AD-A241 628

DTIC
ELECTE
OCT 1 1991
S C D

ARO Report 91-2

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

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.

Sponsored by
The Army Mathematics Steering Committee
on Behalf of

THE CHIEF OF RESEARCH, DEVELOPMENT AND ACQUISITION

91-11979

DTIC
COPY
INSPECTED

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

U.S. Army Research Office

Report No. 91-2

May 1991

PROCEEDINGS OF THE THIRTY-SIXTH CONFERENCE OF THE DESIGN OF EXPERIMENTS

Sponsored by the Army Mathematics Steering Committee

HOSTS

The Army Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland
and
The University of Delaware
Newark, Delaware

17-19 October 1990

HELD AT

The University of Delaware
Newark, Delaware

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

FOREWORD

The Thirty-Sixth Conference of the Design of Experiments in Army Research, Development and Testing was held 17-19 October 1990 on the campus of the University of Delaware. This university served as one of its hosts, the other host being the Ballistic Research Laboratory (BRL). Back in 1987 the Thirty-Third Design Conference had these same hosts. Professor Henry B. Tingey was the Chairperson on Local Arrangements for the University and Mr. Jerry Thomas served in this capacity for BRL. The members of the Army Mathematical Steering Committee (AMSC), sponsor of these conferences, take this opportunity to thank these gentlemen for their excellent handling of the many problems associated with this meeting.

The original format for the Design of Experiments Conferences, which are under the auspices of the AMSC, was outlined by the eminent statistician, Professor Samuel S. Wilks, who served as conference chairman until his death. Through these symposia the AMSC hopes to introduce and encourage the use of the latest statistical and design techniques into the research, development and testing conducted by the Army's scientific and engineering personnel.

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

KEYNOTE ADDRESS

Design of Experiments for Comparing the Performance of Several Multi-Stage Procedures for Selecting the Normal Population Having the Largest Mean When the Populations Have a Common Variance.

Professor Robert E. Bechhofer
Cornell University

Professor David M Goldsman
Georgia Institute of Technology

Speaker and Affiliation	Title of Address
Dr. Paul A. Turkey Bell Communication Research	Some Graphical Techniques for Data Analysis
Professor Howard M. Taylor University of Delaware	Stochastic Models for Particals in a Moving Fluid
Professor Erik V. Nordheim University of Wisconsin	Statistical Consulting
Professor Arthur E. Hoerl University of Delaware	Ridge Regression in Practice

Each year, two days before the start of the conference, a two-day tutorial is presented. This year Professor Russell R. Barton, now at The Pennsylvania State University, presented a tutorial entitled "Graphical Design of Experiments." It was held on the campus of the University of Delaware. His notes on this meeting are being published in these proceedings.

Since no U.S. Army Wilks Award would be given in 1990, the hosts decided to have an invited speaker after the banquet. Professor Tingey invited Dr. John C. Bailar, Department of Health and Human Services, to talk on "John C. Bailar's Laws of Data Analysis." It turned out to be one of those rare talks that all members in the audience could enjoy.

The AMSC has requested that these transactions be published and distributed Army-wide so that the information in them might assist Army scientists with some of their statistical problems. Committee members would like to thank all the speakers for their interesting presentations and also members of the Program Committee for their many contributions to this scientific meeting.

PROGRAM COMMITTEE

Gerald Andersen	Barry Bodt	Carl Bates
J. Robert Burge	Eugene Dutoit	Jock Grynovicki
Carl Russell	Douglass B. Tang	Malcolm Taylor
Jerry Thomas		Henry Tingey

TABLE OF CONTENTS*

Title	Page
Foreword	iii
Table of Contents	v
Agenda	vii
DESIGN OF EXPERIMENTS FOR COMPARING THE PERFORMANCES OF SEVERAL MULTI-STAGE PROCEDURES FOR SELECTING THE NORMAL POPULATION HAVING THE LARGEST MEAN WHEN THE POPULATIONS HAVE A COMMON VARIANCE Robert E. Bechhofer and David Goldsman	1
ARTILLERY COMPUTER METEOROLOGICAL MESSAGE ZONE THICKNESS FOR ALTITUDES ABOVE 20 KM Abel J. Blanco	7
A RANK CORRELATION APPROACH FOR TREND DETECTION OF MILITARY SPARE PARTS DEMAND DATA Barnard H. Bissinger and John R. Boyarski	21
ABRAMS TANK INSPECT AND REPAIR ONLY AS NECESSARY (IRON) ECONOMIC MODEL Albert W. Van Horn	61
A MODEL FOR OPTIMALLY REDUCING UNCERTAINTY Andrew Anderson Thompson III	73
AN ITERATIVE TECHNIQUE FOR TARGET DETECTION AND SEGMENTATION IN IR IMAGING SYSTEMS Duc M. Nguyen	87
COMPARISON DENSITY UNIFICATION OF STATISICAL METHODS FOR CONTINUOUS AND DISCRETE DATA Emanuel Parzen	103
DEVELOPMENT OF DESSERT CAMOFLAGE NETS FOR SAUDI ARABIAN NATIONAL GUARD (SANG) George Anitole, Ronald L. Johnson, and Christopher J. Neubert	111
MATERIAL INCREMENTAL OPTIMALITY WHEN BUILDING SHORTEST EUCLIDEAN TOURS T.M. Cronin	125

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

TABLE OF CONTENTS (continued)

Title	Page
AN ALGEBRAIC DERIVATION OF VARIANCE OF THE GEOMETRIC DISTRIBUTION Richard M. Brugger	143
A LINEAR PROGRAMMING MODEL FOR QUEUEING IN OPERATIONAL AVAILABILITY William C. Hoffman	147
THE MAKING AND USE OF THE "BIG MAC" DATA BASE Fred M. Grimes and John Riemenschneider	153
MODEL SENSITIVITY IN STRESS-STRENGTH RELIABILITY COMPUTATIONS Donald M. Neal, William T. Matthews, and Mark G. Vangel	165
IMPROVING NONPARAMETIC TOLERANCE LIMITS FROM POOLED DATA Donald M. Neal, Mark G. Vangel, and Trevor D. Rudalevige	187
TARGET PRIORITIZATION TO OPTIMIZE EXPECTED UTILITY BASED ON RANDOM FIRE Ann E.M. Brodeen and Douglas H. Frank	223
A PERFORMANCE MODEL FOR A SYSTEM USING RANGE AND ANGLE OF ARRIVAL INFORMATION Andrew Anderson Thompson III	235
RELIABILITY DESIGN PROCEDURES FOR FLEXIBLE PAVEMENTS Yu T. Chou	249
NIMBLE CONSULTANT'S ARTICLES OF ADVICE Rick Nordheim	281
THE SURVIVAL PROBABILITY FUNCTION OF A TARGET MOVING ALONG A STRAIGHT LINE IN A RANDOM FIELD OF OBSCURING ELEMENTS S. Zacks and M. Yadin	283
GRAPHICAL METHODS FOR EXPERIMENT DESIGN Russell R. Barton	297
MAILING LIST	381

AGENDA

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

17-19 October 1990

Hosts: **The Army Ballistic Research Laboratory**
Aberdeen Proving Ground, Maryland

and

The Department of Mathematical Sciences
The University of Delaware
Newark, Delaware

Location: **The University of Delaware**

Wednesday, 17 October 1990

0815 - 0915	**REGISTRATION** - Clayton Hall Lobby
0915 - 0930	**CALL TO ORDER** - Henry B. Tingey, University of Delaware
	OPENING REMARKS - David Roselle, President of the University of Delaware
0930 - 1200	**GENERAL SESSION I**
	Chairperson: Jerry Thomas, U.S. Army Ballistic Research Laboratory
0930 - 1030	**KEYNOTE ADDRESS:**
	Robert Bechhofer, Cornell University
1030 - 1100	**BREAK**
1100 - 1200	SOME GRAPHICAL TECHNIQUES FOR DATA ANALYSIS Paul A. Tukey, Statistical Research Group, Bell Communication Research
1200 - 1300	**LUNCH**

Wednesday (Continued)

1300 - 1500 **CLINICAL SESSION**

Chairperson: Malcolm Taylor, U.S. Army Ballistic Research Laboratory

Panelists: Robert Bechhofer, Cornell University
Arthur E. Hoerl, University of Delaware
Emanuel Parzen, Texas A&M University

A RANK CORRELATION APPROACH FOR TREND DETECTION OF MILITARY SPARE PARTS DEMAND DATA
Barnard H. Bissinger, Hershey Foods Corporation, and John R. Boyarski, Operations Analysis Studies, U.S. Naval Ships Parts Control Center

ARTILLERY COMPUTER METEOROLOGICAL MESSAGE ZONE THICKNESS FOR ALTITUDES ABOVE 20 km
Abel J. Blanco, U.S. Army Atmospheric Sciences Laboratory

ABRAMS TANK INSPECT AND REPAIR ONLY AS NECESSARY (IRON) ECONOMIC MODEL
Albert Van Horn, U.S. Army Tank-Automotive Command

1500 - 1530 **BREAK**

1530 - 1710 **TECHNICAL SESSION 1**

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

A MODEL FOR THE MAXIMUM REDUCTION OF UNCERTAINTY
Andrew Thompson, U.S. Army Ballistic Research Laboratory

AN ITERATIVE TECHNIQUE FOR TARGET DETECTION AND SEGMENTATION IN IR IMAGING SYSTEMS
Duc M. Nguyen, U.S. Army Electronics Command, Center for Night Vision and Electro-Optics

DATA ANALYSIS BY INFORMATION STATISTICS
Emanuel Parzen, Texas A&M University

1830 - 1930 **CASH BAR** - Clayton Hall

Wednesday (Continued)

1930 - 2130 **BANQUET AND INVITED SPEAKER**

BAILAR'S LAWS OF DATA ANALYSIS
John C. Bailar, Office of Disease Prevention and Health Promotion, Department of Health and Human Services

Thursday, 18 October 1990

0830 - 1000 **TECHNICAL SESSION 2**

Chairperson: Jock Grynovicki, U.S. Army Human Engineering Laboratory

DEVELOPMENT OF DESERT CAMOUFLAGE NETS FOR SAUDI ARABIAN NATIONAL GUARD (SANG)
George Anitole and Ronald L. Johnson, U.S. Army Belvoir Research, Development and Engineering Center

MAINTAINING INCREMENTAL OPTIMALITY WHEN BUILDING SHORTEST EUCLIDEAN TOURS
Terence M. Cronin, U.S. Army CECOM Center for Signals Warfare

AN ALGEBRAIC DERIVATION OF THE VARIANCE OF THE GEOMETRIC DISTRIBUTION
Richard M. Brugger, U.S. Army Armament, Munitions, and Chemical Command

A LINEAR PROGRAMMING MODEL FOR QUEUING IN OPERATIONAL AVAILABILITY
William C. Hoffman, Dyncorp, Electromagnetic Environmental Test Facility

1000 - 1030 **BREAK**

1030 - 1200 **TECHNICAL SESSION 3**

Chairperson: Fred M. Grimes, TEXCOM Combined Arms Test Center

MODEL SENSITIVITY IN STRESS-STRENGTH RELIABILITY COMPUTATIONS
Donald M. Neal, William T. Matthews, and Mark G. Vangel, U.S. Army Materials Technology Laboratory

IMPROVING NONPARAMETRIC TOLERANCE LIMITS FROM POOLED DATA
Donald M. Neal, Mark G. Vangel, and Trevor D. Rudalevige

TARGET PRIORITIZATION TO OPTIMIZE EXPECTED UTILITY BASED ON RANDOM FIRE
Ann E.M. Brodeen, U.S. Army Ballistic Research Laboratory and Douglas H. Frank, Indiana University of Pennsylvania

AN APPROACH TO THE CONSTRUCTION OF SEARCH DESIGNS
Sanjay Arora and Parvatha Manikkalingan, Howard University

1200 - 1330 **LUNCH**

1330 - 1440 **APPLICATION SESSION/TECHNICAL SESSION 4**

Chairperson: Eugene Dutoit, U.S. Army Infantry School

A PERFORMANCE MODEL FOR A SENSOR SYSTEM USING RANGE AND ANGLE OF ARRIVAL INFORMATION
Andrew Thompson, U.S. Army Ballistic Research Laboratory

PROBABILISTIC DESIGN OF FLEXIBLE PAVEMENTS
Y. T. Chou, U.S. Army Engineer Waterways Experiment Station

BOOLEAN FACTOR ANALYSIS
Lidia Rejto, University of Delaware

1440 - 1510 **BREAK**

1510 - 1710 **GENERAL SESSION II**

Chairperson: Carl Bates, U.S. Army Concepts Analysis Agency

STOCHASTIC MODELS FOR PARTICLES IN A MOVING FLUID
Howard M. Taylor, University of Delaware

STATISTICAL CONSULTING
Erik V. Nordheim, University of Wisconsin

Friday, 19 October 1990

0830 - 1015 **TECHNICAL SESSION 5**

Chairperson: Barry Bodt, U.S. Army Ballistic Research Laboratory

THE SURVIVAL PROBABILITY FUNCTION OF A TARGET MOVING ALONG A STRAIGHT LINE IN A RANDOM FIELD OF OBSCURING ELEMENTS
Shelemyahu Zacks, State University of New York at Binghamton and Micha Yadin, Israel Institute of Technology

NONNEGATIVE ESTIMATION OF VARIANCE COMPONENTS
Bimal K. Sinha, University of Maryland

IMPROVED ESTIMATORS OF VARIANCE COMPONENTS HAVING SMALLER PROBABILITY
Thomas Mathew, University of Maryland and Robert J. Kelly, Bendix Communications Division

1015 - 1045 **BREAK**

1045 - 1200 **GENERAL SESSION III**

Chairperson: Douglas B. Tang, Walter Reed Army Institute of Research and Chairman of the AMSC Subcommittee on Probability and Statistics

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

RIDGE REGRESSION IN PRACTICE
Arthur E. Hoerl, University of Delaware

ADJOURN

PAPER BY TITLE (It came too late to be on agenda)
THE MAKING AND THE USE OF THE "BIG MAC" DATA BASE
Fred M. Grimes and John Riemenschneider, TEXCOM Combined Arms Test Center

PROGRAM COMMITTEE

Gerald Andersen	Barry Bodt	Carl Bates
J. Robert Burge	Eugene Dutoit	Jock Grynovicki
Carl Russell	Douglas B. Tang	Malcolm Taylor
Jerry Thomas		Henry Tingey

DESIGN OF EXPERIMENTS FOR COMPARING THE PERFORMANCES OF SEVERAL MULTI-STAGE PROCEDURES FOR SELECTING THE NORMAL POPULATION HAVING THE LARGEST MEAN WHEN THE POPULATIONS HAVE A COMMON VARIANCE*

Robert E. Bechhofer
School of Operations Research &
Industrial Engineering
College of Engineering, Cornell University
Ithaca, NY 14853

David M. Goldsman
School of Industrial & Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Abstract

We study the performance characteristics of indifference-zone procedures for selecting the normal population which has the largest mean when the populations have a common known variance. The procedures considered are the open sequential procedure of Bechhofer, Kiefer and Sobel and a truncated version of that procedure by Bechhofer and Goldsman, the closed multi-stage procedure of Paulson which eliminates populations indicated as being non-contending, and an improved version of that procedure by Hartmann. The performance characteristics studied are the achieved probability of a correct selection, the expected number of stages required to terminate experimentation, and the expected total number of observations required to terminate experimentation. All performance characteristics are estimated by Monte Carlo sampling. In addition, the same problem is considered for the case of common unknown variance. Here the competing procedures are the open non-eliminating two-stage procedure of Bechhofer, Dunnett and Sobel, the open eliminating two-stage procedure of Gupta and Kim, and the open eliminating multi-stage sequential procedure of Paulson as improved by Hartmann. Particular emphasis is placed on the fact that when designing the experiments to compare the performance characteristics of these procedures there are many relevant factors which must be varied in the conduct of the experiment in order that meaningful and generalizable results can be obtained.

Research partially supported by the U.S. Army Research Office through the Mathematical Sciences Institute of Cornell University.

Bechhofer (1954) proposed a statistical procedure for selecting the normal population which has the largest population mean with prespecified controls over the probability of a correct selection. Over the years many additional procedures have been proposed to solve this problem. All of these procedures guaranteed the following indifference-zone probability requirement:

$$\text{Prob}\ \{\text{Correct selection}\} \geq P^* \quad \text{whenever} \quad \mu_{[k]} - \mu_{[k-1]} \geq \delta^*.$$

Here $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$ are the ordered values of the population means μ_i $(1 \leq i \leq k)$; the values of the μ_i and $\mu_{[j]}$ $(1 \leq i, j \leq k)$ are assumed to be unknown as is the pairing of the $\mu_{[j]}$ $(1 \leq j \leq k)$ with the k populations. The constants $\{\delta, P^*\}$ with $0 < \delta^* < \infty$, $1/k < P^* < 1$ are specified by the experimenter prior to the start of experimentation.

Each of these procedure at the time that it was suggested possessed certain virtues which made it worthy of consideration. But it was not until recent years that any serious attempt was made to compare the relative merits of these procedures. The ultimate objectives of such a study would be to determine whether any particular procedure dominated any other procedure(s), and if so, in what respect and under what conditions. Such comparative studies could then make it possible to recommend which procedure it would be preferable to use in certain particular environments.

The first such comprehensive study was reported by Bechhofer and Goldsman (1989); in that study which dealt with the cases in which the populations have a common <u>known</u> variance, the critical performance characteristics were compared for the single-stage procedure of Bechhofer (1954), the two-stage procedure of Tamhane and Bechhofer (1977, 1979) in which populations indicated as being non-contending can be eliminated after the first stage, the open non-eliminating sequential procedure of Bechhofer, Kiefer and Sobel (1968), and the truncated version of that procedure by Bechhofer and Goldsman (1987, 1989), and the closed multi-stage procedure of Paulson (1964) in which population can be eliminated at any stage after the first, and an improved version of that procedure by Hartmann (1988). Most recently Bechhofer, Dunnett, Goldsman and Hartmann (1990) studied the same selection

problem for the case of common unknown variance. The procedures considered were the open non-eliminating two-stage procedure of Bechhofer, Dunnett and Sobel (1954), the open two-stage procedure of Gupta and Kim (1984) in which non-contending population can be eliminated after the first stage, and the open multi-stage procedure of Paulson (1964) as improved by Hartmann (1991) in which population can be eliminated at every stage after the first.

In the talk presented, results and conclusions for the case of common known variance were discussed in great detail. It was pointed out why the case of common unknown variance introduced special design problems in terms of the large factorial experiment to be conducted. The final factorial experiment as it was conducted is described below. The results of the experiment are reported in Bechhofer, Dunnett, Goldsman and Hartmann (1990).

Design of a Large Factorial Experiment to Compare the Relative Performances of Four Multi-Stage Procedures for Selecting the Population with the Largest Mean when the Common Variance is Unknown

The procedures:

a) Bechhofer-Dunnett-Sobel (1954)
b) Paulson-Hartmann with $\lambda = \delta^*/4$ (1991)
c) Paulson-Hartmann with $\lambda = \delta^*/2$ (1991)
d) Gupta-Kim (1984)

New underlying factor:

The value of the unknown variance which plays its role in terms of the ratio σ/δ^* where $\delta^* > 0$ is specified.

New design factor:

This is under the control of the experimenter. Common number of observations (n_1) per population which is taken in the first stage.

Factors affecting the performances of each procedure:

a) Number of populations: These were set at three levels: k=3,5,10.
b) Specified P^*-values: These were set at 2 levels: P^*= 0.75, 0.90.
c) Ratios σ/δ^*: These were set at two levels: σ/δ^*=2,3.
d) First-stage sample sizes: These were set at four levels: n_1 = 5, 10, 15, 20.
e) Configuration of the population means: These were set at two levels: least-favorable and equal means.

Thus for each procedure a 5-factor experiment was conducted with $3\times2\times2\times4\times2 = 96$ factor-level combinations. For each combination at least 10,000 independent MC experiments were conducted. The responses recorded for each experiment were:

a) The achieved probability of a correct selection.
b) The estimated expected total number of observations to terminate sampling.
c) The estimated variance of the total number of observations to terminate sampling.
d) The estimated expected number of vector-observations to terminate sampling.
e) The estimated variance of the number of vector-observations to terminate sampling.

References

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

Bechhofer, R.E., Dunnett, C.W. and Sobel, M. (1954). "A two-stage multiple-decision procedure for ranking means of normal populations with a common unknown variance," Biometrika 41, 170-176.

Bechhofer, R.E., Dunnett, C.W., Goldsman, D.M. and Hartmann, M. (1990). "A comparison of the performances of procedures for selecting the normal population having the largest mean when the populations have a common unknown variance," Comm. Statist. -- Simula. Computa. 19 (3), 971-1006.

Bechhofer, R.E. and Goldsman, D.M. (1987). "Truncation of the Bechhofer-Kiefer-Sobel sequential procedure for selecting the normal population which has the largest mean," Comm. Statist. -- Simula. Computa. 16 (4), 1067-1091.

Bechhofer, R.E. and Goldsman, D.M. (1989). "Truncation of the Bechhofer-Kiefer-Sobel sequential procedure for selecting the normal population which has the largest mean (III): supplementary truncation numbers and resulting performance characteristics," Comm. Statist. -- Simula. Computa. 18(1), 63-81.

Bechhofer, R.E. and Goldsman, D.M. (1989). "A comparison of the performances of procedures for selecting the normal population which has the largest mean when the variances are known and equal," Contributions to Probability and Statistics -- Essays in Honor of Ingram Olkin (Eds. L.J. Gleser, M.D. Perlman, S.J. Press, A.R. Sampson). Springer-Verlag, New York, 303-317.

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

Gupta, S.S. and Kim, W.C. (1984). "A two-stage elimination-type procedure for selecting the largest of several normal means with a common unknown variance," Design of Experiments: Ranking and Selection. (Eds. T.J. Santner and A.C. Tamhane), Marcel Dekker, N.Y., 77-94.

Hartmann, M. (1988). "An improvement on Paulson's sequential ranking procedure," Sequential Analysis 7(4), 363-372.

Hartmann, M. (1991). "An improvement on Paulson's procedure for selecting the population with the largest mean from k normal populations with a common unknown variance," Sequential Analysis. To appear.

Paulson, E. (1964). "A sequential procedure for selecting the population with the largest mean from k normal populations," Ann. Math. Statist. 35, 174-180.

Tamhane, A.C. and Bechhofer, R.E. (1977). "A two-stage minimax procedure with screening for selecting the largest normal mean," Comm. Statist. - Theor. Meth. A6, 1003-1033.

Tamhane, A.C. and Bechhofer, R.E. (1979). "A two-stage minimax procedure with screening for selecting the largest normal mean (II): A new PCS lower bound and associated tables," Comm. Statist. -- Theor. Meth. A8, 337-358.

ARTILLERY COMPUTER METEOROLOGICAL MESSAGE ZONE THICKNESS FOR ALTITUDES ABOVE 20 KM

Abel J. Blanco
U.S. Army Atmospheric Sciences Laboratory
White Sands Missile Range, NM 88002-5501

ABSTRACT. The U.S. Army is developing new field artillery systems that have longer range (50 km) capabilities. As a result, extended altitude computer meteorological (met) messages are now required for aiming adjustments that compensate for met effects along the high apogee trajectory. The following experiment is designed to select an optimal zone thickness: aerodynamic parameters for an artillery rocket assisted round are assumed and the Ballistic Research Laboratory general trajectory model is used to simulate impacts computed from using observed met data and observed data averaged into proposed zone thicknesses for the altitudes extending above the current 20 km computer met message height. Measured upper-air wind, temperature, and pressure data collected during the four seasons at White Sands Missile Range, New Mexico, are converted for use in artillery surface-to-surface applications. The 226 rawinsonde flights, including data up to 30 km height, are partitioned into monthly data sets, and tests of significance are used to determine an optimal thickness. The corresponding impact paired differences between the observed data and the proposed zone averaged data yield statistics that demonstrate the 1-km zone average as the optimally selected thickness for extending the computer met message.

1. INTRODUCTION. The U.S. Army Atmospheric Sciences Laboratory (ASL) continues to support the U.S. Army Field Artillery School (USAFAS) in extending the application of the artillery computer meteorological (met) message at altitudes above 20 km. In lieu of unavailable measured data, a software technique (Blanco, 1981) was developed to extend the application to 3 km from the maximum ordinate. By 1983, ASL refined a composite algorithm allowing use of estimated higher altitude met data. Persistence of the last computer met message line data and satellite climatology or available artillery fallout met message were merged into an automated procedure that selects the best options for extending the artillery computer met message. The USAFAS now has a new requirement of extending the computer met message to a maximum ordinate of 30 km. The 10 km extension dictates using measured data instead of the early proposals of using estimated data. The only constraint is that one can easily overload the fire control computer with too much met data. The current met array allows 27 observations for each met parameter--wind direction, windspeed, virtual temperature, and pressure. These 27 lines remain the same and the new fire control computer storage requirements are determined upon selecting the optimal zone thickness that best describes the atmospheric state between 20 km and 30 km above the surface. The only expected change in the field operations is to continue collecting met data up to 30 km above the surface. All other required software changes will be transparent to the user.

This paper presents an optimal zone thickness for the extended artillery computer met message at altitudes above 20 km. The method used in selecting the zone thickness that can be standardized for artillery applications includes the following experiment: aerodynamic parameters for

an artillery rocket assisted round are assumed* and the Ballistic Research Laboratory (BRL) general trajectory model (Lieske and Reiter, 1966) is used to simulate impacts computed from using observed met data and observed data averaged into proposed zone thicknesses for the altitudes extending above the current 20 km computer met message height. The corresponding impact paired statistics between the observed and proposed zone thicknesses are used to document the selection of the optimal zone thickness.

The analysis of these results includes mean component inferences. The Student t-test and the F-ratio analysis of variance comparison are used to determine a significant difference between the observed and the 1-km zone thickness and between the observed and the 2-km zone thickness. A mean vector inference test known as the Hotelling's T^2 is also used for an easier interpretation of results. This test also allows faster calculations when using available utilities from the International Mathematics and Science Library (IMSL). In summary these statistics demonstrate the 1-km zone average as the optimally selected thickness for extending the computer met message. With this zone thickness, 10 new lines are added to the computer met message.

2. DESIGN OF EXPERIMENT. The U.S. Army Field Artillery is developing new weapon systems that have a longer range (assume 50 km) capability. To reach this new range the projectile must go through a higher apogee. The projectile propagates through more atmosphere and spends a longer time under the influence of atmospheric parameters of wind, temperature, and pressure. Since met effects are known to be major contributors to the artillery error budget, deployment of met teams in the battle area is required to measure the atmospheric parameters, prepare formatted messages, and disseminate computer met messages to the fire control centers. These messages are used in making final aiming angle adjustments before ordering fire for effect.

Figure 1 presents a two-dimensional plot of height versus range for the two modes of fire--low quadrant angle and high quadrant angle. New aerodynamic and ballistic coefficients at the higher mach numbers were incorporated in the BRL general trajectory model to simulate firings of a modified M549A1 rocket assisted round. The assumed future configuration of a 155-mm weapon system fired at White Sands Missile Range, New Mexico, did not reach an apogee of 30 km. For the low-angle simulations, the projectile apogee is at 23 km above the surface. For the high-angle simulations, the projectile apogee is at 26 km. Of course the simulation can perform the 30 km apogee; but at this higher angle of fire the BRL test on maximum angle of attack exceeds a predefined criterion, which indicates that the projectile is not trailing the trajectory and there is the possibility of actual round tumble. Following the BRL criteria, this report evaluates results from simulations containing 23 to 26 km apogees corresponding to the low and high quadrant angles.

---

*Robert Lieske, 1990, Memorandum SCLBR-LF-T (340), "Maximum Ordinates for Extended Range Ordnance," U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland.

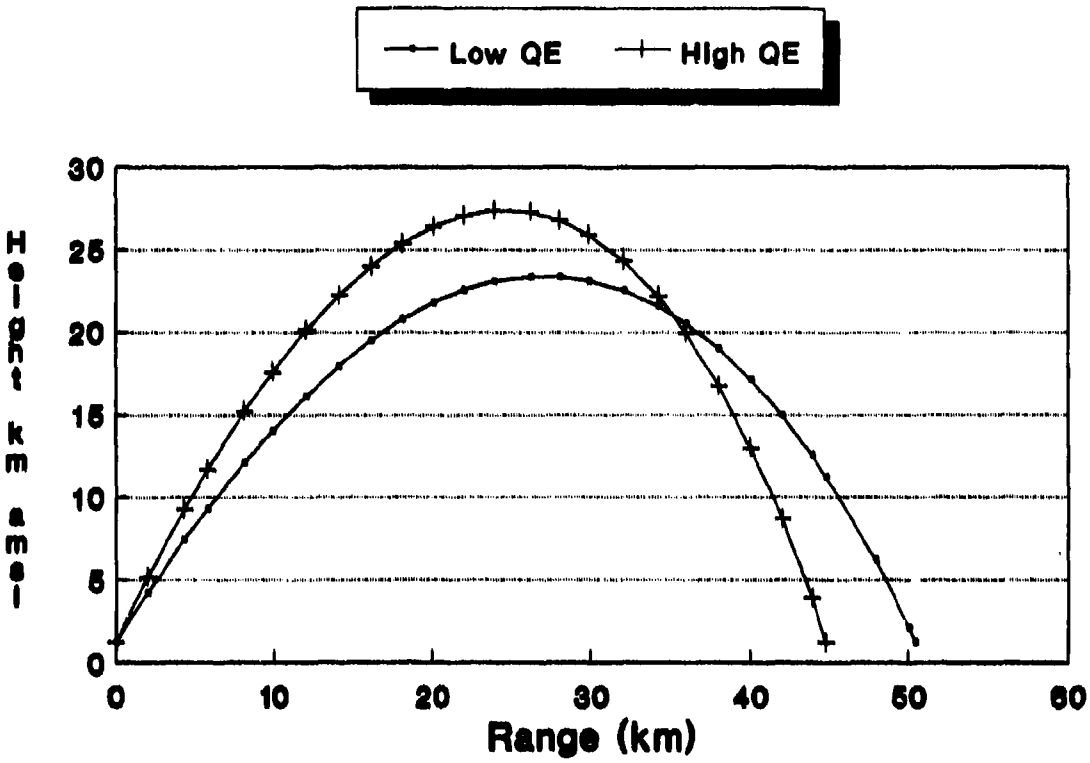


Figure 1. Assumed 155-mm cannon projectile trajectory.

All effective fire includes met adjustments, and table 1 presents the zone thicknesses for three artillery met messages. Today's artillery meteorology still depends on a balloon-borne sensor to provide the appropriate wind, temperature, and pressure measurements. In computing the fallout message for example, the balloon is tracked up to 30 km above the surface. The current computer met message provides met data up to only 20 km because cannon artillery has not exceeded this maximum ordinate. Special artillery rockets have the capability of exceeding this altitude, and the USAFAS has requested methods of extending the application of the available computer met message. To this date the software estimations have been appropriate only because the required extension was within 4 km of the last observed computer line data. With the advent of future artillery systems, there is a need to extend the computer met message to a maximum ordinate of 30 km. Met hardware and software are already available and the only task is to standardize the zone thickness within the 20 to 30 km altitudes.

This task has been a cooperative effort between the UASFAS, BRL, and ASL. ASL recommended all 1-min observations; however, the USAFAS is constrained to available computer storage space in the fire control centers and cannot support this ~300 m thickness. BRL has assumed aerodynamic characteristics and provided ASL with the point mass trajectory model for use in computing the desired long-range cannon fire. An experiment is then designed to prepare met messages that contain 1-min observed data, 1-km zone averages, and 2-km zone averages. These formatted data are then input to the BRL point mass general trajectory model for simulating impacts. All replicates are then paired and statistical results are interpreted to determine if proposed averaged data and the observed data are significantly different.

TABLE 1. FM 6-15 ATMOSPHERIC ZONE STRUCTURE

HEIGHT METERS	LINE NUMBERS		
	BALLISTIC	COMPUTER	FALLOUT
SURFACE	0	0	0
200	1	1	
500	2	2	
1,000	3	3	1
1,500	4	4	
2,000	5	5	
2,500		6	
3,000	6	7	2
3,500		8	
4,000	7	9	
4,500		10	
5,000	8	11	3
6,000	9	12	
7,000		13	
8,000	10	14	4
9,000		15	
10,000	11	16	5
11,000		17	
12,000	12	18	6
13,000		19	
14,000	13	20	7
15,000		21	
16,000	14	22	8
17,000		23	
18,000	15	24	9
19,000		25	
20,000		26	10
* * *			* * *
30,000			15

3. METEOROLOGICAL DATA BASE. An available set of upper-air data containing observations up to 30 km was retrieved from met support provided for special projects at White Sands Missile Range, New Mexico, during 1989. A sample of 226 rawinsonde flights containing representative sets for each of the four seasons is used as the met data base. Table 2 presents all replicate rawinsonde flights with Holloman, White Sands, and Small Missile Range containing the larger samples.

TABLE 2. NUMBER OF RAWINSONDE FLIGHTS COLLECTED DURING 1989

	MONTHLY FLIGHTS												
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	TOT
HOLLOMAN	4	2	3	6	12	3	4	5	8	7	4	6	64
WHITE SANDS	8	2	6	1	6	3	6	5	8	10			55
JALLEN		1		1	7	6							15
STALLION				1	1			3	1		1	1	8
SMR	4	2	7	1	5	5		15	9	8	3	1	60
NW 30			1		1					1			3
TULA SITE	1	1						1	2	1	1		7
OASIS					2	6	2	1			2		12
LAS CRUCES			1										1
TOTAL	17	8	18	10	34	23	12	30	28	27	11	8	
	WINTER			SPRING			SUMMER				FALL		
	43			66			71				46		226

Each rawinsonde flight in this data base was reduced into the 26 computer zone averages by using a linear fit methodology. Given the particular data (W equals wind, temperature, or pressure) versus height, one can interpolate for any information within the adjacent observed data by using the linear coefficients (a and b):

$$a = \frac{W(i+1) - W(i)}{Z(i+1) - Z(i)}, \qquad b = \frac{W(i)\,Z(i+1) - W(i+1)\,Z(i)}{Z(i+1) - Z(i)}, \tag{1}$$

where the range of application is within Z(i) and Z(i + 1). Then for a given height H between Z(i) and Z(i + 1), one can compute the value of the met parameter W by

$$W = a\,H + b. \tag{2}$$

With this linear coefficient method, one can derive the value of the met parameter at the bottom and top of any defined zone thickness. Figure 2 illustrates the position of the data (o) and the location of the artillery zones (H(i)'s). The total AREA defined by the vertical axis, the bottom horizontal line at height H(1), the top horizontal line at height H(2), and the met parameter function within the interval H(1) and H(2) is calculated by the sum of the composite areas. For example,

$$\int\int dwdh = \int (ah + b)\,dh = \frac{ah^2}{2} + bh \Big|_{H(1)}^{Z(2)} \tag{3}$$

computes the area by integration between H(1) and Z(2). Repeating the interation between Z(2) and Z(3) and Z(3) and H(2) yields the composite

areas, and the sum of these areas defines the total AREA. Using the Mean Value Theorem, one can derive the mean value of the particular parameter within the artillery zone thickness by

$$\bar{W} = \frac{AREA}{\Delta H} \quad . \tag{4}$$

Since the atmospheric pressure is exponentially related to the height, the natural log of the pressure is computed before the linear fit and averaging routines are executed. At the end, the exponential function is applied to the output to express the zone pressure average.

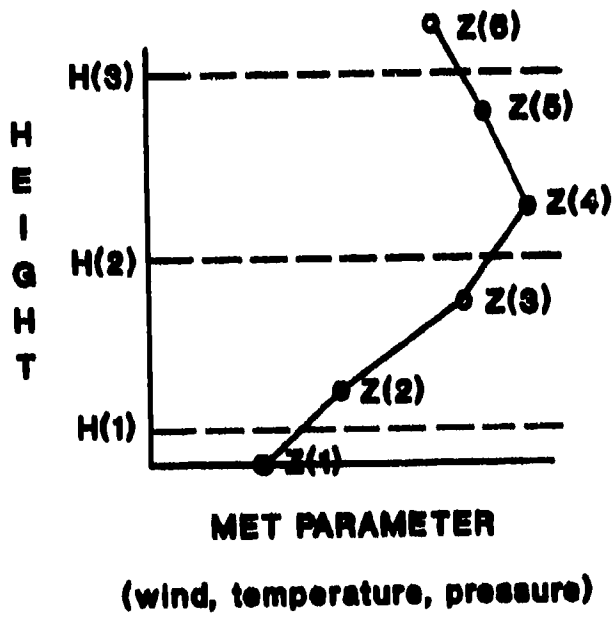


Figure 2. Met data and zone thickness.

In computing the wind zone averages, the artillery interpolates for the x and y balloon positions at the bottom and the top of each artillery zone then divides by the time the balloon ascends from the bottom to the top of the zone. Note from table 1 that the computer zones are not the same thickness. Since actual computer met messages are not available, the best estimate for reproducing the artillery met message from the available rawinsonde data is that described in this section.

Figures 3 through 5 present reduced computer met messages extended to a higher altitude with the actual observed data and the two proposed zone averages. Note that line 0 or the surface is at the Holloman site elevation of 1258 m and the top of line 26 is at 21258 m. In the experiment this structure is fixed and the only variable is the data from the top of line 26 to the last available met observation. These plots present the observed wind component profiles for the December 22 flight at the extended computer met message heights. The observed data represents 1 min or about 300 m interval observations. Using the information in figure 3 as input to the BRL general trajectory model, one can compute a 43506-m range and a 5137-m

cross-component impact for the assumed future projectile fired at a high quadrant angle of 1150 mils. In figure 4, the extended observed data is averaged into 1 km zones, and the simulated impacts demonstrate almost identical results. However, when one averages the observed data into 2 km zones, as represented in figure 5, the simulated impacts demonstrate a -64 m range and +53 m cross difference.

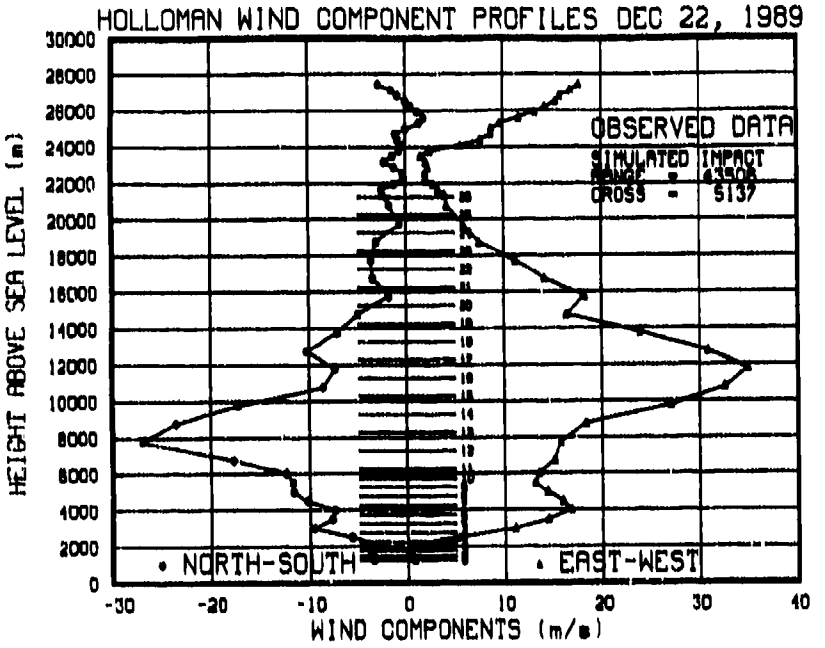


Figure 3. Computer wind components extended with observed data.

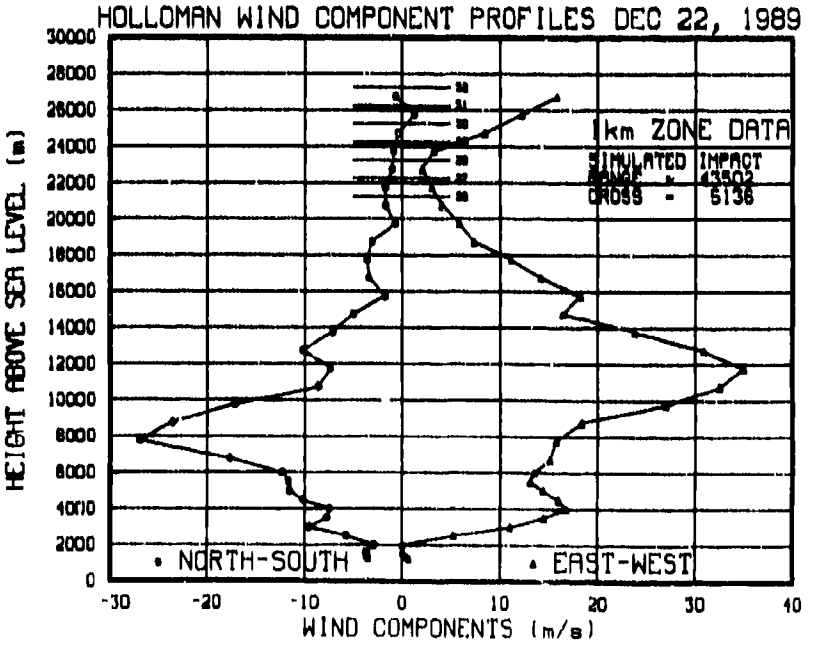


Figure 4. Computer wind components extended with 1 km averaged data.

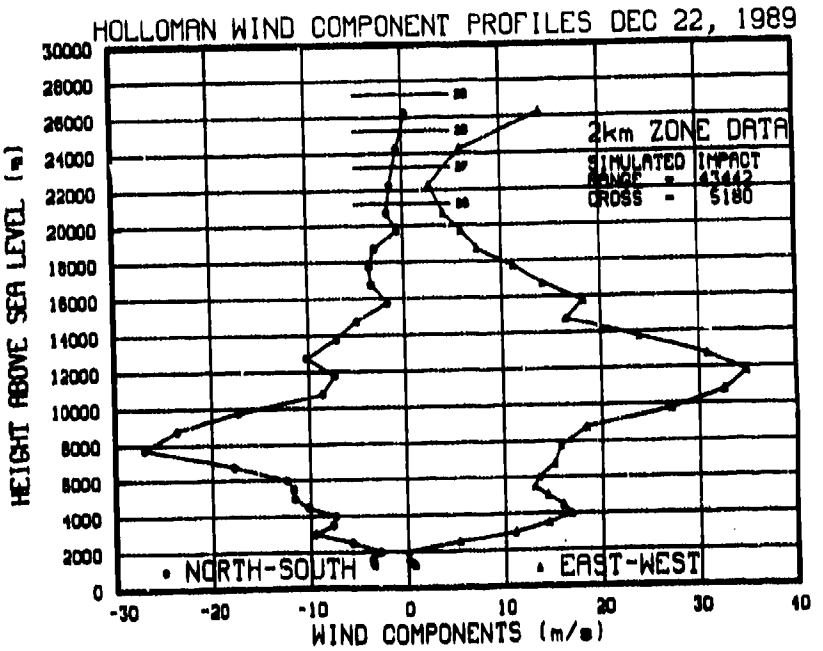


Figure 5. Computer wind components extended with 2 km averaged data.

