physics*, 57 (7): 642-646.
- [95] Reed, I.S. 1944. "The Mellin Type of Double Integral." *Duke Mathematics Journal*, 11: 565-572.
- [96] Schall, R.F. Jr., Kern, C.W., and Tenoso, H.J. 1980. "Test Data Matrix and Results for Linear Regression Analysis by Method of W.E. Deming." *Clinical Chemistry*, 26: 352.
- [97] Schmalensee, R. and Joskow, P.L. 1986. "Estimated Parameters as Independent Variables." *Journal of Econometrics*, 31: 275-305.
- [98] Schneeweis, H. 1976. "Consistent Estimation of a Regression with Errors in the Variables." *Metrika*, 23: 101-115.
- [99] Scott, E.L. 1950. "Note on Consistent Estimates of the Linear Structural Relation Between Two Variables." *Ann. Math. Statist.*, 21: 284-288.
- [100] Seber, G.A.F. *Linear Regression Analysis*, 1977, John Wiley & Sons, New York, NY.
- [101] Smith, M.D. 1989. "On the Expectation of a Ratio of Quadratic Forms in Normal Variables." *Journal of Multivariate Analysis*, 31: 244-257.

- [102] Smith, D.S., Pourfarzaneh, M., and Kamel, R.S. 1980. "Linear Regression Analysis by Deming's Method." *Clinical Chemistry*, 26(7): 1105-1106.
- [103] Sneddon, I.N. *Fourier Transforms*, 1951, McGraw Hill, New York, NY.
- [104] Snedecor, G.W., and Cochran, W.G. *Statistical Methods*, 1967, Iowa State University Press, Ames, IO.
- [105] Snikeris, A.C. 1992. *Bioequivalence As a Linear Structural Relationship: Estimation of Formulation Intersubject and Intrasubject Variability and a Two-Step Method for Assessing Bioequivalence*. Doctoral Dissertation. University of Delaware.
- [106] Sokal, R.R. and Rohlf, F.J. *Biometry*, 1969, W.H. Freeman and Company, San Francisco.
- [107] Springer, M.D. *The Algebra of Random Variables*, 1979, John Wiley & Sons, New York, NY.
- [108] Subrahmanian, K. "On Some Applications of Mellin Transforms to Statistics: Dependent Random Variables." *SIAM Journal on Applied Mathematics*, 19: 658-662.
- [109] Tietjen, G.L. *A Topical Dictionary of Statistics*, 1986, Chapman and Hall, New York, NY.
- [110] Tingey, H.B. "Modeling Gunfire: A Problem in Maximum Likelihood Estimation." ARO Report 92-2 Proceedings of the Thirty Seventh Conference on The Design of Experimentation Army Research, Development and Testing Waterways Experiment Station. Vicksburg, MS.
- [111] Van Uven, M.J. 1939. "Adjustment of a Ratio." *Annals of Eugenics*, 9: 181-202.
- [112] Villegas, C. 1961. "Maximum Likelihood Estimation of a Linear Functional Relationship." *Annals of Mathematical Statistics*, 32: 1048-1062.

- [113] Vormbroek, R. and Helger, R. "Use of Duplicates for Method Comparison Analysis Using Deming's Procedure."
- [114] Wakkers, P.J.M., Hellendoorn, H.B.A., Op De Weegh, G.J., and Heerspink, W. 1975. "Applications of Statistics in Clinical Chemistry: A Critical Evaluation of Regression Lines." *Clinica Chemica Acta*, 64: 173-184.
- [115] Wald, A. 1940-1941. "The Fitting of Straight Lines When Both Variables Are Subject to Error." *Ann. Math. Statist.*, 11: 284-300.
- [116] Weisbrod, I.M. *Statistics for the Clinical Laboratory*, 1985, J.B. Lippincott Co.
- [117] Westgard, J.O. and Hunt, M.R. 1973. "Use and Interpretation of Common Statistical Tests in Method Comparison Studies." *Clinical Chemistry*, 19: 49-57.
- [118] Widder, D.V. *Advanced Calculus*, Second Edition, 1961, Prentice Hall, Inc., Englewood Cliffs, NJ.
- [119] Winsor, C.P. 1946. "Which Regression?" *Biometrics Bulletin*, 2 (6): 101-109.
- [120] Wishart, J. and Bartlett, M. S. 1932. "The Distribution of Second Order Moment Statistics in a Normal System." *Proceedings of the Cambridge Philosophy Society*, 28: 455-459.
- [121] Wolfowitz, J. 1957. "The Minimum Distance Method." *Annals of Mathematical Statistics*, 28: 75-88.
- [122] Zucker, L.M. 1947. "Evaluation of Slopes and Intercepts of Straight Lines." *Human Biology*, 19: 231-259.

**DETERMINATION
OF DESIRED DESIGN AND OPERATIONAL CHARACTERISTICS OF THE
SMALL AREA CAMOUFLAGE COVER
BY GROUND TROOPS**

George Anitole and Ronald L. Johnson
U.S. Army Belvoir Research, Development and Engineering
Center, Fort Belvoir, Virginia 22060-5606

Christopher J. Neubert
U.S. Army Material Command
Alexandria, Virginia 22333-0001

Abstract

The goal of this study was to evaluate the design and operational characteristics of the Small Area Camouflage Cover (SACC), when used by ground soldiers in a tactical environment. The SACC is designed to conceal individuals, small size equipment and fighting positions. Fifty-nine reserve soldiers from the 187th Infantry Brigade, Fort Devens, MA were given the SACC to be incorporated in their training at the Canadian Forces Training Center, New Brunswick, Canada. They were given instructions on the use of the SACC before the start of the maneuvers. Ten days later, at the conclusion of the exercises, the soldiers were presented a questionnaire/survey of twenty-two SACC design and operational characteristics, from which they made individual paired comparisons to determine which of the characteristics were most important. Each characteristic was independently evaluated by each soldier twenty-one times for a total of two-hundred-and thirty-one paired comparisons. A parametric statistical analysis was conducted upon the results of the questionnaire/survey, and six statistically significant ($\alpha = 0.05$) groups of characteristics were determined, with the groups defining a continuum from most to least preferred. This study joined the expertise of an engineer, statistician, and psychologist, and gave the investigators a unique testing challenge of obtaining hard empirical data from a subjective operational field test.

1.0 SECTION 1 - INTRODUCTION

The Small Area Camouflage Cover (SACC) is a continuation of a program begun in 1986 to develop an Individual Camouflage Cover (ICC). The original program was sponsored by FORSCOM and resulted in prototype arctic, woodland, and desert ICCs. The SACC development sponsored by the Soldier Enhancement Program, extended the original design by developing a more effective, durable, and versatile camouflage capability including a cover for tropical backgrounds.

The SACC is designed to provide protection from visual, near-infrared, and radar observation, and in the arctic version, also provides protection from ultra-violet detection. The SACC will conceal individual troops, or can be attached together for use over weapon emplacements, fighting positions, and supply caches.

In designing the SACC, certain characteristics, such as color/textured match with background, lightweight, low shine, and durability were used as guidelines. In addition, 18 other technical and operational design characteristics have been determined for the SACC. In order to finalize the current development and to put emphasis on the most important requirements in future SACC designs, the design characteristics needed to be ranked in their order of importance. To determine the order of importance, a troop test was conducted using soldiers from the 3rd Bn, 35th Inf, 187th Bde from Fort Devens, MA. The test was conducted during exercise Nordic Shield II at the Canadian Forces Combat Training Center near Gagetown, New Brunswick, Canada in August 1992.

2.0 SECTION II - EXPERIMENTAL DESIGN

2.1 Test Target

The test target was a woodland SACC developed at Fort Belvoir, VA. It is reversible with a two-color green pattern on one side and a four-color brown pattern on the reverse. The SACC is made of incised, vinyl coated nylon scrim, weighs less than 518 grams (18 ounces), and is small enough 2.76 x 1.77 meters (4'6" x 7') to be fitted into the pocket of a soldiers uniform. The SACC also has near-infrared and radar camouflage characteristics.

2.2 Test Site

The test site was located at the Canadian Forces Combat Training Center, New Brunswick Canada. The area represented a typical north temperate zone woodland environment, consisting of large open fields of grass land and large tracks of dense coniferous and deciduous forests.

2.3 Test Subjects

A total of 59 reserve troops from the 3rd Bn, 35th Inf, 187th Inf Bde, Fort Devens, MA participated in the study. The troops consisted of enlisted personnel, non-commissioned officers, and commissioned officers.

2.4 Test Procedure

The troops were issued 25 SACCs to be used during their tactical exercise. These SACCs were eventually used by 59 soldiers. At the conclusion of the exercise, a questionnaire/survey Table 1, was given to the troops, in which they made individual comparisons between 22 technical and operational design characteristics. The procedure involved comparing each characteristic to all the others, a pair at a time. The task was to decide which of each pair of characteristics was the most important. Each soldier made a total of 231 paired comparisons, with each characteristic being evaluated 21 times. The comparison was made as follows: If the evaluator preferred the column characteristic over the row characteristic, in Table 1, a one was placed in the box. If the row characteristic was preferred over the column characteristic, a zero was placed in the box. The ones for the row of each characteristic were added along with to the number of zeros for the corresponding column to produce a total acceptance score. The larger the acceptance score the more important the evaluator felt about the characteristic. The soldier was instructed not to skip any comparisons.

3.0 SECTION III - RESULTS

The soldiers answering the questionnaire/survey produced sufficient data to enable a ranking of the subject design and operational SACC characteristics, from most desired to least desired. Table 2 shows the descriptive data with the sample size, mean, standard deviation, standard error, and the 95 percent confidence interval for each characteristic. Table 3 contains the analysis of variance ^{1/ 2/}, while Table 4 shows the Scheffe's Multiple Range Procedure which separates the operational characteristic into statistically different groups. The higher the mean, the greater the preference the evaluator had for the characteristic. In all cases, a code letter is used for the characteristic (see Table 1).

TABLE I

SACC DESIGN AND OPERATIONAL CHARACTERISTICS

CODE LETTER	CHARACTERISTIC
-------------	----------------

- A - Woodland, arctic, tropic and desert SACCs should be reversible (i.e., another seasonal or background color on back)
 - B - Must not make noise when being handled
 - C - Shelf life of 10 years
 - D - SACC should be no bigger than 4 1/2 by 7 feet
 - E - Must be non-flammable
 - F - Weight does not hinder transport
 - G - Does not interfere with vision
 - H - Must not shine or glare
 - I - Must be easily carried
 - J - Must be able to be joined with other SACC units to place over larger objects such as HMMWV or gun position
 - K - Must be fungus resistant
 - L - Does not interfere with hand movement
 - M - Offers protection against visual detection (matches background color, texture, breaks up outline of hull and tracks)
 - N - Must not snag
 - O - Offers protection against radar detection
 - P - Must not greatly increase the body temperature of a soldier under the SACC
 - Q - Offers protection against thermal detection
 - R - Field life (durability, color fading, etc.) of 60 days
 - S - Must not present a health hazard
 - T - Must be easy to use
 - U - Offers protection against near-infrared detection
 - V - Must be easily carried by M1 Tank or Bradley

TABLE 2
MEAN PREFERENCE DESCRIPTIVE DATA FOR
THE SACC DESIGN AND OPERATIONAL CHARACTERISTICS

CHARAC- TERISTIC	SAMPLE SIZE	MEAN	STANDARD	STANDARD	95% CONFIDENCE	
			DEVIATION	ERROR	INTERVAL	LOWER LIMIT
A	1239	.5876	.4925	.0140	.5601	.6150
B	1239	.4931	.5002	.0142	.4653	.5210
C	1239	.5278	.4994	.0142	.5000	.5557
D	1239	.6392	.4804	.0136	.6124	.6660
E	1239	.4560	.4983	.0142	.4282	.4838
F	1239	.5738	.4947	.0141	.5463	.6014
G	1239	.5157	.5000	.0142	.4879	.5436
H	1239	.5343	.4990	.0142	.5065	.5621
I	1239	.5214	.4997	.0142	.4935	.5492
J	1239	.4479	.4975	.0141	.4202	.4757
K	1239	.4318	.4955	.0141	.4042	.4594
L	1239	.4019	.4905	.0139	.3746	.4293
M	1239	.6336	.4820	.0137	.6067	.6604
N	1239	.5093	.5001	.0142	.4814	.5372
O	1239	.5609	.4965	.0141	.5333	.5886
P	1239	.4294	.4952	.0141	.4018	.4570
Q	1239	.4512	.4978	.0141	.4234	.4789
R	1239	.5198	.4998	.0142	.4919	.5476
S	1239	.4213	.4940	.0140	.3938	.4488
T	1239	.4044	.4910	.0139	.3770	.4317
U	1239	.5028	.5002	.0142	.4749	.5307
V	1239	.4366	.4962	.0141	.4090	.4643

TABLE 3
ANALYSIS OF VARIANCE FOR DESIGN
AND OPERATIONAL CHARACTERISTICS PREFERENCE

SOURCE	DEGREES OF	SUM OF	MEAN	SIGNIFICANCE	
	FREEDOM	SQUARES	SQUARES	F-RATIO	LEVEL
Requirement	21	127.4798	6.0705	24.7248	0.000*
Error	27,236	6687.0202	.2455		
TOTAL:	27,257	6814.5000			

Bartlett's Test for Homogeneous Variances

Number Degrees of Freedom = 21
F = 0.306 Significance Level α = 0.999
*Significant at α less than 0.001 level.

Table 3 indicates that there were significant differences in the soldiers preference for the listed design and operational SACC characteristics. The Bartlett's Test indicated that the variance of each characteristic is homogeneous, i.e. not significantly different, so they are from the same population.

The Scheffe's Multiple Range Test (Table 4) was used to determine where these significant differences in preferences occurred. This test separates a set of significantly different means into subsets of homogeneous means.

TABLE 4
SCHEFFE'S MULTIPLE - RANGE TEST - SACC DESIGN AND OPERATIONAL CHARACTERISTICS PREFERENCE

<u>WORST</u> <u>GROUP 1</u>	<u>GROUP 2</u>	<u>GROUP 3</u>	<u>GROUP 4</u>	<u>GROUP 5</u>	<u>BEST</u> <u>GROUP 6</u>
L .4019	S .4213	J .4479	B .4931	R .5198	C .5278
T .4044	P .4294	Q .4512	U .5028	I .5214	H .5343
S .4213	K .4318	E .4560	N .5093	C .5278	O .5609
P .4294	V .4366	B .4931	G .5157	H .5343	F .5738
K .4318	J .4479	U .5028	R .5198	O .5609	A .5876
V .4366	Q .4512	N .5093	I .5214	F .5738	M .6336
J .4479	E .4560	G .5157	C .5278	A .5876	D .6392
Q .4512	B .4931	R .5198	H .5343	M .6336	
E .4560	U .5028	I .5214	O .5609		
B .4931	N .5093	C .5278	F .5738		
U .5028	G .5157	H .5343	A .5876		
N .5093	R .5198	O .5609			
G .5157	I .5214				
	C .5278				
	H .5343				

4.0 SECTION IV - DISCUSSION

The questionnaire was successful in determining which design and operational characteristics were deemed most important and least important as judged by the ground troops (Table 4). The most important characteristics for the SACC were as follows:

- No larger than 4 1/2 by 7 feet
- Offer protection against visual detection
- Woodland, arctic, tropic and desert SACCs should be reversible
- Weight does not hinder transport
- Offers protection against radar detection
- Must not shine or glare
- Shelf life of 10 years

Each group of characteristics differs significantly $\alpha = 0.05$ from each other. The six least important characteristics were:

- Does not interfere with hand movement
- Must be easy to use
- Must not present a health hazard

- Must not greatly increase the body temperature of the soldier under the SACC
- Must be fungus resistant
- Material is durable

Note that most of the characteristics overlap into adjoining groups. However, there were few surprises among the most preferred characteristics. As expected the SACC should be small, lightweight and be able to blend with the background, hence reversible. The least preferred characteristics of not being a health hazard, easy to use, durable, and not interfere with hand movements give an insight into the soldiers thoughts on camouflage. That is, if it works and is not too hard to carry the soldier, will put up with hardships.

The following requirements fell into the middle range of preference, i.e., groups 3 and 4:

- Must not make a noise when being handled
- Offers protection against near-infrared detection
- Must not snag
- Does not interfere with vision
- Field life (durability, color, fading, etc.) of 60 days
- Must be easily carried
- Shelf life of 10 years
- Must not shine or glare
- Offers protection against radar detection
- Weight does not hinder transport
- Must be able to be joined with other SACC units to place over larger objects
- Offers protection against thermal detection
- Must be non-flammable

The proper identification of important and not important characteristics precludes the possibility of incorrectly assigning resources to a characteristic which has little practical importance. A good example of this would possibly be characteristics S (must not present a health hazard) and G (does not interfere with vision).

5.0 SECTION V - SUMMARY AND CONCLUSIONS

A total of 59 soldiers from the 3rd Bn, 35th Inf, 187th Inf Bde, Fort Devens, MA participated in the study. During their field training, they used the SACC to conceal individual troops, weapon emplacements, fighting positions, and supply caches. Upon completion of the exercises, they were given a questionnaire/survey in which the soldiers made individual comparisons between 22 design and operational characteristics. Their task was to decide which of each pair of characteristics was the most important. Each subject made a total of 231 paired comparisons, with each

characteristic being evaluated 21 times. A review of the data indicated that the statistical procedures enabled the investigators to determine the most important and least important characteristics. Logical decisions on how to expend resources on the development of new camouflage can now be determined from what otherwise would be viewed as a large pool of subjective responses out of which little objective conclusions could be determined.

REFERENCES

1. Natrella, Mary C., Experimental Statistics, National Bureau of Standards Handbook 91, U.S. Department of Commerce, Washington, D.C., 1966.
2. Walker, Helen M. and Lev, Joseph, Statistical Inference, Holt, Rinehart, and Winston, New York, 1953.

Tree-structured Statistical Methods

Wei-Yin Loh

Department of Statistics
University of Wisconsin
Madison, WI 53706

Abstract

Recent developments in tree-structured methods are reviewed with emphasis on extensible and computationally efficient strategies.

1 Introduction

Tree-structured methods are compute-intensive statistical procedures that yield decision trees as solutions for classification and regression problems. Two early methods are the AID and THAID (Morgan and Sonquist, 1963; Morgan and Messenger, 1973) computer programs for regression and classification. These methods construct binary decision trees by recursively partitioning a data set. At each stage, all possible splits of the data in the partition are examined to find one that maximally reduces node impurity, where impurity is defined in terms of entropy or mean square error. AID and THAID were later superceded by CART (Breiman, Friedman, Olshen and Stone, 1984), whose most important contribution was a method of "pruning" to get a tree of approximately the right size. CART, however, adopted the slow split-finding strategy of its predecessors.

The FACT (Vanichsetakul, 1986; Loh and Vanichsetakul, 1988) method uses standard linear statistical techniques such as linear discriminant analysis and analysis of variance tests to find splits. It also uses a direct stopping rule similar to that in AID and THAID, instead of pruning. As a result, although FACT usually performs well in many applications, datasets can be constructed to fool it. Further, being based on linear discriminant analysis, FACT does not always give binary splits; it splits each node into as many subnodes as there are classes. On the other hand, the speed of FACT is usually ten to several hundred times faster than CART's.

2 Main results

Several new algorithms have been developed recently at the University of Wisconsin that combine the pruning method of CART with the fast splitting method of FACT. These algorithms share a common philosophy of sacrificing local split optimality for computational speed and ease of extensibility to generalized regression settings. Because of their ability to fit complex models quickly, the statistical accuracy of these methods is typically as good as, if not better than, CART's. Shih (1993) develops a likelihood-based method of split selection for categorical variables and a method of grouping more than two classes into two superclasses to allow binary splits. Chaudhuri, Huang, Loh and Yao (1994) describe a method of tree-structured regression that yields, if desired, smooth estimates of the function and its derivatives. Conditions for asymptotic consistency of the estimates are provided. Chaudhuri, Lo, Loh and Yang (1993), Lo (1993) and Yang (1993) generalize these ideas to tree-structured Poisson regression and logistic regression models. Extensions to stratified regression modeling of censored data using piecewise parametric and nonparametric models (such as Weibull and proportional hazards models) are reported in Loh (1991), Ahn (1992) and Ahn and Loh (1994).

The key ideas may be summarized as follows.

1. Use of a grouping procedure if necessary to combine classes into two superclasses at each node prior to splitting. This ensures binary splits.
2. Use of two-sample t-tests for differences between means and variances to select the variable to split a node, in the case of univariate splits. These tests are also used to detect patterns in residual plots to guide split selection in regression.
3. Use of linear discriminant analysis to determine the best linear combination split or the best univariate split on the selected variable.
4. Use of CART's pruning method to determine the final size of the tree.
5. Use of linear projections with dummy variable coding to convert categorical variables into ordered variables before splitting.
6. Use of maximum likelihood fitting for piecewise generalized regression.
7. Use of weighted averaging to produce smooth estimates of the function and its derivatives.

The details of the algorithms will be reported elsewhere. The practical advantages of this strategy over CART are:

1. Computational speed. CART finds linear combination splits on ordered variables by global optimization over all coefficients in the linear combination. Our method is much more efficient because it uses linear discriminant analysis. In the case of regression, CART fits a model to each subnode for every split considered. Since it examines all possible splits, this process is very time consuming. Our approach fits a model to each subnode only after a split is selected. Hence model fitting is performed only once at each node.
2. Treatment of categorical variables. To find the best split on a categorical variable, CART searches over all subsets of categories. Because the number of such splits increases exponentially with the number of categories, this is also a very time consuming process. Another problem is that this strategy tends to prefer splits on categorical variables with many categories over splits on ordered variables. Our approach of converting each categorical variable into an ordered variable avoids this problem and speeds up split selection.
3. Boolean combination splits on categorical variables. This can be quickly obtained via linear combinations of transformed categorical variables. Global optimization strategies are impractical because of the large number of splits that need to be evaluated.
4. Versatility in model fitting. Because model fitting is performed after split selection, models of arbitrary complexity (such as GLIM or proportional hazards models) may be fitted to each node at little additional cost.

References

- Ahn, H. (1992). *Tree-structured Survival Analysis*, PhD thesis, Department of Statistics, University of Wisconsin, Madison.
- Ahn, H. and Loh, W.-Y. (1994). Tree-structured proportional hazards regression modeling, *Biometrics*. In press.
- Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). *Classification and Regression Trees*, Wadsworth, Belmont.

- Chaudhuri, P., Huang, M.-C., Loh, W.-Y. and Yao, R. (1994). Piecewise-polynomial regression trees, *Statistica Sinica*. In press.
- Chaudhuri, P., Lo, W.-D., Loh, W.-Y. and Yang, C.-C. (1993). Generalized regression trees: Function estimation via recursive partitioning and maximum likelihood, *Technical report*, University of Wisconsin.
- Lo, W.-D. (1993). *Logistic Regression Trees*, PhD thesis, Department of Statistics, University of Wisconsin, Madison.
- Loh, W.-Y. (1991). Survival modeling through recursive stratification, *Computational Statistics and Data Analysis* 12: 295-313.
- Loh, W.-Y. and Vanichsetakul, N. (1988). Tree-structured classification via generalized discriminant analysis (with discussion), *Journal of the American Statistical Association* 83: 715-728.
- Morgan, J. N. and Messenger, R. C. (1973). THAID: A sequential analysis program for the analysis of nominal scale dependent variables, *Technical report*, Institute for Social Research, University of Michigan, Ann Arbor.
- Morgan, J. N. and Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal, *Journal of the American Statistical Association* 58: 415-434.
- Shih, Y.-S. (1993). *Tree-structured Classification*, PhD thesis, Department of Statistics, University of Wisconsin, Madison.
- Vanichsetakul, N. (1986). *Tree-Structured Classification via Recursive Discriminant Analysis*, PhD thesis, Department of Statistics, University of Wisconsin, Madison.
- Yang, C.-C. (1993). *Tree-structured Poisson Regression*, PhD thesis, Department of Statistics, University of Wisconsin, Madison.

COMBINING SIMULATION RESULTS ADDRESSING ARMOR VEHICLE SURVIVABILITY

Paul J. Deason, Ph.D.
U.S. Army TRADOC Analysis Center - WSMR
White Sands Missile Range, N.M. 88002-5502

This is a clinical paper addressing means to combine the results of a number of studies on two simulation models, the desired result of which is to identify a balanced cost-effective set of survivability enhancements for a direct-fire armored weapon system at an acceptable risk, so that these enhancements may be made a part of the engineering specifications for the weapons system. Passive survivability elements consist of ballistic protection measures, and signature reduction in the areas of RF, visual, and thermal spectra. Countermeasures considered cover smoke, receivers, jammers, and active protection systems. The intent is to maximize the use of passive measures, avoiding high technology, high risk solutions, and avoiding highly sophisticated active countermeasures.

SIMULATION MODELS

Simulation models are used in the study to evaluate the effectiveness of the system in combat given the enhancements of the suites of countermeasures, signature reduction, and ballistic protection. The two models are the GROUNDWARS few-on-few direct fire and artillery simulation, and the Combined Arms and Support Task Force Evaluation Model (CASTFOREM), a many-on-many battlefield simulation.

GROUNDWARS, maintained by the Army Materiel Systems Analysis Activity (AMSAA) is used primarily to evaluate weapon system effectiveness by representing land combat between homogeneous forces, where the total number of combatants cannot exceed twenty, and where the systems have a limited representation of sensors and munitions. A statistical terrain is represented. GROUNDWARS is stochastic employing Monte Carlo probability theory as its primary solution technique; three hundred replications of a case are normally employed.

CASTFOREM, maintained by the TRADOC Analysis Center - WSMR (TRAC-WSMR) is a stochastic, event sequenced, force-on-force simulation of ground combat involving up to a BLUE brigade and opposing RED forces. It is used for weapon system trade-off analyses, investigation of alternate tactics, metric analyses of selected weapon system performance parameters, and other similar studies. CASTFOREM is extremely flexible, and can accommodate any terrain or weapon system for which data is available. Terrain used is digitized actual terrain. Weather and ambient light conditions are

constant throughout a battle. Battlefield obscurants, smoke, and dust are modeled as dynamic clouds. Processes are modeled probabilistically using Monte Carlo techniques; the model is stochastic event sequenced, although time-step events are possible. Normally, 21 replications of a case are employed.

COUNTERMEASURES PLAYED IN GROUNDWARS

In the following table are described the variation of countermeasures suites for the system examined through the use of the GROUNDWARS model. Those with a 'Y' in the cell indicates the suite was examined.

The scenarios used are the following:

SCEN A: This scenario represents a BLUE mechanized infantry task force in a prepared defense against an overwhelming modern RED armor attack. The setting is Central Europe, Winter, with snow on the ground and 7 kilometers visibility.

SCEN B: This scenario represents a meeting engagement between a BLUE mechanized infantry brigade and a modern RED tank regiment. The setting again is Central Europe in Winter with snow and 7 kilometers visibility.

SCEN C: This scenario is set in late spring in Southwest Asia, dusty with 14 kilometers visibility. The BLUE force is a mechanized infantry battalion (+) in a hasty defensive posture encountering two Threat tank battalions equipped with current equipment. Threat counter-maneuver artillery is minimal.

The countermeasures described in the table are as follows:

LWR: Laser warning receiver - detects when the system is being lased by a threat rangefinder or detects a missile guidance laser.

MWS: Missile warning system/muzzle flash detector - detects the launch of a missile or the flash of a gun.

RWR: Radar warning receiver - indicates when being painted by radar.

SMK: Signifies the employment of self-protective smoke in the visual, infrared, and millimeter wave spectra in the direction of the perceived threat munition.

JAM: Infrared Jammer - disrupts the infrared tracking beacon on an incoming missile

SLID: Small, low-cost intercept device - a proposed counter-missile system.

SHORTSTOP: A proposed artillery countermeasure device.

- Several suites are not considered for the following reasons:
- Smoke would not be used in the SCEN A prepared defensive position, except if the system were on the move and exposed. Smoke was not used in GROUNDWARS (but was used in CASTFOREM.)
 - In SCEN A, SCEN B and SCEN C the LWR alone would not pick up the threat missiles to JAM, so the cases were not played.
 - In SCEN B and SCEN C the MWS alone is unable to detect the target and cue smoke, so that combination was not played.
 - In SCEN C there was no radar threat portrayed, so RWR was not considered.
 - In Scen C there was no artillery affecting the system, so SHORTSTOP was not played.

COUNTERMEASURE SUITES IN CASTFOREM

Using the same chart for the countermeasure suites played in GROUNDWARS, the suites that were evaluation in CASTFOREM are denoted with a 'C'.

Table 1: COUNTERMEASURE SUITE COMBINATIONS
GROUNDWARS and CASTFOREM

POSSIBLE SUITE	SCEN A	SCEN B	SCEN C
BASELINE (No CM)	YC	YC	YC
LWR	Y	Y	Y
LWR, SMK		Y	Y
LWR, JAM			
LWR, SMK, JAM			
MWS	Y	Y	YC
MWS, SMK			
MWS, JAM	YC	Y	YC
MWS, SMK, JAM			
RWR	Y	Y	
RWR, SMK		Y	
RWR, JAM	Y	Y	
RWR, SMK, JAM		Y	
LWR, MWS	Y	Y	Y
LWR, MWS, SMK		Y	YC
LWR, MWS, JAM	Y	Y	Y
LWR, MWS, SMK, JAM		Y	Y
LWR, RWR	Y	Y	

LWR, RWR, SMK		Y	
LWR, RWR, JAM	Y	Y	
LWR, RWR, SMK, JAM		Y	
MWS, RWR	Y	Y	
MWS, RWR, SMK		Y	
MWS, RWR, JAM	Y	Y	
MWS, RWR, SMK, JAM		Y	
LWR, MWS, RWR	Y	Y	
LWR, MWS, RWR, SMK		Y	
LWR, MWS, RWR, JAM	Y	Y	
LWR, MWS, RWR, SMK, JAM	YC	YC	YC
LWR, MWS, RWR, SMK, SHORTSTOP	YC	YC	
LWR, MWS, RWR, SLID	YC	YC	YC

SIGNATURE REDUCTION IN GROUNDWARS and CASTFOREM

Threat systems employ a variety of means to detect and to bring fire onto our system of concern. These means are direct view optics, low-light television systems, thermal imaging systems, ground surveillance radars, and seekers in "smart" artillery munitions. If the ability of the system to be detected or to be accurately pinpointed were reduced, its survivability would be enhanced. It is possible to reduce the signature of the vehicle through the use of various suites of coatings and shaping. The signature reduction suites are represented on the vehicle by specifying an average detection range reduction achievable against the array of threat sensors represented in the scenarios. Probability of detection by threat seekers as a function of slant range and target signature were determined using Booz-Allen and Hamilton's Desktop Radar and Infrared Signature Model. This reduction is as measured by the NVEOL sensor curves for the threats of interest; the point where the probability of detect curve is 50% ($P_{det}=0.5$) for the range desired was taken as the target reduction criteria. (The Johnson criteria of one cycle was used for detection.) Five levels of signature reduction were played in both the GROUNDWARS and CASTFOREM models, and were designated Level A through Level E.

Table 2: SIGNATURE REDUCTION IN GROUNDWARS and CASTFOREM

DETECTION RANGE	SCEN A	SCEN B	SCEN C
BASELINE SYSTEM	YC	YC	YC
Level A	YC	YC	YC
Level B	YC	YC	YC
Level C	YC	YC	YC
Level D	YC	YC	YC
Level E	YC	YC	YC

BALLISTIC PROTECTION IN GROUNDWARS AND CASTFOREM

The ballistic protection suites added to the vehicle are in addition to the base armor package inherent with the system. This added ballistic protection would be against direct fire kinetic energy and chemical energy munitions, and indirect fire (artillery) munitions. Due to the differences inherent in the penetrations from direct fire rounds and indirect fire munitions cause the ballistic packages to be considered separately, although there would be a carry-over effect (synergy) one to the other. Three levels of ballistic protection were considered, based on the probability of resisting a system kill (as defined by an analysis using the Army Research Laboratory CAD and evaluation models) given a hit, of 50%, 75%, or 95% (given the percentage is higher than the standard armor package.) (an attempt to design a package that would withstand the impact of large calibre direct fire munitions, or a direct impact of artillery HE was not considered.) These packages are limited by the power, weight, and dimensional constraints of the system.

Table 3. BALLISTIC PROTECTION IN GROUNDWARS and CASTFOREM

BALLISTIC PROTECTION	SCEN A	SCEN B	SCEN C
BASELINE SYSTEM	YC	YC	YC
50% DF, 50% IF	YC	YC	YC(DF)
50% DF, 75% IF	YC	YC	
50% DF, 95% IF	YC	YC	
75% DF, 50% IF	YC	YC	YC(DF)
75% DF, 75% IF	YC	YC	
75% DF, 95% IF	YC	YC	
95% DF, 50% IF	YC	YC	YC(DF)
95% DF, 75% IF	YC	YC	

COMBINED RUNS IN CASTFOREM

A number of runs are scheduled in CASTFOREM to evaluate combinations of signature reduction, countermeasures, and ballistic protection. To date, only cases using the countermeasure suites of LWR, MWS, RWR, and SMK at reduced signature levels A, B, and C, using the base level of ballistic protection are scheduled. Other cases will be considered as a result of preliminary analyses; additional combined case runs are welcomed from the panel.

MEASURES OF EFFECTIVENESS

Common measures of effectiveness output by GROUNDWARS and CASTFOREM are system kills, system loss, RED force loss, BLUE force loss, the system exchange ratio (system kills/system loss), force loss exchange ratio (RED force loss/BLUE force loss), and surviving maneuver force ratio (RED maneuver force (initial-final)/BLUE maneuver force (initial-final)). In CASTFOREM these measures are available over time for each replication. A metric which could handle the synergy of the battle over time, and the contribution differences of the system in various parts of the battle by virtue of its survival is envisioned. In these scenarios, the early contribution of a system could cause it to expend its ammunition early, and so not contribute later in the battle. However, because of its early contribution, more BLUE direct fire systems could survive and participate strongly. The more survivable system's contribution could be swamped by the end of the battle due to the synergy. Therefore, a combined metric is visualized.

Then, once the systems providing the most potential are determined (ranked?) by their performance in the simulations, additional factors must be considered, such as the following:

- a. Cost
- b. Weight and size constraints placed on the system
- c. Technological risk and possible fielding date

The intention is to provide the Army with a robust point solution package to enhance the survivability and performance of the system and the force.

EPILOGUE

Since the conference in October 1993, the method of analysis used was to separate GROUNDWARS and CASTFOREM except as the findings were mutually supporting, and use the results from CASTFOREM as the principle effectiveness determiner. The final full factorial CASTFOREM runs matrix consisted of the European Defense and the SWA Meeting Engagement scenarios, two levels of signature reduction, three countermeasure suites, three levels of ballistic protection, which when added to the base level of each factor (no signature reduction, no countermeasure suite, basic level of ballistic protection) resulted in a 96-case matrix (2 X 3 X 4 X 4.) The final product of this analysis, known as the LOSAT Survivability

Requirements Study, should be available after the First of April, 1994 from the Technical Management Division, LOSAT Project Office, U.S. Army Missile Command, Attn: SFAE-ASM-LS, Redstone Arsenal, Huntsville, Alabama 35898-8051

A REDISCOVERY OF THE HODGES-LEHMANN ESTIMATE

PAUL H. THRASHER
Analysis Branch of Analysis Division
Material Test Directorate
White Sands Missile Range, New Mexico 88002-5157

ABSTRACT. The Hodges-Lehmann estimate was proposed in 1963 to balance the risks of estimating too high and too low. It was recently rediscovered in a consideration of what estimate should be used after a one-dimensional null hypothesis is rejected.

1. INTRODUCTION. Use of data to find a point estimate for a parameter is often requested. If no standard or unacceptable value is provided for the parameter, a point estimate is often found to (1) stand alone or (2) serve as the midpoint of a confidence interval. If (1) a standard and unacceptable value exist and (2) the data and agreed upon Type I and Type II risks imply rejection of the null hypothesis that the parameter meets the standard, then the next question is often "How badly does the parameter miss the requirement?" Although a p-value and a post-test Type II risk can answer this question, a point estimate is often requested by managers not versed in statistical language.

If a point estimate is needed, an analyst may well want to present a more statistically justified number than the commonly used average or sample median. One general technique is to extend hypothesis testing. This was (1) done in reliability studies at White Sands Missile Range with arguments described in sections 2-5 of this paper, (2) presented as a clinical paper to the Thirty-Ninth Conference on the Design of Experiments in Army Research, Development, and Testing and (3) recognized by one of the panelists as the Hodges-Lehmann technique.

2. RATIONALE. Any point estimate necessarily has limited information. It should be made as meaningful as possible.

A point estimate might be too high or too low. An intuitive approach is to adopt a goal of equal likelihood; that is, try to equalize the risks of estimating too high and too low.

One way to approach this goal is to think of two hypothesis tests that (1) share the common null hypothesis of "The Desired Parameter Equals The Point Estimate" and (2) have the opposing alternate hypotheses of "The Point Estimate Is Less Than The Desired Parameter" and "The Point Estimate Is Greater Than The Desired Parameter" as the upper and lower alternatives to the null. Since the p-value is the probability of being wrong if the null is rejected, the goal of equal likelihood can be approached by adjusting the point estimate in these two thought hypothesis tests until their p-values from data are as close to each other possible. This forces both p-values toward one half.

Clearly each p-value can be forced to exactly one half if the thought hypotheses tests have p-values that are continuous with the always continuous point estimate. If such thought hypotheses tests are not appropriate for the data, an average can be taken of the two point-estimates that make the two p-values closest to one half.

The resulting estimate can logically be named and described by the acronym p-vulte (p-value upper & lower test estimate). "P-Vulte" can be thought of linguistically as a noun; but it is more informative if it is considered as an adjective in a three word title. For example, "Gaussian p-vulte mean" or "Wilcoxon p-vulte median" denote both the distribution that describes the data and the parameter that is being estimated.

3. NON-INNOVATIVE RESULTS. For data from populations described by some common distributions, use of the p-vulte technique yields nothing new. For example, the Student's t p-vulte mean is simply the sample average.

This result is obtained by considering the two areas, of the probability density function, that are separated by the desired estimate. Adjusting these areas until they are equal makes both of them exactly equal to one half. At this point, the test statistic t is zero. The well known expression for t,

$$\frac{[\text{Sample Average} - \text{Population Mean}]}{[\text{Sample Standard Deviation} / \text{Square Root of Sample Size}]}$$

immediately yields the p-vulte mean to be the sample average.

4. BIASED RESULTS. Non-symmetric probability density functions lead to biased p-vultes. This bias tends to zero as the sample size becomes very large. One example is the binomial p-vulte R where R is the reliability (i.e., the probability of one success in one trial).

Calculation of this p-vulte is direct in concept; but in practice it requires a computer. Equating the two p-values is the same as equating two sums of $b(j;n,p\text{-vulte})$ where b is the function for the binomial probability distribution; x is the number of successes out of n trials; one sum ranges from $j=0$ to $j=x$; and the other sum ranges from $j=x$ to $j=n$. After data is taken, the only unknown in the equation is the p-vulte. Clearly a numerical solution is possible; but the existence of two sums causes difficulties. Calculation is facilitated by (1) pulling the term $b(x;n,p\text{-vulte})$ out of both of the two equal sums, (2) remembering that the sum of $b(j;n,R)$ from $j=0$ to $j=n$ must be one for any R, and (3) arriving at the calculation equation of

$$1 = b(x;n,p\text{-vulte}) + \text{TWICE THE SUM OVER } b(j;n,p\text{-vulte});$$

this sum ranges either from $j=0$ to $j=(x-1)$ or $j=(x+1)$ to $j=n$.

The bias may be illustrated with an example. For a sample size n of 25, different values of x yield the binomial p-vulte R and the maximum likelihood estimator (i.e., x/n) to be

<u>x</u>	<u>p-vulte</u>	<u>x/n</u>
1	.0453	.0400
3	.1247	.1200
7	.2828	.2800
12	.4802	.4800
13	.5198	.5200
18	.7172	.7200
22	.8753	.8800
24	.9547	.9600

All of these p-vultes are biassed towards the central possible value of R (i.e., 0.500). The shifting is greatest for values of x that are farthest from corresponding to $x/n = 0.500$.

As an aside, consider the situation when $x=0$ or $x=n$. There are two possible interpretations for the binomial p-vulte R . Just looking at the summation equations and plots of the distributions for different possible values of R suggest that these p-vultes are unbiased (i.e., identically zero and one). However, looking at the underlying the thought hypothesis tests suggests that these p-vultes are undefined. This occurs because there is no physical alternative that the estimate should be lower than 0 or higher than n . Unless a limiting procedure is considered, the binomial p-vulte R is thus undefined when $x=0$ or $x=n$. One philosophical interpretation is that being undefined is not bad in this situation; that is, neither perfection nor total failure should be claimed for the population just because data from any sample fails to indicate differently.

Finally, another example shows the tendency of the bias to be removed with large sample sizes. Choosing x 's and n 's such that x/n is 1/5 yields

<u>n</u>	<u>p-vulte</u>
5	.2161
10	.2090
25	.2038
75	.2013
250	.2004
1000	.2001

This table exhibits the tendency of consistency.

5. ROBUST AND SENSITIVE RESULTS. Application of the p-vulte technique to the Wilcoxon signed ranks T test yields robust and sensitive results. This should be expected because the Wilcoxon signed ranks T test is well known for its high power.

The Wilcoxon p-vulte median may be calculated without first calculating actual p-values. This is based on the way that the Wilcoxon p-value calculation is done using a thought experiment:

- (1) Consider n chips for the n data of the sample;
- (2) Label each chip with
 - (a) the sample rank of the absolute value of the difference between the standard & the datum and
 - (b) a + sign on top if the standard exceeds the datum but on the bottom if the datum is the biggest;
- (3) Calculate T^+ by summing the ranks on the chips where the standard exceeds the datum (i.e., +'s are up);
- (4) Think of tossing all chips;
- (5) Consider all two to the n th power possible landings;
- (6) Count the number of possible landings for which the sum of the ranks of chips with plus sides up is less than the T^+ result of step 3 (i.e., count possible results that are as bad or worse than the data);
- (7) Find the p-value by dividing the result of step 6 by the result of step 5.

[Note: See the appendix for a discussion of handling ties.]

This finds the probability of being wrong in rejecting the null hypothesis, that the median equals the standard, in favor of the alternate hypothesis that the median is higher than the standard. The p-value for the other alternate hypothesis, that the median is lower than the standard, can be found by changing "less than or equal to" in step 6 to "greater than or equal to". Obviously, the counting in step 6 is tedious and time consuming for even a computer when n is large and the standard is near the middle of the data. Fortunately, it is not necessary to find the p-vulte by the direct approach of guessing "standards" until one is found that yields equal p-values for the two alternate hypotheses. The shortcut is based on features of the number line and the p-value:

- (A) The upper and lower alternate hypotheses both have zero p-values if the standards are outside the data's range;
- (B) In starting with two trial standards on opposite sides of the data and moving them inward, neither p-value changes until the extremes of the data are reached;
- (C) Reaching the extreme data values causes (i) the count in step 6 to increase from zero to one and (ii) the two p-values to increase; both become one divided by two raised to the n th power;
- (D) The other points on the number line that change the counting in step 6 are values of "standards" equaling (i) other data and (ii) pair-wise averages of the data;
- (E) The symmetry of the number line and integer intervals between ranks makes symmetric contributions to the two p-values as the two "standards" are slid in unison over pairs of points identified in property D;
- (F) Equal p-values are retained by crossing pairs of property D points simultaneously;
- (G) The p-vulte is reached when the two "standards" meet.

Thus the Wilcoxon p-vulte median is the sample median of all the pair-wise averages of the data including each datum with itself. If the sum of the sample size and the number of pair-wise averages (i.e., $n + n!/[2!(n-2)!]$) is odd, then the p-vulte is unique. If this sum is even, the p-vulte is somewhere between the innermost pair of data and pair-wise averages on the number line. Although there is no justification, a unique estimate may be obtained by instinctively defining it as the average of the innermost points; this will be called the "even estimate".

Although the calculation of the Wilcoxon p-vulte median is direct in concept, its actual calculation needs a shortcut to be practical for large data sets. Even a modest sample size generates a large number of pair-wise averages. Even medium size computers can have storage difficulties if all $n + n!/[2!(n-2)!] = n(n+1)/2$ averages are stored at once, bubble sorted, and counted off to the middle value. Fortunately, there is a simple technique to avoid the handling of this large array of numbers:

1. Bubble sort the data with the lowest datum at the low end;
2. Think of a triangular array of pair-wise averages of all data including each datum with itself;
3. View the diagonal;
4. Construct and store the averages on the diagonal and the row and column numbers needed to find these averages;
5. Bubble sort the diagonal, discard the lowest average, and replace it with the next largest array average;
 [Note: The location of the replacement average from the discarded average is either (a) immediately to the right on the same row or (b) immediately down the diagonal. Clearly, replacement from the diagonal necessitates another replacement before proceeding to the iteration of step 6.]
6. Repeat step 5 until the sample median of the triangular array can be found.

[Note: For odd $n(n+1)/2$, discarding $[n(n+1)/2 - 1]/2$ values makes the smallest value on the remaining diagonal equal to the Wilcoxon p-vulte median. For even $n(n+1)/2$, discarding $n(n+1)/4 - 1$ values makes the average of the two smallest values on the remaining diagonal equal to the even estimate of the Wilcoxon p-vulte median.]

This technique uses storage for the $3n$ diagonal values and their row and column sources instead of storage for the $n(n+1)/2$ array values.

The sensitivity and robustness of the Wilcoxon p-vulte median may be illustrated by simulations. Either graphs or tables may be used to display the results.

The two graphs display results from a simulation illustration of (1) how rapidly repeated sampling yields convergence and (2) how closely the convergence approaches the input parameter. The three lines are all calculated from the same set of simulations. One sample of size eleven is simulated to find and plot the average, sample median, and Wilcoxon p-vulte median at the left (i.e., #1) ends of the three lines. Each of the following points to the right incorporates another simulation of a sample of size eleven; the three quantities graphed are the average of the averages, the sample median of the sample medians, and the Wilcoxon p-vulte median of the Wilcoxon p-vulte medians.

Since a uniform population between zero and one is used for the simulations, the target value for all three lines is exactly one half. The solid line traces the central limit theorem prediction that the average of averages from different random samples will approach the population mean. The line with long dashes traces the corresponding theorem prediction that the sample median of sample medians from random samples will approach the population median. Finally, the line with short dashes does the analogous process with the Wilcoxon p-vulte median.

The graph from populations with no outliers shows that the average converges best. The Wilcoxon p-vulte median does almost as well; but the sample median exhibits large excursions. Thus the average is most sensitive; the Wilcoxon p-vulte median is quite sensitive; and the sample median is least sensitive.

The graph from populations with outliers shows the sample median to converge best. The Wilcoxon p-vulte median does quite well; but the average is biased toward the weighted average of $(0.95)(0.5) + (0.05)(2.5) = 0.6$. Thus the sample median is most robust; the average is least robust; and the Wilcoxon p-vulte median is bracketed by the sample median and the average.

An analyst is never certain if data has outliers. Thus the Wilcoxon p-vulte median is the best estimate of central tendency.

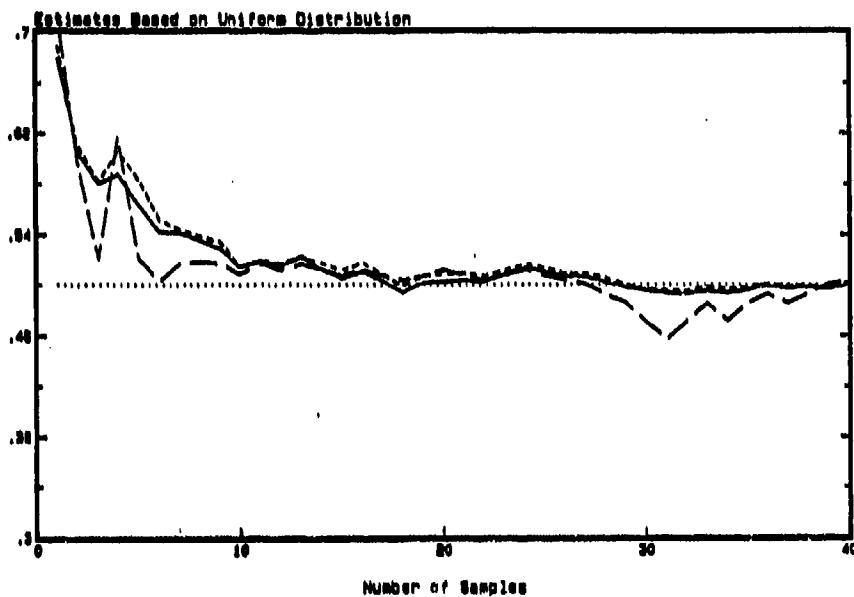
These graphical results need to be repeated many times before they can be generalized. Instead of trying to compare many graphs, repeated simulations can be reported with tables.

Before preparing tables, the investigation should be broadened to include populations other than the uniform. After all, a Gaussian or Student's t probability density function would be expected to have better convergence than the uniform.

For sensitivity investigation, the sample variance of repeated simulations is the quantity that is desired to be minimized. Tabulated results from a set of 200 simulated graphs

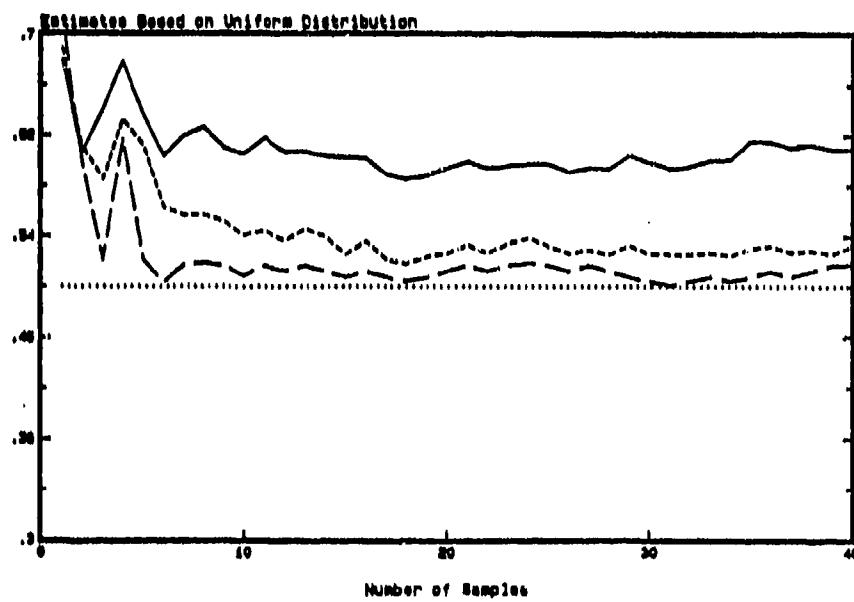
Comparison of Estimates of Central Tendency
Each Sample Has 11 Values; None of Them are Outliers

AVERAGE SAMPLE_MEDIAN PVULTE_MEDIAN PARAMETER



Comparison of Estimates of Central Tendency
Each Sample Has 11 Values; 5% of Them Have a Bias of 8

AVERAGE SAMPLE_MEDIAN PVULTE_MEDIAN PARAMETER



for probability density functions (pdf's) of triangular (t), uniform (u), and sine (s) with no outliers are:

Number of samples of size 11 in estimate	pdf	Sample variance of averages	Sample variance of sample medians	Sample variance of Wilcoxon p-vulte medians
3	t	0.038	0.053	0.040
	u	0.053	0.097	0.059
	s	0.065	0.145	0.073
10	t	0.020	0.029	0.022
	u	0.029	0.055	0.032
	s	0.035	0.085	0.038
60	t	0.007	0.012	0.008
	u	0.010	0.023	0.012
	s	0.013	0.036	0.014

All of the sample variances in the sample median columns are appreciably larger than those in the other two columns. Thus the sample median is the least sensitive. The sample variances in the Wilcoxon p-vulte median column are only slightly larger than those in the average column. Thus the Wilcoxon p-vulte median is almost as sensitive as the average.

For robustness investigation, the quantity of interest is the actual measure of central tendency. Tabulated results from 200 graphs when the target is 0.5 and 5 percent of the population has a bias of 2.0 are:

Number of samples of size 11 in estimate	pdf	Central average of averages	Central average of sample medians	Central average of Wilcoxon p-vulte medians
3	t	0.547	0.515	0.520
	u	0.545	0.533	0.528
	s	0.545	0.539	0.531
10	t	0.546	0.511	0.518
	u	0.544	0.526	0.525
	s	0.546	0.540	0.530
60	t	0.549	0.512	0.520
	u	0.548	0.524	0.528
	s	0.548	0.540	0.532

All of the values in the averages columns are appreciably further from 0.5 than those in the other two columns. Thus the average is the least robust. The values in the Wilcoxon p-vulte median and

sample median columns are comparable. Thus the sample median and Wilcoxon p-vulte median are more robust than the average.

The result of the tabular investigation is thus the same as that of the less extensive graphical analysis. The Wilcoxon p-vulte median is the best estimator of central tendency when the analyst does not know if outliers are present or absent.

6. PROFESSIONAL REVIEW AND ACKNOWLEDGMENTS. All techniques must be reviewed and placed in context with established methods. This review was provided by the Thirty-Ninth Conference on Design of Experiments in Army Research, Development, and Testing at Rice University in October, 1993.

Three important things were recognized at the conference. First and most important, the p-vulte technique is the same as the Hodges-Lehmann estimate. This pioneering work was reported with great mathematical thoroughness in Hodges, J. L. Jr. and Lehmann, E. (1963): Estimates of Location Based on Rank Tests, Ann. Math. Statist. 34, 598-611. Second, Hodges-Lehmann estimation is based in a philosophy that is not in either mainstream of statistical methodology. It is neither frequentist nor Bayesian; it may be most properly described as Fisherian. Third, it has been applied only to one dimensional data. In the thirty years since Hodges-Lehmann estimation was introduced, statistical methodology has made great advances in the more productive area of multidimensional analysis.

Many people associated with the Conference on the Design of Experiments in Army Research, Development, and Testing are deeply appreciated for their valuable contributions. Long before the conference, the clinical session chairperson, W. J. Conover, identified the Wilcoxon T test zero percent confidence interval as the sample median of pair-wise data averages. Also before the conference, program committee member Malcolm Taylor encouraged the presentation of this paper. Another program committee member, Francis Dressel scheduled this paper in a clinical session where it eventually received many constructive comments. Panelist Bernard Harris recognized that the p-vulte has the statistical property of consistency. Panelist Wei-Yin Loh identified the Wilcoxon p-vulte median as the Hodges-Lehmann technique. Panelist J. Sethuraman explained Loh's identification and also identified the reference to the original journal article by Hodges and Lehmann. Program committee member Gerald Andersen specified the section of the Rice University library, where the conference was physically held, that had a textbook description showing clear direct equality of the Wilcoxon p-vulte median and the Hodges-Lehmann estimate. Panelist Nozer Singpurwalla enunciated that the Hodges-Lehmann estimate does not utilize any prior information in a Bayesian analysis. Many other conference participants, especially David W. Scott who taught the tutorial on multivariate density estimation, illustrated that multidimensional techniques have wider applicability than the single dimensional Hodges-Lehmann estimate.

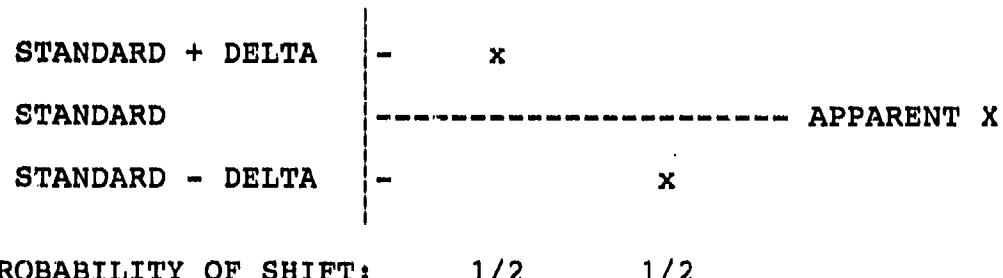
APPENDIX. TIE BREAKING. Certain analyses such as the Wilcoxon signed ranks T+ test have difficulties associated with ties. There are two types of ties. First, groups of data may have one common reported value. Second, the standard may equal the reported value of a datum or group of data. The first of these is easily handled by assigning ranks to each group member in a manner that (1) does not effect the ranks of other data and (2) uses an average rank within the group. [E.G., Assign ranks of 1, 2.5, 2.5, 4, 5, 7, 7, 7, 9, 10, 11, 12 to the absolute values of the differences between the standard and data equaling 2.05, 3.3, 3.3, 3.9, 4.5, 5.5, 5.5, 6.1, 6.2, 6.3, 14.4.] The second type of tie is more difficult and is discussed below.

A trivial method of handling a tie of the standard with a datum or a group of data is to simply discard all zeros in the set of differences of the standard and the data. Unfortunately, this causes the p-value to be non-monotonic with the standard.

A more realistic method of handling these zeros is to recognize, for continuous data, that they really don't exist. They appear to exist only because the data were not measured to a sufficient number of significant digits. Such apparent ties can be removed in a pre-analysis of the data by shifting the data away from the standard. This is analogous to the introduction of "jitter" into data for computerized data viewing in multivariate density estimation. An analytical analysis of this shifted data calculates an expectation value from all possible shifts.

Obviously a probability density function is needed to calculate expectation values. Two possibilities are the binomial and the uniform.

A binomial pre-analysis may be used as a first approximation in breaking of ties. For a single tie, an apparent datum X may be considered as being above or below the standard in the following picture



with the probability density tabulated under the above picture. Two values of the p-value are calculated from the two possible relative locations of the apparent datum. The p-value's expectation value is the sum of the products of possible p-values and the probabilities of those p-values. Since the probabilities are both 1/2 in this binomial pre-analysis for a single tie, this p-value's expectation value is just the average of the p-values.

The picture and table of probabilities for two data, X and Y, that apparently tie the standard are

STD + DEL	-	x,y	x	y	
STD	-				x,y
STD - DEL	-		y	x	x,y
PROB[SHIFT]:		1/4	1/4	1/4	1/4
or					
PROB[SHIFT]:		1/4		1/2	1/4

where the second version of the table of probabilities is for expectation value calculations. (There is no sense in calculating both degenerate p-values for the two center shifts.) In the general case of n data that apparently tie the standard, binomial pre-analysis yields a p-value's expectation value equal to the sum [from j=0 to j=n] of the product of (1) $b(j;n,1/2)$ and (2) the p-value calculated from a data set with j data shifted to one side and n-j shifted to the other of the standard.

The numerical value used in the shift (i.e., delta) does not effect rank analysis results as long as delta is small compared to the smallest separation between two data. However, not all ties of data with itself are broken by binomial pre-analysis when two or more data are tied with the standard. This does have a effect. The effect of false ties of measurements of a continuous random variable can be removed with uniform pre-analysis.

The uniform probability density function is an appropriate description of data that is taken with digital meters. Any reading is necessarily rounded off to the number of digits available on the meter. The meter cannot indicate which way or how far the data value is off. Thus the analyst knows only that the true measurement should be somewhere between in the interval bounded by data plus or minus half the smallest unread digit.

The picture and table of probabilities for uniform pre-analysis of two apparently tied data X,Y are

STD + 2 DEL	-	x	x		
STD + DEL	-	y	y		
STD	-				x,y
STD - DEL	-			y	y
STD - 2 DEL	-		x	x	
PROB[SHIFT]:		1/4	1/4	1/4	1/4

where the stared (i.e., *) non-split absolute values of differences between the standard and split data have been deleted from

	*	*	*	*
-	x x x	y	y	y
-	y	x x x	y	y
-----				X,Y
-	y	y	x x x	y
-	y	y	y	x x x

and the degenerate interchanges of x and y have been removed. Since all four of the probabilities in the table are equal, the p-value's expectation value is the average of the four p-values calculated from the four sets of data less X and Y plus each of the completely-split but non-degenerate x and y. Instead of pictures, tables can be used. For n=2, the table is

DIFFERENCE BETWEEN STANDARD & TOTALLY SPLIT DATA	SIGNS IN 4 SETS OF PAIRS
2 DELTA	+ + - -
1 DELTA	+ - + -

where only the information essential for calculation has been retained. For n=4, the corresponding table is

DIFFERENCE BETWEEN STANDARD & TOTALLY SPLIT DATA	SIGNS IN 16 SETS OF QUADRUPLETS
4 DELTA	+ - + + + - - - + + + - - - + -
3 DELTA	+ + - + + - + + - - + - - + - -
2 DELTA	+ + + - + + - + - + - + - + - - -
1 DELTA	+ + + + - + + - + - + - - + - - - -

where again only the information essential for calculation has been retained. In the general case of n data that apparently tie the standard, uniform pre-analysis yields a p-value's expectation value equal to the sum [over j=0 to j = the total number of ways of choosing 0, 1, ---, n positive signs for the n differences between data and the standard] of the product of (1) the reciprocal on two to the nth power and (2) the p-value calculated from a shifted data set where all n ties have been broken.

Uniform pre-analysis is obviously more complete and time consuming than binomial pre-analysis. Both are improvements over no pre-analysis when rounding off introduces fictitious ties.

Automatic Classification of Research Projects Based on Lexical Content

by
Mc Brown
Army Research Office

Abstract:

To track the scope of industrial and governmental research in a variety of scientific areas of interest to the Army, ARO has previously assigned scientists to read research reports from these organizations and, based on the scientists' knowledge, to classify the research according to a system of classification categories that correspond to Army technologies and operational functions. To avoid this highly labor intensive effort, which needs to be updated annually or bi-annually, an algorithm has been developed that automates this classification. The algorithm performs the classification based only on the aggregate of words that are used in the research report, as calibrated based on previous classifications which had been performed manually.

Introduction:

As part of its ongoing effort to ensure the mission relatedness of its basic research program, and to provide the Army with guidance in its technical base programs, ARO attempts to keep track of research in progress in industry.

A particularly convenient window on industry has been the industry reports prepared in connection with the Independent Research And Development (IRAD) program. The IRAD program provides major DOD contractors with funding that they can use to perform R&D of their own choosing. In return, the contractors have been required to provide reports to DOD describing the research. Until recently, contractors have also been required to provide on-site reviews every 3 years to put portions of their research on display for interested government representatives.

In past years, ARO scientists have examined the written reports and attended some of the on-site reviews, and have manually compiled a data base that summarizes and categorizes the research according to the major Army functions that the research supports. Examples of such functions are logistics, mobility, vulnerability reduction, NBC protection, target acquisition, lethality, C³I, and ECM/ECCM, and their various subfunctions. ARO has also tracked research that support major technical areas such as electronics, materials, manufacturing technology, computers/computer science, and space-related technologies, also broken down by their various sub areas.

For convenience in organizing the data, ARO has assigned alphanumeric labels, called descriptors, to designate the various research categories. Previously, the ARO descriptors have been assigned manually to each of the IRAD projects by ARO scientists through a

tedious, labor intensive effort of reviewing all of the IRAD reports generated by the participating industrial agencies.

Recently, Congress has mandated a number of changes to the IRAD program. Among these changes are the reporting requirements for industry. On-site reviews are now optional on the part of each company, and are less formal. Also, written reports are now to be compressed to a maximum of 5 pages per research project, and are to be reported annually to DTIC for inclusion on a CD ROM that DTIC will make available to DOD agencies.

Details of the information to be included on CD ROM, and formats for the information, are still in flux. However, these are expected to include project titles, company/profit center names, narrative descriptions, and funding levels. Abstracts and keywords may possibly also be included, although these are currently among the items under negotiation between industry, DTIC, and other interested DOD agencies.

In one respect, the new rules make it more difficult to keep track of what industry is doing in its IRAD efforts, as a result of the shortening of the written reports and the severe reduction in on-site reviews. On the other hand, the availability through DTIC of computerized reports promises to enhance ARO's capability to keep track of industry IRAD efforts. In particular, it opens the possibility of automating the process of categorizing the IRAD efforts by Army functions supported, thereby greatly reducing the highly time consuming demands on ARO scientists who have previously performed this categorizing manually.

What follows is an interim report on the development of a computerized technique for automating this categorization based on data supplied by DTIC, and on information derived from manually generated categorizations performed in previous years.

It should be noted that many of the IRAD projects are inherently structured so as to support more than one Army function. Categorization by ARO descriptor is therefore in most cases largely a technical judgement call, and even in principle can be correct only to within broad tolerances. Moreover, the very definitions of the Army functions and, particularly, of the contributing technologies are inherently somewhat fuzzy in their definitions, and constantly changing. Thus, the relationship between ongoing IRAD projects and Army functions that they tend to support, is at best imprecise and changing, even in principle, so that any technique for relating project with functions would necessarily be somewhat imprecise and changing.

Nevertheless, Army managers need basic information as to how the ongoing IRAD projects and the Army's own RDT&E tech base program tend to support the Army's functional needs, even if such information may be less than fully precise and/or stable over time. It would appear, therefore, that categorizing IRAD (and Army tech base) projects according to ARO descriptors promises to be useful to Army management for addressing

policy questions relating to funding the Army tech base, even if the categorizations are performed only to within some residual errors.

Approach: Data

Available data for automating the categorizations consist of two parts.

The first part consists of the sample of data that had been categorized manually in previous years (i.e., the historic data). This consists of written IRAD reports, supplied by industry, of which the project titles and industry-supplied keywords are used, together with ARO-generated characterizations by ARO descriptor.

The second part consists of the DTIC-supplied data that represent the current IRAD projects which are to be classified.

For the purposes of algorithm development, only the historic data is available. DTIC data for current projects is being collected from industry, and is being loaded and formatted onto CD ROMs by DTIC, but is not yet available.

Approach: Model

The categorization algorithm is based on a mathematical model, developed as follows:

Each IRAD report (historic or current) contains several data fields that will be analyzed. The entries in the data fields will be broken into words. In this way, each project report will generate an aggregate of words that have been taken from the various data fields, and the aggregate of words so obtained will be regarded as collectively representing the project. The words will be used as a basis for categorizing the project and associating it with an ARO descriptor. To do this, the procedure must first be calibrated based on historic projects.

Mathematically, the collection of words derived from all of the historic projects (i.e., those used to calibrate the model) will define a multi-dimensional mathematical space. In this space, each word corresponds to one of the dimensions, and vice versa. For convenience, call this space word space.

Now consider an arbitrary historic project. This project defines a vector in word space, as follows: Each coordinate has the value 1 if the corresponding word appears in the word aggregate for that project, and has the value 0 if it does not appear. That is, the coordinate indicates whether or not the word appears at least once in the project report (or, more accurately, in those data fields of the report which are used in the analysis).

Next consider those vectors derived from projects that have a given ARO descriptor assignment. The average of those vectors will be used to represent that particular ARO descriptor. A descriptor vector V_D corresponding to an ARO descriptor D therefore has

the following simple meaning: its typical component V_{DW} corresponding to a word W in the calibration word set measures the observed fraction of projects with descriptor D in which word W appears at least once. Each of the V_{DW} will therefore be between 0 and 1.

Having completed the calibration using historic data, consider the problem of assigning ARO descriptors to a set of current projects. To do this, choose an arbitrary project with vector X where component $X_W = 1$ or 0 according to whether word W appears at least once in the project writeup or not.

Now define a metric in word space. It might seem sufficient to use the metric

$$H_D = \sum_w |(X_w - V_{DW})|$$

to represent the distance between the vector X and a typical descriptor vector V_D .

There is, however, a problem that will cause this metric to need some modification.

As defined above, H_D gives equal weight to all words in the calibrated data base. However, some words clearly have better discriminating power than others, and this needs to be reflected in the definition.

To see how this comes about, consider the two ways that words can fail to discriminate. One way is for a word to appear in almost all project writeups. Such common words as: a, the, it, of, with, and, ... would clearly fail to help to identify projects as to their content. So also would words such as: advanced, novel, ... and others which seem to find their way into most project writeups.

The second way is for a word to appear only once, or at most a very few times, so that the word is likely random and thus not associated strongly with the project writeup's content. Typographical errors might fall into this category.

In the first case, the components V_{DW} for the given word W will be close to 1, for all descriptors D ; in the second case, the components will be close to 0 for all D . It follows that, for word W to be a good discriminator among the D requires that the V_{DW} vary widely over the D . To reflect this, modify the definition of the metric H_D as follows:

$$H_D = \sum_w |(X_w - V_{DW})| G_w$$

where G_w is a weighting function defined as

$$G_w = \max_D (V_{DW}) - \min_D (V_{DW})$$

Using this metric, the project is assigned descriptor D for which the metric is smallest.

In fact, one can also keep track of the smallest, second smallest, third smallest, etc. Using these, there are several possible interpretations. One is the assignment of probabilities that D is correctly assigned. Another is that the descriptor is to be apportioned (by funding level, perhaps) according to the various D in some way. Yet another is a fuzzy set interpretation that assigns partially to each of the D, yet not necessarily requiring partial assignments to add to 1. The question of how best to perform such an interpretation is at this point an open question.

Definition of "Word"

Although the foregoing model suffices, in principle, to define an assignment procedure that can be automated, there is yet one more refinement to add.

The refinement has to do with what it is that constitutes a word. In one sense, the matter is easily settled. A word is simply any string of characters (not itself containing a space) between two spaces. The problem, however, is more subtle.

Consider, for example, the words "optic," "optics," "optical," and "optically." These might appear to be four distinct words, and would according to the above model be treated as four distinct words. Nevertheless, the words are very similar semantically, and for maximum reliability in assigning ARO descriptors be treated as one word.

There are, in principle, several ways to do this. One could, at great effort and expense, compile a table of all English words, augmented by all technical, governmental, and military terms, and assign them to a subset of "root" words. Another way, at perhaps even greater expense, would be to develop rules of English by which one could constructively make the assignments. A much simpler way, though only approximate, is to truncate all but the first k characters, where k is a parameter to be determined. This is the only practical method, as is the one that will be used.

A problem, though, is how to best choose the truncation length k. If k is large, there is no truncation and therefore semantically equivalent variants of a single word will tend to be treated as distinct, as in the example above. If k is too small (e.g., k = 1) then words that are semantically very different will tend to be treated as identical. This is also incorrect. The best value of k will therefore lie between the extremes. To find the best value, tests were run based on a subset of the full data base.

It turned out that k was essentially flat between k = 4 and k = 7. Outside these values, the assignments became progressively erratic. However, k = 3 was not very much less reliable than k = 4.

It seems, based on the test runs, that truncating to the first 4 characters produces the best results, as well as reduces the size of the problem after calibration, and also the computation times for both calibration and descriptor assignment.

In itself, it seems both a surprising and counterintuitive result that a value as small as $k = 4$ should work as well as it does.

Preliminary Results

Test cases were run using small subsets of historic data to calibrate the algorithm. The same subsets of data were used as test data, to be assigned to ARO descriptors. Subsets, rather than the full historic data set, were used in order keep the test computations within a feasible computing time for test purposes. This procedure, in which calibration and assignment data are the same, will of course generate test results that may be distinctly optimistic. As a test of feasibility, however, the procedure serves as a reasonable indicator in the sense that, if the test results are poor when generated in this way then it is unlikely that the algorithm can be made to work under realistic conditions.

The full historic data base consists of 5915 projects that extend over about 540 ARO descriptors. Test cases, randomly selected from the full data base, have consisted of up to about 245 projects.

Tentative results, generated in this way, tend to show an assignment reliability of 98% or greater. This is clearly too optimistic to expect under realistic conditions, but at least demonstrates the feasibility of the procedure. It is to be understood that the research is ongoing, and that the results reported here are only a first cut, and are to be regarded as tentative and, as noted above, biased toward an optimistic outcome.

Computation times for the algorithm have required up to 60 hours on a 486/50 PC, using a Turbo Pascal implementation of the algorithm. The principal reason for the large computing time has to do with memory limitations. The calibration matrix V and other calibration data require too much memory to be kept in RAM. The algorithm was therefore implemented in such a way that the calibrated data was stored on disk. The classification procedure therefore required a large number of hard disk accesses, which are slow and which consumed the overwhelming portion of the total running time.

A production version of the algorithm will attempt to reduce the number of hard disk accesses that are needed.

Moreover, it has been observed that the definitions of the historic ARO descriptors had been chosen in a way that can be improved in two important respects: (1) descriptors can be eliminated or combined where there are found to be few or no project entries; and (2) descriptors can be re-defined to reduce or remove potential ambiguities in the way that they are likely to be assigned. A task is currently in progress to re-define the descriptors

accordingly, and this has largely been done. The revised list contains only 150 ARO descriptors, as compared with 540 previously. This is expected to sharply reduce the memory requirements and computation time, and to significantly improve the reliability of the results.

Tests are in process but results are not yet available.

Implementation Problems

Given, eventually, a successful test implementation with the full historic data base, using the revised ARO descriptors, there will remain several problems that will still need to be addressed. These include the following:

Even as revised, the ARO descriptors may not be optimal. In principle, there probably exists some kind of a natural clustering of the research projects that implies and corresponds to some optimal set of descriptors. The identification of such clustering and the associated descriptors remains to be done.

The assignment procedure depends upon the existence of a number of calibrated vectors v_D that represent the various descriptors D. Among the IRAD projects, however, will be a number which may be of interest to the Navy or to the Air Force, but which are not applicable to Army functions. Corresponding to these, there will be no ARO descriptor, except for the default that identifies them as "Not Applicable," or NA. Unlike the other descriptors, each of which represents projects with some common body of technologies and applications, the NA descriptor represents a broad collection of projects with little in common. The vector v_D that corresponds to descriptor NA will therefore not be "close" to typical vectors of NA projects. Typically, then, NA projects will appear to be closer to other descriptor vectors, and the projects will therefore tend to be misidentified. It may be possible to filter most of the NA projects by requiring, for a project to be identified with a descriptor, not only that the descriptor vector be closest to the project vector, but that the distance between them not exceed some empirically determined threshold.

From one year to the next, technologies change and Army functions (thus ARO descriptors) also change. The calibration that was valid for last year will therefore not be fully valid this year. Annual maintenance of the calibration, both as to technologies and as to Army functions, needs to be addressed.

Other Potential Applications of the Methodology

If successful, the approach used here to classify IRAD projects with respect to their Army functional relevance, as measured by ARO descriptors, might also be applied to other and unrelated problems. Representative examples might include:

In linguistics, it might be interesting to use this approach to study the semantic content of words, parts of words, and sequences of words. The empirical observation, alluded to above, that some significant semantic content is embodied in as few as the first 3 characters of a word, seems relevant.

In a related spirit, one might use this approach to study the psychology of how we humans organize and perceive and understand language. A simple version of such a study might take the form of presenting readers with standard English text, with all words truncated to no longer than k characters, for various k , and to observe the kinds of difficulties that the readers have in interpreting the truncated text.

In literature, forensics, history, and military intelligence, there arise questions of who wrote what. The approach used here might provide a useful approach in cases where literary samples are available from each of the candidates for authorship attribution, and the question were that of identifying the actual author.

An Application of Classification with
Potential Use in Reproductive Toxicology

Barry A. Bodt
Army Research Laboratory

Ronald J. Young
Edgewood Research, Development,
and Engineering Center

ABSTRACT. Mammalian sperm develop distinctive motion patterns during capacitation known as hyperactivated motility. Many studies now point to an association between hyperactivation and in vitro fertilization. A method for the objective determination of hyperactivation is sought as a tool for the clinical assessment of fertility, and as a marker for the investigation of sperm function. Hyperactivated motility is characterized by a change from progressive movement to a highly vigorous, nonprogressive random motion. Historically, the determination of the level of hyperactivated motility has been based on the visual inspection of the cell's path as recorded on film--an extremely lengthy process for a sample containing hundreds of cells. Recent advances in videomicrography allow the cell locations to be tracked by computer systems which record many motility characteristics for each cell (e.g., the straight line velocity). In this presentation we will discuss the application of statistical classification supporting the automated discrimination between hyperactivated and non-hyperactivated cells on the basis of their motility characteristics. We will also preview on-going work where the proportion of hyperactivated cells determined by the classification rule is used as a response in assessing the toxicological effect of certain metals.

1. INTRODUCTION. This work centered on the establishment of an automated procedure to classify rabbit sperm cells as to their motility, hyperactivated or non-hyperactivated. In Figure 1 we show the digitized representation of the swimming paths or tracks of several cells. Hyperactivated motion is described as a highly vigorous, nonprogressive, random motion (e.g., cell tracks 23, 27, 16 and 20 of Figure 1). Hyperactivation is the process of developing from a linear progressive motion (e.g., cell tracks 21, 41, 9, and 12 of Figure 1) to hyperactivated motion. The interest in hyperactivation is that it has been found to be strongly associated with capacitation, the biochemical/bio-physical changes a cell undergoes, enabling it for fertilization (Tesarik et al., 1990). Whereas the components of capacitation are not easily measured, the cell motions can be. Motility classification, supported by these measures, has potential as a marker for capacitation.

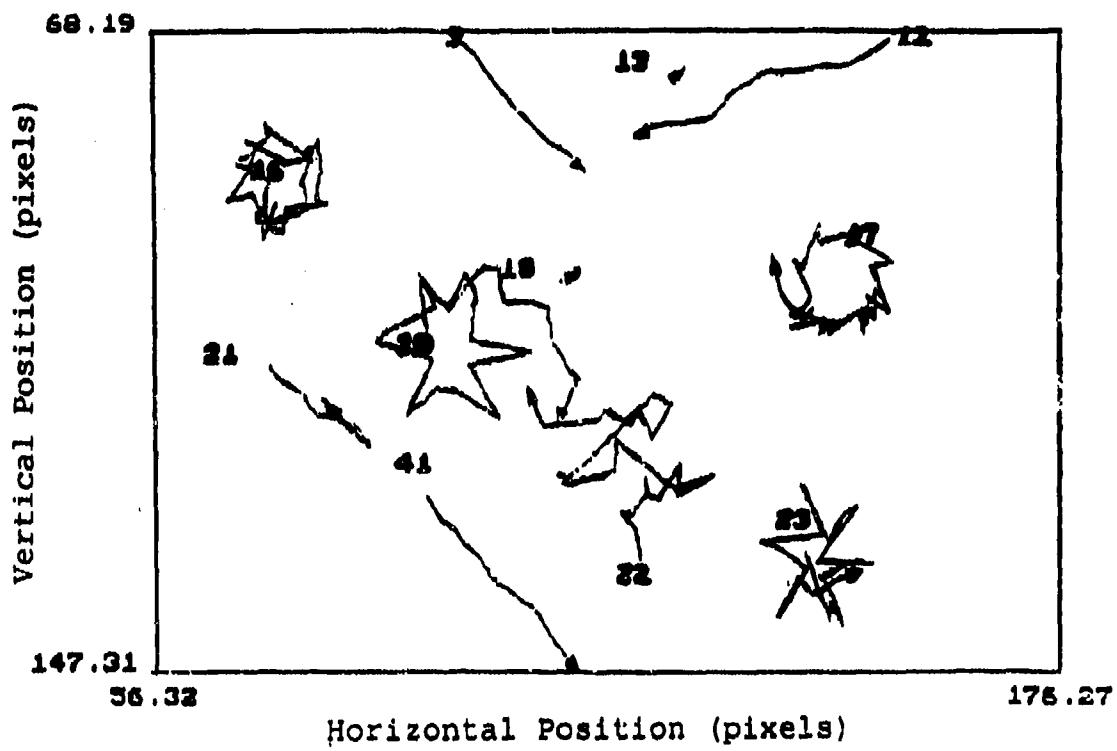


Figure 1. Digitized representation of several cell tracks.

Great improvements have been made in recent years in quantifying cell motions. Several systems are now on the market to support computer assisted videomicrography. This technology provides a VCR tape of the swimming motions of cells, a digitized record of cell motion, and motility parameter values for each cell. With this automation, it is much faster to characterize cells in terms of their motility parameter values--a process once carried out by hand as film frames were successively projected on a grid. The task in this study was to take cell tracks from previously determined hyperactivated or non-hyperactivated cells and establish a rule for classification based on these, now easy to establish, motility parameter values.

2. ESTABLISHING MOTILITY PARAMETERS. The motility parameters of 322 hyperactivated sperm obtained by incubation of sperm from four rabbits under the capacitation conditions of Bracket and Elephant (1975), and 899 non-hyperactivated sperm incubated for one or two hours in T medium were chosen for the statistical analysis. The hyperactivated sperm population contained the major types of hyperactivated motion noted in the literature. The parameters under study were VSL [velocity over a straight-line path ($\mu\text{m/sec}$)], VCL [velocity over a curvilinear path ($\mu\text{m/sec}$)], VAP [velocity over a smoothed curvilinear path; 5-point moving average ($\mu\text{m/sec}$)], LIN (VSL/VCL), STR (VSL/VAP), WOB (VAP/VCL), AALH [average amplitude of the lateral head displacement (μm)], MALH [maximum amplitude of the lateral head displacement (μm)], BCF [beat cross frequency (Hz)], Dance [AALH/LIN (μm)], and Dance Mean [VCL*AALH ($\mu\text{m}^2/\text{sec}$)].

3. PREVIOUS CLASSIFICATION MODELS. Others have attempted to use the motility parameter values for classification (e.g., Mortimer and Mortimer, 1990; Burkman, 1991) with reasonable success. A potential for further analysis was suggested because 1) LIN, a key measure in the decision, would in some cases be misleading, and 2) there was opportunity to employ more sophisticated means of statistical classification.

The first issue was the reliance in decision rules on LIN. In Figure 2, four possible tracks are given with the associated values for VCL, LIN, WOB, and AALH. From Mortimer and Mortimer, high values (> 0.60) for LIN are indicative of a non-hyperactivated or linear progressive motion, and low values (< 0.60) indicate hyperactivated motion. Figure 2a,b show non-hyperactivated and hyperactivated motions, respectively, where the values for LIN are consistent with the rationale for its use (i.e., when VSL and VCL are different, departure from linear progressive motion is present.) Figure 2c,d show non-hyperactivated motions where, because of the looping path of the cell, the values of LIN are in the range for hyperactivated motion. Thus, the measure LIN will in some instances mislead.

A second issue was in the classification methods employed. No article in the biological literature suggested using traditional statistical methods for classification. Techniques

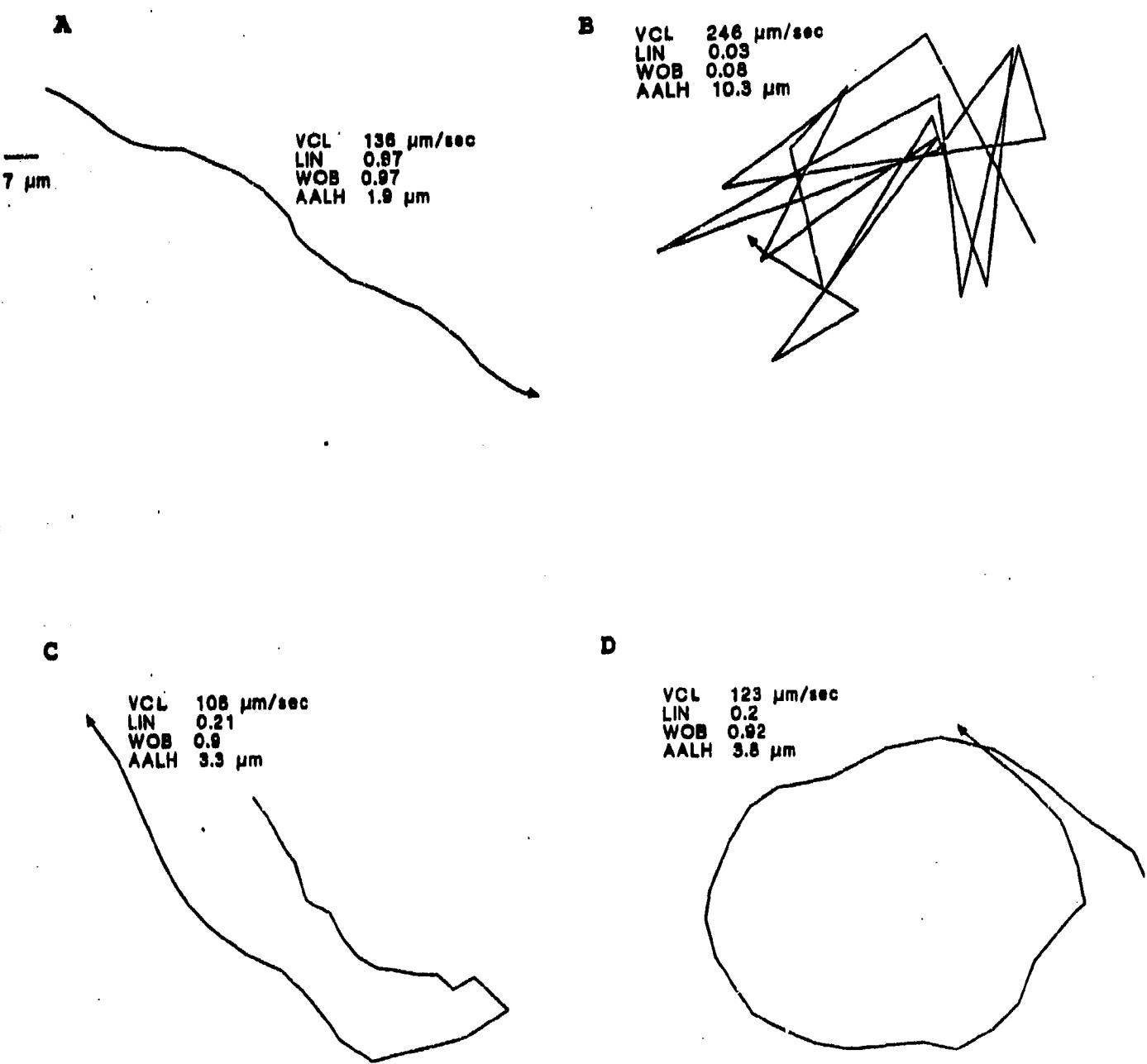


Figure 2. Motility parameter measurements with accompanying cell path display for motion types; a) non-hyperactivated, b) hyperactivated, and c-d) non-hyperactivated.

employed were, for example, comparing the univariate relative frequency histograms associated with the hyperactivated and non-hyperactivated cells, using t-tests to test the difference between mean values from the two groups, and examination of summary statistics from each group. The actual decision rule suggested by Mortimer and Mortimer would have required a cell to satisfy each of three constraints: $VCL > 100 \mu\text{m/sec}$, $LIN < 0.60$, and $AALH \geq 5 \mu\text{m}$. Each were determined individually.

4. GRAPHICAL ANALYSIS. Classification potential was assessed graphically by comparing the motility parameter distributions for hyperactivated and non-hyperactivated cells. Figure 3 shows unmodified parallel boxplots of the hyperactivated (H) and non-hyperactivated (N) class distributions for each motility parameter. Data were normalized to support examination of the relative classification potential among motility parameters. For LIN and WOB, at most 25% of the non-hyperactivated cells show values in the same range as those of the hyperactivated class, indicating that both have strong potential for classification. Based on the degree of separation for the inner 50% of the data, it is likely that AALH, MALH, and $VCL \cdot AALH$ would also be reasonable classifiers. Note that a classification rule based on VC alone did not appear promising.

The relative frequency distributions for LIN, VCL, AALH, and WOB are given for each motility class in Figure 4. LIN, VCL, and AALH were selected for display because of their prominence in the literature (2-4), and WOB for its importance in this study. Hyperactivated cells were absent in the range (0.8 - 1.0) for both LIN and WOB, and conversely high percentages of non-hyperactivated cells, LIN, 75.5% and WOB, 94.3% were found over this range. This strongly suggests good classifying potential for each. AALH shows only minimal distribution overlap. VCL has considerably more. The individual concomitants of hyperactivation suggested by Mortimer and Mortimer are reasonably consistent with these results despite the fact that rabbit sperm, not human sperm, values are reported here.

As a starting point for improvement, the rules suggested by Mortimer and Mortimer were implemented on our data. The results appear as Figure 5. In Figure 5a it can be seen that cells satisfying the VCL and AALH constraints (partitions have been overlaid) for hyperactivated motion are very likely hyperactivated, but a good number of cells not satisfying those constraints are also hyperactivated. VCL and AALH are linearly associated. In Figure 5b all three conditions are shown. Again, a number of hyperactivated cells do not meet the decision criteria. Of course the rules are being implemented on a species for which they were not intended. Further investigation of rules based on these parameters and our data was warranted.

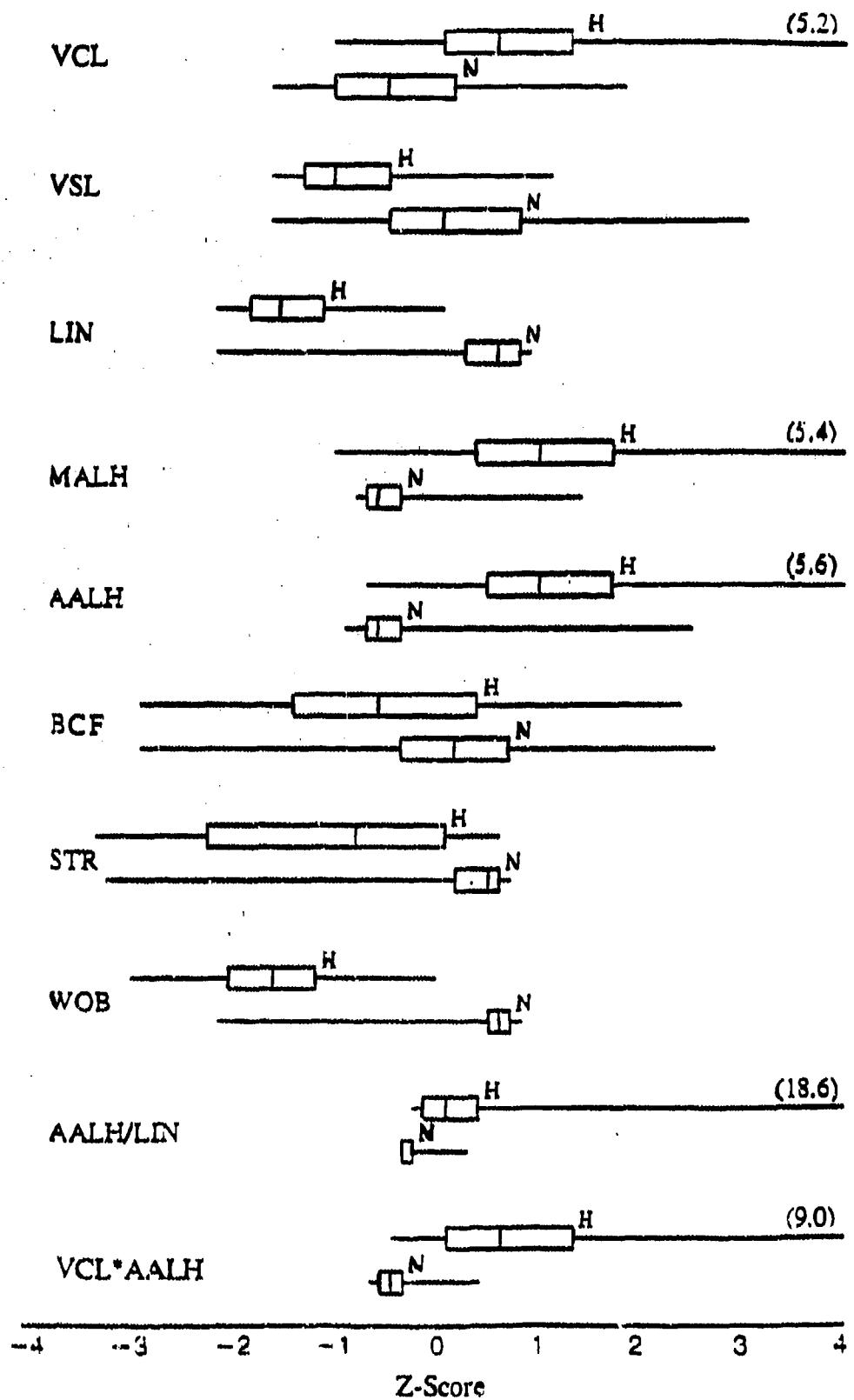


Figure 3. Parallel boxplots of motility parameter measures for hyperactivated (H) and non-hyperactivated (N) cells.

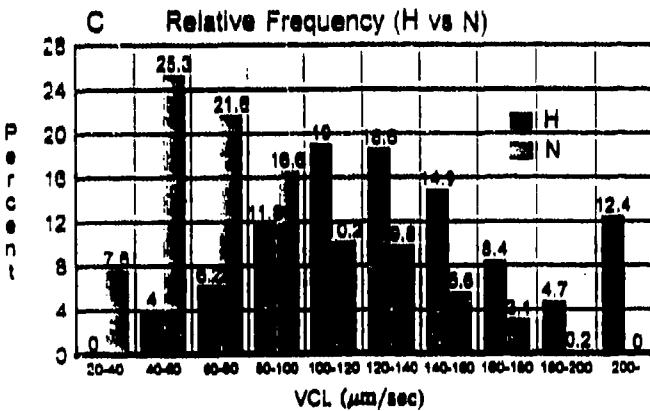
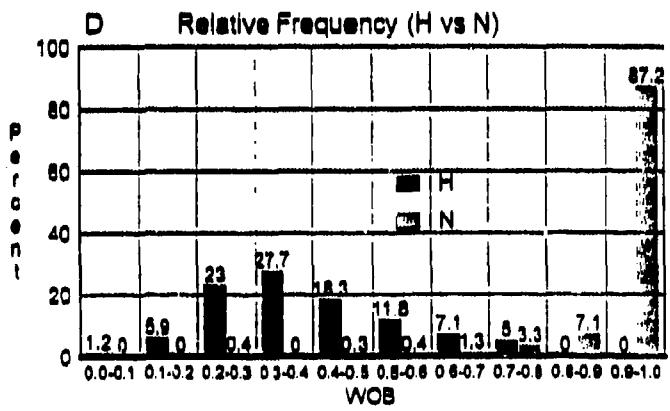
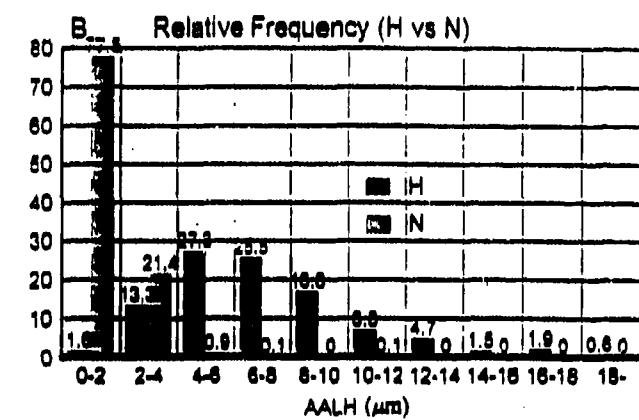
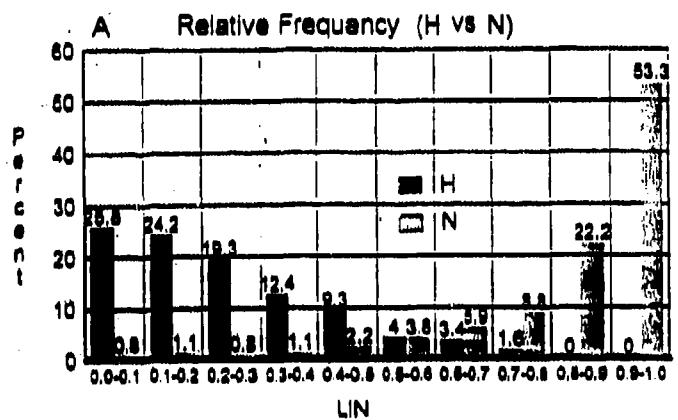


Figure 4. Relative frequency distributions for a) LIN, b)AALH, c) VCL, and d) WOB under hyperactivated (H) and non-hyperactivated (N) motility.

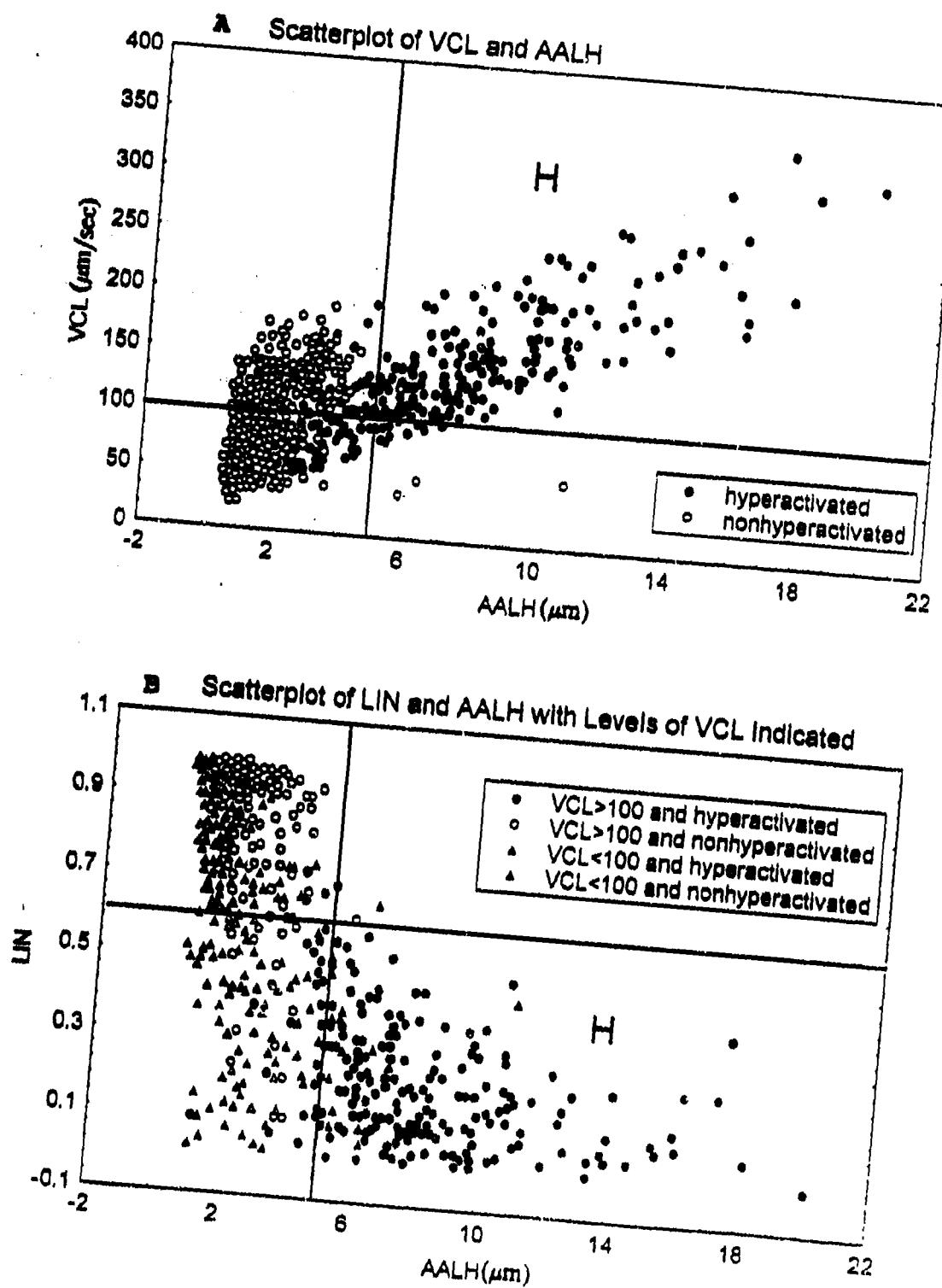


Figure 5. Scatterplots showing the association between actual hyperactivated motility and predicted hyperactivated motility, where the quadrant H signifies predicted hyperactivated motility based on VCL and AALH in a) and based on LIN and AALH with additional VCL information in b).

5. DISCRIMINANT ANALYSIS. Discriminant analysis and regression, on a (0,1) class variable, were used to explore models for classification. All possible subsets regression was used to select the best models for each number of independent variables. BMDP programs supporting stepwise discriminant analysis and regression were used in the analysis. Table 1 lists the results for individual motility parameters, the best two-parameter models, and the best three-parameter models. Though many models perform well, it is clear that WOB is the key motility parameter. Model 11, based on WOB and VCL, was judged to be the best. It had the highest efficiency, and included motility parameters which were not strongly linearly associated, correlation -0.47. (Interestingly, multiple correlation made the use of AALH, LIN, and VCL together an undesirable choice from a prediction standpoint.) The discriminant form of the model was selected but did not differ markedly from the regression model. Cells were classified as hyperactivated if

$$WOB < 0.596 + VCL * (6.76 * 10^{-4}).$$

Jackknifed cross validation procedures using BMDP software reinforced this model selection.

6. CART MODEL. Tree-structured methods were also used in developing a model (CART™, version 1.1, 1985 California Statistical Software, Inc.) The CART routine offers many options; only the defaults were used. Generally, CART works as follows for univariate partitions. Each possible predictor variable (motion parameter) for class is examined individually. For a specific variable, the program searches over all the values, resting at each one to see how efficient it would be to partition the data into hyperactivated and non-hyperactivated classes based on that value. (In our data set this requires over 1200 assessments of efficiency for each variable.) The routine notes the best value for that variable based on classification efficiency. The variable which partitions the data in the most efficient manner is selected and its value is used as the first partition of the data, creating two nodes, one each for hyperactivated and non-hyperactivated classes. Within each one, some cells may be misclassified. The routine then searches among the variables looking to further partition the two nodes to increase efficiency. The routine eventually settles on a decision tree for classification with maximum efficiency subject to the constraint that the tree complexity should not be great.