4. ANALYSIS OF SIMULATED IMPACTS. For the six available Holloman December met messages, one can compute 18 impacts representing the observed, 1 km, and 2 km averaged data. Figure 6a presents these simulated impacts. The impacts derived from using the December 22 averaged profiles presented in the last three figures are now identified with a 6. Note that the dark 6 represents two impacts and the light 6 represents the other. One can see the same general results for the other numbered impacts. For this month the simulated impacts derived from using the observed and 1 km averaged data are almost identical. Pairing the impacts resulting from the observed data with the corresponding impacts from the 1 km and 2 km averaged data, one can see a significant difference in using the proposed zone thicknesses as illustrated in figure 6b. Not all results include this easily derived graphical interpretation; therefore, tests for the mean component inferences were automated (Rickmers and Todd, 1967).

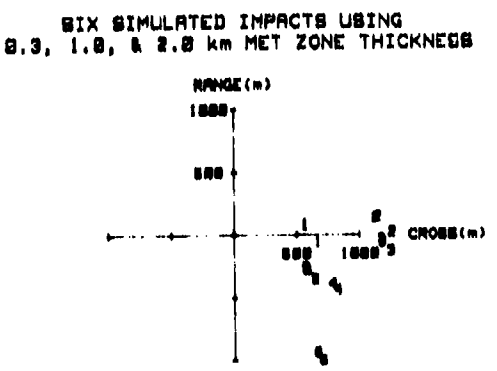


Figure 6a. Simulated impacts.

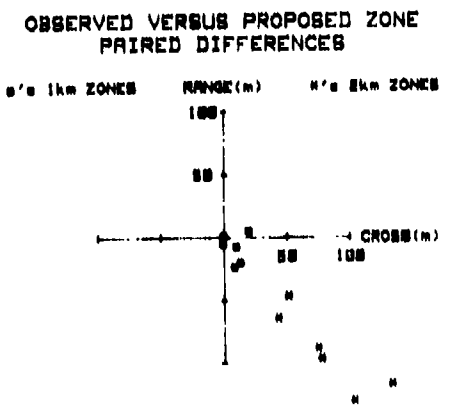


Figure 6b. Paired differences.

The first statistical test is the Student t-test. The paired differences (n, average, and sigma) are used to derive a t-value

$$t_{computed} = \frac{\bar{d} - \delta}{\sigma_{\bar{d}}} \tag{5}$$

where

$\bar{d}$ = average difference between pairs

δ = hypothesized pair difference

$\sigma_{\bar{d}}$ = standard error = $\frac{\sigma_d}{\sqrt{n}}$

In this paper and in general δ is made equal to 0. By comparison with the t-table value, one can determine presence of a significant difference using the 5-percent alpha risk. If the computed value is within the ± table value, then one can accept the null hypothesis that no difference exists.

Another test is the F-ratio analysis of variance. Theoretically the F distribution is the square of the Student t and a conclusion for the two tests is the same. However, the F-ratio analysis of variance provides other information that the Student t does not. In computing the different ratios,

$$F = \frac{\text{variance of means}}{\text{variance for errors}}, \tag{6}$$

a significant difference among the six met rawinsonde flights was revealed. In the selected two-factor setup of the F-ratio test, each observation is assumed to be determined by the following four possible effects: the general mean of the data, a possible zone thickness effect, a possible atmospheric (rawinsonde flights) effect, and the effect of the error in our hypothetical mathematical model of the sum of the variances.

These tests were performed on the component impact differences. Analyses of the results were difficult to interpret because, as illustrated in table 3, the cross-component 1-km average contains the computed t- and F-values outside the corresponding table values. The conclusion is that the 1-km average provides significantly different cross-component impacts than those impacts derived from the observed data. Different conclusions from the individual comparisons for each component impacts were not encouraging. The graphical results in figure 6b indicate that the observed and 1-km zone averaged data provide similar impacts and should not be considered as significantly different. This is certainly true when considering the lethal radius of the delivered munition. To maintain agreement with the graphical results, a third statistical test was automated.

The Hotelling T^2 test uses the mean vector impact differences and allows easier interpretation of the inference results. Table 3 now reveals that there is no significant difference between the observed and the 1 km averaged data for the December flights collected at Holloman. However, the comparison between impacts derived from the observed and the 2 km averaged

data reveal that there is a significant difference. Under these atmospheric conditions and this high quadrant angle of fire, the met data must be formatted into 1 km averages to allow accurate artillery fire.

TABLE 3. INFERENCE TEST RESULTS

USING HQE SIMULATED IMPACTS FOR DEC HOLLOMAN MET DATA

	STUDENT t		F RATIO		HOTELLING T^2
	cross	range	cross	range	vector (F)
TABLE VALUE	2.57	2.57	6.61	6.61	6.94
1 km ZONE	2.71	-1.55	7.33	2.41	3.77
2 km ZONE	5.90	-7.06	34.82	49.83	20.04

Not only are the Hotelling results easier to interpret but this test also allows faster computations by using subroutines from the IMSL. No new software development or table lookup is required. The computed T^2 value is defined (Johnson and Wichern, 1989) as the quadratic form; and knowing the mean and the covariance matrixes, one can derive the result by utilizing the IMSL subroutine, "blinf."

$$T^2 = \text{quadratic form "blinf"} \tag{7}$$

$$T^2 = n\,(\bar{X} - M)^t\, S^{-1}\,(\bar{X} - M) \tag{8}$$

where

$$\underset{(p\times 1)}{\bar{X}} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

$$\underset{(p\times p)}{S} = \frac{1}{n-1}\sum_{j=1}^{n} (X_j - \bar{X})\,(X_j - \bar{X})^t$$

$\underset{(p\times 1)}{M}$ = a plausible value for the mean

p = 2 (the range and cross variables)

X = the matrix of paired impact differences.

The distribution of T^2 has the following F distribution relationship:

$$\frac{(n - 1)\ p}{(n - p)}\ F_{p,n-p}$$

where

n is the number of paired impact differences and
p is 2 representing the range and cross variables.

Solving for F in the above equation, one can compare the computed value with the table value obtained by using the "fin," IMSL subroutine. Another method to determine the presence of a significant difference is to compute the probability = 1 - "fdf" where "fdf" (IMSL subroutine) is the cummulative distribution. The alpha risk is then used as the criteria to demonstrate presents of the null hypothesis, no significant difference.

5. EXPERIMENTAL RESULTS. In general the statistical comparisons yield results from the following categories: (1) during stable weather conditions all proposed zone averages demonstrate agreement with observed data; (2) during unstable weather conditions all proposed zone averages demonstrate disagreement with observed data; and (3) under certain weather conditions one zone average exists that is a better observed data estimator than the other proposed zone average.

The Hotelling results for each month are listed in table 4. Beginning with the Holloman, December, high quadrant angle results which have already been discussed in the previous two sections, let's define the "N" for no significant difference and the "Y" for a significant difference. By reviewing the results in table 3, one can see that the computed F for the 1-km zone is within the ± table values and that the computed F for the 2-km zone is not within the ± table values. Therefore, for the high quadrant angle we note an "N" for the 1-km zone and a "Y" for the 2-km zone during the month of December. For February there are only two rawinsonde flights (n) and since the Hotelling test requires that the number of variables be equal to the degree of freedom (p), the (n - p) computation in the denominator of the Hotelling distribution leads to a division by zero. So the other symbol, "-," in the table indicates that there is not enough data to complete the test. For the other stations, "-" may also indicate that there is no data for completing any mean inference test.

Overall, the majority (30 cases) of the Hotelling mean vector inference test of the paired impacts derived from the met flights averaged into the different zones reveal no significant difference. In the minority (12 cases), particularly for low quadrant angle of fire, none of the proposed zone met averages are representative of the observed data. This observation is true for all months during the year but seems to occur more often at the SMR station, which is located closer to complex terrain. Eight of the remaining 11 cases show that the 1-km-zone averages yield better impact results than the 2-km-zone averages. This further demonstrates that the 1-km-zone averages produce almost identical simulated impacts as those produced by using the 1-min (~300 m) observed met data.

TABLE 4. SIGNIFICANT DIFFERENCE BETWEEN OBSERVED AND PROPOSED ZONE

LOW QUADRANT ANGLE
USING HOL SIMULATED IMPACTS FOR MONTHLY MET DATA

	1	2	3	4	5	6	7	8	9	10	11	12
1 km ZONE	N	-	N	N	N	N	N	Y	Y	N	N	N
2 km ZONE	N	-	N	N	Y	N	N	Y	Y	N	N	N
HIGH QUADRANT ANGLE												
1 km ZONE	N	-	N	N	Y	N	N	N	N	N	N	N
2 km ZONE	N	-	N	N	Y	N	N	N	N	N	Y	Y

USING SMR SIMULATED IMPACTS FOR MONTHLY MET DATA
LOW QUADRANT ANGLE

	1	2	3	4	5	6	7	8	9	10	11	12
1 km ZONE	Y	-	Y	-	Y	Y	-	Y	N	N	N	-
2 km ZONE	Y	-	Y	-	Y	Y	-	Y	Y	Y	N	-
HIGH QUADRANT ANGLE												
1 km ZONE	N	-	N	-	N	N	-	Y	N	N	N	-
2 km ZONE	N	-	N	-	N	N	-	N	Y	N	-	-

USING WSD SIMULATED IMPACTS FOR MONTHLY MET DATA
LOW QUADRANT ANGLE

	1	2	3	4	5	6	7	8	9	10	11	12
1 km ZONE	Y	-	Y	-	N	N	Y	N	Y	N	-	-
2 km ZONE	Y	-	Y	-	Y	N	Y	N	Y	N	-	-
HIGH QUADRANT ANGLE												
1 km ZONE	N	-	N	-	N	Y	N	Y	N	N	-	-
2 km ZONE	Y	-	N	-	N	N	N	N	N	N	-	-

6. SUMMARY AND RECOMMENDATION. From the majority of the 226 rawinsonde flights evaluated, it is revealed that the atmospheric state can be accurately represented by either 1 or 2 km zone averages. Under this condition the artillery may use the already available fallout zone extension to the computer met message. However, this paper also identifies cases where the atmospheric state could not be represented by any of the proposed (1 km or 2 km) zone averages. Under this condition the artillery must use all available observation to allow accurate artillery fire. Since the 1 min (~300 m) would require some 33 new lines to extend the current computer met message to the new height of 30 km, the 300-m-zone thickness is not a possible solution because of the limited fire direction computer storage constraint. The upper minority of cases reveal that the 1 km is representative of the observed met data. This result and the majority of cases that reveal no difference in zone thickness indicate that the 1-km zone is the optimal of the two proposed zone thicknesses.

In conclusion, it is recommended that 10 new lines be required to describe the atmosphere between 20 and 30 km above the surface. Future fire direction computers shall have storage available for 37 lines of met information. Each line shall contain an average value for wind direction, windspeed, virtual temperature, and pressure at the designated height. The extended computer met message zone structure for long-range artillery surface-to-surface fire shall be composed of ten 1-km zones averages.

REFERENCES

Blanco, Abel, 1981, Extending Application of the Artillery Computer Meteorological Message, ASL-TR-0087, U.S. Army Atmospheric Sciences Laboratory, White Sands Missile Range, New Mexico.

Johnson, Richard, and Dean Wichern, 1989, Applied Multivariate Statistical Analysis, Second Edition, Prentice Hall, Englewood Cliffs, NJ, pp 169-205.

Lieske, Robert, and Mary L. Reiter, 1966, Equations of Motion for a Modified Point Mass Trajectory, TR-1314, Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland.

Rickmers, Albert, and Hollis Todd, 1967, Statistics An Introduction, McGraw-Hill Book Company, New York, pp 62-178.

A RANK CORRELATION APPROACH FOR TREND DETECTION OF MILITARY SPARE PARTS DEMAND DATA

Barnard H. Bissinger, Ph.D.

John R. Boyarski
Director, Operations Analysis Studies
Navy Ships Parts Control Center
Mechanicsburg, PA

KEY WORDS: TREND, RANK CORRELATION, DEMAND, SPARE PARTS, KENDALL's S

ABSTRACT

The prediction of future demands for each of over 200,000 demand-based repair parts is crucial to the readiness of the Navy, and certainly other branches of the Armed Services. Item managers with only eight quarters of history are responsible for ordering that which is sufficient to fill material requirements over an average eight quarter procurement leadtime and simultaneously staying within economic bounds and excess material constraints.

Pivotal in demand prediction is the ability to determine whether the demand time series are stable or, if trending, are they increasing or decreasing.

Former methods do not seem to do the job now, and one reason is that the average procurement leadtime is nearly twice what it was 10 years ago.

The present approach has turned to the use of nonparametric analysis, in particular, using Kendall's S (or T). The following factors and levels are considered:

- Window Size: N = 4, 6, 8, 10 sample sizes of recorded demand
- Demand Level: Poisson, low, medium, high, very high average quarterly demand $\bar{x}$
- Ratio of $\hat{\sigma}/\bar{x}$: Five levels dependent on the level of demand

Extensive simulation, as well as actual demand data, are used to estimate the degree of type 1 and type 2 errors and the expected performance of the "S" statistic for various combinations of demand and variability levels. Kendall's "S" does a good job confirming no trend when there is no trend. Also, it is satisfactory to detecting trend when there is trend and discriminating between increasing and decreasing trends.

The weakest link in the chain, so to speak, is detecting the commencement of a trend in demand immediately following a nontrending period, as well as identifying when a trend has terminated and the process generating demand is now stable. A lag of maybe three periods can occur. However, proper selection of window length plays an important role here in minimizing the reaction time.

I. INTRODUCTION. Prior to this research, the Navy had been using a trigger-trend statistic to identify situations where demand was trending; namely,

$$T = \frac{2\ (D_1 + D_2)}{D_1 + D_2 + D_3 + D_4} \tag{1}$$

where D_1, D_2, D_3, and D_4 are, respectively, the last four quarterly demands, D_1 being the most recent. Thus, the trend statistic "T" was simply the ratio of twice the sum of the last two quarters of demand compared to the sum of the

Preceding Page Blank

last four quarters of demand. The value of this statistic could range from a minimum of zero to a maximum of two.

When this statistic's value lay between .9 and 1.1, it was concluded that no appreciable change in the demand pattern had occurred and the last demand, D_1, was exponentially smoothed with the last forecast to compute a new forecast. The nontrending smoothing weights were:

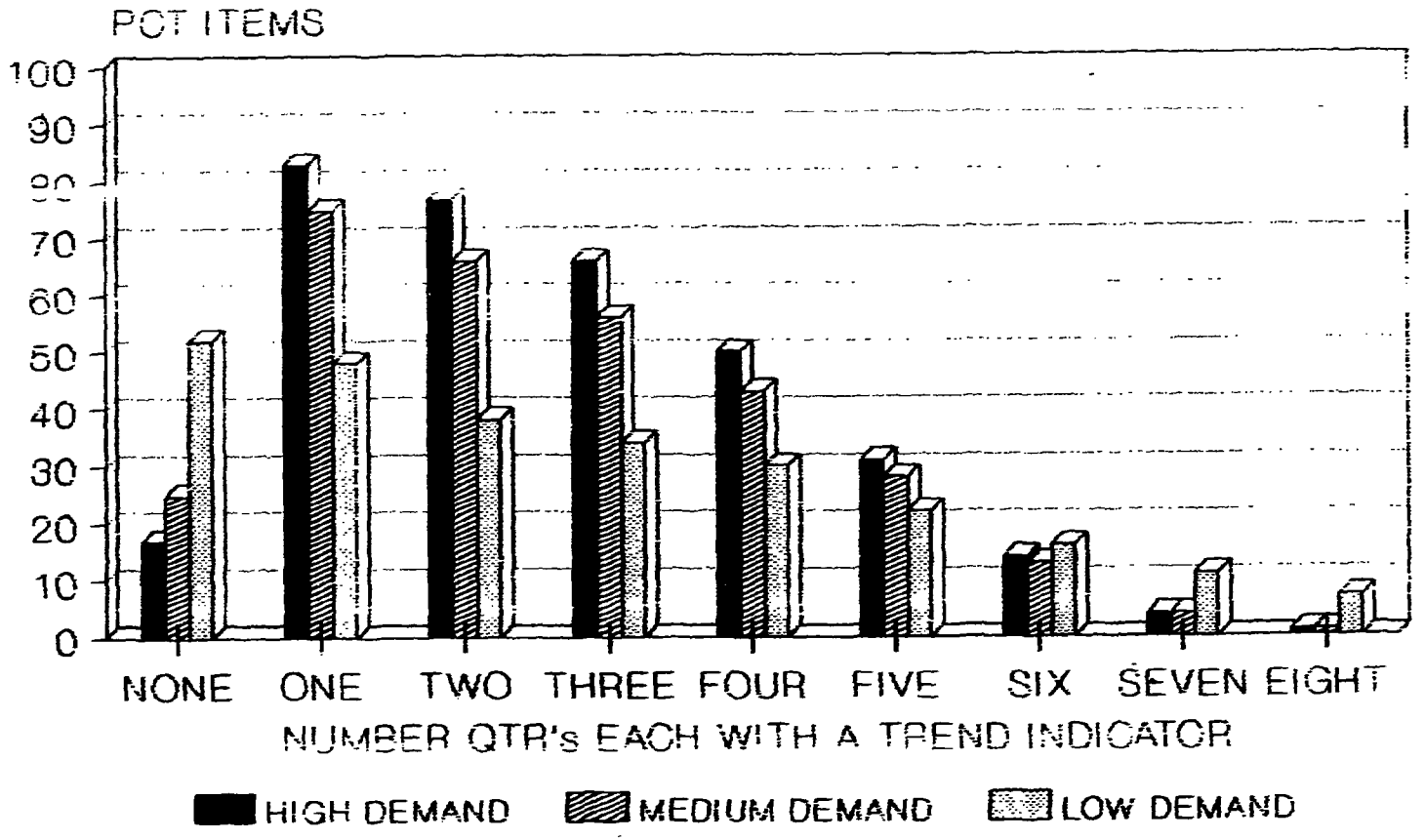
$$.1\ (D_1) + .9\ (\text{Last Forecast}) = \text{New Forecast}$$

On the other hand, when this statistic was less than .9 or greater than 1.1, it was concluded that a trend existed and the smoothing weights were increased to .3 and .7, thereby placing more emphasis on recent demand.

Experience over the last decade showed an increase in the percent of contracts for material that had to be cancelled because excess material was on-hand or on-order. The counterpart to this, insufficient material on-hand or on-order, was also on the rise. Additionally, procurement leadtimes in the past 10 years have risen from a year on the average to two years, significantly increasing the material forecast horizon.

Both top management and individual item managers believed that this trigger was overreacting, that it was too sensitive to random noise in the system. This study was undertaken to examine the trend detection method more closely. Chart I portrays evidence confirming these management concerns from the quarterly trend statistics for the period June 1987 through March 1989.

CHART I
CUMULATIVE PCT ITEMS DEMAND TRENDING BEFORE FORECASTING INITIATIVES



SOURCE: SPCC MAR'89 LEVELS STATS

Our first examination of formula (1) was done by using a table of random numbers to obtain a sample frequency distribution of the trend statistic. The demands were separated into three categories; 0 to 1, 1 to 5, and 6 to 9 units of quarterly demand.

Using a sample of size 40, we obtained for the random nontrending (low usage rates) demand of size 0 or 1 the following pattern of values for the statistic:

Trend Statistic	0	.67	1.00	1.33	2.00
Frequency	5	3	19	4	9

It is easily seen that using trigger values of .9 and 1.1 would conclude that demand was trending 53% of the time for low demand items. Furthermore, it is interesting to consider the 16 possible permutations of 0s and 1s over the eight quarters. Here we find the values for the trend statistic to be:

Trend Statistic	0	2/3	1	4/3	2
Frequency	3	2	5	2	4

So much depends on the case of four consecutive periods of zero demand. This situation has no numerical value, but common sense would indicate that we should assign these cases the value of one. If this is done, the statistic is biased to the right of one. Also, it is interesting to note what percentage would be declared trending using different lower and upper bounds on the trigger statistic.

Trigger Statistic Threshold Bounds			Percent Triggered
Low		High	
.9	to	1.1	70%
.8	to	1.2	70%
.7	to	1.3	70%
.6	to	1.4	43%

Thus, the trigger trend statistic would still indicate trend in 43% of the cases even using for a lower bound any number from 0 to .6, and for the upper bound, any value from 1.4 to 2. At this point, it became clear that no statistic such as this should be used on a string of zeros and ones. Yet, the majority of items managed by the Navy have characteristically low demand rates.

A similar analysis was conducted for the medium demand category (one to five demands per quarter) using a random sample of size 20. The following table was obtained for the same statistic (values were rounded off to the nearest tenth):

Trend Statistic	.3	.6	.8	1.0	1.1	1.3	1.4	1.6
Frequency	1	3	3	4	1	5	1	2

Again, the mean is slightly larger than one and the present lower and upper bounds on the trigger (.9, 1.1) would signal trend on the lower side about 35% of the time and about 40% on the upper side. A simple examination of the data suggested changing the thresholds to .6 and 1.6 for low and high cutoffs to obtain 80% assurance that a trend is not declared for common stationary demand situations.

Finally, for the high demand (values of six to nine demands per quarter) using a random sample of size 30, we found the following values:

Value	.87	.90	.91	.93	.94	.96	.97	1.00	1.03	1.04	1.07	1.13	1.14
Frequency	1	1	2	3	2	3	6	4	4	1	1	1	1

For high demand items, it appears that the trend statistic becomes more stable and lies within a decreased computational range. The threshold values (.9, 1.1) hardly come into play. So in these situations, the probability of a Type I error is relatively small and may be tolerable.

In September 1989, action was taken to widen the trigger trend threshold values from (.9, 1.1) to (.6, 1.4) in an effort to reduce the Type I error. CHART II details the reduction in the number of items designated as trending after these changes were accomplished. Significant reductions were achieved for high and medium demand items, but low demand items were relatively unaffected, experiencing almost a flat (45,000 to 50,000 items) trend declaration level. This finding is consistent with our theoretical conclusions discussed earlier relative to the trigger trend statistic expected performance for low demand items with a large percentage of zero observations.

This relatively minor forecasting change, based only on straightforward extensions of probability theory and operations analysis techniques, led to an immediate 10 million dollar per quarter reduction in inventory requirements changes. Requirements changes (churn) due to quarterly reforecasting of demand was measured in June 1989 as approximately 30 million dollars per quarter at SPCC. Of this amount, 10 million dollars worth of changes was attributed directly to the prior overly sensitive trend trigger thresholds of (.9, 1.1) invoking (.3, .7) smoothing on about 11,000 high and medium demand items.

These reductions were primarily due to decreases in the amount of quarter to quarter change in the key inventory requirements areas of economic order quantity and reorder point computations. Operationally, whenever unnecessary or unwarranted requirements changes are reduced, better estimates of future needs and budget requirements can be made and rework of procurement orders are reduced. Likewise, significant reductions can be achieved in excess material on-order as forecasts become more stable.

From the theoretical aspect, we recognize in formula (1) that by dividing the numerator and denominator of the statistic by four, that it is essentially a quotient of the means of a sample of size two and a sample of size four. Furthermore, the two means are correlated in that the two most recent quarters of demand appear in both the numerator and denominator.

Much has been written on the product and quotient of two random variables. It is a theoretically complicated subject. Only in the restricted situation of two normally distributed variables that are correlated has an explicit density function been derived. In this case, E. C. Fieller in 1932 showed that if x and y are bivariate normal with correlation ρ, the statistic

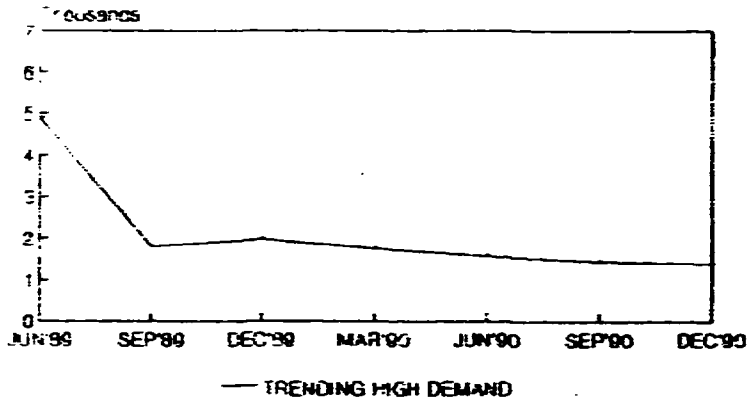
$$Z = \frac{x - \mu_x}{\sigma_x} \Big/ \frac{y - \mu_y}{\sigma_y}$$

has frequency function

$$g(Z) = \frac{(1-\rho^2)^{\frac{1}{2}}\, dz}{\Pi\,(1 - 2\rho Z + Z^2)}\ , \quad -\infty \leq Z \leq \infty$$

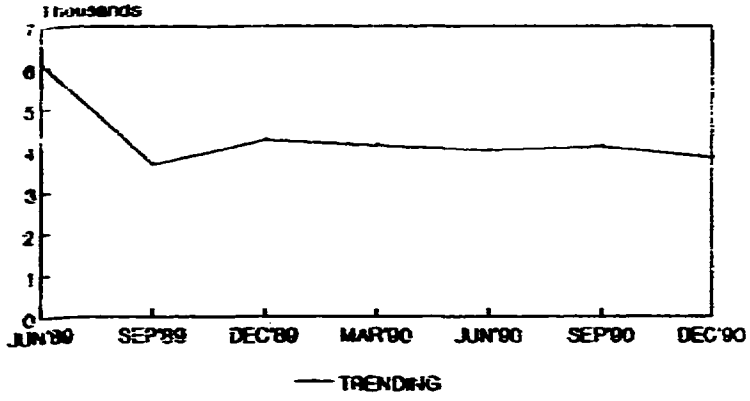
It doesn't seem defensible to try to invoke such theory in our present situation. Now that it was clear that the previous method used to detect trends was problematic, our attention and research turned to alternative methods for trend detection and estimation.

CHART II
HIGH DEMAND ITEMS
TRENDING TRACK



AVG DEMAND 3 OR MORE REQN PER QTR
SPOC BASELINE JUNE'89

CHART II
MEDIUM DEMAND ITEMS
TRENDING TRACK



AVG DEMAND 1 TO 3 REQN PER QTR
SPOC BASELINE JUN'89

II. **FIRST ATTEMPTS TO IMPROVE.** Having concluded that the trigger-trend statistic had serious operational limitations (poor trend detection and poor stability confirmation) we began research into alternative methods for trend detection in time series data. Our goal being to select and evaluate a method which was:

- Simple to use and interpret by Inventory Managers.
- Robust.
- Invariant to distribution assumptions.
- Excellent ability to confirm stationary demand.
- Relatively good ability to detect substantial trends and process change.
- Could be used effectively with small sample sizes.

We sought first to identify a suitable method to detect trends and then subsequently select a method to ascertain the level of trend present.

The work of Gini provided a fundamental understanding of the relationship of the variance between individual observations and the variance of the series of observations. From this research, we come to the common sense, but much overlooked conclusion, that for a nontrending time series, each of the observations are equally likely events. That the highest and lowest values may be first, last, or in the middle of the series and yet no trend may be present.

Tintner's work on the variate difference method provided an important insight and understanding to our most prevalent demand pattern ... the ZIG-ZAG demand pattern. From his work, we were able for the first time to realize that the majority of our items were not trending but rather oscillating between values of high to low observations. Many trend detection methods experience difficulty in handling oscillating demand patterns and, in particular, our trend-trigger statistic is overly sensitive to such common demand relationships.

Closely related to all this is what we call the mean squared successive differences. Here we compute the average of the squares of the (n-1) successive differences between successive elements in a random sample of size n. It has been shown that the expected value of this statistic, namely of

$$\delta^2 = \frac{\Sigma\,(x_{i+1} - x_i)^2}{n-1} \tag{2}$$

is $2\sigma_x^2$, when the base population distribution is normal. But the expected value of the ordinary sample variance, namely of

$$s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1}$$

is σ_x^2. Therefore, we can say the ratio

$$\eta = \frac{\delta^2}{s^2}$$

has expected value 2.

Dr. von Neumann gave us the distribution function of η, and in 1942 Dr. Hart provided a table of its values. However, as has happened in so many other situations, Dr. von Neumann assumed the sample came from a normal distribution. So, in order to use this statistic, we are parametrically bound. When data has an upward trend, δ^2 will increase much less than s^2, so η would be less than 2. On the other hand, if the data rapidly goes up and down (a situation which we now face to which our present trigger seems to be sensitive), δ^2 will increase proportionally greater than s^2. Then η will be greater than 2 and would be unsatisfactory.

Some attention was also given to using a sign test or a runs test using the median. The analysis showed such examination would not detect nonrandomness when η did.

III. TREND DETECTION USING RANK CORRELATION APPROACH. In this work, it is difficult, if not impossible, to defend any assumption of normality. A measure which is robust is desired. Kendall's S (Tau) is based on the ranks of observations as opposed to the magnitudes associated with them. What is important is that the "S" distribution is independent of the distributions of the variables of interest, e.g., the independent and dependent variable in linear regression. Interesting and useful is the fact that if two variables have a monotonic relationship, their ranks will have a regression that is apparently linear. Hence, the ranks can be considered to be transformed variables.

Let us now discuss the concept of rank correlation as it applies in our case to demand observations over a period of time. Consider a set of data values ordered in time; i.e., demand observations ranked by quarter:

$$(x_i, y_i; x_{i+1}, y_{i+1}; \ldots; x_n, y_n)$$

where

x =	1	2	3	4	5	6	7	8	(time in quarters)
y =	A	B	C	D	E	F	G	H	(observed values)

If there is a trend, there will appear to be an arrangement or ordering to the data. That is, if the ranks are correlated, then as "x" increases "y" will generally increase or decrease. Where there is no trend, we should expect the values to be randomly arranged. Kendall refers to the first case as "array" and the second as "disarray" in the data.

Step:

(1) Let us first arrange one rank in its natural increasing order:

$$\text{Time} = x_1, x_2, x_3 \ldots x_n$$

Then we compare all possible $\binom{n}{2}$ demand observation values two at a time. For example, for the first paired observations:

where $y_1 = A$ and $y_2 = B$

(2) Upon inspection, if we find that $y_2 > y_1$, implying that $B > A$, then Kendall would call the pair concordant, since the values are in agreement or arrayed with respect to the ordering of time. For each such pair assign a score of +1.

(3) However, if we find that $y_2 < y_1$, implying that $B < A$, then the pair is called discordant, since the values are in disagreement or disarray with respect to time. For each such pair, assign a score of -1.

(4) In the special case where $y_2 = y_1$, that is $B = A$, the pair is considered to be tied in ranks and a score of zero is assigned.

(5) The Kendall "S" statistic is simply the sum of the number of concordant pairs less the sum of the number of discordant pairs. It is a measure of the rank correlation between the order of the variables in time and their order in magnitude. The following table displays the probabilities that the absolute value of "S" attains or exceeds a specified value for a sample size of eight observations:

KENDALL "S" PROBABILITIES
SAMPLE SIZE OF 8 OBS

• PROBABILITY "S" ATTAINS OR EXCEEDS A SPECIFIED VALUE

	|S|	PROB	|S|	PROB
	0	.548	14	.054
	1	.500	15	.043
	2	.452	16	.031
	3	.408	17	.024
	4	.360	18	.016
	5	.317	19	.012
	6	.274	20	.0071
	7	.237	21	.005
	8	.199	22	.0028
	9	.169	23	.0018
	10	.138	24	.0007
	11	.114	25	.0005
	12	.089	26	.00019
------>	13	.0715	27	.00011
			28	.000025

The Kendall "S" statistic measures the degree of rank correlation on ordering present in the data. Its range is equal to the maximum number of pairs of comparisons that can be found between the two rankings. In the present case, "S" will range from (-28, +28):

$$\binom{n}{2} = \frac{n(n-1)}{2} = \frac{8(8-1)}{2} = 28$$

In this case, if we assume a nontrending stationary demand time series with 28 possible rank combinations, one would expect from theory that exactly 14 pairs would be concordant and 14 will be discordant yielding ("S"=+14-14=0) an "S" value of zero. Inspection of the above table would suggest that this condition would occur about 55% of the time.

Where the data is perfectly arranged in increasing or decreasing order, "S" would attain values of (+28, -28), respectively. Between these extremes of perfect to zero rank correlation probabilities can be assigned for observed values of "S" which can form the basis for hypothesis testing against the existence of trend.

(6) Construct an hypothesis test against the existence of trend.

- H_0: No trend exists
- H_1: Trend is present in the data

We will reject the hypothesis H_0 that no trend is present in the data if the probability of "S" equals or exceeds a specified critical value. In the table above, we selected a critical value of about seven percent yielding an "S" value of 13 which must be achieved or exceeded before we would reject the hypothesis of no trend present in the data.

For the reader not familiar with Kendall's S (Tau) in trend analysis, we present a numerical example:

NAVY PART NIIN 00-025-3457

Demand	1	2	2	3	2	3	7	2
Quarter	1	2	3	4	5	6	7	8
	Oldest							Latest

<u>Rule</u>: Subtract every demand from each one that precedes it. If greater, assign +1; if less, assign -1. If equal, assign 0. Sum the positive and negative scores to calculate "S".

	1						
	2	+1					
	2	+1	C				
	3	+1	+1	+1			
	2	+1	0	0	-1		
	3	+1	+1	+1	0	+1	
	7	+1	+1	+1	+1	+1	+1
	2	+1	0	0	-1	0	-1	-1
S=Σ		+7	+3	+3	-1	+2	0	-1

$S = +13$

$Pr\ (S \leq 13) \approx .958$

since

$Pr\ (S \leq 12) = .946$ $\quad$ $Pr\ (S \geq 12) = .089$

$Pr\ (S \leq 14) = .969$ $\quad$ $Pr\ (S \geq 14) = .054$ $\quad$ $S \geq 13 \rightarrow .072$

Owen's Tables $\quad$ Kendall's Table

Thus, Kendall's "S" with an error probability of about 7% suggests a high probability of an upward trend present in the data. Kendall's Tau (α) likewise, is a simple linear transformation of "S" and is calculated by dividing "S" by 28. With an effective operating range of (-1, +1), Kendall's Tau (α) parallels the range of the Pearson correlation coefficient of parametric statistics.

IV. NONPARAMETRIC TREND ESTIMATION. It is well known that an outlying observation can have an appreciable effect on the position of the parametric least squares fit. On the other hand, the regression that passes through the sample median of X and, simultaneously, the sample median of Y instead of the centroid, is less sensitive to outlier observations. If, in addition following Sen, we use for the slope the median of the set of slopes determined from all possible pairs of points, we will have induced more stability. For "n" data points, the slope set will have n(n-1)/2 slopes and this seems to have something in common with the calculation of Kendall's S (Tau) to detect trend. It is interesting to note that the classical least squares estimate of slope can be obtained from a weighted average of these individual slopes as can be done for the intercept.

The 28 slopes have a median of .29, while the least squares estimate is .40. The centroid is (4.5, 2.75) while the median point is (4.5, 2). It is simply accidental that both intercepts equal .93.

From Sen we have N = 21 (nonzero), n = 8, $\binom{n}{2} = 28$ and using Kendall's $\epsilon_n = .054$ and $U^* = 14/28$, we calculate

$$N^* = \left\{ 21 \binom{8}{2} \right\}^{\frac{1}{2}} * \frac{14}{28} = 12.12 \qquad (3)$$

Then, the lower and upper 90% rank confidences values are

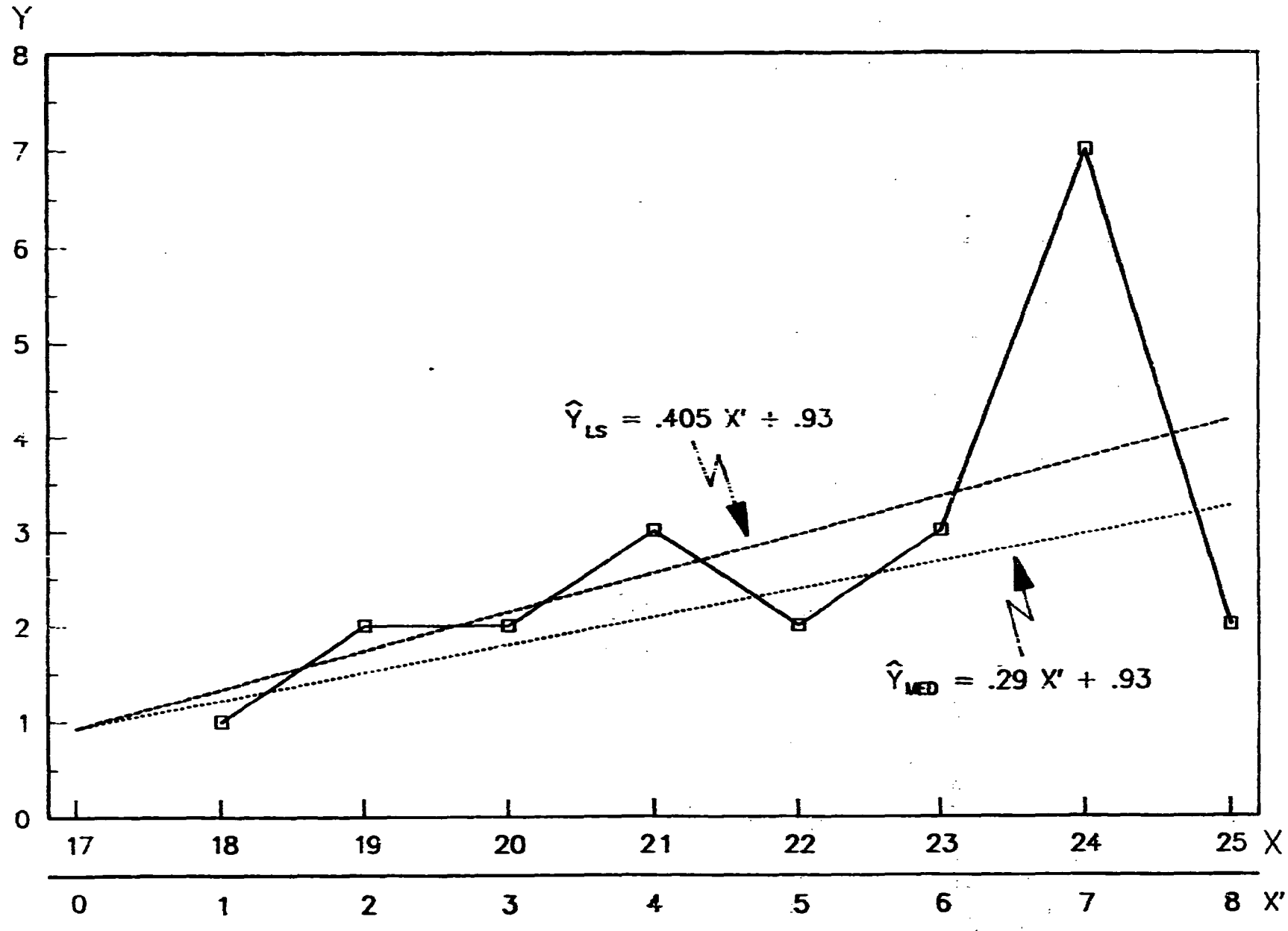
$$M_1 = \frac{1}{2}(21 - 12.12) = 4.44 \sim 4 \tag{4}$$

$$M_2 = \frac{1}{2}(21 + 12.12) = 16.56 \sim 17 \tag{5}$$

where we rounded down on the lower rank and up on the upper rank, as recommended by Conover. So the 4th ranked slope and the 17th ranked slope gives us our 90% two-sided confidence interval of the slope; i.e., (-.25, .5).

The following graph provides a comparison of classical least squares regression and the nonparametric methods in fitting the time series data from our example. It is clear from the graph that a positive trend component exists in the data; however, the median regression method provides a more conservative trend estimate. Thus, this method is less sensitive to the one outlier observation sustained in period seven. This attribute may be of partial significance in those situations where the trend component effect is used to compute future average demand values eight or more periods into the future as is the case with Navy inventory control systems.

LEAST SQUARES vs MEDIAN REGRESSION



V. **EXPERIMENTAL DESIGN**. Our overall objective was to evaluate the performance of the Kendall "S" trend test across a broad range of expected demand, variability and trend levels, while maintaining control of key variate factors to ensure the results reflected intrinsic expected model behavior/performance. Specifically, we sought to access the ability of the Kendall "S" statistic to:

- Detect demand trends.
- Transition effectively between alternating trend/no trend periods.
- Confirm demand stability in the absence of trend.
- Determine the average lag and delay in identifying trend commencement/trend termination.

To maintain maximum control of the key design factors, we initially chose simulation. By using simulated data we could specify certain factors and levels precisely:

- The underlying mean demand of the base population.
- The degree of variability about the mean.
- When a trend was introduced and its level/magnitude.
- The random variable noise level present in the data via the standard deviation to mean ratio.

These types of factors and levels usually affect the ability to detect the presence of a trend or to confirm the lack of a trend (demand stability).

Our trend generating function is:

New Mean Demand = [Mean Demand at Beginning of Trend]

$$\times\ [1 + (\underbrace{\text{Period Multiplier}}_{\text{"Alpha"}} \times (\text{Period Into Trend})^{\underbrace{\text{Period Power}}_{\text{"Beta"}}})]$$

In algebraic symbols this is

$$y_x = y_o\ [1 + \alpha(x)^{\beta}]$$

The sign of α determines the direction and magnitude of the expected equipment population increases or decreases. The Beta exponent or period power is used to adjust the rate of increase or decrease in an item's individual expected failure rate per installed population.

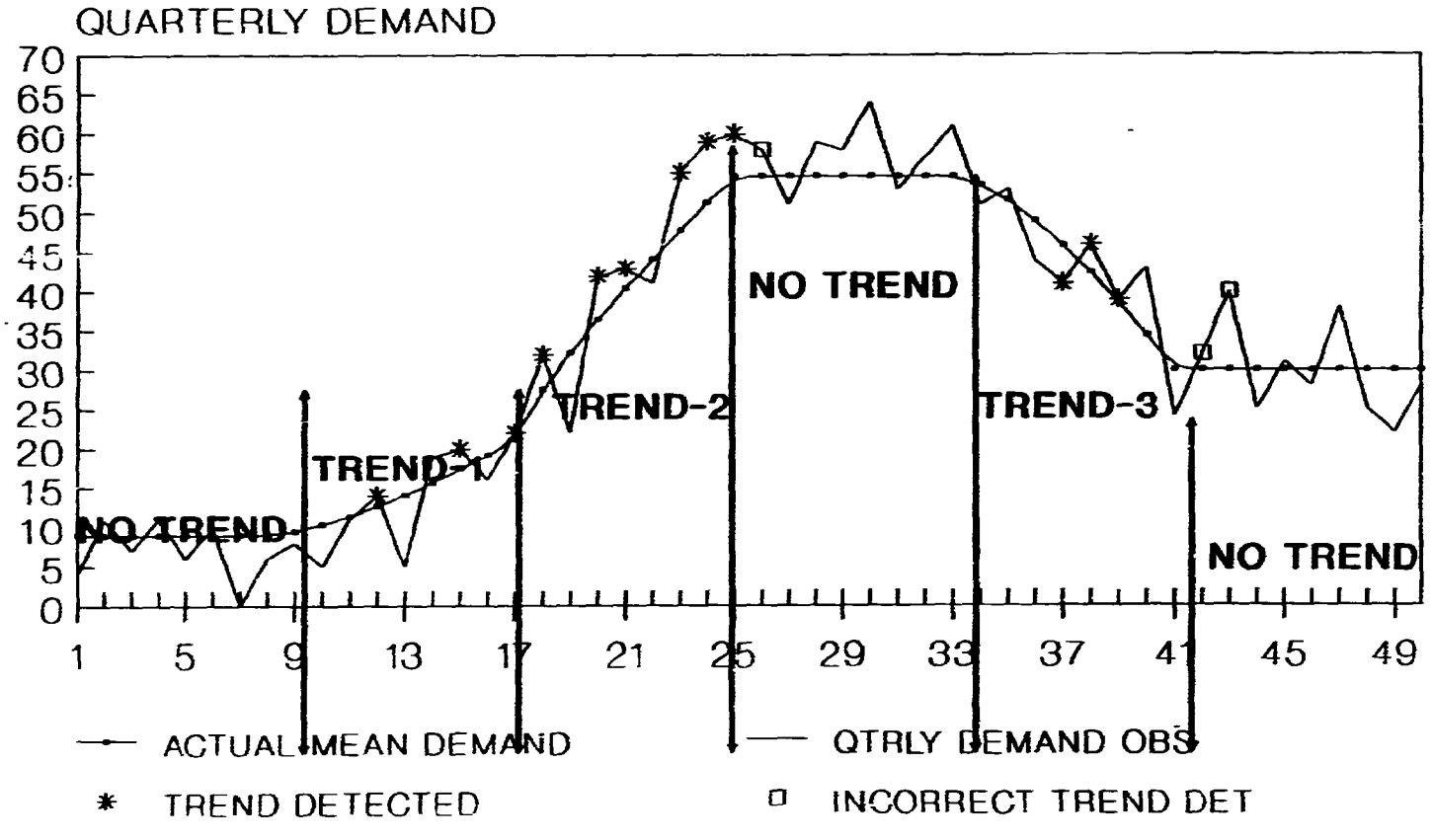
Each item started with a stable demand period - initialization phase - followed by two periods of trend, first a moderate trend and then a stronger trend to reflect increasing population and failure rate - a growth phase. The trend was then turned off to represent the "top of the ramp" steady state typical middle phase of the system support life cycle. Next we induced a strong negative trend to reflect the effect of de-installations; i.e., population "decline phase". Finally, the above decline was followed by a stable "final residual support phase" period.

For medium, high, and very high demand, we used:

Periods 1 - 8	no trend
Periods 9 - 17	$\alpha = .05$, $\beta = 1.5$
Periods 18 - 25	$\alpha = .3$, $\beta = .8$
Periods 26 - 33	no trend
Periods 34 - 41	$\alpha = -.02$, $\beta = 1.5$
Periods 42 - 50	no trend

The following graph shows the general shape of such a demand activity. We need to explain next how we induced random variation - noise.

TRENDING SIMULATION EXAMPLE
HIGH DEMAND ... MEDIUM VARIABILITY



TREND SCORE 46% CORRECT
NO TREND SCORE 95% CORRECT
WINDOW SIZE = 6

We generated two random numbers from a uniform distribution on (0,1). Call them R_1 and R_2. Then by M. Muller's method, we calculated

$$V = (-2 \log R_1)^{.5} \times (\cos 2 \Pi R_2)$$

for normal deviation on (0,1). To convert V to a normal distributed variable with mean μ and standard deviation σ, we used the transform

$$V_1 = V \times \sigma + \mu$$

Poisson random variable noise was generated using Martin's method based on a method suggested by Kahn. Briefly, one begins by generating N uniform random numbers U_i and successively multiplying them until the following inequality is satisfied:

$$\prod_{1}^{N} \mu_i < e^{-np}$$

Then (N-1) is the desired Poisson random variable value with a mean of np. When the first random variable number generated satisfies the inequality, then the Poisson variable is assigned the value zero.

In the previous Section III, we illustrated the calculation of Kendall's S (Tau) for a window of (sample size) 8. It is apparent this statistic can be calculated for other size windows and we felt it necessary to examine the reaction of this statistic to various size windows. Obviously, a longer window is not as susceptible to change as a shorter window. As will be seen, we will take our known generated pattern and utilize various window lengths on such a time series by dropping off the earliest number in the window and adding on the latest one. We will do this for window sizes of 4, 6, 8, 10, and 12.

The second factor for which we seek the effects at various levels comes from scientific judgment and experience in this field of work. We have seen patterns or the lack of such according to the volume of demand for a part. Consequently, we have classified demand into five categories or levels: very low, low, medium, high, and very high, given by:

Demand Category	Average Demand
Very low demand	.5 to 4 per year
Low demand	1 per quarter
Medium demand	3 per quarter
High demand	9 per quarter
Very high demand	20 per quarter

The third factor of concern is demand variability which we need to simulate and control. After much analysis of actual data, it was decided to use for the measure the ratio of the average variability to the average demand, the standard deviation to mean ratio or coefficient of variation.

The selection of levels for this factor appeared not to be independent of the levels of the second factor. A sort of Pareto distribution analysis of over 9000 randomly selected parts gave us the following two-way table:

	Variability Level Standard Deviation/Mean Ratio				
Demand Catagory	VL	LO	MD	HI	VHI
VL	.49	.64	.71	1.00	1.60
LO	.40	.70	.90	1.21	1.26
MD	.16	.30	.55	.91	1.13
HI	.14	.28	.53	.93	1.12
VHI	.14	.28	.53	.93	1.06

VI. FINDINGS - RESULTS - RECOMMENDATIONS. The 50 various data sets used in the study each consisted of 50 quarters of demand; that is, for each combination of demand and variability, a simulated demand data set was generated. In addition, this was done for Case (1) - a set with three periods of trending/no trend and Case (2) - a set without trends (stationary). TABLEs I - III provide graphs of the simulated demand observation data sets used in the study. The Kendall "S" test was applied to each successive set of observations in the window (moving through time) in an attempt to identify trending and nontrending processes.

NORMAL DISTRIBUTION
NON-TRENDING DATA SETS

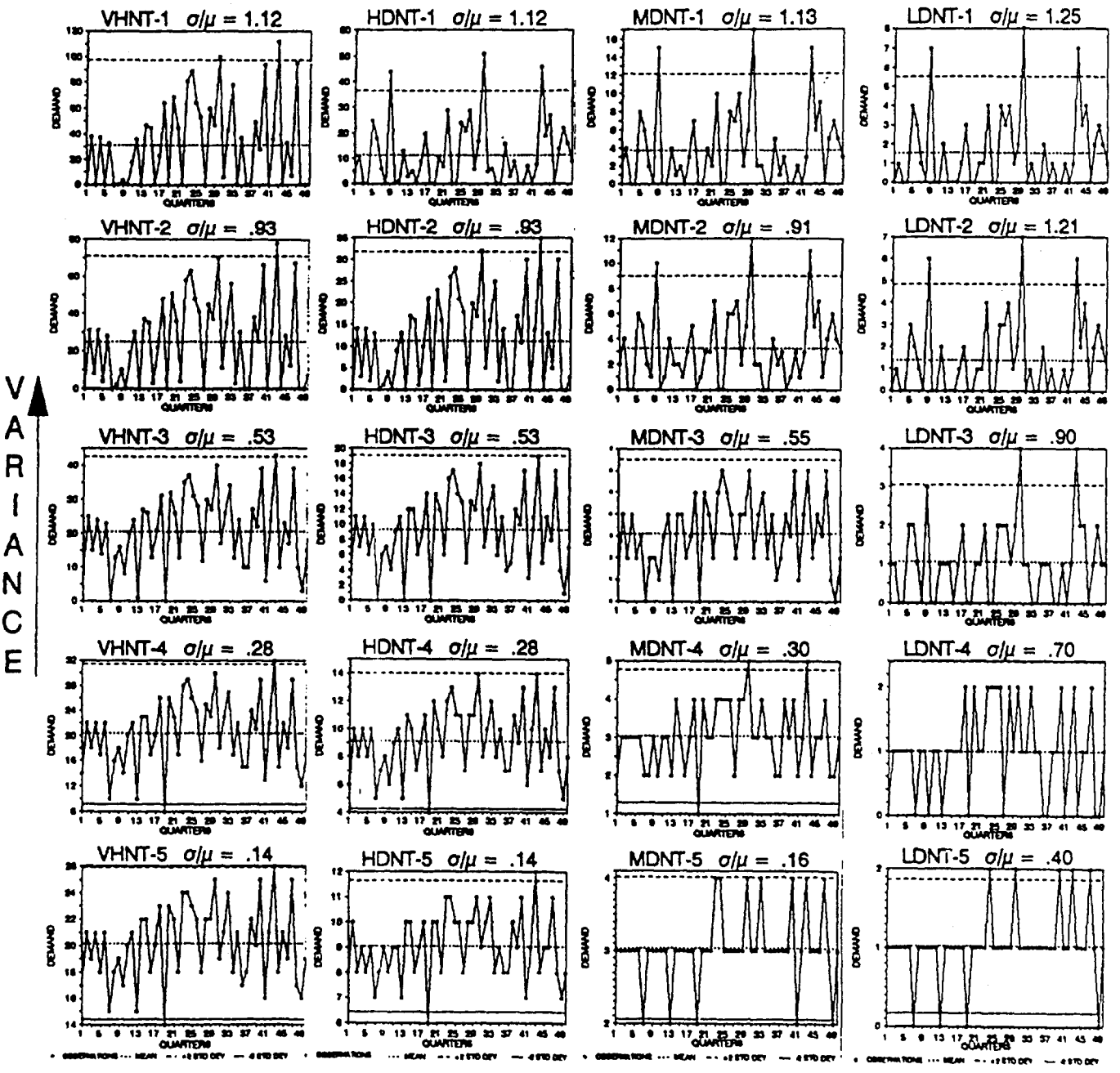


TABLE I

NORMAL DISTRIBUTION TRENDING DATA SETS

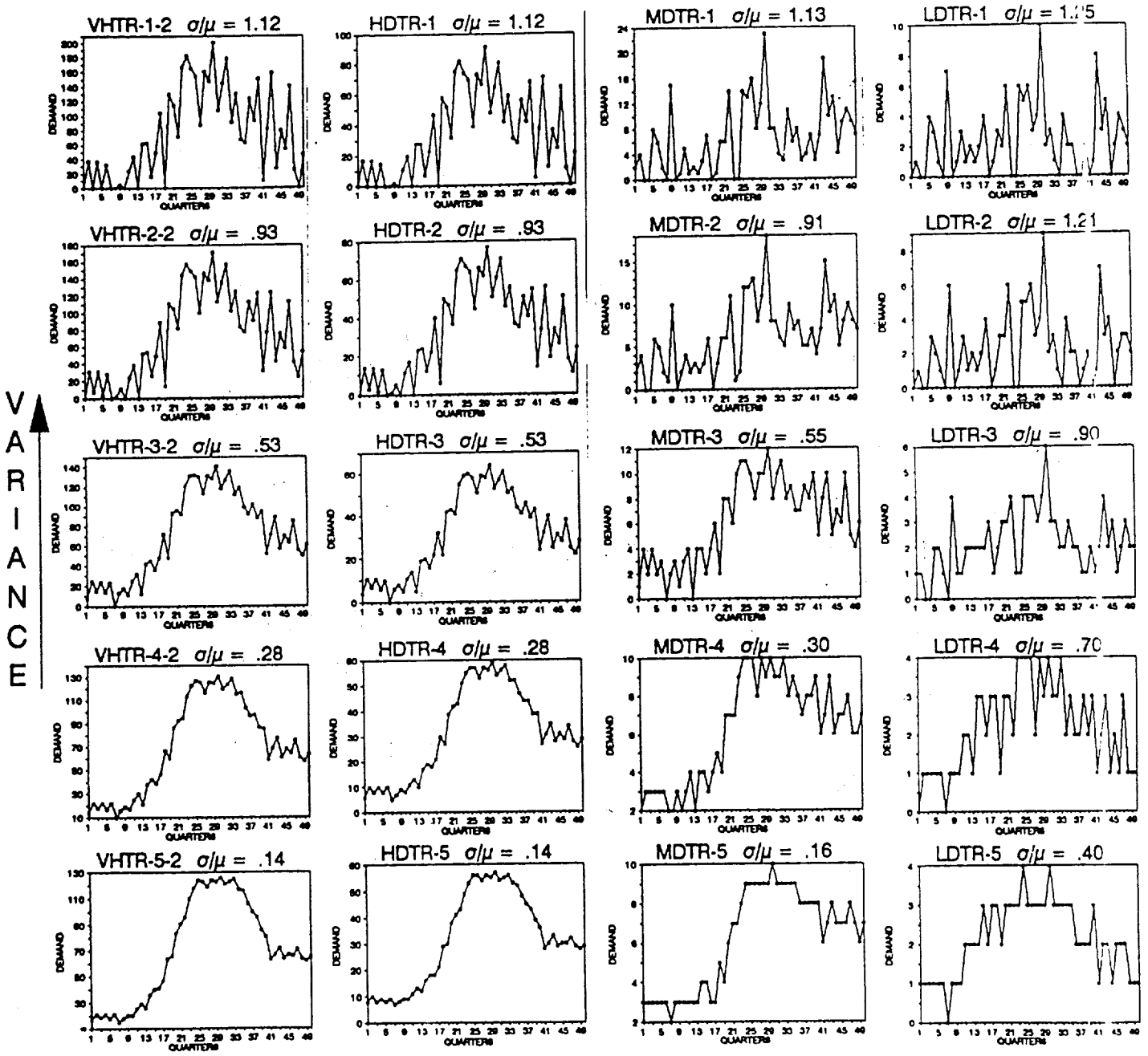
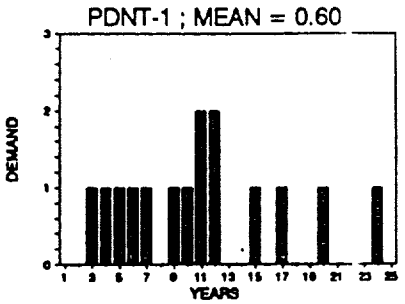


TABLE II

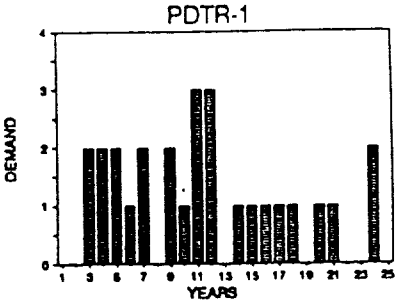
POISSON DISTRIBUTED DEMAND

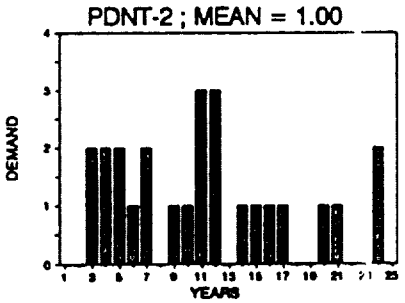
STATIONARY DEMAND — TRENDING DEMAND

VARIANCE

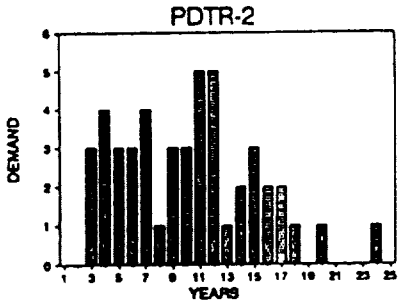


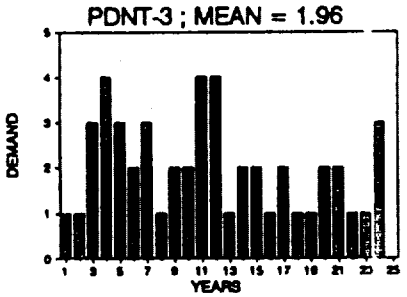
$\sigma/\mu = 2.7$



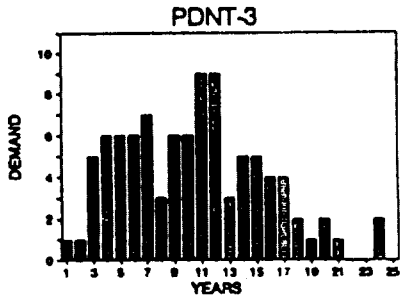


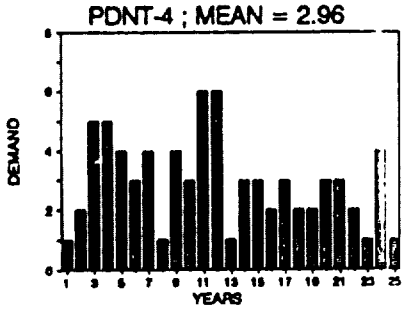
$\sigma/\mu = 2.0$



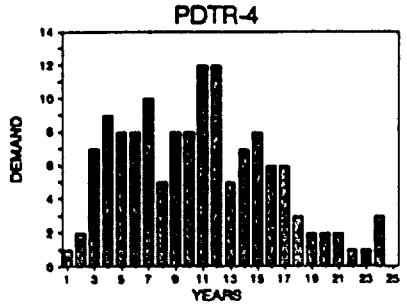


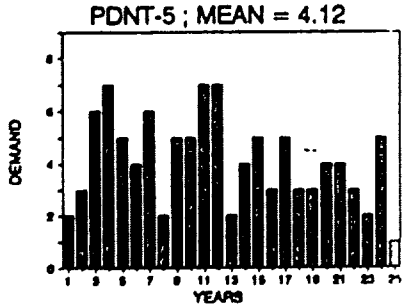
$\sigma/\mu = 1.4$





$\sigma/\mu = 1.0$





$\sigma/\mu = 1.1$

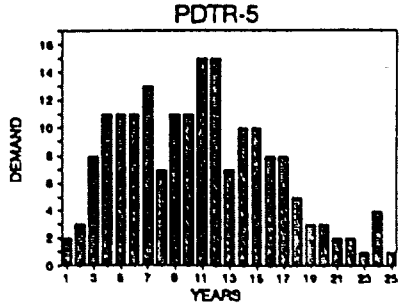


TABLE III

For each observation set within the moving window across the 50 data points, we counted how often the Kendall "S" correctly identified the presence of trend and how often it correctly identified the lack of trend; i.e., confirmed stability.

The evaluation to be used follows the following example for each window size across the 50 data points.

		ACTUALITY	
		TREND	NO TREND
PREDICT	TREND	12	1
	NO TREND	13	19
	TOTAL	25	20
	SCORE	48%	95%

We call a Type I error the classifying of actual nontrending as trending and a Type II error the classifying of trending as nontrending. In the above table, we predicted 19 of the 20 no trending situations which means the probability of a Type I error is 5%; whereas, of the 25 trending situations we predicted only 12, thereby having a probability of a Type II error of 52%.

So our immediate goal is to determine optimal window sizes in order to minimize Type I and Type II errors. Initially, we set performance goals:

(1) Maintain very high capability to confirm stability for stationary demand patterns - more than 88% of the time.

(2) Maintain relatively high capability to detect that trend influence has ceased - more than 80% of the time.

(3) Select window size maximizing trend detection while meeting (1) and (2).

Practically this led us to minimize Type II errors for a specified Type I goal. Now in using the Kendall "S", we tried to hold to cutoff values that were exceeded about 5%. However, tabular values closest to 5% had to be used. These are given at the bottom of each column in the following tables.

The following five tables give the percentages that Kendall's "S" achieved in being correct. Entries that are shadowed appear best for that window size. For example, look at the table for very high demand. Note the situation apparently improves as you go from right to left and from top to bottom. But it must be remembered that for rows above the last one, it is best to use a wider window; i.e., larger sample size for the time series.

The reader will note that observations like those in the previous paragraph are not as easily made for lower demand items. For the low demand items, it appears we cannot come close to our goal for detecting trend. The same can be said for the very slow Poisson case.

The sixth table to follow, "Overall Performance", comprises the five preceeding tables by washing out window length and sigma/mu categories.

KENDALL "S" PERFORMANCE ANALYSIS
VERY HIGH DEMAND ITEMS MU = 20/QTR

	W=4			W=6			W=8			W=10			W=12		
SIGMA/MU	TR	NT	ST	TR	NT	ST	TR	NT	ST	TR	NT	ST	TR	NT	ST
1.12	0	95	98	16	90	93	28	94	88	48	94	95	68	65	87
.93	0	95	98	16	95	89	44	89	88	50	88	88	73	35	82
.53	4	100	100	44	90	89	72	83	91	67	41	93	77	12	85
.28	20	100	98	76	85	89	80	56	91	71	35	93	82	0	90
.14	76	100	94	84	80	96	84	50	88	76	30	93	82	0	95
	S=5			S=9			S=13			S=18			S=22		

KENDALL "S" PERFORMANCE ANALYSIS
HIGH DEMAND ITEMS MU = 9/QTR

	W=4			W=6			W=8			W=10			W=12		
SIGMA/MU	TR	NT	ST	TR	NT	ST	TR	NT	ST	TR	NT	ST	TR	NT	ST
1.12	0	95	98	16	90	91	32	94	88	42	94	88	68	65	77
.93	0	100	98	16	95	89	44	88	87	54	88	90	73	41	82
.53	4	100	98	48	95	89	71	60	88	66	41	93	73	12	85
.28	36	95	96	80	90	96	80	50	95	71	35	95	82	0	90
.14	76	91	94	84	80	93	84	50	98	83	35	95	91	0	95
	S=5			S=9			S=13			S=18			S=22		

READ PCT CORRECT: TRending, Non Trending, STationary

KENDALL "S" PERFORMANCE ANALYSIS
MEDIUM DEMAND ITEMS MU = 3/QTR

	W=4			W=6			W=8			W=10			W=12		
SIGMA/MU	TR	NT	ST	TR	NT	ST	TR	NT	ST	TR	NT	ST	TR	NT	ST
1.13	0	95	96	9	90	91	12	94	86	20	94	88	22	94	74
.91	0	95	87	16	100	93	28	100	84	25	94	88	36	76	80
.55	4	91	96	20	100	96	16	94	98	38	82	100	45	65	95
.30	12	95	94	32	95	98	40	83	93	58	71	94	78	65	87
.16	16	100	100	24	90	98	50	66	98	54	47	100	59	24	97
	S=5			S=9			S=13			S=18			S=22		

KENDALL "S" PERFORMANCE ANALYSIS
LOW DEMAND ITEMS MU = 1/QTR

	W=4			W=6			W=8			W=10			W=12		
SIGMA/MU	TR	NT	ST	TR	NT	ST	TR	NT	ST	TR	NT	ST	TR	NT	ST
1 25	4	95	94	4	100	96	9	89	93	21	100	88	27	88	74
1 21	12	86	91	4	100	96	14	83	93	25	94	88	27	83	74
.90	12	91	91	8	95	93	32	88	93	40	100	90	50	82	77
.70	0	100	94	8	100	96	24	89	93	41	91	98	64	94	85
.40	0	100	100	16	95	96	40	100	100	50	100	100	77	71	99
	S=5			S=9			S=13			S=18			S=22		

READ PCT CORRECT: TRending, Non Trending, STationary

KENDALL "S" PERFORMANCE ANALYSIS
POISSON DEMAND ITEMS MU = .5-1/YR

			W=4			W=6			W=8		
MU	SIGMA	RATIO	TR	NT	ST	TR	NT	ST	TR	NT	ST
.5	.71	1.42	0	100	95	0	100	95	0	93	94
1.0	1.0	1.00	30	93	95	20	87	100	30	80	83
2.0	1.4	.70	40	87	91	30	73	95	20	93	83
3.0	1.7	.58	40	93	86	33	73	95	40	53	89
4.0	2.0	.50	50	80	86	30	73	100	50	53	94
			S=5			S=9			S=13		

READ PCT CORRECT: TRending, Non Trending, STationary

KENDALL "S" RESULTS
OVERALL PERFORMANCE

DATA SET	HIGH	MEDIUM	LOW	POISSON
T	40 - 75%	20 - 40%	20 - 50%	30 - 50%
NT	80 - 95%	80 - 95%	90 - 100%	80 - 90%
ST	88 - 95%	88 - 100%	88 - 100%	85 - 90%

"T" = TRENDING DATA SET

"NT" = NONTRENDING BUT PREVIOUSLY TRENDING

"ST" = STATIONARY DATA SETS

PERCENT IDENTIFIED CORRECTLY

VII. THE LAG. The next five pages use the same set of data we used earlier for high demand, trending and not trending, and low variability, symbolized HDTR-4. On each computer printout, the 50 demand observations are blocked in five blocks, as the induced demand goes from no trend to trend (1), from trend (1) to trend (2), from trend (2) to no trend, from no trend to trend (3), and finally to no trend. Then, using an S with a small probability - around 7% - we followed S values in an interval and across intervals.

For example, the first page for a window of size 4 shows the five calculated values of S for the eight values in the beginning no trend period indicate correctly no trend. On the other hand, starting with period 9 and through period 17, when we have a trend, the S statistic fails to pick up trend throughout the entire period. Moreover, for four periods into the third stage, the S statistic still fails to suggest trend. Finally, at period 22, the S statistic indicates a trend and correctly predicts through period 25. We have indicated with a dash (-) when S is incorrect and with a star (*) when it is on target. For this particular run of 50 demands, we see nine correct calls out of a possible 25 for trending. However, out of 22 nontrending cases, our statistic S is correct 21 times, and so, is 95% correct, or if you will, has a probability of a Type I error of 5%. On the other hand, our statistics selected only 36% of trending cases; hence, we face a sample probability of a Type II error of 64%.

A similar explanation can be given for the other four cases of window lengths 6, 8, 10 and 12. It is apparent that window size of 6 or 8 does better.

This delay or lag puts limitations on successful use of the S statistic for forecasting as summarized in the chart following the five computer printouts.

It seems we are faced with a problem which lies within that general purview described by Bennett and Franklin on page 688 entitled "Choice of Tests for Non-Randomness".

> It will be observed that the sensitivity of the various tests for nonrandomness of observations depends upon the type of nonrandomness which is present. In cases where this can be anticipated from technical considerations, the appropriate test may be selected in advance, but, if the choice of tests is based on an examination of the results, the significance levels are clearly biased. An obvious alternative, that of subjecting all data to the same series of tests, is not free from criticism since many of the tests are not independent, although the degree of dependence is not yet known. It would appear that without technical guidance, the best procedure is probably to use a series of tests, interpreting the significance levels as a general indication rather than a specific prediction.

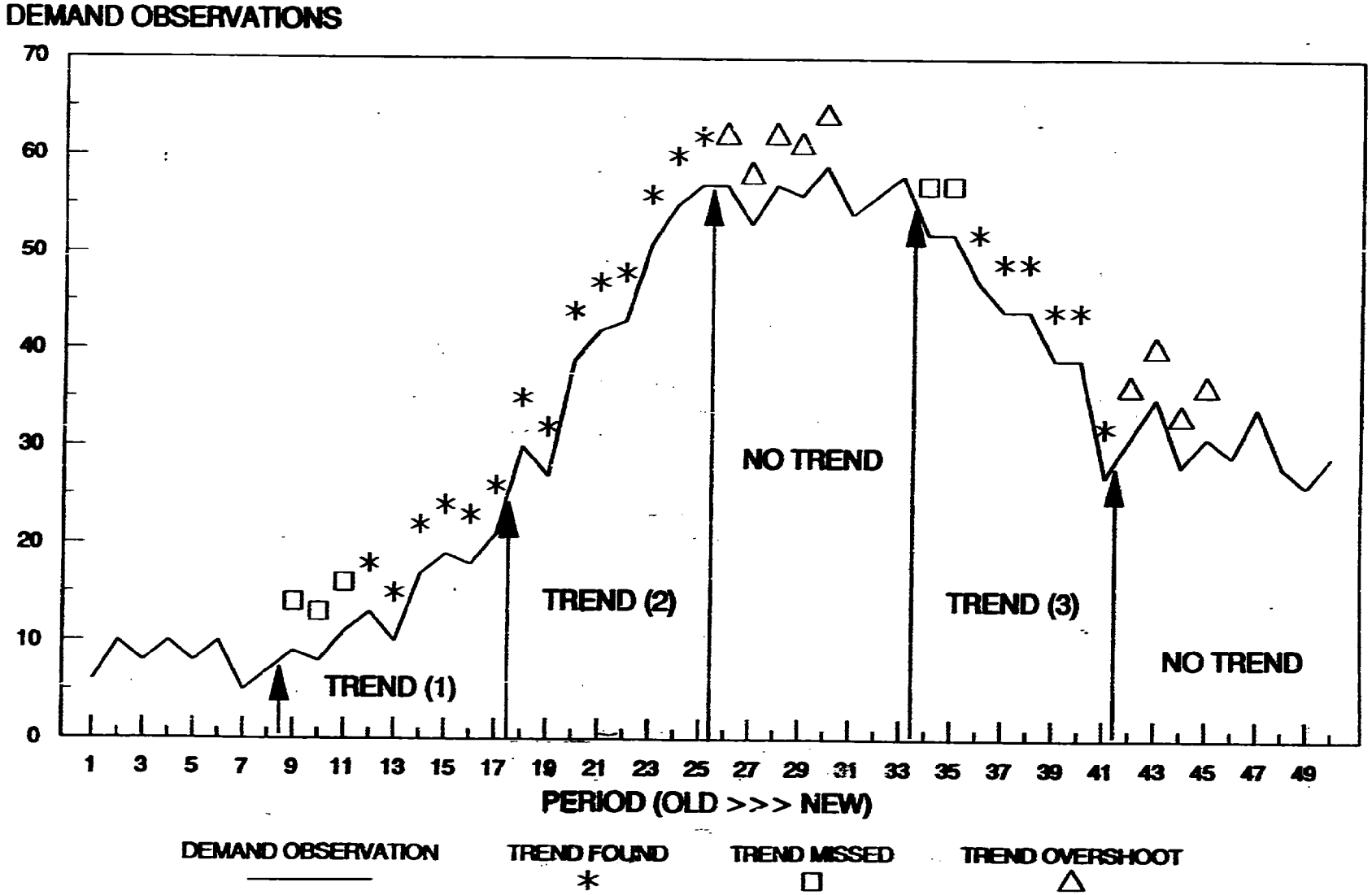
We wish to thank John Price of Hershey Foods who in the early stages programmed for us the Kendall S with the assistance of Brent Burkholder. This was later taken over by Emerson Evelhock who expanded the work to useful

graphs. The demand forecast and trend simulation model was developed by CDR Tom Bunker, SC, USN. Throughout all this, Charla Scheaffer typed and retyped and stayed the course.

DEMAND OBSERVATIONS

HDTR-4

WINDOW SIZE = 8



IDENTIFICATION OF LIMITATIONS

- ERRATIC DEMAND (DATA): SIGMA/MU RATIOS GT>1.25
  - EXHIBIT SO MUCH VARIATION EVEN KNOWN TRENDS ARE VERY DIFFICULT TO DETECT

- OVERSHOOT PHENOMENA:
  - WHEN: TREND PERIOD FOLLOWED BY NO TREND PERIOD
  - DEGREE: NORMALY 2 -3 PERIODS
  - CONDITION:
    - USING LARGE WINDOW SIZES (G.E. 8) TO INCREASE TREND DETECTION CAUSES OVERSHOOT OF TURNING POINT ... LARGE TYPE (I) ERRORS
  - WHERE:
    MOST EVIDENT FOR HIGH DEMAND LOW VARIANCE ITEM

- LAG OR DELAY TO DETECT TREND COMMENCEMENT
  - WHEN: NO TREND PERIOD FOLLOWED BY TREND
  - DEGREE: NORMALLY 3 - 4 PERIODS
  - CONDITION:
    - USING LARGE WINDOW SIZES WILL INCREASE MODEL SENSITIVITY BUT DELAYS MAYBE UNACCEPTABLE

KENDALL'S S STATISTIC
WINDOW = 4
HDTR-4

PERIOD	DEM OBS	KENDALL'S S	PROBABILITY	
1	6			
2	10			
3	8			
4	10	3	0.271	
5	8	-2	0.375	NO TREND
6	10	2	0.375	
7	5	-3	0.271	
8	7	-2	0.375	
9	9	- 0	0.625	
10	8	- 4	0.167	
11	11	- 4	0.167	
12	13	- 4	0.167	
13	10	- 2	0.375	TREND (1)
14	17	- 2	0.375	
15	19	- 4	0.167	
16	18	- 4	0.167	
17	21	- 4	0.167	
18	30	- 4	0.167	
19	27	- 4	0.167	
20	39	- 4	0.167	
21	42	- 4	0.167	
22	43	* 6	0.042	TREND (2)
23	51	* 6	0.042	
24	55	* 6	0.042	
25	57	* 6	0.042	
26	57	- 5	0.1045	
27	53	-1	0.5	
28	57	-1	0.5	
29	56	-1	0.5	
30	59	4	0.167	NO TREND
31	54	-2	0.375	
32	56	-1	0.5	
33	58	0	0.625	
34	52	- 0	0.625	
35	52	- -3	0.271	
36	47	* -5	0.1045	
37	44	* -5	0.1045	
38	44	* -5	0.1045	TREND (3)
39	39	* -5	0.1045	
40	39	- -4	0.167	
41	27	* -5	0.1045	
42	31	-3	0.271	
43	35	0	0.625	
44	28	2	0.375	
45	31	-1	0.5	
46	29	-2	0.375	
47	34	4	0.167	NO TREND
48	28	-2	0.375	
49	26	-4	0.167	
50	29	-2	0.375	

CRITICAL VALUE S=5

		IV ACTUALITY T	IV ACTUALITY NT
PRED	T	9	1
	NT	16	21
		25	22
		36%	95%

KENDALL'S S STATISTIC
WINDOW = 6
HDTR-4

CRITICAL VALUE S=9

PERIOD	DEM OBS	KENDALL'S S		PROBABILITY	
1	6				
2	10				
3	8				
4	10				
5	8				NO TREND
6	10		5	0.235	
7	5		-5	0.235	
8	7		-5	0.235	
9	9	-	-4	0.2675	
10	8	-	0	0.559	
11	11	-	5	0.235	
12	13	*	13	0.0083	
13	10	*	9	0.068	TREND (1)
14	17	*	9	0.068	
15	19	*	11	0.028	
16	18	*	9	0.068	
17	21	*	11	0.028	
18	30	*	13	0.0083	
19	27	*	11	0.028	
20	39	*	11	0.028	
21	42	*	13	0.0083	
22	43	*	13	0.0083	TREND (2)
23	51	*	13	0.0083	
24	55	*	15	0.0014	
25	57	*	15	0.0014	
26	57	-	14	0.00485	
27	53		8	0.102	
28	57		6	0.1855	
29	56		0	0.559	
30	59		2	0.4	NO TREND
31	54		0	0.559	
32	56		2	0.4	
33	58		0	0.559	
34	52	-	-4	0.2675	
35	52	-	-8	0.102	
36	47	*	-8	0.102	
37	44	*	-12	0.01815	
38	44	*	-13	0.0083	TREND (3)
39	39	*	-13	0.0083	
40	39	*	-13	0.0083	
41	27	*	-13	0.0083	
42	31	-	-11	0.028	
43	35		-8	0.102	
44	28		-6	0.1855	
45	31		-2	0.4	
46	29		2	0.4	
47	34		0	0.559	NO TREND
48	28		-4	0.2675	
49	26		-4	0.2675	
50	29		-6	0.1855	

		IV ACTUALITY T	IV ACTUALITY NT
PRED	T	20	2
	NT	5	18
		25	20
		80%	90%

KENDALL'S S STATISTIC
WINDOW = 8
HDTR-4

CRITICAL VALUE S=13

PERIOD	DEM OBS	KENDALL'S S		PROBABILITY	
1	6				
2	10				
3	8				
4	10				
5	8				NO TREND
6	10				
7	5				
8	7		-4	0.36	
9	9	-	-8	0.199	
10	8	-	-4	0.36	
11	11	-	2	0.452	
12	13	*	13	0.0715	
13	10	*	13	0.0715	
14	17	*	22	0.0028	TREND (1)
15	19	*	22	0.0028	
16	18	*	20	0.0071	
17	21	*	22	0.0028	
18	30	*	22	0.0028	
19	27	*	22	0.0028	
20	39	*	24	0.00087	
21	42	*	24	0.00087	
22	43	*	24	0.00087	TREND (2)
23	51	*	26	0.00019	
24	55	*	26	0.00019	
25	57	*	26	0.00019	
26	57	-	27	0.0001075	
27	53	-	21	0.00495	
28	57	-	19	0.01155	
29	56	-	13	0.0715	
30	59	-	13	0.0715	NO TREND
31	54		1	0.5	
32	56		-4	0.36	
33	58		4	0.36	
34	52	-	-1	0.5	
35	52	-	-12	0.089	
36	47	*	-16	0.031	
37	44	*	-21	0.00495	
38	44	*	-20	0.0071	TREND (3)
39	39	*	-24	0.00087	
40	39	*	-25	0.00053	
41	27	*	-25	0.00053	
42	31	-	-24	0.00087	
43	35	-	-20	0.0071	
44	28	-	-18	0.016	
45	31	-	-14	0.054	
46	29		-10	0.138	NO TREND
47	34		-1	0.5	
48	28		2	0.452	
49	26		-12	0.089	
50	29		-10	0.138	

		IV ACTUALITY T	IV ACTUALITY NT
PRED	T	20	9
	NT	5	9
		25	18
		80%	50%

KENDALL'S S STATISTIC
WINDOW = 10
HDTR-4

CRITICAL VALUE S=18

PERIOD	DEM OBS	KENDALL'S S		PROBABILITY	
1	6				
2	10				
3	8				
4	10				
5	8				NO TREND
6	10				
7	5				
8	7				
9	9				
10	8	-	-3	0.431	
11	11	-	-1	0.5	
12	13	-	13	0.146	
13	10	-	13	0.146	
14	17	*	25	0.014	TREND (1)
15	19	*	30	0.00345	
16	18	*	37	0.00018	
17	21	*	37	0.00018	
18	30	*	37	0.00018	
19	27	*	37	0.00018	
20	39	*	37	0.00018	
21	42	*	39	0.000058	
22	43	*	41	0.000015	TREND (2)
23	51	*	41	0.000015	
24	55	*	41	0.000015	
25	57	*	43	0.0000028	
26	57	-	42	0.0000089	
27	53	-	36	0.000325	
28	57	-	36	0.000325	
29	56	-	30	0.00345	
30	59	-	30	0.00345	NO TREND
31	54	-	18	0.066	
32	56		9	0.242	
33	58		7	0.3	
34	52	-	-7	0.3	
35	52	-	-12	0.168	
36	47	*	-17	0.078	
37	44	*	-29	0.0046	
38	44	*	-32	0.0017	TREND (3)
39	39	*	-37	0.00018	
40	39	*	-36	0.000325	
41	27	*	-40	0.0000365	
42	31	-	-40	0.0000365	
43	35	-	-36	0.000325	
44	28	-	-35	0.00047	
45	31	-	-30	0.00345	
46	29	-	-26	0.01115	NO TREND
47	34		-17	0.078	
48	28		-14	0.127	
49	26		-15	0.108	
50	29		-6	0.332	

		IV ACTUALITY T	IV ACTUALITY NT
PRED	T	17	13
	NT	7	6
		24	17
		71%	35%

KENDALL'S S STATISTIC
WINDOW = 12
HDTR-4

CRITICAL VALUE S=22

PERIOD	DEM OBS	KENDALL'S S		PROBABILITY	
1	6				
2	10				
3	8				
4	10				
5	8				NO TREND
6	10				
7	5				
8	7				
9	9				
10	8				
11	11				
12	13		18	0.125	
13	10		13	0.21	TREND (1)
14	17	*	28	0.031	
15	19	*	34	0.01	
16	18	*	44	0.001	
17	21	*	49	0	
18	30	*	58	0	
19	27	*	56	0	
20	39	*	56	0	
21	42	*	58	0	
22	43	*	58	0	TREND (2)
23	51	*	60	0	
24	55	*	62	0	
25	57	*	62	0	
26	57	-	61	0	
27	53	-	57	0	
28	57	-	55	0	
29	56	-	49	0	
30	59	-	51	0	NO TREND
31	54	-	39	0.0035	
32	56	-	30	0.022	
33	58	-	28	0.031	
34	52		8	0.319	
35	52		-13	0.21	
36	47	*	-27	0.037	
37	44	*	-33	0.013	
38	44	*	-37	0.0055	TREND (3)
39	39	*	-49	0	
40	39	*	-52	0	
41	27	*	-57	0	
42	31	-	-55	0	
43	35	-	-55	0	
44	28	-	-55	0	
45	31	-	-50	0	
46	29	-	-47	0	NO TREND
47	34	-	-37	0.0055	
48	28	-	-34	0.01	
49	26	-	-35	0.0085	
50	29	-	-26	0.043	

		IV ACTUALITY T	IV ACTUALITY NT
PRED	T	18	17
	NT	4	0
		22	17
		82%	0%

REFERENCES

The reader is referred to the excellent list of references in Practical Nonparametric Statistics, 2nd edition, by W. J. Conover, 1980, J. Wiley & Sons, pp 394-426. We used some of them along with the book.

During the preparation of this paper, we used several papers not in Conover's list. These are given immediately below.

- Bennett, C. and Franklin, N. (1954). Statistical Analysis in Chemistry and the Chemical Industry, John Wiley and Sons
- Box, G. E. P. and Muller, M. E. (1958). A Note in the Generation of Normal Deviates, Annals of Mathematical Statistics, 28, pp 610-611
- Brown, B. and Maritz, J. (1982). Distribution-Free Methods in Regression, Australian Journal), pp 318-331
- Bunker, Tom CDR, SC, USN (1987) Demand Forecast Simulator, Navy Ships Parts Control Center, O. A. Report.
- Dietz, E. J. (1989). Teaching Regression in a Nonparametric Statistics Course, The American Statistician, 43, pp 35-40
- Fieller, E. C. (1932). The Distribution of the Index in a Normal Bivariate Population, Biometrika V-34, pp. 428-440
- Griffin, H. D. (1958). Graphic Computation of Tau as a Coefficient of Disarray, Journal of the American Statistical Association, 53, pp 441-447.
- Kaarsemaker, L. and van Wijngaaden, A. (1953). Tables for use in Rank Correlation, Statistica Neerlandica, 7, pp 41-54
- Kahn, Herman (1956). Application of Monte Carlo, R17-1237-AEC, The RAND Corporation
- King, J. (1979). Fitting the Median Regression Line, Team Easy Analysis Methods (TEAM), 6(1), pp 1-3
- Lieberson, S. (1961). Non-Graphic Computation of Kendall's Tau, The American Statistician, 15, pp 20-21
- Mann, H. B. (1945). Nonparametric Tests Against Trend, Econometrica, 13, pp 245-259
- Martin, F. F. (1968). Computer Modeling and Simulation, John Wiley and Sons
- Muller, M. E. (1959). Comparison of Methods for Generating Normal Deviation on Digital Computers, Journal of the Association of Computing Machinery, 6(3), pp 376-383

- Owen, D. B. (1962). Handbook of Statistical Tables, Addison-Wesley.

- Sillitto, G. P. (1947). The Distribution of Kendall's Tau Coefficient of Rank Correlation in Rankings Containing Ties, Biometrika, 34, pp 36-40

- Terpstra, T. J. (1952). The Asymptotic Normality and Consistency of Kendall's Test Against Trend when Ties are Present in our Ranking, Proceedings, Nerderlands, Akademie Wettenschappen, Series A, 55, 329-333.