In running CART, all the motility parameters considered earlier as possible predictors were included. The result was that CART chose only WOB and VCL, with the rule: classify as hyperactivated if

$$VCL > 51 \text{ and } WOB < 0.79.$$

The decision tree is illustrated as Figure 6. Of the 1221 cases examined, only 12 non-hyperactivated cells and 2 hyperactivated

Table 1. Summary of best models using discriminant (D)/regression (R) analysis
based on 322 hyperactivated (H) and 899 nonhyperactivated (N) cells

Model	Variables	Misclassified (No.)		Efficiency (%)	
		H (D/R)	N (D/R)	D/R	R ²
1	WOB	22 / 31	19 / 15	96.64 / 96.23	0.838
2	LIN	21 / 34	63 / 46	93.12 / 93.45	0.702
3	AALH	59 / 91	5 / 4	94.76 / 92.22	0.639
4	MALH	63 / 109	16 / 9	93.53 / 90.34	0.600
5	VCL*AALH	113 / 188	4 / 0	90.42 / 84.60	0.411
6	STR	132 / 171	78 / 36	82.80 / 83.05	0.356
7	VCL	108 / 209	205 / 56	74.37 / 78.30	0.261
8	VSL	56 / 210	301 / 26	70.76 / 80.67	0.235
9	AALH/LIN	190 / 283	1 / 0	84.36 / 76.82	0.140
10	WOB, AALH	21 / 32	16 / 10	96.97 / 96.56	0.856
11	WOB, VCL	23 / 30	12 / 11	97.13 / 96.64	0.847
12	VCL LIN. AALH	23 / 34	34 / 29	95.33 / 94.84	0.757
13	VCL LIN. MALH	24 / 38	40 / 30	94.76 / 94.43	0.746
14	VCL LIN. VCL*AALH	24 / 40	50 / 36	93.78 / 93.78	0.729

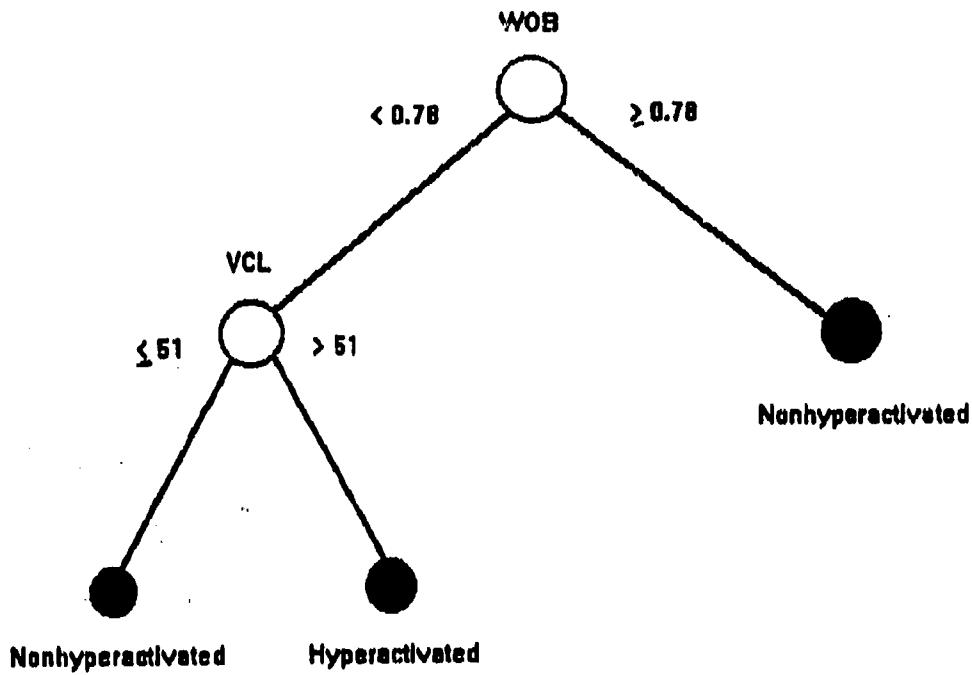


Figure 6. CART decision tree for determining hyperactivation.

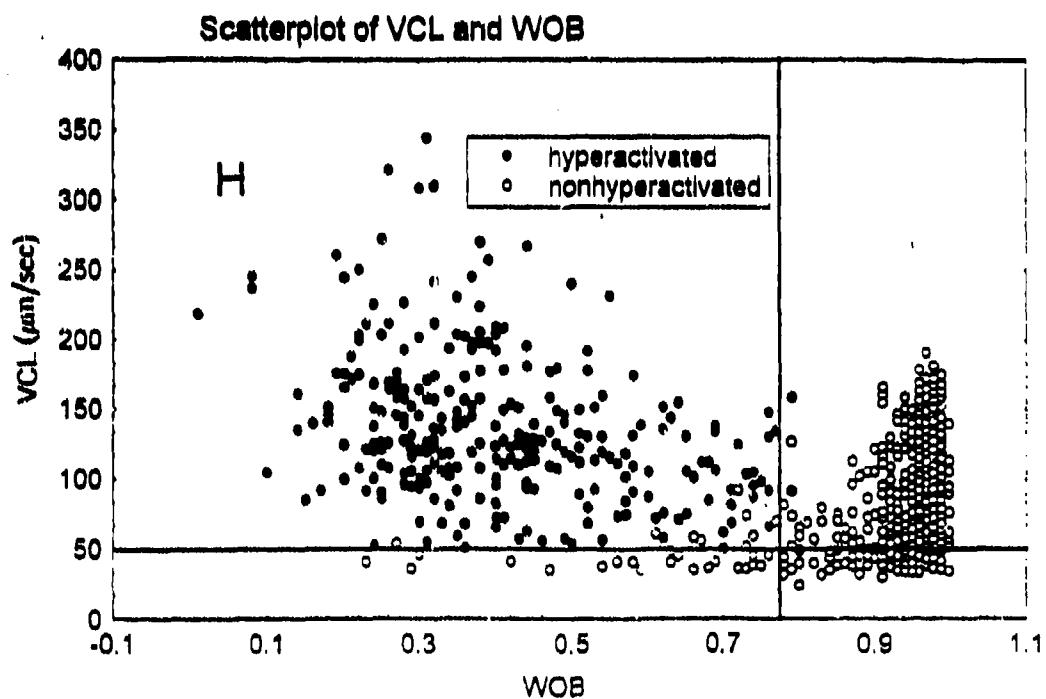


Figure 7. Scatterplot showing the association between actual hyperactivated motility and predicted hyperactivated motility using the CART model, where the quadrant H signifies predicted hyperactivated motility.

cells were misclassified for an efficiency of 98.9%. Cross validation attempts, holding out randomly selected subsets, consistently identified WOB and VCL as the important motility parameters. Application of this decision rule to our data appears as Figure 7.

The use of LIN, AALH, and VCL was also investigated. CART did not choose VCL. The tree was slightly more complex, having five nodes instead of three as above. The classification efficiency was 96.5%. When a model based on WOB and AALH was attempted, CART did not choose to use AALH, opting instead for a rule based only on WOB for an efficiency of 97.0%. Other runs using linear combinations of variables were attempted but resulted in more complex decision trees.

7. MODEL COMPARISON. Figure 8 illustrates decision criteria delivered by the discriminant and CART models using VCL and WOB. To understand the model differences we have partitioned the point set $WOB \times VCL$, where WOB ranges from 0.0 to 1.0 and VCL ranges from 0 to 350, according to the hyperactivity decision rules for each model. A cell whose WOB and VCL values locate it in a shaded region would be classified non-hyperactivated by CART. The unshaded region corresponds to a hyperactivated classification delivered by CART. The bold line represents the discriminant model. Points falling below that line would be classified as hyperactivated, and above, non-hyperactivated. Within each region we have indicated the true number of hyperactivated and non-hyperactivated cells present. From this one can assess their relative performance, and will find the CART model to be slightly better.

8. APPLICATION. An experiment was conducted in which sperm cells were exposed to metal ions in four concentrations over four different time periods. This factorial design was run within blocks (different rabbit donors). Cells were identified as hyperactivated or non-hyperactivated by the CART model established above. Initial graphical analysis (Figure 9) suggests an adverse effect induced by increased exposure to lead on the percentage of motile cells which exhibited hyperactivated motion. Since lead is known to be a reproductive toxicant, this might suggest that one impact is in its inhibition of hyperactivation.

9. SUMMARY. A classification problem in reproductive toxicology was approached using well known statistical procedures. We found classification criteria involving a different set of motility parameters than what had been suggested in the literature. Further, the combination of WOB and VCL performed better than the popular set of VCL, AALH, and LIN. Application of the new model is now being made to help uncover potential reproductive toxicants.

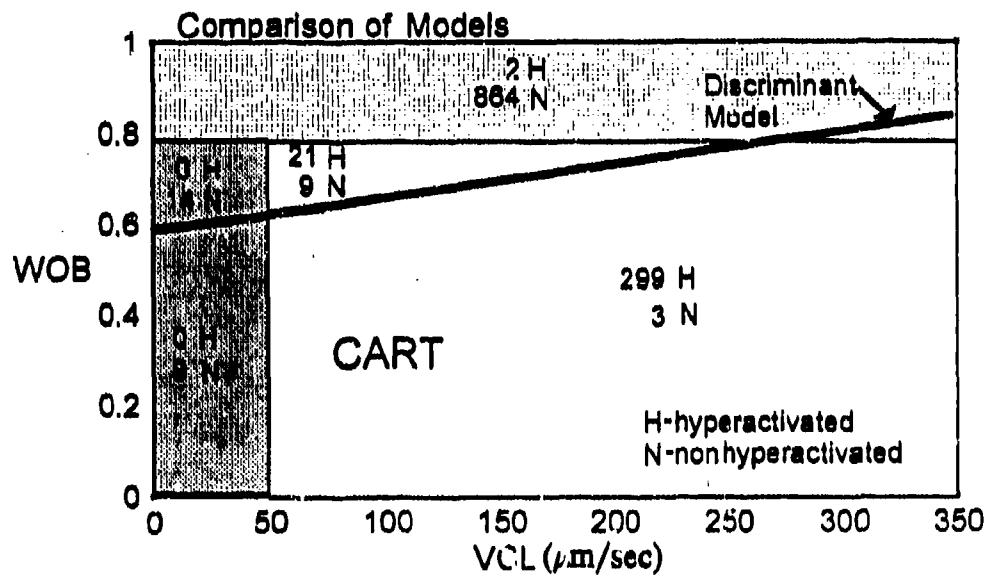


Figure 8. Comparison of models, where the unshaded region and the half-plane below the discriminant model denote regions for predicted hyperactivated motility by CART and discriminant analysis, respectively, and the actual counts for hyperactivated (H) and non-hyperactivated (N) motility are given.

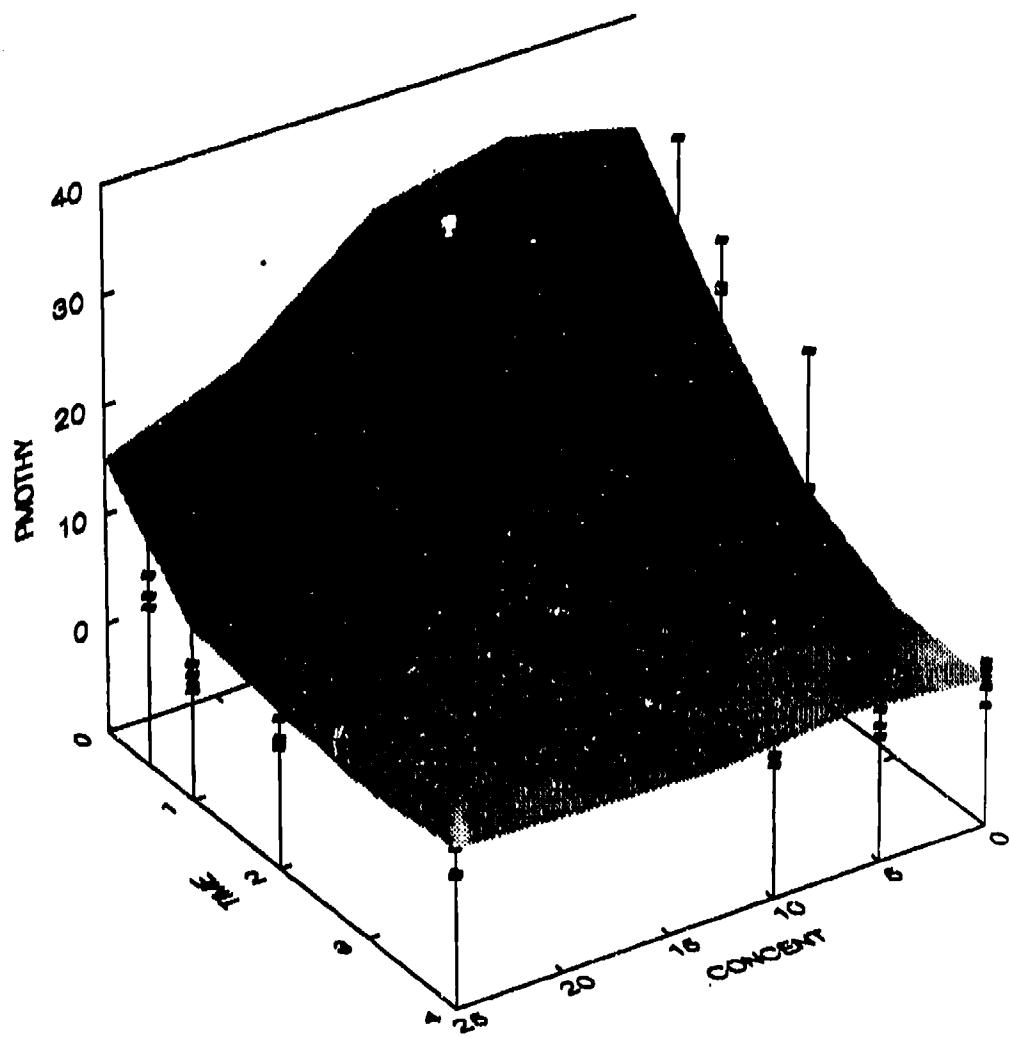


Figure 9. Smoothed scatterplot showing a possible interaction effect on the percent of motile cells which are hyperactivated (PMOTHY) attributable to lead exposure expressed in terms of time and concentration .

Literature Cited

1. Tesarik J, Mendoza Oltras C, Tesart J (1990): Effect of the human cumulus oophorus on movement characteristics of human capacitated spermatozoa. *J Reprod Fertil* 88:665-675.
2. Brackett B, Elephant G (1975): Capacitation of rabbit spermatozoa in vitro. *Biol Reprod* 12:260-274.
3. Mortimer, S, Mortimer, D (1990): Kinematics of human spermatozoa incubated under capacitating conditions. *J Androl* 11:195-203.
4. Burkman, L (1992): Discrimination between nonhyperactivated and classical hyperactivated motility patterns in human spermatozoa using computerized analysis. *Fertil Steril* 55:363-371.

Confidence Intervals and Tests of Hypotheses for Normal Coefficients of Variation

Mark G. Vangel

National Institute of Standards and Technology
Statistical Engineering Division, Bldg. 101/A-337
Gaithersburg, MD 20899-0001*

* this work was completed while the author was at the Army Research Laboratory, Materials Directorate, Watertown, MA 02172-0001

Abstract

This article presents an analysis of the small-sample distribution of a class of approximate pivotal quantities for a normal coefficient of variation which contains the approximations of McKay (1932), David (1949), the 'naïve' approximate interval obtained by dividing the usual confidence interval on the standard deviation by the sample mean, and a new interval closely related to McKay (1932). For any approximation in this class, a series is given for $e(t)$, the difference between the cdfs of the approximate pivot and the reference distribution. Let κ denote the population coefficient of variation. For McKay (1932), David (1949), and the 'naïve' interval $e(t) = O(\kappa^2)$, while for the new procedure $e(t) = O(\kappa^4)$. Examples involving strength data for a composite material are discussed.

Key Words: Noncentral t distribution, chi-squared approximation, McKay's approximation

1 Introduction

If X is a normal random variable with mean μ and variance σ^2 , then the parameter

$$\kappa \equiv \frac{\sigma}{\mu} \quad (1)$$

is called the population coefficient of variation. Let X_i for $i = 1, \dots, n$ be an independent random sample, with $X_i \sim N(\mu, \sigma^2)$ for each i . In terms of the usual

sample estimates of the normal parameters

$$\bar{X} = \sum_{i=1}^n X_i/n \quad (2)$$

and

$$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n - 1), \quad (3)$$

a point estimate of (1) is

$$K \equiv S/\bar{X}. \quad (4)$$

This statistic is widely calculated and interpreted, often for very small n , usually without an accompanying confidence interval. An exact method for confidence intervals on κ based on the noncentral t distribution is available (Lehmann, 1986, p. 352) but it is computationally cumbersome; hence the need for approximate intervals. In this article, we will investigate an approximate pivotal quantity which can be used to easily calculate confidence intervals and perform hypothesis tests on κ which attain very nearly the nominal confidence level or size. These calculations require only standard tables.

Let Y_ν denote a χ^2 random variable with $\nu = n - 1$ degrees of freedom, and define $W_\nu \equiv Y_\nu/\nu$. For $\alpha \in (0, 1)$, let $\chi_{\nu, \alpha}^2$ denote the 100α percentile of the distribution of Y_ν , and let $t \equiv \chi_{\nu, \alpha}^2/\nu$ be the corresponding quantile of W_ν . Define the random variable

$$Q \equiv \frac{K^2(1 + \kappa^2)}{(1 + \theta K^2)\kappa^2}, \quad (5)$$

where $\theta = \theta(\nu, \alpha)$ is a known function. If we choose θ so that

$$\Pr(Q \leq t) \approx \Pr(W_\nu \leq t) \quad (6)$$

then, since the distribution of W_ν is known and free of κ , we can use Q as an approximate pivot for constructing hypothesis tests and confidence intervals for κ . We define the accuracy of the approximation (6) to be $e(t) \equiv p - \alpha$, where $p \equiv \Pr(Q \leq t)$. Note that p is the actual confidence level of a one-sided confidence interval for κ^2 , based on Q , having nominal confidence α . In Section 2, we give a Taylor series expansion for $e(t)$ in powers of κ^2 , leaving the details to the Appendix. We then consider four choices for θ : corresponding to the approximations of McKay (1932) and David (1949), to the 'naïve' approximate interval obtained by dividing the usual confidence interval on the standard deviation by the sample mean, and to a new interval closely related to McKay (1932).

McKay (1932) proposed that Q and W_ν are approximately equal in distribution when $\theta = \nu/(\nu + 1)$, but he was unable to investigate the small-sample distribution of Q . Consequently, Fieller (1932) and Pearson (1932) performed a simulation study, with satisfactory results. David (1949) proposed McKay's approximation with $\theta = 1$; this suggestion has received much less attention than McKay (1932). Much later, Iglewicz and Myers (1970) compared selected quantiles of the approximate distribution for K , obtained from Q with McKay's choice of θ , with the corresponding exact values obtained using the noncentral t distribution. This numerical investigation demonstrated that McKay's approximation is very good, at least for $n \geq 10$ and $0 < \kappa \leq .3$. Instead of examining differences in quantiles numerically, we will investigate differences in cdfs analytically, and thereby develop a deeper understanding of the small-sample properties of these approximations.

2 A Taylor Series for $e(t)$

Denote the distribution of W_ν by $H_\nu(\cdot)$ so that, for $0 < \alpha < 1$, $H_\nu(t) = \alpha$. Since $u(x) = x/(\theta x + 1)$ is a monotone function with inverse $u^{-1}(y) = y/(1 - \theta y)$,

$$\begin{aligned} \Pr \left[\left(\frac{K^2}{1 + \theta K^2} \right) \left(\frac{1 + \kappa^2}{\kappa^2} \right) \leq t \right] &= \Pr \left(\frac{K^2}{1 + \theta K^2} \leq t \frac{\kappa^2}{1 + \kappa^2} \right) \\ &= \Pr \left[\left(\frac{K}{\kappa} \right)^2 \leq \frac{t}{1 + (1 - \theta t)\kappa^2} \right] \equiv p. \end{aligned} \quad (7)$$

For a given choice of $\theta(\nu, \alpha)$, we have defined the accuracy of the corresponding approximation to be $e(t) \equiv p - \alpha$. In the Appendix, we show that

$$\begin{aligned} e(t) &= t H'_\nu(t) \left\{ \left[\theta t - 1 + \frac{(1-t)\nu - 1}{\nu + 1} \right] \kappa^2 \right. \\ &\quad + \left[\frac{-6 + 11\nu - 6\nu^2 + \nu^3 - 3\nu t + 6\nu^2 t - 3\nu^3 t + 3\nu^3 t^2 - \nu^3 t^3}{4(1+\nu)^2} \right. \\ &\quad + \frac{(1-\nu+2\nu t)(1-\theta t)}{1+\nu} + \frac{(1-\nu+\nu t)(2-\nu+\nu t)(1-\theta t)^2}{2(1+\nu)} \\ &\quad \left. \left. + \frac{(2+\nu-\nu t)(1-\theta t)^2}{4} \right] \kappa^4 + O(\kappa^6) \right\}. \end{aligned} \quad (8)$$

For McKay's approximation, $\theta = \theta^{(1)} \equiv \nu/(\nu + 1)$, and (8) becomes

$$\begin{aligned} e_1(t) &\equiv tH'_\nu(t) \left\{ \frac{-2\kappa^2}{\nu+1} + \left[\frac{2+9\nu+2\nu^2-3\nu^3+2\nu^4}{2(1+\nu)^3} \right. \right. \\ &+ \left. \left. \frac{-3\nu t+9\nu^2 t+5\nu^3 t-7\nu^4 t-5\nu^2 t^2-2\nu^3 t^2+9\nu^4 t^2-5\nu^4 t^3+\nu^4 t^4}{2(1+\nu)^3} \right] \kappa^4 \right. \\ &+ O(\kappa^6) \}. \end{aligned} \quad (9)$$

David (1949) proposed Q with $\theta = \theta^{(2)} \equiv 1$ as an approximate pivot for a normal coefficient of variation. The accuracy of David's approximation is

$$\begin{aligned} e_2(t) &\equiv tH'_\nu(t) \left\{ \frac{(t-2)\kappa^2}{\nu+1} + \left[\frac{4+14\nu-10\nu^2+4\nu^3}{4(1+\nu)^2} \right. \right. \\ &+ \left. \left. \frac{-16t+4\nu t+18\nu^2 t-14\nu^3 t+6t^2-15\nu t^2-6\nu^2 t^2+18\nu^3 t^2}{4(1+\nu)^2} \right] \kappa^4 \right. \\ &+ \left. \frac{5\nu t^3-4\nu^2 t^3-10\nu^3 t^3+2\nu^2 t^4+2\nu^3 t^4}{4(1+\nu)^2} \right] \kappa^4 \\ &+ O(\kappa^6) \}. \end{aligned} \quad (10)$$

Another reasonable choice for θ is $\theta^{(3)} \equiv 1/t$. Confidence intervals are obtained for this choice of θ by simply dividing the endpoints of the usual confidence interval for σ by \bar{X} . The corresponding approximation has accuracy

$$\begin{aligned} e_3(t) &\equiv tH'_\nu(t) \left\{ \frac{\nu(1-t)-1}{\nu+1} \kappa^2 \right. \\ &+ \left. \left[\frac{-6+11\nu-6\nu^2+\nu^3-3\nu t+6\nu^2 t-3\nu^3 t+3\nu^3 t^2-\nu^3 t^3}{4(1+\nu)^2} \right] \kappa^4 + O(\kappa^6) \right\}. \end{aligned} \quad (11)$$

Finally, note that if

$$\theta = \theta^{(4)} \equiv \frac{2}{(\nu+1)t} + \frac{\nu}{\nu+1} = \frac{\nu}{\nu+1} \left[\frac{2}{\chi_{\nu,\alpha}^2} + 1 \right], \quad (12)$$

then the $O(\kappa^2)$ term in (8) is zero, and we have an approximation with accuracy

$$\begin{aligned} e_4(t) &\equiv tH'_\nu(t) \left\{ \left[\frac{-2-3\nu+12\nu^2-9\nu^3+2\nu^4+\nu t-15\nu^2 t+21\nu^3 t-7\nu^4 t}{2(1+\nu)^3} \right. \right. \\ &+ \left. \left. \frac{5\nu^2 t^2-16\nu^3 t^2+9\nu^4 t^2+4\nu^3 t^3-5\nu^4 t^3+\nu^4 t^4}{2(1+\nu)^3} \right] \kappa^4 + O(\kappa^6) \right\}. \end{aligned} \quad (13)$$

We will refer to the approximations corresponding to these four choices of θ as Approximations 1-4, or the McKay, David, Naïve, and Modified McKay approximations, respectively.

3 Discussion

If κ is small, as is usually the case in practice, $e_j(t)$ will also be small for $j = 1, \dots, 4$, so any of the above approximations will be satisfactory. For large samples, $\theta(j) \approx 1$ for $j \neq 3$; hence the three corresponding methods are asymptotically equivalent. Investigation of $e_1(t)$ and $e_2(t)$ demonstrates that David's approximation is not clearly better than McKay's, and, in any case, McKay's method is much more often used than David's. Also, Approximation 3, though adequate if κ is sufficiently small, is substantially less accurate than the other three approximations. We will therefore not consider David's and the Naïve approximation further, and restrict attention primarily to the McKay and Modified McKay approximations.

Denote $e(\cdot)$ regarded as a function of α by $\tilde{e}(\cdot)$, that is $\tilde{e}(\alpha) \equiv e[H_\nu^{-1}(\alpha)]$. The difference $\varphi(\alpha) = |\tilde{e}_1(\alpha)| - |\tilde{e}_4(\alpha)|$ will be positive when the Modified McKay approximation is more accurate than McKay's approximation, and negative otherwise. Hence this difference provides a means for comparing these two methods. Using the noncentral t distribution, it is straightforward to evaluate $\varphi(\alpha)$ exactly, and this is preferable to using the approximate formulas of the previous section. In Figure 1, results are displayed of computing $\varphi(\alpha)$ numerically, for 20 values of κ between .025 and .5; for sample sizes of 2, 5, 10, and 25; and for α equal to .01, .05, .95, and .99. Note that the Modified McKay method is usually more accurate than McKay's method. What is not clear from these differences is that, particularly when κ is small, the Modified McKay approximation is often extremely accurate: in fact, virtually exact. This point is made by Figure 2, which shows the accuracies of these two methods (as determined from the noncentral t distribution), as functions of α , for a sample size of 5, and for $\kappa = .05$ and $\kappa = .25$, respectively.

4 Confidence Intervals and Hypothesis Tests

In this section, we illustrate how the approximate pivot (5) can be used for approximate confidence intervals and one- and two-sample hypothesis tests. We assume that κ is positive, and that the probability of X being negative is negligible.

A $100(1 - \alpha)\%$ approximate confidence interval based on (5) is

$$\Lambda = \left[\frac{K}{\sqrt{t_1(\theta_1 K^2 + 1) - K^2}}, \frac{K}{\sqrt{t_2(\theta_2 K^2 + 1) - K^2}} \right], \quad (14)$$

where $t_1 \equiv \chi_{\nu, 1-\alpha/2}^2 / \nu$ and $t_2 \equiv \chi_{\nu, \alpha/2}^2 / \nu$. One-sided intervals can be determined similarly. If we let $u_i \equiv \nu t_i$, for $i = 1, 2$, then we can write the McKay and Modified McKay confidence intervals as

$$\Lambda_1 = \left\{ K \left[\left(\frac{u_1}{\nu+1} - 1 \right) K^2 + \frac{u_1}{\nu} \right]^{-1/2}, K \left[\left(\frac{u_2}{\nu+1} - 1 \right) K^2 + \frac{u_2}{\nu} \right]^{-1/2} \right\} \quad (15)$$

and

$$\Lambda_4 = \left\{ K \left[\left(\frac{u_1+2}{\nu+1} - 1 \right) K^2 + \frac{u_1}{\nu} \right]^{-1/2}, K \left[\left(\frac{u_2+2}{\nu+1} - 1 \right) K^2 + \frac{u_2}{\nu} \right]^{-1/2} \right\}, \quad (16)$$

respectively.

Since $(1 + \kappa^2)/\kappa^2$ in (5) is a monotone function of κ^2 , we can also use (5) to test the null hypothesis $H_0 : \kappa = \kappa_0$, for some known κ_0 . An endpoint of the interval (14) does not exist if $t(\theta K^2 + 1) - K^2 \leq 0$, or equivalently,

$$K^2 \geq \theta \frac{t}{1-t}. \quad (17)$$

In order for (17) to hold for the choices of θ considered in this article, either K^2 must be large or t must be small. Neither of these conditions are likely to occur in practice except possibly when n and t are both very small. If K is small but (17) holds, then one can either reduce the confidence level, increase the sample size, or else use the exact method based on the noncentral t distribution. Note that if $\theta_i = \theta_i^{(3)} = 1/t_i$ for $i = 1, 2$, then (14) becomes

$$\Lambda_3 = \left(K/\sqrt{t_1}, K/\sqrt{t_2} \right), \quad (18)$$

which is the usual interval on σ , with the endpoints divided by X .

Assume that we are given two independent random samples of sizes n_1 and n_2 , having population coefficients of variation κ_1 and κ_2 , with sample estimates K_1 and K_2 , respectively. From (6) we see that

$$\frac{K_1^2(1 + \theta K_2^2)}{K_2^2(1 + \theta K_1^2)} \sim \frac{\kappa_1^2(1 + \kappa_2^2)}{\kappa_2^2(1 + \kappa_1^2)} F_{\nu_1, \nu_2} \equiv \tau F_{\nu_1, \nu_2}, \quad (19)$$

where F_{ν_1, ν_2} denotes an F random variable with $\nu_1 = n_1 - 1$ numerator and $\nu_2 = n_2 - 1$ denominator degrees of freedom. When $\kappa_1 = \kappa_2$, $\tau = 1$, and it is easy to show that τ is monotone increasing in $\rho \equiv \kappa_1^2/\kappa_2^2$. In fact, since we are assuming that both κ_1 and κ_2 are small, $\tau \approx \rho$. Hence, we have an approximate F test for the equality of two coefficients of variation, analogous to the usual F test for the equality of variances.

5 Examples

The tensile strength of five specimens of a composite material are as follows (in 1000 Psi): 326, 302, 307, 299, 329. We have $X_1 = 312.6$ and $S_1 = 13.94$, so that $K_1 = .045$, $u_1 = \chi^2_{2,975} = 11.14$, and $u_2 = \chi^2_{4,025} = .4844$. Equations (15) and (16) lead to confidence intervals on κ by the McKay and Modified McKay methods, respectively. For this example the Modified McKay procedure gives the 90% confidence interval (.0269, .1299), which differs from the McKay interval only in the fourth decimal place.

Five specimens of the same material are tested in shear, giving shear strengths as follows (in 1000 Psi): 9.7, 9.6, 9.4, 9.4, 10.9. For these shear data, $X_2 = 9.8$, $S_2 = .6285$, and $K_2 = .064$. To test the null hypothesis that the population coefficient of variation for tensile strength equals the corresponding value for shear strength, we compute (for the McKay method)

$$\frac{K_1^2[1 + \nu_2/(\nu_2 + 1)K_2^2]}{K_2^2[1 + \nu_1/(\nu_1 + 1)K_1^2]} = \frac{.045^2[1 + (.8)(.064^2)]}{.064^2[1 + (.8)(.045^2)]} \doteq .495. \quad (20)$$

Since the probability that an $F_{4,4}$ random variable is less than .495 is .256, there appears to be insufficient evidence to reject this null hypothesis. Note that the Modified McKay method is not appropriate for this significance test since $\theta^{(4)}$ is a function of α .

6 Conclusion

A class of approximate pivotal quantities for a normal coefficient of variation related to the approximation of McKay (1932) has been investigated analytically, with particular emphasis on four special cases. The most important results are that, if κ denotes the population coefficient of variation, then the difference between the actual and nominal levels of McKay's (1932) confidence interval are of $O(\kappa^2)$, and that a very slight modification of McKay's method leads to an appar-

ently new $O(\kappa^4)$ method which is usually superior to McKay (1932), and which is recommended.

Appendix: Derivation of Equation (8)

For most applications κ will be small, so our plan is to let

$$q = q(\kappa^2) \equiv \frac{t}{1 + (1 - \theta t)\kappa^2} \quad (21)$$

and to expand $\Pr[(K/\kappa)^2 \leq q]$ in a Taylor series in κ^2 , then to expand each term in this series again in powers of κ^2 , using (21). We assume throughout that q is nonnegative; this imposes slight restrictions on α and κ which are not important in practice.

The random variables X and S are equal in distribution to

$$\bar{X} = \mu + Z\sigma/\sqrt{n} \quad (22)$$

and

$$S = \sigma\sqrt{W_\nu}, \quad (23)$$

respectively; where $Z \sim N(0, 1)$, and Z and W_ν are independent. Hence

$$\left(\frac{K}{\kappa}\right)^2 = W_\nu (1 + \kappa Z/\sqrt{n})^{-2} = \left[\frac{T_{\nu,\delta}}{\delta}\right]^{-2}, \quad (24)$$

where $T_{\nu,\delta}$ denotes a noncentral-t random variable with degrees of freedom ν and noncentrality parameter $\delta \equiv \sqrt{n}/\kappa$. By conditioning on Z and expanding in a Taylor series about $Z = 0$, we have that

$$\begin{aligned} p = \Pr\left[\left(\frac{K}{\kappa}\right)^2 \leq q\right] &= E\left\{H_\nu\left[q\left(1 + \frac{\kappa Z}{\sqrt{n}}\right)^2\right]\right\} = H_\nu(q) + qH'_\nu(q) \cdot \{ \\ &\left[\frac{(1-q)\nu - 1}{\nu + 1} \kappa^2 + \frac{-6 + 11\nu - 6\nu^2 + \nu^3 - 3\nu q + 6\nu^2 q - 3\nu^3 q + 3\nu^3 q^2 - \nu^3 q^3}{4(\nu + 1)^2} \kappa^4 \right] \\ &+ O(\kappa^6)\}. \end{aligned} \quad (25)$$

Using (21), the terms in (25) can now be expanded in powers of κ^2 about $\kappa^2 = 0$, giving

$$\begin{aligned} H_\nu(q) &= H_\nu(t) + t(\theta t - 1)H'_\nu(t)\kappa^2 \\ &+ t(\theta t - 1)^2/2 [2H''_\nu(t) + tH'_\nu(t)]\kappa^4 + O(\kappa^6), \end{aligned} \quad (26)$$

$$H'_\nu(q) = H'_\nu(t) + t(\theta t - 1)H''_\nu(t)\kappa^2 + O(\kappa^4), \quad (27)$$

and

$$\frac{(1-q)\nu - 1}{\nu + 1} = \frac{(1-t)\nu - 1}{\nu + 1} + \frac{(\theta t - 1)(\nu - 1 - 2\nu t)}{\nu + 1}\kappa^2 + O(\kappa^4). \quad (28)$$

Using the identity

$$t^2 H''_\nu(t) = \left[\frac{\nu}{2}(1-t) - 1 \right] t H'_\nu(t), \quad (29)$$

substituting (26), (27), and (28) into (25), and collecting terms in κ^2 leads to (8).

References

- [1] David, F. N. (1949), "Note on the Application of Fisher's k -statistics", *Biometrika*, 36, 383-393.
- [2] Fieller, E., (1932), "A Numerical Test of the Adequacy of A. T. McKay's Approximation", *Journal of the Royal Statistical Association*, 95, 699-702.
- [3] Iglewicz, B. and Myers, R. H. (1970), "On the Percentage Points of the Sample Coefficient of Variation", *Technometrics*, 12, 166-169.
- [4] Lehmann, E. L. (1986), *Testing Statistical Hypotheses*, 2nd. edition, John Wiley and Sons, New York.
- [5] McKay, A. T. (1932), "Distribution of the Coefficient of Variation and the Extended 't' Distribution", *Journal of the Royal Statistical Society*, 95, pp. 695-698.
- [6] Pearson, E. S. (1932), "Comparison of A. T. McKay's Approximation With Experimental Sampling Results", *Journal of the Royal Statistical Association*, 95, 703-703.

Figure 1: $\Phi(\alpha)$ Vs. κ for $n=2, 5, 10$ and 25

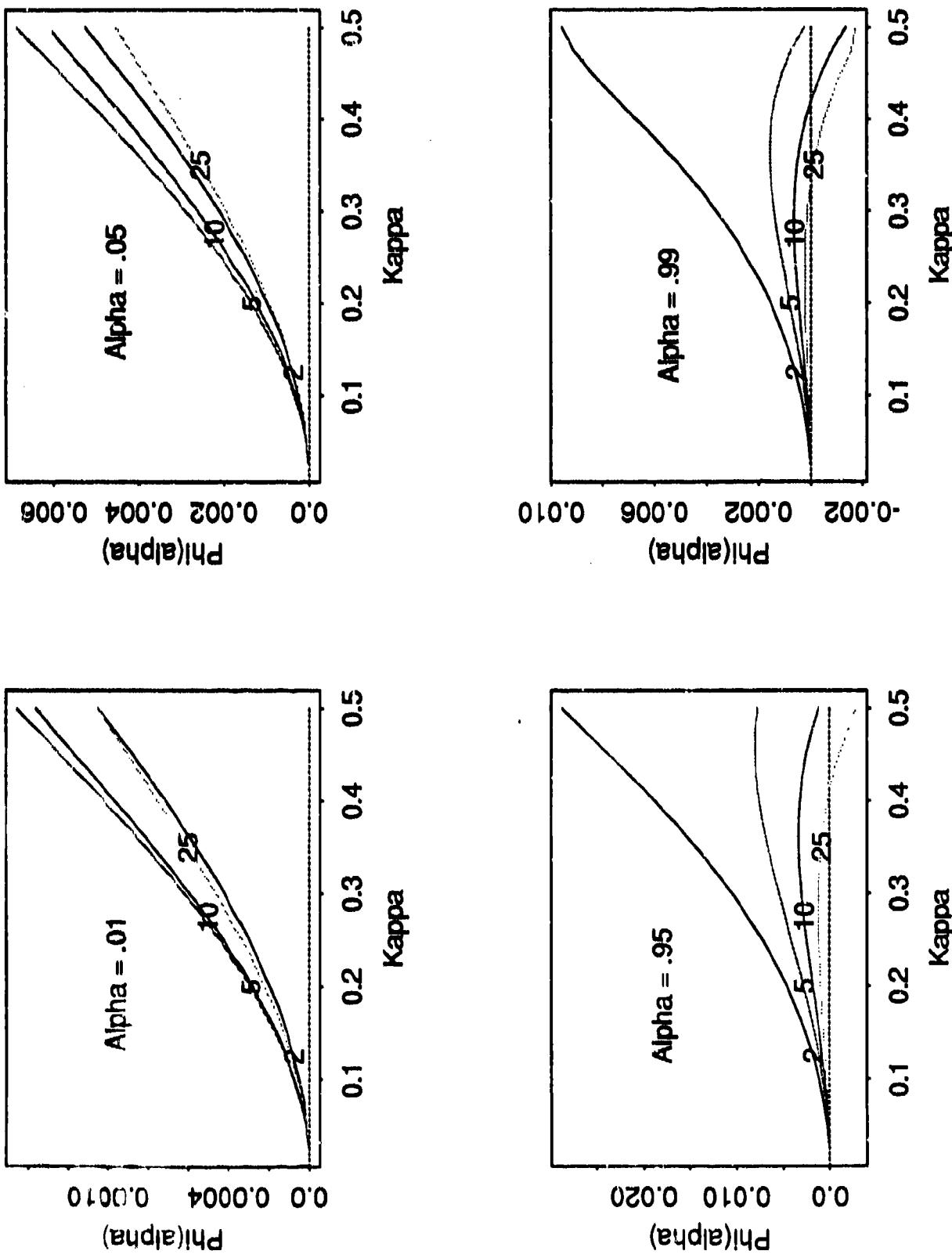
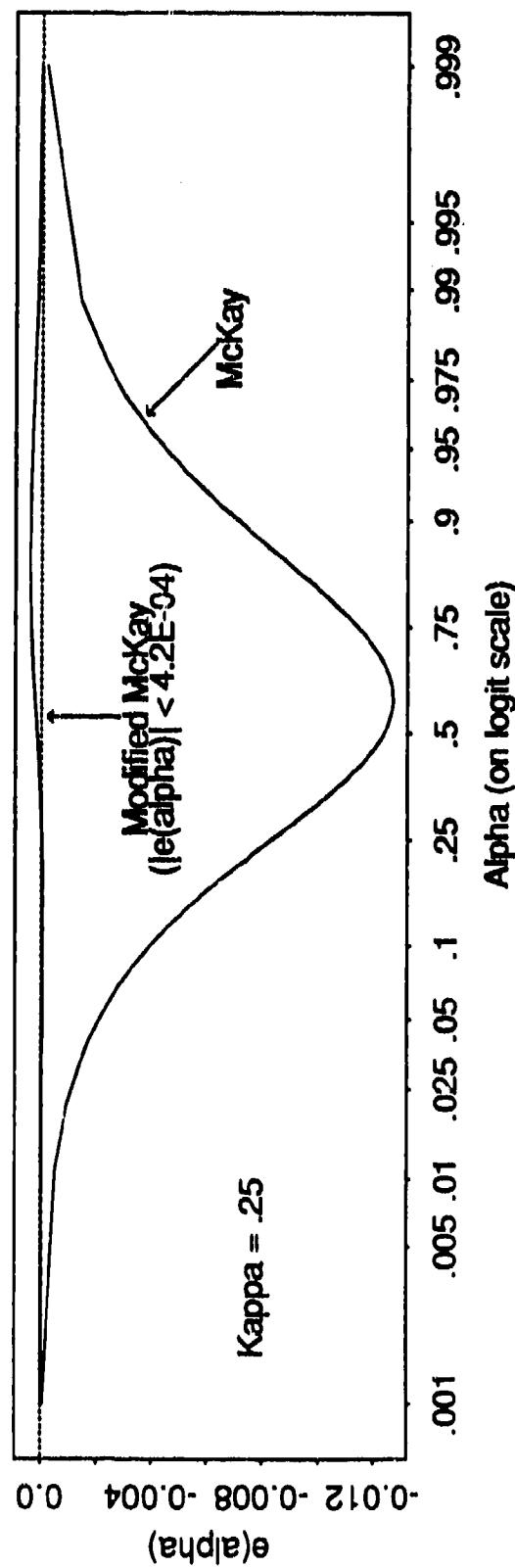
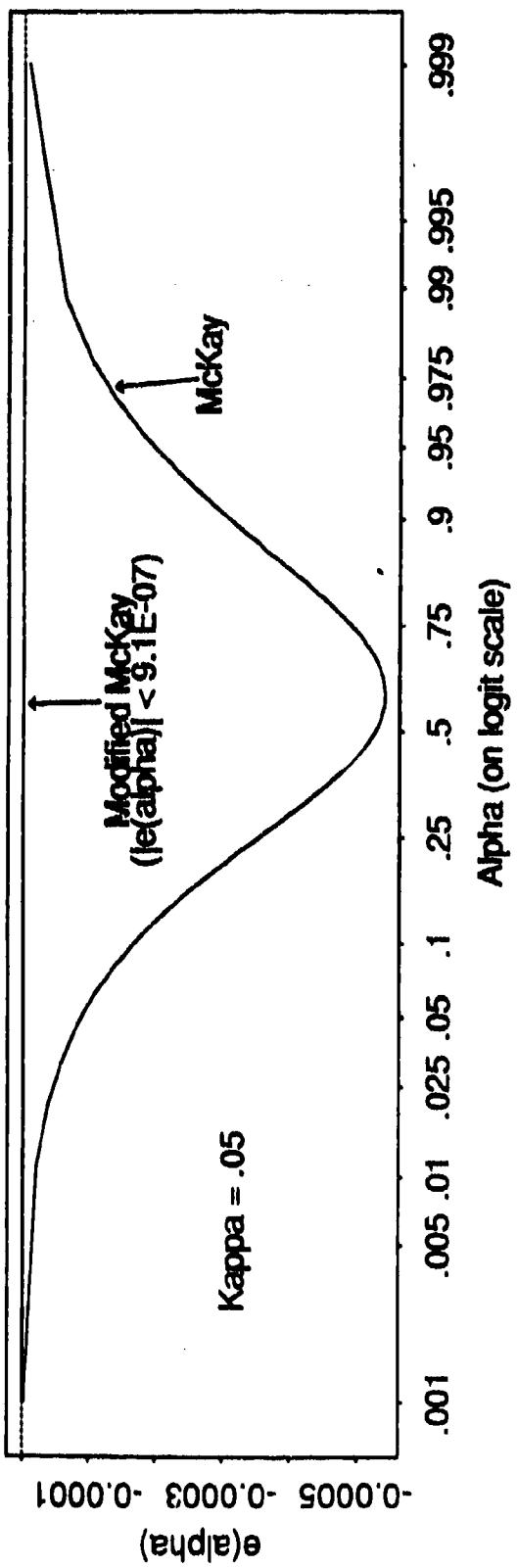


Figure 2: Exact Accuracies of McKay and Modified McKay Vs. Alpha
 $(n=5, \kappa=.05 \text{ and } .25)$



Improved periodogram-based estimators of frequency for the Cosinor model

By R. JOHN WEAVER, MARSHALL N. BRUNDEN and JONATHAN RAZ

The Upjohn Company, Kalamazoo, Michigan 49001

and the University of Michigan, Ann Arbor, Michigan 48109

SUMMARY

This paper discusses the estimation of the parameters of the Cosinor model. The standard periodogram-based approach of Walker (1971) produces biased estimates of all the parameters, and it will be shown that the bias in the frequency estimate can be substantial. An alternative periodogram-based estimator of frequency is proposed and is shown to have minimal bias.

Some key words : Cosinor model; Frequency estimation; Periodogram; Time series.

1. INTRODUCTION

Suppose we have a time series $y_t : t = 1, 2, \dots, n$, where y_t is the observation taken at time t . The model we consider is

$$y_t = \alpha_0 \cos(\omega_0 t) + \beta_0 \sin(\omega_0 t) + \varepsilon_t, \quad (1)$$

where the errors ε_t are assumed to be independent with $E(\varepsilon_t) = 0$ and $\text{Var}(\varepsilon_t) = \sigma^2$ for all t . This model was proposed for the analysis of biological rhythms by Halberg, Tong and Johnson (1965), who called it the Cosinor model. Further details have been given in Halberg, et al. (1972), Nelson, et al. (1979) and Bingham, et al. (1982). This model has been extensively used and reported in the chronobiology literature, and computer programs for its implementation have been published by Monk and Fort (1983) and Vokac (1984).

In matrix notation, this model can be expressed as

$$\mathbf{y} = \mathbf{X}_0 \boldsymbol{\phi}_0 + \boldsymbol{\varepsilon}$$

where \mathbf{y} is the $n \times 1$ vector of the observations, \mathbf{X}_0 is the $n \times 2$ design matrix with $\cos(\omega_0 t)$ in row t of the first column, and $\sin(\omega_0 t)$ in row t of the second column. The 2×1 vector $\boldsymbol{\phi}_0$ equals $(\alpha_0, \beta_0)^T$ and $\boldsymbol{\varepsilon}$ is the $n \times 1$ vector of error terms. In the conventional Cosinor model, the frequency (or equivalently the period) is considered to be fixed and known, and the subscript 0 indicates that \mathbf{X}_0 is a function of the true parameter value ω_0 . In this case, the model is linear in the unknown parameters α_0 and β_0 , and the usual least squares estimates of $\boldsymbol{\phi}_0$ apply.

Many times, however, we may not know the true frequency and we need a way to estimate all the parameters simultaneously. A common approach is to use the method of nonlinear least squares estimation. For the Cosinor model, these methods have problems with converging to local rather than global minima, and are extremely sensitive to the choice of starting values. Alternative methodology is desirable.

2. STANDARD PERIODOGRAM ESTIMATORS FOR THE COSINOR MODEL

A natural approach to this estimation problem would be to use a method based on the periodogram, or some function of it. The periodogram has long been used in the hidden periodicity problem, and E.T. Jaynes (1987) has demonstrated that it is a sufficient statistic for inferences about a single stationary frequency, when the errors are normally distributed.

Periodogram-based estimators for the Cosinor model were proposed by Whittle (1952), and have been extensively discussed by Walker (1971), who derived many of their properties. The estimators are

$$\hat{\alpha}(\hat{\omega}) = \frac{2}{n} \sum_{i=1}^n y_i \cos(\hat{\omega}t) \quad (2)$$

$$\hat{\beta}(\hat{\omega}) = \frac{2}{n} \sum_{i=1}^n y_i \sin(\hat{\omega}t) \quad (3)$$

where $\hat{\omega}$ is such that

$$I_n(\hat{\omega}) = \max_{0 < \omega < \pi} I_n(\omega)$$

and

$$I_n(\omega) = \frac{2}{n} \left| \sum_{i=1}^n y_i e^{i\omega t} \right|^2 = \frac{n}{2} (\hat{\alpha}(\omega)^2 + \hat{\beta}(\omega)^2).$$

The expression $I_n(\omega)$ is one of the usual definitions of the periodogram, and we will refer to it as the Standard periodogram. The estimator $\hat{\omega}$ is defined as the value of ω that results in the absolute maximum of $I_n(\omega)$ for $0 < \omega < \pi$. This estimate of ω is used in (2) and (3) to get the estimates of α_0 and β_0 . These estimators are equal to the least squares estimates when ω_0 is known and $n = cP_0$, and are called approximate least squares estimates (Bloomfield, 1976). Walker proved that they are consistent and asymptotically normal, and gave expressions for their asymptotic variance matrix. Rice and Rosenblatt (1988) show that the estimates of α_0 and β_0 will be consistent only when ω_0 is estimated with precision $\sigma(n^{-1})$, and that the asymptotic theory should be used cautiously.

In obtaining the Walker estimates of frequency, Diggle (1990) suggested considering all frequencies $0 < \omega < \pi$, not just the Fourier frequencies. The algorithm searches for the ordinate that results in a maximum along a grid of specified length, centered on the Fourier frequency that produces the maximum periodogram. This type of approach is also discussed in Rice and Rosenblatt (1988) and Zhao-Guo (1988).

The Walker estimators have good asymptotic properties, but can be significantly biased for a moderate length time series. If the model holds and $E(y_i) = \alpha_0 \cos(\omega_0 t) + \beta_0 \sin(\omega_0 t)$, then

$$E(\hat{\alpha}(\omega) | \omega) = \frac{\alpha_0}{n} [C_n(\omega + \omega_0) + C_n(\omega - \omega_0)] \\ + \frac{\beta_0}{n} [S_n(\omega + \omega_0) - S_n(\omega - \omega_0)],$$

where

$$S_n(u) = \sin\left(\frac{n+1}{2}u\right) \frac{\sin(\frac{n u}{2})}{\sin(\frac{u}{2})}$$

$$C_n(u) = \cos\left(\frac{n+1}{2}u\right) \frac{\sin(\frac{n u}{2})}{\sin(\frac{u}{2})}$$

This expression simplifies to the following if $\omega = \omega_0$

$$E(\hat{\alpha}(\omega) | \omega) = \alpha_0 + \frac{\alpha_0}{n} C_n(2\omega_0) + \frac{\beta_0}{n} S_n(2\omega_0)$$

Similar expressions hold for β_0 .