- Thesen, Sun, Wang (1984). Some Efficient Random Number Generators for Micro-Computers, Proceedings of the 1984 Winter Simulation Conference edited by Sheppard, Pooch, Pegsen

- Tintner, G. (1940). The Variate Difference Method, Principia Press

- Wong, R. (1988). Kendall's Tau and Tests for Trend in Laboratory Data, Proceedings of the Biopharmaceutical Section of the American Statistical Association, pp 122-124

Abrams Tank Inspect And Repair Only As Necessary (IRON) Economic Model

Albert W. Van Horn

Systems and Cost Analysis Directorate (AMSTA-V)
U. S. Army TACOM
Warren, Michigan 48397-5000

Preceding Page Blank

Abrams Tank Inspect And Repair Only As Necessary (IRON) Economic Model

1. Introduction

On 22 March 1989 Lieutenant General Jimmy D. Ross the Deputy Chief of Staff for Logistics proposed an initiative to establish a point in the life of an Abrams tank when a depot level inspect and repair only as necessary (IRON) program may be appropriate to extend the service life of the tank. He recommended that TACOM develop an Economic Model to provide the cost-benefit of this IRON program. This model should predict the point in time when a tank becomes an IRON candidate. As proof of the benefit, the model must predict the extension of service life and the reduction of operating and sustainment costs through this remaining life. Finally, the results should be validated in a hardware demonstration.

In response to this request TACOM and PM Abrams have developed the Reliability Centered IRON (RCIRON) program. (The RCIRON program differs from the IRON program in that only certain candidate components are inspected and repaired.) For the purpose of a validation test 60 M1IP tanks were transferred from Germany to the National Training center (NTC), Fort Irwin, California. Fourteen of these tanks were selected for the RCIRON demonstration. These tanks were sent to Anniston in December 1989 where the RCIRON inspection was performed. These 60 tanks are currently being tested at the NTC. The goal of the RCIRON program is "preventive correction of impending failure" which can

Decrease the maintenance burden on field units.
Extend "combat life"
Decrease field Operating and Support (O&S) costs.

Our problem is to develop an optimum maintenance policy for the implementation of the RCIRON for the entire M1/M1A1 fleet. This requires the establishment of periodic inspections for a system composed of many parts, each of which has its own failure rate and its own required frequency of repair. Since each item has its own expected time between failures, each will have its own optimum time for periodic inspection.

To maintain a system at the desired level of operation a regular routine of test and inspection must be established. This type of inspection is normally performed at the organizational level. The maintenance organization may also perform preventive periodic inspections and repairs on a scheduled list of items or subsystems. There are, however, some parts of the system which may require inspection and repair at a higher echelon, such as a depot. It is assumed that by scheduling the inspection of item a little before the expected failure, the number of failures and the costs of unscheduled downtime can be reduced enough to offset the cost of scheduled maintenance. This is the basis of the RCIRON program. Fifty parts have been identified as candidates for the RCIRON Program. If there is evidence of failure or deterioration at the time of inspection these parts will be repaired or replaced.

2.0 Replacement models

A large number of the models reviewed during this study develop the optimum replacement time for a single type of part or component. A model proposed by Samuel B. Richmond was suggested as an economic useful life model for the RCIRON program.
The model equation is
were

$$\frac{f(T)}{q}\int_0^T tf(t)dt+Tf(t)-p=\frac{B}{A-B}$$

T = the preventive maintenance period
p = probability of failure before time T
f(t) = cumulative probability of failure
A = cost of overhaul after failure
B = cost overhaul before failure

This model considers that if the cost a scheduled replacement is less than the cost of a field replacement then the scheduled replacement is cost-effective. The model assumes that an item will always be replaced at RCIRON inspection, and the main cost will be the downtime required to perform the maintenance. The nature of the RCIRON process is to inspect a component and replace it only if necessary. This type of model does not meet the requirements of the RCIRON program.

3.0 Deteriorating Markov Process Model

A more realistic model is presented by A. S. Goldman and T. B. Slattery in their 1964 book Maintainability. With some modification this model can be adapted for the RC-IRON program. This model assumes that periodic inspections will occur at scheduled periods of time, miles or rounds. At the time of these periodic inspections there is a probability that the item will have deteriorated and should be replaced. There also exists a probability that the item will have failed prior to the scheduled inspection. Associated with these inspections and replacements are incremental costs which contribute to the total maintenance costs.

3.1 Model Requirements

The goal of a maintenance model is to develop an optimum maintenance policy. In order to develop a realistic model and to optimize the system retain certain system characteristics must be determined.
These characteristics are,

For each item:

1. Deterioration probability.
2. Probability of failure during deteriorated state.
3. Failure probability during normal state.
4. Average hours for preventive inspection.
5. Average hours for preventive repair.
6. Average elapsed time for field diagnosis and repair

7. Average hours for field repair

For the system:

1. Cost of setting up for preventive inspection.
2. Depot costs per hour.
3. Field costs per hour.
4. Cost of downtime

When this information is available the mathematical model can be developed. The objective is to develop a method of applying mathematics and economic theory to preventive maintenance problems. The method described applies only to preventive maintenance situation where the failure pattern is predictable. This pattern is generally associated with wear out failure distributions.

3.2 Mathematical Model

The periodic inspection model can be represented by the following expression:

$$C_i = \frac{(C1_i + C2_i \times P_i + C3_i \times E_i)}{T}$$

Where:

- T = Scheduled inspection period. (Time, miles, rounds)
- i = ith item to be inspected.
- Ci = Unit cost of maintenance and repair of the ith item of the system.
- C1i = Cost of periodic inspection of the ith item.
- C2i = Cost of preventive repair of the ith item.
- C3i = Cost of field failure of the ith item.
- Pi = Probability of repair of ith item.
- Ei = Expected number of failures between periodic inspections

The Goldman/Slatterly (G/S) model considers that the main cost of maintenance is the cost of lost downtime. This is also the main cost driver in the AMSAA model. The AMSAA model also includes the cost of the average hours required to replace the item. The G/S model also includes hours, but in a different way. The G/S model considers hours more as an overhead factor. For the RC-IRON program the AMSAA method will be used.

The cost of inspection and repair can then be written as:

$$C1_i = C_D \times I1_i + CI1_i \times TI1_i$$

$$C2_i = C_D \times D2_i + CR2_i \times TR2_i$$

$$C3_i = C_D \times D3_i + CR2 \times TR3_i$$

Where:

- C1i = Cost of periodic inspection of the ith item.
- C2i = Cost of preventive repair of the ith item.
- C3i = Cost of field failure of the ith item.
- CD = Cost of downtime. ($1781/day).
- I1i = Downtime due to inspection.
- CI1i = Cost/hour of inspectors.
- TI1i = Hours required for inspection.
- D2i = Downtime due to depot repair.
- CR2i = Cost/hour of depot repair.($46.69/Hr).
- TR2i = Hours required for depot repair.
- D3i = Downtime due to field repair.
- CR3i = Cost/hour of field repair.($104/Hr)
- TR3i = Hours required for field repair.

3.3 Total Systems Cost

The total systems cost is made up of the cost of setting up the inspection process, the overhead costs of maintaining the inspection and repair facility and the sum of all the part inspection and repair costs. If the inspection and repair facility was dedicated to the M1 program the total overhead costs would have to be included in the cost of preventive inspection and repair. The depots however are funded separately and the facility cost can be left out of the cost equation. The expression for total system cost can be written as:

$$C = \frac{C_0}{T} + \sum_{i}^{N} (C_i)$$

Were:

- C = Total System Cost
- Co = Cost of setting up inspection
- Ci = Inspection and repair costs of the ith item.
- T = Scheduled inspection period. (Time, miles, rounds)

3.4 Failure and Deterioration Probability

In order for the RC-IRON program to be successful items must be replaced before they have failed. One of the goals of the RC-IRON program is to develop methods of detecting the degree of deterioration of the items inspected. When this has been done an item will be in one of three states at the end of a time interval. The action taken will depend on which state the item is in at the time

of inspection. The state and actions taken are shown in Table I.

Table I

State of Item	Action Taken
1. Good	None.
2. Deteriorated	Repair at next scheduled inspection.
3. Failed	Repair immediately.

The data currently available from Sample Data Collection (SDC) provides an estimate of the probability that an item has either failed or is in good condition. This can be expressed as:

$$R+Q=1$$

Where:

$$R = \exp(-\alpha x t^{\beta})$$

and
alpha and β = parameters of Weibull distribution.
t = time, miles or rounds.
R = Reliability.
Q = Probability of a failure.

A model of the RCIRON process must include a third probability and the state of the item can be expressed as:

$$p1 + p2 + p3 = 1$$

Where:

p 1 = probability that item is in good condition.
p 2 = probability that item is in deteriorated state.
p3 = probability that item is in a failed state.

The method used to solve for the failure and deterioration probabilities in the G/S model is a deteriorating Markov process. For a given interval of time γ each item can be characterized by the probabilities $p_{i,j}$ that it starts in the time interval in state i and ends in state j. As an example $p_{1,1}$ is the probability that the item remains in the good condition through the interval and $p_{1,3}$ is the probability that item fails during the time period. These probabilities can be arranged in a matrix as a special form of a Markov process. This matrix is known as the transition matrix.

Arranged in a matrix these probabilities are:

$$\begin{array}{cc} & \text{FINALSTATE}(j) \\ \text{INITIALSTATE}(i) & \begin{vmatrix} P_{1,1} & P_{1,2} & P_{1,3} \\ 0 & P_{2,2} & P_{2,3} \\ 1 & 0 & 0 \end{vmatrix} \text{FOR A T} \end{array}$$

The process which can be described in this manner is a special case of a Markov chain. The probabilities that an item will be in the ith state at the end of the n+1th time interval are given by

$$p1(n), p2(n), p3(n) \begin{vmatrix} P_{1,1} & P_{1,2} & P_{1,3} \\ 0 & P_{2,2} & P_{2,3} \\ 1 & 0 & 0 \end{vmatrix} p1(n+1), p2(n+1), p3(n+1)$$

Which can be abbreviated as

$$[Pi(n)](pij) = [Pi(n+1)]$$

Since the item is assumed to be good at the beginning the condition for the item at the start can be written as

$$[Pi(0)] = [1,0,0]$$

The condition of the item after n transitions can then be written as

$$[Pi(n)] = [Pi(0)](pi)n$$

If T is the period between preventive inspections and γ is the period of transition the number of periods between inspections is equal to

$$N = T / \gamma$$

The number of preventive repairs during the life of the tank would then be equal to

$$R = p12 \times T1 / T$$

Were

P1 = number of preventive repairs
p12 = probability of transition to the deterioration state.
T1 = Useful life(time,miles,rounds)
T = Period between inspections

The expected number of failures can be computed using the expression

$$E_1 = \sum_{(j=1)}^{N_1} P3(j)$$

4. Some Concerns

A major area of concern is the applicability of the deteriorating Markov to a system were the time between inspections is large. Most applications of this model require that the time τ should be short enough so that only one change of state is likely but should be sufficiently long so that a repair could be done in τ time. Another problem is the use of a constant transition matrix regardless of the age of the system. These factors do not affect the validity of the basic model but limit the usefulness of the methods of calculating the failure and deterioration probabilities.

The model as presented represents a simplified solution to the preventive maintenance problem and presents the basics involved. There is very little data available on the deterioration rate of M1 components. It is hoped that the results of the RCIRON will provide some clues as to which components exhibit deterioration. Without this data the model given is of little practical value. The example, however, serves to illustrate the elements involved in a more general treatment of the RC-IRON problem.

An extensive literature search failed to turn up an analytical model that satisfies all the requirements of the RC-IRON process. To over come this and to develop an idea of how the system would preform it was decided to develop a simulation model.

5. Simulation Model

The data required by a simulation model is the same as that used in the Markov process model. The input requirements for the RCIRON model listed in paragraph 2.2 will provide the information necessary for a simulation model.

For each item the following parameters are required:

1. Miles to deterioration = T_i
2. Miles to failure during deteriorated state = D_i
3. Miles to failure during normal state = R_i
4. Average hours for preventive inspection = DTI_i
5. Average hours for preventive repair = RTD_i
6. Average down time at depot = DTD_i
7. Average down time for field diagnosis and repair = DTF_i
9. Average hours for field repair RTF_i
10. Miles at next failure = $NT_{j,i}$
11. Miles at next deterioration = $NR_{j,i}$
12. Cost of preventive inspection = $INSPC_i$

For the system:

1. Cost of setting up for preventive inspection =INSUP

2. Depot repair costs per hour = $46.69
3. Field repair costs per hour = $104.00
4. Cost of downtime = $1784.00
5. Inspection schedule (miles) = INSPECT
6. Miles at next inspection = NEXTINSP
7. Total vehicle mileage = TMILES
8. Vehicle life = 20 years.
9. Average annual mileage = YMILES
10. LIFE = 20*YMILES

A component will be in either of three states:

1. Operational
2. Deteriorated
3. Failed

If a component fails it is repaired and replaced immediately and returned to the operational state. If a component is in the deteriorated state it remains in that state until the next scheduled inspection. Each component can remain in the deteriorated state for a limited number of miles. If the deterioration miles plus the deterioration period is less than the miles for the next inspection the part fails and is repaired or replaced. The simulation begins by calculating the miles to the next failure and the next change to the deterioration state. This is represented by:

$NT_{1,i} = T_i$
$NR_{2,i} = R_i$

The times to failure are computed using a Weibull distribution. The expression for this calculation is given as:

T = EXP(LOG(LOG(1/(1-RND)))-LOG(ALPHAi)/BETAi)

where

ALPHA is the scale parameter
BETA is the shape parameter
RND is a random uniform deviate

The time to deterioration and the duration of the deteriorated state are not well known. Little or no data has been collected showing these factors. In order to run the model assumptions have been made to generate the deterioration time and duration of the deteriorated state. The first assumption is that the time for a component to enter the deteriorated state is less than the time to failure. Components are likely to deteriorate before they fail. The time spent in the deteriorated state would depend on a variety of maintenance factors. The deterioration could be so subtle that the tank crew may not realize that a problem exists. Other deterioration would be so obvious

that the crew would repair or replace the component after a very short deterioration period. For the sake of demonstrating the model it is assumed that the deterioration distribution is also a Weibull distribution with a scale parameter 2 times greater than the ALPHA of the failure distribution. The expression for the time to deterioration is:

$$T_i = \exp(\log(\log(\frac{1}{1-rnd}))) - \log(\frac{2\alpha_i}{\beta_i})$$

The life of the tank is assumed to be twenty years and the average yearly mileage is about 1000 miles. It is also assumed that a deteriorated component would not go undetected for 6 months or 500 miles. The time for a component to remain in the deteriorated state is the difference between the failure time and the deterioration time.

The rules for failing or passing inspection are:

1. If NT1i and NR1i < NEXTINSP then the component will fail and TMILES will be equal to either NT1,i or NR,i which ever is the smallest and a new NT1,i and a new NR1,i will be calculated as:

NT2,i = TMILES + Ti
NR2,i = TMILES + Ri

2. If NT1,i > NEXTINSP and NR1,i < NEXTINSP then the component will fail at the scheduled inspection and b e repaired or replaced. TMILES will be set equal to NEXTINSP and a new NTJ,I and a new NRj,i will be calculated as:

NT2,i = TMILES + Ti
NR2,i = TMILES + Ri
TMILES = NEXTINSP

Then a new NEXTINSP will be calculated as:

NEXTINSP = NEXTINSP + INSPECT

3. If NT1,i < NEXTINSP and NR1,i > NEXTINSP then the component will fail and be replaced immediately. TMILES will be set equal to NT1,i and a new NTj,i and a new NRj,i will be calculated as:

NT2,i = TMILES + Ti
NR2,i = TMILES + Ri

If NT2,i > NEXTINSP then:

TMILES = NEXTINSP

And a new NEXTINSP will be calculated as:

NEXTINSP = NEXTINSP + INSPECT

At each step in the calculation the cost associated with the inspection, field and depot repair are accumulated using the expressions:

FCOSTi = FCOSTi * DTFi +$1784 + RFTi * $104

RCOSTi = RCOSTi * DTDi +$1784 + RTDi * $104

ICOSTi = ICOSTi +INSPCi

The iteration continues until TMILES >= LIFE. After each run the total cost of maintenance for the vehicle life is calculated and printed in a file for further analysis. The scheduled miles between inspections is incremented and the simulation is repeated until the scheduled maintenance period is equal to the life of the vehicle.

6. Conclusions

Preliminary runs of the model were made using various deterioration factors and inspection times from 1000 miles to 20000 miles. The results obtained from running the model were disappointing. The model did not demonstrate that there was any advantage to the RCIRON process. Since we were dealing with an unproven model using fictitious data we decided to wait until the data from Anniston Depot and the tests at the NTC were available before presenting the results obtained by running the model.

The data obtained from Anniston did not provide us with any additional estimates of deterioration times.

The data from the tests at NTC however were very revealing. The tests at NTC were performed by three different organizations. Each organization was assigned to NTC for about two weeks during which time it was planned that the tanks would travel about 300 miles. Each of these periods is called a rotation. During the first rotation the raw data showed that RCIRON tanks had a mean miles between maintenance actions (MMBMA) of almost twice that of the control tanks.

There were several sources of variation affecting the results of the tests. The primary source of variation was the differences between the RCIRON and the control tanks.

The other known sources are the difference between rotations and the fact that each tank was driven a different number of miles. In order to determine the degree of variation between the RCIRON tanks and the control tanks the other sources of variation must be removed.

Several statistical test were performed to determine the degree of the RCIRON tanks and the control tanks. The difference between the two sample during rotation #1 if very significant. The difference during the other two rotations is much less significant.

During rotation #3 the test show that we can be more than 90% confident that the two sample are the same. From this we can conclude that the benefit of the RCIRON process is dissipated within the first 1000 miles after completion. These results are similar to those of the simulation model.

A Model for Optimally Reducing Uncertainty

Andrew Anderson Thompson III
Ballistic Research Laboratory

Abstract

This paper presents a model that can be used for the optimal reduction of uncertainty. There are three major components of the model. An a priori surface used to describe the original state of the world represents the measure that is to be minimized. A function representing the reduction of the measure as a function of the resource needs to be supplied. Based on these two functions a surface describing the optimal allocation of the resource can be derived.

Introduction

This paper presents a model that can be used for the optimal reduction of uncertainty. The components of the model will be presented with some examples of their interactions and then several successful interpretations of the model will be discussed. The paper concludes by suggesting some possible interpretations of the model. Hopefully, the reader will be able to find some useful interpretations of this model.

There are three major components of the model. An a priori surface used to describe the original state of the world represents the measure that is to be minimized. A function representing the reduction of the measure as a function of the resource needs to be supplied. Based on these two functions a surface describing the optimal allocation of the resource can be derived.

Uncertainty Surface

The uncertainty function represents the distribution of the attribute that is to be minimized. In the case of a lifeboat, a bivariate normal distribution with a variance dependent on time provides a useful realization of the uncertainty function as the probable unknown location. In many cases the uncertainty will be based on odds and thus will be a likelihood surface.

Information Function

The information function reduces the uncertainty as a function of the resource applied. This is the mechanism that transfers the measure of the uncertainty function to a discrete category representing gain. This gain

can be thought of as knowledge. A desirable property to require is that the remaining uncertainty at a point is the same for two applications of resource to a point and for an amount of resource equal to the sum of the two but only applied once. Mathematically, the following property is required

$$f(l(p),r_1+r_2)=f(f(l(p),r_1),r_2).$$

In the above equation, l(p) represents the likelihood or uncertainty at a point and r represents an amount of resource. A function that has this property is the exponential function. Using the exponential function as the information function and w(p) to represent the amount of reduction at a given point the amount of uncertainty remaining can be represented as

$$l(p)e^{-w(p)r}.$$

The amount transferred to expected information is

$$l(p)(1-e^{-w(p)r}).$$

Optimal Resource Surface

Koopman (1979) presents a theorem, based on methods used by J. Willard Gibbs in thermodynamics, to define the properties of a likelihood surface after an optimal resource allocation is completed. After an optimal allocation, the likelihood density will have equal values in the area of allocation and be less in the other areas. Performing an optimal allocation amounts to passing a plane, that is parallel to the independent variables through the target likelihood density so that the mass above the plane is equal to the available resource. An optimal allocation divides the potential area into an area of resource application and an area to be ignored. In the area of allocation, the effort expended at each location will be proportional to the amount of likelihood mass that is above the "cutoff" plane. One way to visualize this is to consider the change in the gain of information, and always apply resource to the regions associated with the largest gains. First consider

$$\frac{\partial}{\partial r}l(p)(1-e^{-w(p)r})\Big|_{r=0}=l(p)w(p)$$

Imagine a gridlike partition of $l(p)w(p)$.

Allocate resource to the cells with the largest value until they hit the level of the next lower level.

Repeat this until all the resource is used.

The result is a surface that is flat in the region of uncertainty reduction. One can imagine that the grid becomes the set of rational numbers.

Example 1.

As an illustration of the above ideas, three methods for solving a single problem will be discussed. Consider a 3x3 matrix, where the value in cell (i,j) [i and j running from 1 to 3] represents the uncertainty associated with cell (i,j).

i=1	0.20	0.10	0.05
=2	0.10	0.30	0.05
=3	0.05	0.10	0.05
	j=1	=2	=3

The probability associated with the region corresponding to cell (2,2) is 0.3. Assume there is a fifty-percent reduction in the probability if a unit of resource is applied to a cell. Fifty percent is transfered to the catagory information gain. What sort of resource allocation will maximize the probability for twenty units of resource?

1. Method One: Proceed sequentially using a maximum-likelihood method; apply each unit to the area with the highest probability. After each unit is applied, replace the original cell [call it C] with "C (1-DF)," or, here, (0.3 x 1/2 =) 0.15. (That is, some of the probability is moved to the category of knowledge.) For the allocation of the twenty rounds, the procedure would follow this pattern:

Step 1: Choose cell (2,2), since it contains the greatest probability.
Replace 0.3 with (0.5 x 0.3 =) 0.15.

Step 2: Choose cell (1,1), since it now contains the greatest probability.
Replace 0.2 with (0.5 x 0.2 =) 0.1.

Step 3: Choose cell (2,2), since it now contains the greatest probability.
Replace 0.15 with (0.5 x 0.15 =) 0.075

Now several cells have the equal maximum probability of 0.10. Here, one can randomly choose any one of those cells. (When this possibility exists, one can only speak of "an" optimal solution.)

Method One not only indicates the number of units to be applied to each cell but also indicates the best sequence for delivery. This straightforward method can be used when one has a discrete uncertainty surface.

2. Method Two. This method introduces the techniques used when one has a continuous uncertainty surface. The method is to find the amount of resource to apply to the highest-valued cell to reduce it to the level of the cell containing the next lower value. Then apply the resource to both of those cells until level of the next lower cell is reached. In this situation, we assume that the effort is also continuous (i.e., that the resource can be applied in fractional amounts). This assumption is necessary in order to calculate the optimal-resource surface.

If a unit is applied to cell (2,2), then the probability remaining would be 0.15. Since this value is lower than 0.2, the reduction has gone too far. The proper amount to apply to cell (2,2) is a fractional amount, enough to have reduced it to 0.2 and no lower. Thus:

Step 1: Reduce the value of cell (2,2) to the value of cell (1,1)

$$0.2 = 0.3*0.5**n \rightarrow n = 0.585.$$

Step 2: Find the amount of effort that will reduce 0.2 to 0.1

$$0.1 = 0.2*0.5**n \rightarrow n = 1$$

Step 3: Find the amount of effort that will reduce 0.1 to 0.05

$$0.05 = 0.1*0.5**n \rightarrow n = 1$$

The derivation thus far has not made use of the fact that the resource is limited. In this continuous case, the optimal allocation of ammunition can now be derived by means of an "Effort Matrix" expressed in terms of "E" -- the "minimal-effort value." The value in each cell represents the effort or resource to be applied to the corresponding area.

For the Effort Matrix corresponding to the given problem, each of the cells with probabilities of 0.05 has some undetermined amount of

effort E in it. The cells corresponding to the probabilty cells with entries of 0.1 have one full unit more of effort in them, so they carry the value 1+E. The cell corresponding to the value of 0.2 has yet one more unit of effort, so its entry is 2+E. Finally, the cell with a value of 0.3 has an additional 0.58 units of effort (rounded to two decimal places). Its entry is 2.58+E.

2 + E	1 + E	E
1 + E	2.58 + E	E
E	1 + E	E

The total effort expended is to be equal to twenty so by summing the cells of the matrix and setting the total equal to twenty, we can find the optimal effort for each cell:

$$20 = 9E + 7.58$$
$$E = 1.38$$

Thus, where the resource can be applied continuously, the optimal solution is:

3.38	2.38	1.38
2.38	3.96	1.38
1.38	2.38	1.38

Returning to the original case of 20 integral units, if the resource could not be subdivided, it would be necessary to enter integer numbers of rounds in each cell. As an approximation, we would round off the values in the cells to integer amounts and make further adjustments to ensure that the sum is twenty. One plausible solution in this case is

3	3	1
3	4	1
1	3	1

3. Method Three. This method uses the ideas of Koopman; Gibbs (1928) originally applied these ideas to a physics problem. Koopman derives two formulas that can be used to find the optimal resource surface. The first formula is used to find the area to apply resource. After this area has been defined, the second formula is used to determine the amount to apply to each point. This method is valid when the TLD is continuous and is a formalization of the technique used in Method 2. The equations used are:

$$\Phi = \iint_A \left[(\ln (p(x,y)\, w(x,y)) - \ln \lambda)/w(x,y) \right] dxdy \qquad (1)$$

and

$$\phi(x,y) = \frac{1}{w(x,y)} \ln\left[\frac{p(x,y)\,w(x,y)}{\lambda}\right] \tag{2}$$

Φ is the total effort ϕ (x,y) is the resource density at x,y

p(x,y) is the probability at x,y

λ is the height of the cut off plane

A is the area of application

$e^{-w(x,y)}$ is the residual probability located at (x,y).

Equation 1 divides the uncertainty region into the two areas based on the amount of resource available. Equation 2 is used to determine the allocation at each point. As applied to the current example the steps are as follows:

First, express the residual probability as an exponential. As we apply more resource to a specific region the returns on each unit diminish in proportion to the probability that remains in that region. This diminishing rate of return is captured by the exponential function.

Second, solve Equation 1 for λ. λ is the height of the plane that cuts the uncertainty surface at the level appropriate for that amount of effort. (Note that integration can be replaced by summation for this discrete case).

Third, find the amount of effort at each of the nine points using Equation 2.

Fourth, find an integer solution.

Implementing these steps yields the following:

$$e^{-w(x,y)} = p_k \rightarrow w(x,y) = -ln.5 = .6931$$

Note that w(x,y) is constant and can be replaced by w.

$$20 = \sum_{i=1}^{9} (\ln(p_i * w) - \ln\lambda)/w$$

$$\ln\lambda = \frac{20w - \sum \ln p_i\, w}{-9} \rightarrow \lambda = .0133$$

Using $\phi_i = \frac{1}{w} \ln \left(\frac{p_i * w}{\lambda} \right)$ we get the optimal effort matrix

3.3819	2.3818	1.3817
2.3818	3.9669	1.3817
1.3817	2.3818	1.3817

Integerization of this solution yields the same result as the previous example. Note that this method will work for a continuous uncertainty surface.

Next an example using continuous resource on the bivarite normal distribution is presented. It is hoped that this example will serve to clarify all the above ideas.

Bivariate Normal

In the continuous case Koopman's result guides us in finding the value for the cutoff plane for a given amount of effort. In using this equation note that the resource surface must be greater than zero at all points; we cannot use negative amounts of resource at an unlikely area and counterbalance this by applying more resource to more probable area. The value z = ln (λ) is the cutoff plane of the surface ln (p(x,y) w(x,y)). All regions where ln (p(x,y) w) < ln (λ) are ignored. Assuming w(x,y) = w and ln (λ) can be expressed as ln ($p(x',y')$ w) Equation 1 can be expressed as

$$\Phi = \iint_A \Big[\ln (p(x,y)\, w) - \ln (p(x',y')\, w) \Big] dxdy / w. \tag{3}$$

This can be rewritten as

$$w\,\Phi = \iint_A \Big[\ln (p(x,y)) + \ln w - \ln(p(x',y')) - \ln w \Big] dxdy \tag{4}$$

$$= \iint_A (\ln p(x,y) - \ln p(x',y'))dxdy$$

In this situation Equation 4 shows ln λ is inversely related to w. Equation 2 can be written as

$$\phi(x,y) = \frac{1}{w} [\ln p(x,y) - \ln p(x',y')]. \tag{5}$$

This indicates the amount of resource applied to each point is proportional to the difference between the distribution and the cutoff plane. These ideas are applied to the following problem.

Example 2:

Given a circular normal probability surface and constant information function, describe the optimal resource surface. The uncertainty is described by

$$p(x,y) = (2\pi\sigma^2)^{-1} e^{\frac{-(x^2+y^2)}{2\sigma^2}} .$$

Taking the natural log we have

$$\ln p(x,y) = \ln (2\pi\sigma^2)^{-1} + \frac{-(x^2+y^2)}{2\sigma^2}.$$

Rewriting Equation 1 for this problem we have

$$\Phi = \iint_A \left[\ln (2\pi\sigma^2)^{-1} + \frac{-(x^2+y^2)}{2\sigma^2} + \ln w - \ln \lambda \right] dxdy/w.$$

The integrand is the equation of a concave downward parabaloid.

Next change to polar coordinates

$$w\,\Phi = \int_0^{2\pi}\int_0^{A} \left[\ln (2\pi\sigma^2)^{-1} + \frac{-R^2}{2\sigma^2} + \ln w - \ln \lambda \right] RdRd\Theta \qquad (6)$$

The value in brackets as previously mentioned must be greater or equal to zero. In terms of A we will use the following expression for $\ln \lambda$

$$\ln \lambda = \ln (2\pi\sigma^2)^{-1} + \frac{-A^2}{2\sigma^2} + \ln w.$$

Equation 6 can be written

$$w\,\Phi = \int_0^{2\pi}\int_0^{A} \left[\ln (2\pi\sigma^2)^{-1} + \frac{-R^2}{2\sigma^2} + \ln w - \ln (2\pi\sigma^2)^{-1} - \frac{-A^2}{2\sigma^2} - \ln w \right] RdRd\Theta$$

$$= \int_0^{2\pi}\int_0^{A} \left[\frac{-R^3}{2\sigma^2} + \frac{A^2R}{2\sigma^2} \right] dRd\Theta$$

$$= \int_0^{2\pi} \left(\frac{-A^4}{8\sigma^2} + \frac{A^4}{4\sigma^2} \right) d\Theta$$

$$= \frac{2\pi A^4}{8\sigma^2} = \frac{\pi A^4}{4\sigma^2} \rightarrow A^4 = \frac{4 w \Phi \sigma^2}{\pi}$$

$$\rightarrow \lambda = (2\pi\sigma^2)^{-1} e^{\frac{-1}{\sigma}} \left(\frac{w\Phi}{\pi}\right)^{1/2} w.$$

The optimal resource surface will be zero outside the circle of radius A; within the circle the allocation is given by Equation 2 which simplifies to

$$(A^2 - R^2)/(w\, 2\sigma^2)$$

The information gain for this example can be found by the following method. Note from Figure 1 that P(information) is the volume bound by the curve p(R,Θ) and the cutoff plane z = λ or the difference between Figure 1a and 1b. The volume under the circular normal distribution is given by

$$1 - e^{\frac{-R^2}{2\sigma^2}}$$

This volume contains a cylinder of radius A and height p(A); so we must remove this volume from the previous value. The volume of the cylinder is

$$\pi A^2 h = 2\pi\sigma \left(\frac{\Phi w}{\pi}\right)^{1/2} \frac{1}{2\pi\sigma^2} e^{\frac{-1}{\sigma}\left(\frac{\Phi w}{\pi}\right)^{1/2}}$$

so the expression for p(information) is

$$1 - \left(1 + \frac{1}{\sigma}\left(\frac{\Phi w}{\pi}\right)^{1/2}\right) e^{\frac{-1}{\sigma}\left(\frac{\Phi w}{\pi}\right)^{1/2}} \qquad (7)$$

Next we extend this problem to quantify the relationship between intelligence or reduced variance of the uncertainty distribution and increasing the amount of resource. Assume a circular normal density with sigma of 100. Let w = .5 (the probability of missing the target is $e^{-.5}$ or .61) and suppose there are ten units of resource, each unit having an effective radius of thirty meters. Notice that both the standard deviation of the uncertainty and the total amount of effort need to be in the same units; thus the total effort needs to be an area. For this example the total

effort available (Φ) is the number of resource units multiplied by the area each covers or $10 * \pi * 30^2$. From Equation 7 above the probability of hit is

$$1 - \left(1 + \frac{1}{100}\left(\frac{10*.5*900*\pi}{\pi}\right)^{1/2}\right) e^{\frac{-1}{100}(4500)^{1/2}} = .146$$

If we instead use 20 units, the information gain is .246; however, if we had reduced the target location error by fifty percent so sigma was 50 for 10 units the p(information) would be .388. A reduction in sigma to 70 increases the probability of information to the same level as doubling the amount of resource. These observations give guidelines for analysis of the benefits of intelligence versus increasing the resource allocation.

Theory of Search

Koopman worked for the Navy during WWII and directed his energies toward solving problems associated with getting convoys across the Atlantic, finding submarines, and directing maritime rescue operations. During the course of his work he developed a formal theory of search. The model presented by this paper is a slight generalization of search theory. In order to describe the a priori location density of a target, knowledge of the situation based on past experience or functional relationships must suffice. For maritime rescue operations, the bivarite normal distribution proved to be an adequate model. The detection function was based on models of human observation from airplanes and several adequate glimpse models were derived. Using these functions it was posible to allocate search time judisously. There was a major problem getting convoys across the Atlantic. Using destroyer escort time as a resource to minimize the probability of attack by a submarine and reasoning about submarine tactics to build up probability of attack surfaces; it was possible for Koopman and his associates to make reasonable allocations of destroyer time and substantially decrease the sucessful the effects of the submarines.

Artillery Effectiveness

Problems associated with increasing the effectiveness of indirect artillery fire have been addressed by Sandmeyer(1986). Targeting errors, target size, and errors associated with a round need to be considered.

In order to see the connection between theory of search problems and artillery problems, think of an artillery round as the detection function. The round searches for the target. For a given number of rounds, the goal

is to diminish the target likelihood density as much as possible. The expected damage of a round is the convolution of the precision errors and the damage function of the round. For artillery systems this convolution will usually involve the bivariate normal distribution and the von Neuman - Carlton damage function. The target likelihood density is the convolution of the target area, the target location errors, and the artillery mean point of impact errors. The latter two are both assumed to be bivariate normal distributions, while the target area function is usually a rectangle whose height is the reciprocal of the target area. The optimal amount of target damage can be found by using these two convoluted functions and Koopman's theorem.

Our total effort is packaged into a discrete number of artillery rounds, or expected damage functions, each of which have the same area effects. It is not possible to apply the indicated optimal effort or damage to each point; thus, it is impossible to achieve the optimal amount of damage. This optimal damage surface will serve as an upper bound on performance.

Recalling that the target likelihood density is broken into a part to be searched and an area to be ignored, several theoretical observations can be made. First, the amount of damage occurring outside the optimal area is wasted effort in the sense that it could be applied more effectively elsewhere. Next, damage exceeding that indicated by the cutoff plane in the search area is wasted effort. The best approach is to try to approximate the optimal effort surface by the summation of translated expected damage functions. In most cases this will result in a target likelihood density with ripples near the level of the optimal cutoff plane.

Richard Sandmeyer (1986) of the AMSAA developed a technique to optimally approximate the optimal effort surface. There are three major sections to his method.

1. Calculate the upper bound and optimal effort surface.
2. Calculate an approximate solution by using either integer programming or least squares techniques on a restricted domain. The restricted domain consists of all the points defined by the intersections of grid that includes the optimal effort surface.
3. Starting with the previous solution, use a steepest descent technique to arrive at the solution.

The theoretic upper bound is found based on Koopman's theorem. The optimal effort surface can be calculated functionally for each point. In finding this surface, it is assumed there are no restrictions on the effort applied at each point.

Step two of Sandmeyer's method involves finding a close approximation of the best set of translated damage functions. This is accomplished by looking at a discrete number of points and optimizing the approximation over this set.

The third step of his method starts with the result of step two and allows the aimpoints to drift into the positions of the best approximation. Computationally the third step is the most expensive. Happily, for most situations the result of the second step is within a few percent of the upper bound. This method requires too much CPU time to be used in the field. As Sandmeyer has already suggested the performance of this method can be used to evaluate the effectiveness of other methods. It is clear that the approach developed by Sandmeyer is not limited to solving problems associated with artillery.

Medical Diagnosis

In this situtation, an unknown desease presents itself through a subject or patient as a set of symptoms. The goal is to identify the desease or short of that find a treatment that gets rid of the symptoms. The parameter space is the set of all known medical ailments; the goal is to reduce the uncertainty as to the patient's condition in some optimal fashion. The resource available to the physician is money, which is translated into information via various microbiological and chemical tests and through response to particular medical treatments. An a priori likelihood surface can be built up from existing desease frequencies, knowledge of the patient, and risk factors. After this a sequence of treatment and testing can be developed that would be optimal in reducing the uncertainty surface associated with the patient. This would give both the patient and the physician a formal method that explains a particular course of action.

Funding Problems

This class of problems arise when there is a limited amount of resource to spend on a number of problems. The uncertainty surface can be interpreted as the likelihood of benefit to society. The different programs will have different reduction rates and these can be estimated from historical data (or from pilot studies). In this case funding would allocated automatically after the politicians charactorized the uncertainty surface and the social scientists evaluated the effectiveness of the various programs.

Conclusion

It is my hope that the ideas developed by search theory will find a wider

application when presented as a general model for reasoning. Sandmeyers work has increased the domain of the problems by developing approximation methods. On an intuitive level this model of reasoning can be seen to operate in many fields. The advantages of having a formal model are to be seen in a wider scope of applications and in more efficient problem solving.

An Iterative Technique for Target Detection and Segmentation in IR Imaging Systems

Duc M. Nguyen
US Army, Center for Night Vision and Electro-Optics
Fort Belvoir, VA 22060-5677
Revised Nov 9th, 1990

Abstract

This paper describes a method to detect and segment ground targets in IR imaging systems using statistical models for target and background. The detection method is an iterative technique that subtracts the background from the scene, using a double window filter with the assumed independent distribution function for background. The outer size of the double window is determined iteratively by the largest connected component remaining in the scene. The segmentation or boundary detection methods employs the well-known statistical theory of change detection in speech segmentation. The likelihood ratio test is modified to improve the performance. The design method was initially applied on laser radar range imagery, but simulation results showed that the method works well for both flir and laser radar range image, hence promising an advanced technique for multisensor fusion algorithm. Hypothesis modelling for target and background is analyzed and a brief discussion of the fusion algorithm is also included.

1. - Introduction

The double window filtering technique has been used by many investigators for target detection for several years[1]. The double window filter approach is a general technique which tests a statistic calculated from the target and the background. The sizes of the inner and the outer windows of the filter are usually fixed, according to an estimation of the smallest and the largest target sizes. Several factors limit the performance of the filter, such as the requirement of knowledge of target size , time consuming and the erratic behavior of the pixel classification near the target boundary or inside a bimodal target. In this paper, we propose a technique that attacks all those above problems by : i) updating apriori knowledge of target size by a feedback loop, ii) speed up the computational time of the filter by jumping the window instead of sliding it pixel by pixel, and iii) testing the statistic of a pixel by generating several outer window statistics instead of using its neighboring pixel. The detection method is further improved by incorporating a second test, which measures the statistic of a connected component against its immediate surrounding pixels to reject possible clutter components. The detected boundary is then refined using the log-likelihood ratio

Preceding Page Blank

test which detects the boundary without requiring a threshold. This likelihood ratio test is modified with apriori knowledge (i.e., knowing the direction of changes) to improve the performance.

2. - Hypothesis modelling for target and background

Theorem [2] : let $x1, x2, \ldots\ldots, xn$ be independent and identically distributed (iid) random variables

and $Sn = x1 + x2 + \ldots\ldots + xn$

let $$Y = \frac{Sn - E(Sn)}{\sigma(Sn)}$$

where $E(Sn)$ is the estimation of Sn and $\sigma(Sn)$ is the standard deviation of Sn,
then random variable Y
a) has mean 0, variance of 1
b) is the sum of small independent random variables.

2.1 - Double window filter

The double window filter provides a technique to test the statistics of the pixels inside the inner window against the pixels inside the outer window, see Figure 1.

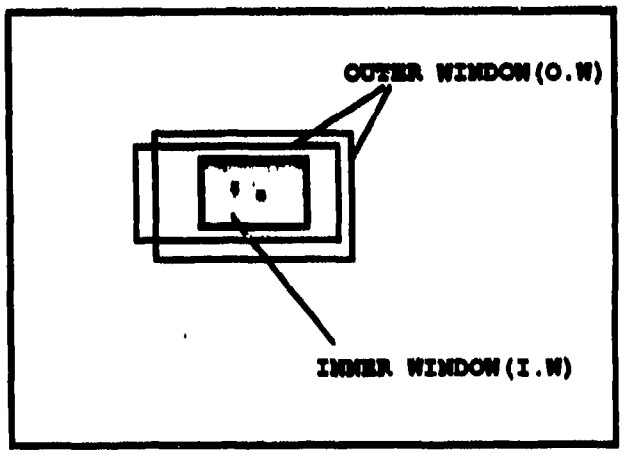


FIGURE 1

Let $$Sij = \frac{Xij - E(Xij \text{ such that } ij \in O.W)}{\sigma(Xij \text{ such that } ij \in O.W)} \quad \text{for } ij \in I.W$$

where Xij are random variables of the array or image.

If the pixels inside the inner and the outer window are iid then, Sij has mean 0 and variance of 1.

2

2.2 - Hypothesis models

Let Ho denote the hypothesis for background
H1 denote the hypothesis for target

and

Hypothesis Ho : $(R_{ij} / \theta = \theta_0)$

versus $\theta_0 \neq \theta_1$

H1 : $(R_{ij} / \theta = \theta_1)$

Here Rij represents the intensity of the ijth pixel and θ_{ij} represents a feature extracted from Rij that we may use to distinguish the target from the background. In flir imagery, the thermal intensity of the target is usually different from the surrounding background. In laser radar range imagery, at low depression angle, the target can be distinguished from the background by observing the variation of the measured ambiguous range, because it is often reasonable to assume the background consist of rough surfaces. Therefore, a 3X3 window is used to convert the range image to the variance image as a preprocessing (or feature extraction) step before using the double window filter.

Hence,

for each pixel inside the inner window, compute

$$S_{ij} = \frac{\theta_{ij} - E (\theta_{ij} \text{ such that } ij \in O.W)}{\sigma (\theta_{ij} \text{ such that } ij \in O.W)}$$

O.W : outer window

where θ_{ij} is the mean of Rij for flir images and θ_{ij} is the 3x3 window variance of Rij for laser radar range images.