Any bias or estimation error in $\hat{\omega}$ may result in additional bias in the estimates of α_0 and β_0 , and even if we are able to estimate ω_0 exactly, our estimates of α_0 and β_0 will still be biased if n is not an integer multiple of the true period.

3. BIAS IN THE STANDARD PERIODOGRAM ESTIMATOR OF FREQUENCY

There has been little published on the bias of the Walker estimator of ω_0 . The exact bias has not been determined, but Bloomfield gives an indication of the approximate bias, credited to Whittle (1952), as

$$E(\hat{\omega}) = \omega_0 + \text{terms involving } \frac{1}{n}.$$

Rice and Rosenblatt also discuss the bias of the frequency estimate, and show for a moderate length data series, the bias can be significant. In a simulation with $\alpha_0 = 8$, $\beta_0 = 0$, $\omega_0 = 0.5$, and $n=100$, the bias was shown to be .0013, which is more than twice the standard error indicated by the asymptotic theory.

An analytic expression for the bias of the frequency estimator cannot be derived. As an alternative, we will consider an approximation suggested by the work of Rice and Rosenblatt (1988). To measure the bias in the estimator of ω , we will approximate $E(\hat{\omega})$ by the value of ω that maximizes the expected value of the periodogram.

To derive the expectation, the periodogram is reexpressed in matrix notation. As in Section 1, define the $n \times 2$ matrix

$$X(\omega) = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ \cos(2\omega) & \sin(2\omega) \\ \vdots & \vdots \\ \cos(n\omega) & \sin(n\omega) \end{pmatrix}$$

For simplicity, we will write it as X in the discussion that follows, remembering that it is actually a function of ω . Again, let us denote $X(\omega_0)$ as X_0 . The Walker estimators of α_0 and β_0 can then be written in matrix notation as

$$\hat{\phi}_w = \begin{pmatrix} \alpha_w \\ \beta_w \end{pmatrix} = \frac{2}{n} X'y$$

and the Standard periodogram can be expressed as

$$I_n(\omega) = \frac{n}{2} \hat{\phi}_w' \hat{\phi}_w$$

which is

$$I_n(\omega) = \frac{n}{2} \left[\frac{2}{n} y'X \right] \left[\frac{2}{n} X'y \right] = \frac{2}{n} y'XX'y$$

The periodogram in matrix notation has a familiar form from the study of linear models. The matrix XX' is square ($n \times n$) and symmetric, making the periodogram a positive semidefinite quadratic form $y'Ay$ with matrix $A = (2/n)XX'$.

The expectation is easily derived using properties of quadratic forms. The Cosinor model can be written as $y_t = s_t + \epsilon_t$, for $t = 1, 2, \dots, n$, with $s_t = \alpha_0 \cos(\omega_0 t) + \beta_0 \sin(\omega_0 t)$. In this form, the vector y is composed of a signal vector s , with t^{th} element s_t , and a vector of errors ϵ . For the general signal plus noise model, the expectation of the corresponding quadratic form is

$$E(y'Ay) = \sigma^2 \text{trace}(A) + s'A s$$

Applying this to the Standard periodogram we have

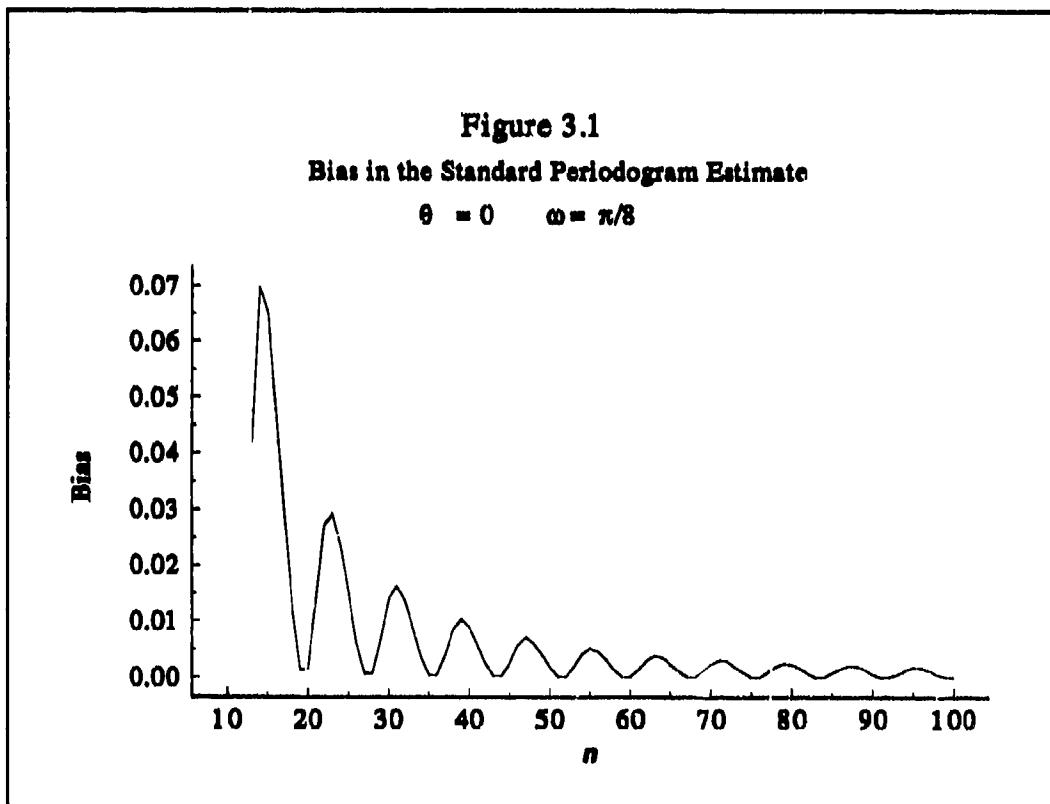
$$E\left(\frac{2}{n} y'XX'y\right) = 2\sigma^2 + \frac{2}{n} s'XX's.$$

The first term has no effect on the location of the maximum, so only the second term will need to be considered.

The expected periodogram is a function of the true parameter values α_0 , β_0 , ω_0 and the sample size n . Given values of these parameters, we can find the ω that results in the maximum and obtain an approximation to the bias. In general, for given values of ω_0 and n , it can be shown that the bias will be the same for all α_0 and β_0 such that $\alpha_0 = k\beta_0$. The bias depending on k is equivalent to the bias depending on the phase $\theta_0 = \arctan 2(-\beta_0/\alpha_0)$, since $\tan(\theta_0) = -k$ when $\alpha_0 = k\beta_0$. We consider θ_0 equal to $0, \pi/4, \pi/2$ and $3\pi/4$. The values selected for the true frequency ω_0 are the midpoints of each quarter of the interval

$[0, \pi]$. Sample sizes of n ranged from 10 to 500. In order to ensure that we find the global maximum of the expected periodogram, we use a grid search algorithm. The expected periodogram is first calculated at the Fourier frequencies, and the Fourier frequency that results in the largest periodogram ordinate is identified. This frequency is ω_m . Using ω_m as the center of the grid, a refined search with a grid mesh of 0.0001 is made from ω_{m-2} to ω_{m+2} , representing a range of four Fourier frequencies. Extensive simulations show that the expected periodogram approach gives excellent estimates of the true bias.

Figure 3.1 gives a graphical picture of the bias in the estimator of frequency for one set of parameters, $\theta_0 = 0$ and true frequency $\omega_0 = \pi/8$. It shows that the bias itself is a periodic function of sample size. This type of pattern was seen for all combinations of parameter values we looked at, though the bias was sometimes negative, or alternated from positive to negative with increasing n .



4. IMPROVED ESTIMATORS FOR THE COSINOR MODEL

We note the similarity between Walker's estimators and least squares estimators in Section 1. For non-Fourier frequencies, the Walker estimates of α_0 and β_0 are approximately equal to usual least squares estimates when ω is known, and are equal when $n = cP_0$ or when ω is a Fourier frequency. This relationship suggests another definition of the periodogram for non-Fourier frequencies, where we replace the "approximate" Walker estimates by the actual ordinary least squares estimators.

The ordinary least squares estimators of α_0 and β_0 are given by

$$\hat{\phi}_{LS} = \begin{pmatrix} \hat{\alpha}_{LS} \\ \hat{\beta}_{LS} \end{pmatrix} = (X'X)^{-1} X' y$$

Define the "least squares" periodogram as

$$I_n^{LS}(\omega) = \frac{n}{2} (\hat{\alpha}_{LS}^2 + \hat{\beta}_{LS}^2) = \frac{n}{2} \hat{\phi}_{LS}' \hat{\phi}_{LS}$$

or equivalently as the quadratic form

$$I_n^{LS}(\omega) = y' A y, \quad A = \frac{n}{2} (X(X'X)^{-1}(X'X)^{-1}X').$$

Since the actual least squares and Walker estimators are equal at the Fourier frequencies, the two periodograms are also equal at these frequencies. As we have replaced "approximate" estimators with "exact" estimators in the new periodogram, we might expect our estimate of frequency to improve also.

Again, the bias of the least squares periodogram estimator of frequency is approximated by maximizing the expected periodogram. The least squares periodogram is also a quadratic form with $A = n/2 (X(X'X)^{-1}(X'X)^{-1}X')$. Unlike the Standard periodogram, we initially cannot ignore the contribution of the error term, since

$$\text{trace}(A) = \frac{2n^2}{n^2 - D_n^2(2\omega)}.$$

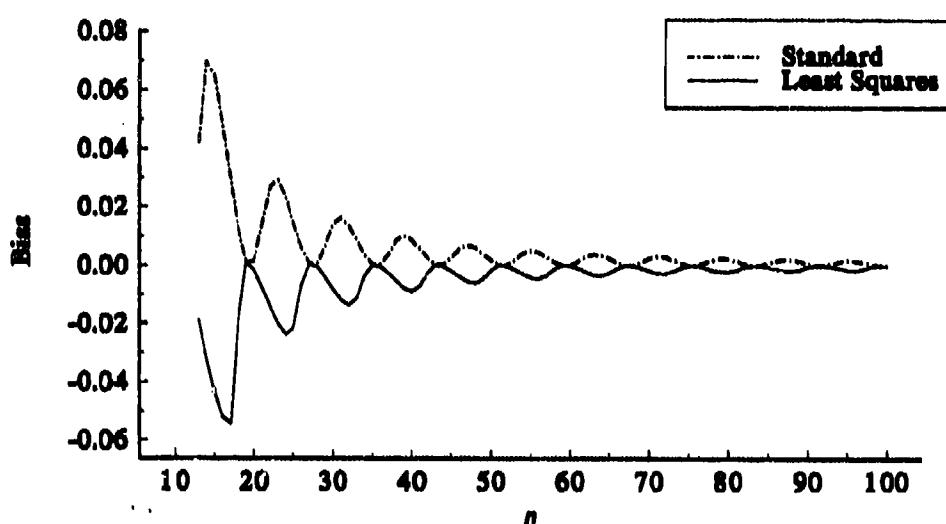
where

$$D_n(\mu) = \frac{\sin(\frac{n\mu}{2})}{\sin(\frac{\mu}{2})}$$

This term has negligible effect unless the signal to noise ratio is very small, n is small or ω_0 is close to zero or 2π . The calculations of bias are performed for the same values of θ_0 , ω_0 and n as for the Standard periodogram.

The results show that estimation of the frequency based on the least squares periodogram is also biased and does not really offer an improvement over conventional Walker estimates. This is illustrated in Figure 4.1, which reveals an interesting and possibly useful relationship. In the cases examined, the bias of the Least Squares periodogram estimate is always of the opposite sign as the bias in the Walker periodogram estimate, and of approximate equal magnitude. This suggests using a periodogram that is in some sense a combination of the two periodograms.

Figure 4.1
Bias in the Least Squares Periodogram Estimate
 $\theta = 0 \quad \omega = \pi/8$



The simplest combination is the arithmetic average of the two periodograms, a quadratic form with

$$A = X \left[\frac{1}{n} I + \frac{n}{4} (X'X)^{-1} (X'X)^{-1} \right] X'.$$

Based on the biases we have seen with the Standard and Least Squares periodograms, the estimate of frequency based on the average periodogram is expected to give an unbiased estimator of ω_0 , or at least one with reduced bias. Calculations and simulations show that the latter is true. The new average periodogram estimator has less bias than both the Walker or Least Squares estimators, and the bias approaches zero faster as n becomes large. See Figure 4.2.

Another approach is to use a geometric average. Let us define the "composite" periodogram by

$$I_n^C = \frac{n}{2} \hat{\phi}_W' \hat{\phi}_{LS}.$$

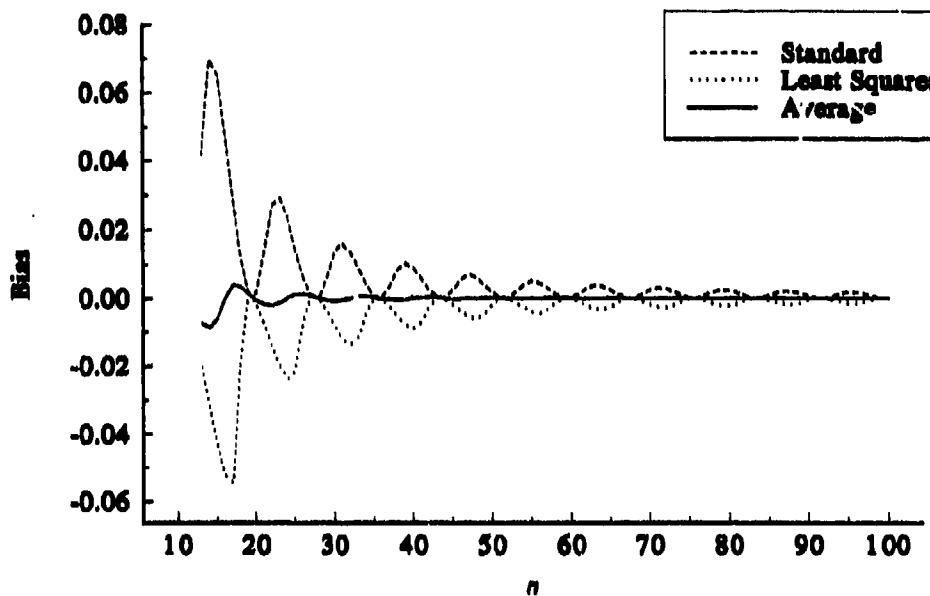
Substituting in the matrix representations of the estimates gives

$$I_n^C(\omega) = y' A y, \quad A = X(X'X)^{-1} X'.$$

The matrix A is the familiar hat matrix from linear regression, and $I_n^C(\omega)$ has a form similar to the

Figure 4.2
Bias in the Average Periodogram Estimate

$$\theta = 0 \quad \omega = \pi/3$$



regression sum of squares in linear models. Again, when ω is a Fourier frequency, $(X'X)^{-1} = 2/n I$, and the Composite periodogram will be equal to the Standard periodogram.

We will show that the value that maximizes the expectation of the composite periodogram is always ω_0 . Consider the expression

$$\begin{aligned} R_n(\omega) &= s's - s'X(X'X)^{-1}X's \\ &= s'(I - H)s \end{aligned}$$

where s is the signal vector and $H = X(X'X)^{-1}X'$. This is the difference between the total sum of squares of the signal and the value of the periodogram at ω . The following are true

$$s's > 0 \quad s \neq 0$$

$$s'Hs \geq 0$$

$$s'(I - H)s \geq 0$$

The first statement is immediate since the expression $s's$ is a sum of squares, while the second and third follow because H and $I - H$ are both idempotent. From these we can also conclude $s's \geq s'Hs$. This means that the absolute maximum value $s'Hs$ can possibly attain is $s's = s'Hs$. For the Cosinor model the signal is $s = X_0\phi_0$, so when $X = X_0$ we have

$$s' H s = \phi_0' X_0' X_0 (X_0' X_0)^{-1} X_0' X_0 \phi_0 \\ = s' s .$$

Thus, an absolute maximum occurs when $X = X_0$, i.e. when $\omega = \omega_0$. It can also be shown that the absolute maximum is unique.

The key portion of this proof is that $s'(I - H)s$ is idempotent, and thus positive semidefinite. This does not occur for the other periodograms. In fact, it can be shown by numeric example that $s'A_s$ can exceed $s's$ in all of the other periodograms.

The performance of the Composite Periodogram estimator of frequency was examined by simulation study. From the Cosinor Model with $\alpha_0 = 8.0$, $\beta_0 = 0.0$ and $\omega_0 = 0.5$ a total of 1000 data series, each of length $n = 100$, were generated with random Gaussian noise of mean 0 and variance 1. The results of the simulation are reported in Table 4.1. The Standard periodogram produced biased estimates of frequency, as well as bias in the estimates of the other two parameters. The Composite periodogram gave an unbiased frequency estimate and the estimates of α_0 and β_0 were greatly improved, but still biased. The biases were expected, since in section 2, we saw that even with ω estimated without bias or error, we still obtain biased estimates for α_0 and β_0 . Using this fact, we construct the bias-corrected estimates

$$\alpha^* = \alpha - \frac{\alpha}{n} C_n(2\omega) D_n(2\omega) - \frac{\beta}{n} S_n(2\omega) D_n(2\omega)$$

and

$$\beta^* = \beta - \frac{\alpha}{n} S_n(2\omega) D_n(2\omega) + \frac{\beta}{n} C_n(2\omega) D_n(2\omega)$$

We will call these the adjusted Composite Periodogram estimators. Simulations, reported in Table 4.2, indicate that the adjusted estimators are now approximately unbiased.

5. SUMMARY AND CONCLUSIONS

We have considered four definitions of the periodogram, given by

$$I_n(\omega) = \frac{n}{2} \hat{\phi}_W' \hat{\phi}_W$$

$$I_n^{LS}(\omega) = \frac{n}{2} \hat{\phi}_{LS}' \hat{\phi}_{LS}$$

$$I_n^{AVG}(\omega) = \frac{1}{2} (I_n^W(\omega) + I_n^{LS}(\omega))$$

$$I_n^C(\omega) = \frac{n}{2} \hat{\phi}_W' \hat{\phi}_{LS}$$

These are called the Standard, Least Squares, Average and Composite periodograms. The periodograms are equal at the Fourier frequencies, differing only in how they approximate the periodogram between these frequencies. They may be expressed as quadratic forms in y , allowing us to easily compute their expectations and to compute the approximate bias in the frequency estimates based on these periodograms.

Based on maximizing the expected periodogram, the Standard and Least Squares periodograms

Table 4.1

**Simulation Study and Comparison
of the Standard Walker Periodogram with
the Composite Periodogram Estimators**

1000 Replicates

Parameter	True Value	Standard Periodogram			Composite Periodogram		
		Mean	Std. Dev.	MSE	Mean	Std.Dev.	MSE
α_0	8.000	7.9461	0.1415	0.1514	7.9554	0.1424	0.1492
β_0	0.000	0.4786	0.2923	0.5523	-0.0125	0.2855	0.2858
α_0	0.500	0.5012	0.0006	0.0013	0.5000	0.0006	0.0006
σ^2	1.000	1.0521	0.1456	0.1546	1.0133	0.1455	0.1461

Table 4.2
Simulation Study and Comparison
of the Composite Periodogram with the
Adjusted Composite Periodogram Estimators
1000 Replicates

Parameter	True Value	Standard Periodogram			Adjusted Composite Periodogram		
		Mean	Std. Dev.	MSE	Mean	Std.Dev.	MSE
α_0	8.000	7.9554	0.1424	0.1492	7.9977	0.1426	0.1426
β_0	0.000	-0.0125	0.2855	0.2858	-0.0027	0.2843	0.2843
ω_0	0.500	0.5000	0.0006	0.0006	0.5000	0.0006	0.0006
σ^2	1.000	1.0133	0.1456	0.1461	1.0123	0.1454	0.1460

produce biased frequency estimators for moderate n . The Average periodogram estimators are also biased, but to a lesser degree than either component. The estimators based on the Composite periodogram are, on the other hand, unbiased for all combinations of true parameter values and all n . The Composite periodogram also has a familiar interpretation in terms of the least squares problem of fitting Cosine curves, making it easy to implement.

We also propose bias-adjusted estimators of α_0 and β_0 , using the Composite periodogram estimator of frequency. Simulations show that these estimators are approximately unbiased, and that the Standard and Composite Periodogram estimators have similar variances. Based on these results, we would strongly recommend using our new estimators for fitting the Cosinor model to individual data series.

6. REFERENCES

- Bingham, C., Arbogast, B., Cornelissen, G., Lee, J.K. and Halberg F. (1982). Inferential statistical methods for estimating and comparing Cosinor parameters. *Chronobiologia* 9, 397-439.
- Diggle, P.J. (1990). *Time Series : A Biostatistical Introduction*. Oxford Science Publications, Clarendon Press, Oxford.
- Halberg, F., Tong, Y.L. and Johnson, E.A. (1965). Circadian system phase - an aspect of temporal morphology. In *The Cellular Aspects of Biorhythms*, 20-48. Springer-Verlag, Berlin.
- Jaynes, E.T. (1987). Bayesian spectrum and chirp analysis. In *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*, 1-37. Smith, C.R. and Erickson, G.J. (editors). D. Reidel, Dordrecht-Holland.
- Monk, T.H. and Port, A. (1983). "COSINA" : A Cosine curve fitting program suitable for small computers. *International Journal of Chronobiology* 8, 193-224.
- Nelson, W., Tong, Y.L., Lee, J.K. and Halberg, F. (1979). Methods for Cosinor rhythmometry. *Chronobiologia* 6, 305-323.
- Rice, J.A. and Rosenblatt, M. (1988). On frequency estimation. *Biometrika* 75, 477-484.
- Vokac, M. (1984). A comprehensive system of Cosinor treatment programs written for the Apple II microcomputer. *Chronobiology International* 1, 87-92.
- Walker, A.M. (1971). On the estimation of a harmonic component in a time series with stationary independent residuals. *Biometrika* 58, 21-36.
- Whittle, P. (1952). The simultaneous estimation of a times series harmonic component and covariance structure. *Trabajos de Estadistica* 3, 43-57.
- Zhao-Guo, C. (1988). An alternative consistent procedure for detecting hidden frequencies. *Journal of Time Series Analysis* 9, 301-317.

Meta-Analysis of Gas Flow Resistance Measurements Through Packed Beds

Malcolm S. Taylor and Csaba K. Zoltani¹

Measurements of the resistance to flow through packed beds of inert spheres have been reported by a number of authors through relations expressing the coefficient of drag as a function of Reynolds number. A meta-analysis of the data using improved statistical methods is undertaken to aggregate the available experimental results. For Reynolds number in excess of 10³ the relation log F_v = 0.49 + 0.90 log Re' is shown to be a highly effective representation of all available data.

Nomenclature

D_b = spherical particle (bead) diameter

D_c = test chamber diameter

f_{s'} = $\frac{F_v}{Re/(1-\phi)}$, friction factor

F_v = $\frac{\Delta P}{L} \frac{D_b^2}{\mu \bar{U}} \left(\frac{\phi}{1-\phi} \right)^2$, coefficient of drag

F_{v_i} = i-th observed value of the drag coefficient

F_{v_i'} = predicted drag coefficient corresponding to
the i-th observed value

L = length scale

Re = Re_pφ = ρUD_bφ/μ, Reynolds number

Re' = Re/(1 - φ)

Re_p = Reynolds number based on particle size

U = average gas velocity

β_i, i = 0, 1, 2 = model coefficient

ΔP = change in pressure

ρ = density

φ = porosity of the packed bed

(1 - φ) = solids loading

μ = gas viscosity

1. U.S. Army Research Laboratory, Aberdeen Proving Ground, Maryland 21005-5066. This material also appears in ARL-TR-301, November 1993.

1. Introduction

Experimental results are cumulative if in aggregate they unify and extend empirical relations and theoretical structures which may be obscured in individual investigations. Empirical cumulativeness, which Hedges (1987) describes as "... the degree of agreement among replicated experiments or the degree to which related experimental results fit into a simple pattern that makes conceptual sense," is the focus of this paper. Glass (1976) was among the first to recommend the use of quantitative procedures in integrative research reviews and to introduce the term "meta-analysis" to cover the collection of such procedures. Meta-analysis claims certain classical statistical procedures, as well as approaches developed specifically for research synthesis, and has found application in the social and biological sciences. The unification of experimental results obtained by different investigators, operating independently with their own experimental protocol and sometimes using different methods of analysis, is the kernel of meta-analysis. A comprehensive treatment of this subject is given by Hedges and Olkin (1985).

Measurement in the physical sciences is generally regarded as highly accurate, and although some variability is inevitable, the variation itself is thought to be insignificant from a practical standpoint. Counterexamples to this notion are plentiful, even in carefully conducted experiments. Consider, for instance, the situation described by Touloukian (1975) involving two sets of measurements taken on the thermal conductivity of gadolinium. These data, shown in Figure 1, "... are for the same sample, measured in the same laboratory two years apart in 1967 and 1969. The accuracy of curve 1 was stated as within 1% and that of curve 2 as 0.5% ..." and yet, the curves differ by more than several hundred percent at higher values of temperature.

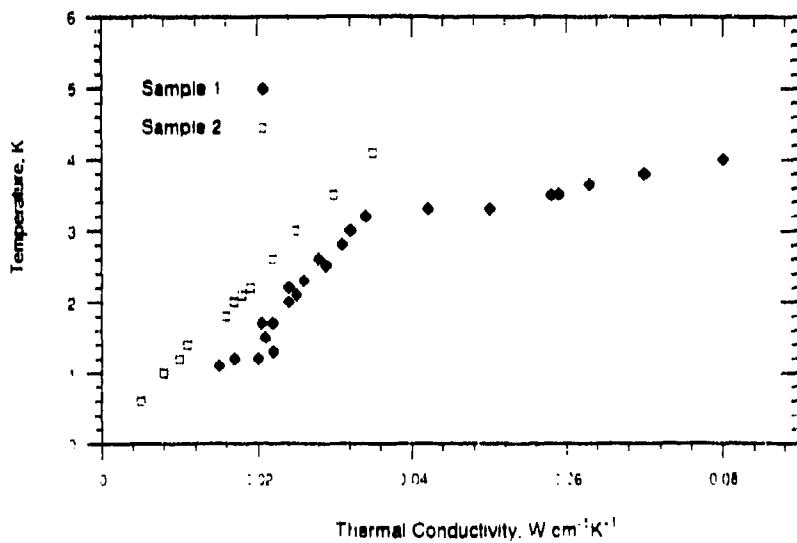


Figure 1. Thermal conductivity of gadolinium.

Physical scientists normally bring a careful qualitative analysis to their research studies. If prudently employed, interrogative statistics, which are part of meta-analysis, have a contribution to make in the physical sciences as well.

After data has been collected according to a carefully constructed experimental design (e.g., see Montgomery 1991) the main reason for determining a correlation (a regression analysis) is to examine the effects that some variables exert, or appear to exert, on others. Even when no intuitive physical relationship is apparent, regression analysis may provide a convenient summary of the data. The summary can be accomplished in a number of ways and has been an active area of investigation since the time of A.M. Legendre (1752-1833), who published the first account of regression by least squares in 1805. Section 2 of this paper reviews the correlations that have been advanced for steady flow through inert spherically packed beds and some of the consequences of the attendant data analysis. In Section 3, a meta-analysis of the gas flow resistance measurements is undertaken. Section 4 contains a summary and main conclusions.

2. Regression Analysis of Gas Flow Resistance Measurements

Ergun (1952), Kuo and Nydegger (1978), and Jones and Krier (1983) have proposed models relating coefficient of drag to Reynolds number for steady flow through packed beds of inert spheres. However, the correlations were developed under different experimental regimens. Robbins and Gough (1978) also investigated coefficient of drag at high Reynolds number but presented their results in terms of a friction factor $f_v' = \frac{F_v}{Re/(1-\phi)}$, which is the ratio of coefficient of drag F_v , and Reynolds number Re scaled by a solids loading factor $(1-\phi)$.

In comparing Ergun's relation

$$F_v = 150 + 1.75 \left(\frac{Re}{1 - \phi} \right), \quad (1)$$

to that of Kuo and Nydegger

$$F_v = 276.23 + 5.05 \left(\frac{Re}{1 - \phi} \right)^{0.87}, \quad (2)$$

or of Jones and Krier

$$F_v = 150 + 3.89 \left(\frac{Re}{1 - \phi} \right)^{0.87}, \quad (3)$$

a slight notational difference portends substantial complications. Equation (1) is a simple linear model. Equations (2)-(3) are nonlinear in the sense that one or more parameters appear nonlinearly. Nonlinearity complicates the statistical analysis of the data since determining appropriate choices for the parameters in equations (2)-(3) becomes a computationally intensive optimization procedure, and inference about the resultant relation and

parameters becomes much more tentative. The mathematical underpinnings of nonlinear regression will not support as much in the way of statistical inference or hypothesis testing as is available for linear regression. In general, nonlinear models should be avoided unless there is a compelling reason for their use. Draper and Smith (1981) discuss this issue in greater detail.

Standard regression procedures are developed under several assumptions. Fundamental among these is that the response (here, F_v) is measured with error but the predictor(s) (here, Re and ϕ) are measured without error. Jones and Krier provide estimates of error for F_v , Re , and ϕ , confirming that this assumption is not met, and call into question the efficacy of the resultant correlations. Sometimes an attempt to circumvent this requirement is undertaken by arguing that the error in predictor measurement is sufficiently small as to be ignored when compared to the range of the predictor variable. If this claim is invoked, reliance upon any resultant representation must be tempered accordingly.

Since a correlation provides a convenient representation of the available data, a direct attempt at evaluating the adequacy of a regression equation involves an examination of the differences between the measurements taken and the values predicted by the equation. These differences, $F_{v_i} - F_{v'_i}$, $i = 1, 2, \dots, n$, are called residuals; F_{v_i} is an experimentally determined value of drag coefficient, and $F_{v'_i}$ is the corresponding value predicted by the regression equation. A residual plot for equation (3) is shown in Figure 2. These plots may serve as a diagnostic tool in addition to assessing the adequacy of a fitted regression model.

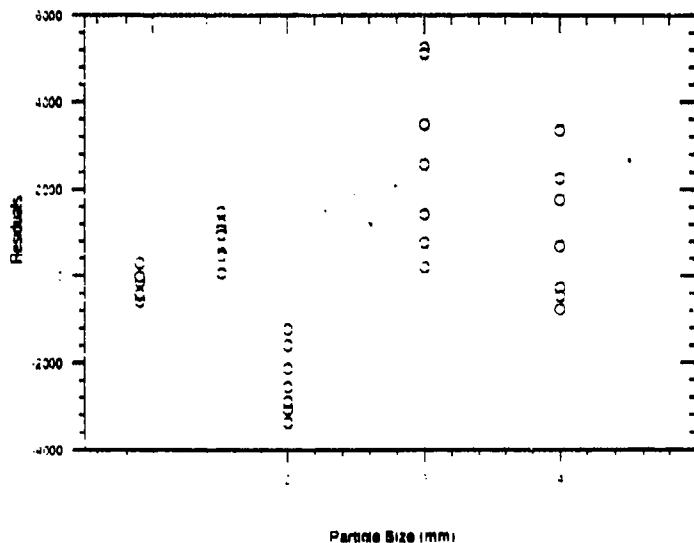


Figure 2. Residuals vs. particle diameter D_b ; Jones and Krier data with 6 mm beads excluded.

Figure 2 strongly suggests that another crucial regression assumption is not satisfied. The variance of the residuals does not appear constant over the range of $Re' = Re/(1 - \phi)$; and moreover, the departure from the fitted equation is systematic with bead diameter, D_b . Jones and Krier recommend reverting to the relation (2) proposed by Kuo

and Nydegger to describe their own measurements taken for 6 mm beads. This recommendation is data specific and is difficult to justify in general. They conjecture that an interaction between bead size and tube diameter may be present, but this requires quantitative substantiation. In general, weighted least squares, or a transformation on the observations F_v , before regression, are potential corrective procedures suggested by this residual pattern.

2.1 Regression Analysis Revisited

Nonlinear regression algorithms normally seek to minimize the sum of the squared residuals—as in ordinary linear regression—in attempting to determine the "best" choice of parameters to model the data. These procedures have previously been cited as computationally intensive. More specifically, they are iterative and may diverge or converge to local extrema, depending upon the choice of initial conditions. Through a systematic selection of initial conditions, the authors determined that the equation

$$F_v = 61 + 2.7 \left(\frac{Re}{1 - \phi} \right)^{0.91}, \quad (4)$$

provides an improved representation of the data reported by Jones and Krier.

The root mean square error (RMSE), an estimate of the standard deviation of the residuals and a commonly used measure for adequacy of fit, is reduced by 20% compared to that corresponding to equation (3). The measurements taken on the 6 mm beads, the chief contributor to heterogeneity of variance, have been excluded from the regression, making the comparison with Jones and Krier direct. A reduction of one-fifth in RMSE is not by itself a stunning improvement, but it does focus more sharply on the underlying physical process. The residual plot for equation (4) still exhibits the undesirable pattern of under(over) fitting categories of bead diameter, but is an improvement compared to the display in Figure 2.

The data collected by Robbins and Gough (1978, 1979), which "... correspond to several tests performed on several occasions" for beds of spheres, right circular cylinders, and multiperforated cylinders, may be transformed into units appropriate for comparison through the relationship $f_v' = \frac{F_v}{Re/(1 - \phi)}$. The authors confined the analysis to data taken on 1.27 mm diameter lead shot and on 4.76 mm and 7.94 mm diameter steel spheres, and determined the equation

$$F_v = -237 + 3.14 \left(\frac{Re}{1 - \phi} \right)^{0.89}, \quad (5)$$

for representation of flow through spherically packed beds. Equations (4)-(5) are shown, along with the previously established correlations (1)-(2), in Figure 3.

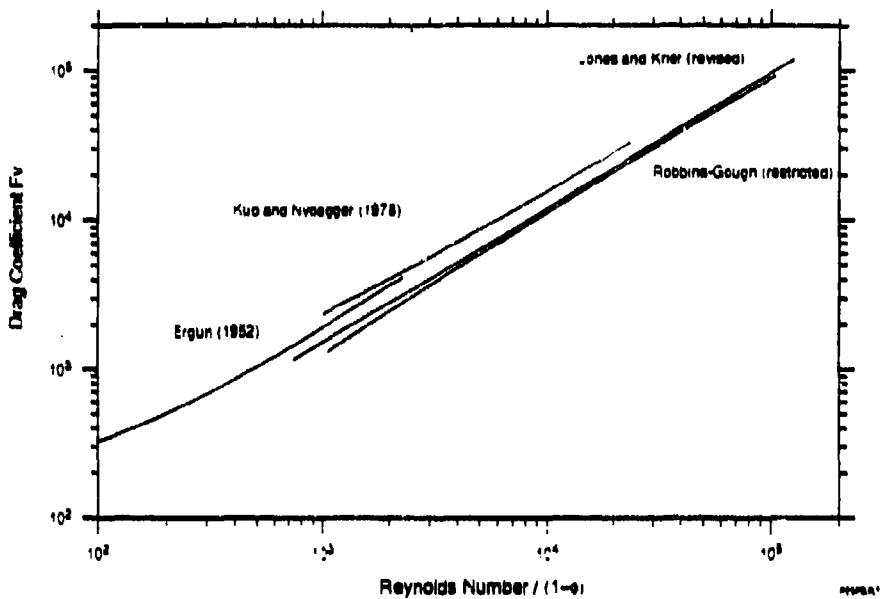


Figure 3. Proposed models for relating coefficient of drag and Reynolds number.

Transforming the variables (Re' , F_v) by taking logarithms, which was suggested by the residual plot in Figure 2, effectively linearizes the data. In regression analysis, a measure of precision of the regression line which is used in addition to RMSE, is given by a statistic denoted as R^2 . R^2 assumes values in the unit interval [0, 1] and quantifies the amount of variation in the response accounted for by the regression line. Values close to one are highly desirable, indicating that the regression has effectively accounted for most of the variation in the response. The regression line determined after logarithmic transformation of the Jones and Krier data has $R^2 = 0.98$. The transformed Robbins and Gough data have $R^2 = 0.99$. These values are so close to 1.0 that pursuit of a nonlinear model is difficult to justify mathematically.

Comparison between linear models and nonlinear models is difficult. RMSE values cannot be compared across the transformation, and a well-defined R^2 statistic for nonlinear models does not exist.

3. Meta-Analysis of Gas Flow Resistance Measurements

Consider in aggregate the correlations that have been advanced for gas flow resistance measurements through spherically packed beds. For the nonlinear models, a statistical resampling plan is applied, whose goal is to extract information from a set of data through repeated inspection. The procedure is called the "bootstrap," named to convey its self-help attributes, and it attempts to address an important problem in data reduction—having computed an estimate of some parameter, what accuracy can be attached to the estimate? Accuracy here refers to the " \pm something" that often accompanies statistical esti-

mates, and may be conveyed through such devices as variance, RMSE, or confidence interval. For the log-linear model, the available data are directly combined.

The authors are hindered in fully exploiting a meta-analysis approach by the inability to obtain all of the pertinent experimental data. It is unfortunate that experimental data are not routinely archived after collection; otherwise, additional information that it may hold is lost to extraction by subsequent investigations and by alternative statistical methods. The data of Jones and Krier and of Robbins and Gough were accessible. With these data, this paper proceeds as far as statistical prudence permits.

3.1 Bootstrapping Regression Correlations

Detailed descriptions of the bootstrap and accounts of its successful applications are amply documented (e.g., Efron (1979, 1982)), Efron and Tibshirani (1985), LePage and Billard (1992). The computational contrivance that the bootstrap procedure exploits is the generation of perturbed data sets from a single set of data through sampling with replacement. Specific to this study, the set of paired observations taken on coefficient of drag and Reynolds number, $\{(F_{v_i}, Re_1'), \dots, (F_{v_n}, Re_n')\}$, that is the basis for a reported correlation, is sampled with replacement to generate another set $\{(F_{v_1}^*, Re_1^{**}), \dots, (F_{v_n}^*, Re_n^{**})\}$ whose elements are copies (with duplication) of the original measurements. This set is called a bootstrapped data set. The process of sampling with replacement to generate bootstrapped data sets is repeated many times.

If a correlation is determined for each bootstrapped data set and its equation plotted, an indication of the sensitivity of the regression line to perturbation of the original data comes into focus. In Figure 4, the results of 1000 replications of this process are pictured. The outermost lines indicate boundaries within which the correlation (5) might be expected to lie if the original data set were simply perturbed. They were obtained from the maxima and minima of the drag coefficient predicted for particular values of Re' .² The envelope constructed for correlation (5) contains correlation (4). This suggests that no significant difference between these empirical relations exists. Similar results are obtained if we begin with correlation (4); correlation (5) will lie within the corresponding confidence envelope. Consideration of perturbed data is highly appropriate here, since experimental results cannot be expected to be reproduced, even if the experiment is replicated under tightly controlled conditions. The theoretical justification for the use of bootstrapped data is given by Efron (1982).

The relationship of Kuo and Nydegger, for which the experimental data was not accessible, was determined for a single diameter bead, $D_b = 0.83$.

2. More precisely, the values represent extreme quantiles after all of the F_v 's have been ranked; their values are not essentially different from maxima and minima.

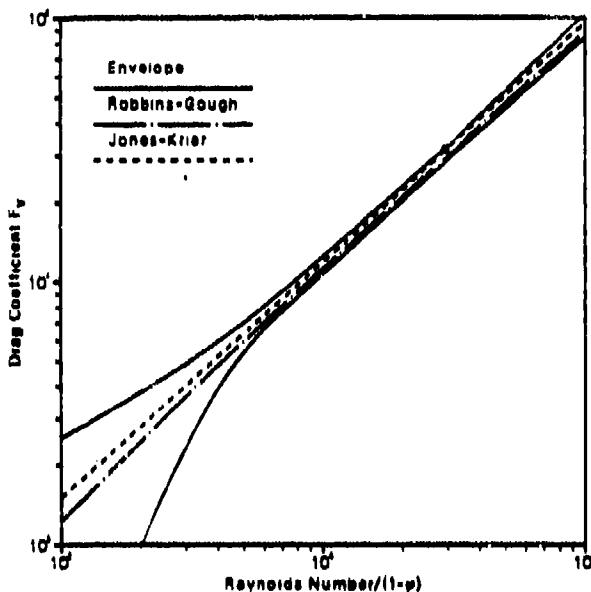


Figure 4. Bootstrapped confidence envelopes for nonlinear regression
(based on Robbins-Gough data).

3.2 Log-linear Regression

Figure 5 displays the logarithmic transformed data of Jones and Krier, and Robbins and Gough, combined. The fitted line for these data is

$$\log F_v = 0.49 + 0.90 \log Re'; \quad (6)$$

included in the regression are the data taken on 6 mm beads which were previously excluded.

Visually, the data appear linear after transformation. Statistically, the R^2 -value for the regression is 0.99, making the fitted line a highly satisfactory representation of these data for all practical purposes.

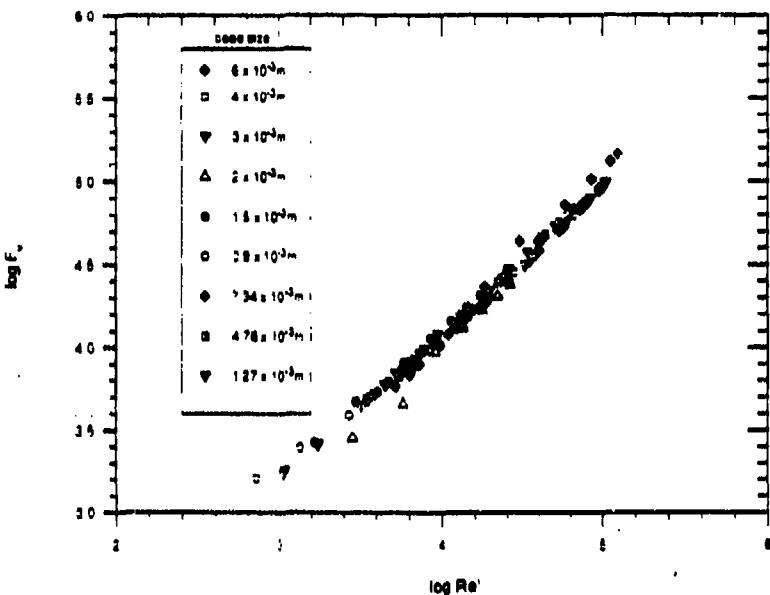


Figure 5. Drag coefficient vs. Reynolds number/(1- ϕ);
Jones and Krier, Robbins and Gough data, combined.

4. Summary and Conclusions

For Reynolds number exceeding 10^3 , a more effective representation and data analysis than presently available can be obtained after logarithmic transformation of the data. This linearizes the data and removes the necessity for nonlinear regression techniques. The equation

$$\log F_v = 0.49 + 0.90 \log Re' \quad (7)$$

is an effective description of the available experimental data.

If a representation of the form

$$F_v = \beta_0 + \beta_1 (Re/1 - \phi)^{\beta_2} \quad (8)$$

is required, then Jones and Krier's results are more effectively reflected through the equation

$$F_v = 61 + 2.7 \left(\frac{Re}{1 - \phi} \right)^{0.91}, \quad (9)$$

and Robbins and Gough's data restricted to spherically packed beds provide the relation

$$F_v = -237 + 3.14 \left(\frac{Re}{1 - \phi} \right)^{0.89}, \quad (10)$$

but here again, approximate confidence envelopes constructed with the aid of the bootstrap suggest that these relations can be combined without loss of underlying physical insight. In total, the statistical analysis supports the combination of the various correlations, for the stated test conditions, into a single relationship.

While it is quite reasonable to suspect an interaction between the geometry of tube and packing, perhaps reflected through the ratio D_c/D_b , more extensive testing is required to establish this relation. Hopefully this will be done in accordance with a formal statistical experimental design to minimize testing and maximize extraction of information.

G.E.P. Box, an important contemporary statistician, has remarked that "No model is correct, but some are useful." In this spirit these remarks are offered along with the hope for an incremental move toward a more useful model.

Acknowledgements

The authors express their appreciation to Prof. Herman Krier of the University of Illinois and to Mr. F.W. Robbins of the U.S. Army Research Laboratory for their cooperation and for making available the raw data which was the basis of this meta-analysis. We happily acknowledge Mr. Monte W. Coleman of the U.S. Army Research Laboratory for performing the bootstrap computations, and for producing their summary in Figure 4.

References

- Draper, N.R., and Smith, H., *Applied Regression Analysis*, 2nd Ed., John Wiley & Sons, New York, 1981, pp. 458-459.
- Efron, B., "Computers and the Theory of Statistics: Thinking the Unthinkable," *SIAM Review*, 21, 1979, pp. 460-480.
- Efron, B., *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM Monograph 38, Philadelphia, 1982.
- Efron, B., and Tibshirani, R., "The Bootstrap Method for Assessing Statistical Accuracy," Department of Statistics Tech. Rept. No. 4, University of Toronto, 1985.
- Ergun, S., "Fluid Flow Through Packed Columns," *Chemical Engineering Progress*, Vol. 48, No. 2, Feb. 1952, pp. 89-94.
- Glass, G.V., "Primary, Secondary, and Meta-Analysis of Research," *Educational Researcher*, 5, 1976, pp. 3-8.
- Hedges, L.V., "How Hard is Hard Science, How Soft is Soft Science? The Empirical Cumulativeness of Research," *American Psychologist*, Vol. 42, No. 2, 1987, pp. 443-455.
- Hedges, L.V., and Olkin, I., *Statistical Methods for Meta-Analysis*, Academic Press, New York, 1985.
- Jones, D.P., "Friction Factors for Packed Beds at Reynolds Numbers Greater Than 10^3 ," Master's Thesis, University of Illinois, 1980.
- Jones, D.P., and Krier, H., "Gas Flow Resistance Measurements Through Packed Beds at High Reynolds Numbers," *Trans. ASME Journal of Fluids Eng.*, Vol. 105, 1983, pp. 168-173.
- Kuo, K.K., and Nydegger, C.C., "Flow Resistance Measurements and Correlation in a Packed Bed of WC 870 Ball Propellants," *Journal of Ballistics*, Vol. 2, No. 1, 1978, pp. 1-25.
- LePage, R., and Billard, L., *Exploring the Limits of Bootstrap*, John Wiley & Sons, New York, 1992.
- Montgomery, D.C., *Design and Analysis of Experiments*, 3rd Ed., John Wiley & Sons, New York, 1991.
- Robbins, F.W., and Gough, P.S., "Experimental Determination of Flow Resistance in Packed Beds of Gun Propellant," Proceedings of the 15th JANNAF Combustion Meeting, CPIA Pub. 297, Sept. 1978.
- Robbins, F.W., and Gough, P.S., "Influence of Length and Diameter of Cylinders on Packed Bed Flow Resistance," Proceedings of the 16th JANNAF Combustion Meeting, CPIA Pub. 308, Dec. 1979.
- Touloukian, Y.S., "Reference Data on Thermophysics," In H.A. Skinner (Ed.), *International Review of Science—Physical Chemistry: Vol. 10, Thermochemistry and Thermodynamics*, Butterworth, London, 1975, pp. 119-146.

DESKTOP MODELS FOR WEAPONS ANALYSIS

**JOHN D'ERRICO
EUGENE F. DUTOIT**

DISMOUNTED WARFIGHTING BATTLE LAB

A. INTRODUCTION

The purpose of this report is to provide a collection of simple, desktop computer models for operations research analysts and others within the combat developments area.

It was not intended that any of these models should replace the more complex models available to combat developers. There seems to be no lack of complex models, or efforts to produce more of the same. This report, on the other hand, attempts to attack the other end of the modeling spectrum.

Many operations research analysts, authors of new concepts, action officers who are developing operational requirements documents, and others, are not well served by the large, complex models which demand much in the way of resources, time, knowledge, and money, in order to use them.

On the other hand, there has been a substantial void in the number of models available to combat developments action officers to help them in their day-to-day work. This is the area which this report attempts to resolve, at least to some extent.

Each model in this report has been thoroughly researched, developed, and tested. Ample references to source documents have been cited. The format used to describe each model was based on ease of understanding and use.

A 3.5" disk , containing all the models, sample data files, and programs described herein, can be obtained by sending a blank, DOS-formatted, 3.5" double density or high density disk to:

Commandant
U.S. Army Infantry School
ATTN: ATSH-WCS (Mr. D'Errico)
Fort Benning, GA 31905-5400

B. PROBABILITY OF HIT MODEL

1. Introduction.

a. Description. This model, developed by Mr. John D'Errico, calculates the probability of hit for direct fire, single shot weapons, against a point target. Inputs required are the weapon's biases and dispersions, the target dimensions, the firer's aimpoint on the target, and the range to the target. The resulting probability of hit is displayed on the screen.

b. Limitations. This program computes the probability of hit based on the measurements of a two-dimensional target. It uses a process similar to the one used in many wargame models.

c. Applications. Desktop analytic tool for studying weapon accuracies and calculating probabilities of hit based on biases and dispersions provided by AMSAA and JMEM.

d. Setup. This model runs on a DOS-based computer. Data can be entered into the model in a few minutes, and results are displayed in less than one minute.

2. Guide to Operation.

a. Equipment Required.

(1) IBM compatible PC computer.

(2) 3.5" disk drive.

b. Installation.

* (1) Turn on the computer and get to the DOS prompt.

(2) Insert the 3.5" disk containing the PHCALC model into your computer's disk drive.

* (3) From the DOS prompt, enter the command: A:PHCALC (or B:PHCALC if you're using the B: disk drive).

c. Definitions.

(1) Horizontal shift in aimpoint from target center, in centimeters: If the firer's intended aimpoint is to the left of the target's center, the user should input the number of centimeters to the left as a negative number (e.g., -23). If the firer's intended aimpoint is to the right of the target's center, enter the number of centimeters to the right as a positive number (e.g., 12).

(2) Vertical shift in aimpoint from target center, in centimeters: If the firer's intended aimpoint is below the target's center, enter the number of centimeters from the center as a negative number. If the firer's intended aimpoint is above the target's center, enter the number of centimeters from the center as a positive number.

d. Operation.

(1) This program determines the probability of hit on a rectangular, two-dimensional target. If you wish to obtain the probability of hit on a target composed of two rectangles, (a vehicle consisting of a hull and turret), you merely need to keep in mind the location of the single aimpoint on the target, and run this program twice—once for each rectangle—and manually add the two resulting probabilities.

(2) For example, assume you are firing a missile at a tank 300 meters away. Its frontal measurements are: 300 cm wide by 200 cm high for the hull, and 200 cm wide by 100 cm high for the turret. Your aimpoint is the junction between the turret and the hull. The missile's biases and dispersions are shown in Screen 1. Using this program you will determine the separate probabilities of hit for the turret and the hull, keeping in mind that your aimpoint for both is the turret ring. When you enter the data for determining the probability of hit against the turret, your vertical shift in aimpoint from the center of mass is -50cm because your actual aimpoint is 50 cm below the turret's center of mass. In determining the probability of hit for the hull, you must indicate an upward shift of +100 cm from the hull's center of mass. Adding both probabilities will give you the probability of hit against the target.