Then,

$$H_0 : S_{ij} \sim N (0 , 1)$$

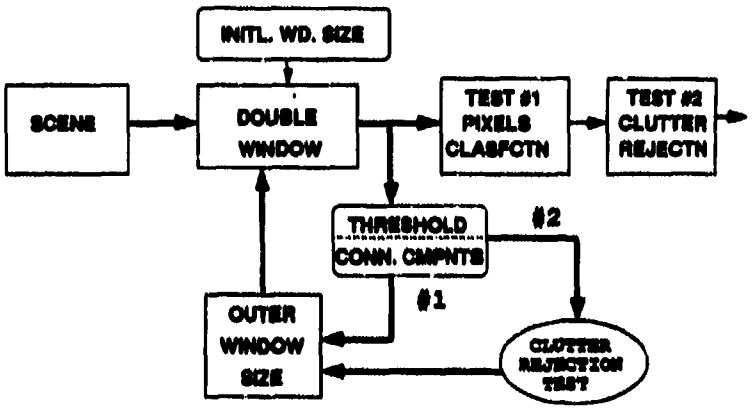
$$H_1 : S_{ij} \text{ not belong to } H_0$$

where N(0,1) denotes a normal distribution with mean 0 and variance of 1. Note that the approximated normal distribution is reasonable for both sensors because: i) for flir imagery, the thermal intensities can be assumed to be small independent random variables; ii) for laser radar range imagery, the 3X3 window variances of the ambiguous ranges were approximated to be multivariate Gaussian variables[3].

3. - Implementation of the detection algorithm

Figure 2 illustrates the implementation of the feedback loop for an ROI (region-of-interest) detection algorithm with two options labeled as #1 and #2. The first option determines the size of the largest connected component in the scene, whereas the second option determines the largest sizes, in horizontal and vertical, of the connected components remaining in the scene after performing

FIGURE 2 - BLOCK DIAGRAM OF THE DETECTION METHOD



a simple clutter rejection test. The outer window size is first initialized and then the double window moves across the entire image in such a way that its inner windows, which its size is half of the outer window size, do not overlap, but are adjacent to each other. This speeds up the computational time of the double window algorithm, compared to the conventional method which moves pixel by pixel[1]. Each pixel inside the inner window is tested against the statistics of pixels inside the outer window, for which the connected components are formed by pixels passing a threshold value which is normalized to the outer window statistics (explained in section 3.1 below). The sizes of these connected components are then determined and fed back to the double window filter process. This process is iteratively repeated while accumulating the statistic for each pixel for each iteration. Pixel classification is then performed to form ROI connected components. Each ROI is now tested against its immediate surrounding background pixel statistic to reject possible clutter component.

3.1 - Hypothesis test #1 (pixel classification)

Recall that Sij is assumed to have mean 0 and variance of 1 approximately if Sij belonged to the background. Therefore, a threshold can be chosen base on the justification of the probability of error. That is, the probabilistic value of Sij can be estimated corresponding to the total number of iterations, and is tested to determine whether or not Sij belongs to the background, which has mean 0 and variance of 1.
For example, a Student-t-test[4] may be formed as

$$\text{test} \quad \frac{E(Sij) - z_{\alpha}}{\sigma(Sij)/\sqrt{n}} < t_{\alpha} \quad \text{(Student-t-test)}$$

where,

4

z_α : critical value of z for areas under the normal curve
t_α : critical value of the t-distribution
n : total number of iterations
α : degree of confidence level

The number of iterations is usually greater than 10 when using the student-t-test. Moreover, this requires the accumulation of all the statistical values of Sij.

For more convenience, fewer computations and storages, a binomial trial method can be used to test the hypothesis. For example,

$$\text{let} \quad Xij = 0 \quad \text{if} \quad Sij < z_\alpha$$
$$= 1 \quad \text{otherwise}$$

$$\text{and} \quad P(Xij = 0) = P(Sij < z_\alpha) = p$$

then for n trials, the probability that X = 0 occurred k times is,

$$P(k) = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$

Hypothesis test : accept H0 if $k > r$

H1 if $k \leq r$

then

$$P(H1/H0) = \sum_{k=0}^{r} P(k)$$

From standard statistical tables which are available in most statistical text books[4], we can justify n, r, and p to obtain the desired value of P(H1/H0). For example,

From Table IV, page 513 of Reference[4], if $z_\alpha = 0.9$ then p = 0.815 at the significance level of 0.0005; From Table II, if n = 5 and r = 1 then P(H1/H0) < 0.0067.

Since the assumption that the pixel values of background are independently distributed may not be true in practice, we have to justify the value of parameters based on the experiment. The selected value of z_α can be determined either from mathematic model or from the histogram.

With the selected value of z_α , connected components are formed by ijth pixel such that the value of Sij is greater than z_α for each iteration. The outer window size can be determined according to the sizes of these connected components for each iteration, and the process is repeated with this new value.

3.2 - Hypothesis Test #2 (clutter rejection)

From the previous process, the connected components are formed by adjacent pixels such that for n iterations, P(Xij = 1) > 1 -

P(H1/H0). In this process, we will test the statistics of these components against the immediate surrounding background pixels to reject possible clutter components.

Let Ti and Bi be set of pixels as shown in Figure 3,

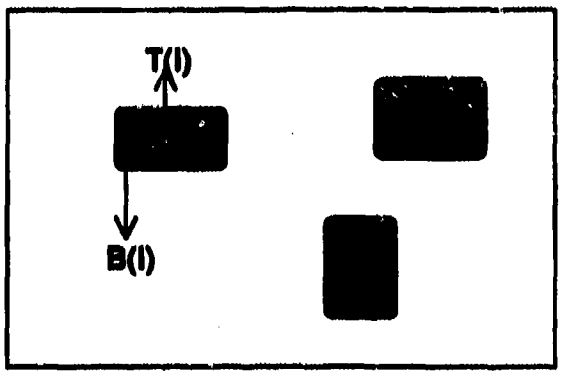


FIGURE 3

and let $$D_i = \frac{E(\Theta_{Ti}) - E(\Theta_{Bi})}{\sqrt{\sigma^2(\Theta_{Ti}) + \sigma^2(\Theta_{Bi})}}$$

where

$$E(\Theta_Z) = E(\ \Theta ij \text{ such that } ij \in Z\ )$$

$$\sigma(\Theta_Z) = \sigma(\ \Theta ij \text{ such that } ij \in Z\ )$$

The Student-t-test may be used in this case. Once again, justification must be based on the real data experiments, and this test (SNR test) works as a simple clutter rejection method.

Suppose we have,

Hypothesis HO : $D \leq 1$

versus H1 : $D > 1$

then $P(H1/H0) \sim 1 - P(z_\alpha < 1) \sim 1 - 0.8413 = 0.1587$

This simple clutter rejection test uses the SNR and estimates the sizes and shapes of the connected components to reject possible clutter components. This process is illutrated in Figure 2 as the option #2 in the feedback loop.

However, the final clutter rejection test is more important. This test is derived from the analysis of the target and background statistics to enhance the performance.

For example, let D is a function of means and variances of target and background,

$$D = f(m_T, \sigma_T, m_B, \sigma_B)$$

Assume that target mean and variance are greater than background mean and variance, respectively. Then,

$$D = (m_T \circ \sigma_T) * (m_B \circ \sigma_B)$$

where $\circ$ denotes a proportional operator and $*$ denotes an improportional operator, i.e,

$$\text{if } m_T > m_B \text{ then } m_T - m_B > 0 \text{ or } m_T/m_B > 1 .$$

Therefore, the improportional operator can be a subtraction or a division.

For flir imagery,

$$\text{assume} \quad m_T > m_B , \quad \sigma_T \neq \sigma_B$$

$$\text{then} \quad D = \frac{m_T - m_B}{\sqrt{\sigma_T^2 + \sigma_B^2}} + \sqrt{\frac{|\sigma_T^2 - \sigma_B^2|}{\sigma_T^2 + \sigma_B^2}}$$

For laser radar imagery,

$$\text{assume} \quad m_B > m_T , \quad \sigma_T < \sigma_B$$

$$\text{then} \quad D = \frac{m_B - m_T}{\sigma_T} + \frac{\sigma_B}{\sqrt{\sigma_T^2 + \sigma_B^2}}$$

$$\text{Hypothesis} \quad H0 : D \leq \lambda$$

$$H1 : D > \lambda$$

where λ is a constant, to be determined from the empirical distribution. Note that, this constant can be modeled as a function of range and weather conditions. In fact, these factors make the target distribution function varied relatively to the background distribution. Therefore, a constant setting threshold will have false alarms or miss some detections in some images, i.e., close or long ranges images.

4. - Boundary Detection Method

4.1 - Statistical Theory of Change Detection [5]

The use of the statistical theory of change detection in speech segmentation has been known and very successful. The results are impressive and well documented. Since the methods are general, they are well suited for application to the detection of the boundaries of objects in IR images. The likelihood ratio test is the most common hypothesis test used to detect the time change of an unknown autoregressive (AR) model.

$$\text{Let} \quad y_n = \sum_{i=1}^{p} a_i y_{n-i} + v_n$$

7

where $\quad var(v_n) = \sigma^2$

$$\Theta = (a_1, a_2, \ldots\ldots, a_p ; \sigma^2)$$

Hypothesis $\quad$ H0 : $\quad \Theta = \Theta_0 \quad 1,2,\ldots\ldots,n$

H1 : $\quad \Theta = \Theta_1 \quad 1,2,\ldots\ldots,r \qquad r?$

$\Theta = \Theta_2 \quad r+1,\ldots\ldots,n$

r and Θ's unknown

Here H0 denote a detection of no change and H1 denote a detection of change.

Likelihood ratio test follows

$$\max_{r} \; \max_{\Theta_1, \Theta_2} \; \min_{\Theta_0} \; \log L(H1/H0) > \lambda$$

which yields,

$$\max_{r} \; P(r) > \lambda$$

$$P(r) = n \log \hat{\sigma}_0^2 - r \log \hat{\sigma}_1^2 - (n-r) \log \hat{\sigma}_2^2$$

$$\hat{\sigma}^2(W) = \min_{\Theta} \frac{1}{(\#W)} \sum_{k \in W} \left(y_k - \sum_{i=1}^{p} a_i \, y_{k-i} \right)^2$$

$$\hat{\Theta}(W) = \arg \min_{\Theta} \sum_{k \in W} \left(y_k - \sum_{i=1}^{p} a_i \, y_{k-i} \right)^2$$

where W denotes any one of the 3 window $(1,2,\ldots\ldots,n)$, $(1,2,\ldots\ldots,r)$, $(r+1,\ldots\ldots,n)$ and # denotes number of the elements of the referred set.

4.2 - Modified likelihood ratio test

For our application, only H1 is considered and the log likelihood ratio test can be simplified to,

$$\min \; P(r) = r \log \sigma^2(r) + (n-r) \log \sigma^2(n-r)$$

where, $\sigma^2(r)$: sampled variance of the set $(1,2,\ldots\ldots,r)$
$\sigma^2(n-r)$: sampled variance of the set $(r+1,\ldots\ldots,n)$

Appendix A gives an analytically proof of the log likelihood ratio test base on the "ideal-stochastic" case. From the analysis of the sampled variances, the following heuristic tests are developed to improve the performance.

Hypothesis $H0 : \theta \sim N(m0, \sigma^2 0)$ $1,2,\ldots\ldots,n$

$H1 : \theta \sim N(m1, \sigma^2 1)$ $1,2,\ldots\ldots,r$

$\theta \sim N(m2, \sigma^2 2)$ $r+1,\ldots\ldots,n$

$(mi, \sigma^2 i)$ are sampled mean and variance of the referred set.

case 1 : $m1 > m2$, $\sigma 1 \neq \sigma 2$

$$\min \; P(r) = r(2m0 - m1 + \log \sigma 1) + (n-r)(m2 + \log \sigma 2)$$

case 2 : $m1 < m2$, $\sigma 1 \neq \sigma 2$

$$\min \; P(r) = r(m1 + \log \sigma 1) + (n-r)(2mo - m2 + \log \sigma 2)$$

case 3 : $m1 \neq m2$, $\sigma 1 > \sigma 2$

$$\min \; P(r) = r(2\sigma 0 - \sigma 1 + \log m1) + (n-r)(\sigma 2 + \log m2)$$

case 4 : $m1 \neq m2$, $\sigma 1 < \sigma 2$

$$\min \; P(r) = r(\sigma 1 + \log m1) + (n-r)(2\sigma 0 - \sigma 1 + \log m2)$$

Note that the direction of changes are known. The improvement of these modified tests over the log likelihood ratio test was illustrated in the previous report[6] by mean of the simulation.

4.3 - Implementation of the Method

As previously described, the detection process generates a binary ROI image. Since it is difficult to apply the test in two dimensional data, we can now use a "spoke"[1], as shown in figure 4 below, collocated with each ROI to obtain sequences of data in one dimension and then apply the hypothesis test above to detect the boundary.

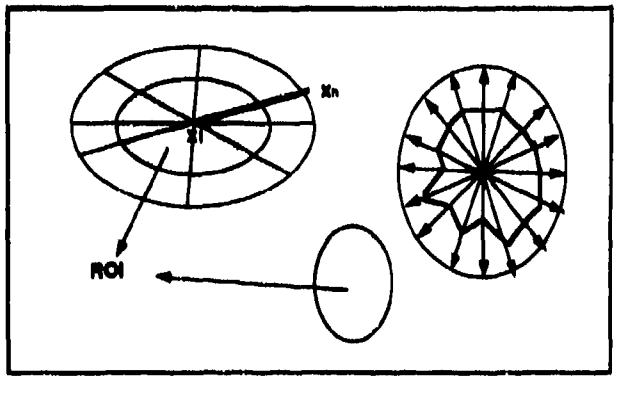


FIGURE 4

1) For each line of the spoke, obtain a sequence of data from the original image, which is correspondingly positioned along the line and where its length is twice the distance from the centroid to the boundary of its corresponding ROI. 2) Compute the value of P(r) along that line and detect the boundary point as a point at which P(r) is minimum. 3) Repeat step 1) and 2) to obtain N-detected boundary point. 4) Connect these N(usually 64)-points to form an object boundary. Repeat the process for each ROI. Note that the data must be interpolated, especially for small targets, to acquire a sufficient amount of data for the test.

4.4 - Iteration Method

The outer window size is initialized and then iteratively determined to feed back to the double window process. In the boundary detection method, the length of a spoke line is initialized from the ROI boundary, thus it can be iteratively determined to feed back to the boundary detection process. For both processes, these iterations update the target size and the surrounding background region.

5. - Multisensor Fusion Algorithm

The availability of multisensor data raises the interesting problem of how to fuse information obtained from different sensors. The main difficulty is making a decision when the single-sensor algorithms disagree with each other. Recall that the hypothesis modeled for background distribution in different sensors all have mean 0 and variance of 1, after performing a double window filter process on the data. Therefore, we can fuse data from multiple sensors after that process, but before the detection.

Futhermore, segmentations from multisensor (or different segmentors) can also be fused using the log-likelihood ratio test.

For example, suppose that A1 and A2 are segmented objects from two sensors (or segmentors)

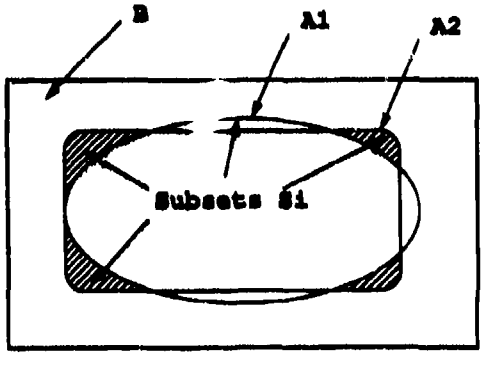


FIGURE 5

Let B be the box as shown in Figure 5 and assume that there exists an optimum segmented region that satisfies:

$$r_{opt} = (\#S_{opt})*\log \sigma S_{opt} + (\#(B-S_{opt}))*\log \sigma (B-S_{opt}) \text{ is minimum}$$

10

for all $S_i \subset B$, where #X denotes number of pixels inside region X and σX denotes standard deviation of pixels inside X.

Then we can apply the above formula to fuse the subsets of A1 union with A2 for obtaining a better segmentation. This is the implementation of testing the log-likelihood ratio test in two dimensions.

6. - Simulation Results and Discussion

Figure 6 illustrates the unprocessed original laser radar images and their corresponding boundary detections. The connected components are ROIs resulting from the detection and, as we shall see, the pixel detection is just about 60 to 90 percent of the target area but the boundary detection is as good as a hand segmentation. This is an advantage of boundary detection over thresholding for processing a noisy image or large deviation data. Figure 7 illustrates a) the original flir image, b) the segmentation of a), c) the zooming of data from b) with zooming factor of 3 (also use for d) and f)), d) the zooming of data from a) using bilinear interpolation, e) the segmentation of d), and f) the down zooming segmentation from e). The segmentations seem to work well as judging it by the human eye. Note that at present time, the measurement of the segmentation accuracy is still in controversy because of the difficulty of accurately located target pixels in real images. For the zoomed case, the segmentation shows to have an improvement over the unzoomed case, and thus for small targets, data must be interpolated before testing the hypothesis to detect the boundary.

However, most of the algorithms have a failing case and this algorithm fails if the hypothesis is false or if the target shape is not a simple N-side polygon. Therefore, a fusion algorithn for segmentors is a must for seeking an optimal segmentation.

7. - Conclusions

This paper has presented i) an iterative technique that uses a double window filter for target detection; ii) the implementation of a log-likelihood ratio test to refine the boundary of a detected object without requiring a threshold. For this latter case, the concept of partitioning an object into several subregions (just as we have done with a spoke) demonstrated that, an object ought to be thresholded at several levels, rather than by a single value. That is, the object is first estimated by segmenting with a threshold value, and then partitioned into subregions to find a threshold value for each subregion. This concept should be implemented by a technique that uses a normalized threshold value for segmentation to improve the performance. Even though this is true, the optimal performance can not be gained by a single method due to the presence of clutter components near the target. Therefore, a fusion algorithm for segmentation is needed to reach the goal, and the log-likelihood ratio test appear to be a good candidate of hypothesis test for this fusion algorithm.

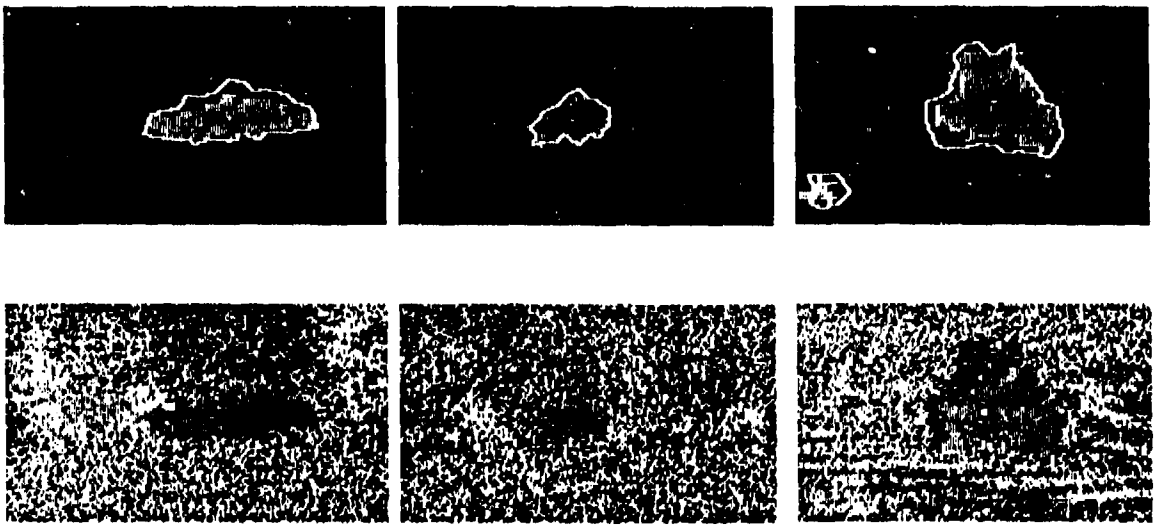


FIGURE #6

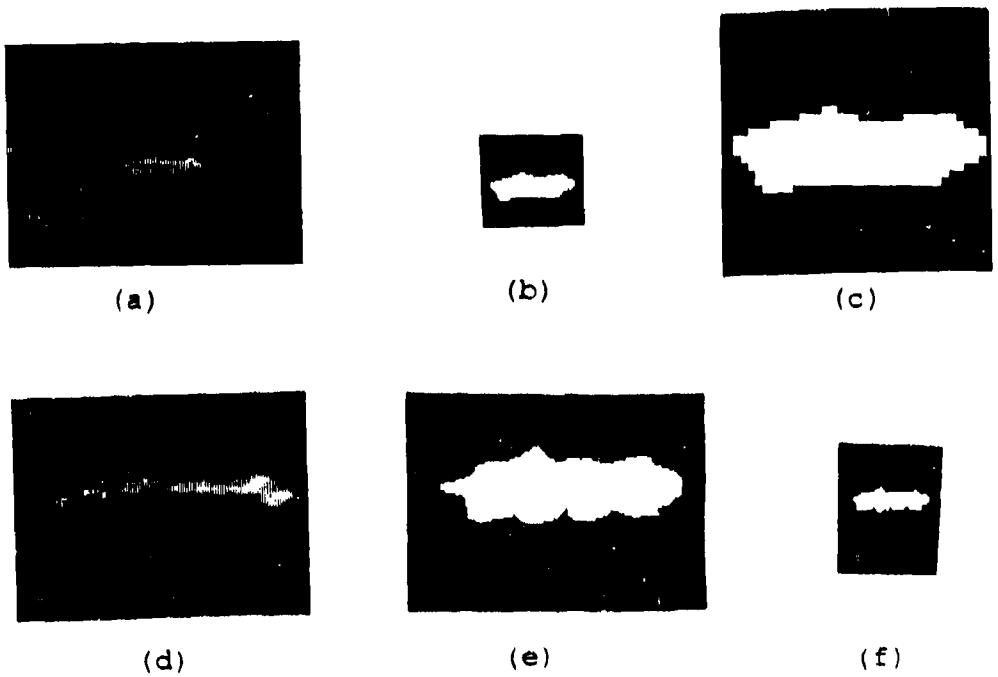


FIGURE #7

Acknowledgement : The author would like to thank Dr. Vincent Mirelli for contributing the idea of applying the log-likelhood ratio test into the image processing. Also thanks to Dr. Lynn Garn, Cyndi Thompson and members of the Image and Signal Processing Division for their support. A special thank to Brian Sadler of Harry Diamond Laboratories for his useful comments.

References :

[1] - Bruce Schachter, "Analysis and Evaluation of IR Target Detection, Segmentation, and Classification", Westinghouse Electric Corporation, Contract report for Army CNVEO, Revised July 20 1988.

[2] - Leo Breiman,"Statistics with a View Toward Applications", Houghton Mifflin Company - Boston, 1973.

[3] - Ahuja, Poor and Preparata, "Discrete Mathematics Applied to ATR Design", Coordinated Science Laboratory, Contract report for Army CNVEO, Jan 29th 1988.

[4] - Walpole & Myers,"Probability and Statistics for Engineers and Scientists", Macmillan Publishing Co.,Inc., 2nd Ed., 1978.

[5] - Basseville & Beveniste, "Detection of Abrupt Changes in Signals and Dynamical Systems", Lecture Notes in Control and Information Science, Spinger Verlag, Vol. 77, 1986.

[6] - Duc Nguyen, "A Statistical Change Detection Method for Image Segmentation in Laser Radar Imagery, CNVEO report, July 1988.

APPENDIX A

Proof of the log-likelihood ratio test for an "ideal-stochastic" case

Problem : let S be a set of random variables $\{s1, s2, \ldots\ldots, sn\}$

hypothesis $\theta = \theta 1$ $\{s1, s2, \ldots\ldots, sk\}$

$\theta = \theta 2$ $\{sk+1, \ldots\ldots, sn\}$ $\theta 1 \neq \theta 2$

prove that $y(r) = r\log \sigma^2(r) + (n-r)\log \sigma^2(n-r)$ is minimum at $r = k$ in these two cases,

case 1 : $\theta 1 = E\{s1, s2, \ldots\ldots, sk\}$

$\theta 2 = E\{sk+1, \ldots\ldots, sn\}$

$var\{s1, s2, \ldots, sk\} = var\{sk+1, \ldots, sn\} = \sigma_o^2$

case 2 : $\theta 1 = var\{s1, s2, \ldots\ldots, sk\}$

$\theta 2 = var\{sk+1, \ldots\ldots, sn\}$

$E\{s1, s2, \ldots, sk\} = E\{sk+1, \ldots, sn\} = E\{s1, s2, \ldots\ldots, sn\} = 0$

where $\sigma^2(r)$: sampled variance of the set $\{s1, s2, \ldots, sr\}$

$\sigma^2(n-r)$: sampled variance of the set $\{sr+1, \ldots\ldots, sn\}$

$0 < r < n$, $0 < k < n$, $n > 0$.

Notice that the meaning of the "ideal-stochastic" in this problem is,

we assume that $\sigma(r) = \sigma(k)$ for all r less than k

and $\sigma(n-r) = \sigma(n-k)$ for all r greater than k .

<u>case 1:</u> $\theta 1 = E(s1, s2, \ldots\ldots sk)$

$\theta 2 = E(sk+1, \ldots\ldots, sn)$

$var(s1, s2, \ldots, sk) = var(sk+1, \ldots, sn) = \sigma_0^2$

proof :

Since $var(si) = \sigma_0^2$ for $\forall\ i \in n$ and $E(si) \neq E(sj)$ for $i<k<j$

then, $\sigma^2(r) \geqslant \sigma_0^2$ and $\sigma^2(n-r) \geqslant \sigma_0^2$ for all $r \in n$.

at $r = k$, $\sigma^2(r) \dashrightarrow \sigma_0^2$ $\sigma^2(n-r) \dashrightarrow \sigma_0^2$

$\Longrightarrow$ $y(r) = r*\log \sigma^2(r) + (n-r)\log \sigma^2(n-r)$ is minimum at $r= k$

<u>case 2:</u> $\theta 1 = var(s1, s2, \ldots\ldots, sk) = \sigma_1^2$

$\theta 2 = var(sk+1, \ldots\ldots, sn) = \sigma_2^2$ $\sigma_1^2 \neq \sigma_2^2$

$E(s1, s2, \ldots, sk) = E(sk+1, \ldots\ldots, sn) = E(s1, s2, \ldots, sn) = 0$

$y(r) = r*\log \sigma^2(r) + (n-r)*\log \sigma^2(n-r)$

for this case, to prove $y(r)$ is minimum at $r = k$, we have to prove that $dy/dr = y'(r) < 0$ for $r \leqslant k$

and $y'(r) > 0$ for $r > k$

proof :

$$\sigma^2(r) = \sigma_1^2 \quad \text{for} \quad r \leqslant k$$
$$= [k*\sigma_1^2 + (r-k)*\sigma_2^2]/r = \alpha(r) \ , \quad r > k$$

$$\sigma^2(n-r) = [(k-r)*\sigma_1^2 + (n-k)*\sigma_2^2]/(n-r) = \beta(r) \quad r \leqslant k$$
$$= \sigma_2^2 \quad r > k$$

$$y(r) = r*\log \sigma_1^2 + (n-r)*\log(\beta(r)) \quad r \leqslant k$$
$$= r*\log(\alpha(r)) + (n-r)*\log \sigma_2^2 \quad r > k$$

COMPARISON DENSITY UNIFICATION OF STATISTICAL METHODS FOR CONTINUOUS AND DISCRETE DATA

by Emanuel Parzen
Department of Statistics, Texas A&M University

Abstract

The analysis of a univariate continuous sample is proposed as a probability model identification process consisting of four steps, each of which yields a distribution function, respectively denoted $F, F^{\sim}, F^{\wedge}, F^{\wedge\sim}$; each represents the outcome of successive steps in the analysis. Step 0: F, the true distribution function; Step 1: $F^{\sim}$, the fully non-parametric sample distribution; Step 2: $F^{\wedge} = F(.;\theta^{\wedge})$, a smooth distribution function obtained from a parametric model $F(.;\theta)$ whose parameter θ is efficiently estimated by $\theta^{\wedge}$; Step 3: Comparisons of observations $F^{\sim}$ and expectations $F^{\wedge}$ are provided by suitable sample comparison density functions (denoted $d^{\sim}$) which are fully non-parametric estimators of a comparison density (denoted d or $d(u)$, $0 \leq u \leq 1$) which compares F and $F^{\wedge}$; Step 4: $F^{\wedge\sim}$, a smooth-parametric estimator of F, is obtained from a smooth estimator $d^{\wedge}$ which we recommend as the optimal way to provide a goodness of fit test of $F^{\wedge}$ to $F^{\sim}$. Examples demonstrate the insights obtainable from this approach.

1. Introduction

We propose the concept of unification of statistical methods in order to develop a general philosophy of statistical data analysis. We propose that ways of thinking about statistical ends (goals) and means (procedures) are needed that provide a framework for implementing and comparing several different approaches to a data analysis problem. We believe that unification has benefits which include: existing (often parametric) methods will be better understood; many new (often nonparametric) methods will be developed. The new methods are usually computer intensive; consequently unification of statistical methods can be considered to be closely related to computational statistics. We define computational statistical methods as characterized by being graphics intensive and number crunching intensive.

This paper provides an introductory account of our approach to unification for the case of observations $Y_1, \ldots, Y_n$ which are a random sample of a continuous random variable Y with true unknown distribution function F, quantile function F^{-1}, and probability density function f. The estimation of these functions is called the Probability Model Identification Problem.

Our approach implements our favorite definition of statistics: Step 1: make and summarize observations; Step 2: form expectations; Step 3: compare observations and expectations (goodness of fit of model to data); Step 4: if model does not fit observations, revise model to fit (by estimation of the comparison density).

I believe that one should try to answer the philosophical question: what makes statistical inference possible (what makes it possible for us to be able to infer from data probability models that fit the data)? We propose that the answer includes the following fact: if G is a distribution function which one considers as a model for the true F, the transformed random variable $W = G(Y)$ has distribution function $F(G^{-1}(u))$ and quantile function $G(F^{-1}(u))$, both of which equal the identity function u when $G = F$ and F is continuous.

Probability Model Identification Fundamental Approach: Base estimation criteria on the idea that for a continuous random variable the closeness of a model G to the true F is judged by measures which test the hypothesis that $W = G(Y)$ is Uniform[0,1].

To test a hypothesis H_0 about a random variable W of which one has observed $W_1, \ldots, W_n$ (a random sample) early researchers (going back to the first hypothesis testers such as Laplace (1754)), proposed (1) computing a suitable test statistic T, (2) determining exactly or asymptotically the distribution under H_0 of the statistic T; (3) using this distribution to determine a rejection critical region R of values of the statistic T, assuming a specified probability of rejection α; (4) reporting rejection of the null hypothesis H_0 if the observed value of T belongs to R, or at least reporting a p-level of the observed value of T (defined to be the largest value of α, probability of rejection, whose rejection critical region contains the observed value of the test statistic T). This paper proposes to form critical regions that represent differences from 1 of the probability density or quantile density of transfomed random variables W on the unit interval.

2. Parametric Probability Model Identification

Step 1 of data analysis is to form fully nonparametric estimators: the sample distribution function $F^{\sim}(x)$ and the sample quantile function $Q^{\sim}(u) = F^{\sim -1}(u)$, where

$$F^{\sim}(x) = \text{ fraction of sample } \leq x, -\infty \leq x \leq \infty$$

The sample probability density $f^{\sim}$ does not exist

as an ordinary function but the notation is used as a symbolic function which has meaning as an integrand.

The sample expectation $E^n[g(Y)]$ is defined by

$$E^n[g(Y)] = (1/n)\sum_{j=1}^{n} g(Y_j)$$

which can also be expressed as an expectation with respect to F^n:

$$E^n[g(Y)] = \int_{-\infty}^{\infty} g(y)dF^n(y)$$

Step 2 forms fully parametric estimators $F^\wedge(x)$ and $Q^\wedge(u)$ of the form

$$F^\wedge(x) = F(x;\theta^\wedge),$$
$$Q^\wedge(u) = Q(u;\theta^\wedge)$$

where $\theta^\wedge$ is an optimal estimator of the (vector) parameter θ of a parametric model which we express as a hypothesis H_0 given by

$$H_0 : F(x) = F(x;\theta)$$

for some value of the vector parameter θ.

The maximum likelihood principle of estimation forms $\theta^\wedge$ as the value of θ maximising $L^n(\theta)$, the sample (average log) likelihood, defined by

$$L^n(\theta) = (2/n)\log f(Y_1,\ldots,Y_n;\theta)$$
$$= (2/n)\sum_{j=1}^{n}\log f(Y_j;\theta)$$

Expressions for $L^n(\theta)$ that we find useful for interpretation are

$$L^n(\theta) = 2E^n[\log(Y;\theta)]$$
$$= 2\int_{-\infty}^{\infty}\log f(y;\theta)dF^n(y)$$
$$= 2\int_0^1 \log f(Q^n(u);\theta)du$$

Sample likelihood is an estimator given the observations of the population likelihood

$$L(\theta) = 2\int_{-\infty}^{\infty}\log f(y;\theta)dF(y)$$
$$= 2\int_0^1 \log f(Q(u);\theta)du$$
$$= -H(F;F(.;\theta)),$$

defining cross entropy of two continuous distributions F and G

$$H(F;G) = (-2)\int_{-\infty}^{\infty}(\log g(y))f(y)dy.$$

We define entropy of F by

$$H(F) = (-2)\int_{-\infty}^{\infty}(\log f(y))f(y)dy.$$

We define Kullback information divergence by

$$I(F;G) = (-2)\int_{-\infty}^{\infty}(\log(g(y)/f(y))f(y)dy$$
$$= H(F;G) - H(F).$$

Important inequalities are $I(F;G) \geq 0, H(F;G) \geq H(F)$.

The procedure of maximum likelihood estimation can be expressed in terms of information theory concepts. The population likelihood is negative cross entropy. If the true $f(y)$ equals $f(y;\theta_0)$ then $-L(\theta)$ has its minimum value at $\theta = \theta_0$, and the minimum value is the entropy $H(F(.;\theta_0))$. Similarly the sample likelihood $L^n(\theta)$ has its maximum value at $\theta = \theta^\wedge$, and the maximum value satisfies:

$$L^n(\theta^\wedge) = -H(F^n, F_{\theta^\wedge}) \leq -H(F^n).$$

When the parametric family $f(.;\theta)$ is an exponential model,

$$L^n(\theta^\wedge) = \max_\theta L^n(\theta) = -H(F(.;\theta^\wedge))$$

the neg-entropy of the fitted distribution.

3. Comparison density functions

For F fixed, minimimising population cross-entropy $H(F;F(.;\theta))$ is equivalent to minimising population information divergence $I(F;F(.;\theta))$. One can regard this as a mathematical measure of how closely one can approximate F by a member of the parametric family $F(.;\theta)$. But we prefer a more statistical interpretation in terms of the comparison density function of two continuous distributions F and G, defined as follows:

$$d(u;F,G) = gF^{-1}(u)/fF^{-1}(u)$$

One can show that

$$d(u;F,G) = D'(u;F,G),$$

defining

$$D(u:F,G) = G(F^{-1}(u)).$$

One can show that

$$I(F;G) = (-2)\int_0^1 -\log d(u:F,G)du.$$

We interpret $I(F,G)$ as a measure of the closeness to 1 of the density $d(u:F,G)$.

The information and comparison density interpretation of maximum likelihood estimation that we are proposing can be expressed as follows: population likelihood is related to $I(F;F(.;\theta))$ which measures the closeness of $F(.;\theta)$ to F by measuring the closeness to 1 of $d(u;F;F(.;\theta))$, the derivative of $D(u;F,F(.;\theta)) = F(F^{-1}(u);\theta)$ which is the quantile function of the random variable $W_\theta = F(Y;\theta)$; maximum likelihood estimation of θ is equivalent to finding the value of θ of the transformation from Y to W_θ which is closest to a Uniform[0,1] distribution as measured by the distance of the sample quantile function $F(F^{\wedge -1}(u);\theta)$ from $D(u) = u$.

The estimation process in Step 2 uses quantile functions. Step 3 uses distribution functions for goodness of fit tests; one measures how close the fitted model $F(.;\theta^\wedge) = F^\wedge(.)$ is to $F^\sim$ by how close to uniform is the sample distribution function $D^\sim(u) = F^\sim(F^{\wedge -1}(u))$, also denoted $D^\sim = F^\sim(Q^\wedge)$

To understand why the sample distribution function is more convenient note that it is an estimator of $D(u) = F(F^{\wedge -1}(u)$ with derivative

$$d(u) = f(F^{\wedge -1}(u))/f^\wedge(F^{\wedge -1}(u)).$$

An estimator $d^\wedge(u)$ leads to a revised estimator $f^{\wedge\wedge}(y)$ by

$$f^{\wedge\wedge}(y) = d^\wedge(F^\wedge(y))f^\wedge(y).$$

Step 3 of our approach to statistical data analysis studies various diagnostics of $D^\sim(u)$ which measure the significance of its difference from u, and forms sequences of smooth approximations $d^\wedge(u)$ of the symbolic derivative $d^\sim(u)$. Step 4 chooses an optimal $d^\wedge(u)$, which could be identically 1; if this is not the case, one obtains a revised probability density estimator $f^{\wedge\wedge}(y)$.

4. Diagnostics of comparison distributions

The sample distribution function $D^\sim = F^\sim(Q^\wedge)$ of $W = F^\wedge(X)$ is called a sample comparison distribution, estimating $D = F(Q^\wedge)$ which measures how well the true F is approximated by the fitted model $F^\wedge$.

Classical goodness of fit statistics are portmanteau statistics, such as the Cramer-von Mises and Anderson-Darling tests in the continuous case and the original Karl Pearson Chi-Square test in the discrete case. We do not recommend them because it is difficult to determine their (asymptotic) distributions when parameters are estimated, and they do not satisfy the vital criteria of (1) looking at the data and (2) providing insight into how to revise the model when it does not fit.

The "components" approach chooses score functions $J_1(u),\ldots,J_m(u)$ satisfying for $j = 1,\ldots,m$

$$\int_0^1 J_j(u)du = 0, \int_0^1 J_j^2(u)du = 1,$$
$$\int_0^1 J_j(u)J_k(u)du = 0 \text{ for } j \neq k.$$

Components are statistics $T^\sim$ ("linear detectors") which are linear functionals in $d^\sim(u)$ of the form

$$T^\sim(J_j) = [J_j, d^\sim] = \int_0^1 J_j(u)d^\sim(u)du = \int_0^1 J_j(u)dD^\sim(u)$$
$$= (1/n)\sum_{i=1}^n J_j(W_i) = E^\sim[J_j(W)]$$

Components often can be shown to be asymptotically normal (under the null hypothesis) independent $N(0,1/n)$ random variables. Component tests judge significance of $n^{.5}T^\sim(J_j)$ and $nS_{k,m}$, defining chi-squared test statistics

$$S_{k,m} = \sum_{j=k}^m |T^\sim(J_j)|^2.$$

The Cramer-von Mises goodness of fit test is a "quadratic detector" in the sense that it can be expressed as a weighted sum of squares of components:

$$\int_0^1 (D^\sim(u) - u)^2 du = \sum_{j=1}^\infty w_j^2[\phi_j, d^\sim]^2$$

where

$$\phi_j(u) = 2^{.5}\cos(j\pi u), w_j = 1/j\pi.$$

The Anderson-Darling goodness of fit test is a quadratic detector:

$$\int_0^1 (D^\sim(u) - u)^2/u(1-u)du = \sum_{j=1}^\infty w_j^2[\phi_j, d^\sim]^2$$

where

$$\varphi_j(u) = (2j+1)^{.5}p_j(2u-1), w_j = 1/j(j+1)^{.5}, p_j(t)$$

are Legendre polynomials on [-1,1].

Hermite polynomial goodness of fit test can be defined by a quadratic detector with $w_j = 1/j$,

$$\phi_j(u) = (j!)^{-.5} H_j(\Phi^{-1}(u)), H_j(x)$$

are Hermite polynomials.

A quadratic detector can be interpreted as

$$\int_0^1 (d^\wedge(u) - u)^2 du$$

a distance from 1 of a smooth density estimator

$$d^\wedge(u) = 1 + \sum_{j=1}^{\infty} w_j [\phi_j, d^\sim] \phi_j(u)$$

5. Comparison density estimation

The novel elements of this paper are the role of the comparison density function, and especially the proposition that estimation of the comparison density function can be used to motivate and interpret components. Interpret sample components $T^\sim(J_j)$ as estimators of

$$T(J_j) = \int_0^1 J_j(u) d(u) du = \int_{-\infty}^{\infty} J_j(F^\wedge(x)) dF(x)$$

Consider all $d(u)$ obeying the constraints that for $j = 1, \ldots, m$

$$T(J_j) = T^\sim(J_j);$$

one determines $d_{\lambda,m}$, defined as the density obeying the constraints which minimizes

$$IR_1(d) = \log \int_0^1 d^2(u) du, \text{ for } \lambda = 1,$$

$$IR_0(d) = 2 \int_0^1 \{\log d(u)\} d(u) du, \text{ for } \lambda = 0,$$

$$IR_{-1}(d) = 2 \int_0^1 \{-\log d(u)\} du \text{ for } \lambda = -1,$$

$$IR_\lambda(d) = \{2/\lambda(1+\lambda)\} \log \int_0^1 \{d(u)\}^{1+\lambda} du \text{ for } \lambda.$$

We call $IR_\lambda(d)$ Renyi information of index λ.

An explicit formula for $d_{1,m}$ is the truncated Fourier series:

$$d_{1,m} = \sum_{j=1}^{m} [J_j, d^\sim] J_j(u).$$

An exponential model with parameters $\theta_1, \ldots, \theta_m$ is the form of $d_{0,m}$:

$$\log d_{0,m}(u) = \sum_{k=1}^{m} \theta_k J_k(u) - \Psi(\theta_1, \ldots, \theta_m)$$

where Ψ is the integrating factor that guarantees that $d_{0,m}$ is a density.

6. Order determination

The problem of goodness of fit can be regarded as a problem of determining the optimal order $m^\wedge$ and whether the optimal order $m^\wedge = 0$. As m increases, $d_{0,m}(u)$ converges to $d^\sim(u) = d(u; F^\wedge, F^\sim)$ and $IR_0(d_{0,m})$ increases to

$$IR_0(d^\sim) = IR_{-1}(d(u; F^\sim, F^\wedge))$$

which is the Moran goodness of fit, or non-parametric entropy, test.

The combination of $F^\wedge(.) = F(.; \theta^\wedge)$ and $d_{0,m^\wedge}$ is regarded as an estimator $F^{\wedge\wedge}$. A random sample from $F^{\wedge\wedge}$ can be generated from a random sample from $F^\wedge$ using $d_{0,m^\wedge}$ and the *rejection method of simulation.*

7. Examples of One Sample Continuous Data Analysis

National Bureau of Standards NB10 Measurements: Freedman, Pisani, Purves in their textbook on Statistics (p.94) report 100 measurements of the 10 gram check-weight NB10 made at the National Bureau of Standards. They report: "The normal curve does not fit at all well. The normal curve does fit the data with three outliers removed. The normal curve fitted to these measurements has an average of 404 micrograms below 10 grams, and a standard deviation of about 4 micrograms. But in a small percentage of cases, the measurements are quite a bit farther away from the average than the normal curve suggests. The overall standard deviation of 6 micrograms is a compromise between the standard deviation of the main part of the histogram (4 micrograms) and the three outliers, representing deviations of 18, -30, and 32 micrograms. In careful measurement work, a small percentage of outliers is expected. The only unusual aspect of the NB10 data is that the National Bureau of Standards reported its outliers; many investigators don't. Realistic performance parameters require the acceptance of all data that cannot be rejected for cause."