(3) Actual prompts and sample inputs for the turret are shown in Screen #1.

```
Horizontal fixed bias (mils)? 0
Vertical fixed bias (mils)? 0

Select one of the following:
 1 - Total horizontal & total vertical dispersions
 2 - Separate variable & random error dispersions.

(Enter 1 or 2 from the keyboard)? 1

Total horizontal dispersions (mils)? 3
Total vertical dispersions (mils)? 3

Target width (centimeters)? 200
Target height (centimeters)? 100
Distance to target (meters)? 300
Horizontal shift in aimpoint from target center (cm)? 0
Vertical shift in aimpoint from target center (cm)? -50
```

Screen #1

- (4) The following result will be displayed on your screen.

The probability of hit, P(H), is .2754702

Screen #2

- (5) To run the program again, enter A:PHCALC or B:PHCALC

Horizontal fixed bias (mils):? 0
Vertical fixed bias (mils):? 0

Select one of the following:
1 - Total horizontal & total vertical dispersions
2 - Separate variable & random error dispersions.

(Enter 1 or 2 from the keyboard)? 1

Total horizontal dispersions (mils):? 3
Total vertical dispersions (mils):? 3

Target width (centimeters):? 300
Target height (centimeters):? 200
Distance to target (meters):? 300
Horizontal shift in aimpoint from target center (cm):? 0
Vertical shift in aimpoint from target center (cm):? 100

Screen #3

- (6) The probability of hit for the hull will be displayed as follows:

The probability of hit, P(H), is .4444504

Screen #4

- (7) Adding the results for the turret (Screen 2) and the hull (Screen 4) gives the probability of hit on the tank.

.2754702
.4444504
.7199206

e. Explanation.

(1) This model uses the probability density function for a random variable having a normal distribution: $f(x) = [1/(\sigma\sqrt{2\pi})] \{ \exp[-(x-\mu)^2/(2\sigma^2)] \}$, where μ is the population mean and σ is its standard deviation. In calculating probability of hit, the fixed bias is taken as the mean, and the total variable biases and dispersions are taken as the variance.

(2) Since a rectangular target has two dimensions, width and height, the problem becomes one of determining the joint probabilities of hitting the target within its horizontal boundaries and, simultaneously, within its vertical boundaries.

(3) To keep measurements consistent, the biases and dispersions are converted from mils to centimeters on target at the given range to the target. This conversion is by the equation

$$200(\text{Range in meters})\tan[(\theta/2)(.00098175)]$$

where the constant .00098175 is used to convert mils to radians, and θ represents the mean or standard deviation in mils.

(4) Given that the mean and standard deviation in the horizontal direction are converted to centimeters, and the width of the target is in centimeters, the distances of the horizontal boundaries of the target are transformed to standard form using the equation $z = (x-\mu)/\sigma$. The probability density function of the standard normal variable then becomes

$$\phi(z) = (1/\sqrt{2\pi})\exp(-z^2/2)$$

(5) Integration of the probability density function with the z -scores as the limits of integration yields the probability of hit in the horizontal direction.

(6) The same process is used to determine the probability of hitting the target within the vertical boundaries of the target.

(7) Finally, the probability of hitting the target within the horizontal boundaries is multiplied by the probability of hitting the target within the vertical boundaries, and the result is the probability of hitting the target.

(8) The trapezoidal method of integration, with 100 intervals, is used in this model.

C. PROBABILITY OF HIT PLOTTING MODEL

1. Introduction.

a. Description. This model, developed by Mr. John D'Errico, plots the hits for direct fire, single shot weapons, against a point target. Inputs required are the weapon's biases and dispersions, the target dimensions, the range to the target, and the number of iterations (i.e., the number of single shots to be plotted). The results are displayed graphically on the screen.

b. Limitations. This program plots the strike of each bullet relative to a two-dimensional target.

c. Applications. Desktop analytic tool for studying weapon accuracies. In effect, this model is a graphic, stochastic version of the PHCALC probability of hit model.

d. Setup. This model runs on a DOS-based PC computer. Data can be entered into the model in a few minutes, and results are displayed in less than one minute.

2. Guide to Operation.

a. Equipment Required.

(1) IBM compatible PC computer.

(2) 3.5" disk drive.

b. Installation.

(1) Turn on the computer and get to the DOS prompt.

(2) Insert the 3.5" disk containing the PHCALC model into your computer's disk drive.

(3) The process for printing the display with a printer depends on the version of DOS being used, the type keyboard, and the type of printer, but you must prepare for it now.

(a) For DOS 5.0, type the command GRAPHICS GRAPHICS (the word "graphics," typed twice, separated by a space) from the DOS prompt, before running this program. With an enhanced keyboard, pressing the [Print Screen] key, or the [Shift] + [Print Screen] keys, should print the display on your printer.

(b) If your version of DOS is older than 5.0, you should type the command GRAPHICS at the DOS prompt before running this program. If you have an unenhanced keyboard, the keys [Shift] + [Prt Scn] should print the screen on the printer.

(4) From the DOS prompt, enter the command: A:PHPLOT (or B:PHPLOT if you're using the B: disk drive).

c. Operation.

(1) This model places a target on the screen, scaled to the height and width inputs, and then displays the impact of each round in the target area, according to the weapon's biases and dispersions and range to the target.

(2) Program prompts and sample inputs are shown below, in screen #1.

```
Enter the horizontal fixed bias (mils) of the weapon system? 0
Enter the vertical fixed bias (mils) of the weapon system? 0
Total horizontal variable biases & dispersions (mils)? 3
Total vertical variable biases & dispersions (mils)? 3
Enter the weapon-target range in meters? 300
Enter the height of the target in meters? 2
Enter the width of the target in meters? 1
Enter the number of single rds to be fired? 200
```

Screen #1

(3) Upon entering the last input, a result similar to the one shown on the following page will appear on the screen. The display remains on the screen until any key is pressed, in case the user wishes to print the display on his printer. Pressing any key (except the print screen keys) will clear the screen, and return the user to the DOS prompt.

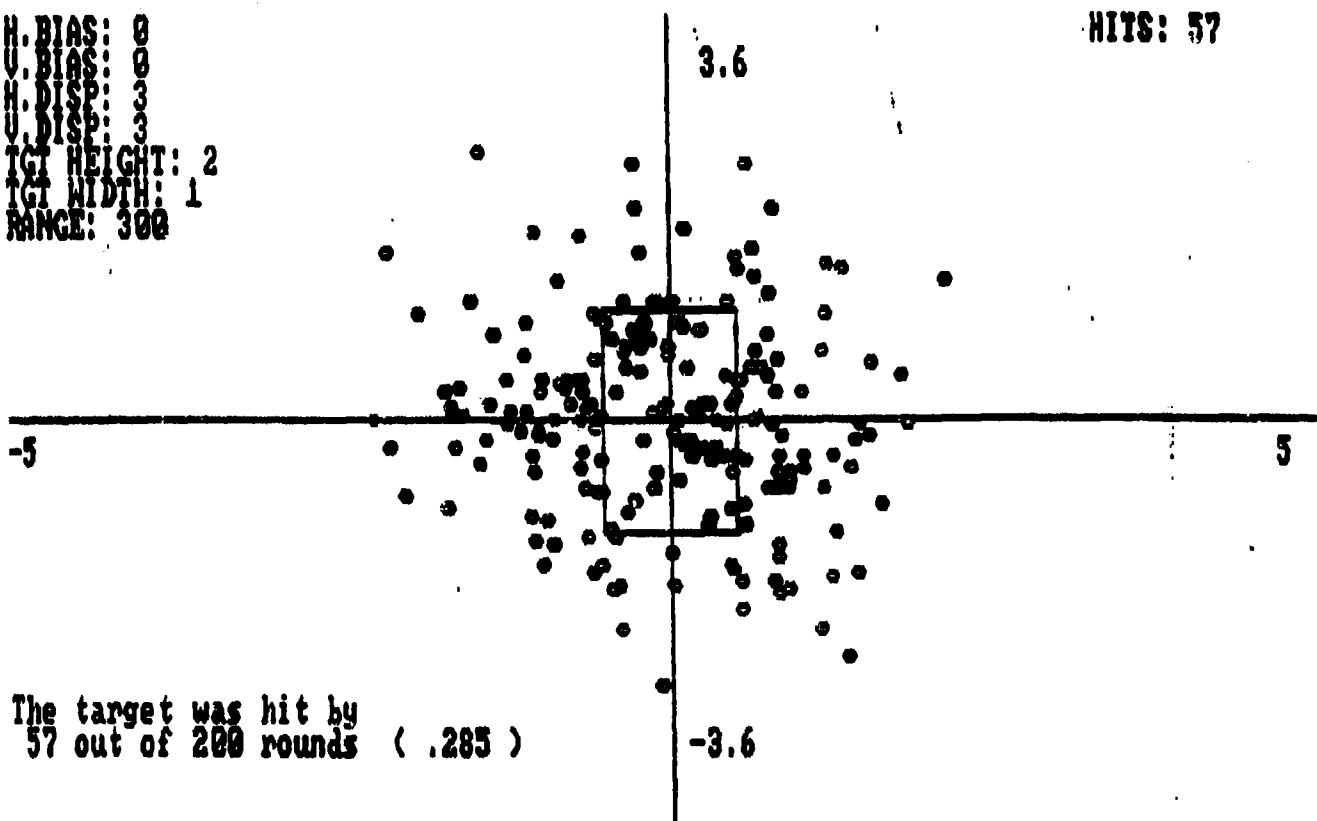
(4) To run the program again, enter A:PHPLOT or B:PHPLOT, whichever is appropriate.

d. Explanation. Given the biases and dispersions, this model samples from a normal probability distribution for the accuracy of each round, then determines the impact point based on the range to the target. The method used to generate normally distributed (pseudo) random numbers was proposed by Marsaglia and Bray in 1964. (Reference 8)

SAMPLE PRINTOUT OF PHPLOT PROGRAM

H.BIAS: 0
U.BIAS: 0
H.DISP: 3
U.DISP: 3
TGT HEIGHT: 2
TGT WIDTH: 1
RANGE: 300

HITS: 57



D. FORCE EFFECTIVENESS INDICES MODEL

1. Introduction.

a. Description. This model, from the TATAWS study (Reference 7), programmed and modified by Mr. John D'Errico, determines the value of each weapon system in a wargame, based on input from the killer-victim scoreboards. In general, the value of a weapon in a wargame is based on the values and quantities of opposing forces killed by that weapon.

b. Limitations. Force effectiveness indices are not commonly used, and as such, are not familiar to decision makers.

c. Applications. Desktop analytic tool for evaluating the effectiveness of a weapon within a wargame.

d. Setup. This model runs on an IBM compatible PC computer. Data can be entered into the model in a few minutes, and results are printed in less than one minute.

2. Guide to Operation.

a. Equipment Required.

(1) IBM PC compatible computer.

(2) 3.5" disk drive.

(3) Dot matrix Printer.

b. Installation.

(1) Turn on the computer and get to the DOS prompt.

(2) Insert the 3.5" disk containing the FEI2 model into your computer's disk drive.

(3) Turn on your printer, and make sure that it is "on line."

(4) From the DOS prompt, enter the command. A:FEI2 (or B:FEI2 if you're operating from a B: disk drive).

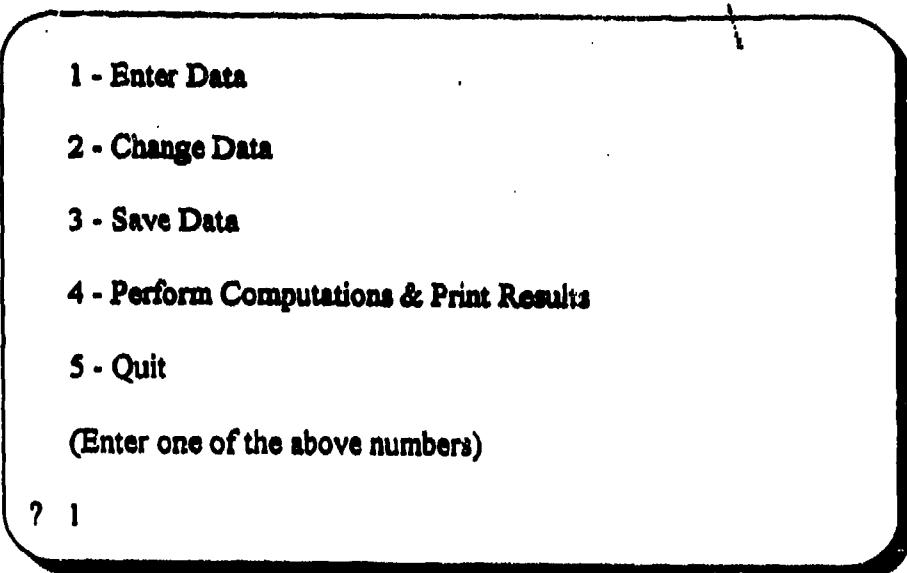
c. Operation.

(1) Essentially, you will be asked to enter the names of Red and Blue weapon systems, and data from a killer-victim scoreboard. Whether you are conducting a trial run or

not, you should save your data via main menu item #3--it will keep you from being frustrated in case you exit the program unintentionally (power interrupt, etc...) and have to enter the data all over again. Also, if you make a mistake while entering data, continue to enter the remainder of the data since you will be able to make any changes when you're done.

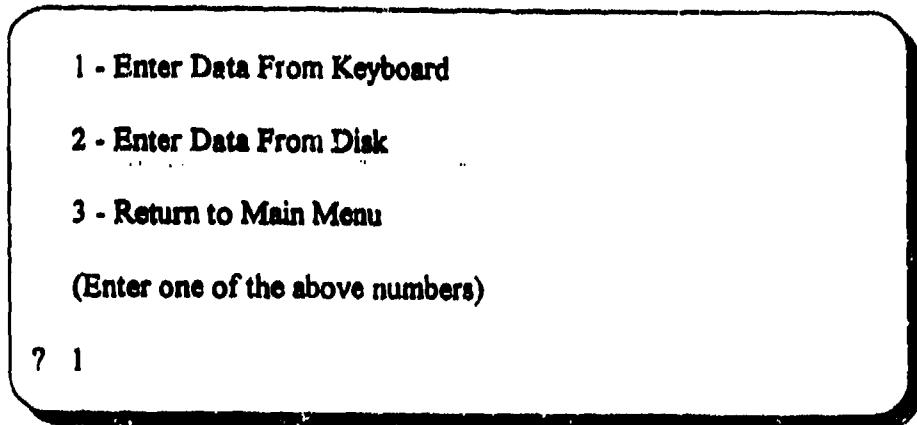
(2) The prompts you will see, and sample responses, are shown on the following facsimile screens.

(3) The first menu, also called the main menu, is as follows



Screen #1

(4) Entering the number 1, in Screen #1, leads to the next menu.



Screen #2

(5) Having selected the method for entering data, you will next be asked to enter the title for this case. The response in this example is "Test Case #1, 22 Dec 92, John D'Errico."

ENTER TITLE OF GAME. Test Case #1, 22 Dec 92, John D'Errico

Screen #3

(6) You will now be asked to enter the number of Blue weapon types, followed by the name of each Blue weapon type. For example, the killer-victim scoreboard might show three types of weapon systems: tanks, Bradley Fighting Vehicles, and Improved TOW vehicles. Therefore, you would enter 3 for the number of Blue types, followed by the name of each type. When entering the names of the weapon systems, try to use no more than five or six characters, such as M1A1 or BFV-1 or TANK1; otherwise, the printout becomes too crowded and difficult to read.

(7) It is suggested that you enter every weapon system on the killer-victim scoreboard. Later on, this program will allow you to choose those Blue and Red weapons you wish to have included in the force effectiveness ratios.

ENTER NO. OF BLUE WEAPON TYPES. ? 3

ENTER THE NAME OF BLUE WEAPON # 1
? TANK

ENTER THE NAME OF BLUE WEAPON # 2
? BFV

ENTER THE NAME OF BLUE WEAPON # 3
? HMMWV

Screen #4

(8) Similarly for the Red weapon types.

ENTER NO. OF RED WEAPON TYPES. ? 3

ENTER THE NAME OF RED WEAPON # 1
? TANK

ENTER THE NAME OF RED WEAPON # 2
? BMP

ENTER THE NAME OF RED WEAPON # 3
? BRDM

Screen #5

(9) Now you will be asked to fill in the data from the killer-victim scoreboard. Read each prompt carefully, and you should have no trouble entering the correct data. If you make a mistake, keep going with the correct data. You will be able to make corrections later by selecting the "Change Data" item from the main menu. Keep in mind that the last input requested on each screen is the total number of that type weapon system at the start of the wargame.

ENTER NO. OF RED TANK KILLED BY BLUE TANK
? 3

ENTER NO. OF RED BMP KILLED BY BLUE TANK
? 4

ENTER NO. OF RED BRDM KILLED BY BLUE TANK
? 2

ENTER NO. OF BLUE TANK
? 6

Screen #6

(10) Similar displays will request the remainder of the Blue vs Red and Red vs Blue results, as follows.

(11) Red vehicles killed by Blue BFV:

ENTER NO. OF RED TANK KILLED BY BLUE BFV
? 5

ENTER NO. OF RED BMP KILLED BY BLUE BFV
? 3

ENTER NO. OF RED BRDM KILLED BY BLUE BFV
? 1

ENTER NO. OF BLUE BFV
? 12

Screen #7

(12) Red vehicles killed by Blue HMMWV:

ENTER NO. OF RED TANK KILLED BY BLUE HMMWV
? 0

ENTER NO. OF RED BMP KILLED BY BLUE HMMWV
? 2

ENTER NO. OF RED BRDM KILLED BY BLUE HMMWV
? 0

ENTER NO. OF BLUE HMMWV
? 2

Screen #8

(13) Blue vehicles killed by Red tank:

ENTER NO. OF BLUE TANK KILLED BY RED TANK
? 2

ENTER NO. OF BLUE BFV KILLED BY RED TANK
? 4

ENTER NO. OF BLUE HMMWV KILLED BY RED TANK
? 0

ENTER NO. OF RED TANK
? 12

Screen #9

(14) Blue vehicles killed by Red BMP:

ENTER NO. OF BLUE TANK KILLED BY RED BMP

? 3

ENTER NO. OF BLUE BTV KILLED BY RED BMP

? 3

ENTER NO. OF BLUE HMMWV KILLED BY RED BMP

? 1

ENTER NO. OF RED BMP

? 12

Screen #10

(15) Blue vehicles killed by Red BRDM:

ENTER NO. OF BLUE TANK KILLED BY RED BRDM

? 1

ENTER NO. OF BLUE BTV KILLED BY RED BRDM

? 3

ENTER NO. OF BLUE HMMWV KILLED BY RED BRDM

? 0

ENTER NO. OF RED BRDM

? 4

Screen #11

(16) The program now returns to the main menu. (SAVE YOUR DATA!)

1 - Enter Data

2 - Change Data

3 - Save Data

4 - Perform Computations & Print Results

5 - Quit

(Enter one of the above numbers)

? 3

Screen #12

(17) Save the data immediately after data entry. You could always change it and then save it again later. Saving it as soon as possible will preclude having to re-enter all the data in case of a mishap. When you select item #3, above, to save the data, you will be prompted for a drive and filename. You may save the data to any drive and any normal filename (beginning with an alphabetic letter and having no more than eight letters and numerical digits, with no spaces in it). Remember, if you want the data saved in a particular directory on the c: drive, you must specify the entire path in the filename. For example, if you want to save your data on your fixed disk, in an ORSA\BFV-COEA directory, with a file name of BFV-RUN3.DAT, then your file name would be C:\ORSA\BFV-COEA\BFV-RUN3.DAT. For saving your data to a floppy disk, usually something like A:BFV-RUN3.DAT is sufficient, since most people don't create different directories on their floppy disks. You may want to use a different floppy for data.

Enter Drive:filename

(For example, C:test_1, is drive C: and filename TEST_1)

? A:FEI-TEST.DAT

Screen #13

(18) The data will now be saved, and the main menu will reappear.

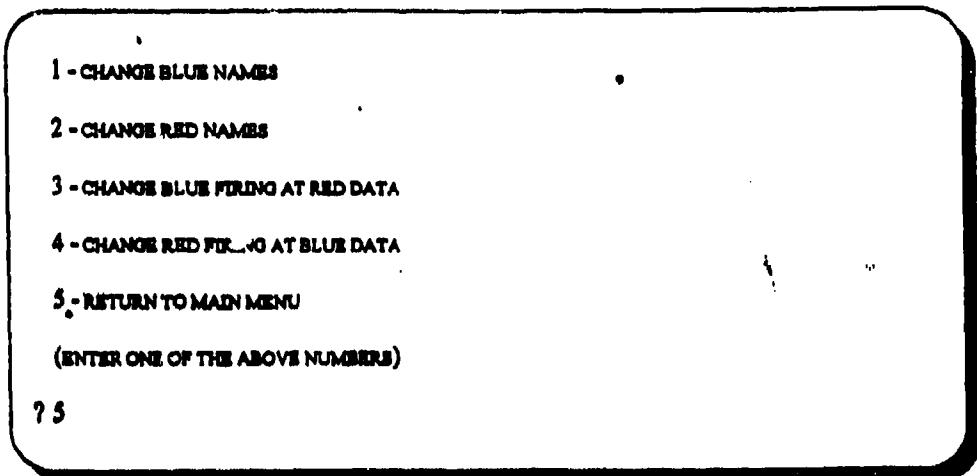
- 1 - Enter Data
- 2 - Change Data
- 3 - Save Data
- 4 - Perform Computations & Print Results
- 5 - Quit

(Enter one of the above numbers)

? 2

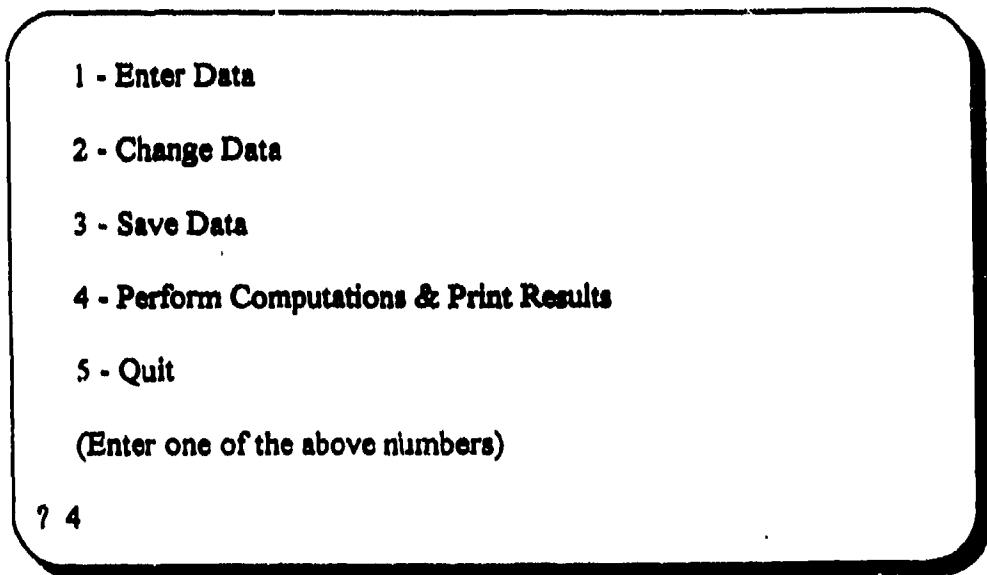
Screen #14

(19) Item #2, to change data, was selected in this example simply to display the change data menu which appears below.



Screen #15

(20) Selecting #5 returns you to the main menu.



Screen #16

(21) Selecting item #4, above, does not immediately initiate the calculations and printing of results. The user is first given an opportunity to select whether or not standard force effectiveness ratios should be included in the results (force exchange ratio, loss exchange ratio, system exchange ratio, percent system contribution, and percent force remaining, in addition to this model's force effectiveness indicators. See page 19 for definitions.

(22) Select type of results desired.

- 1 - Print Standard Effectiveness Ratios
 - 2 - Do Not Print Standard Effectiveness Ratios
- (Enter 1 or 2)

? 1

Screen #17

(23) Before doing the calculations and printing the results, the user is given the chance to select the types of forces to be included in the calculations and force effectiveness ratios. Consequently, the list of Blue forces will be displayed, followed by the list of Red forces, and the user selects the forces to be counted in the resulting ratios.

(24) Select forces to be included in computation of results.

- 1 TANK
- 2 BFV
- 3 HMMWV

Enter the numbers (one at a time, pressing the enter key after every selection) you want included in the standard force ratios.

Enter -9 when all selections have been made.

Enter -1 to select all the items.

? -1

Screen #18

(25) After making the Blue weapon system selections, the list of Red weapon systems will be displayed, as in Screen #19.

- 1 TANK
- 2 BMP
- 3 BRDM

Enter the numbers (one at a time) you want included in the standard force ratios.

Enter -9 when all selections have been made.
Enter -1 to select all the items.

? -1

Screen #19

(26) Having made the Blue force and Red force selections, the program will automatically perform the calculations and print the results. After the results have been printed, the main menu will be displayed. Selecting "5" to quit the program will also send the necessary control codes to your printer to return it to normal (after printing the results in small print).

(27) The following printout is a result of the inputs used in the above example.

DESCRIPTION: Test Case #1, 22 Dec 92, John D'Errico
FILEN/ME: A:FEI-TEST.DAT

KILLER : VICTIM MATRICES

RED VICTIM

BLUE KILLER	NUM	TANK	BMP	BRDM	TOTAL
TANK	6.0	3.0	4.0	2.0	9.0
BFV	12.0	5.0	3.0	1.0	9.0
HMMWV	2.0	0.0	2.0	0.0	2.0

BLUE VICTIM

RED KILLER	NUM	TANK	BFV	HMMWV	TOTAL
TANK	12.0	2.0	4.0	0.0	6.0
BMP	12.0	3.0	3.0	1.0	7.0
BRDM	4.0	1.0	3.0	0.0	4.0

TOTAL VALUE OF BLUE = 0.7216

STANDARD LER = 1.1763 (Reds Killed)/(Blues Killed)

TOTAL VALUE OF RED = 0.8773

STD BLUE PFR = 0.1500

FORCE EFFECTIVENESS RATIO (FER) = 0.8223

STD RED PFR = 0.2857

(Total Blue Value)/(Total Red Value)

STANDARD FER = 0.8403

FEI VALUES
(Value of one weapon)

STD. SER VALUES

FRACTIONAL PARTICIPATION INDICES

WEAPON	WEI	WEAPON	WEI
TANK	0.0367	TANK	0.0249
BFV	0.0256	BMP	0.0328
HMMWV	0.0374	BRDM	0.0442

SYSTEM	SER	PSC
TANK	1.5000	0.4900
BFV	1.2857	0.4500
HMMWV	0.5000	0.1000

LETHALITY SURVIVABILITY

	LETHALITY	SURVIVABILITY
TANK	1.5000	0.0000
BFV	0.7500	1.1111
HMMWV	1.0000	3.3333

d. Definitions.

Standard Loss Exchange Ratio (LER): (Red Losses)/(Blue Losses).

Standard Force Exchange Ratio (FER): $\frac{(\text{Red Losses})}{(\text{Blue Losses})} / \frac{(\text{Red Initial Strength})}{(\text{Blue Initial Strength})}$

Standard Specific Exchange Ratio SER: (Red Losses From Specific Blue System)/(Specific Blue System Losses)

Standard Percent System Contribution: (Red Losses Due to Specific Blue System)/(Total Red Losses)

Percent Force Remaining: (Total Number of Blue Survivors)/(Total Initial Number of Blue Forces)

Lethality: (Red Losses by Specific Blue System)/(Initial Number of Specific Blue Systems)

Survivability: ((Blue System Survivors)/(Total Blue Survivors))/((Initial Blue Systems)/(Total Initial Blue Forces))

Total Value of Blue: Sum of the value of each Blue weapon times the initial number of that Blue weapon.

Total Value of Red: Sum of the value of each Red weapon times the initial number of that Red weapon.

E. SINGLE SHOT BURSTING MUNITIONS MODEL

1. Introduction.

a. Description. This computer model was developed by Mr. John D'Errico, Dismounted Warfighting Battle Laboratory, U.S. Army Infantry School, Fort Benning, Georgia. It displays the results of firing one or more bursting (exploding) munitions from a single-shot weapon, such as the M203 grenade launcher, at an area target. Personnel in the target area may be deployed in a line, file, column, or wedge formation. Inputs required are: the biases and dispersions of the weapon; the projectile velocity; the weapon-target range; radius of damage; number of single rounds to be fired at the target; and the number, spacing, and formation of personnel in the target area.

b. Limitations. The targets depicted in this model are stationary, standing, two dimensional, personnel targets.

c. Applications. Desktop analyses involving small arms, small arms munitions, and their effects on personnel area targets.

d. Setup. This model runs on any IBM compatible PC computer. Run time depends on the number of iterations desired, with one to fifteen minutes being typical. Each iteration takes about one second.

2. Guide to Operation.

a. Equipment Required.

- (1) IBM compatible PC computer.
- (2) 3.5" disk drive.
- (3) Printer (optional).

b. Installation.

- (1) Turn on the computer and get to the DOS prompt.
- (2) Insert the 3.5" disk containing the SSBURST model into your disk drive.
- (3) See paragraph C.2.b.(3), Probability of Hit Plotting Model, for printing graphics.
- (4) From the DOS prompt, enter the command: A:SSBURST (or B:SSBURST if you're using the B: disk drive).

c. Definitions. For this model, one "iteration" refers to firing one set of rounds against the target. For example, if the number of rounds to be fired at the target is four, then each iteration will fire four rounds at the target. For trial purposes, five or ten iterations is sufficient to see the model work. For more accurate results, 200 to 1000 iterations is recommended.

d. Operation.

(1) You will be prompted for input. The first prompt will ask you to enter the horizontal fixed bias of the weapon system. Entering a zero indicates that the weapon has been zeroed for the range to the target. Biases and dispersions are in mils.

(2) The next prompt will ask you to enter the vertical fixed bias of the weapon system. Entering a zero indicates that the weapon has been zeroed for the range to the target.

(3) A total of ten prompts will appear on the screen, and you must enter a response for each one. Biases and dispersions are in mils, and distances are in meters. All ten prompts and sample responses are shown below.

(4) Keep in mind that the wedge formation was constructed for nine personnel only. If you plan to select a wedge, enter a 9 in the eighth prompt below. The other formations can accept any number of personnel.

Enter the horizontal fixed bias of the weapon system? 0
Enter the vertical fixed bias of the weapon system? 0
Enter the total horizontal variable biases and dispersions? 10
Enter the total vertical variable biases and dispersions? 10
Enter the projectile velocity in meters per sec? 60
Enter the weapon-target range in meters? 250
Enter the radius of damage in meters? 5
Enter the number of personnel in the target area? 11
Enter the space between personnel? 5
Enter the number of single rds to be fired? 4

Screen #1

(5) After you enter the last of these ten responses, the next (and last) set of prompts will appear. These prompts, and their sample responses are shown in screen #2.

(6) When selecting the type formation, keep in mind that the wedge was constructed for nine personnel only.

(7) Select formation.

Target formation:
1 - Line
2 - Column
3 - Wedge (9-man squad)
4 - File

(Enter 1 - 4)? 2

(8) After you input the number of iterations, the model will begin to graphically display each iteration's results, one iteration at a time.

(9) Each iteration's results are displayed on the screen, for a brief time. When the last iteration has been completed, the picture will remain on the screen until you either print the screen to a printer, or press any other key to return to the MS-DOS prompt.

(10) You will automatically get a printout of results, showing how each round did against each target. The printout will include the number of the round, the number of the target, the effect of each round on the target area, the average number of targets killed by each round, and the average results for the cumulative effect of all rounds. For 1000 iterations, the results are highly repeatable.

(11) Although you may select practically any number of personnel for a line, column, or file formation, the wedge currently applies to only nine personnel. Keep in mind that the scale of the display on the screen depends on the number of personnel in the target area and their separation distance. Choosing a large number of personnel separated by 10 meters will make the personnel, and possibly the bursting radius, very small or invisible.

(12) The circle which represents the bursting radius on the screen may appear to enclose a target without killing it (killed targets are shown as solid white squares). This is because the screen's vertical-to-horizontal scale may not allow a circle to look like a circle. Sometimes the bursting radius circle will appear as an oval, or ellipse. The mathematics, however, are correct, and all targets within the bursting radius are killed.

(13) A complete example, from prompts and responses to results, follows.

(14) Opening screen:

This program was authored by:
John D'Errico
Dismounted Warfighting Battle Lab
U.S. Army Infantry School
Fort Benning, GA 31905
(706) 545-7611/7000
DSN 835-7611/7000

(Press the [Enter] key to continue)?

Screen #3

(15) Description of inputs:

This program will require you to enter the following:
Horizontal and vertical fixed biases (zeroes if the weapon is assumed to be zeroed on the target).
Total horizontal variable biases and dispersions.
Total vertical variable biases and dispersions.
Muzzle or average projectile velocity.
Range to the target area.
Bursting munition's radius of damage.
Number of personnel in the target area.
Separation distance between personnel.
The number of single rounds to be fired.
The target formation: line, column, wedge or file (not applicable to a single person point target).
Number of iterations (not applicable to point targets.)

The last picture plotted on the screen remains until you press a key, in case you want to first print it with [PrtSc].

(PRESS THE ENTER KEY TO BEGIN THE PROGRAM/INPUTS)?

Screen #4

(16) Select printer option:

Do you want the program results sent to your printer? Y

Turn on your printer and press [Enter] when ready?

Screen #5

(17) Inputs:

Enter the horizontal fixed bias of the weapon system? 0
Enter the vertical fixed bias of the weapon system? 0
Enter the total horizontal variable biases and dispersions? 10
Enter the total vertical variable biases and dispersions? 10
Enter the projectile velocity in meters per sec? 60
Enter the weapon-target range in meters? 250
Enter the radius of damage in meters? 5
Enter the number of personnel in the target area? 11
Enter the space between personnel? 5
Enter the number of single rds to be fired? 4

Screen #6

(18) Target formation:

Target formation:
1 - Line
2 - Column
3 - Wedge (9-man squad)
4 - File

(Enter 1 - 4)? 2

Screen #7

(19) Since only the last screen will remain on display, until you either print it to the printer or press any other key, only the last screen in this example is shown on the following page. On page 26 the results are printed out.

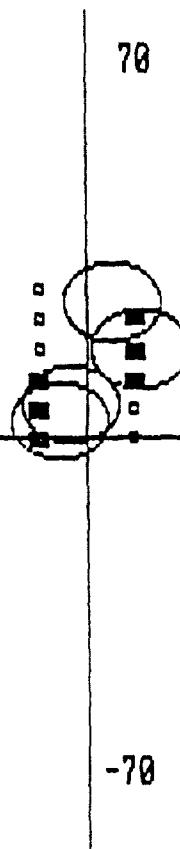
SAMPLE PRINTOUT OF THE SSBURST PROGRAM

H.Bias: 0
V.Bias: 0
H.Disp: 10
V.Disp: 10
Veloc: 60
Range: 250
Damage: .5
Tgts: 11
Rounds: 4

KILLS: 6

-70

70



Targets killed: 6
Kill ratio: .5454546

(20) The following results are based on the example above. Since only 10 iterations were used, you can expect substantially different results if you run the same example.

NUMBER OF ITERATIONS: 10

ROUND	TGT	NUM KILLED	AVG/ITERATION
1	1	4	0.400
1	2	4	0.400
1	3	4	0.400
1	4	3	0.300
1	5	2	0.200
1	6	1	0.100
1	7	0	0.000
1	8	1	0.100
1	9	0	0.000
1	10	0	0.000
1	11	0	0.000
1	ALL	19	1.900
2	1	1	0.100
2	2	2	0.200
2	3	1	0.100
2	4	2	0.200
2	5	4	0.400
2	6	4	0.400
2	7	3	0.300
2	8	1	0.100
2	9	0	0.000
2	10	1	0.100
2	11	0	0.000
2	ALL	19	1.900
3	1	1	0.100
3	2	0	0.000
3	3	1	0.100
3	4	2	0.200
3	5	1	0.100
3	6	5	0.600
3	7	1	0.100
3	8	4	0.400
3	9	1	0.100
3	10	1	0.100
3	11	1	0.100
3	ALL	18	1.800
4	1	0	0.000
4	2	0	0.000
4	3	0	0.000
4	4	0	0.000
4	5	1	0.100
4	6	0	0.000
4	7	3	0.300
4	8	2	0.200
4	9	3	0.300
4	10	3	0.300
4	11	2	0.200
4	ALL	14	1.400
CUMULATIVE		70	7.000

(21) Positioning of personnel in the various formations is according to the following format. The numbers indicate the actual number assigned each person in the target area, and match the numbers referred to in the printout of results.

Line

... 6 5 4 3 2 1

Column

11
9 10
7 8
5 6
3 4
1 2

Wedge

8
7
9 6
5
3
2
1
2 4

File

7
6
5
4
3
2
1

F. ANALYTIC HIERARCHY PROCESS MODEL

1. Introduction.

a. Description. This program was authored by Mr. John D'Errico, U.S. Army Infantry School. The Analytic Hierarchy Process (AHP) was developed by Thomas L. Saaty in the early 1970's. It is a method for ranking a set of alternatives based on multiple levels of characteristics. For example, performance and cost may be two characteristics on one level, and they might each consist of several other characteristics on a lower level. In turn, each of these characteristics could be further defined by characteristics on even lower levels. Each characteristic's value may be based on physical data such as seconds, inches, pounds, dollars, probability of hit, etc..., or on subjective evaluations. The AHP can also assist the user when developing subjective values.

b. Limitations. This model is primarily intended for first-time users of Saaty's Analytic Hierarchy Process. It is considered more as a tutorial which will enable the user to make an easy transition to the use of a spreadsheet program such as Lotus 1-2-3. Spreadsheet software would be much faster and more flexible for a complex AHP analysis.

c. Applications. The AHP has been applied to a large variety of problems in the areas of education, management of energy, political candidacy, transportation planning, and others. It has also been in use at the Pentagon. At the Infantry School the AHP was used in the combat boot analysis, multipurpose bayonet analysis, and TOW warhead improvement analysis and selection.

d. Setup. Mr. John D'Errico has developed two BASIC language programs for the Analytic Hierarchy Process. These programs will run on any IBM compatible PC. Data sorting and transformations usually take one or two days. Runs can occur at the rate of one every ten minutes. Lotus 1-2-3 can also be used to run the AHP, in which case the user gains much flexibility and speed in sensitivity analyses and run time.

2. Guide to Operation.

a. Equipment Required.

(1) IBM PC compatible computer.

(2) A 3.5" disk drive.

(3) A printer.

(4) GWBASIC . This programming language can usually be found on the DOS disks if you have a DOS version earlier than 5.0. It is also provided on the modeling disk. If using DOS version 5.0 or more recent, use the GWBASIC on the modeling disk. (A:GWBASIC)

b. Installation.

- (1) Turn on the computer and get to the DOS prompt.
- (2) If you are using the a: drive, enter the command A:GWBASIC
- (3) You will know that GWBASIC has been loaded when you see a screen with the OK prompt at the top and the ten function keys along the bottom.
- (4) Enter the command LOAD"A:AHP (You will receive another "OK" prompt).
- (5) Enter the command RUN
- (6) You will now see the prompting messages, and requests for data, according to the facsimile screens shown at the end of this section.

c. Example. The following example shows the mechanics of the AHP process and it should help to explain both the process itself and the terminology associated with it. It will also serve as a basis for describing some of the practical applications in which the AHP has been used, and the various ways of setting up the AHP to fit the problem at hand. This example assumes that there are three alternatives (ALT1, ALT2, ALT3) and five characteristics (CHAR1, CHAR2, CHAR3, CHAR4, CHAR5) which will be used to evaluate the alternatives.

(1) STEP1. Compare each characteristic to every other characteristic. Comparative values or weights may be based on either real data such as seconds, pounds, feet, or dollars, or based on subjective determinations. A matrix for these pairwise comparisons of characteristics would be set up as follows.

	CHAR1	CHAR2	CHAR3	CHAR4	CHAR5
CHAR1	1.00				
CHAR2		1.00			
CHAR3			1.00		
CHAR4				1.00	
CHAR5					1.00

(a) The 1's on the main diagonal indicate that each characteristic is equal to itself in importance. To fill in the remainder of the matrix, ask yourself how much more important or better is the item in the left column than the item across the top row. For the use of subjective data, Saaty recommends a scale of one to nine, where the number 1 indicates equality, and three, five, seven, and nine indicate that the item on the left is weakly more important, strongly more important, demonstrably more important, and absolutely more important than the item across the top. In this example, we assume that we have physical measurements which we are comparing. Accordingly, we know that CHAR1 is five times better than CHAR2, three times better than

CHAR3, three times better than CHAR4, and nine times better than CHAR5. Adding these comparative values to the matrix results in the following.

	CHAR1	CHAR2	CHAR3	CHAR4	CHAR5
CHAR1	1.00	5.00	3.00	3.00	9.00
CHAR2		1.00			
CHAR3			1.00		
CHAR4				1.00	
CHAR5					1.00

(b) After each row is filled, the reciprocal of each number in the row is entered in the symmetrically opposite cell across the main diagonal. For example, since the intersection of CHAR1 and CHAR3 is a 3, meaning CHAR1 is three times better than CHAR3, then the intersection of CHAR3 and CHAR1 is 1/3, or 0.33, meaning CHAR3 is one-third as good as CHAR1, as follows.

	CHAR1	CHAR2	CHAR3	CHAR4	CHAR5
CHAR1	1.00	5.00	3.00	3.00	9.00
CHAR2	0.20	1.00			
CHAR3	0.33		1.00		
CHAR4	0.33			1.00	
CHAR5	0.11				1.00

(c) Since we're not using subjective evaluations, we can actually fill in all cells based on the relationships established in the first row. Since CHAR1 is five times better than CHAR2 and three times better than CHAR3, then CHAR2 must be 3/5 as good as CHAR3. Similarly, since CHAR1 is five times better than CHAR2 and nine times better than CHAR5, then CHAR2 must be 9/5 (1.80) times better than CHAR5, and so on. Consequently, the matrix will be filled as follows, based on the relationships established in the first row.

	CHAR1	CHAR2	CHAR3	CHAR4	CHAR5
CHAR1	1.00	5.00	3.00	3.00	9.00
CHAR2	0.20	1.00	0.60	0.60	1.80
CHAR3	0.33	1.67	1.00	1.00	3.00
CHAR4	0.33	1.67	1.00	1.00	3.00
CHAR5	0.11	0.55	0.33	0.33	1.00

(2) STEP 2. Compute the priority vector. Mathematically, this is roughly equivalent to normalizing the principal eigenvector.

(a) For each row, take the nth root of the product of the n numbers in the row , as follows. This is all done automatically in the model, but to translate this to Lotus 1-2-3 you must know the process occurring within the model.

	CHAR1	CHAR2	CHAR3	CHAR4	CHAR5		
CHAR1	1.00	5.00	3.00	3.00	9.00	405.00	3.32
CHAR2	0.20	1.00	0.60	0.60	1.80	0.13	0.66
CHAR3	0.33	1.67	1.00	1.00	3.00	1.65	1.11
CHAR4	0.33	1.67	1.00	1.00	3.00	1.65	1.11
CHAR5	0.11	0.55	0.33	0.33	1.00	0.007	0.37

(b) Normalize this last vector by dividing each number by the sum of all the numbers. In this case, the sum of the numbers is $3.32 + 0.66 + 1.11 + 1.11 + 0.37 = 6.57$; so the normalized numbers would be as follows.

CHAR1	3.32/6.57	0.51
CHAR2	0.66/6.57	0.10
CHAR3	1.11/6.57	0.17
CHAR4	1.11/6.57	0.17
CHAR5	0.37/6.57	0.06

(c) This priority vector is really a statement of the weights attributed to each of the characteristics according to the pairwise values given in the above matrices. In other words, CHAR1 is considered to be the most important characteristic, with a score of .51, and it is five times as important as either CHAR3 or CHAR4, which each have a value of 0.17. Except for the mathematic rounding errors, the characteristics have maintained their original relationship. But this is because we have not used subjective values. Had we used subjective data, we would not have taken the first row of data in the initial matrix and automatically formed reciprocals. Instead we would have continued to enter raw subjective entries for each cell, without regard to previously implied relationships. When using purely subjective means to acquire the entries, we could very well end up saying that CHAR1 is five times better than CHAR2 and three times better than CHAR3 (which implies that CHAR2 is 3/5 as good as CHAR3) and then say that CHAR2 is half as good as CHAR3.

(3) STEP 3. Estimate the consistency of the priority vector. This will be our measure or indication of how consistently the characteristics were compared to each other during development of the original matrix of pairwise comparisons. Again, since we have not used subjective data, our matrix of pairwise comparisons should be consistent. An example of inconsistency was given at the end of the paragraph above.

(a) Multiply the matrix of comparisons by the priority vector.

COMPARISON MATRIX					PRIORITY VECTOR	V1
1.00	5.00	3.00	3.00	9.00	0.51	2.57
0.20	1.00	0.60	0.60	1.80	0.10	0.51
0.33	1.67	1.00	1.00	3.00	0.17	0.86
0.33	1.67	1.00	1.00	3.00	0.17	0.86
0.11	0.55	0.33	0.33	1.00	0.06	0.28

(b) Obtain a new vector V2 by dividing the first number in V1 by the first element of the priority vector; the second element of V1 by the second element of the priority vector; and so on, as follows.

	V2
2.57/.51	5.04
.51/.10	5.10
.86/.17	5.06
.86/.17	5.06
.28/.06	4.67

(c) Add the elements in V2 and divide this sum by the number of elements (i.e., average the numbers in V2). In our example, $(5.04 + 5.10 + 5.06 + 5.06 + 4.67)/5 = 4.99$. This number, 4.99, is an approximation of the maximum (or principal) eigenvalue, abbreviated as λ_{\max} , and it is used to estimate the consistency of the pairwise comparisons. The closer λ_{\max} is to the number of rows or columns in the matrix of comparisons, the more consistent the pairwise comparisons were.

(d) How close is close? A method of evaluating the consistency follows.

- Obtain the consistency index by dividing $(\lambda_{\max} - n)$ by $(n - 1)$. In our example, the consistency index would be $(4.99 - 5)/(5 - 1) = -.01/4 = -.003$. Since we're only interested in the magnitude of the difference, and not its direction, we'll call it .003.

- Divide the consistency index by the appropriate random index, shown in (3) below, to obtain the consistency ratio. A consistency ratio of 0.10 or less is considered acceptable. In our case, the consistency ratio would be $.003/1.12 = .003$, indicating that we were consistent in our pairwise comparisons. If we had been using subjective judgements for all our comparisons, the consistency ratio would help us catch significant errors in transitivity, such as: A is as good as B, B is twice as good as C, and A is as good as C.

- Random indices for comparison matrices of up to 15 rows (or 15 columns).

Number of Rows	Random Index
3	.58
4	.90
5	1.12
6	1.24
7	1.32
8	1.41
9	1.45
10	1.49
11	1.51
12	1.48
13	1.56
14	1.57
15	1.59

(4) STEP 4. Much of the above work was done to obtain a measure of consistency for the pairwise comparisons made in the original matrix. The priority vector, however, is what we were after. Now we have to repeat the process for the matrix of alternatives as shown below.

(a) The set of alternatives must now be evaluated in light of the above characteristics. In order to do so, pairwise comparisons must be made with respect to each characteristic above. This means we will have five sets of matrices, one for each characteristic. In the first matrix of pairwise comparisons, the question we are asking is: with respect to CHAR1, how much better or more important is ALT1 than ALT2, and so on.

(b) These matrices, along with their priority vectors, maximum eigenvalues, consistency indices (C.I.), and consistency ratios (C.R.), are shown below.

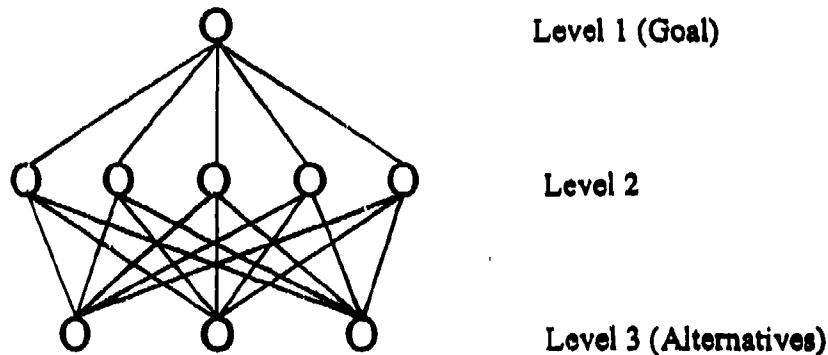
				PRIORITY VECTOR	λ_{MAX}	C.R.
CHAR1	ALT1	ALT2	ALT3			
	ALT1	1.00	0.50	2.00	0.29	3.00
	ALT2	2.00	1.00	4.00	0.57	
	ALT3	0.50	0.25	1.00	0.14	
CHAR2	ALT1	ALT2	ALT3		3.00	0.00
	ALT1	1.00	1.00	0.50	0.25	
	ALT2	1.00	1.00	0.50	0.25	
	ALT3	2.00	2.00	1.00	0.50	
CHAR3	ALT1	ALT2	ALT3			
	ALT1	1.00	2.00	2.00	0.50	3.00
	ALT2	0.50	1.00	1.00	0.25	
	ALT3	0.50	1.00	1.00	0.25	
CHAR4	ALT1	ALT2	ALT3			
	ALT1	1.00	1.00	2.00	0.40	3.00
	ALT2	1.00	1.00	2.00	0.40	
	ALT3	0.50	0.50	1.00	0.20	
CHAR5	ALT1	ALT2	ALT3			
	ALT1	1.00	1.00	1.00	0.33	3.00
	ALT2	1.00	1.00	1.00	0.33	
	ALT3	1.00	1.00	1.00	0.33	

(5) STEP 5. The matrix of priority vectors from the pairwise comparisons of the alternatives is now multiplied on the right by the priority vector from the characteristics.