The NB10 data illustrates the statistical analysis strategy that we propose be routinely applied to data. Step 1. Specify a parametric probability model for the data (here the model is normal). Step 2. Estimate parameters of the model (here mean and standard deviation) to be 10 grams-404 micrograms and 6 micrograms respectively. Step 2*. Robust parameter estimation by Renyi information of index between 0 and 1 obtains as estimators of a normal model (fitted to the part of the data that can be well fitted by a normal model) the same mean and a standard deviation of 4 micrograms. Step

4: Goodness of fit test of normality by traditional tests. Step 5: Maximum entropy estimator of comparison density $d(u;$ normal model, data) clearly indicates the nature of the data; a poor fit of normal model to data. Shape of $d^{\wedge}(u)$ in interior of interval (0,1) can be interpreted as expected curve if $d^{\wedge}(u)$ estimates

$$d(u; N(0,(6)^2), N(0,(4)^2))$$
$$= \kappa \exp\left\{-.5\left(\kappa^2-1\right)\left(\Phi^{-1}(u)\right)^2\right\}, \kappa = 6/4.$$

Peaks of $d^{\wedge}(u)$ at $u = 0, 1$ indicate longer tails than normal. In general, one must decide whether to consider these tails in $d^{\wedge}(u)$ as outliers or as evidence that a longer tailed distribution than the normal should be used to model the data. In Figure 1 two graphs illustrate the comparison density estimation process: the raw estimator $d^{\sim}(u)$ superimposed on a smooth estimator $d^{\wedge}(u)$; the exponential model smooth estimator $d_{0,4}{}^{\wedge}$, the orthogonal polynomial estimator $d_{1,4}{}^{\wedge}$, and a naive step function estimator d^{*} representing increments of $D^{\sim}(u)$ on 8 equal subintervals. Diagnostic tools at step 1 which help identify probability models for the data are illustrated by a IQQ plot of the sample quantile function of the data versus the quantile function of a normal with density $f(x) = \exp(-\pi x^2)$. The informative quantile function of the sample is defined $QI^{\sim}(u) = \{Q^{\sim}(u) - Q^{\sim}(.5)\}/2\{Q^{\sim}(.75) - Q^{\sim}(.25)\}$.

Breaking Stress of Beam: Cheng and Stephens (1989) give a data set of breaking stress of 41 beam specimens cut from a single carbon block of graphite H590, and discuss goodness of fit tests of the hypothesis that the data is normal. Let $F(\cdot;\theta^{\wedge})$ denote the normal distribution with maximum likelihood estimated value of θ. They show that Moran's statistic, which is equivalent to $IR_0(d(u; F(\cdot;\theta^{\wedge}), F^{\sim})$ "correctly" rejects the hypothesis that the sample is normal, in contrast to more traditional empirical distribution based statistics (such as Kolmogorov-Smirnov and Cramer-von Mises) which accept the hypothesis of normality for the sample tested. The comparison density estimation approach indicates the nature of the data; an excellent fit of normal model in interior of interval (0,1) but peaks at $u = 0, 1$ indicate outliers or long tails (clearly evident in stem and leaf table of the data). One conjectures that a symmetric extreme value distribution would be a more appropriate model. Figure 2 illustrates the comparison density estimation process for a normal model $F(\cdot;\theta^{\wedge})$. The graph of $D(u; F(\cdot;\theta^{\wedge}), F^{\sim})$ is graphically well fitted by a uniform distribution, and therefore passes traditional goodness of fit tests. The raw estimator $d(u; F(\cdot;\theta^{\wedge}), F^{\sim})$ is superimposed on a smooth estimator. The exponential model smooth estimator $d^{\wedge}(u)$ is superimposed on a step function estimator computed from increments of $D(u; F(\cdot,\theta^{\wedge}), F^{\sim})$ over 8 sub-intervals.

Cheng and Stephens Break Stress Data
(Stem and Leaf)

27	.55							
28								
29	.89							
30	.07	.65						
31	.23	.53	.53	.82				
32	.23	.28	.69	.98				
33	.28	.28	.74	.74	.86	.86	.86	
34	.15	.15	.15	.44	.62	.74	.74	
35	.03	.03	.32	.44	.61	.61	.73	.90
36	.20	.78						
37	.07	.36	.36	.36				
38								
39								
40	.28							

8. One Sample Discrete Data Analysis

Step 1: Identify a parametric family of probability mass functions $p(x;\theta)$ to model the sample probability mass function $p^{\sim}(x)$.

Step 2: Parameter estimation. Maximum likelihood estimator $\theta^{\wedge}$ can be obtained by minimising

$$IR_{-1}(d(u; F^{\sim}, F(\cdot;\theta)))$$
$$= (-2)\sum_{x} \log\{p(x;\theta)/p^{\sim}(x)\}p^{\sim}(x)$$

A parametric estimator of p is $p^{\wedge}(x) = p(x;\theta^{\wedge})$. Minimum chi-square estimation uses the modified chi-squared distance

$$IR_1(d(u; F^{\sim}, F(\cdot;\theta))) = \sum_{x}\{(p(x;\theta)/p^{\sim}(x)) - 1\}^2 p^{\sim}(x)$$

Step 3: Parametric hypthesis testing. To test a hypothesis H_0 about the parameter θ, let $\theta_{H_0}{}^{\wedge}$ denote the minimum-modified chi square estimator of θ under H_0; equivalent to likelihood ratio tests is the test statistic

$$IR_1(d(u; F^{\sim}, F(\cdot;\theta_{H_0}{}^{\wedge}))) - IR_1(d(u; F^{\sim}, F(\cdot;\theta^{\wedge})))$$

Step 4: Goodness of fit test of $H_0 : p = p(\cdot;\theta^{\wedge})$ or equivalently $H_0 : d(u; F(\cdot;\theta^{\wedge}), F) = 0$. Test the significance of the difference from zero of

$$IR_1(d(u; F(\cdot;\theta^{\wedge}), F^{\sim})) = IR_{-2}(d(u; F^{\sim}, F(\cdot;\theta^{\wedge}))).$$

Step 5: Maximum entropy goodness of fit tests and estimators $d_{0,m}{}^{\wedge}(u)$ of $d(u; F(\cdot;\theta^{+}), F)$ are obtained by

minimising $IR_0(d^*)$ among densities $d^*(u)$ satisfying, for $k = 1, \ldots, m$ and specified score functions $J_k(\cdot)$,

$$[J_k, d^*] = [J_k, d^*]$$

defining $d^*(u) = d(u; F(\cdot; \theta^*), F^*)$. For m large enough $d_{0,m}^*(u)$ equals $d^*(u)$ and $IR_0(d_{0,m}^*)$ increases to a test statistic (alternative to that of Step 4) $IR_0(d(u; F(\cdot; \theta^*), F^*))$.

Step 6: Rejection simulation nonparametric estimation of F. Use an order determining criterion to determine an order m^* with the properties: if $m^* = 0$, accept H_0; if one rejects H_0 use $d_{0,m^*}^*(u)$ as the density to be used in the rejection method of simulating a random sample from F. The combination of $F(\cdot; \theta^*)$ and $d_{0,m^*}^*(u)$ is regarded as an estimator F^{**}.

REFERENCES

Alexander, William (1989) "Boundary kernel estimation of the two-sample comparison density function" Texas A&M Department of Statistics Ph.D. thesis.

Aly, E. A. A., M. Csorgo, and L. Horvath (1987) "P-P plots, rank processes, and Chernoff-Savage theorems" in *New Perspectives in Theoretical and Applied Statistics* (ed. M. L. Puri, J. P. Vilaplann, W. Wertz) New York: Wiley 135–156.

Boos, Dennis D. (1986) "Comparing k populations with linear rank statistics", *Journal of the American Statistican Association*, 81, 1018–1025.

Cheng, R. C. H. and Stephens, M. A. (1989) "A goodness of fit test using Moran's statistic with estimated parameters", *Biometrika*, 76, 385–392.

Chui, C. K., Deutsch, F., Ward, J. D. (1990) "Constrained best approximation in Hilbert space," *Constructive Approximation*, 6, 35–64.

Eubank, R. L., V. N. LaRiccia, R. B. Rosenstein (1987) "Test statistics derived as components of Pearson's Phi-squared distance measure", *Journal of the American Statistical Association*, 82, 816–825.

Freedman, D., Pisani, R., Purves, R. (1978) *Statistics*, New York: Norton.

Parzen, E. (1979) "Nonparametric statistical data modelling", *Journal of the American Statistical Association*, 74, 105–131.

Parzen, E. (1989) "Multi-sample functional statistical data analysis," in *Statistical Data Analysis and Inference*, (ed. Y. Dodge). Amsterdam: North Holland, pp. 71–84.

Rayner, J. C. W. and Best, D. J. (1989). *Smooth Tests of Goodness of Fit*, Oxford University Press, New York.

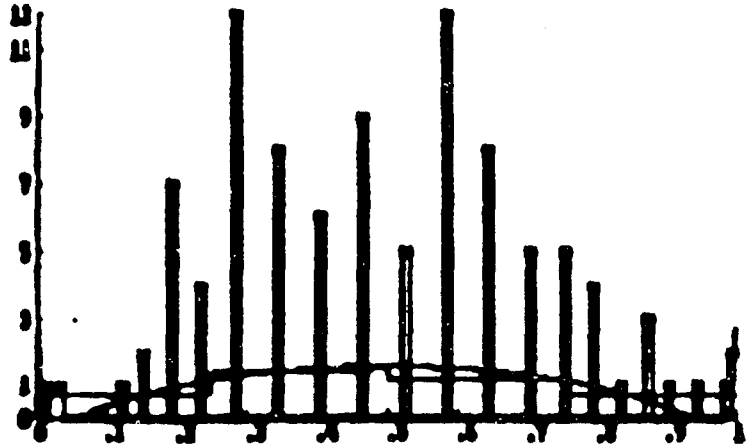
Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness of Fit Statistics for Discrete Multivariate Data*, Springer Verlag, New York.

Renyi, A. (1961). "On measures of entropy and information." *Proc. 4th Berkeley Symp. Math. Statist. Probability, 1960*, 1, 547-461. University of California Press: Berkeley.

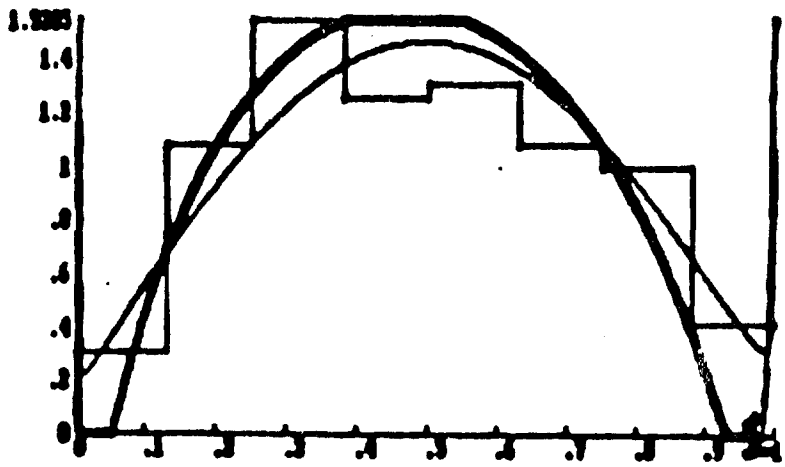
Shorack, Galen and John Wellner (1986) *Empirical Processes With Applications to Statistics*, New York: Wiley.

Research supported by the U. S. Army Research Office.

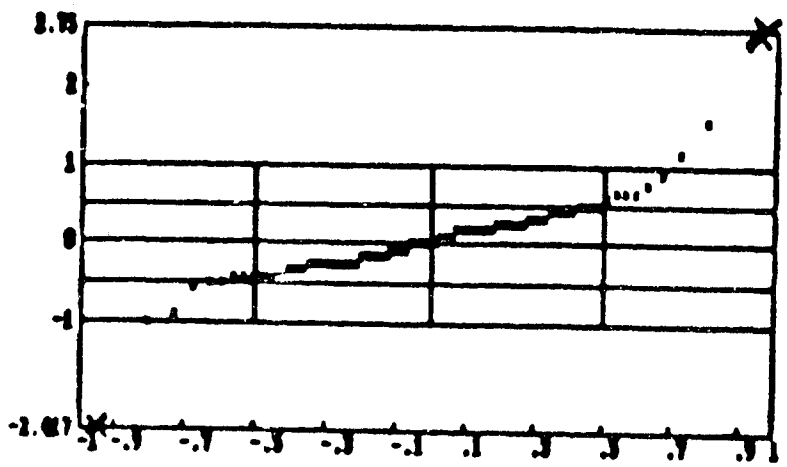
Figure 1
Test NB10 Measurements for Normality



Raw $d^{\sim}(u)$ to test normality, smooth by $d^{\wedge}(u)$

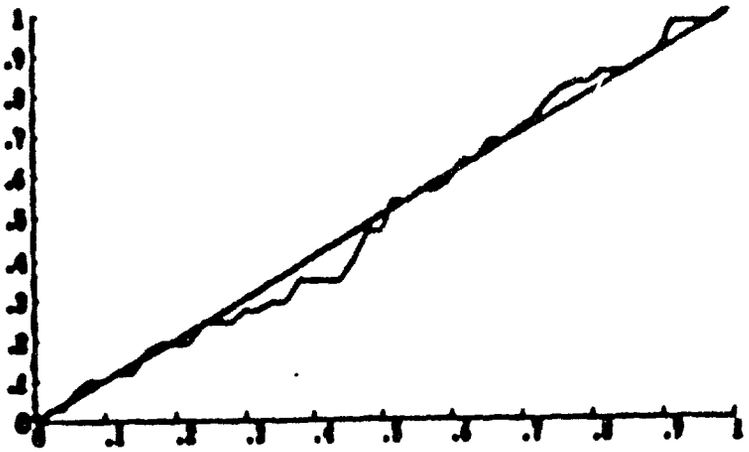


Estimators $d^{\sim}(u;$ normal, data) Orthogonal Polynomial;
Exponential Model (graph closest to graph of step function):

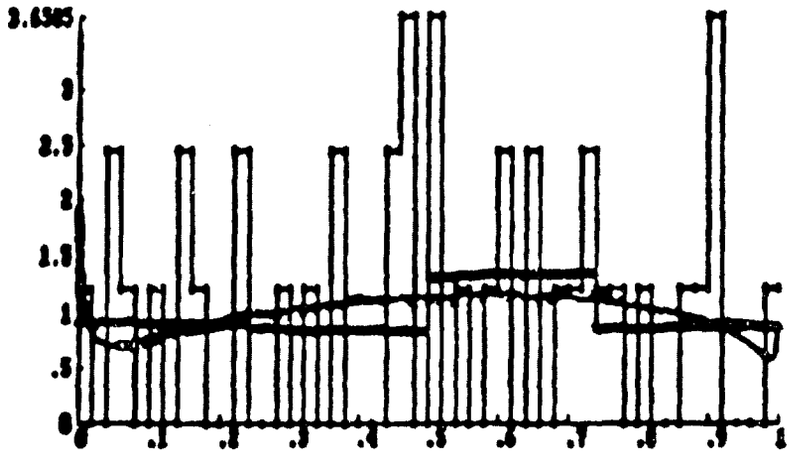


IQQ Plot $(Q_0 I, Q^{\sim} I)$ Q_0 Normal

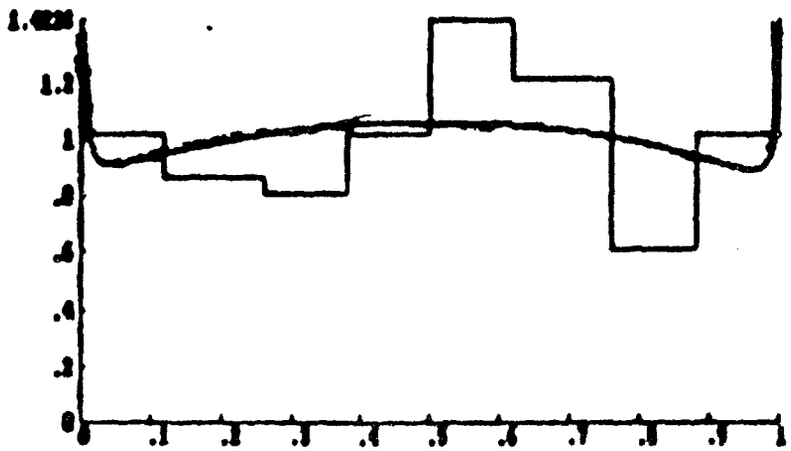
Figure 2
Test Breaking Stress Measurements for Normality



$D(u; F(\cdot;\theta^{\wedge}), F^{\sim})$ to test normality



Raw $d^{\sim}(u)$ to test normality, smooth by orthogonal polynomial $d^{\wedge}(u)$



Estimation $d^{\wedge}(u;$ normal, data) Exponential Model,
Step function

DEVELOPMENT OF DESERT CAMOUFLAGE NETS FOR SAUDI ARABIAN NATIONAL GUARD (SANG)

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 Materiel Command
Alexandria, Virginia 22333-0001

ABSTRACT

The objective of this program was to determine the best camouflage net colors and texture for Saudi Arabian desert backgrounds. The program began with a fact-finding visit to Saudi Arabia. Two of the findings of this visit were that the standard U.S. Army desert net was too dark, and that a light-colored monotone net would best fit in with the backgrounds. Using this information, and spectrophotometric readings of soil samples, a series of test nets were constructed for a field test in Saudi Arabia. Twenty-four nets of different colors were constructed for use in the full-scale test. The standard U.S. Army desert camouflage net was used for comparison purposes. Five test sites, representative of terrain colors generally found in the SANG operational areas around Riyadh and Hofuf, were selected. The test procedure involved a selection and ranking process as to their ability to blend with the background. The nets were narrowed down to six final candidates, and then these six were ranked in their order of preference. This data was statistically analyzed, and the final color/texture recommendations made. This project joined the expertise of an engineer, statistician, and psychologist into a working operational research team to develop a new camouflage net for SANG.

1.0 SECTION I -- INTRODUCTION

The Belvoir Research, Development and Engineering Center (BRDEC) was requested by the U.S. Army Project Manager, Saudi Arabian National Guard (PM, SANG) to provide assistance in developing a camouflage program. A fact-finding team was dispatched to Saudi Arabia to determine the specific requirements for the program. The primary areas of interest were the vicinities around the capital, Riyadh, and Daman-Hofuf, the center of the oil-producing area along the Persian Gulf. It was found during the visit that the U.S. standard desert camouflage net was too dark for the light desert backgrounds of these areas. As a result, a program to develop a new camouflage net for SANG was begun.

Spectrophotometric readings of soil samples taken of the areas of interest during the fact-finding visit, along with color slides of the areas, were analyzed and studied to determine a spectrum of desert colors for use in the construction of a series of camouflage nets for a field evaluation in Saudi Arabia. The field evaluation narrowed the spectrum of colors. The remaining colors were further refined, and other colors were added for a second field test. Eleven desert tones were developed for the second field test. All the nets constructed for the second field test were monotones since it was determined during the first field test that monotones blended better with the Saudi Arabian light-colored deserts.

In addition to color, the variable of texture was also investigated. Texture is defined as the degree of perceived roughness of the camouflage net. Texture is caused through the incision cuts in the material

making up the net. The shadows caused by the incisions gray out the color of the net. The larger the incision cuts, the more the net color is affected. The incision size used on the standard U.S. Army net was too large and tended to gray out the light garnish colors which would blend well into the desert background. Two smaller incisions were developed for testing the eleven selected garnish colors. These incisions were identified in this report as "small" and "smaller". This report describes the second field evaluation of the camouflage nets that took place in Saudi Arabia, with the objective of determining the most effective camouflage net colors and texture (incision pattern) for use in the Riyadh and Hofuf regions of Saudi Arabia.

2.0 SECTION II -- EXPERIMENTAL DESIGN

2.1 Test Nets

The same color was used on pairs of nets, i.e., one for the small incision size net and one for the still smaller size incision net. In addition to these nets, a two-color net (i.e., one color on each side) was developed for each incision size for evaluation in both light and darker desert areas. Two standard U.S. Army desert nets were also included for control and comparison purposes, making a total of twenty-six hexagon nets. The small incision nets were identified by Roman numerals I through XIII. The smaller incision nets were identified by the lower case alphabetical letters a through m. Note that test nets V and e had a different color on each side, and that test nets VIII and h were the standard U.S. camouflage nets with the standard incision size.

2.2 Test Sites

A total of five sites were selected, four at Riyadh and one at Hofuf. The site locations and color designations were as follows:

- Site 1 – Tan; Riyadh
- Site 2 – Light Tan; Riyadh
- Site 3 – Brownish Tan; Riyadh
- Site 4 – Grayish Brown; Riyadh
- Site 5 – Reddish Tan; Hofuf

The exact site locations were selected so that the nets were always observed against the terrain, and not highlighted against the sky.

2.3 Experimental Design

The nets were set up at each site, separated by incision size reading "I" through "XIII" for the small incision group and "a" through "m" for the smaller incision group. A red flag was placed on the first, sixth, and tenth nets, thus subdividing each of the two incision size groups into three smaller groups of five, four, and four.

The test procedure involved a selection process during which the thirteen nets from each incision size grouping were narrowed down to three final candidates within each incision size, and the final three from each incision size were grouped together and ranked in the order of preference. The subjects first selected three nets from each incision size subgroup (i.e., five, four, and four) which best blended with the background. The results were tabulated in the field, and the best three nets were kept, with the others being taken down. This left three subgroups of three nets each. The next iteration involved selecting the best two nets from each new subgroup. Again the results were tabulated in the field, and

the worse net for each subgroup was taken down. This left a total of six nets standing for each incision size group. These six nets were then rated as to color blend and texture, with one being the best rating and six the worst. The results were tabulated and the three best nets from each incision size setup were identified and moved to make one final group of six nets. The ranking in order of preference as to color and texture was performed as before, to determine which color and incision pattern best blended with the background. Note that this one-through-six rating was the first time the small and smaller incision size nets were evaluated as one group. For all the ratings, described above, no ties were allowed.

A briefing was read to the observers, at the start of each test, describing the sequential setups of the nets, and pertinent instructions for each particular phase of the test. All evaluations were performed between the hours of 1200 and 1400 for proper sun angle and minimum shadows. Each day was clear and hot with temperature ranges between 118-130 degrees Fahrenheit. This procedure was repeated for each of the five test sites.

2.4 Subjects

A total of sixteen male subjects (seventeen for site 1) were used as observers for each of the five sites. The subjects were screened for visual acuity using a reading card produced by the American Optical Corporation. Pseudo-Isochromatic Plates were used to screen each potential observer for color vision. In order to participate in this study, both screening tests had to be passed. All subjects were employees of the Vinnell Corporation, stationed in Riyadh.

3.0 SECTION III -- RESULTS

Of the twenty-six color/texture combination desert camouflage nets tested, thirteen advanced to the final selection process in at least one of the five sites. Table 1 summarizes the placement of these nets for each site. Net "c" was added only to aid in computer data analysis, by completing the pairing off of the color/texture combinations. It was not used in the final grouping for any of the five sites.

Table 1
Final Placement of the Desert Camouflage Nets for Each of the Five Test Sites

Net	Sites					Total
	1	2	3	4	5	
I	No	No	No	Yes	No	1
III	No	Yes	No	No	Yes	2
V	Yes	No	Yes	Yes	No	3
IX	Yes	Yes	Yes	No	Yes	4
XI	No	No	No	Yes	No	1
XII	Yes	No	Yes	No	No	2
XIII	No	Yes	No	No	Yes	2
a	No	No	No	Yes	Yes	2
c	No	No	No	No	No	0
e	Yes	No	Yes	Yes	No	3
i	Yes	Yes	Yes	No	Yes	4
k	No	No	Yes	Yes	No	2
l	Yes	Yes	No	No	No	2
m	No	Yes	No	No	Yes	2

3.1 Net Color

The same colors, with the exception of nets V and e, appear for each of the two textures. Nets V and e were reversed to show their darker side when used on sites 3 and 4 in order to evaluated the side which matched the background closer. Normally, the color perceived of a one-color net is a result of a visual integration of the garnish color and the incision shadows producing a third color. In the case of reversible nets, with different colors on each side, the color perceived is actually a fourth color. This is the result of the visual integration of some of the color visible from the reverse side of the net. Therefore color IX on one side of net V or e is perceived slightly altered by the addition of color XI on the reverse side caused by the convoluting garnish. It is this "fourth" color that was evaluated during the net tests.

Using the numerical values of one being the best, six the least preferred, and seven as not being in the final set of nets for color/texture evaluation at each of the five sites, the mean value of acceptance with the associated 95% confidence interval was determined. This descriptive data is shown in Table 2 and pictured graphically in Figure 1. The higher the value of preference, the less preferred is the net color/texture.

Table 2

Mean Preference and 95-Percent Confidence Intervals for the Final Desert Net Colors, Small and Smaller Incisions, Averaged Across All Sites

Net Color	Sample Size	Mean	Standard Deviation	Standard Error	95% Confidence Interval Lower Limit		Upper Limit
I/a	162	6.43	1.2355	.0970	6.24	to	6.62
III/c	162	6.38	1.3606	.1069	6.17	to	6.59
V/e	162	4.92	2.0851	.1638	4.60	to	5.24
IX/i	162	3.60	2.2881	.1797	3.24	to	3.96
XI/k	162	5.91	1.8776	.1475	5.62	to	6.20
XII/l	162	5.52	2.0828	.1636	5.20	to	5.84
XIII/m	162	5.70	1.8556	.1458	5.41	to	5.99

Colors IX/i and V/e are the most preferred colors with mean preferences of 3.60 and 4.92 respectively. The associated confidence interval states that there is 95% confidence that the true mean preferences rest between the upper and lower limits shown in the table. An analysis of variance[1/] was performed upon the data in Table 2 to determine if there were significant differences between colors when compared with sites (Table 3).

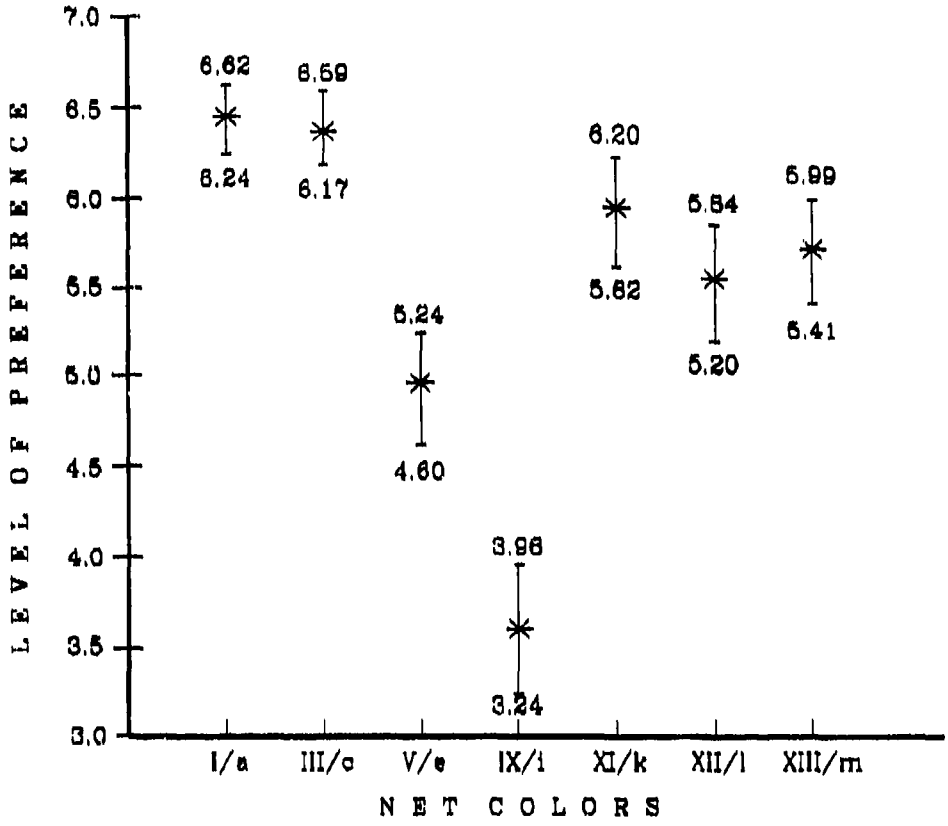


Figure 1. Graphic Display of Table 2

Table 3
Analysis of Variance for Net Colors as to Blend With Sites

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-Ratio	Significance Level
Colors	6	938.9630	156.4938	45.4500	0.000*
Error	1127	3880.4938	3.4432		
Total	1133	4819.4568			

Bartlett's Test for Homogeneous Variances

Number Degrees of Freedom = 6

F = 15.428 Significance Level α = 0.000*

*Significant at α less than 0.001 level

Table 3 indicates that there were significant differences in the ability of the desert net colors to blend with the background. The Bartlett's Test indicated that the variances of each color were not homogeneous, i.e., significantly different, so they are not necessarily from the same population.

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

Table 4
Duncan's Multiple-Range Test – Desert Net Colors

Best							Worst
Subset 1		Subset 2		Subset 3		Subset 4	
Color	Mean	Color	Mean	Color	Mean	Color	Mean
IX/i	3.60	V/e	4.92	XII/l	5.52	III/c	6.38
				XIII/m	5.70	I/a	6.43
				XI/k	5.91		

Table 4 showed that the best color was IX/i for blending with the desert backgrounds.

Table 5 contains the nonparametric[2/] results of the Kruskal-Wallis One-Way ANOVA as a double check upon the Duncan's Test. The Bartlett's Test indicated that the variances for each level of color were not homogeneous. Parametric tests, such as Duncan's, assume homogeneity of variance.

Table 5
Kruskal-Wallis One-Way ANOVA – Desert Net Colors

Color	Sample Size	Mean Rank
I/a	162	708.72
III/c	162	704.22
V/e	162	471.28
IX/i	162	308.15
XI/k	162	638.24
XII/l	162	575.41
XIII/m	162	566.49
	Total 1134	

Sample Size	Chi-Squared	Significance Level
1134	181.5825	0.000*

	Corrected for Ties	
Sample Size	Chi-Squared	Significance Level
1134	223.8419	0.000*

*Significant at α less than 0.001

These results were in agreement with Table 3 above.

3.2 Incision Size

Table 6 showed the descriptive data for the small and smaller incisions, which is the second variable of interest.

Table 6
Mean Preference and 95-Percent Confidence Intervals
for the Small and Smaller Net Incisions, Averaged Across All Sites

Incision Size	Sample Size	Mean	Standard Deviation	Standard Error	95% Confidence Interval Lower Limit		Upper Limit
Small	567	5.16	2.30	.0963	4.97	to	5.35
Smaller	567	5.83	1.74	.0731	5.69	to	5.97

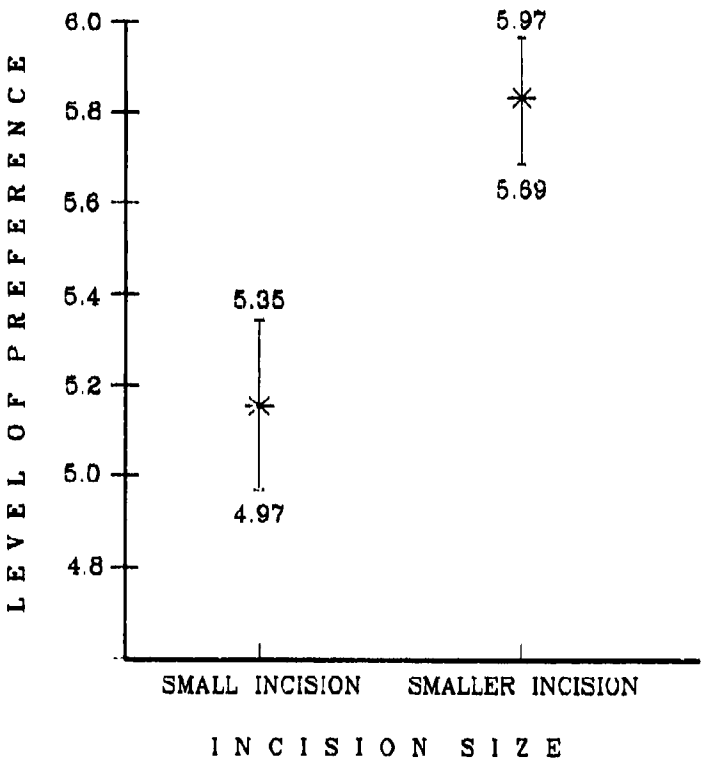


Figure 2. Graphic Display of Table 6

Table 6 and Figure 2 showed that the small incision was preferred over the smaller incision. Table 7 contains the analysis of variance performed upon the data of Table 6 to determine if these differences are significant.

Table 7

Comparison of the Mean Preference of the Small Incision vs the Smaller Incision

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-Ratio	Significance Level
Incision	1	128.6808	128.6808	31.0538	0.000*
Error	1132	4690.7760	4.1438		
Total	1133	4819.4568			

Bartlett's Test for Homogeneous Variances

Number Degrees of Freedom = 1

F = 42.300 Significance Level α = 0.000*

*Significant at α less than 0.001 level

Table 7 indicated that the two textures, small and smaller incisions, differed significantly ($\alpha < 0.001$) from each other. The small incision blended better with the desert background than the smaller incision. The Bartlett's Test indicated that the variances of the two textures were not homogeneous, i.e., significantly different, so they are not necessarily from the same population.

Table 8 contains the nonparametric results of the Kruskal-Wallis One-Way ANOVA Test as a check upon the Duncan's data. The Bartlett's Test indicated that the variances for the two levels of texture were not homogeneous. Parametric tests, such as Duncan's, assume homogeneity of variance.

Table 8

Kruskal-Wallis One-Way ANOVA -- Textures Small and Smaller

Incision	Sample Size	Mean Rank
Small	567	532.04
Smaller	567	602.96
	Total 1134	

Sample Size	Chi-Squared	Significance Level
1134	13.2978	0.0003

Corrected for Ties

Sample Size	Chi-Squared	Significance Level
1134	16.3925	0.0001

These results are in agreement with those of Table 7.

3.3 Color and Incisions

It is important to determine if the variables color and texture were independent of each other. To determine this, a two-way analysis of variance was performed, and the results are seen in Table 9.

Table 9

Two-Way Analysis of Variance for Color and Texture

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-Ratio	Significance Level
Colors	6	938.963	156.494	48.666	0.000*
Incisions	1	128.681	128.681	40.016	0.000*
Interaction	6	150.233	25.039	7.786	0.000*
Error	1120	3601.580	3.216		
Total	1133	4819.457			

*Significant at α less than 0.001

This table verifies the results of Tables 3 and 6 with significant F values for colors and incisions. It also shows a significant interaction between colors and textures. This means that for some nets, the effect of either color or incision was more important than for other nets.

3.4 Nets

The following analysis was done to determine the best net(s) in their ability to blend with the desert background. Table 10 contains the descriptive data.

Table 10

Mean Preference and 95-Percent Confidence Interval for Final Camouflage Nets on Ability To Blend With Desert Backgrounds

Nets	Sample Size	Mean	Standard Deviation	Standard Error	95% Confidence Interval Lower Limit		Upper Limit
Small Incision							
I	81	6.42	1.2027	.1336	6.15	to	6.69
III	81	5.75	1.7141	.1905	5.37	to	6.13
V	81	4.32	2.4332	.2704	3.78	to	4.86
IX	81	2.81	2.3298	.2589	2.30	to	3.33
XI	81	6.23	1.5594	.1733	5.90	to	6.58
XII	81	5.31	2.2397	.2489	4.81	to	5.80
XIII	81	5.25	2.1362	.2374	4.77	to	5.72
Smaller Incision							
a	81	6.44	1.2748	.1416	6.16	to	6.73
c	81	7.00	0.0000	.0000	7.00	to	7.00
e	81	5.52	1.4501	.1611	5.20	to	5.84
i	81	4.38	1.9658	.2184	3.95	to	4.82
k	81	5.58	2.1087	.2343	5.11	to	6.05
l	81	5.73	1.9040	.2116	5.31	to	6.15
m	81	6.16	1.2694	.1410	5.88	to	6.44

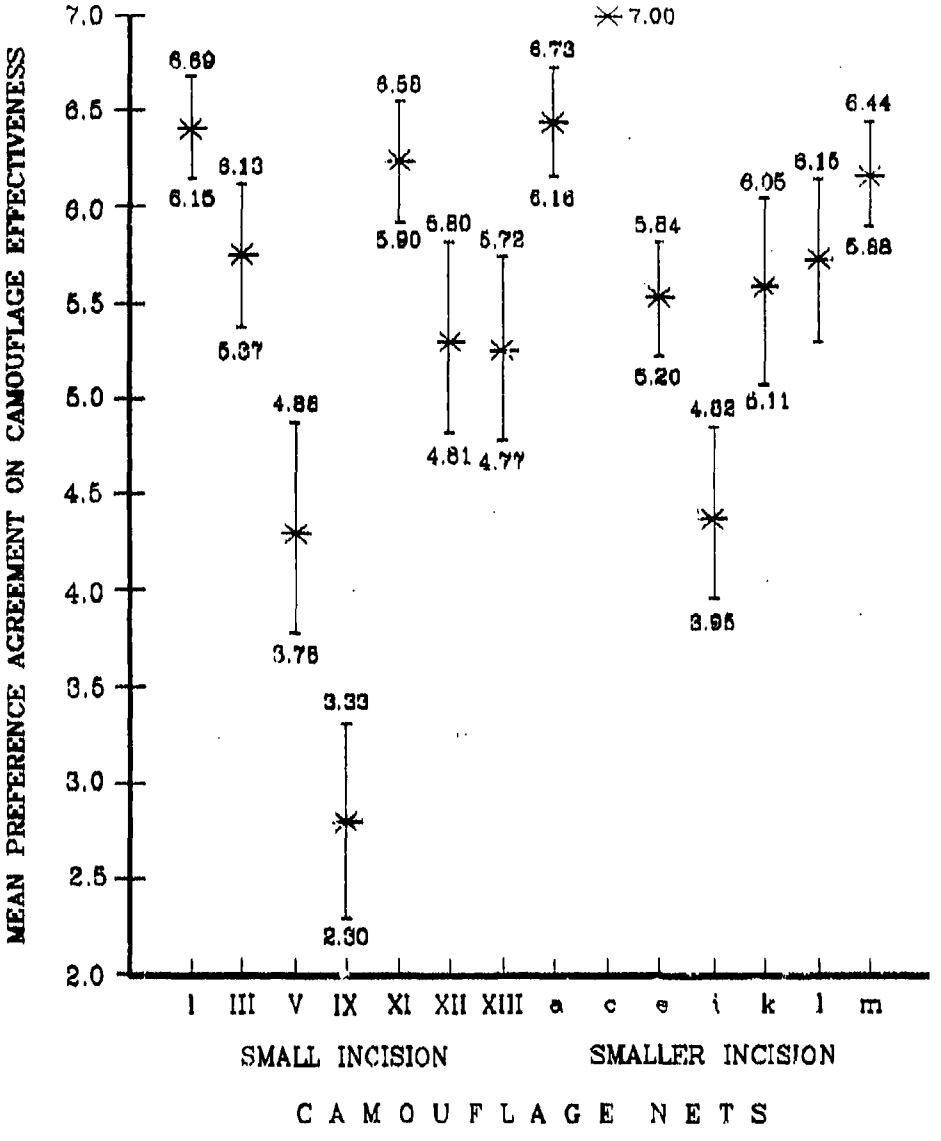


Figure 3. Graphic Display of Table 10

An analysis of variance was performed upon the data in Table 10 to determine if there were significant differences between the camouflage nets in their ability to blend with the desert backgrounds.

Table 11

Analysis of Variance of Nets as to Blend with Sites

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-Ratio	Significance Level
Nets	13	1217.8765	93.6828	29.1330	0.000*
Error	1120	3601.5803	3.2157		
Total	1133	4819.4568			

Bartlett's Test for Homogeneous Variances

Number Degrees of Freedom = 13

F = 9.120 Significance Level α = 0.000*

*Significant at α less than 0.001 level

Table 11 indicated that there were significant differences in the ability of the final desert nets to blend with the background. The Bartlett's Test indicated that the variances of each net were not homogeneous, i.e., significantly different, so they were not necessarily from the same population.

The Duncan's Multiple-Range Test (Table 12) was used to separate this set of significantly different means into subsets of homogeneous means.

Table 12

Duncan's Multiple-Range Test -- Final Camouflage Nets

Best													Worst
Subset 1		Subset 2		Subset 3		Subset 4		Subset 5		Subset 6		Subset 7	
Net	Mean	Net	Mean	Net	Mean	Net	Mean	Net	Mean	Net	Mean	Net	Mean
IX	2.81	V	4.32	XIII	5.25	k	5.58	l	5.73	m	6.16	I	6.42
		i	4.38	XII	5.31	l	5.73	III	5.75	XI	6.23	a	6.44
				e	5.52	III	5.75	m	6.16	I	6.42	c	7.00
				k	5.58	m	6.16	XI	6.23	a	6.44		
				l	5.73								
				III	5.75								

Net IX was the most preferred by the ground observers as to its ability to blend with the desert backgrounds. Nets V and i were in the second-best group. Note that net IX and i were the same color. The only difference between the two is the size of the incision. Net IX has small incisions, while net i has smaller incisions.

Table 13 contains the nonparametric results of the Kruskal-Wallis One-Way ANOVA as a double-check upon the Duncan's Test. This was done, because the Bartlett's Test indicated non-homogeneity of variance and parametric tests, such as Duncan's, assume homogeneity of variance.

Table 13

Kruskal-Wallis One-Way ANOVA--Desert Nets

Net	Sample Size	Mean Rank
I	81	707.63
III	81	597.93
V	81	422.30
IX	81	244.13
XI	81	692.50
XII	81	558.17
XIII	81	501.58
a	81	709.80
c	81	810.50
e	81	520.25
i	81	372.17
k	81	583.98
l	81	592.65
m	81	631.41
	Total 1134	

Sample Size	Chi-Squared	Significance Level
1134	219.7193	0.000*

	Corrected for Ties	
Sample Size	Chi-Squared	Significance Level
1134	270.8542	0.000*

*Significant at α less than 0.001

These results are in agreement with Table 11.

4.0 SECTION IV -- DISCUSSION

Table 1 identified the seven nets, their color, and type of incision, that made the final six preferred nets for at least one of the five desert sites, on their ability to blend with the background. An inspection of Tables 2-5 indicated that color IX/i was the most preferred color in the ability to blend with the desert backgrounds. It was significantly ($\alpha < 0.001$) better than the second most preferred color V/e. The data in Tables 6-8 indicated that the small incision was significantly ($\alpha < 0.001$) preferred over the smaller incision in blending with the desert background. Table 9 shows a significant ($\alpha < 0.001$) interaction between the variables of color and incision. This is interpreted to mean that for some nets, the effect of color and/or incision size is more important than for other nets. All colors are going to appear even lighter when textured with the smaller incision. The smaller the incision size, the more garnish material is visible, and the less "graying" out of the color will result, due to the smaller shadows. Conversely, the larger the incision cut, the less garnish surface is visible, and the more "graying" out of the visible color surface by the effect of shadows. The effects of the interaction of color and texture are also seen in Tables 10-12, which identify which camouflage nets were the most preferred in blending with the desert. Net IX is the same color as

net i, only it has the small incision cut, while net i has the smaller incision cut. Yet net IX is all alone in the group that best blended with the desert background. Net i is in the second-best group, along with net V, which has the small incisions. Net e, which placed in the third group, had the same color as net V, only it was textured with the smaller incision cut. For darker colors such as XI/k, the smaller incision cut created a lighter overall color and enabled the net to blend better with the desert than the small incision cut.

The parametric Duncan's Multiple-Range Tests and the nonparametric Kruskal-Wallis One-Way ANOVA tests were in close agreement with each other. This occurred even though the Duncan's Multiple-Range Test yielded a significant F value. Previous work by Neubert et al[3/] had found similar results. Thus it appears that the assumption that the variances must be homogeneous for parametric statistics can be overlooked without obtaining invalid data.

5.0 SECTION V -- SUMMARY AND CONCLUSIONS

Eleven colors were selected for field evaluation in Saudi Arabia. Two incision sizes, small and smaller, were also investigated. A total of twenty-four nets were made for this study. Twelve of the nets had the small incision, and twelve nets had the smaller incision. The same color was used on pairs of nets, i.e., one for the small incision and one for the still smaller incision. In addition to these nets, a two-color net (i.e., one color on each side) was developed for each incision size for evaluation in both light and darker desert areas. Two standard U.S. Army desert nets were also included for control and comparison purposes, making a total of twenty-six hexagon nets. The small-incision nets were identified by Roman numerals I through XIII. The smaller-incision nets were identified by the lower-case alphabetical letters a through m. Test nets V and e had a different color on each side, and nets VIII and h were the standard U.S. camouflage nets. A total of five different sites were selected in Saudi Arabia. They were viewed by sixteen male subjects (seventeen for site 1). The nets were evaluated on their ability to blend with the desert background. A statistical analysis of these ratings produced the following conclusions:

A. The most effective color was IX/i for blending with the desert background.

B. The small incision produced a net that blended better with the desert background. An exception to this finding can occur when the net color is dark, such as color XI/k (Tables 10 and 12).

C. Net IX was the most effective in blending with the desert background.

6.0 SECTION VI -- EPILOGUE

As stated in Section I, this study was done for SANG. The standard U.S. Army desert net, identified in the test as nets VIII and h, remained in use with U.S. forces with minor changes. However, with the advent of Desert Shield, the U.S., based on the results of the SANG program, began a special purchase of net IX for shipment to Saudi Arabia.

REFERENCES

1. Natrella, Mary G., *Experimental Statistics*, National Bureau of Standards Handbook 91, U.S. Department of Commerce, Washington, D.C., 1966.

2. Siegel, Sidney, *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Book Company Inc., 1956.

3. Neubert, Christopher J., Anitole, George, and Johnson, Ronald L., *Comparison of Parametric Versus Nonparametric Evaluation upon Non-Homogeneous Field Data*, Twenty-eighth U.S. Army Operations Research Symposium, Fort Lee, Va., 10 October 1989.

Maintaining Incremental Optimality when Building Shortest Euclidean Tours

T.M. Cronin
CECOM Center for Signals Warfare
Warrenton VA 22186-5100

Abstract.

Reported upon are experimental runs of an algorithm designed to maintain incremental optimality when building tours for the Euclidean traveling salesman problem. Unlike the Lin-Kernighan edge exchange or Padberg-Rinaldi branch-and-cut techniques which begin with suboptimal tours and proceed by iterating in an attempt to converge upon or exceed the Held-Karp lower bound, the new algorithm strives to maintain optimality as each city is inserted. In previous Army research at the CECOM Center for Signals Warfare, proofs were obtained to show that the underlying search space for the Euclidean traveling salesman problem is piecemeal quartic and hyperbolic. To exploit this new knowledge, the author has developed a dynamic programming algorithm which begins with a baseline tour consisting of the outer (inner) convex hull of cities, and proceeds by adding a city at a time to the interior (exterior). How the city is inserted into the existing tour is dictated by a set of quartic and hyperbolic loci which separate existing and hypothesized subtours from each other. The insertion may involve three different non-linear operations: hyperbolic extension, quartic shunting, and quartic interchange. To test the efficacy of these operators with regard to type 1 and 2 statistical errors, the algorithm as currently implemented is run against a benchmark of city databases for which the optimal tours are known. For those runs which result in a suboptimal solution, an explanation is sought to facilitate fixes to the formal design specification, and the code is subsequently changed. In this paper, the most recent set of runs is analyzed and reported upon, and a prognosis for scaling up to large databases is forecast. The theory predicts that a run should consume time as a function of n^3, where n is the number of cities; this bound is checked empirically by plotting city size vs. CPU time for several databases.

Background.

The Euclidean traveling salesman problem [ETSP] is a long-standing problem in optimization, having roots and primary development in the field of operations research, with ancillary developments in the fields of computational geometry and graph theory. As is the case with many obtuse problems in mathematics, the ETSP may be succinctly stated. Given a set of cities and the distances between each pair, find the shortest tour which visits each city exactly once, except the start city, which is revisited at tour's end. A tour is simply a closed loop connecting all the cities; the formal mathematical name for a tour is a Hamiltonian cycle. One of the interesting facts discovered early on is that a tour is not permitted to cross itself [F1]. There are (n-1)! / 2 possible tours through n cities, which is a combinatorially prohibitive number of operations to perform by brute force, so it is therefore desirable to find an algorithm which arrives at a solution in polynomial time. The ETSP is a special case of the general traveling salesman problem, the former bearing the distinction that the metrics involved are Euclidean distances rather than arbitrary costs or weights.

To date, the Euclidean traveling salesman problem remains unsolved. By "unsolved", it is meant that no one has developed a formal proof of optimality for a polynomial-time algorithm guaranteed to produce the shortest tour. In the mid-seventies, it was proven that the ETSP is NP-hard [G1]. This is a somewhat less damning complexity result than that obtained for the general traveling salesman problem, which belongs to the NP-complete class of problems [G2]. There have been two camps of researchers working on the Euclidean version of the problem, with the earliest computational work dating back to the end of the second world war [L2]. The first camp has striven to produce an exact solution to the problem, and in doing so has pioneered advances in the field of linear programming, including such techniques as the simplex algorithm, branch-and-bound, and branch-and-cut [P1]. An exact approach favors precision at the cost of performance. The second camp of researchers has settled for an approximate approach, by resorting to heuristics which produce high quality solutions per unit of processing time. The principal heuristic techniques are k-opt edge exchange (the most advanced of which is the iterated Lin-Kernighan), simulated annealing, genetic algorithms, elastic bands, and neural nets [J1]. Generally, the approximate techniques develop a solution with more speed than exact approaches, at the cost of precision. However, even this

generality is moot, because some of the heuristic approaches render solutions orders of magnitude faster than others, with only marginally inferior solutions.

Applications.

There are a myriad of applications for the traveling salesman problem. Among them are job scheduling, resource-constrained scheduling, optimal component placement, minimal hookup wire, lowest transmission power, and path-constrained network flow [L2, J1]. The structural similarities between the problems may be somewhat subtle. For example, to map the traveling salesman problem onto job scheduling, the names of the cities are replaced by job names, and the costs between cities are replaced by the job setup times between respective jobs. The job times themselves are considered to be constants; the setup times between jobs turn out to be the crucial factor.

Some of the applications have been shown to be NP-complete problems (e.g., resource-constrained scheduling and path-constrained network flow), and therefore have failed to yield to polynomial-time algorithms [G2]. Thus, it would seem preferable at this point in time to approach such problems with approximate techniques. However, in the long term, if a polynomial-time exact solution may be obtained for the Euclidean traveling salesman problem, then it may be feasible to map the resultant algorithm onto one of the harder problems in such a way that a high quality (albeit suboptimal) solution is achieved. This mapping is arguably most suitable for that class of problems for which the triangle inequality is valid.

Verifying the Optimality of a Tour.

To test a ETSP algorithm (whether it be exact or approximate) against large databases, it is necessary to have at hand some technique to verify an optimal solution in polynomial time. For city databases of size one hundred or less, it is possible to use a variant of branch-and-bound to check optimality in reasonable computer time [J1]. However, when n becomes much larger than one hundred, certifying optimality begins to consume unreasonable amounts of time. It is for this reason that a technique based on computing a lower bound on optimal tour length has been developed [H1]. This quantity, known as the Held-Karp lower bound, is computable in polynomial time, and empirical results indicate that it is consistently within two percent of optimal [J1]. Scientists in the field of operations research have made good use of the bound. Rather than strive for an optimal tour, researchers instead attempt to come within a reasonable neighborhood of the Held-Karp bound.

The Discovery of the Non-linear Search Space for the ETSP.

Despite over forty years of intense study by computer scientists and operations research analysts, the search space for the Euclidean traveling salesman problem remained unspecified as of 1990 (i.e., it was not known whether the mathematics of tour construction was linear, non-linear, or transcendental in the number of cities). This lack of knowledge prompted the author to conduct experiments during the winter of 1990, in an attempt to characterize the space by leveraging the recently developed field of computational geometry upon the problem. In 1968, researchers at the Johns Hopkins University reported upon a slight modification to a theorem due to Barachet to show that an optimal tour must preserve the order of the convex hull of cities - the shortest tour must contain these cities in the order in which they appear about the perimeter [B1, B2]. This fact suggested that an experiment which inserts an arbitrary city into a hull could serve as a valuable testbed in which to discover the geometric locus of equal hull perturbation. A *perturbation* is a subtour which leads into the interior of the hull through two adjacent hull vertices, to capture cities which do not lie on the hull. In conjunction with a perturbation we introduce the *elliptic distance* between a segment and a point p, which is defined to be the sum of the distances from the endpoints of the segment to p, minus the length of the segment.

When comparing a perturbed hull segment against another perturbed segment, one is actually comparing a confocal system of ellipses against another, under a continuous spectrum of elliptic distances. The foci of the two families of ellipses are respectively the two endpoints of the hull segments being perturbed. In Army research at the CECOM Center for Signals Warfare performed during the 1990 fiscal year, it was discovered that the search space induced by the intersection of the two confocal systems of ellipses is in general fourth order (quartic), and in special cases hyperbolic [C2]. The same non-linear behavior is manifested as more cities are added to the interior, which means that the general search space is piecemeal fourth and second order regardless of the number of cities added to

the tour from within the hull. Dynamic programming immediately suggested itself as an approach to the problem which might provide the framework to keep track of the quartic and hyperbolic boundaries of equal tour perturbation when a new city is added to the existing space. Armed with the the new information about the non-linear search space, the author has proceeded to develop a dynamic programming algorithm to maintain incremental optimality when building shortest Euclidean tours.

A Dynamic Programming Algorithm for the ETSP.

The algorithm is based on a principle of incremental optimality: the shortest tour containing k cities is a quartic and hyperbolic function of the shortest tour containing k-1 cities. Beginning with a baseline tour consisting of the convex hull of cities (the smallest bounding polygon containing all of the cities), one city at a time is inserted into the tour in an attempt to preserve the original optimality guaranteed by the hull. As currently specified by the algorithm, there are three types of topologies to maintain in parallel when developing a tour. First, the tour may be simply extended by inserting the new city into the space between those two cities for which the elliptic distance is smallest; this topology is termed *extension space*. A second topology is one in which the new city causes a shunt to be formed between two existing perturbations, to form a new perturbation between the two older ones; this structure is called *shunt space*. The final topology is one which deals with interchanges between perturbations which are "across the hull" from each other; this structure is termed *interchange space*. Extension space is designed to capture the hyperbolic discriminator inherent to extending an existing perturbation, whereas shunt and interchange space are models of the quartic discriminator instructing when to perform a global merger of perturbations.

A nested hull decomposition is computed during a preprocessing step. The decomposition may be computed in O [n * log n] time, as proven by Chazelle [C1]. The nested hull structure, also known as the "onion", is devised to control the order in which the interior cities are inserted. To limit the generation of greedy perturbations, those cities nearest the outer hull are installed first. The set of hulls is visited one at a time, and each hull is traversed in a counterclockwise fashion, until the set of all interior cities is exhausted. Therefore the order of insertion is dictated by a major key equal to the ordinal number of the hull in which a city resides, with a minor key equal to the relative counterclockwise position within the hull (N.B., there are exceptions based on the angle which the city forms with the tour). An alternative strategy is to begin with the innermost hull (the core of the onion) and probe outwards one hull at a time until all exterior cities are processed. Since the quartic and hyperbolic boundaries extend both inside and outside the boundary defined by the current tour, the theory guarantees that it *is* legitimate to process the nested hull decomposition in either direction, with the same optimal solution produced regardless of the processing order. An example of bi-directional processing is demonstrated in the appendix for the capitals of the forty-eight contiguous states of America.

The City Databases.

Seven sets of data (Fig. 1) are currently being used as a testbed for the dynamic programming algorithm. The first is a ten city problem published by Barachet in 1957 [B1]. The optimal tour for this small problem is discussed and derived below. The second set is a sixteen city problem which appears in a seminal computational geometry textbook [P2]. The third, fourth, and fifth sets are databases of twenty, thirty-seven, and forty-one cities which were generated to exhibit non-random behavior; they respectively represent a hull containing a single loop of interior cities; a block letter "E"; and a block letter "S". For these three databases, the shortest tours are not known with certainty (insufficient resources precluded certifying optimality with the branch-and-bound technique utilized by the operations research community), but it is conjectured that they consist of the visually-obvious structured boundaries of the hand-crafted figures. The loop dataset is discussed below, and a temporal history of the conjectured optimal tour is contained in the appendix. The sixth dataset is a forty-eight city problem solved to optimality by AT&T Bell Laboratories in 1985 [A1]. The development of its optimal solution is also contained in the appendix; both an inside-out and outside-in nested hull traversal are graphically portrayed, with the same optimal solution being obtained. The seventh and last dataset is a one hundred twenty-seven city problem formulated by the University of Augsburg in 1989 [R1]; this dataset has recently been solved to optimality by the new algorithm, but a detailed description of the optimal tour is not included here, since it will serve as a primary example in the development of a theorem to be published in a forthcoming paper [C3].

Also described below is a set of experiments in which eighty sets of cities are generated at random to be used as databases to test the analytically-derived time complexity bound for the dynamic programming algorithm.

For these eighty databases, the optimal tour lengths are unknown but are not required, for the sole purpose in using the cities is to statistically test the run time performance of the current implementation of the algorithm, independent of the fact that the solutions developed may not be admissible.

Database	Optimality Known?	Length of Optimal Tour	n =
Barachet	Yes	948	10
Preparata	Yes	1774	16
Loop	No	1919	20
E-figure	No	1258	37
S-figure	No	1770	41
Capitals	Yes	3352	48
Augsburg	Yes	4731	127

Figure 1. The seven benchmark databases, with associated optimality information. The tour lengths are expressed as a function of pixels of the computer raster.

An Analysis of the Barachet Dataset.

Figure 2 is a visual graphic of a shortest tour evolving in time as interior cities are incrementally processed. The data is the Barachet dataset, published in 1957 [B1]. The original constellation of cities is shown in the upper left corner, followed by a graphic of the convex hull, which in this case is simply a square. The dynamic programming algorithm then proceeds to add each of the five interior cities to the tour. The simple extensions, represented by the H_E operator, turn out to be not very interesting. The insertions which produce the most profound changes are the interchange operators, designated H_S. As an example, the last state (frame nine) is produced by an interchange. The extension shown in the next-to-last frame causes the upper perturbation to yield two cities to the extended perturbation as at frame nine, while at the same time producing a new perturbation from the top, which was seen once before at frame number three. Of course, the sequence would look quite different if the interior cities were to be inserted in an order other than that dictated by the process of nested hull traversal, but the final tour would look the same.

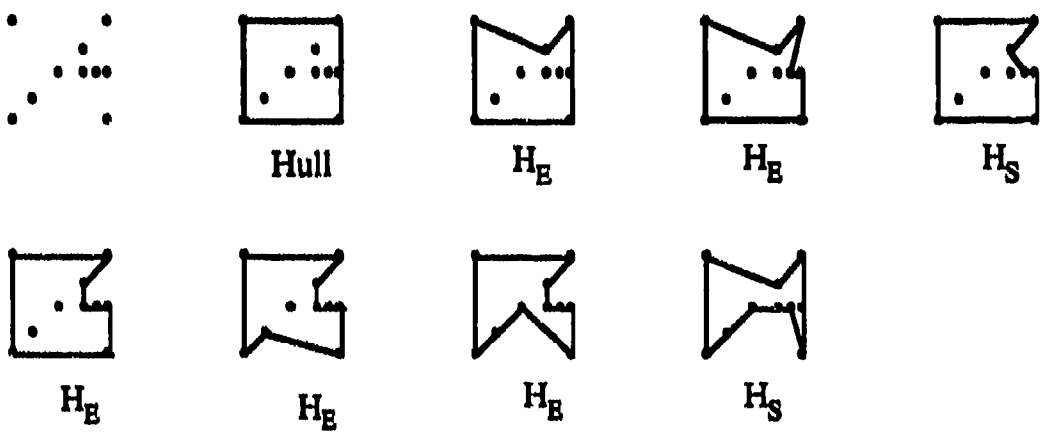


Figure 2. Incremental Optimality Portrayed for the Barachet Data

An Analysis of the Loop Dataset.

Figure 3 is a tabular description of the algorithmic logic manifested when processing the twenty city loop figure (a graphic temporal history of the logic is contained in the appendix). Although the extension and shunting operations are well represented, there are no cross-hull interchanges which occur in this database. The deferral

operations occur because the city under consideration forms a more acute angle with the current tour than does some other sample in the queue of interior cities. In such cases, the city forming the more acute angle is placed back in the queue, and the other city is brought forward for processing. Actually, the deferral and extension operations are mundane when compared to shunts and interchanges. The two interesting insertions for this dataset are the shunts introduced by the addition of cities 19 and 112, both of which radically alter the global tour shape. In particular, the insertion of city 112 causes the lower right portion of the tour to change from a "fishtail" shape to a concave loop. It should be emphasized that the insertion of the cities in some other order might cause the ultimate loop behavior to be displayed earlier, but the algorithm is designed to display the shortest tour for only the cities which are currently entered. A partial tour for k cities may or may not structurally resemble the shortest tour for all n cities.

Entered City	Insertion Operation	Relevant Subtour
15	deferral	-
119	extension	(4,19,20)
15	extension	(4,5,19,20)
116	extension	(20,16,2)
115	extension	(20,16,15,2)
114	extension	(20,16,15,14,2)
113	deferral	-
112	deferral	-
117	extension	(20,17,16,15,14,2)
113	extension	(20,17,16,15,14,13,2)
112	deferral	-
118	extension	(20,18,17,16,15,14,13,2)
112	deferral	-
17	extension	(4,7,5,19,20)
112	deferral	-
16	extension	(4,7,6,5,19,20)
112	deferral	-
18	extension	(4,8,7,6,5,19,20)
112	deferral	-
19	leftsided shunt	(3,9,8,7,6,4,5,19,20)
112	leftsided shunt	(4,5,6,7,8,9,12,13, 14,15,16,17,18,19,20)
110	extension	(4,5,6,7,8,9,10,12,13, 14,15,16,17,18,19,20)
111	extension	(4,5,6,7,8,9,10,11,12,13, 14,15,16,17,18,19,20)

Figure 3. The dynamic programming result for the twenty city loop figure. The insertions of cities 19 and 112 produce back-to-back shunting operations, each of which radically alters the visual appearance of the optimal tour. A temporal history of this example (the loop dataset) is contained in the appendix.

An Analysis of the United States Capitals Dataset.

The appendix concludes with two graphics which depict the development of the shortest tour for the forty-eight capitals of the contiguous United States. The first graphic demonstrates the same approach described above for the Barachet and loop datasets: i.e., a baseline tour consisting of the outer hull is established, and and the interior cities are inserted incrementally by probing inward one hull at a time until all cities are exhausted. The first interesting behavior occurs at row three, column five, with the introduction of Little Rock: Oklahoma City and Jackson are interchanged into Little Rock's new perturbation. Another interchange occurs at row four, column two, when Frankfort is extended into Charleston's perturbation, which subsequently causes Montgomery to be interchanged from below. Yet another interchange occurs in row four, column six, when the introduction of Cheyenne first causes extension space to transpose Bismarck with Pierre, and then forces the interchange of Salt Lake City. The final interchange occurs in row five, column three, when the newly introduced city of Lansing

compels the cities of Albany and Harrisburg to be absorbed into Lansing's perturbation. By far, the most dramatic behavior is encountered at row six, column seven, when the introduction of Springfield forces a left shunt. Springfield is originally attached by extension space between Nashville and Frankfort, but the shunt operator then links it to Jefferson City and synthesizes a new perturbation issuing from the hull segment with endpoints consisting of Baton Rouge and Tallahassee. The final tour shown in row six, column 2 was proven optimal by AT&T Bell Laboratories in 1985 [A1].

Turning to the second graphic in the appendix concerning the forty-eight capitals, the alternative convex hull approach is utilized. This time, the initial tour consists of the innermost hull in the nested decomposition, with vertices comprised of Des Moines, Springfield, Indianapolis, and Columbus. The nested hulls exterior to this hull are subsequently processed, beginning with the one nearest to the inner hull. Because the quartic and hyperbolic loci remain valid regardless of the processing order, the same optimal tour is ultimately obtained at row six, column two.

Some Remarks about the Augsburg Dataset.

The Augsburg dataset consists of the locations of one hundred twenty-seven beer gardens in the city of Augsburg, Germany. This dataset has been solved to optimality by German researchers at the University of Augsburg, using a variant of branch-and-cut [R1]. The same optimal tour is obtained by the dynamic programming algorithm described in this paper. However, analysis of this dataset will not be described here, since it will serve as the primary example in the development of a theorem to be published in a forthcoming paper [C3].

Scaling Up: a 532-City Dataset.

A five hundred thirty-two city dataset was developed by Shen Lin when he was employed at AT&T Bell Laboratories, and represents the locations of AT&T telephone offices in the contiguous United States. A certificate of optimality has been obtained for this data by the originators of the branch-and-cut algorithm [P1]. This database is intriguing because it is the largest database certified to date for which the cities are randomly positioned in the plane (a 2392-city dataset has been solved, but the constellation of cities is formed by repeating the same small pattern of cities several times). The dynamic programming algorithm has not yet been brought to bear upon this database, but it may be feasible to describe the result of its application in the same paper in which the one hundred twenty-seven city solution is discussed.

Time Complexity: A Worst-case Analysis.

The dynamic programming algorithm is continuing to evolve as a research and development tool, and as such remains suboptimal. Nevertheless, it is instructive to perform a worst-case analysis of the code as currently implemented. A condensed algorithmic flowchart is shown at Figure 4. The label "In" is the input loop, in which a new city is input from the front of the queue of unprocessed interior cities. Upon entry from the queue, the city is processed by a routine which checks for intra-perturbation optimality. The new city is first compared against every segment in the current tour to discover the segment of least elliptic distance. This segment may or may not contain the new city's nearest neighbor, so a subroutine is called to check the tour length if an alternative hypothesis allows the connection to occur. To encourage the gradual introduction of cities relative to the perturbed hull, if some other interior city forms a larger angle with the current tour, it is brought forward for processing and the candidate city is put on hold. The intra-perturbation routine concludes by reordering the city's perturbation if necessary to achieve optimality.

Next, a global heuristic is applied to determine if some tour segment forms a larger angle with the newly inserted city than the segment to which it was attached locally via the elliptic distance computation. The global interchange operator attaches the city to such a segment if it exists, and triggers a quadratic matching operation in an attempt to absorb cities from other perturbations. Next in the processing sequence is the synthesis of the left and right shunt topologies. The extended perturbation is attached to both the nearest perturbation on the left and the nearest perturbation on the right, and new perturbations are generated respectively to the left and the right, between the perturbations maintained by the extension space. Once the shunts are computed, the tour lengths for the extension space, the left shunt space, and the right shunt space are compared, and the minimal topology is preserved.

At this point, the interchange operator is invoked once again to absorb cities from other perturbations, using the left and right extension edges of the new city's position as a baseline perturbation. To wrap up the processing of the city, some housekeeping operations are performed to commit the tour and its length to computer memory, before returning to the input loop to process any remaining interior cities.

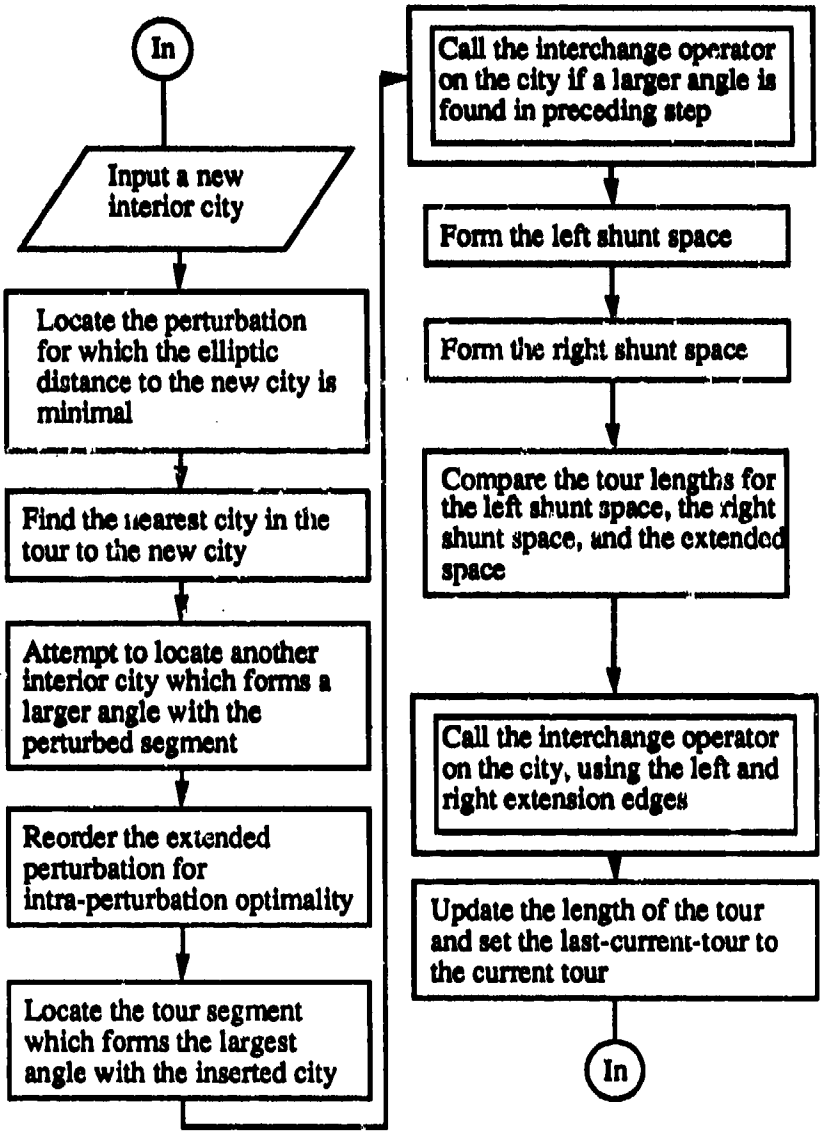


Figure 4. A high level flowchart of the dynamic programming algorithm as currently implemented. For the sake of brevity, several comparison operations of complexity O[1] have been omitted. An interchange operation (double box) is relatively expensive; each interchange entails a quadratic matching, four sorts, and four linear searches.

As the kth city is processed, it is possible for it to trigger two searches of quadratic complexity, eight sorts, each of complexity k*log k, and fifteen searches of linear complexity. All of the boxes in the flowchart represent processes of linear complexity or faster, except for the two double boxes representing the interchange processes. The interchanges are more expensive, in that they involve quadratic matching and sorting in a quest to globally merge perturbations which may lie on opposite sides of the hull. Finally, there is a constant overhead ζ_{OS} associated with

the computer operating system and hardware suite. Therefore, the worst-case processing time t_W is bounded by the following cubic expression:

$$
\begin{aligned}
t_W &= \Sigma\ [\,2k^2 + 8k * \log k + 15k + \zeta_{os}\,] \\
&\leq 2/6\, n * (n+1) * (2n+1) + 8/2\, n * (n+1) * \log n + 1/2\, n * (n+1) + \zeta_{os}\, n \\
&= 2n^3 + 4n^2 * \log n + 7/2\, n^2 + 4n * \log n + (3/2 + \zeta_{os})\, n\,.
\end{aligned}
$$

Complexity theorists refer to such a bound as a ceiling function, because it is derived empirically from an algorithm which has not yet been proven to be optimal, and in general must be considered inferior to a theoretical bound on performance. Conversely, a floor function is obtained analytically from worst-case analysis of an algorithm known to terminate with an optimal solution (usually, once a floor function is established by theory, progress is rapid in bringing an algorithmic ceiling function down to converge upon the floor function). Since a floor function has to date not been established for the Euclidean traveling salesman problem, it is necessary to attempt to empirically lower the ceiling function by resorting to heuristic techniques. The operators depicted in the flowchart are heuristic techniques designed to model the non-linear search space reported upon at [C2]. The intrapath operators at the left represent the hyperbolic portion of the locus, while those on the right approximate the discriminator for the fourth-order components. Many of the minor processing steps which are of sublinear complexity are intentionally omitted, to afford the reader as concise a view as possible of the global logic. The author wishes to stress that the implementation is at best a stopgap measure, which is a useful research tool only until more geometric facts about the search space become available. Indeed, the suggested implementation is already obsolete, due to a new theorem with the potential to dispatch a significant portion of the interior cities during a fast preprocessing step [C3].

An Experiment to Test the Validity of the Analytic Cubic Bound.

Tables 1-8 on the next page are a compilation of a set of experiments designed to test the validity of the cubic bound developed in the preceding section of the paper. The algorithm as currently implemented was tasked against sets of cities randomly distributed on a computer screen (the author used the computer mouse to rapidly input a set of random points to the screen, which were then utilized as coordinates for a city database). The number of cities simulated was allowed to vary from ten to forty-five, in increments of five. For a specific number of cities n, ten sets of random data of size n were generated. Each set was processed by the dynamic programming algorithm, and the following parameters were monitored by the computer operating system: *space* (the number of Lisp cons cells, or computer words, consumed by the run); *time* (the number of seconds of central processing unit time consumed by the run); and *allocation* (the number of seconds of CPU time devoted to dynamic reclaiming of memory, using the Lisp garbage collector).

Only the CPU time (the central column of each dataset) was analyzed statistically. The sample mean, variance, and standard deviation were computed for each set of CPU time data. In addition the best and worst run time outliers were selected for each set. It was anticipated that the worst case outlier would be a good candidate to compare against the cubic bound predicted by the analysis.

Figure 5 is a line graph of the experimental results. For the eighty runs of the algorithm listed in the appendix, the best-case, average-case, and worst-case running times are plotted for each of the eight groupings of ten cities. Also included in the same plot is the cubic bound; the bound is computed for each value of n, and is scaled by the constant .0075 to render the graphic more compact in the ordinate dimension.

It is perhaps imprudent to extrapolate for values of n larger than those shown, but the cubic bound predicted by the theory appears to be a reasonable ceiling function for the worst-case performance of the algorithm. Although there is a gap between the bound and the worst-case outlier, there are valid explanations. One explanation is that an insufficient number of samples were selected to see true worst-case behavior. Another explanation is that the author was overly conservative when conducting a worst-case analysis of the computer code, causing the cubic bound to be somewhat inflated. Yet another explanation is that a more judicious selection of a scalar multiplier of the cubic expression could close the gap. The important thing to note is that the bound is visually well-correlated with the worst-case plot, and that the general behavior of the two curves is markedly similar.

	Space	Time	Allocation
1	7337	12.599	3.380
2	5943	9.739	3.689
3	3271	5.717	1.480
4	2772	4.879	0.000
5	2539	4.342	1.267
6	4726	8.155	2.326
7	3641	6.437	1.579
8	3314	5.714	1.630
9	3245	6.021	1.427
10	4192	7.116	1.399

$n = 10,\ \overline{X} = 7.712,\ S^2 = 10.130,\ S = 3.183$

1	10841	16.515	5.122
2	11926	16.248	8.381
3	13918	20.285	7.593
4	13347	18.926	8.388
5	17042	23.476	9.775
6	15003	20.873	9.793
7	9214	13.009	5.688
8	20864	31.220	11.403
9	9755	13.709	6.302
10	13369	19.800	7.806

$n = 15,\ \overline{X} = 19.406,\ S^2 = 27.906,\ S = 5.283$

1	31893	39.953	18.919
2	39848	52.327	23.654
3	31961	41.991	19.001
4	33496	44.131	19.732
5	28844	36.980	18.365
6	41600	50.265	28.959
7	27420	36.007	17.767
8	27191	34.494	17.022
9	29502	37.315	17.940
10	28340	35.103	18.456

$n = 20,\ \overline{X} = 40.857,\ S^2 = 39.732,\ S = 6.303$

1	63375	72.603	38.143
2	50002	47.761	41.782
3	48373	56.510	27.231
4	52179	67.577	30.072
5	53730	65.259	32.730
6	63618	77.347	39.047
7	67187	81.024	40.680
8	64343	78.275	37.596
9	62891	75.275	35.368
10	49037	56.832	28.396

$n = 25,\ \overline{X} = 67.846,\ S^2 = 123.389,\ S = 11.108$

	Space	Time	Allocation
1	102662	121.708	58.073
2	101337	125.221	61.622
3	92058	98.096	66.107
4	108662	132.876	64.330
5	104230	129.015	62.486
6	108849	134.460	66.508
7	82841	97.734	46.839
8	103552	68.177	119.213
9	84362	133.353	285.494
10	101140	124.738	59.313

$n = 30,\ \overline{X} = 116.538,\ S^2 = 469.822,\ S = 21.675$

1	150861	185.943	101.698
2	121924	152.557	86.844
3	161438	197.643	114.430
4	186448	246.866	129.780
5	163579	204.787	114.469
6	161428	212.644	112.123
7	178192	227.529	121.439
8	214286	295.458	152.270
9	180710	196.557	105.282
10	156222	167.035	64.046

$n = 35,\ \overline{X} = 208.702,\ S^2 = 1671.621,\ S = 40.885$

1	209331	224.835	98.215
2	241452	267.132	99.419
3	206614	212.946	85.847
4	263413	295.444	115.531
5	295821	348.702	136.283
6	239564	257.743	110.176
7	229468	263.191	106.129
8	223627	239.093	97.718
9	216796	249.739	96.802
10	210261	244.238	91.558

$n = 40,\ \overline{X} = 260.306,\ S^2 = 1492.335,\ S = 38.631$

1	329234	389.722	244.246
2	354376	411.498	156.177
3	382554	490.126	179.530
4	426650	593.475	208.267
5	293465	301.532	114.221
6	404882	505.157	173.208
7	384530	530.231	278.536
8	341033	413.296	213.390
9	335337	389.147	196.894
10	278591	344.297	178.995

$n = 45,\ \overline{X} = 436.848,\ S^2 = 8147.415,\ S = 90.263$

Tables 1-8. Space, Time, and Allocation Complexity for Eight Sets of ETSP Experiments

(n is the number of cities per experiment; *Space* is the number of Lisp cons cells consumed by a run; *Time* is the number of seconds of CPU time consumed by a run; and *Allocation* is the time dedicated to the Lisp garbage collector)

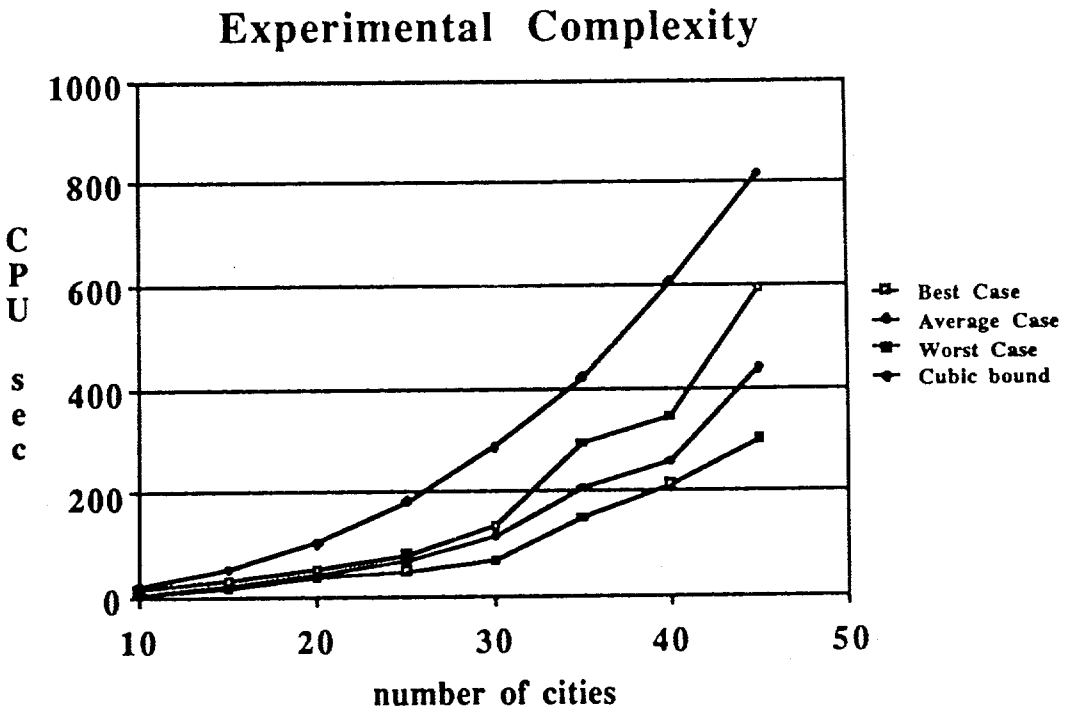


Figure 5. A line plot of the time complexity of the dynamic programming algorithm as a function of the number of cities processed.

Summary.

Some very preliminary statistical experiments on the worst case behavior of an algorithm designed to provide an exact solution for the Euclidean traveling salesman problem indicate that the run time of the algorithm is in fact bounded by an expression cubic in the number of cities. The algorithm is based on some recent theoretical results pertaining to the non-linear search space for the Euclidean traveling salesman problem, and as such the computer code has not yet reached an optimal level of maturity. Nevertheless, it has proven useful to statistically compare the performance of the dynamic programming algorithm against the cubic bound predicted by a cursory examination of the currently implemented software. The worst-case statistical outliers compiled for each set of experiments are indeed bounded by the cubic expression developed analytically from the current formal design specification of the algorithm. It is apparent that the science of statistics is invaluable with regard to gauging the probabilistic performance of an algorithm versus its analytic time complexity bound.

Future Directions of the Research.

An entirely independent issue is whether or not the algorithm is admissible: i.e., whether it terminates with an optimal solution. It is desirable to attempt to prove that the dynamic programming technique is admissible; a proof by induction seems promising. Thus far, the implementation is proceeding in the spirit of the Hungarian mathematician Lakatos, who contended that a theory is never truly proven until sufficient time has passed such that the community at large accepts the theory, based on the fact that counterexamples cease to be forthcoming from empirical testing [L1]. The implementation discussed in the paper is at a stage where counterexamples can still be found. However, the author feels that the counterexamples are sufficiently trivial to be local rather than global, which indicates that the problems remaining to be ironed out are details of implementation rather than profound issues of conceptualization. It seems important to pursue the proof of optimality; otherwise, the new algorithm will be vulnerable to the same kinds of criticism which plague all heuristic approaches to problem solving.

The algorithm will continue to undergo empirical testing, as the number of cities is scaled up. A good source of benchmarks is maintained at reference [R1]. As mentioned above, a one hundred twenty-seven city database

has recently been solved to optimality by the dynamic programming algorithm. It is desirable in 1991 to move ahead to a five hundred thirty-two city certified benchmark [P1]. Unfortunately, there are only a handful of large databases for which a certificate of optimality has been obtained.

In preparation is a paper which describes a new geometric result pertaining to the aspect angle which an interior city forms with the convex hull, and the positive implication of the result as a preprocessing step for the dynamic programming algorithm [C3]. It is premature to forecast the utility of the new theorem, but empirical testing indicates that on the average a surprisingly large percentage of cities interior to the hull may be inserted into the tour in a fast preprocessing step.

Acknowledgments

The research has benefited from technical discussions with the following individuals: Gerald Andersen, Richard Antony, Robert Bixby, Chris Bogart, Paul Broome, Mel Brown, Jagdish Chandra, Douglas Chubb, Ben Cummings, Michael Dillencourt, Francis Dressel, Herbert Edelsbrunner, Geoffrey Fox, Ray Freeman, Martin Groetschel, Andrew Harrell, Robert Hein, William Jackson, David S. Johnson, Shen Lin, Sanjoy Mitter, Jay Sethuraman, J. Michael Steele, Andrew Thompson, Paul Tseng, Franz-Erich Volter, and David Willshaw. I am deeply indebted to Richard Antony for his comprehensive technical review, requests for elaboration, and patient editing of the final draft. A special note of gratitude to Geoffrey Fox and Robert Bixby for an invitation to a productive workshop on computational aspects of the traveling salesman problem (held at Rice University in April of 1990); to David S. Johnson for constructive criticism of the dynamic programming algorithm, and for providing the coordinates for the forty-eight capitals; and to Herbert Edelsbrunner for insights into computing the convex hull. I also wish to thank the management of the CECOM Center for Signals Warfare for establishing a scientific environment in which fundamental research in mathematics and computer science is both encouraged and supported.

Bibliography

[A1] Anonymous, "Here's one for the road, and then some", *Discover*, July 1985.

[B1] Barachet, L.L., "Graphic Solution of the Traveling Salesman Problem", Operations Research 5, 1957, pp. 841-845.

[B2] Bellmore, M., and G.M. Nemhauser, "The Traveling Salesman Problem: a Survey", Operations Research 16, 1968, pp. 538-558.

[C1] Chazelle, B., "On the convex layers of a convex set", IEEE Trans. Info. Theory IT-31, 1985.

[C2] Cronin, T.M., "The Voronoi Diagram for the Euclidean Traveling Salesman Problem is Piecemeal Quartic and Hyperbolic", *Transactions of the Eighth Annual Army Conference on Applied Mathematics and Computing*, Cornell University, June 1990.

[C3] Cronin, T.M., "A Note on the Aspect Angle formed between the Convex Hull and its Interior Points, within the Context of Shortest Euclidean Tours", in preparation.

[F1] Flood, M.M., "The Traveling Salesman Problem", Operations Research 4, 1956, pp. 61-75.

[G1] Garey, M.R., R.L. Graham, and D.S. Johnson, "Some NP-complete Geometric Problems", Eighth Annual Symp. on Theory of Comp., May 1976, pp. 10-22.

[G2] Garey, M.R., and D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman and Company, New York NY, 1979.

[H1] Held, M., and R.M. Karp, "The Traveling Salesman Problem and Minimum Spanning Trees: Part II", Mathematical Programming 1, 1971, pp. 6-25.

[J1] Johnson, D.S., Private communication, and set of viewgraphs entitled "How to Beat Lin-Kernighan", *Workshop on Computational Aspects of the Traveling Salesman Problem*, Rice University, Houston TX, April 1990.

[J2] Johnson, D.S., "Local Optimization and the Traveling Salesman Problem", to appear in the *Proceedings of the Seventeenth Colloquium on Automata, Languages, and Programming*, Springer-Verlag, 1990, pp. 446-461.

[L1] Lakatos, I., Proofs and Refutations: The Logic of Mathematical Discovery, Cambridge University Press, New York NY, 1976.

[L2] Lawler, E.L., J.K. Lenstra, A.H.G. Rinnooy Kan, and D.B. Shmoys, eds., The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization, John Wiley and Sons, New York NY, 1985.