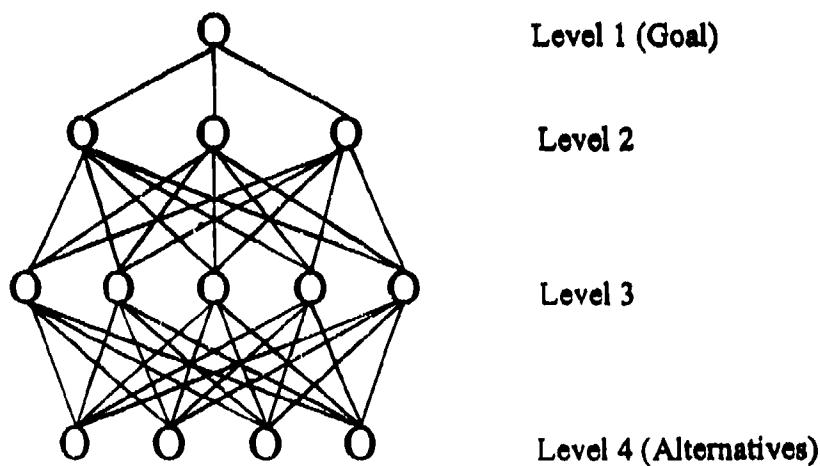
					0.51
0.29	0.25	0.50	0.40	0.33	0.10 0.340 (ALT1)
0.57	0.25	0.25	0.40	0.33	0.17 0.443 (ALT2)
0.14	0.50	0.25	0.20	0.33	0.17 0.217 (ALT3)
					0.06

This last vector, the solution vector, shows the final values of alternative 1 through alternative 3.

d. Additional Notes. In this example, there was only one level of characteristics. Additional levels may be considered in the same problem by simply repeating the above process. This is comparatively easy in a complete hierarchy, in which every item on one level is related to every item on the next higher level. Our example is a three-level, complete hierarchy, depicted by the following diagram. Level one is the solution; level two consists of the set of characteristics; and level three contains the set of alternatives.

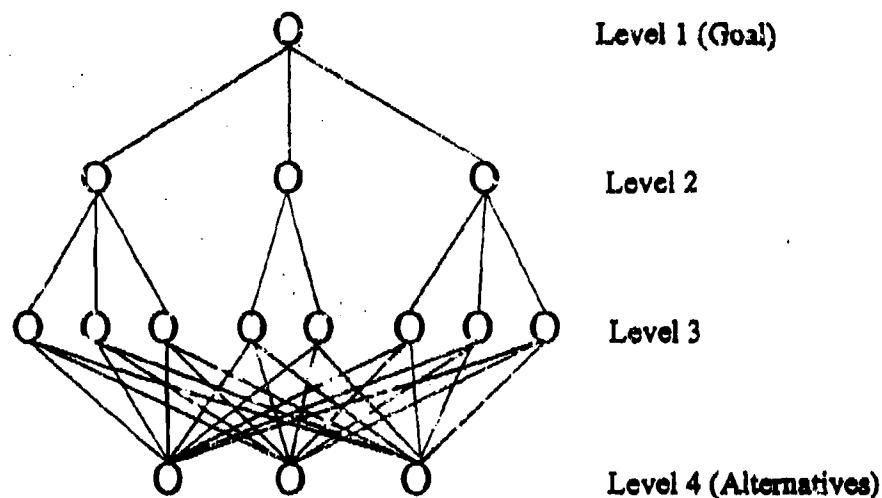


A four-level complete hierarchy might look like the following, containing the solution on level one, sets of characteristics on levels two and three, and the alternatives on level four.



In any case, the procedure remains the same. You compare the alternatives with respect to each of the characteristics in the next higher level; then compare the characteristics with respect to each of the superior characteristics in the next higher level, and so on, developing a set of priority vectors at each level. Then you compare the highest level of characteristics with respect to the solution. Finally, you must multiply each set of priority vectors in the correct order, to obtain the solution vector. The correct order of multiplication is as follows: Put the lowest level's set of priority vectors on the left (this will be the set of vectors resulting from comparing the alternatives to each other), and place each successively higher set of vectors to the right. Then multiply the matrices and vectors from left to right.

A slightly more complicated hierarchy is an "incomplete" one, where each item on one level is not necessarily related to every item on the level above, as shown below.



The easiest way to solve this type of hierarchy is to convert it into a complete hierarchy by putting zeros in the matrix of comparisons to indicate no relationship between characteristics. After that, the problem is solved as a complete hierarchy.

If any questions or problems arise from the use of this method, or the AHP program, contact Mr. D'Errico at 3914 Eve Ct, Columbus GA 31909. Office phone (706) 545-7611.

e. Displays. The following facsimile screens display all the prompts, inputs, and menus, based on the example in the text, above.

Load GWBASIC (assuming it's in the DOS directory).

C:\DOS\GWBASIC

RUN

Load the AHP model from disk drive A, and enter command "RUN".

LOAD"A:AHP

Make sure your printer is on.

Turn on your printer. This program will not run without it.

(Press the [Enter] key when ready)?

Enter three lines to describe this run. Press [Enter] to leave a line blank.

DESCRIPTION OR TITLE FOR THIS ANALYSIS (Enter 3 lines for title)

AHP-Test #1 [Enter]

20 Dec 92 [Enter]

John D'Errico [Enter]

Select "Enter Data" from the main menu.

MAIN MENU

1 - ENTER DATA

2 - CHANGE DATA

3 - SAVE DATA (As soon as you have entered all data!)

4 - PERFORM COMPUTATIONS

5 - END PROGRAM

(SELECT ONE OF THE ABOVE NUMBERS)

? 1

Enter data from the keyboard (K), unless previously saved on a disk(D).

ENTER DATA FROM KEYBOARD OR DISK ? (K/D)

? D (Enter "D" and use the sample data provided on the modeling disk)

When retrieving data from a disk, be prepared to enter the disk/path/filename.

ENTER NAME OF DISK: FILE

? a:ahpltest.dat (This file was included on your modeling disk)

After the data has been entered or changed, re-save it, and select item 4.

MAIN MENU

- 1 - ENTER DATA
- 2 - CHANGE DATA
- 3 - SAVE DATA
- 4 - PERFORM COMPUTATIONS
- 5 - END PROGRAM

(SELECT ONE OF THE ABOVE NUMBERS)

? 4

Printing will stop after the title, alternatives, and characteristics are printed, in case you want to start printing the results on a new page for a cleaner look.

DO YOU WANT TO SKIP TO NEXT PAGE ?
(Y/N)? Y

After the results are printed, select "5" to end the program.

MAIN MENU

- 1 - ENTER DATA
- 2 - CHANGE DATA
- 3 - SAVE DATA
- 4 - PERFORM COMPUTATIONS
- 5 - END PROGRAM

(SELECT ONE OF THE ABOVE NUMBERS)

? 5

Enter the command "SYSTEM" (without quotation marks) to return to DOS.

MAIN MENU

- 1 - ENTER DATA**
- 2 - CHANGE DATA**
- 3 - SAVE DATA**
- 4 - PERFORM COMPUTATIONS**
- 5 - END PROGRAM**

(SELECT ONE OF THE ABOVE NUMBERS)

? 5
Ok
SYSTEM

f. Results. The following results were based on the example given in the text above.

AHP TEST #1

20 Dec 92
John D'Zurico

ALTERNATIVES EVALUATED:

ALT1
ALT2
ALT3

CHARACTERISTICS CONSIDERED:

CHAR1
CHAR2
CHAR3
CHAR4
CHARS

CHARACTERISTIC VALUES

	CHAR1	CHAR2	CHAR3	CHAR4	CHARS
CHAR1	1.00	5.00	3.00	3.00	9.00
CHAR2	0.20	1.00	0.60	0.60	1.00
CHAR3	0.33	1.67	1.00	1.00	3.00
CHAR4	0.33	1.67	1.00	1.00	3.00
CHARS	0.11	0.55	0.33	0.33	1.00

EIGENVECTOR MAX EIGENVALUE: 4.99 CONSISTENCY RATIO: -0.00250
0.51
0.10
0.17
0.17
0.06

CHAR1	ALT1	ALT2	ALT3	EIGENVECTOR	EIGENVALUE
ALT1	1.00	0.50	2.00	0.29	3.00
ALT2	2.00	1.00	4.00	0.57	
ALT3	0.50	0.25	1.00	0.14	

CONSISTENCY RATIO: 0.00000

CHAR2	ALT1	ALT2	ALT3	EIGENVECTOR	EIGENVALUE
ALT1	1.00	1.00	0.50	0.25	3.00
ALT2	1.00	1.00	0.50	0.25	
ALT3	2.00	2.00	1.00	0.50	

CONSISTENCY RATIO: 0.00000

CHAR3	ALT1	ALT2	ALT3	EIGENVECTOR	EIGENVALUE
ALT1	1.00	2.00	2.00	0.50	3.00
ALT2	0.50	1.00	1.00	0.25	
ALT3	0.50	1.00	1.00	0.25	

CONSISTENCY RATIO: 0.00000

CHAR4	ALT1	ALT2	ALT3	EIGENVECTOR	EIGENVALUE
ALT1	1.00	1.00	2.00	0.40	3.00
ALT2	1.00	1.00	2.00	0.40	
ALT3	0.50	0.50	1.00	0.20	

CONSISTENCY RATIO: 0.0000

CHAR5	ALT1	ALT2	ALT3	EIGENVECTOR	EIGENVALUE
ALT1	1.00	1.00	1.00	0.33	3.00
ALT2	1.00	1.00	1.00	0.33	
ALT3	1.00	1.00	1.00	0.33	

CONSISTENCY RATIO: 0.0000

RANKING OF ALTERNATIVES:

ALT2 = 0.443

ALT1 = 0.340

ALT3 = 0.217

G. DATA RANKING.

1. Introduction.

- a. Description.** This program, developed by Mr. John D'Errico, takes any set of numerical data as input, sorts it into ascending and descending orders, and provides the ranks associated with each.
- b. Limitations.** This program can accept a maximum of 1000 data points.
- c. Applications.** Desktop tool for data analysis.
- d. Setup.** This model runs on a DOS-based computer. Data entry consists solely of entering the numbers to be sorted and ranked. Sorting and ranking will usually take less than a minute.

2. Guide to Operation.

a. Equipment Required.

- (1) IBM compatible PC computer.
- (2) 3.5" disk drive.
- (3) Printer.

b. Installation.

- (1) Turn on the computer and get to the DOS prompt.
- (2) Insert the 3.5" disk containing the RANKDATA program into your computer's disk drive.
- (3) From the DOS prompt, enter the command A:RANKDATA (or B:RANKDATA if you're using the B: disk drive).

c. Definitions.

- (1) Rank. After a set of numbers is put into order (ascending order, for example) the rank of each number is simply the number of its position in the ordered list. However, when the same number is repeated on the list, their rank is determined by averaging the numbers of their positions. For example, assume that the numbers 12, 3, 17, 11, 12, 6, 42, 3 must be ranked.

The first task is to sort the numbers into (for this example) ascending order. The sorted list of numbers then becomes 3, 3, 6, 11, 12, 12, 17, 42. The number 3 occupies position 1 and 2, so each 3 gets a rank of 1.5 since $(1 + 2)/2 = 3/2 = 1.5$. The number 6 gets a rank of 3 since it holds position 3; the number 11 gets a rank of 4 since it holds position 4; the numbers 12, occupying positions 5 and 6, each get a rank of 5.5 since that's the average of position numbers 5 and 6. Number 17 gets a rank of 7 since it holds position 7, and the number 42 gets a rank of 8 since it holds position 8 in the ordered list of eight numbers.

(2) Ascending Order. Numbers in ascending order are listed with the smallest number at the top of the list and the largest number at the bottom of the list.

(3) Descending Order. Numbers in ascending order are listed with the largest number at the top of the list and the smallest number at the bottom of the list.

d. Operation.

(1) The first display is as follows:

This program accepts up to 1000 numbers, then prints the numbers as entered, followed by the numbers in ascending and descending orders and their associated ranks. Ranks are assigned from 1 to n.

Tied scores are assigned the mean of the ranks for which they are tied.

(Press RETURN to begin the program)

Screen #1

(2) The second display prompts you to enter score (number) #1, score #2, score #3, etc..., with the instruction to enter the number -99 when you have no more numbers to enter.

Enter score # 1 ? 3

(Enter -99 after last score has been entered)

Screen #2

(3) The final display will prompt you to make sure that your printer is turned on.

Make sure that your printer is on

Press the [Enter] key when ready.

?

Screen #3

(4) The following is a sample printing from this program.

Data Entered: 5 3 11 2 7 28 5 2 9 24 35 17 12 7 9
16 12 3 7 9 2 4 6

ASCENDING		DESCENDING	
DATA	RANKS	DATA	RANKS
2	2	35	1
2	2	28	2
2	2	24	3
3	4.5	17	4
3	4.5	16	5
4	6	12	6.5
5	7.5	12	6.5
5	7.5	11	8
6	9	9	10
7	11	9	10
7	11	9	10
7	11	7	13
9	14	7	13
9	14	6	15
11	16	5	16.5
12	17.5	5	16.5
12	17.5	4	18
16	19	3	19.5
17	20	3	19.5
24	21	2	22
28	22	2	22
35	23	2	22

(5) At this point the program ends and returns you to the DOS prompt.

H. LAGRANGE INTERPOLATION.

1. Introduction.

- a. Description. This model, developed by Mr. John D'Errico, uses Lagrange polynomials to interpolate between two points on a nonlinear curve.
- b. Limitations. This method is subject to error if using a large number of known data points as a basis for the interpolation.
- c. Applications. Desktop tool for data analysis.
- d. Setup. This program runs on an IBM compatible PC computer. It takes approximately one minute to enter five data points, and less than a minute to display the interpolation.

2. Guide to Operation.

a. Equipment Required.

- (1) IBM compatible PC computer.
- (2) 3.5" disk drive.

b. Installation.

- (1) Turn on the computer and get to the DOS prompt.
- (2) Insert the 3.5" disk containing the LAGRANGE program into your disk drive.
- (3) From the DOS prompt enter the command: A:LAGRANGE

c. Explanation.

(1) Given a set of data points such as those in Table 1, there is often a need to determine a data point which is not listed in the table. For example, we might need to estimate the probability of hit [P(H)] at a range of 1.2 kilometers, based on the data in Table 1.

RANGE	P(H)
0.1	0.90
0.5	0.88
1.0	0.58
2.0	0.18

Table 1

(2) A common practice, due to its simplicity and speed, is to interpolate linearly between two given data points. Using Table 1 this would mean interpolating between the ranges of 1.0 and 2.0 kilometers in order to find the value not given--the $P(H)$ at 1.2 kilometers. However, when the given data does not fall along a straight line, linear interpolation is subject to gross errors, particularly if the data points within which the interpolation is done are not close together.

(3) The method described herein uses Lagrange's form of interpolation polynomials. This is a widely used form for interpolation within a set of given data points. The given data points may be equally or unequally spaced, and may lie along a nonlinear curve.

(4) This method is also subject to errors, particularly if using a large number of given data points to make the interpolation. A way to minimize the error is to take a couple of data points on either side of the value to be interpolated, ignoring the data points which are farther away.

(5) This method is presented as an alternative to linear interpolation, not as a substitute, and it is to be used when a straight line would be substantially off the true curve of known data points, as shown in Figure 1.

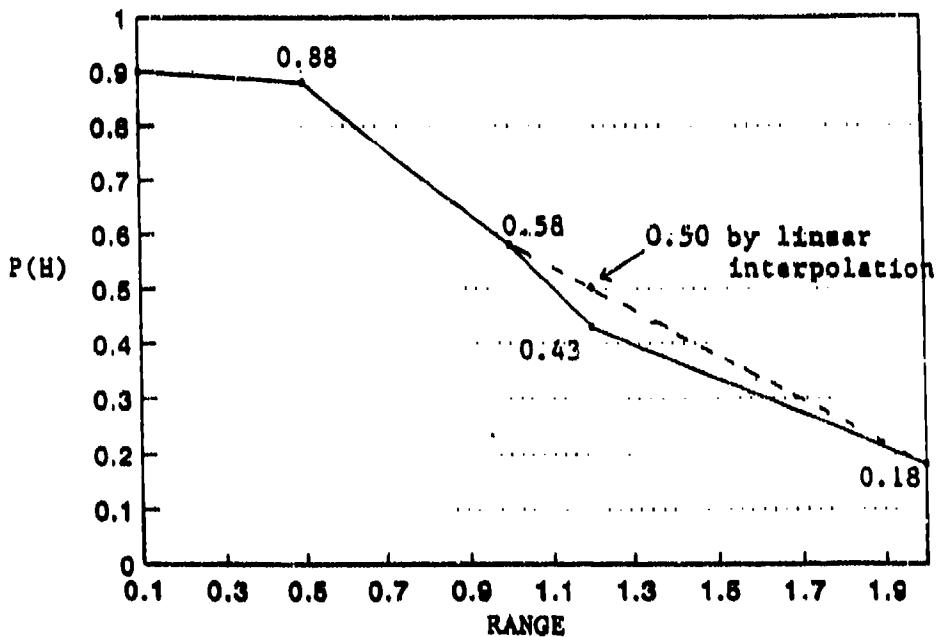


FIGURE #1

(6) This method should only be used to interpolate within the range of given data points. There are better methods for interpolating (or extrapolating) outside the range of the given data points; namely, Newton's forward and backward differences, among others.

(7) There is more than one way to derive the approximating interpolative polynomial used herein, such as the method of undetermined coefficients; however, the method selected here is straight forward, and was easy to program.

(8) Equations.

(a) Since the derivations, proofs, and uniqueness theorems are readily available in a multitude of books on numerical analysis, these are not duplicated in this paper.

(b) Given a set of $n+1$ data points of the form $(x_i, f(x_i))$, the collocation polynomial (the n th degree polynomial fitting those points) is

$$p(x) = \sum f(x_j)L_j(x)$$

where, for each j , $0 < j < n$, $f(x_j)$ is the given value along the y -axis associated with the given x_j value along the x -axis. $L_j(x)$ is the n th degree polynomial defined as

$$L_j(x) = \frac{(x-x_0)(x-x_1)\dots(x-x_{j-1})(x-x_{j+1})\dots(x-x_n)}{(x_j-x_0)(x_j-x_1)\dots(x_j-x_{j-1})(x_j-x_{j+1})\dots(x_j-x_n)} = \prod_{i=0}^{n-1} \frac{(x-x_i)}{(x_j-x_i)}$$

d. Operation.

(1) Using the data in Table 1, assume that we want to estimate the probability of hit at 1.2 kilometers. Table 1 is repeated below.

RANGE	P(H)
0.1	0.90
0.5	0.88
1.0	0.58
2.0	0.18

Table 1

(2) The LAGRANGE program's first screen asks if you want a program description and explanation displayed on the screen. For this example we will select Y(es).

If you want a brief program description/explanation, enter
Y or y and press [Enter]; otherwise, simply press [Enter].
? y

Screen #1

(3) The next three displays consist of the explanation.

DESCRIPTION

This program accepts any number of (x,y) coordinates, determines the Lagrange form of interpolative polynomial which fits the (x,y) data points, and then asks the user to enter any number of x-values for which a y-value must be predicted.

It is recommended that this program be used to interpolate only between the lowest and highest known x-values (don't extrapolate) and that only 3 to 6 coordinates of known points be used for this interpolation.

(Press [Enter] to continue)

Screen #2

(4) Continued description:

When entering the first set of data, simply enter the x-value and y-value, separated by a comma, and press [Enter] after each pair of coordinates. For example, to enter the coordinates (1,2), (2,4), and (3,9) you would first enter a "3" in response to "Enter the number of known (x,y) data points." Then you would enter the three coordinates as follows:

1,2 [Enter]
2,4 [Enter]
3,9 [Enter]

(Press [Enter] to continue)

Screen #3

(5) Final screen of descriptions.

After you have entered the known (x,y) coordinates, you will be asked to enter the number of x-values for which you need predicted y-values. Simply enter the number of x-values for which you need y-values interpolated. Finally, you will be asked to enter the x-values, one at a time, pressing the [Enter] key after each x-value entry.

Screen #4

(6) Now you will be prompted to enter the number of known data points.

Enter the number of known (x,y) data points.

? 4

Screen #5

(7) The next four prompts ask you to enter the data points. Only the first prompt is shown here, since the remaining three are identical except for the coordinate entered.

Enter X,Y for Data Point #1

? .1,.9

(Enter the X and Y values, separated by a comma)

Screen #6

(8) The next prompt asks for the number of x-values for which you need a y-value. Only one y-value is requested in this example--the P(H) at 1.2 kilometers.

Enter the number of x-values for which you need
a y-value predicted.

? 1

Screen #7

(9) Now you must enter the single x-value. For this example the response is 1.2, representing 1.2 kilometers.

Enter X value # 1

? 1.2

Screen #8

(10) The final display lists the x-values and y-values you entered, followed by the x-value and y-value you needed interpolated.

X	Y
.1	.9
.5	.88
1	.58
2	.18
1.2	.4347836

c:\>

Screen #9

(11) As you can see from Screen #9, the program has ended and returned you to the prompt you started with--in this case, the root directory.

I. FUNDAMENTAL DUEL

i. Introduction.

a. Description. This model (Reference 4, chapter 17) depicts the outcome of two opposing, single shot, direct fire weapon systems, each having an unlimited amount of ammunition. Inputs required are each weapon's reliability, rate of fire, and probability of kill given a single shot. The results are displayed in terms of the probability that the Blue weapon wins the duel and the probability that the Red weapon wins the duel.

b. Limitations. This model evaluates the outcome of a simple one-on-one duel, based on rates of fire and exponentially distributed firing times between rounds.

c. Applications. Desktop analytic tool for applying a simple concept to evaluations of single shot weapon systems.

d. Setup. This model runs on an IBM compatible PC computer equipped with Lotus 1-2-3. Data can be entered into the model and results displayed in less than a minute.

2. Guide to Operation.

a. Equipment Required.

- (1) IBM PC compatible computer.
- (2) 3.5" disk drive.
- (3) Lotus 1-2-3 spreadsheet software.

b. Installation.

- (1) Turn on your computer and activate Lotus 1-2-3.
- (2) Insert the 3.5" disk containing the FUNDUEL model into your computer's 3.5" disk drive.
- (3) From the Lotus 1-2-3 menu, load the A:FUNDUEL.WK1 model (or B:FUNDUEL if you're working from the b: drive) by entering /FR, then backspace to erase the default path, and enter A: and press the [Enter] key. After the Lotus 1-2-3 files are shown, select the FUNDUEL file.

c. Operation.

(1) Move the cursor to the cell you wish to change. This should be cell B2, B3, B4, B6, B7, B8, B10, or B11.

(2) As you make a change in one cell, the probabilities of Blue and Red winning (cells F5 and F6 respectively) are automatically recalculated for a practically instantaneous answer.

THE FUNDAMENTAL DUEL		
REL-B	1	
PR HIT-B	0.6	
PR K/H-B	0.7	
PSSK-B	0.42	PROB-B 0.567568
REL-R	1	
PR HIT-R	0.8	PROB-R 0.432432
PR K/H-R	0.8	
PSSK-R	0.64	CHECK 1
ROF-B	2	
ROF-R	1	

Screen #1

d. Definitions.

REL-B: Reliability of the Blue weapon system.

PR HIT-B: Blue weapon's probability of hitting the Red target.

PR K/H-B: Blue weapon's probability of kill, given a hit, on the Red target.

PSSK-B: Blue weapon's probability of kill given a single shot at the Red target. It is the product of the probability of hit and the probability of kill given a hit.

REL-R: Reliability of the Red weapon system.

PR HIT-R: Red weapon's probability of hitting the Blue target.

PR K/H-R: Red weapon's probability of kill, given a hit, on the Blue target.

PSSK-R: Red weapon's probability of kill given a single shot at the Blue target. It is the product of the probability of hit and the probability of kill given a hit.

ROF-B: Blue weapon's rate of fire.

ROF-R: Red weapon's rate of fire.

PROB-B: The probability that Blue wins the duel.

PROB-R: The probability that Red wins the duel.

CHECK: Verification of the equations. It is the sum of PROB-B and PROB-R, which should be equal to 1.00.

c. Explanation.

(1) Background. In a fundamental duel, it is hypothesized that two duelists, Blue (B) and Red (R), fire at each other until one is put out of action. The firing times, or time between rounds, for each duelist is considered to be of a random character with known probability density functions, the parameters for which may be different for Blue and Red. At the start of the engagement, each contestant loads, aims, and fires his first round at his opponent. Thus, in the fundamental duel, both start with unloaded weapons. It is also assumed here that each time Blue and Red fire at each other they have constant single shot kill probabilities, although such kill probabilities of Blue and Red may be different. Both Blue and Red have unlimited ammunition supplies, so that a kill is certain.

(2) Definitions.

ρ_B = mean rate of fire of Blue (B)

ρ_R = mean rate of fire of Red (R)

p_B = single shot kill probability of Blue against Red

p_R = single shot kill probability of Red against Blue

$P(B)$ = chance that B wins the duel

$P(R)$ = chance that R wins the duel = $1 - P(B)$

(3) The mean rates of fire, ρ_B and ρ_R , are, respectively, the reciprocals of the mean times between rounds fired by Blue and Red.

(4) The single shot chances of kill, p_B and p_R , may be built up or determined by taking the product of the chance of a hit and the conditional probability that a hit is a kill; i.e.,

$$p_B = p_B(h) p_B(k|h) \text{ and } p_R = p_R(h) p_R(k|h).$$

(5) Finally, we make an assumption that appears of practical value; namely, that the time to fire the first round and the times between rounds fired for B and R follow single

parameter negative exponential distributions. So, for random times t

$$f(t) = \rho \exp(-\rho t), \text{ where } \rho = \rho_B \text{ or } \rho_R, \text{ as needed. Mean time between rounds} = 1/\rho.$$

(6) Since the exponential distribution is equivalent to the chi-square distribution with two degrees of freedom, this means that the time at which the nth round is fired is the sum of n independent selections from the above equation, or the chi-square distribution with $2n$ degrees of freedom (or the gamma distribution) given by

$$f(t_n) = \rho^n (t_n)^{n-1} \exp(-\rho t_n) / (n-1)!$$

Then, the chance that Blue wins is: $P(B) = \frac{P_B \rho_B}{P_B \rho_B + P_R \rho_R}$

and the chance that Red wins is $P(R) = 1 - P(B) = \frac{P_R \rho_R}{P_B \rho_B + P_R \rho_R}$

(7) Consequently, for exponentially distributed firing times between rounds, the chance that a side wins is the kill rate for that side divided by the sum of the kill rates for both sides, which is a rather simple outcome. Hence, the value of kill rate as a key measure of effectiveness is evident. Note that if the single shot kill probabilities of B and R are equal, then their rates of fire take over; and if their rates of fire also are equal, each B and R have a 50% chance of winning. The chance of a draw, or both being killed, is zero.

J. FUNDAMENTAL DUEL WITH LIMITED AMMUNITION FOR BLUE

1. Introduction.

a. Description. This model (Reference 4, chapter 17) depicts the outcome of two opposing, single shot, direct fire weapon systems when the Blue weapon system has a limited amount of ammunition. Inputs required are each weapon's reliability, rate of fire, and probability of kill given a single shot. The results are displayed in terms of the probability that the Blue weapon wins the duel and the probability that the Red weapon wins the duel.

b. Limitations. This model evaluates the outcome of a simple one-on-one duel, based on rates of fire, probabilities of kill, and exponentially distributed firing times between rounds.

c. Applications. Desktop analytic tool for applying a simple concept to evaluations of single shot weapon systems.

d. Setup. This model runs on an IBM compatible PC computer equipped with Lotus 1-2-3. Data can be entered into the model and results displayed in less than a minute.

2. Guide to Operation.

a. Equipment Required.

- (1) IBM PC compatible computer.
- (2) 3.5" disk drive.
- (3) Lotus 1-2-3 spreadsheet software.

b. Installation.

- (1) Turn on your computer and activate Lotus 1-2-3.
- (2) Insert the 3.5" disk containing the LIMAMMOB model into your computer's 3.5" disk drive.
- (3) From the Lotus 1-2-3 menu, load the A:LIMAMMOB model (or B:LIMAMMOB if you're working from the b: drive) by entering /FR, then backspace to erase the default path, and enter A: and press the [Enter] key. After the Lotus 1-2-3 files are shown, select the LIMAMMOB file.

c. Operation.

(1) Move the cursor to the cell you wish to change. This should be cell B2, B3, B4, B6, B7, B8, B10, B11, or B12.

(2) As you make a change in one cell, the probabilities of Blue and Red winning (cells E3 and E4 respectively) are automatically recalculated for a practically instantaneous answer.

FUNDAMENTAL DUEL (LIMMITED AMMO FOR BLUE)

REL-B	1	PROB-B	0.558272
PR HIT-B	0.6	PROB-R	0.441728
PR K/H-B	0.7	CHECK	1
PSSK-B	0.42		
REL-R	1		
PR HIT-R	0.8		
PR K/H-R	0.8		
PSSK-R	0.64		
ROF-B	2		
ROF-R	1		
ROUNDS-B	5		

Screen #1

d. Definitions.

REL-B: Reliability of the Blue weapon system.

PR HIT-B: Blue weapon's probability of hitting the Red target.

PR K/H-B: Blue weapon's probability of kill, given a hit, on the Red target.

PSSK-B: Blue weapon's probability of kill given a single shot at the Red target. It is the product of the probability of hit and the probability of kill given a hit.

REL-R: Reliability of the Red weapon system.

PR HIT-R: Red weapon's probability of hitting the Blue target.

PR K/H-B: Red weapon's probability of kill, given a hit, on the Blue target.

PSSK-R: Red weapon's probability of kill given a single shot at the Blue target. It is the product of the probability of hit and the probability of kill given a hit.

ROF-B: Blue weapon's rate of fire.

ROF-R: Red weapon's rate of fire.

PROB-B: The probability that Blue wins the duel.

PROB-R: The probability that Red wins the duel.

ROUNDS-B: The number of rounds available to the Blue weapon system.

CHECK: Verification of the equations. It is the sum of PROB-B and PROB-R, which should be equal to 1.00.

e. Explanation.

(1) This model uses the same parameters as the fundamental duel where both sides have unlimited amounts of ammunition, except that now Blue is limited by N-rounds.

(2) When Blue has a fixed number of rounds equal to N, and Red has an unlimited supply of ammunition, then for the assumption of exponential firing times between rounds, the chance that Blue wins is given by

$$P(B) = [p_B \rho_B / (p_B \rho_B + p_R \rho_R)] (1 - [{}_R q_B \rho_B / (p_R \rho_R + \rho_B)]^N)$$

and

$$P(BR) = 0$$

Note: ${}_R q_B = 1 - p_B$ = single shot survival probability for Red when fired on by Blue.

$P(BR)$ = chance of a draw (B and R kill each other).

(3) See paragraph I.2.e. for additional explanations of the fundamental duel.

J. FUNDAMENTAL DUEL WITH LIMITED AMMUNITION FOR RED

1. Introduction.

a. Description. This model (Reference 4, chapter 17) depicts the outcome of two opposing, single shot, direct fire weapon systems when the Red weapon system has a limited amount of ammunition. Inputs required are each weapon's reliability, rate of fire, and probability of kill given a single shot. The results are displayed in terms of the probability that the Blue weapon wins the duel and the probability that the Red weapon wins the duel.

b. Limitations. This model evaluates the outcome of a simple one-on-one duel, based on rates of fire, probabilities of kill, and exponentially distributed firing times between rounds.

c. Applications. Desktop analytic tool for applying a simple concept to evaluations of single shot weapon systems.

d. Setup. This model runs on an IBM compatible PC computer equipped with Lotus 1-2-3. Data can be entered into the model and results displayed in less than a minute.

2. Guide to Operation.

a. Equipment Required.

(1) IBM PC compatible computer.

(2) 3.5" disk drive.

(3) Lotus 1-2-3 spreadsheet software.

b. Installation.

(1) Turn on your computer and activate Lotus 1-2-3.

(2) Insert the 3.5" disk containing the LIMAMMOR model into your computer's 3.5" disk drive.

(3) From the Lotus 1-2-3 menu, load the A:LIMAMMOR model (or B:LIMAMMOR if you're working from the b: drive) by entering /FR, then backspace to erase the default path, and enter A: and press the [Enter] key. After the Lotus 1-2-3 files are shown, select the LIMAMMOR file.

c. Operation.

(1) Move the cursor to the cell you wish to change. This should be cell B2, B3, B4, B6, B7, B8, B10, B11, or B12.

(2) As you make a change in one cell, the probabilities of Blue and Red winning (cells E3 and E4 respectively) are automatically recalculated for a practically instantaneous answer.

BASIC DUEL (LIMITED AMMO FOR RED)		
REL-B	1	
PR HIT-B	0.6	
PR K/H-B	0.7	
PSSK-B	0.42	PROB-B 0.567582
REL-R	1	PROB-R 0.432418
PR HIT-R	0.8	CHECK 1
PR K/H-R	0.8	
PSSK-R	0.64	
ROF-B	2	
ROF-R	1	
ROUNDS-R	5	

Screen "1

d. Definitions.

REL-B: Reliability of the Blue weapon system.

PR HIT-B: Blue weapon's probability of hitting the Red target.

PR K/H-B: Blue weapon's probability of kill, given a hit, on the Red target.

PSSK-B: Blue weapon's probability of kill given a single shot at the Red target. It is the product of the probability of hit and the probability of kill given a hit.

REL-R: Reliability of the Red weapon system.

PR HIT-R: Red weapon's probability of hitting the Blue target.

PR K/H-B: Red weapon's probability of kill, given a hit, on the Blue target.

PSSK-R: Red weapon's probability of kill given a single shot at the Blue target. It is the product of the probability of hit and the probability of kill given a hit.

ROF-B: Blue weapon's rate of fire.

ROF-R: Red weapon's rate of fire.

PROB-B: The probability that Blue wins the duel.

PROB-R: The probability that Red wins the duel.

ROUNDS-R: The number of rounds available to the Red weapon.

CHECK: Verification of the equations. It is the sum of PROB-B and PROB-R, which should be equal to 1.00.

e. Explanation.

(1) This model uses the same parameters as the fundamental duel where both sides have unlimited amounts of ammunition, except that now Red is limited by N-rounds.

(2) When Red has a fixed number of rounds equal to M, and Blue has an unlimited supply of ammunition, then for the assumption of exponential firing times between rounds, the chance that Blue wins is given by

$$P(B) = \frac{P_B \rho_B}{P_B \rho_B + P_R \rho_R} + \left[\frac{P_R \rho_R}{P_B \rho_B + P_R \rho_R} \right] \times \left[\frac{\rho_B \rho_R}{P_B \rho_B + \rho_R} \right]^M$$

and

$$P(BR) = 0.$$

Note: $\rho_R = 1 - p_R$ = single shot survival probability for Blue when fired on by Red.

$P(BR)$ = chance of a draw (B and R kill each other).

(3) See paragraph I.2.e. for additional explanations of the fundamental duel.

K. LANCHESTER'S SQUARE LAW AS A FUNCTION OF TIME

1. Introduction.

a. Description. This model (Reference 4, chapter 28) determines the remaining Blue forces and remaining Red forces at any given time during a battle between homogeneous forces. Inputs required for each side are the total number of weapon systems, and each weapon's probability of hit, probability of kill given a hit, and rate of fire.

b. Limitations. This model evaluates the outcome of one set of identical weapon systems against an opposing set of identical weapon systems. It is based on each weapon's constant kill rate of opposing forces.

c. Applications. Desktop analytic tool for evaluating homogeneous force effectiveness in terms of a basic concept.

d. Setup. This model runs on an IBM compatible PC computer equipped with Lotus 1-2-3. Data can be entered into the model and results displayed in less than a minute.

2. Guide to Operation.

a. Equipment Required.

- (1) IBM PC compatible computer.
- (2) 5.25" disk drive.
- (3) Lotus 1-2-3 spreadsheet software.

b. Installation.

- (1) Turn on your computer and activate Lotus 1-2-3.
- (2) Insert the 3.5" disk containing the LANBASIC model into your computer's 3.5" disk drive.
- (3) From the Lotus 1-2-3 menu, load the A:LANBASIC model (or B:LANBASIC if you're working from the b: drive) by entering /FR, then backspace to erase the default path, and enter A: and press the [Enter] key. After the Lotus 1-2-3 files are shown, select the LANBASIC file.

c. Operation.

(1) The model parameters are shown below.

	BATTLE	BLUE	RED
INITIAL FORCE SIZE*		100	50
P(H)*		0.5	0.5
P(K H)*		0.25	0.5
RATE OF FIRE*		0.4	0.4
CONSTANT KILL RATE		0.05	0.1
TIME ELAPSED (T)*	2		
STRENGTH AT TIME T		90.97	40.47
FORCE ADVANTAGE RATIO		2.25	0.44
LER AT TIME T	1.05		
TIME OF ANNIHILATION		ERR	12.4645
0.141421 0.141893 1.010017			
1.414214 0.707107			

Screen #1

(2) Move the cursor to the cell you wish to change. You may change only the items in Screen #1 which are marked with an asterisk. In Screen #1 an error (ERR) is shown in cell F13 because the given data has the Red force annihilated before the Blue force; consequently, Blue cannot be annihilated (no force remaining).

d. Definitions.

Initial Force Size: Number of identical weapons on the Blue or Red side.

P(H): A weapon's probability of hit against the opposing force target.

P(K|H): Probability of kill given a hit against the opposing force target.

Rate of Fire: The rate of fire in rounds per time unit (usually minutes).

Constant Kill Rate: The constant rate at which a single weapon kills an opposing force target.

Time Elapsed: The battle time.

Strength at Time T: The number of Blue or Red weapons remaining at the end of time T.

Force Advantage Ratio: The number of friendly weapons divided by the number of opposing force weapons after time T.

LER at Time T: The number of Red losses divided by the number of Blue losses.

Time of Annihilation: The time at which there are no remaining weapons on that side. An error, indicated by "ERR" will be displayed on the side of the force which has weapons remaining after the opposing force has been annihilated. This is because the winning force cannot be annihilated after all opposing weapons have been destroyed.

Note: The data appearing in rows 14 and 15 are intermediate calculations.

c. Explanation of Lanchester's square law as a function of time.

(1) Definitions.

B_0 = Initial Blue strength

R_0 = Initial Red strength

B = Size of the Blue force at any time t

R = Size of the Red force at any time t

ρ = Constant rate at which a single Blue weapon kills a Red weapon

β = Constant rate at which a single Red weapon kills a Blue weapon

(2) The remaining Blue forces $B(t)$ and remaining Red forces $R(t)$ at any time t are given by

$$B(t) = B_0 \cosh \sqrt{\rho\beta} t - \sqrt{\rho\beta} R_0 \sinh \sqrt{\rho\beta} t$$

$$R(t) = R_0 \cosh \sqrt{\rho\beta} t - \sqrt{\rho\beta} B_0 \sinh \sqrt{\rho\beta} t$$

(3) The time t_R at which Red is annihilated (i.e., $R(t) = 0$) is given by

$$t_R = [1/(2\sqrt{\rho\beta})] \ln [(\sqrt{\rho}B_0 + \sqrt{\beta}R_0)/(\sqrt{\beta}B_0 - \sqrt{\rho}R_0)]$$

(4) Similarly, if Red wins, then Blue's time of annihilation (i.e., $B(t) = 0$) is given by

$$t_B = [1/(2\sqrt{\rho\beta})] \ln [(\sqrt{\rho}B_0 + \sqrt{\beta}R_0)/(\sqrt{\beta}R_0 - \sqrt{\rho}B_0)]$$

L. DUEL WHEN BLUE HAS A WEAPON FAILURE RATE

1. Introduction.

a. Description. This model (Reference 4, chapter 17) depicts the outcome of two opposing, single shot, direct fire weapon systems, each having an unlimited amount of ammunition, as in the fundamental duel, but including the idea of weapon failure times. Inputs required are: each weapon's probability of hit, probability of kill given a hit, round reliability, and rate of fire; the number of Blue weapons, and the Blue weapon failure rate. Results are displayed in terms of the probability that Blue wins the duel and the probability that Red wins the duel.

b. Limitations. Only homogeneous forces are used in this model. Blue and Red have unlimited ammunition supplies; Blue has a limited number of weapons, and Red has a failure-free weapon.

c. Applications. Desktop analytic tool for applying simple failure rates to evaluations of single shot weapon systems.

d. Setup. This model runs on an IBM PC compatible computer equipped with Lotus 1-2-3. Data can be entered into the model and results displayed in less than a minute.

2. Guide to Operation.

a. Equipment Required.

(1) IBM PC compatible computer.

(2) 5.25" disk drive.

(3) Lotus 1-2-3 spreadsheet software.

b. Installation.

(1) Turn on your computer and activate Lotus 1-2-3.

(2) Insert the 3.5" disk containing the DLFAILB model into your computer's 3.5" disk drive.

(3) From the Lotus 1-2-3 menu, load the A:DLFAILB model (or B:DLFAILB if you're working from the b: drive) by entering /FR, then backspace to erase the default path, and enter A: and press the [Enter] key. After the Lotus 1-2-3 files are shown, select the DLFAILB file.

c. Operation.

- (1) The model parameters are shown below.

DUELS WITH WEAPON FAILURE RATES FOR BLUE	
REL OF BLUE RD*	1
PROB HIT BLUE RD*	0.6
PROB K H BLUE RD*	0.7
PSSK BLUE RD	0.42
REL OF RED RD*	1
PROB HIT RED RD*	0.8
PROB K H RED RD*	0.8
PSSK RED RD	0.64
ROF BLUE*	2
ROF RED*	1
NUM BLUE WPNS*	1
BLUE WPN FAIL RATE*	0.02 0.04
PROB BLUE WINS	0.552632

Screen #1

- (2) Move the cursor to the cell you wish to change. You may change only the items in Screen #1 which are marked with an asterisk.

d. Definitions.

REL OF BLUE RD: Reliability of the Blue round.

PROB HIT BLUE RD: Blue weapon's probability of hitting the Red target.

PROB K|H BLUE RD: Blue round's probability of killing the Red target given a hit.

PSSK BLUE RD: The product of the above three inputs.

REL OF RED RD: Reliability of the Red round.

PROB HIT RED RD: Red weapon's probability of hitting the Blue target.

PROB K|H RED RD: Red round's probability of killing the Blue target given a hit.

PSSK RED RD: The product of the above three inputs.

ROF BLUE: The Blue weapon's rate of fire.

ROF RED: The Red weapon's rate of fire.

NUM BLUE WPNS: The number of Blue weapons in the Blue force.

BLUE WPN FAIL RATE: Failure rate of the Blue weapon.

PROB BLUE WINS: The probability that Blue wins the duel.

PROB RED WINS: The probability that Red wins the duel.

e. Explanation.

(1) Definitions.

$P(B)$ = Probability that Blue wins the duel.

ρ_B = Mean rate of fire for a Blue weapon.

ρ_R = Mean rate of fire for a Red weapon.

p_B = Single shot kill probability of Blue against Red.

p_R = Single shot kill probability of Red against Blue.

μ_B = Mean failure rate for a Blue weapon.

μ_R = Mean failure rate for a Red weapon.

(2) Blue's and Red's weapon failure times are assumed to be exponentially distributed with mean failure times $1/\mu_B$ and $1/\mu_R$, respectively, or mean failure rates of μ_B and μ_R . If we further assume that Blue and Red have unlimited ammunition supplies, Blue has a limited number of weapons N , and Red has a failure-free weapon ($\mu_R = 0$), then the chances that Blue and Red win are

$$P(B) = \frac{\rho_B p_B}{\rho_B p_B + \rho_R p_R} \left[1 - \left(\frac{\mu_B}{\mu_B + \rho_B p_B + \rho_R p_R} \right)^N \right]$$

$$P(R) = 1 - P(B), \quad P(BR) = P(\text{Draw}) = 0$$

M. ESTIMATING OPERATIONAL AVAILABILITY

1. Introduction.

a. Description. A method for estimating operational availability based on a combination of test data and parameter estimates from other sources. This model was developed by Fred Bernstein, Eugene Dutoit, and Greg Meyers (Reference 5).

b. Applications. Desktop model for estimating operational availability. It also gives the reliability analyst the opportunity to determine the sensitivity of operational availability to changes in the parameters that contribute to this measure of readiness.

c. Setup. This model runs on a DOS-based computer. On-hand data can be entered into the model and results displayed in a few minutes.

2. Guide to Operation.

a. Equipment Required.

(1) IBM compatible PC computer.

(2) 3.5" disk drive.

(3) Lotus 1-2-3 spreadsheet software.

b. Installation.

(1) Turn on the computer and activate Lotus 1-2-3.

(2) Insert the 3.5" disk containing the OPERAO model into your 3.5" disk drive.

(3) From the Lotus 1-2-3 menu, load the A:OPERAO model (or B:OPERAO if you're working from the b: drive) by entering /FR, then backspace to erase the default path. Enter A: (or B:) and press the [Enter] key. After the Lotus 1-2-3 files are shown, select the OPERAO file.

c. Operation.

(1) The model parameters are shown below.

(2) Move the cursor to the cell you wish to change. You are allowed to change only the data marked by an asterisk--other data represents calculations made by the model.

ESTIMATING OPERATIONAL AVAILABILITY	
OT	10
TT	30
MR	0.3
K	1
ALDT	5
MTBOMF	100
	0.3
	0.05
	0.35
	0.333333
	0.116667
A _o	0.883333

Screen #1

d. Definitions for Screen #1.

OT: Operating Time.

TT: Total Time.

MR: Maintenance Ratio.

K: Ratio of Maintenance Manhours to Clock Hours.

ALDT: Administrative Logistic Downtime.

MTBOMF: Mean Time Between Operational Mission Failure.

e. Explanation:

(1) Definitions.

A_o Operational Availability.

ALDT Administrative Logistic Downtime.

DT	Downtime.
K	Ratio of Maintenance Manhours to Clock Hours.
MR	Maintenance Ratio.
MTBOMF	Mean Time Between Operational Mission Failure.
OT	Operating Time.
ST	Standby Time.
TALDT	Total Administrative Logistic Downtime.
TCM	Total Corrective Maintenance.
TPM	Total Preventive Maintenance.
TT	Total Time.

(2) The basic relationship that is used to estimate Operational Availability (A_o) is:

$$A_o = (OT + ST)/(OT + ST + TCM + TPM + TALDT) \quad (1)$$

(3) The entire denominator of equation (1) is Total Time (TT). The last three terms of the denominator account for all the downtime (DT). The numerator of this equation represents the total uptime (UT) for the system. An alternate way to express uptime is to subtract the DT from the TT. Equation (1) can then be written as:

$$A_o = (TT - DT)/TT = 1 - DT/TT \quad (2)$$

(4) Equation (2) can be expressed in terms of the "Downtime" components as:

$$A_o = 1 - (TCM + TPM + TALDT)/TT \quad (3)$$

(5) The Maintenance Ratio (MR) is the total number of man-hours of maintenance of direct labor in some particular time period divided by the total operating time in this same time period. This can be expressed as:

$$MR = K * (TCM + TPM)/OT \quad (4)$$

where K is the ratio of Maintenance Manhours to Maintenance Clock Hours. For example, if two maintenance men work from 12:00 noon to 5:00 PM (10 Maintenance Manhours during a 5 clock hour period of time) then $K = 10/5 = 2$. Equation (4) can also be written as:

$$TCM + TPM = (MR) * (OT)/K. \quad (5)$$

(6) TALDT can be estimated by considering the total number of failures in some given time period multiplied by the average logistical down time for each failure (ALDT). This relationship can be stated as:

$$TALDT = (OT) * (ALDT)/MTBOMF. \quad (6)$$

(7) Equations (5) and (6) can be substituted into equation (3). By factoring (OT) and (TT) as common terms, the following estimating relationship is obtained:

$$A_o = 1 - (OT/TT) * ((MR)/K + (ALDT/MTBOMF)) \quad (7)$$

(8) Equation (7) can be used to assess the A_o of a system based on a combination of test data and parameter estimates from other sources. The ratio of (OT/TT) can be obtained from the operational mode summary and mission profile for the system. The estimates for the MR and MTBOMF can be obtained from testing and engineering analysis. The values for ALDT and K can be estimated from additional logistical analysis, testing and field reports for existing but similar systems. Equation (7) also gives the reliability analyst the opportunity to determine the sensitivity of A_o to changes in the parameters that contribute to this measure of readiness. This can help determine which factors can be traded off against A_o and still have the system meet the operational requirement of readiness.

N. INDIRECT FIRE EFFECTS

1. Introduction.

a. Description. This model was developed by the Joint Munitions Effectiveness Manuals, Surface to Surface, and published under the authority of the Joint Technical Coordinating Group for Munitions Effectiveness. It calculates the effects of artillery and mortar fires for high explosive and improved conventional munitions. Inputs required are: number of volleys; number of rounds per volley; round reliability; lethal area; submunition reliability, volley pattern dimensions; target area dimensions; number of submunitions per round; angle of fall; mean point of impact and precision errors; target location error, and pattern adjustment factor. Results are displayed/printed in terms of fractional damage (amount of target destroyed) for the number of volleys used, or the number of volleys required to achieve a desired fractional damage.

b. Limitations. Effectiveness estimates for a large number of volleys may be unreliable due to the methodology used in this model.

c. Applications. Desktop analytic tool for determining artillery and mortar effects on personnel and materiel targets.

d. Setup. This model runs on an IBM compatible PC computer. Data is readily available from Joint Technical Coordinating Group publications or the Army Materiel Systems Analysis Activity at Aberdeen Proving Ground. Data can be entered and results displayed or printed in a few minutes.

2. Guide to Operation.

a. Equipment Required.

(1) IBM compatible PC computer.

(2) 3.5" disk drive.

(3) Printer (optional).

b. Installation.

(1) Turn on the computer and get to the DOS prompt.

(2) Insert the 3.5" disk containing the SUPERQUICKIE II program into your disk drive.

(3) From the DOS prompt, enter the command A: (or B: if you're operating from the B: disk drive).

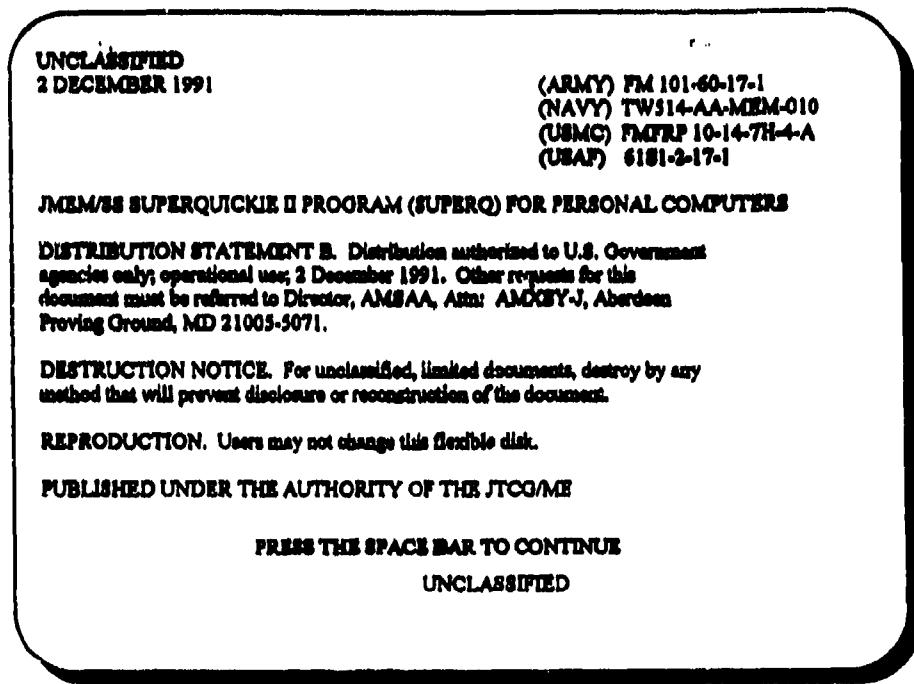
(4) Enter the command CD SQ. (to change to the SQ directory on the 3.5" disk)

(5) Enter the command SUPERQ

(6) If you receive the message, "Enter run time file path," it is probably because you are not in the SQ directory of the A: (or B:) drive. You cannot run this program simply by entering the name of the executable file and path (A:\SQ\SUPERQ.EXE).

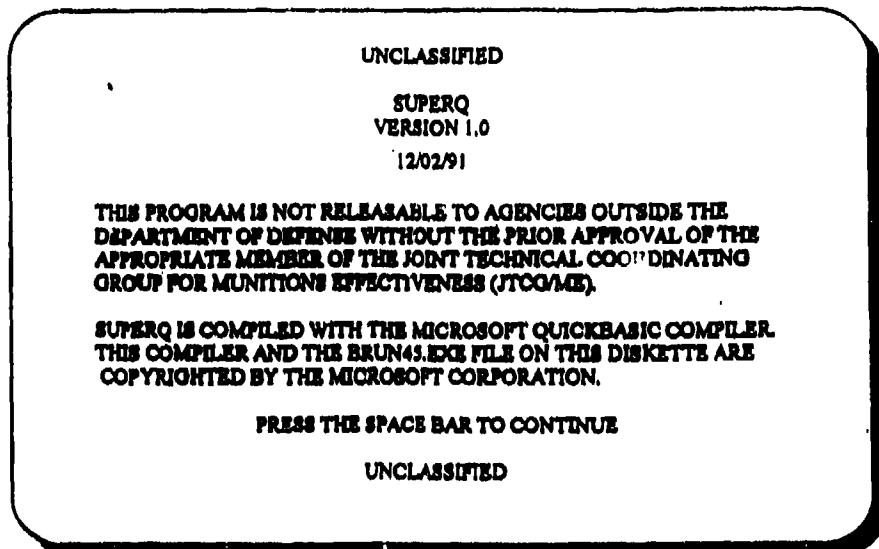
c. Operation.

(1) The first two screens contain publication, destruction, and copywrite information. Please take the time to read these screens.



Screen 1

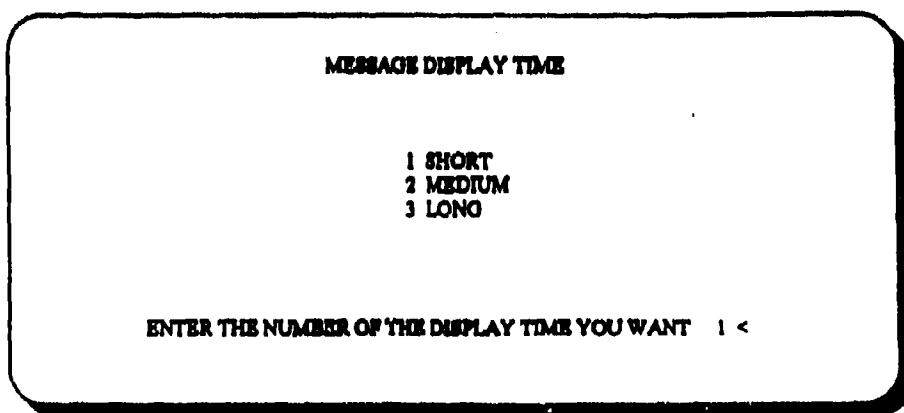
(2) Super Quickie II is for use by the Department of Defense only.



UNCLASSIFIED

Screen 2

(3) The next prompt asks you to select the amount of time you want messages displayed. "Short" is recommended.



Screen 3

(4) Next you will be asked if you have a color monitor.



Screen 4

(5) If you have a color monitor, you will be given the opportunity to change colors.

DO YOU WANT TO CHANGE THE COLORS - Y/N N <

Screen 5

(6) The next prompt warns you to make sure that the 3.5" disk's write-protect tab is disabled. It is disabled (will allow writing to it) if the write-protect tab is covering the small, rectangular hole on your disk. If you can see through the hole, slide the tab over the hole. Additionally, you are asked to enter the drive that has the Super Quickie II program on it. Do not enter a colon after the drive letter (do not enter A:, for example, just the letter A, B, or C.)

* NOTE *

DISKETTE MUST NOT HAVE A WRITE-PROTECT TAB INSTALLED

ENTER THE DRIVE YOU ARE WORKING ON - A, B, OR C A

Screen 6

(7) The following notices will be displayed next.

!! NOTE !!

UNITS OF MEASUREMENT MUST BE CONSISTENT

PRESS THE ESCAPE KEY AT ANY TIME TO EXIT PROGRAM

PRESS THE SPACE BAR TO CONTINUE

Screen 7

(8) The next display gives you the options you have with Super Quickie II. Basically, you can choose HE or ICM, and you can choose to input the number of volleys and have Super Quickie II determine the fractional damage to the target area, or you can input the fractional damage desired and have SuperQuickie II determine the number of volleys required.

OPTIONS

- 1 HE/FD - DETERMINES THE EFFECTIVENESS OF HE WEAPONS WHERE THE NUMBER OF VOLLEYS/SALVOS IS INPUT AND THE EXPECTED FRACTIONAL DAMAGE OR CASUALTIES IS OUTPUT.
- 2 ICM/FD - DETERMINES THE EFFECTIVENESS OF ICM'S WHERE THE NUMBER OF VOLLEYS OR SALVOS IS INPUT AND THE EXPECTED FRACTIONAL DAMAGE/CASUALTIES IS IS OUTPUT.
- 3 HE/NV - DETERMINES THE EFFECTIVENESS OF HE WEAPONS WHERE THE DESIRED FRACTIONAL DAMAGE/CASUALTIES IS INPUT AND THE REQUIRED NUMBER OF VOLLEYS/SALVOS IS OUTPUT.
- 4 ICM/NV - DETERMINES THE EFFECTIVENESS OF ICM'S WHERE THE DESIRED FRACTIONAL DAMAGE/CASUALTIES IS INPUT AND THE REQUIRED NUMBER OF VOLLEYS OR SALVOS IS OUTPUT.

ENTER THE NUMBER OF THE OPTION YOU WANT TO RUN 1

Screen 8

(9) The next set of displays request the inputs for the option you chose above. If you decide during the inputs that you have made an error on a previous entry, don't worry; you will get a chance to make corrections later--just continue with the remainder of the entries. Another point worth remembering: some entries will require an additional prompt at the bottom of your screen, and you could be frustrated if you don't notice it. The prompt may be waiting for a yes or no response and you'll be trying to enter a regular numerical input which won't be accepted.

(10) The following entries pertain to option #1, selected above in Screen 8. This option calls for a number of volleys of high explosive (HE) rounds, and will obtain a result in terms of the fraction of the target area destroyed. Fractional damage of a target area is a decimal number which equates to the fraction of the total number of personnel or materiel targets in the target area which were destroyed by the indirect fire. For example, if the lethal areas entered below are for personnel, a result of .23 means that 23% of the personnel in the target area were killed. It doesn't matter how many personnel are actually in the target area. Similarly for materiel targets. If the lethal areas entered are for tanks, then a result of .19 means that 19% of the tanks in the target area were destroyed.

(11) The first input is the number of rounds per volley. Press [Enter] after each input.

HE/FD	
NUMBER OF ROUNDS PER VOLLEY/SALVO	6
NUMBER OF UNIQUE VOLLEY/SALVO SETS (MAX = 5)	00000
ROUND RELIABILITY	00000
SUBMUNITION RELIABILITY	00000
VOLLEY/SALVO PATTERN LENGTH (RNG)	00000
VOLLEY/SALVO PATTERN WIDTH (DEFL)	00000
AREA TARGET LENGTH (RNG) OR RADIUS	00000
AREA TARGET WIDTH (DEFL)	00000
NUMBER OF SUBMUNITIONS PER ROUND	00000
ANGLE OF FALL, DEGREES	00000
SUBMUNITION RECTANGULAR PATTERN LENGTH (RNG) OR RADIUS	00000
SUBMUNITION RECTANGULAR PATTERN WIDTH (DEFL)	00000
MPI RANGE ERROR PROBABLE OR CEP	00000
MPI DEFLECTION ERROR PROBABLE	00000
PRECISION RANGE ERROR PROBABLE OR CEP	00000
PRECISION DEFLECTION ERROR PROBABLE	00000
TARGET LOCATION ERROR (CEP)	00000
PATTERN ADJUSTMENT FACTOR (K)	00000

Screen 9

(12) The second input is for the number of unique volley sets. For example, if you would like to obtain fractional damage results for firing 3 volleys and 12 volleys into the target area, then you have two unique volley sets--one set of 3 volleys and one set of 12 volleys. This model will automatically add an additional result for firing one volley. In this example we will enter two volley sets--3 volleys and 12 volleys--and the model will give us results for three volley sets: that is, results for 1 volley, 3 volleys, and 12 volleys. Note that there is a maximum of five unique volley sets. You can enter five volley sets, and the model will add the sixth result for one volley.

(13) The second input (the number of unique volley sets) is one of those inputs which will produce an almost inconspicuous prompt at the bottom of the screen. This prompt will ask you to enter the number of volleys you want fired for each unique volley set. In our case, we're going to enter two unique volley sets, and the prompt at the bottom of the screen will appear, asking us to enter the number of volleys for volley set 1, and then another prompt will appear in the same place, asking us to enter the number of volleys for volley set 2.

(14) The second entry (number of unique volley sets) and the first prompt at the bottom of the screen looks like Screen 10.

HE/FD	
NUMBER OF ROUNDS PER VOLLEY/SALVO	6
NUMBER OF UNIQUE VOLLEY/SALVO SETS (MAX = 5)	2
ROUND RELIABILITY	00000
SUBMUNITION RELIABILITY	00000
VOLLEY/SALVO PATTERN LENGTH (RNG)	00000
VOLLEY/SALVO PATTERN WIDTH (DEFL)	00000
AREA TARGET LENGTH (RNG) OR RADIUS	00000
AREA TARGET WIDTH (DEFL)	00000
NUMBER OF SUBMUNITIONS PER ROUND	00000
ANGLE OF FALL, DEGREES	00000
SUBMUNITION RECTANGULAR PATTERN LENGTH (RNG) OR RADIUS	00000
SUBMUNITION RECTANGULAR PATTERN WIDTH (DEFL)	00000
MPI RANGE ERROR PROBABLE OR CEP	00000
MPI DEFLECTION ERROR PROBABLE	00000
PRECISION RANGE ERROR PROBABLE OR CEP	00000
PRECISION DEFLECTION ERROR PROBABLE	00000
TARGET LOCATION ERROR (CEP)	00000
PATTERN ADJUSTMENT FACTOR (K)	00000
ENTER VOLLEY/SALVO SIZE NUMBER 1 = 3	

Screen 10

(15) The second prompt at the bottom of the screen will ask for the second volley size. In our example, 12 volleys will be entered for the size of the second volley set, as follows.

HE/FD	
NUMBER OF ROUNDS PER VOLLEY/SALVO	6
NUMBER OF UNIQUE VOLLEY/SALVO SETS (MAX = 5)	2
ROUND RELIABILITY	00000
SUBMUNITION RELIABILITY	00000
VOLLEY/SALVO PATTERN LENGTH (RNG)	00000
VOLLEY/SALVO PATTERN WIDTH (DEFL)	00000
AREA TARGET LENGTH (RNG) OR RADIUS	00000
AREA TARGET WIDTH (DEFL)	00000
NUMBER OF SUBMUNITIONS PER ROUND	00000
ANGLE OF FALL, DEGREES	00000
SUBMUNITION RECTANGULAR PATTERN LENGTH (RNG) OR RADIUS	00000
SUBMUNITION RECTANGULAR PATTERN WIDTH (DEFL)	00000
MPI RANGE ERROR PROBABLE OR CEP	00000
MPI DEFLECTION ERROR PROBABLE	00000
PRECISION RANGE ERROR PROBABLE OR CEP	00000
PRECISION DEFLECTION ERROR PROBABLE	00000
TARGET LOCATION ERROR (CEP)	00000
PATTERN ADJUSTMENT FACTOR (K)	00000
ENTER VOLLEY/SALVO SIZE NUMBER 2 = 12	

Screen 11

(16) The next input is for round reliability. Notice at the bottom of the list the volley sizes are now displayed, including 1 volley added by the model.

HE/FD	
NUMBER OF ROUNDS PER VOLLEY/SALVO	6
NUMBER OF UNIQUE VOLLEY/SALVO SETS (MAX = 5)	2
ROUND RELIABILITY	0.96
SUBMUNITION RELIABILITY	00000
VOLLEY/SALVO PATTERN LENGTH (RNG)	00000
VOLLEY/SALVO PATTERN WIDTH (DEFL)	00000
AREA TARGET LENGTH (RNG) OR RADIUS	00000
AREA TARGET WIDTH (DEFL)	00000
NUMBER OF SUBMUNITIONS PER ROUND	00000
ANGLE OF FALL, DEGREES	00000
SUBMUNITION RECTANGULAR PATTERN LENGTH (RNG) OR RADIUS	00000
SUBMUNITION RECTANGULAR PATTERN WIDTH (DEFL)	00000
MPI RANGE ERROR PROBABLE OR CEP	00000
MPI DEFLECTION ERROR PROBABLE	00000
PRECISION RANGE ERROR PROBABLE OR CEP	00000
PRECISION DEFLECTION ERROR PROBABLE	00000
TARGET LOCATION ERROR (CEP)	00000
PATTERN ADJUSTMENT FACTOR (K)	00000
VOLLEY/SALVO SIZES 1, 3, 12	

Screen 12

(17) The next item in the list, submunition reliability, will now display N/A in the right column because we selected the HE option. Submunition reliability is used for ICM only.

HE/FD	
NUMBER OF ROUNDS PER VOLLEY/SALVO	6
NUMBER OF UNIQUE VOLLEY/SALVO SETS (MAX = 5)	2
ROUND RELIABILITY	0.96
SUBMUNITION RELIABILITY	N/A
VOLLEY/SALVO PATTERN LENGTH (RNG)	00000
VOLLEY/SALVO PATTERN WIDTH (DEFL)	00000
AREA TARGET LENGTH (RNG) OR RADIUS	00000
AREA TARGET WIDTH (DEFL)	00000
NUMBER OF SUBMUNITIONS PER ROUND	00000
ANGLE OF FALL, DEGREES	00010
SUBMUNITION RECTANGULAR PATTERN LENGTH (RNG) OR RADIUS	00000
SUBMUNITION RECTANGULAR PATTERN WIDTH (DEFL)	00000
MPI RANGE ERROR PROBABLE OR CEP	00000
MPI DEFLECTION ERROR PROBABLE	00000
PRECISION RANGE ERROR PROBABLE OR CEP	00000
PRECISION DEFLECTION ERROR PROBABLE	00000
TARGET LOCATION ERROR (CEP)	00000
PATTERN ADJUSTMENT FACTOR (K)	00000
VOLLEY/SALVO SIZES 1, 3, 12	

Screen 13

- (18) The volley/salvo pattern length (in the range direction) and width (in deflection) are entered next. These are the dimensions of the volley pattern in the impact area.

HE/FD	
NUMBER OF ROUNDS PER VOLLEY/SALVO	6
NUMBER OF UNIQUE VOLLEY/SALVO SETS (MAX = 5)	2
ROUND RELIABILITY	0.96
SUBMUNITION RELIABILITY	N/A
VOLLEY/SALVO PATTERN LENGTH (RNG)	250
VOLLEY/SALVO PATTERN WIDTH (DEFL)	95
AREA TARGET LENGTH (RNG) OR RADIUS	00000
AREA TARGET WIDTH (DEFL)	00000
NUMBER OF SUBMUNITIONS PER ROUND	00000
ANGLE OF FALL, DEGREES	00000
SUBMUNITION RECTANGULAR PATTERN LENGTH (RNG) OR RADIUS	00000
SUBMUNITION RECTANGULAR PATTERN WIDTH (DEFL)	00000
MPI RANGE ERROR PROBABLE OR CEP	00000
MPI DEFLECTION ERROR PROBABLE	00000
PRECISION RANGE ERROR PROBABLE OR CEP	00000
PRECISION DEFLECTION ERROR PROBABLE	00000
TARGET LOCATION ERROR (CEP)	00000
PATTERN ADJUSTMENT FACTOR (K)	00000
VOLLEY/SALVO SIZES 1, 3, 12	

Screen 14

- (19) The next entry is for the area target length (range direction), or the radius of the target area. A prompt at the bottom of the screen asks you if you entered a radius or not.

HE/FD	
NUMBER OF ROUNDS PER VOLLEY/SALVO	6
NUMBER OF UNIQUE VOLLEY/SALVO SETS (MAX = 5)	2
ROUND RELIABILITY	0.96
SUBMUNITION RELIABILITY	N/A
VOLLEY/SALVO PATTERN LENGTH (RNG)	250
VOLLEY/SALVO PATTERN WIDTH (DEFL)	95
AREA TARGET LENGTH (RNG) OR RADIUS	100
AREA TARGET WIDTH (DEFL)	00000
NUMBER OF SUBMUNITIONS PER ROUND	00000
ANGLE OF FALL, DEGREES	00000
SUBMUNITION RECTANGULAR PATTERN LENGTH (RNG) OR RADIUS	00000
SUBMUNITION RECTANGULAR PATTERN WIDTH (DEFL)	00000
MPI RANGE ERROR PROBABLE OR CEP	00000
MPI DEFLECTION ERROR PROBABLE	00001
PRECISION RANGE ERROR PROBABLE OR CEP	00000
PRECISION DEFLECTION ERROR PROBABLE	00000
TARGET LOCATION ERROR (CEP)	00000
PATTERN ADJUSTMENT FACTOR (K)	00000
VOLLEY/SALVO SIZES 1, 3, 12	
DID YOU ENTER A RADIUS - Y/N N	

Screen 15

(20) The target width (deflection) is 100 meters for this example. As soon as you enter this number, the next line, for number of submunitions per round will be shown as N/A.

HE/FD	
NUMBER OF ROUNDS PER VOLLEY/SALVO	6
NUMBER OF UNIQUE VOLLEY/SALVO SETS (MAX = 5)	2
ROUND RELIABILITY	0.96
SUBMUNITION RELIABILITY	N/A
VOLLEY/SALVO PATTERN LENGTH (RNG)	250
VOLLEY/SALVO PATTERN WIDTH (DEFL)	95
AREA TARGET LENGTH (RNG) OR RADIUS	100
AREA TARGET WIDTH (DEFL)	100
NUMBER OF SUBMUNITIONS PER ROUND	N/A
ANGLE OF FALL, DEGREES	00000
SUBMUNITION RECTANGULAR PATTERN LENGTH (RNG) OR RADIUS	00000
SUBMUNITION RECTANGULAR PATTERN WIDTH (DEFL)	00000
MPI RANGE ERROR PROBABLE OR CEP	00000
MPI DEFLECTION ERROR PROBABLE	00000
PRECISION RANGE ERROR PROBABLE OR CEP	00000
PRECISION DEFLECTION ERROR PROBABLE	00000
TARGET LOCATION ERROR (CEP)	00000
PATTERN ADJUSTMENT FACTOR (K)	00000
VOLLEY/SALVO SIZES 1, 3, 12	

Screen 16

(21) The angle of fall is entered next, and the next two inputs will be shown as N/A.

HE/FD	
NUMBER OF ROUNDS PER VOLLEY/SALVO	6
NUMBER OF UNIQUE VOLLEY/SALVO SETS (MAX = 5)	2
ROUND RELIABILITY	0.96
SUBMUNITION RELIABILITY	N/A
VOLLEY/SALVO PATTERN LENGTH (RNG)	250
VOLLEY/SALVO PATTERN WIDTH (DEFL)	95
AREA TARGET LENGTH (RNG) OR RADIUS	100
AREA TARGET WIDTH (DEFL)	100
NUMBER OF SUBMUNITIONS PER ROUND	N/A
ANGLE OF FALL, DEGREES	47
SUBMUNITION RECTANGULAR PATTERN LENGTH (RNG) OR RADIUS	N/A
SUBMUNITION RECTANGULAR PATTERN WIDTH (DEFL)	N/A
MPI RANGE ERROR PROBABLE OR CEP	00000
MPI DEFLECTION ERROR PROBABLE	00000
PRECISION RANGE ERROR PROBABLE OR CEP	00000
PRECISION DEFLECTION ERROR PROBABLE	00000
TARGET LOCATION ERROR (CEP)	00000
PATTERN ADJUSTMENT FACTOR (K)	00000
VOLLEY/SALVO SIZES 1, 3, 12	

Screen 17

- (22) For the mean point of impact (MPI) errors, the first entry will cause a prompt at the bottom of the screen, asking you if you entered a circular error probable. In this case, yes.

HE/FD	
NUMBER OF ROUNDS PER VOLLEY/SALVO	6
NUMBER OF UNIQUE VOLLEY/SALVO SETS (MAX = 5)	2
ROUND RELIABILITY	0.96
SUBMUNITION RELIABILITY	N/A
VOLLEY/SALVO PATTERN LENGTH (RNG)	250
VOLLEY/SALVO PATTERN WIDTH (DEFL)	95
AREA TARGET LENGTH (RNG) OR RADIUS	100
AREA TARGET WIDTH (DEFL)	100
NUMBER OF SUBMUNITIONS PER ROUND	N/A
ANGLE OF FALL, DEGREES	47
SUBMUNITION RECTANGULAR PATTERN LENGTH (RNG) OR RADIUS	N/A
SUBMUNITION RECTANGULAR PATTERN WIDTH (DEFL)	N/A
MPI RANGE ERROR PROBABLE OR CEP	40
MPI DEFLECTION ERROR PROBABLE	00000
PRECISION RANGE ERROR PROBABLE OR CEP	00000
PRECISION DEFLECTION ERROR PROBABLE	00000
TARGET LOCATION ERROR (CEP)	00000
PATTERN ADJUSTMENT FACTOR (K)	00000
VOLLEY/SALVO SIZES 1, 3, 12	
DID YOU ENTER A CEP - Y/N Y	

Screen 18

- (23) Similarly for the precision errors.

HE/FD	
NUMBER OF ROUNDS PER VOLLEY/SALVO	6
NUMBER OF UNIQUE VOLLEY/SALVO SETS (MAX = 5)	2
ROUND RELIABILITY	0.96
SUBMUNITION RELIABILITY	N/A
VOLLEY/SALVO PATTERN LENGTH (RNG)	250
VOLLEY/SALVO PATTERN WIDTH (DEFL)	95
AREA TARGET LENGTH (RNG) OR RADIUS	100
AREA TARGET WIDTH (DEFL)	100
NUMBER OF SUBMUNITIONS PER ROUND	N/A
ANGLE OF FALL, DEGREES	47
SUBMUNITION RECTANGULAR PATTERN LENGTH (RNG) OR RADIUS	N/A
SUBMUNITION RECTANGULAR PATTERN WIDTH (DEFL)	N/A
MPI RANGE ERROR PROBABLE OR CEP	40
MPI DEFLECTION ERROR PROBABLE	00000
PRECISION RANGE ERROR PROBABLE OR CEP	42
PRECISION DEFLECTION ERROR PROBABLE	00000
TARGET LOCATION ERROR (CEP)	00000
PATTERN ADJUSTMENT FACTOR (K)	00000
VOLLEY/SALVO SIZES 1, 3, 12	
DID YOU ENTER A CEP - Y/N Y	

Screen 19

(24) The target location error is entered as 0 meters.

HE/FD	
NUMBER OF ROUNDS PER VOLLEY/SALVO	6
NUMBER OF UNIQUE VOLLEY/SALVO SETS (MAX = 5)	2
ROUND RELIABILITY	0.96
SUBMUNITION RELIABILITY	N/A
VOLLEY/SALVO PATTERN LENGTH (RNG)	250
VOLLEY/SALVO PATTERN WIDTH (DEFL)	95
AREA TARGET LENGTH (RNG) OR RADIUS	100
AREA TARGET WIDTH (DEFL)	100
NUMBER OF SUBMUNITIONS PER ROUND	N/A
ANGLE OF FALL, DEGREES	47
SUBMUNITION RECTANGULAR PATTERN LENGTH (RNG) OR RADIUS	N/A
SUBMUNITION RECTANGULAR PATTERN WIDTH (DEFL)	N/A
MPI RANGE ERROR PROBABLE OR CEP	40
MPI DEFLECTION ERROR PROBABLE	00000
PRECISION RANGE ERROR PROBABLE OR CEP	42
PRECISION DEFLECTION ERROR PROBABLE	00000
TARGET LOCATION ERROR (CEP)	0
PATTERN ADJUSTMENT FACTOR (K)	00000
VOLLEY/SALVO SIZES 1, 3, 12	

Screen 20

(25) The pattern adjustment factors may be obtained from the JTACG. In our example the pattern adjustment factor is 4.

HE/FD	
NUMBER OF ROUNDS PER VOLLEY/SALVO	6
NUMBER OF UNIQUE VOLLEY/SALVO SETS (MAX = 5)	2
ROUND RELIABILITY	0.96
SUBMUNITION RELIABILITY	N/A
VOLLEY/SALVO PATTERN LENGTH (RNG)	250
VOLLEY/SALVO PATTERN WIDTH (DEFL)	95
AREA TARGET LENGTH (RNG) OR RADIUS	100
AREA TARGET WIDTH (DEFL)	100
NUMBER OF SUBMUNITIONS PER ROUND	N/A
ANGLE OF FALL, DEGREES	47
SUBMUNITION RECTANGULAR PATTERN LENGTH (RNG) OR RADIUS	N/A
SUBMUNITION RECTANGULAR PATTERN WIDTH (DEFL)	N/A
MPI RANGE ERROR PROBABLE OR CEP	40
MPI DEFLECTION ERROR PROBABLE	00000
PRECISION RANGE ERROR PROBABLE OR CEP	42
PRECISION DEFLECTION ERROR PROBABLE	00000
TARGET LOCATION ERROR (CEP)	0
PATTERN ADJUSTMENT FACTOR (K)	00000
VOLLEY/SALVO SIZES 1, 3, 12	

Screen 21

(26) Having filled the first set of prompts, the entries for lethal areas are next. After you enter the lethal areas, you will get a prompt at the bottom of the screen asking if you entered personnel lethal areas rather than lethal areas for materiel targets.

LETHAL AREAS	
LETHAL AREAS MUST BE ENTERED IN DECREASING ORDER	
LETHAL AREA OF POSTURE 1 OR MATERIEL TOT 1	530.0
LETHAL AREA OF POSTURE 2 OR MATERIEL TOT 2	400.0
LETHAL AREA OF POSTURE 3 OR MATERIEL TOT 3	150.0
DID YOU ENTER PERSONNEL LETHAL AREAS - Y/N Y	

Screen 22

(27) The last set of entries pertain to the percent of personnel in each of the above postures during the first and subsequent volleys. For example, in the above screen you might have entered three lethal areas for personnel who are standing, crouching, and prone. The scenario might dictate that when the first volley lands, 80% of the personnel in the target area are standing, 10% are crouching, and 10% are in a prone position. However, for subsequent volleys, 10% of the personnel are still standing (running, moving to a new position, etc...), 50% are crouching, and 40% are in a prone position. Entries to match this scenario are as follows.

LETHAL AREAS	
LETHAL AREAS MUST BE ENTERED IN DECREASING ORDER	
LETHAL AREA OF POSTURE 1 OR MATERIEL TOT 1	530.0
LETHAL AREA OF POSTURE 2 OR MATERIEL TOT 2	400.0
LETHAL AREA OF POSTURE 3 OR MATERIEL TOT 3	150.0
FOR POSTURE SEQUENCING, THE FRACTION OF PERSONNEL IN EACH POSTURE MUST LIE BETWEEN 0.0 AND 1.0. THE SUM OF ALL THE POSTURES MUST EQUAL 1.0	
FRACTION OF PERSONNEL POSTURE 1 DURING FIRST VOL/SAL	0.80
FRACTION OF PERSONNEL POSTURE 2 DURING FIRST VOL/SAL	0.10
FRACTION OF PERSONNEL POSTURE 3 DURING FIRST VOL/SAL	0.10
FRACTION OF PERSONNEL POSTURE 1 FOR SUBSEQUENT VOL/SAL	0.10
FRACTION OF PERSONNEL POSTURE 2 FOR SUBSEQUENT VOL/SAL	0.50
FRACTION OF PERSONNEL POSTURE 3 FOR SUBSEQUENT VOL/SAL	0.40

Screen 23

(28) All inputs will now be displayed, and you are given the opportunity to make changes in the data prior to calculating the results.

INPUTS			
1 NUMBER OF RDS PER VOL/SAL	6	15 PRECISION CEP	42.00
2 NUM OF UNIQUE VOL/SAL SETS	2	16 PRECISION DEP	N/A
3 ROUND RELIABILITY	.96	17 TARGET LOCATION ERROR	0.00
4 SUBMUNITION RELIABILITY	N/A	18 PATTERN ADJ FACTOR (K)	4
5 VOL/SAL PATTERN (RNG)	250.00	19 POSTURE 1 LETHAL AREA	530.00
6 VOL/SAL PATTERN (DEFL)	95.00	20 POSTURE 2 LETHAL AREA	400.00
7 TARGET LENGTH (RNG)	100.00	21 POSTURE 3 LETHAL AREA	150.00
8 TARGET WIDTH (DEFL)	100.00	22 POSTURE 1 FIRST VOL/SAL	0.80
9 NUM OF SUBMUNITIONS PER RD	N/A	23 POSTURE 2 FIRST VOL/SAL	0.10
10 ANGLE OF FALL	47.00	24 POSTURE 3 FIRST VOL/SAL	0.10
11 SUBMUNITION PATTERN (RNG)	N/A	25 POSTURE 1 AFTER FIRST VOL/SAL	0.10
12 SUBMUNITION PATTERN (DEFL)	N/A	26 POSTURE 2 AFTER FIRST VOL/SAL	0.50
13 MPI CEP	40.00	27 POSTURE 3 AFTER FIRST VOL/SAL	0.40
14 MPI DEP	N/A	28 VOL/SAL SIZES 1 3 12	

DO YOU WANT TO MAKE A CHANGE - Y/N N

Screen 24

(29) If you had wanted to make a change in the inputs, you would have responded with a "Y" to the prompt at the bottom of Screen 24. However, assuming that no changes need to be made, a response of "N", as in Screen 24, will produce the desired results in terms of fractional damage.

RESULTS				
EXPECTED FRACTIONAL DAMAGE/CASUALTIES				
VOLLEY/SALVO	POS 1	POS 2	POS 3	MIXED
1	0.0362	0.0281	0.0113	0.0329
3	0.1047	0.0819	0.0335	0.0752
12	0.3570	0.2891	0.1275	0.2401

NOTE: 6 ROUND(S) PER VOLLEY/SALVO

PRESS P TO PRINT OR PRESS THE SPACEBAR TO CONTINUE

Screen 25

(30) Whether you press "P" to print the results shown on the screen, or press the spacebar to continue, the following screen is displayed.

- 1 ENTER ALL NEW INPUT (RETURN TO THE OPTIONS MENU)
- 2 CHANGE EXISTING INPUT
- 3 EXIT TO SYSTEM

ENTER THE NUMBER OF THE OPTION YOU WANT 3

Screen 26

(31) Entering a 3, above, will return you to the a:> prompt after the next screen.

UNCLASSIFIED

RECOMMENDED CHANGES, COMMENTS OR CORRECTIONS TO
IMPROVE THIS PROGRAM SHOULD BE ADDRESSED TO:

DIRECTOR
U.S. ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY
ATTN: AMXSY - J
ABERDEEN PROVING GROUND, MD 21005-5071

PRESS THE SPACE BAR TO RETURN TO SYSTEM

UNCLASSIFIED

Screen 27

(32) Upon pressing the space bar, you will be returned to the A:\\$Q> prompt. If you need to return to the C: drive and prompt, simply type C: and press the [Enter] key.

References

1. Hergert, Douglas. *Microsoft QuickBASIC*. 2nd Ed. Redmond, WA: Microsoft Press, 1988.
2. Johnson, Lee W., and R. Dean Riess. *Numerical Analysis*. 2nd Ed. Reading, MA: Addison-Wesley, 1982.
3. Mendenhall, William, Richard L. Scheaffer, and Dennis D. Wackerly. *Mathematical Statistics with Applications*. 2nd Ed. Boston: Duxbury Press, 1981.
4. US Army Materiel Development and Readiness Command. *Engineering Design Handbook: Army Weapon Systems Analysis, Part One*. DARCOM Pamphlet 706-101, November 1977.
5. Berstein, Fred, Eugene F. Dutoit, and Greg Meyers. *The Affect of Administrative Logistic Downtime on the Operational Availability of Weapon Systems*. Orange Paper-Report No. RWG-JMEM/SS-1, Reliability Working Group Joint Munitions Effectiveness Manual/Surface to Surface, Aberdeen Proving Ground, MD: 31 October 1992.
6. Saaty, Thomas L. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York: McGraw-Hill, 1980.
7. U.S. Army. *Tank AntiTank Armor Weapons System (TATAWS) Study*. 1968.
8. Marsaglia, G., and T. A. Bray. *A Convenient Method for Generating Normal Variables*. SIAM Review, Vol. 6, No. 3, pp. 260-264, 1964.

An Application of Generalized p-Values
in Tank Gun Accuracy Research

David W. Webb

US Army Research Laboratory
Weapons Technology Directorate
Aberdeen Proving Ground, MD 21005

By optimally rotating a tank cannon to counteract gravity droop and the cannon's dynamic response during firing, the idea of "dynamic indexing" was believed to be a major step in the reduction of between-tube variability, σ^2_r . Using an indirect approach to compare the between-tube variance components for dynamically indexed tubes (DIT's) and standard tubes (ST's), an earlier analysis of the test data failed to show a difference between σ^2_{r-DIT} and σ^2_{r-ST} . Seeking a more direct comparison of independently obtained between-tube variance components, Zhou and Mathews proposed a test variable based on the recently developed concept of generalized p-values. This paper describes how this generalized test variable is employed to compare two between-tube variance components taken from independent mixed models. Finally, a comparison is made between the conclusions drawn from the original analysis and a reanalysis of the field test using Zhou and Mathews' generalized p-value approach.

Introduction

U.S. Army experiments conducted in the late 1980's showed that between-tube variability is a significant contributor to the overall error in the M1A1 series tank. In an attempt to reduce this variability, researchers took advantage of the fact that each gun tube has its own unique curvature by proposing that gun tubes be dynamically indexed (Schmidt, et. al., 1988). That is, each gun tube is rotated about the center boreline so that its curvature counteracts both the gravitational droop and the whipping motion of the tube immediately after trigger pull. This whipping motion (more properly referred to as the dynamic response) is caused by a vertical difference in the centers-of-gravity of the gun tube and

the breech block which supports the gun tube. Normally, a gun tube is only rotated to lessen the effects of the gravitational droop. This is known as standard indexing.

In 1990, the U.S. Army Ballistic Research Laboratory (now part of the U.S. Army Research Laboratory) conducted a large-scale field test whose primary purpose was to determine if dynamic indexing would reduce the between-tube variability of the M1A1 series tank (Webb, et. al., 1991). This costly experiment included four types of ammunition, four tanks, twenty standard tubes (ST's), and twenty dynamically indexed tubes (DIT's). The response recorded from each round was its horizontal and vertical jump, where jump is defined as the distance from the aimpoint to the impact point after all known corrections (such as wind and muzzle velocity) have been applied.

A separate and independent analysis of jump was performed for all eight (2×4) combinations of direction and ammunition type. Table 1 shows an arrangement of the data collected for each subset of the entire test. In this table, we see that the fixed factor Tube Type and the random factor Tank were crossed, while the random factor Tube was nested within Tube Type. Three rounds were fired per cell.

To obtain an estimate of between-tube variability for both dynamically indexed and standard tubes, two independent mixed linear models were applied to Table 1 (one for each tube type). For each type of tube, the linear model is:

$$z_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{k(ij)},$$

where

- 1) z_{ijk} is the jump of the k^{th} round from the j^{th} tube on the i^{th} tank, measured in mils;
- 2) μ is the overall mean;
- 3) α_i is the effect of the i^{th} tank for $i = 1, 2, 3, 4$;
- 4) $\beta_{j(i)}$ is the effect of the j^{th} tube on the i^{th} tank for $j = 1, 2, 3, 4, 5$;

Tube Type	Tank	Tube ID	Jump
1	1	1	z z z
		2	z z z
		3	z z z
		4	z z z
		5	z z z
	2	6	z z z
		7	z z z
		8	z z z
		9	z z z
		10	z z z
	3	11	z z z
		12	z z z
		13	z z z
		14	z z z
		15	z z z
	4	16	z z z
		17	z z z
		18	z z z
		19	z z z
		20	z z z
2	1	21	z z z
		22	z z z
		23	z z z
		24	z z z
		25	z z z
	2	26	z z z
		27	z z z
		28	z z z
		29	z z z
		30	z z z
	3	31	z z z
		32	z z z
		33	z z z
		34	z z z
		35	z z z
	4	36	z z z
		37	z z z
		38	z z z
		39	z z z
		40	z z z

Table 1. Data matrix for each combination of direction and ammunition type. Each "z" represents a jump value.

- 5) $\beta_{j(i)} \sim N(0, \sigma^2_T);$
- 6) $\epsilon_{k(ij)}$ is the error associated with the k^{th} round from the j^{th} tube on the i^{th} tank for $k = 1, 2, 3$, and;
- 7) $\epsilon_{k(ij)} \sim N(0, \sigma^2_R).$

Comparison of Between-Tube Variabilities: An Indirect Approach

The goal of the statistical analysis was to conduct a one-sided hypothesis test comparing the between-tube variabilities, namely,

$$H_0: \sigma^2_{T-ST} \leq \sigma^2_{T-DIT} \quad \text{vs.} \quad H_a: \sigma^2_{T-ST} > \sigma^2_{T-DIT},$$

or, equivalently,

$$H_0: \frac{\sigma^2_{T-ST}}{\sigma^2_{T-DIT}} \leq 1 \quad \text{vs.} \quad H_a: \frac{\sigma^2_{T-ST}}{\sigma^2_{T-DIT}} > 1.$$

Superficially, the ratio

$$F^* = \frac{MS_{T-ST}}{MS_{T-DIT}} = \frac{SS_{T-ST}}{SS_{T-DIT}},$$

where F^* follows an F distribution with 16 numerator and denominator degrees of freedom, may appear to be a proper test statistic for H_0 . However, examination of the expected mean squares for each model shows that F^* is actually a test statistic for

$$H_0^*: \frac{\sigma^2_{T-ST} + 3\sigma^2_{R-ST}}{\sigma^2_{T-DIT} + 3\sigma^2_{R-DIT}} \leq 1 \quad \text{vs.} \quad H_a^*: \frac{\sigma^2_{T-ST} + 3\sigma^2_{R-ST}}{\sigma^2_{T-DIT} + 3\sigma^2_{R-DIT}} > 1.$$

Under the assumption that σ^2_{R-DIT} and σ^2_{R-ST} are equal, F^* serves as an indirect test statistic for H_0 , since significantly large values of F^* would be attributable to differences in the between-tube variabilities and not the between-round variabilities.

The assumption of equal between-round variabilities can be tested by the statistic

$$F^t = \frac{MS_{R-ST}}{MS_{R-DIT}} = \frac{SS_{R-ST}}{SS_{R-DIT}} ,$$

where F^t follows an F distribution with 40 degrees of freedom in both the numerator and denominator. If this assumption is not rejected, then one may proceed with the computation of F^* to test H_0 .

On the other hand if the assumption of equal between-tube variance is rejected, then F^* is more prone to Type I or Type II errors. These errors are due to the presence of the nuisance parameters, σ^2_{R-DIT} and σ^2_{R-ST} , in the expected value of the test statistic. How should the analyst proceed if this is the case?

Comparison of Between-Tube Variabilities: The Generalized p-value Approach

As described by Tsui and Weerahandi (1989), classical one-sided hypothesis tests of the form $H_0: \theta \leq \theta_0$ versus $H_1: \theta > \theta_0$, utilize a test statistic $T(X)$ that is simply a function of the sample space, X . For the observed response, x , the critical region, C_x , is defined as

$$C_x = \{X: T(X) \geq T(x)\} .$$

The p-value associated with the hypothesis test is then given as

$$p = \sup_{\theta \in H_0} Pr(X \in C_x | \theta) .$$

However, if this probability is dependent upon some nuisance parameter, η , then the p-value may not be calculable. This is exactly the problem that exists with the dynamically indexed tube experiment.

Tsui and Weerahandi proposed the idea of a generalized p-value (GPV) for one-sided hypothesis tests when nuisance parameters are present. In lieu of a test statistic, a generalized test variable is used, which is not only a function of the sample space, but also the sample data and the parameters. The generalized test variable,

$T(X; x, \theta, \eta)$, is chosen so that for all fixed values of x , the following conditions hold:

- 1) $t = T(x; x, \theta_0, \eta)$ is free of η ;
- 2) the distribution of $T(X; x, \theta_0, \eta)$ is free of η ; and
- 3) for fixed η , $\Pr(T(X; x, \theta, \eta) \geq t)$ is nondecreasing in θ ;

In addition, the critical region is replaced by the generalized extreme region, $C_x(\theta, \eta)$, whose domain includes the nuisance parameter, and is defined to be:

$$C_x(\theta, \eta) = \{X: T(X; x, \theta, \eta) \geq T(x; x, \theta, \eta)\}.$$

Finally, the GPV is given as:

$$\begin{aligned} p &= \Pr(X \in C_x(\theta, \eta) | \theta = \theta_0) \\ &= \Pr(T(X; x, \theta_0, \eta) \geq t). \end{aligned}$$

With the above definitions, Tsui and Weerahandi showed that the GPV is independent of the nuisance parameters and can hence be used as evidence against the null hypothesis.

The construction of generalized test variables is not a trivial task and unfortunately little guidance is given in the few papers that have been published on this topic. Zhou and Mathew (1993) derived a generalized test variable that is used to compare variance components obtained from two independent mixed hierachial models. This methodology was directly applied to the between-tube variability comparison. Their generalized test variable is given by:

$$T(X; x, \theta, \eta) = T(X; x, \sigma_{R-DIT}^2, \sigma_{T-ST}^2, \sigma_{R-DIT}^2, \sigma_{T-ST}^2)$$

$$= \frac{\begin{pmatrix} \sigma_{R-DIT}^2 + 3\sigma_{T-ST}^2 \\ \sigma_{T-ST}^2 & \sigma_{T-DIT}^2 \\ \hline \sigma_{R-DIT}^2 + 3 & \sigma_{T-DIT}^2 \\ \sigma_{T-DIT}^2 & \end{pmatrix} (\sigma_{R-DIT}^2 + 3\sigma_{T-DIT}^2) \frac{SS_{T-DIT}}{SS_{T-DIT}} + \sigma_{R-ST}^2 \frac{SS_{R-ST}}{SS_{R-ST}}}{(\sigma_{R-ST}^2 + 3\sigma_{T-ST}^2) \frac{SS_{T-ST}}{SS_{T-ST}} + \sigma_{R-DIT}^2 \frac{SS_{R-DIT}}{SS_{R-DIT}}},$$

where each SS term is the random variable for the appropriate sums-of-squares and each ss term is the realized value of SS . These ss values are taken directly from standard analyses of variance of the field test data.

Although this generalized test variable appears to be very cumbersome, the calculation of a GFV is straightforward. Under H_0 ,

$$\begin{aligned}
 T(X; x, \theta_0, \eta) &= \frac{\left(\begin{array}{c} \sigma_{R-DIT}^2 + 3(1) \\ \sigma_{T-DIT}^2 \\ \hline \sigma_{R-DIT}^2 + 3 \\ \sigma_{T-DIT}^2 \end{array} \right) (\sigma_{R-DIT}^2 + 3\sigma_{T-DIT}^2) \frac{SS_{T-DIT}}{SS_{T-DIT}} + \sigma_{R-ST}^2 \frac{SS_{R-ST}}{SS_{R-ST}}}{(\sigma_{R-ST}^2 + 3\sigma_{T-ST}^2) \frac{SS_{T-ST}}{SS_{T-ST}} + \sigma_{R-DIT}^2 \frac{SS_{R-DIT}}{SS_{R-DIT}}} \\
 &= \frac{(\sigma_{R-DIT}^2 + 3\sigma_{T-DIT}^2) \frac{SS_{T-DIT}}{SS_{T-DIT}} + \sigma_{R-ST}^2 \frac{SS_{R-ST}}{SS_{R-ST}}}{(\sigma_{R-ST}^2 + 3\sigma_{T-ST}^2) \frac{SS_{T-ST}}{SS_{T-ST}} + \sigma_{R-DIT}^2 \frac{SS_{R-DIT}}{SS_{R-DIT}}} \\
 &= \frac{k_1}{\Phi_1} + \frac{k_2}{\Phi_2}, \\
 &\quad \frac{k_3}{\Phi_3} + \frac{k_4}{\Phi_4},
 \end{aligned}$$

where each k_i is an observed sum-of-squares (a constant) and each Φ_i is a chi-square random variable.

Furthermore, if $X = x$, then $SS_{T-DIT} = ss_{T-DIT}$, $SS_{R-DIT} = ss_{R-DIT}$, $SS_{T-ST} = ss_{T-ST}$, and $SS_{R-DIT} = ss_{R-DIT}$, so that,

$$\begin{aligned}
 t &= T(x; x, \theta_0, \eta) \\
 &= \frac{(\sigma_{R-DIT}^2 + 3\sigma_{T-DIT}^2)(1) + \sigma_{R-ST}^2(1)}{(\sigma_{R-ST}^2 + 3\sigma_{T-ST}^2)(1) + \sigma_{R-DIT}^2(1)} \\
 &= \frac{\sigma_{R-DIT}^2 + 3\sigma_{T-DIT}^2 + \sigma_{R-ST}^2}{\sigma_{R-ST}^2 + 3\sigma_{T-ST}^2 + \sigma_{R-DIT}^2} \\
 &= 1, \quad (\text{since } \sigma_{T-DIT}^2 = \sigma_{T-ST}^2).
 \end{aligned}$$

Finally, the expression for the GPV simplifies to

$$P = \Pr \left(\frac{\frac{k_1}{\Phi_1} + \frac{k_2}{\Phi_2}}{\frac{k_3}{\Phi_3} + \frac{k_4}{\Phi_4}} \geq 1 \right);$$

that is, the GPV is the relative frequency with which this function of chi-square random variables exceeds unity.

For known values of the sums-of-squares, this probability can be determined by simulation. A FORTRAN program simulated 50,000 values of the generalized test variable and counted the number of times that it exceeded unity to obtain the GPV.

Results

Due to security classification restrictions, the sums of squares derived from the data cannot be divulged in this report. However, p-values from both the indirect and GPV approach hypothesis tests for all combinations of direction and ammunition type are presented in Table 2.

Direction	Ammunition	Indirect Test Approach	Generalized Test Approach
Azimuth	A	.993 (.433)	.989
	B	.858 (.111)	.676
	C	.017 (.016)	.036
	D	.149 (.373)	.122
Elevation	A	.981 (.760)	.972
	B	.779 (.745)	.759
	C	.253 (.560)	.226
	D	.439 (.873)	.453

(p-values for the test of $H_0: \sigma^2_{R-ST} = \sigma^2_{R-DIT}$ are in parentheses)

Table 2. P-values for the tests of $H_0: \sigma^2_{T-ST} \leq \sigma^2_{T-DIT}$.

The two columns of p-values for $H_0: \sigma_{T-ST}^2 \leq \sigma_{T-DIT}^2$ are quite similar. This indicates that both analysis approaches arrive at the same basic conclusion, namely that dynamic indexing fails to consistently reduce between-tube variability. Only in one case out of eight (Ammunition C in Azimuth), was σ_{T-DIT}^2 determined to be significantly lower than σ_{T-ST}^2 .

It is also interesting to note a few differences in the two columns of p-values for $H_0: \sigma_{T-ST}^2 \leq \sigma_{T-DIT}^2$. For Ammunition C in azimuth, the GPV is more than double that of the p-value obtained via the indirect approach. Also, for Ammunition B in azimuth, the absolute difference in p-values is rather large. These differences may be due to the unequal between-round variabilities associated with these data sets (note the low parenthesized p-values in Table 2). Recall that the indirect approach requires that the between-round variabilities are equal, whereas the GPV approach does not require this assumption and is therefore an exact hypothesis testing procedure. Violation of this assumption may result in unreliable p-values reported via the indirect approach.

Summary

For this particular data set, both procedures arrived at the same conclusions to the dismay of the engineers behind the dynamic indexing concept. Some minor differences in the p-values highlighted potential problems in using the indirect approach to test $H_0: \sigma_{T-ST}^2 \leq \sigma_{T-DIT}^2$.

The procedures for testing independent between-tube variabilities presented in this paper each have their particular advantages and drawbacks. The indirect approach is simple to apply, as it requires only the use of F ratios based on sums-of-squares taken directly from independent analyses of variance. However, this approach relies on assumptions made about the nuisance parameters, σ_{R-ST}^2 and σ_{R-DIT}^2 . Failure to meet the assumptions may increase either the Type I or Type II error probabilities.

The GPV approach is independent of the nuisance parameters, and is therefore an exact test for the null hypothesis. The main disadvantage to this approach is that there is little guidance in the statistical literature on the derivation of a proper

generalized test variable. Furthermore, computation of the GPV requires computer simulation of the generalized test variable. If the analyst can obtain a proper generalized test variable, the exactness of the GPV approach makes it the more desirable of the two analytical strategies.

References

- Schmidt, E., J. Bornstein, J. Thomas and T. McCloskey (1988), "A Method for Indexing Tank Cannon", BRL-IMR-912.
- Tsui, K.W. and S Weerahandi (1989), "Generalized p-Values in Significance Testing of Hypotheses in the Presence of Nuisance Parameters", *Journal of the American Statistical Association*, 84, 602-607.
- Webb, D. W., J. Thomas, and R. Carter (1991), "M1A1 Tank Dynamic Indexed Tube Experiment: Analysis and Results", U.S. Army Ballistic Research Laboratory, BRL-MR-3962.
- Zhou, L. and T. Mathew (1993), "Some Tests for Variance Components Using Generalized p-Values", unpublished manuscript.

IDENTIFYING THE CRITICAL FACTORS IN AN ADAPTIVE NETWORK

Ann E. M. Brodeen, Barbara D. Broome, George W. Hartwig, Jr., and Maria C. Lopez
Advanced Computational and Information Sciences Directorate
Aberdeen Proving Ground, Maryland 21005-5067

ABSTRACT

In the ideal communications network each node would be smart enough to monitor network performance and, when necessary, adapt itself to better accommodate its workload. The adaptive network node would employ a decision algorithm to modify configuration, routing and protocol parameters based on measured network performance and system requirements. This paper describes continuing research into feasible approaches to developing an adaptive network for use in battlefield command and control systems. The initial approach entails the collection of message traffic information into a deductive database from which network performance is assessed and compared to system requirements. Inadequate performance would trigger identification and assessment of alternatives for improvement. The project emphasizes use of actual hardware and controlled experiments to explore alternatives for parameter settings. This paper describes an initial attempt to identify baseline performance data for a prototype communications network and to determine those factors to which the system is most sensitive.

BACKGROUND

Decentralized battlefield command and control requires reliable and timely distribution of information. At present, information distribution is limited by noisy channels and protocols that do not meet traffic demands, forcing commanders to make decisions from out of date or incomplete information. To solve this problem, our research addresses control of noise and interference on communication channels and construction of network protocols that will be effective on the modern battlefield.

Currently the civilian sector is experiencing a communications revolution; however, civilian applications often assume a physical infrastructure, such as towers and high power base stations, that is not always feasible in a military environment. Our research takes into account the special problems of the battlefield: mobility, bandlimited channels, arbitrary or intentional interference, multimedia data, and rapid pace of operations. The networks that are of particular interest to the Army have nodes with high computing power but weak, noisy, shared communication links. For this reason, our approach to communication emphasizes working intelligently at each node to limit or redirect the amount of information that must be passed along the communication channel. Each node is assumed to act independently to improve the effectiveness of the information exchange between nodes. Such a system of controls requires that each node be able to: monitor the network traffic; decide whether performance is inadequate; and if so, make an appropriate adjustment to the protocol.

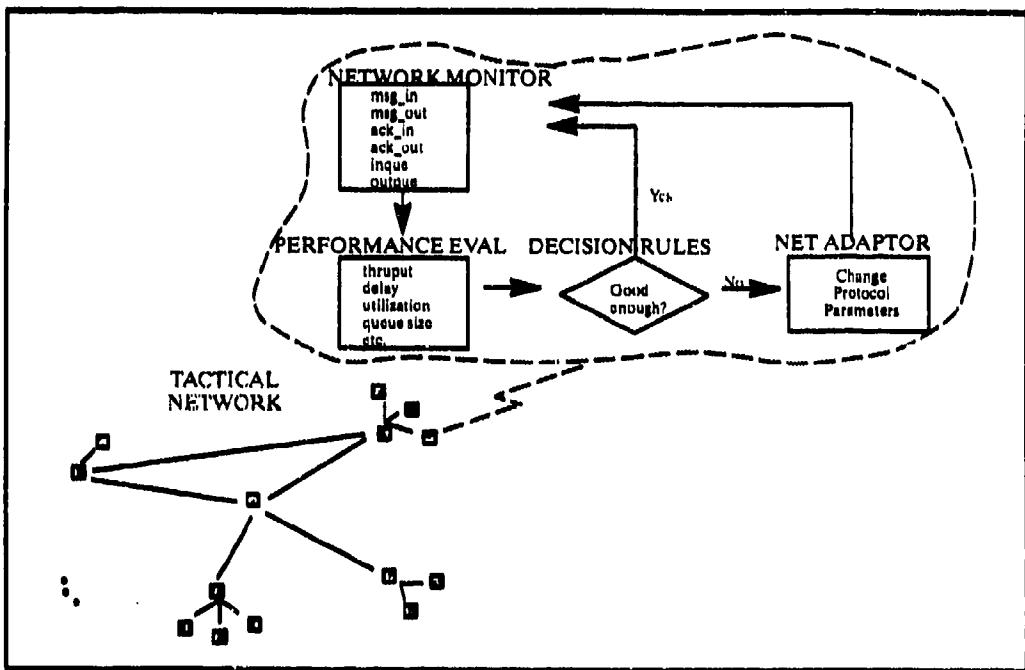


Figure 1. An Adaptive Tactical Network.

OBJECTIVE

Protocol parameters such as packet size, coding technique and channel access algorithm could be adjusted to improve or possibly optimize information transfer. In general, the objectives are to maximize throughput and minimize delay in the delivery of information to the end user, where throughput and delay are defined as follows:

Network throughput is the average number of bits per second that are successfully transmitted and acknowledged over a one hour test cell. This does not include such overhead as acknowledgements, error detection/correction codes, synchronization characters, or, in the event of collisions, message retransmissions.

Network delay is the average time interval that passes between a message's arrival at the host's modem and the host's receipt of the message acknowledgement. Messages that are not completely serviced during the running of the test cell will not be considered in computing network delay.

The question of how best to adapt to a particular situation is extremely difficult to address. Research into network protocols and communication channels will provide the underlying foundation required to identify appropriate network adaptations. However, because of the complexity of these protocols, theoretical research must be supported with carefully designed and controlled experiments to determine which network parameters are most useful in moderating network congestion.

APPROACH

Based on previous research, several parameters were selected for a sensitivity analysis: retry interval, the time to wait for an acknowledgement before retransmitting; window size, the number of outstanding messages permitted before transmission is blocked; message length, the number of characters in each message; and arrival rate, the number of messages per hour queued for transmission at each node.

A prepilot test has been conducted to determine thresholds for retry interval, window size, message length and arrival rate. Next a pilot test will be executed to screen each of the four parameters for possible elimination. Finally an experiment will be designed and executed to measure throughput and delay under each of the test cell conditions.

EXPERIMENTAL CONFIGURATION

The experimental hardware consists of the equipment shown in Figure 2. The computers are Tadpole SPARCbook I's each with 32 megabytes of memory. These are connected to a Harris Black Box Radio Emulator via Harris Tactical Data Buffers (TDB). The TDB provides an interface between VHF transceivers and digital computer equipment.

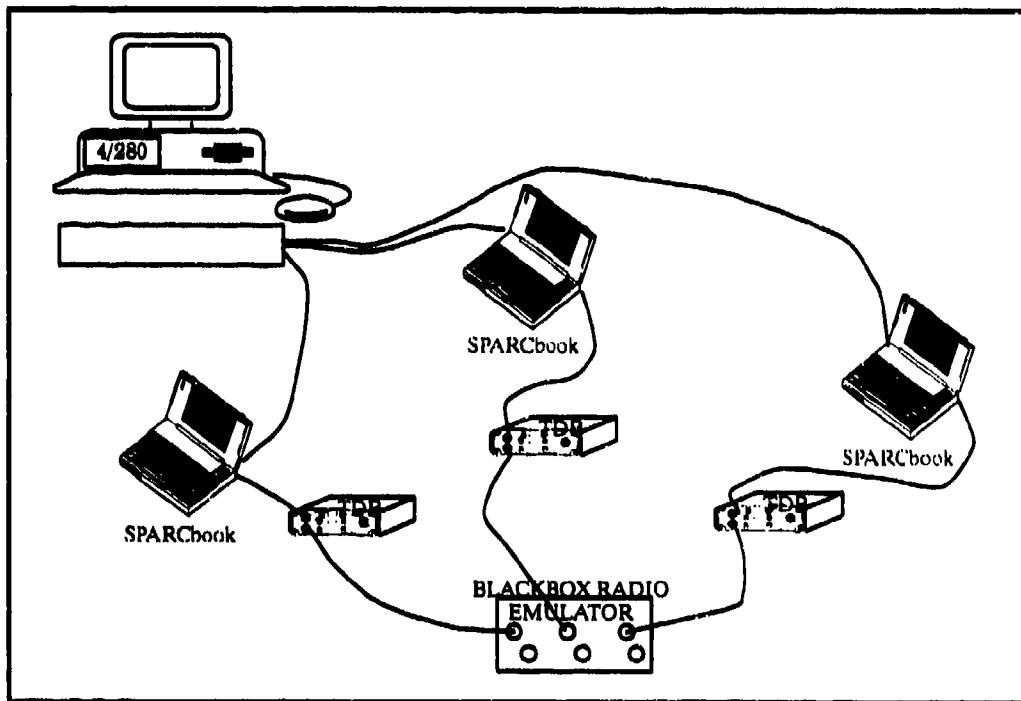


Figure 2. Experimental Configuration

In providing this service the TDB performs the following tasks: data modulation/demodulation, error detection/correction, and compensation for unequal terminal and radio link data rates.

The SPARCheckers are connected to a SUN 4/280 serving as the data storage and data reduction machine. The software residing on the SUN generates messages, logs message traffic, and identifies message retries and delay. To minimize blocking and possible errors, input is read from text files in a predefined order. Through the software, the experimenter can interactively select which test cell and iteration to execute next.

PREPILOT RESULTS

The prepilot test was conducted to determine thresholds for retry interval, window size, message length and arrival rate and to explore limitations of the software. Figure 3 illustrates the various factors explored.

During this period it was found a window size of one resulted in overflow errors that prevented data transmission. Average message delays were computed over one minute intervals to insure the sampling was sufficient to identify the warm up period. Software requires further development to support fully automated execution of an entire replication and to accommodate more nodes in the network.

FACTOR	LEVELS	
Retry Timeout (seconds)	10	40
Window Size (messages)	8	50
Message Length (characters)	80	240
Arrival Rate (messages/node)	200	600

Figure 3. Prepilot Study Factors and Levels

FUTURE WORK

When software modifications are completed, the pilot test will be conducted to explore the need to eliminate or refine the levels of investigation for any of the factors. The number of replications will be dependent upon the duration of the test cells and the amount of automation introduced. A full factorial design will be implemented. The parameters selected for this test are those which can be easily modified. Future experiments will consider more complex protocol modifications.

MEASURES OF EFFECTIVENESS
FOR
MONTE CARLO SENSITIVITY ANALYSES

Andrew W. Harrell
U.S. Army Engineer Waterways Experiment Station
Vicksburg, MS 39180-0631

ACKNOWLEDGMENTS. The tests described and the resulting data presented herein, unless otherwise noted were obtained from research conducted under the MILITARY RESEARCH DEVELOPMENT TEST AND EVALUATION PROGRAM of the United States Army Corps of Engineers by the U.S. Army Engineer Waterways Experiment Station. Permission was granted by the Chief of Engineers to publish this information.

I wish to thank Mr. Newell R. Murphy, Jr., and Dr. Niki C. Deliman for their valuable comments and suggestions which have improved the exposition of this paper.

ABSTRACT. The Army Mobility model (AMM) developed, at the U.S. Army Engineer Waterways Experiment Station, uses the data from about a hundred factors that describe a vehicle terrain unit, road unit, or linear feature to predict vehicular speeds. Recently, Monte Carlo simulations were conducted for several wheeled and tracked vehicles and different areas, varying some selected groups of these factors plus and minus 10 percent about their nominal values. The results of these simulations have been studied to develop empirical relationships that allow the expression of confidence measures for the speed predictions on an entire mobility map. As a first step, programs have been written to test methods to estimate the value of continuous statistical parameters (the mode and its standard deviation) of a discrete histogram. This allows theorems of mathematical statistics to be applied to the confidence levels around the values of the parameters. The method uses a variation I made on E. Parzen's formula for the location of the mode of the continuous distribution associated with a discrete histogram.¹ The formula works by estimating the rate of an associated statistical process by discrete windows (Jth waiting times). The incomplete gamma function and a maximum likelihood product is then used to estimate the parameters.² This approach has been tested for a range of Monte Carlo generated discrete approximations to gamma distributions. It was then applied to the histograms of possible errors in speed predictions of tactical vehicles moving across areas on different mapsheets. These histograms were generated previously in the course of the work by Lessem and Ahlvin and are discussed in reference [6].

¹See Parzen, Emanuel, "Stochastic Processes", Holden Day, 1962, and Press, W., Flannery, B. et al., "Numerical Recipes in C," 2nd ed., Cambridge U. Press, 1988.

² Ibid.

In trying to determine how to organize the sensitivity trials in this particular set of programs and data there are several approaches that can be taken. Because the speed prediction program uses a series of lookup tables and flow chart, "yes" or "no", go and no-go cutoff rules, points at which the program computes a no-go output are natural areas to investigate its sensitivity to errors in the data. Error measures can be associated with "critical regions" in the data around these points. Determining the modes and moments in the discrete non-parameteric histograms generated by the sensitivity trials gives a way of characterizing and reproducing the confidence in information contained in the program's output involving these regions. One approach, which measures the program's "inherent sensitivity" to errors, is terrain-independent and vehicle dependent. It examines the code in the program to find the 1-factor critical regions in the outputs of the Monte Carlo trials. It then adjusts the values of the other factors in a detrimental direction of the lookup table values until the 2,3 and higher multi-factor critical regions are identified. Another approach is "project specific" and is both terrain dependent and vehicle dependent. It looks at the areas on the speed prediction maps where no-gos occur. It then goes back to the input files to determine the values of the data at the terrain units where these no-gos occur. This is the approach that will be taken in this paper.

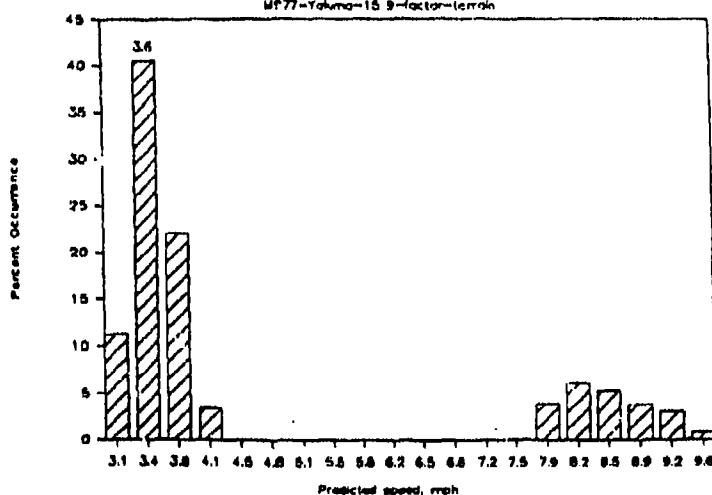
After the procedure for conducting the trials is determined it is important to consider ways to examine confidence levels for the parameters that are estimated. One approach to this, which recently has gained popularity, is the technique of bootstrapping. This technique conducts Monte Carlo trials of the Monte Carlo trials. The algorithm resamples not from the original data, but from a smoothed kernel estimate of the data (see MathCad [8] for the details of the algorithm and Efron, Hall and Tittleman, and Scott for the theory behind formulas for the variance of the sampled estimate of the parameter). Smoothed kernel formulas, introduced by Parzen and others (see Scott [12], Parzen [9]) allow better resolution of modes and other information in the data using a given histogram bin size or window. In order to estimate the second moment or the variance of the kernel estimate, it is necessary to write programs to compute the second derivative of the frequency polygon of the histogram (see Scott [12]). Bootstrapping confidence intervals can then also be computed from this information.

In this paper a somewhat simplified approach is taken. A leave-one-out maximum likelihood product of smoothed kernels over different possible bin widths is taken. The product is taken over a choice of possible bin widths. Once the best bin width is determined the variance of the kernel associated with this bin width is computed (see Numerical Recipes in C, 2nd ed. [10]). This aggregates the data in a one dimensional histogram and does not give you as much information as in the more complicated multi-dimensional approach.

Figures 1, 2, 3, and 4 show the results of a series of Monte Carlo error sensitivity trials run on some vehicle speed

NRMM-Areal: Project-Spec. Sensitivity

MF77=Volume=15.9=factor=Terrain



NRMM-Areal: Project-Spec. Sensitivity

MF77=Volume=77.6=factor=Terrain

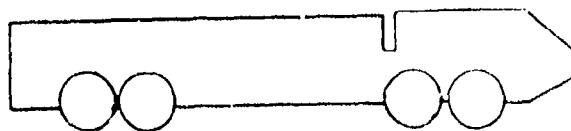
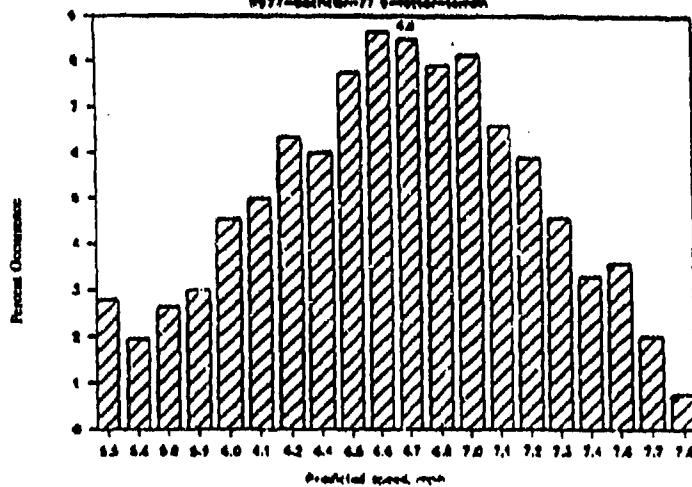
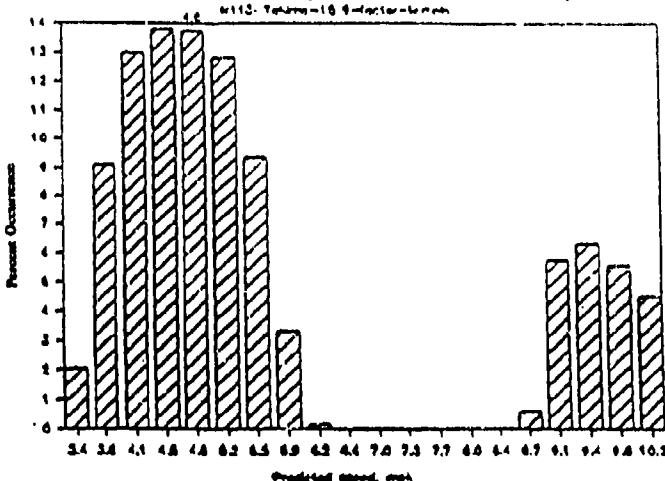


Figure 1 Monte Carlo Sensitivity Runs, HEMTT

NRMM-Areal: Project-Spec. Sensitivity



NRMM-Areal: Project-Spec. Sensitivity

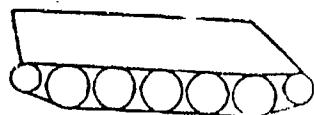
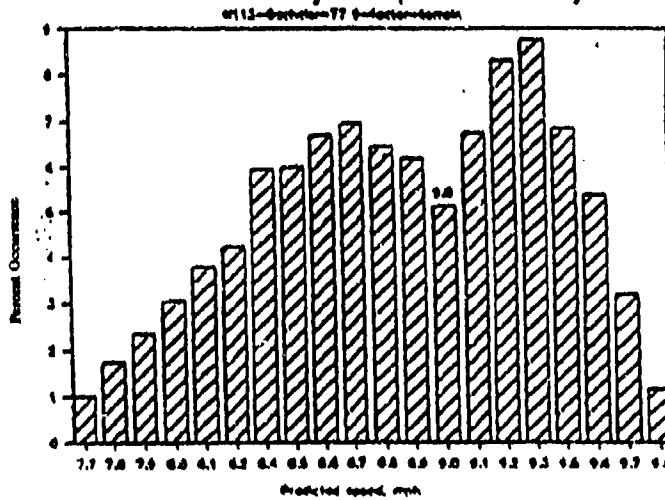


Figure 2 Monte Carlo Sensitivity Runs, M113

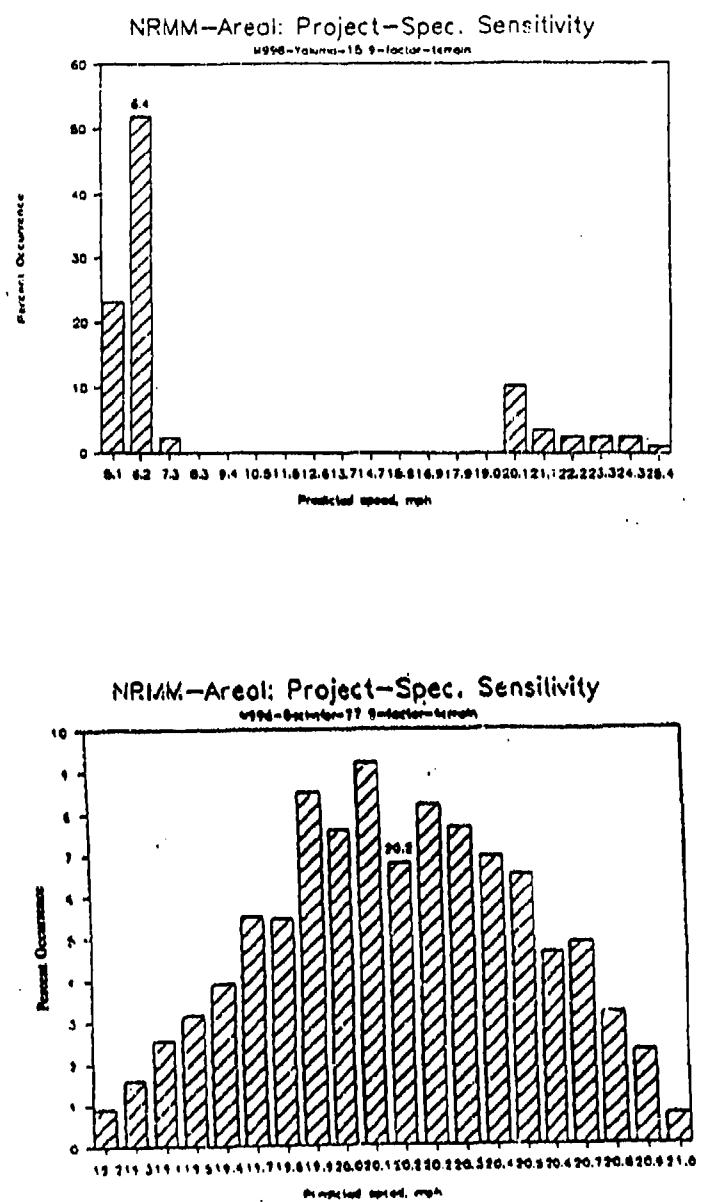


Figure 3 Monte Carlo Sensitivity Trials, HMMV

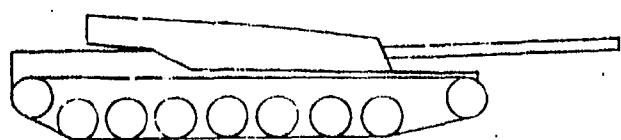
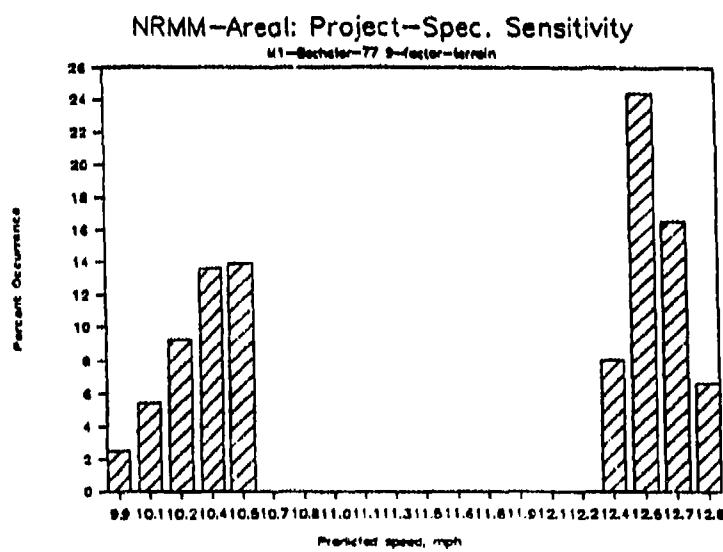
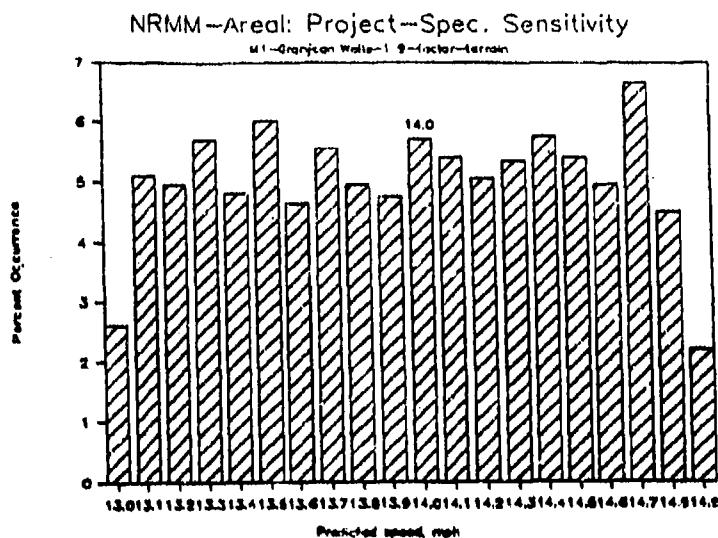


Figure 4 Monte Carlo Sensitivity Trials, M-1

predictions by Lessem et al. [6]. They display the speeds predicted for the M998 High Mobility Multi-Purpose Vehicle (HMMV), the M977 10-Ton Heavy Expanded Mobility Tactical Truck (HEMTT), M113 Armoured Personnel Carrier, and the M-1 tank. The terrain areas tested are in Yakima, Washington, Granjean Wells, New Mexico and Bachelor, Australia. The graphs have predicted speeds plotted on the horizontal axis. The speeds were computed by varying nine factors: soil strength, slope, surface roughness, visibility, vegetation type, and four other attributes dealing with obstacle characteristics around their nominal values in a certain terrain unit. The nominal values for that terrain unit were chosen as the points around which the vehicle's performance on the mapsheet terrain units changed most noticeably. The points were determined by referring both to the output that the program computed and to the tables in the speed computation program where the performance changed significantly. On the vertical axis is a count of the number of occurrences of a given speed for that vehicle, that terrain unit, and for the range of Monte Carlo trials used. Both uniform and normal density functions were used to compute the random numbers used in the Monte Carlo sensitivity trials. Thus the graph displays the areal sensitivity of the speed predictions for that vehicle in that area. Notice that the results don't appear to have a common probability density function. The WES technical reports by Lessem et. al. [6] and [7] contain a more detailed discussion of the features of the mobility programs which cause the histograms to assume these shapes.

In general, these histograms will separate into several parts each with distinct characteristics. In this particular case parts of the graphs associated to each single mode were separated out. Let us assume this has already been done. We arrange the results of the Monte Carlo simulations in a histogram of N bins with the number of Monte Carlo hits (test items) in the i th bin equal to $hist_i$. In order to estimate the number of Monte Carlo trials necessary to reproduce the probability density function from which these results give samples we have to use an unstructured or nonparametric approach.³ Let us define

$$(1.1) \quad p(t+1/2*J) = \frac{\sum_{i=t}^{t+J} hist_i}{N*J}$$

where t = bin number around which estimate is centered
 J = integer ≥ 1
 N = total number of observations

³ Keinosuke Fukunaga, "Introduction to Statistical Pattern recognition," Academic Press, 2nd ed., 1991.

According to the reference by Fukunaga [3], this formula gives the Parzen density estimate for the value of this probability density function at the point $k = t + J/2$.⁴ In this formula we are using a local region defined by a window of size J around the point to estimate the number of hits in a counting process in terms of the histogram values located in this region. This formula gives estimates for the values of the density function at $N-J$ points. Sorting these estimates and picking out the middle and highest values then gives the best prediction of the mode and the mean of the histogram using windows of size J . On page 261 of this reference the value of the standard deviation of this estimate is calculated to be:

$$(1.2) \quad \sigma(t+1/2*J) = \frac{\sum_{i=t}^{t+J} hist_i}{J*\sqrt{J*N}} .$$

Note that the value of this standard deviation refers to an interval around a point on the x-axis of the histogram and not around the height of the histogram or number of Monte Carlo values in that bin.

These formulas and theorems allow a leave-one-out procedure along with a maximum likelihood product to be used to estimate the value of the window size which gives the smallest error in estimating the parameters.⁵

Using our procedure for computing estimates of the value of the probability distribution, at the point k defined in equation 1.1 the function $p(k)$ is proportional to the amount the cumulative distribution function changes in this interval... so, the larger it is, the better is the chance for a local maximum of the probability distribution function at that point. The program computes estimates of the continuous modes for different window sizes, where J = window size, x_i = bin# of largest of these estimates, $p(k)$ = weighted estimate of mode at this bin = (sum of # of distribution hits in the bins inside a window of width J centered at k)/(total # shots)* J . In the case where the

⁴ Actually, this is the density function of a "renewal counting process" as defined in Parzen [9].

⁵ See besides the Numerical Recipes in C reference also the Introduction to Statistical Pattern Recognition text referred to above. These same procedures can be used to characterize the histogram distribution of pixel intensities in digital images. Such a characterization allows the use of various neural network learning procedures to be used to identify the images.

distribution function is suspected to be bimodal, this procedure will identify at least the top two modes when it is iterated over different window sizes.

Let $\delta(J)$ = the range of data values around the candidate for mode calculated using a window size of J .

$$\text{Thus, } \delta(J) = \sum_{i=x_0 - \frac{J}{2}}^{i=x_0 + \frac{J}{2}} \text{hist}_i .$$

Then, in this notation, the probability distribution $p_J(k)$ of the smoothed estimate of the original data is given by:⁶

$$p_J(k) = \frac{\delta(J)}{N J} .$$

Let

$$(2.1) \quad \delta_J(k) = \sum_{i=k - \frac{J}{2}}^{i=k + \frac{J}{2}} \text{hist}_i .$$

Let $H(J)$ = the hypothesis that the true mode x , has been identified by considering a window of size J . We want to consider how likely it is that the range around x , should be shorter than it is observed to be. Let $P(a, x)$ be the incomplete gamma function:

$$P(a, x) = \frac{1}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt$$

where:

⁶ See the discussion in Numerical Recipes in C edition 1 and also the book by Parzen, pages 133-134.

$$\Gamma(a) = \int_0^{\infty} e^{-t} t^{a-1} dt .$$

Thus, $P(a, x)$ is the cumulative Poisson probability distribution function, $\text{Prob}(X \leq a)$ for the Poisson probability distribution X . It is defined as the probability that the number of Poisson random events occurring will be between 0 and $a - 1$. Each of these random events will have a probability of occurrence of $N p_j$.

The probability that the range around x_j is actually shorter than observed to be if $H(n)$ is true instead of $H(J)$ is⁷:

$$\int_0^{\delta_j(J)} (N p_j(n)) \frac{(N p_j(n) t)^{n-1}}{(n-1)!} e^{-N p_j(n) t} dt .$$

If we let:

$$\begin{aligned} y &= N p_j(n) t \\ a &= n \\ x &= \delta_j(J) \end{aligned}$$

in the above equation ,

then it is equal to:

$$P(n, \frac{\delta_j(J)}{N p_j(n)})$$

which is the same as:

$$P(n, J \frac{\delta_j(J)}{\delta_j(n)}) .$$

Taking the product of all these factors for each mode x_n then gives the likelihood that the range around x_n should be shorter than the range observed around x_j for all n other than J . Thus the likelihood function is defined by $\text{Likelihood}(H(J))$:

⁷ Parzen, Ibid pp. 133-134.

$$(2.2) \quad L(n|J) = \prod_{n=0}^N P(n, J * \delta_J(J) / \delta_J(n)) .$$

The program then computes the value of this window size J that maximizes the likelihood function, given a set of arrayed a posteriori error sizes.

More precisely, the steps in the computation are:

1) Compute the

$$\delta_J(n)$$

according to equation (2.1) for the points corresponding to each bin.

2) Compute the maximum likelihood products according to equation (2.2) in order to determine the optimal window size.

3) Compute the weighted sums $p(k)$ according to equation (1.1) and the standard deviations according to equation (1.2) for the points corresponding to each bin.

Because of the nonparametric form of the Parzen density estimate, the procedures will work for any empirically determined histogram. A discrete sorting procedure normally gives a pretty good estimate of the value of the mean and mode (even assuming the actual distribution is continuous). However, in order to approximate the size of the standard deviation in the estimation of the parameters, it is necessary to use the maximum likelihood estimators. These estimators of the best window sizes will result in good approximations of the parameters.

An example of how these parameter estimates work is shown as it is applied to the results of Monte Carlo sensitivity runs in Figures 1,2,3,4. The simulations shown in the figures were conducted for four vehicles the HMMV, the M997 trailer transporter, the M113 APC, and the M-1 tank. The top charts show the results for a mapsheet Yakima Proving Grounds and the bottom charts those for a mapsheet including Batchelor Australia. These figures show the results of varying the parameter values plus and minus 10 per cent around their nominal values. Nominal values are defined as the vehicle parameters plus the specific parameter values in each terrain unit. For, this analysis, we considered the particular values for which that vehicle experiences a go, no-go situation, as the values around which variations were made.

Data from the M997, M113, and M998 runs were extracted directly from the top row of histograms in Figures 1,2, and 3 respectively. Programs were written to expand the information into a 20 bin histogram and to scale the data. This turned out to be a good range for the incomplete gamma function to discriminate the maximum likelihood estimates. The results of the program runs are shown below. First the program calculates a value for the mode by simply sorting the columns of the histogram. This is called a discrete estimate. The abscissa of this point is called mode1. Then the program computes the optimal window size for smoothing the data using the leave-one-out maximum likelihood procedure explained above and determines a continuous estimate for the mode along with a standard deviation. Both of these numbers are computed using this optimal window size.

The results are shown below:

histogram of Monte Carlo error runs
M998 Yakima-15 9-factor-terrain (mode#1)

x	p(x)	graph:
4.0000	0.1820	*****
4.5500	0.3275	*****
5.1000	0.4731	*****
5.6500	0.6186	*****
6.2000	0.3311	*****
6.7500	0.0437	**
7.2500	0.0218	*
7.7500	0.0000	
8.3000	0.0000	

Data drawn from a histogram of Monte Carlo sensitivity to errors in terrain factors

Discrete estimate of mode of data set is 42.500000

Discrete estimated value of mode1= 5.650001

Probability of mode detected at window size 3 is 0.229365

Probability of mode detected at window size 4 is 0.253296

Probability of mode detected at window size 5 is 0.256476

Probability of mode detected at window size 6 is 0.014268
Probability of mode detected at window size 7 is 0.014372

Most likely window size is 5 value of mode is 32.50000

Continuously estimated value of modei=5.10000

Standard deviation of the continuous estimate (for this window size) is 0.607092

histogram of Monte Carlo error runs
M998 Yakima-15 9-factor-terrain (mode#2)
x p(x) graph:

16.8000	0.0000
17.3500	0.0000
17.9000	0.0000
18.4500	0.0000
19.0000	0.7143 *****
19.5500	1.4286 *****
20.0500	0.9740 *****
20.5500	0.5195 *****
21.1000	0.4545 *****
21.6500	0.3896 *****
22.2000	0.3831 *****
22.7500	0.3766 *****
23.2500	0.3766 *****
23.7500	0.3766 *****
24.3000	0.2597 *****
24.8500	0.1429 *****
0.0000	0.0000

Data drawn from a histogram of Monte Carlo sensitivity to errors in terrain factors

Discrete estimate of mode of data set is 11.0000000

Discrete estimated value of modei=19.549995

Probability of mode detected at window size 3 is 0.204653
Probability of mode detected at window size 4 is 0.039479
Probability of mode detected at window size 5 is 0.116221
Probability of mode detected at window size 6 is 0.065450
Probability of mode detected at window size 7 is 0.129556
Most likely window size is 3 value of mode is 11.0000000
Standard deviation of the continuous estimate (for this window size) is 0.269430

Continuously estimated value of modei=11.00000

histogram of Monte Carlo error runs
M997 Yakima-15 9-factor-terrain (mode#1)

x p(x) graph:

2.8000	-0.0387
2.9500	0.2062
3.1000	0.4510
3.2500	0.6959
3.3833	0.5956
3.5167	0.4954
3.6500	0.3952
3.8000	0.2234
3.9500	0.0515
4.0833	0.0344
4.2167	0.0172
4.3500	0.0000

Data drawn from a histogram of Monte Carlo sensitivity
to errors in terrain factors

Discrete estimate of mode of data set is 40.500000

Discrete estimated value of modei= 3.250000

Probability of mode detected at window size 3 is 0.282627
 Probability of mode detected at window size 4 is 0.064773
 Probability of mode detected at window size 5 is 0.076770
 Probability of mode detected at window size 6 is 0.083600
 Probability of mode detected at window size 7 is 0.084213
 Most likely window size is 3 value of mode is 40.500000
 Standard deviation of the continuous estimate (for this window
 size) is 1.253331

Continuously estimated value of modei=3.250000

histogram of Monte Carlo error runs
M977 Yakima-15 9-factor-terrain (mode#2)

x	p(x)	graph:
6.9000	-0.0032	*****
7.0500	0.0000	*****
7.2000	0.0032	*****
7.3500	0.0065	*****
7.4833	0.1775	*****
7.6167	0.3485	*****
7.7500	0.5195	*****
7.9000	0.6494	*****
8.0500	0.7792	*****
8.2000	0.7143	*****
8.3500	0.6494	*****
8.4833	0.6061	*****
8.6167	0.5628	*****
8.7500	0.5195	*****
8.9000	0.4545	*****
9.0500	0.3896	*****
9.1833	0.3030	*****
9.3167	0.2165	*****
9.4500	0.1299	*****
0.0000	0.0000	*****

Data drawn from a histogram of Monte Carlo sensitivity
to errors in terrain factors

Discrete estimate of mode of data set is 6.000000

Discrete estimated value of modei= 8.049999

Probability of mode detected at window size 3 is 0.007822
Probability of mode detected at window size 4 is 0.023633
Probability of mode detected at window size 5 is 0.082934
Probability of mode detected at window size 6 is 0.183745
Probability of mode detected at window size 7 is 0.298017
Most likely window size is 7 value of mode is 5.500000
Standard deviation of the continuous estimate (for this window
size) is 0.103940

Continuously estimated value of modei=8.199999

histogram of Monte Carlo error runs
M113 Yakima-15 9-factor-terrain (mode#1)

x p(x) graph:

3.0000	-0.0862	
3.2000	0.1149	*****
3.4000	0.3161	*****
3.6000	0.5172	*****
3.7500	0.6322	*****
3.9000	0.7471	*****
4.1000	0.7701	*****
4.3000	0.7931	*****
4.4500	0.7902	*****
4.6000	0.7874	*****
4.8000	0.7644	*****
5.0000	0.7414	*****
5.1500	0.6322	*****
5.3000	0.5230	*****
5.5000	0.3563	*****
5.7000	0.1897	*****
5.8500	0.0977	****
6.0000	0.0057	
0.0000	0.0000	
0.0000	0.0000	

Data drawn from a histogram of Monte Carlo sensitivity
to errors in terrain factors

Discrete estimate of mode of data set is 13.799999

Discrete estimated value of modei= 4.300000
standard deviation is 0.283068

Probability of mode detected at window size 3 is 0.003206
Probability of mode detected at window size 4 is 0.020749
Probability of mode detected at window size 5 is 0.114160

Probability of mode detected at window size 6 is 0.132713
Probability of mode detected at window size 7 is 0.369124
Most likely window size is 7 value of mode is 13.799999
Standard deviation of the continuous estimate (for this window size) is 0.283068

Continuously estimated value of mode=13.79999

Summary of Mode Estimates for data

Discrete estimate of mode of data set 1 is point 5.650001 at 42.500000
continuous estimate of mode of data set 1 is point 5.100000 with value 32.500000
A window of size 5 was used to estimate this

Discrete estimate of mode of data set 2 is point 19.549995 at 11.000000
continuous estimate of mode of data set 2 is point 19.549995 with value 11.000000
A window of size 3 was used to estimate this

Discrete estimate of mode of data set 3 is point 3.250000 at 40.500000
continuous estimate of mode of data set 3 is point 3.250000 with value 40.500000
A window of size 3 was used to estimate this

Discrete estimate of mode of data set 4 is point 8.049999 at 6.000000
continuous estimate of mode of data set 4 is point 8.199999 with value 5.500000
A window of size 7 was used to estimate this

Discrete estimate of mode of data set 5 is point 4.300000 at 13.799999
continuous estimate of mode of data set 5 is point 4.300000 with value 13.799999
A window of size 7 was used to estimate this

In summary, using this technique of estimation for finding modes there is in one case (data set 1) about a 10 percent increase in the accuracy of the determination of its location. This makes available a more accurate fix on the NOGO program vehicle speed values around which to do the sensitivity analyses. Also, determination of the optimal window size to use in the estimate, gives a means to non-parametrically estimate the standard deviation of the sensitivity analyses results. This then tells us how many Monte Carlo trials should be used to explore the program's sensitivity to variations in the values in its internal tables and input data. For example, for the two runs concerning the M977 performance, one mode has a determination with a standard deviation of 1.253 and the other with a standard deviation of .1039. After determining this, you could then go back and run 10 times more

Monte Carlo trials around the first mode. Similarly, although it was not analyzed for this paper, the determination of a mode in the case of the M-1 tank is much less well defined. Looking at the Monte Carlo sensitivity histogram in the top part of Figure 4, it is clear that in this case the predictions will be less accurate.

REFERENCES

1. Dahlquist, G. and Bjorck, A., "Numerical Methods," Prentice Hall, 1974.
2. Efron, B., "The Annals of Statistics," Vol 7, No 1 (1979).
3. Fukunaga, K., "Introduction to Statistical Pattern Recognition," Academic Press, 2nd ed., 1991.
4. Hall, P. and DiCiccio, T. J. and Romano, J. P., "On Smoothing and the Bootstrap," *Annalen Statistics* 17 692-704, 1989.
5. Hall, P. and Titterington, D. M., "On Confidence Bands in Nonparametric Density Estimation and Regression," *J. Multivariate Analysis* 26 59-88, 1988.
6. Lessem, A. S., Ahlvin, R., and Mason, G., "Stochastic Vehicle Mobility Forecasts Using the NATO Reference Mobility Model," Report 1, "Basic Concepts and Procedures," U.S. Army Engineer Waterways Experiment Station, Technical Report, GL-91-11.
7. Lessem, A. S., Ahlvin, R., Mlakar, P., and Stough, W., "Stochastic Vehicle Mobility Forecasts Using the NATO Reference Mobility Model," Report 2, "Extension of Procedures and Application to Historic Studies," U.S. Army Engineer Waterways Experiment Station, Technical Report, GL-93-15.
8. MathCad, Inc., "Statistics I, Tests and Estimation," Cambridge, MA, 1991.
9. Parzen, E., "Stochastic Processes," Holden Day, 1962.
10. Press, W., Flannery, B. et al., "Numerical Recipes in C," 2nd ed., Cambridge U. Press, 1988.
11. Press, W., Flannery, B. et al., "Numerical Recipes in C," 3rd ed., Cambridge U. Press, 1993.
12. Scott, D. W., "Multivariate Density Estimation: Theory, Practice, and Visualization," John Wiley, 1992.

Attended the Short Course and the Meetings

Dr. Andrew Harrell
Waterways Experiment Station
Vicksburg, Ms 39180

Dr. Paul J. Deason
U. S. Army Tradoc Analysis Center- WSMR
White Sands Missile Range, NM 88002 -5502

Dr. Barney Bissinger
Hershey Foods
Technical Center, Hershey, PA 17033

David Scott
NSA/Rice
12123 Gray Star Way,
Columbia, MD 21044

Margaret Gelder Ehm
Rice University, Dept. of Statistics
P. O. Box 1892 Houston, TX 77251

Barbara Wainwright
Salisbury State University
Department of Mathematics
Salisbury Md. 21801

Henry B. Tingey
University of Delaware
Department of Mathematical Science
Newark DE 19716-2552

Linda L.C. Moss
U. S. Army Research Laboratory
Attn: AMSRL-SL-BV,
Aberdeen proving Ground, MD 21006-
5068

Calvin Parker, Jr.
U. S. Army Missile command, RDEC
USAMICOM, RDEC, AMSMI-RD-AS-MM
Redstone Arsenal, AL 35898-5248

Malcolm S. Taylor
U.S. Army Research Lab
Aberdeen Proving Ground, MD

Barry A. Bodt
Army Research Laboratory
Aberdeen Proving Ground

Paul H. Thrasher

**U. S. Army
STEWS-MTD-OA, White Sands Missile
Range, NEW Mexico 88002**

**Ann E. M. Brodeen
U. s. Army Research Laboratory
Attn AMSRL-CI-CC
Aberdeen Proving Ground, MD 21005-
5067**

**Royal J. Weaver
The Upjohn Company
301 Henrietta Street
Kalamazoo, MI**

**Marshall H. Brunden
The Upjohn Company
301 Henrietta Street
Kalamazoo, MI 49007**

**Judy L. Palmer
U.T. M. D. Anderson Cancer Center
Department of Biomathematics
Box 237
Houston, Texas 77030**

**Thomas Walker
STECS -EN-BA
Aberdeen Proving Ground, MD. 21005**

**Barbara Kaschenbach
STECS-EN-BA (CSTA)
Aberdeen Proving Ground, MD 21005**

**Deloris Testerman
TEXCOM, Attn: CSTE-TAV
Fort Hood, Texas 76544-5065**

**Eugene Dutoit
U. S. Army Infantry School
Fort Benning, GA 31907-5400**

**Stephan Sain
Rice University
P.O. Box 1892
Houston, Texas 77251**

**Keith Baggerly
Rice University
P.O. Box 1892
Houston, Texas 77251**

Marc Elliott

Rice University
P.O. Box 1892
Houston, Texas 77251

James R. Thompson
Rice University
P.O. Box 1892
Houston, Texas 77251

Dennis Cox
Rice University
P.O. Box 1892
Houston, Texas 77251

Mark Vangel
National Institute of Standards and TechStatistical Engineering Division,
Gaithersburg Md 20899

Monnie McGee - Rice University
P.O. Box 1892
Houston, Texas 77251

Robert Burge (Home Address)
7134 Smooth Path
Columbia, MD 21045

Douglas Tang (Home Address)
10809 Meadowhill Road
Silver Spring, MD. 20901

Jack Lee
U.T. M. D. Anderson Cancer Center
Department of Biomathematics
Box 237
Houston, TX 77030

John F. Elder
Rice University
Department of Computational and Applied Mathematics
P. O. Box 1892
Houston, Texas 77251

Peter W. M. John
University of Texas
Department of Mathematics
University of Texas
Austin Texas 78712

D. Richard Cutler
Utah State University
Department of Mathe and Stat
Logan, UT 84322-3900

Sue Bass
Dept Math Sci, U.S. Military Academy
West Point, NY 10996

Col Rickey A. Kolb
Department Math Sci
U. S. Military Academy
West Point, NY 10996

Nozer Singpurwalla
George Washington University
707 22nd Street, N.W.
Washington D. C. 20052

Al Feiverson
Johnson Space Center
Mail Code ET 2
Houston, Texas 77058

Barry W. Brown
U.T. M.D. Anderson Cancer Center
Dept. of Biomathematics
Box 237
Houston, Texas 77030

Martin Lawera - Rice Univ
Rice University
P.O. Box 1892
Houston, Texas 77251

Monnie McGee
Rice University
P.O. Box 1892
Houston, Texas 77251

Denise Bullock
USA Corps of Engineers,
Waterways Experiment Station
3909 Halls Ferry Rd.,
Vicksburg, MS

Nancy Renfroe

**U.S.A. Corps of Engineers
Waterways Experiment STation**

**Bernard Harris
University of Wisconsin - Statistics
1210 W. Dayton Street
Madison, Wisconsin 53706**

**J. Sethuramans
Department of Statistics
B-167
Florida State Univ.
Tallahassee, FL 32306-3033**

**Katherine Ensor
Rice University
P.O. Box 1892
Houston, Texas 77251**

**Joe Ensor
Empirical Science
3506 Hwy. 6 so.,
Suite 302
Sugar Land, Texas**

Attended the Conference only

David F. Crues, Ph.D
U.S.U.H.S. Medical School
6123 Camelback Lane
Columbia, Maryland 21045-3812

Mel Brown
ARO
P. O. Box 12211, RTP, NC 27709-2211

Gerald R. Anderson
ARO
P. O. Box 12211 RTP NC 27709

Francis Drexel
U. S. Army Research Office
P. O. Box 12211, Research Triangle Park,
NC. 27709-2211

Wei-Yin Loh
University of Wisconsin
1210 W. Dayton Street
Madison, WI 53706

Richard P. Heydorn
NASA - Johnson Space Center
Houston, Texas 77085

LTC. Ronald Scotka
TEXCOM ATTN: CSTE-TAV
Fort Hood, Texas 76544-5065

Emanuel Parzen
Texas A & M
Department of Statistics
College Station, Texas 77843-3143

Dennis Cox
Rice University
P.O. Box 1892
Houston, Texas 77231

Mark Vangel
National Institute of Standards and
TechStatistical Engineering Division,
Gaithersburg Md 20899

Jerry Thomas
U. S. Army Research Lab
AMSRL-WT -WE
Aberdeen Proving Ground, MD 21005-5066

Attended the Shourt Course only

Joni Dublin
2400 Westheimer Apt 222E
Houston, Texas 77098

Mary York
University of Houston
(no address)

Ewa Wlodarczyk
School of Public Health
(no Address)

David Sullivan
Texas Natural Resources Conservation Commission
P. O. Box 13087
Austin, TX 78711-3087

Kathleen Callahan
HQ, TEXCOM
CSTE-TCD-C Fort Hood, TX 76544-5063

Richard Scribner
HQ Texcom
CSTE-TCD-A Fort Hood, TX 76544-5065

Rebecca Richards-Kortum
University of Texas
718 Shady Tree Dr.
Austin, TX 78748

Terri Hood-- Rice University
Rice University
P.O. Box 1892
Houston, Texas 77251

Rachel Boeckenhauer -- Rice University
P.O. Box 1892
Houston, Texas 77251

Philip Chen -- Rice Univ
P.O. Box 1892
Houston, Texas 77251

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

18. REPORT SECURITY CLASSIFICATION Unclassified		19. RESTRICTIVE MARKINGS	
20. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release: distribution unlimited.	
21. DECLASSIFICATION/DOWNGRADING SCHEDULE		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6. PERFORMING ORGANIZATION REPORT NUMBER(S) ARO REPORT 94-2		7a. NAME OF MONITORING ORGANIZATION	
6a. NAME OF PERFORMING ORGANIZATION Army Research Office	6b. OFFICE SYMBOL (if applicable) AMXRO-MCS	7b. ADDRESS (City, State, and ZIP Code)	
8c. ADDRESS (City, State, and ZIP Code) P.O. Box 12211 Research Triangle Park, NC 27709		8a. NAME OF FUNDING/SPONSORING ORGANIZATION AMSC	
8c. ADDRESS (City, State, and ZIP Code)		8b. OFFICE SYMBOL (if applicable)	
		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
		10. SOURCE OF FUNDING NUMBERS	
		PROGRAM ELEMENT NO.	PROJECT NO.
		TASK NO.	WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) Proceedings of the Thirty-Ninth Conference on the Design of Experiments in Army Research, Development and Testing			
12. PERSONAL AUTHOR(S)			
13a. TYPE OF REPORT Technical	13b. TIME COVERED FROM Jan 94 TO Feb 94	14. DATE OF REPORT (Year, Month, Day) 1994 June	15. PAGE COUNT 333
16. SUPPLEMENTARY NOTATION			
17. CCGATI CODES	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) This is a technical report of the Thirty-Ninth Conference on the Design of Experiments in Army Research, Development and Testing. It contains most of the papers presented at this meeting. These articles treat various Army Statistical and design problems.			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION	
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Francis G. Dressel		22b. TELEPHONE (Include Area Code) 919-549-4319	22c. OFFICE SYMBOL AMXRO-MCS

DD Form 1473, JUN 86

Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

★ U.S GOVERNMENT PRINTING OFFICE: 1994 530-520