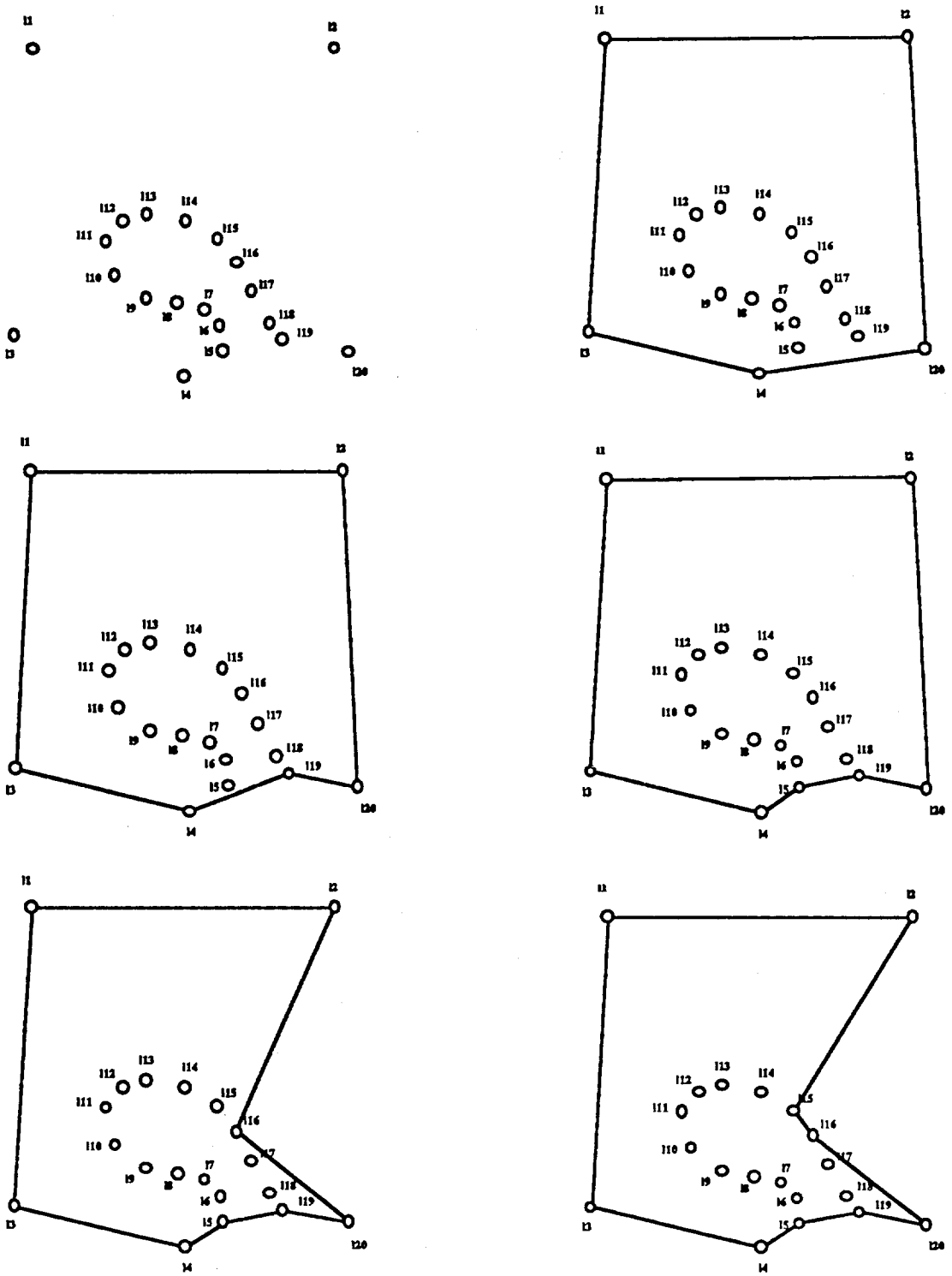
[L3] Lin, S., and B.W. Kernighan, "An effective heuristic algorithm for the traveling salesman problem", Operations Research 21, 1973, pp. 498-516.

[P1] Padberg, M., and G. Rinaldi, "Optimization of a 532-city Symmetric Traveling Salesman Problem by Branch and Cut", Op. Res. Letters, Vol 6, No.1, March 1987.

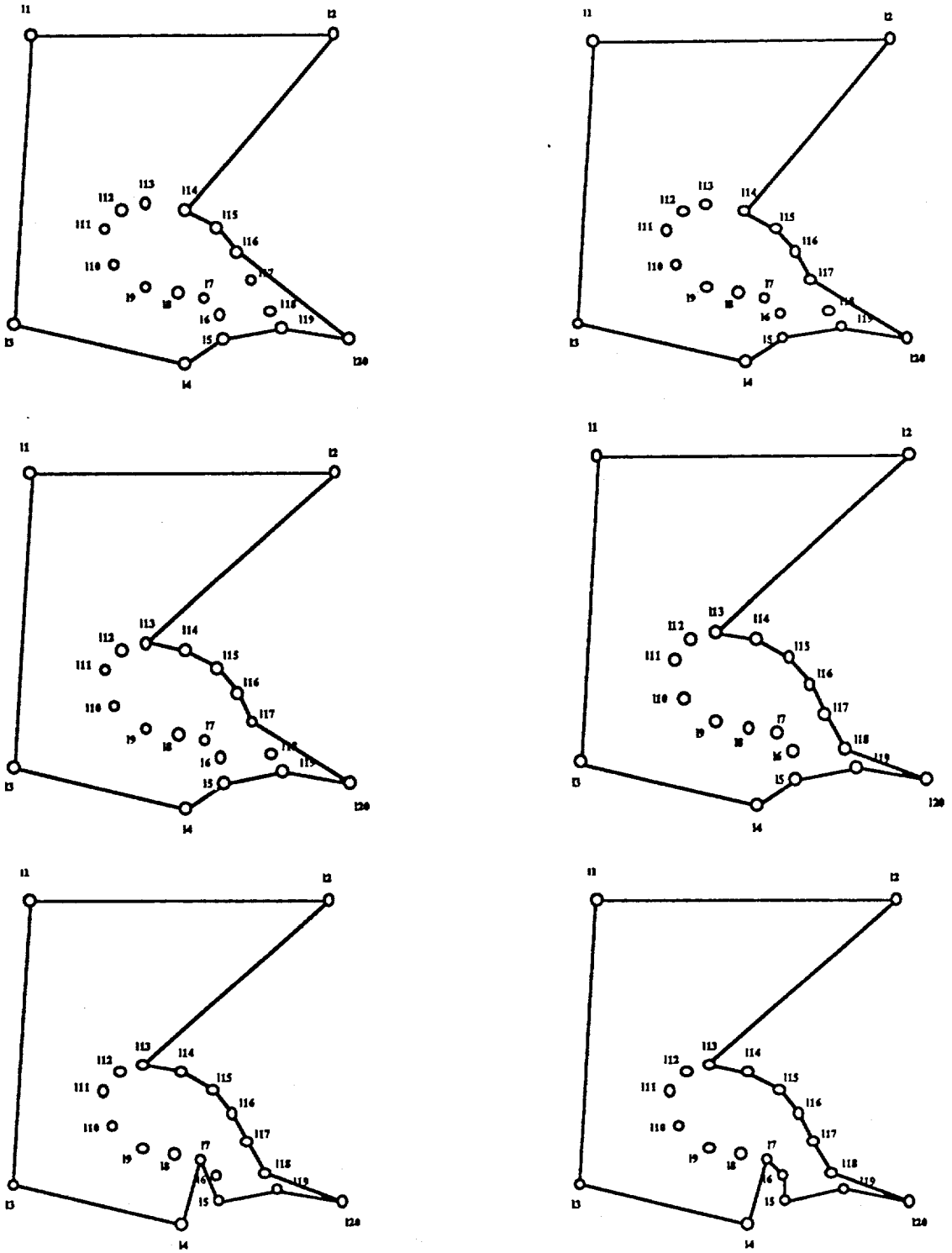
[P2] Preparata, F.P., and M.I. Shamos, Computational Geometry: An Introduction, Springer-Verlag, New York NY, 1985.

[R1] Reinelt, G., "TSPLIB - A Traveling Salesman Problem Library", Institute of Mathematics, University of Augsburg, Augsburg Germany, 20 March 1989.

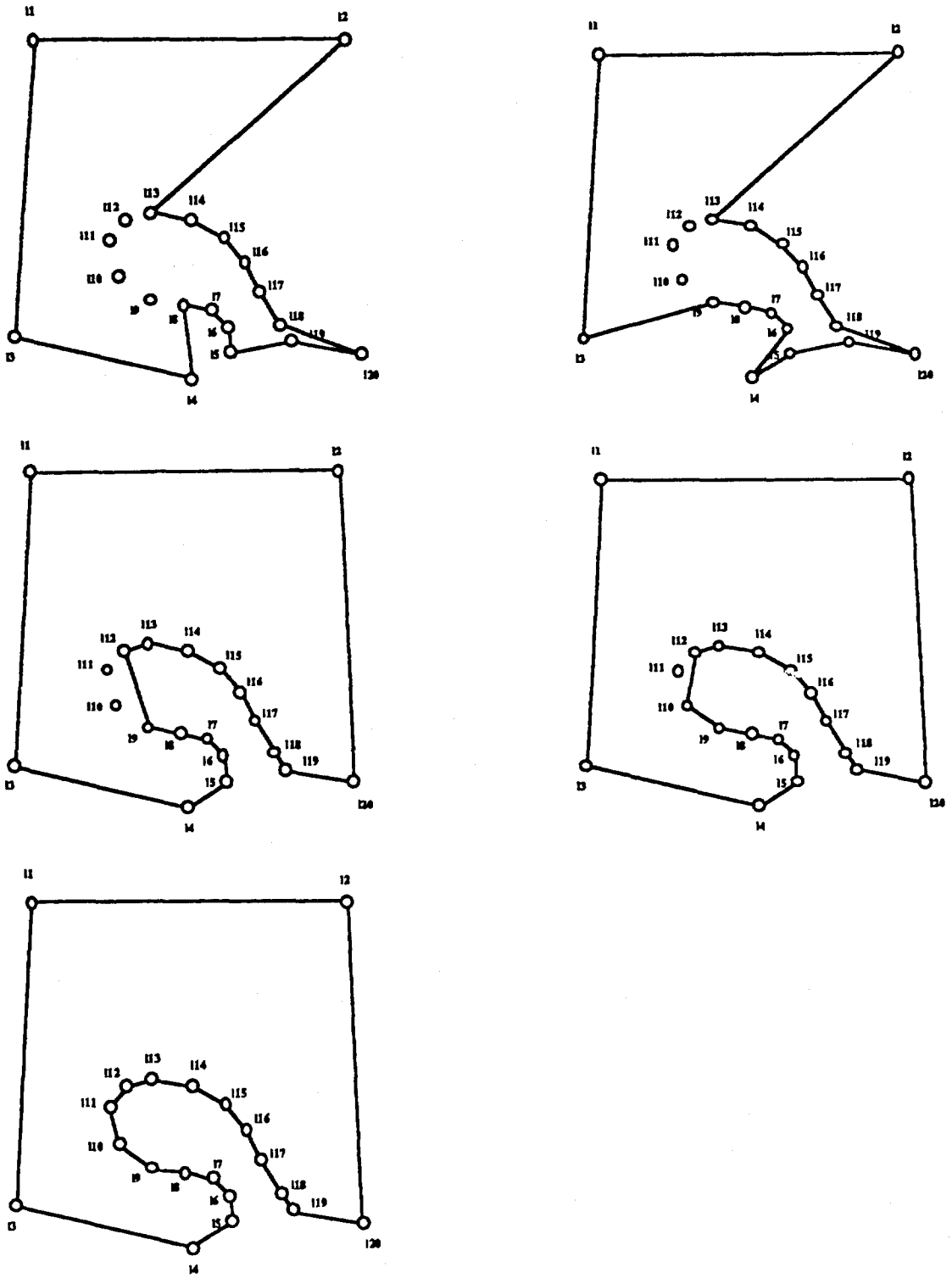
[W1] Walpole, R.E., and R.H. Myers, Probability and Statistics for Engineers and Scientists, Macmillan Publishing Company, Inc., New York NY, 1972.



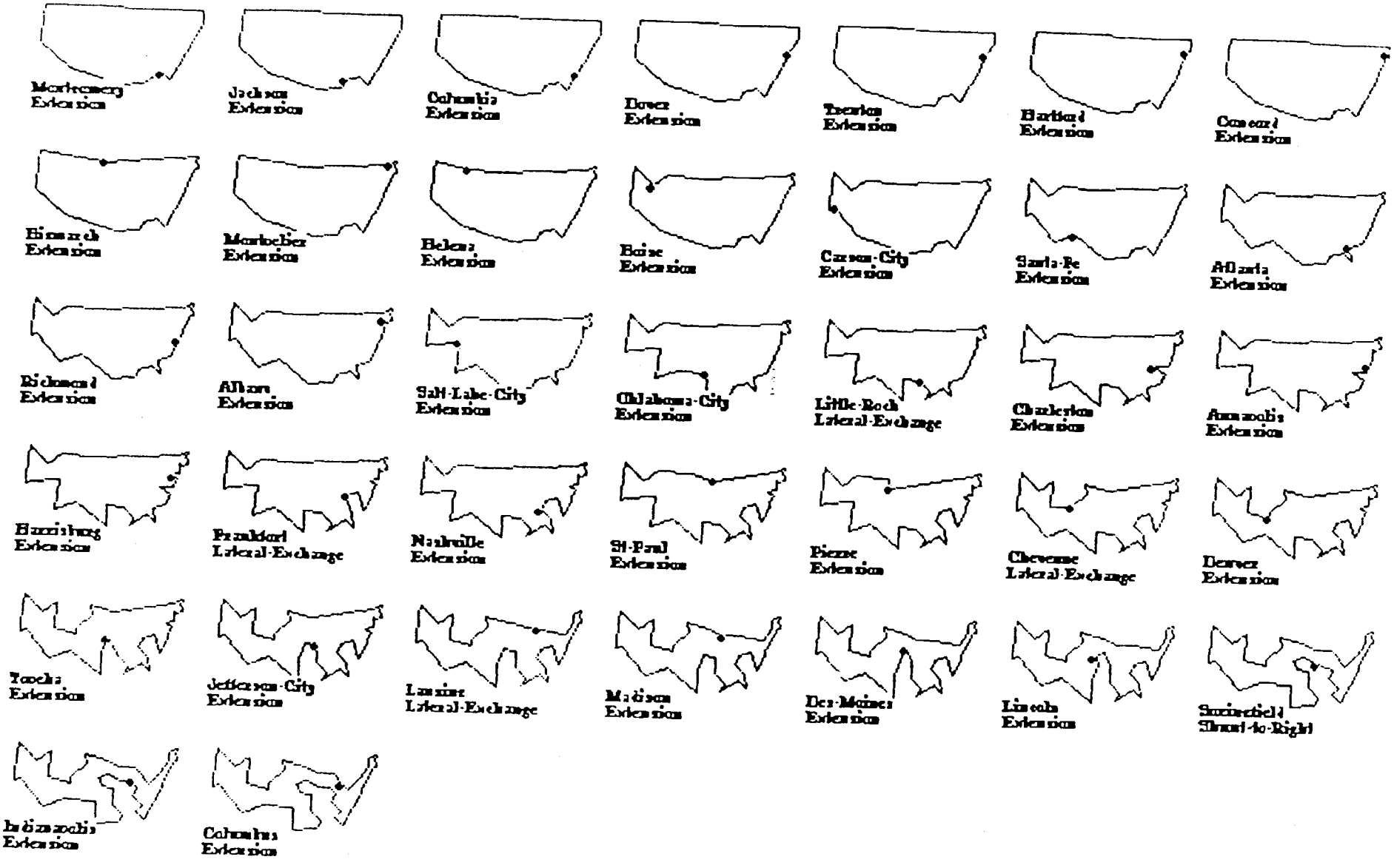
Loop dataset: original data; convex hull;
dynamic programming insertion of interior cities l19; l5; l16; l15.

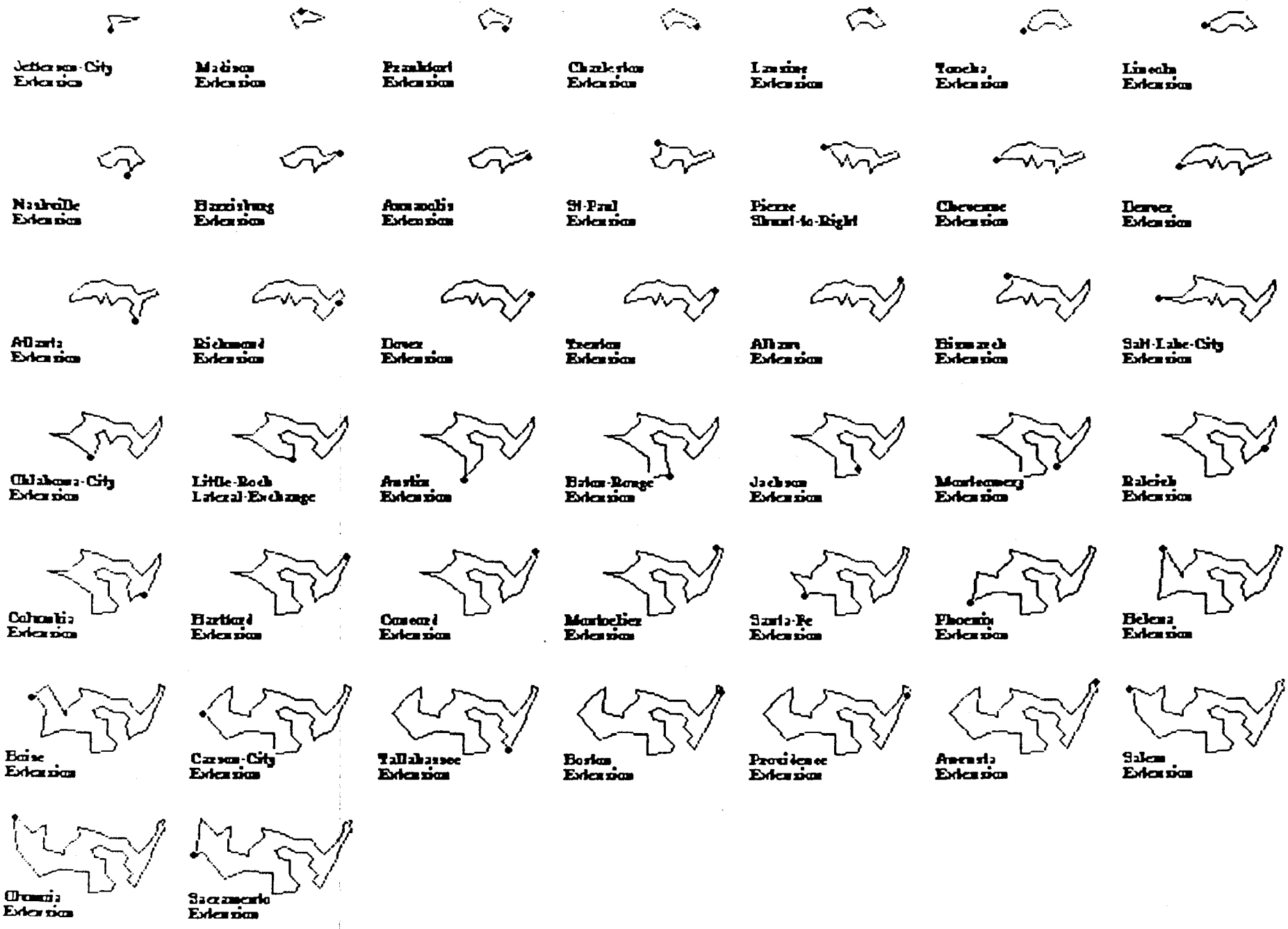


Loop dataset: dynamic programming insertion of interior cities i14; l17; l13; l18; l7; l6.



Loop dataset: dynamic programming insertion of interior cities l8; l9; l12; l10; l11.


Montgomery Extension
Jackson Extension
Columbia Extension
Dover Extension
Trenton Extension
Hartford Extension
Concord Extension
Bismarck Extension
Montpelier Extension
Helena Extension
Boise Extension
Carson City Extension
Santa Fe Extension
Atlanta Extension
Richmond Extension
Albany Extension
Salt Lake City Extension
Oklahoma City Extension
Little Rock Lateral Exchange
Charleston Extension
Annapolis Extension
Harrisburg Extension
Frankfort Lateral Exchange
Nashville Extension
St Paul Extension
Pierre Extension
Cheyenne Lateral Exchange
Denver Extension
Topeka Extension
Jefferson City Extension
Lansing Lateral Exchange
Madison Extension
Des Moines Extension
Lincoln Extension
Springfield Thrust-to-Right
Indianapolis Extension
Columbus Extension


Jefferson-City Extension
Madison Extension
Frankfort Extension
Charleston Extension
Lansing Extension
Topeka Extension
Lincoln Extension
Nashville Extension
Harrisburg Extension
Annapolis Extension
St-Paul Extension
Pierre Shunt-to-Right
Cheyenne Extension
Denver Extension
Atlanta Extension
Richmond Extension
Dover Extension
Trenton Extension
Albany Extension
Bismarck Extension
Salt-Lake-City Extension
Oklahoma-City Extension
Little-Rock Lateral-Exchange
Austin Extension
Baton-Rouge Extension
Jackson Extension
Montgomery Extension
Raleigh Extension
Columbia Extension
Hartford Extension
Concord Extension
Montpelier Extension
Santa-Fe Extension
Phoenix Extension
Helena Extension
Boise Extension
Carson-City Extension
Tallahassee Extension
Boston Extension
Providence Extension
Annapolis Extension
Salem Extension
Olympia Extension
Sacramento Extension

AN ALGEBRAIC DERIVATION OF VARIANCE OF THE GEOMETRIC DISTRIBUTION

Richard M. Brugger
Product Assurance and Test Directorate
U.S. Army Armament, Munitions and Chemical Command
Rock Island, Illinois 61299-6000

ABSTRACT. Typically, the variance of the geometric distribution is derived using generating functions or moment generating functions. This paper provides an alternative using algebra only.

1. INTRODUCTION. Consider an unbounded sequence of Bernoulli trials, each trial having probability p of success and probability $q = 1 - p$ of failure. Let X denote the trial number of the first success. Then X is distributed geometrically.

The geometric distribution plays an important part in many probabilistic considerations. For example, it is the underlying reliability distribution for an item that functions discretely with an assumed constant failure rate, such as a gun. It also plays an important part in sampling plans. Dodge's original work on continuous sampling plans, described below, incorporated geometric distributions extensively.

2. DERIVATION OF DODGE'S h. Dodge (1943) used the following derivation of a geometric distribution curtailed after i trials (where i is fixed, not an index). For the remainder of this paper $0 < p < 1$ and thus $0 < q < 1$.

$$h = \frac{1}{1-q^i}\left(p+2pq+3pq^2+4pq^3+\ldots+ipq^{i-1}\right) \tag{1a}$$

$$= \frac{p}{1-q^i}\left(1+2q+3q^2+4q^3+\ldots+iq^{i-1}\right) \tag{1b}$$

$$= \frac{p}{1-q^i}\cdot\frac{d}{dq}\left(1+q+q^2+q^3+\ldots+q^i\right) \tag{1c}$$

$$= \frac{p}{1-q^i}\cdot\frac{d}{dq}\frac{(1-q^{i+1})}{1-q}\ . \tag{1d}$$

Considering only the derivative portion yields

$$\frac{d}{dq}\cdot\frac{(1-q^{i+1})}{1-q} = \frac{(1-q)(-(i+1)q^i)-(1-q^{i+1})(-1)}{p^2} \tag{2a}$$

$$= \frac{p(-i-1)q^i+(1-q^{i+1})}{p^2} \tag{2b}$$

$$= \frac{1-q^i(1+pi)}{p^2}\ . \tag{2c}$$

Preceding Page Blank

So

$$h = \frac{p}{1-q^i} \cdot \frac{1-q^i(1+pi)}{p^2} \tag{3a}$$

$$= \frac{1}{p(1-q^i)} \cdot (1-q^i(1+pi)). \tag{3b}$$

Brugger (1989) provided the following. Returning to the first expression:

$$h = \frac{1}{1-q^i} (p+2pq+3pq^2+4pq^3+\ldots+ipq^{i-1}) . \tag{1}$$

Since $p = 1-q$, this becomes

$$h = \frac{1}{1-q^i} (1-q+2(1-q)q+3(1-q)q^2+4(1-q)q^3+\ldots+i(1-q)q^{i-1}) \tag{4a}$$

$$= \frac{1}{1-q^i} (1-q+2q-2q^2+3q^2-3q^3+4q^3-4q^4+\ldots+iq^{i-1}-iq^i) \tag{4b}$$

$$= \frac{1}{1-q^i} (1+q+q^2+q^3+q^4+\ldots+iq^{i-1}-iq^i) \tag{4c}$$

The first i terms within the parentheses follow a geometric progression, so the above becomes

$$h = \frac{1}{1-q^i} \cdot \frac{1-q^i}{p} - \frac{iq^i}{1-q^i} \tag{5a}$$

$$= \frac{1}{1-q^i} \cdot \frac{1-q^i}{p} - \frac{piq^i}{p(1-q^i)} \tag{5b}$$

$$= \frac{1}{p(1-q^i)} (1-q^i(1+pi)) . \tag{5c}$$

This is the same as Dodge's result (see equation 3b), but in this latter case only algebra was used to obtain the result.

3. <u>DERIVATION OF THE MEAN OF A GEOMETRIC DISTRIBUTION</u>. Using the method described above, it is straightforward to derive the mean, E(X), of a geometric distribution:

$$E(X) = \sum_{j=1}^{\infty} jpq^{j-1} \tag{7a}$$

$$= 1p+2pq+3pq^2+4pq^3+\ldots \tag{7b}$$

$$= 1-q+2q-2q^2+3q^2-3q^3+4q^3-4q^4+\ldots \tag{7c}$$

$$= 1+q+q^2+q^3+q^4+\ldots \tag{7d}$$

$$= \frac{1}{1-q} = \frac{1}{p} \tag{7e}$$

4. DERIVATION OF THE VARIANCE. By definition,

$$\mathrm{Var}\,(X) = E(X^2)-(E(X))^2. \tag{8}$$

From equation 7e above

$$(E(X))^2 = 1/p^2 \tag{9}$$

is obtained.

$$E(X^2) = \sum_{j=1}^{\infty} j^2 pq^{j-1} \tag{10a}$$

$$= p+4pq+9pq^2+16pq^3+\ldots \tag{10b}$$

$$= 1-q+4q-4q^2+9q^2-9q^3+16q^3-16q^4+\ldots \tag{10c}$$

$$= 1+3q+5q^2+7q^3+\ldots \tag{10d}$$

It is claimed that the series in equation 10d will continue with coefficients equal to the odd integers. This is easily proved. Let n be any positive integer. Then

$$n^2-(n-1)^2 = n^2-n^2+2n-1 \tag{11a}$$

$$= 2n-1 \tag{11b}$$

where the expression in 11b is clearly an odd integer. Summing, we obtain

$$E(X^2) = \sum_{j=1}^{\infty} (2j-1)q^{j-1} \tag{12a}$$

$$= 2\sum_{j=1}^{\infty} jq^{j-1} - \sum_{j=1}^{\infty} q^{j-1} \tag{12b}$$

$$= \frac{2}{p}\sum_{j=1}^{\infty} jpq^{j-1} - \frac{1}{p} \tag{12c}$$

$$= \frac{2}{p}\cdot\frac{1}{p} - \frac{1}{p} \tag{12d}$$

$$= \frac{2}{p^2} - \frac{1}{p} \tag{12e}$$

$$= \frac{2-p}{p^2} \tag{12f}$$

$$= \frac{q+1}{p^2}\,. \tag{12g}$$

Recalling that

$$\mathrm{Var}\,(X) = E(X^2)-(E(X))^2 \tag{8}$$

it is seen that

$$\mathrm{Var}\,(X) = \frac{q+1}{p^2} - \frac{1}{p^2} \tag{13a}$$

$$= \frac{q}{p^2} \tag{13b}$$

5. EXTENSION TO THE NEGATIVE BINOMIAL DISTRIBUTION. If one is interested in some positive integer n (n>1) of successes in a series of Bernoulli trials with constant probability p of success, the trial number of the n'th success is distributed according to a negative binomial distribution, which is an extension of the geometric distribution.

Then

$$E(X_1+X_2+\ldots+X_n) = nE(X_1) \tag{14a}$$

$$= n/p \tag{14b}$$

$$\text{and } \mathrm{Var}\,(X_1+X_2+\ldots+X_n) = n\mathrm{Var}(X_1) \tag{15a}$$

$$= nq/p^2 \tag{15b}$$

REFERENCES

1. Brugger, Richard M., "Eliminating calculus depending in the derivation of Dodge's u", Presented at the Thirty-fifth Conference on Design of Experiements in Army Research, Development and Testing, October, 1989.

2. Dodge, Harold F., "A sampling inspection plan for continuous production", Annals of Mathematical Statistics, 1943, pp 264-279.

A LINEAR PROGRAMMING MODEL FOR QUEUEING IN OPERATIONAL AVAILABILITY

William C. Hoffman
P. O. Box 2005
Sierra Vista, AZ 85635

ABSTRACT. Operational Availability A_0 is the term used to refer to readiness of a military system for battlefield use. The servicing of such a system can be divided into several stages, each of which may be modeled by an Erlang type queue $M/E_k/s$, s being the number of service personnel. Maximum Operational Availability is equivalent to minimization of Total Corrective Maintenance Time (TCM) plus Total Administrative and Logistic Downtime (TALDT). Under appropriate constraints the maximum A_0 problem can thus be formulated as a linear programming problem.

1. INTRODUCTION. TRADOC/AMC Pamphlet 70-11, RAM RATIONALE REPORT HANDBOOK, defines (p. E.18) Operational Availability A_0 by the following formula:

$$A_0 = (OT + ST)/(OT + ST + TCM + TPM + TALDT). \qquad (1)$$

Here OT = Operating Time, i.e., time during a mission profile when the system is "on" and actively performing at least one of its missions;
ST = Standby Time, i.e., the time during a given period when a system is inoperative but is operable;
TCM = Total Corrective Maintenance Time, i.e., the total time spent on restoring a failed item to operational condition;
TPM = Total Preventive Maintenance Time spent in activities to maintain an item in a specified condition;
and TALDT = Total Administrative and Logistics Downtime spent waiting for parts, arrival of maintenance personnel, or transportation, during a given period.

In a tactical situation Standby Time ST and Preventive Maintenance Time PCM can frequently be neglected, in which case the above formula reduces to

$$A_0 = 1/[1 + (TCM/OT) + (TALDT/OT)].$$

It is clear from this expression that for given OT, A_0 is maximized by minimizing TCM + TALDT:

$$\max A_0 \sim \min \{TCM + TALDT\}.$$

The right hand side of this expression will provide the basis for a linear programming approach to optimization of Operational Availability.

The present approach was suggested by a requirement for deciding among competing maintenance systems for a deployed IEW network along the FLOT (Forward Line of Own Troops). The maintenance structure consisted of as many as four stages: mobile contact units (BMCT) which provided minimal repairs, usually in the form of simple replacment of defective line replaceable units (LRU's), at the FLOT; shop repairs (BS) at a forward level; mobile contact units (LMCT's) dispatched from a rear light electronics maintenance company to handle more complex diagnosis and repairs; and a rear echelon repair shop (LS) for major diagnosis and repair/replacment. Associated with each level is a characteristic MTTR (Mean Time To Repair) and an ALDT (Administrative and Logistic Downtime) for maintenance. These parameters will play key roles in the queueing models for the maintenance process(es). Both the MTTR sequence and the ALDT sequence may be modeled as a four-stage, series Erlang queue $M/E_4/s$, where M represents "customers" (i.e., FLOT system elements with defective parts (LRU's)) occurring at random according to a Poisson process, E_4 denotes an Erlang distribution representing a four-stage FIFO service protocol, and s denotes the number of servers, i.e., maintenance personnel at any given level.

<u>2. Estimation of Administrative and Logistic Downtime for Maintenance (ALDTm)</u>. The expected waiting time at a defective LRU site for maintenance personnel is given by the formula [Gaver & Thompson, 1973, p. 505]

$$E\{W_{ALDTm}\} = \Sigma\ (MTTR)_i\ u_i/(1 - u_i), \qquad (2)$$

where the summation on i runs over the 4 stages and the "utilization", or "traffic load factor", is defined to be

$$u_i = MTTR_i/(s_i\ MTBF), \qquad (3)$$

Here s_i is the number of "servers" at stage i, and MTBF is the Mean Time Between Failures. If u_i turns out to be greater than one, the system saturates. Hence $0 < u_i < 1$ will be assumed in the sequel. Since then $u/(1 - u) = u + \Sigma\ u^n \geq u$, Eq. (2) may be written in the series form

$$E\{W_{ALDTm}\} \geq \Sigma\ (MTTR)_i\ u_i = (1/MTBF)\ \Sigma\ (MTTR_i)^2\ (1/s_i). \qquad (4)$$

Apart from costs, the defining of an organizational structure essentially resides in the assignment of personnel and facilities. To utilize (4) in a linear programming formulation of optimality for A_0 we therefore introduce as a new variable x_i, the reciprocal of s_i, i.e., the reciprocal of the number of servers at stage i.

Inequality (4) then becomes

$$E\{W_{ALDTm}\} \geq (1/MTBF)\ \Sigma\ (MTTR_i)^2\ x_i. \qquad (5)$$

The introduction of the new variable x_i makes possible formulation as a linear programming problem, but it is not an unmixed blessing. Being a reciprocal, the variate x_i will accentuate the tendency of a linear programming problem to give solutions in non-integer form and may also lead to instabilities for small assignments.

3. Estimation of Administrative and Logistic Downtime for Transport ALDTt. In the present model, the Administrative and Logistic Downtime awaiting parts from CONUS is neglected since it exceeds the hypothesized time scale. It only remains therefore to formulate ALDTt. Since transport is involved, only the mobile contact teams, the BMCT's and the LMCT's will contribute. The time involved may be estimated by distributing the total number of manhours spent by each in travel between base and FLOT and equating this to the TALDTt requirement:

$$\begin{aligned} E\{TALDTt/F\} = {} & \text{Probability}\{\text{service by BMCT}\}\cdot(\text{Mean roundtrip time from BN to FLOT})x_1 + \\ & + \text{Probability}\{\text{service by LMCT}\}\cdot(\text{Mean roundtrip time from LEMCO to FLOT})x_3. \end{aligned} \qquad (6)$$

These expected roundtrip times may be estimated either from data or by dividing the doubled distance by the average speed of the mobile contact unit. The probabilities of service by BMCT's or LMCT's are estimated from appropriate decision trees.

4. Estimation of Corrective Maintenance Time. The average time required for corrective maintenance is the Total Corrective Maintenance Time, TCM, divided by the mean number of failures: TCM/F. In equilibrium this is the same as the MTTR divided by the number of servers at any particular stage of the 4-level maintenance process:

$$E\{W_{CM}\} = \Sigma\ (MTTR_i)/s_i = \Sigma\ (MTTR_i)\ x_i. \qquad (7)$$

5. The Objective Function. Combining the results of Eqs. (4), (5), (6), and (7) leads to the following objective function for the maximization of A_0:

$$\begin{aligned} \min\ \{TALDT + TCM\} \approx \min\ \{ & [(MTTR_1)^2/MTBF + MTTR_1 + \\ & + \text{distance}_{BS\text{-}FLOT}/\text{kmph}_{BMCT}]x_1 + \\ & + \Sigma_{i=2,4}[(MTTR_i)^2/MTBF + MTTR_i]x_i + \\ & + [(MTTR_3)^2/MTBF + MTTR_3 + \text{distance}_{LS\text{-}FLOT}/\text{kmph}_{LMCT}]x_3. \end{aligned} \qquad (8)$$

5. The Constraints. It remains to formulate the constraints. We need for this purpose the following result from elementary algebra: If a_1/b_1, $a_2/b_2, \ldots a_n/b_n$ are unequal fractions all of whose denominators are of the same sign, then

$$\min(a_i/b_i) \le \Sigma a_j/\Sigma b_j \le \max(a_i/b_i), \qquad (i = 1,2,\ldots,n). \quad (9)$$

Relation (9) leads to the first of the constraints, the personnel constraint:

$$\Sigma_{i=1,2,3,4} x_i \ge (\text{number of stages})/(\text{total number of servers}). \quad (10)$$

The TALDTm constraint is given by (5); the transport constraint by (6), and the TCM constraint by (7). There is also a constraint imposed by limitation upon the number of vehicles available for assignment as MCT's.

6. A Numerical Example. The application led to the following linear programming problem (LPP):

$$\min\{1.354x_1 + 0.1648x_2 + 2.115x_3 + 0.466x_4\}$$

subject to

TALDTm: $0.04x_1 + 0.0098x_2 + 0.005x_3 + 0.066x_4 \ge 0.07$
TALDTt: $x_1 + 2x_3 \ge \min \text{TALDTt}/F = 1.24$
TCM: $0.314x_1 + 0.155x_2 + 0.11x_3 + 0.40x_4 \ge 0.5$
No. personnel: $x_1 + x_2 + x_3 + x_4 \ge 0.36$

Using STATGRAPHICS, the LPP solution was reached after 5 pivots. The objective function took the minimum value 1.81 hours. The solution vector for the x_i was

$$x = (0.00000,\ 0.27556,\ 0.62000,\ 0.97272).$$

Taking reciprocals, these values correspond to personnel assignments of

$$s_{BMCT} \sim \infty,\ s_{SS} = 4\ (\approx 3.6),\ s_{LMCT} = 2\ (\approx 1.6),\ \text{and}\ s_{LS} = 1,$$

which except for the BMCT value are eminently reasonable and do correspond to assignments that were actually made in practice. The continuous nature of the LPP solution is of course evident. Software for an integer programming solution was not available, but might have led to better results. The impossible value for s_{BMCT} presumably corresponds to the sensitivity of reciprocals to small values. However, such an extreme value is consistent with the intuition that in the short-term AirLand Battle, under a replacement rather than repair philosophy ("design for discard"), the major part of maintenance should be concentrated in mobile units closest to where needed. Error estimates for this aspect of

the problem remain open. There is also the matter of the problem being ill-scaled in that the coefficients of the objective function differ by as much as two orders of magnitude from those in the TALDTm constraint, for instance. This feature is known on occasion to make computer solution of LPP's impossible.

REFERENCE

Gaver, D. P. & G. L. Thompson (1973). Programming and Probability Models in Operations Research. Monterey, California: Brooks/Cole Publishing Co.

The Making and Use of the "BIG MAC" Data Base

Fred M. Grimes
John Riemenschneider
Combat Arms Test Directorate
TEXCOM Combined Arms Test Center
Fort Hood, Texas 76544-5065

I. Introduction

The ultimate test of new Army equipment is the field test by user troops. Engineering tests can measure certain aspects of the equipment's performance, however, it is only in the field that the Army finds out about the true value of the equipment. But getting meaningful test data from the field is greatly more difficult than getting data from engineering tests. Engineering tests can be closely controlled so that there is flexibility in collecting test data by manual and instrumented means. During field tests manual and instrumented data can only be collected if it does not interfere with the troops using the equipment. Further the troops must be using the equipment in a certified operational manner. This constraint inhibits collecting all desired data. This constraint becomes a big time problem when test data is required from force-on-force engagements such as was encountered in the Combat Vehicle - Combat Performance Operational Assessment (CV-CPOA) which the Combat Arms Test Directorate (CATD) was tasked to perform in the spring of 1987.

Although the mission was simple:

"Gather empirical data on the survivability of the Bradley Fighting Vehicle (BFV) during simulated combined arms force-on-force combat operation against a realistically configured threat".

The execution was difficult.

This paper discusses this operational test and how the vital test data was collected and reduced to a large computer data base which contained many data elements. The computer data base is called the "BIG MAC".

The Combat Arms Test Director who personally directed this test had the foresight to assign an experienced test officer to gather video and film data of all the events occurring during the execution of the test so that a composite video film could summarize these events. A seventeen minute video resulted from this effort and is available for official use. The video gives an excellent overview of the test.

Preceding Page Blank

II. Background

The test was specifically designed to collect data on the survivability of the currently configured BFV and three proposed modifications.

The CV-CPOA test design methodology was driven by the evaluation concept in the Operational Test and Evaluation Agency (OTEA) Independent Evaluation Plan. The Congressional defense authorization bill mandating the BFV assessment specified live-fire and operational comparison testing of the Army version of an enhanced survivability BFV (M2A1 and M3A1 HS) and the Department of Defense version of an enhanced BFV (the ASTB). The Army's assessment of Congressional intent and implied taskings revealed the need for a data base that contained those elements of survivability data that would facilitate a comparison of proposed enhancements with a baseline using the basic BFV. An analysis of all factors that contribute to or degrade the survivability of a combat vehicle resulted in the following key data elements being identified by CATD and OTEA for inclusion in a data base:

a. Distribution of direct fire aiming points and engagement angles.

b. Exposure of target.

c. Motion of firer and target.

d. Frequency of indirect fire engagements.

e. Ammunition and personnel on board at time of engagement.

f. Suppressive effects of firing port weapons.

g. Target acquisition and engagement frequencies.

h. Distribution of engagement ranges.

Research conducted by OTEA and CATD revealed a dearth of such survivability data on the BFV. Available data included recent live-fire studies that had been criticized by the Department of Defense, the General Accounting Office, and Congress as being unrealistic and biased. Other data existed that was based upon modeling analyses that did not have credibility with the same audience. Virtually no empirical operational data existed. Even though the Congressional bill directed only a comparison test of the two BFV versions, the absence of empirical survivability data on the basic BFV led OTEA and CATD to design a test which would provide such baseline data. This data had to be the product of battalion-sized combined arms forces operating in a realistic, simulated combat environment. The combat operations had to be representative of those described in AirLand Battle doctrine, and the OPFOR had to be structured and trained to represent the projected threat circa 1987-92 in a European SCORES V scenario

setting. Phase I trials were based on this concept. The principal data output required of phase I was the actual distribution of all direct fire engagement angles and aiming points.

In order to respond specifically to Congressional inquiries, direct comparative testing of the two BFV versions was necessary. The optimal test design would have had the two enhanced versions of the BFV used in the same type operations with the same size forces and weapon types as those in phase I. However, this was infeasible due to nonavailability of prototypes and surrogates, instrumentation limitations, and time constraints. Consequently, phase II was designed to be a scaled down examination of a Blue platoon-sized force using the vehicles to conduct offensive missions in a simulated combat environment. This scaled down examination included comparison of the basic BFV, the HS vehicle, and the ASTB vehicle, all with 500-horsepower engines. The Chief of Staff, Army, later directed the inclusion of an ASTB version equipped with a 600-horsepower engine. Phase II trials were based on this concept. As with phase I, the primary data elements of interest were engagement angles and aiming points.

OTEA's evaluation concept was to examine the comparative performance data produced in the phase II platoon trials and scale up to the baseline data of the phase I battalion-sized trials. The data from phases I and II was used as inputs into the US Army Materiel Systems Analysis Activity (AMSAA) and the Ballistic Research Laboratory survivability models to lend further insights into the BFV survivability. The correlation of the phase II data with the phase I baseline data and the outputs of the models and other testing will allow the Army to predict the performance of the various types of BFV's with respect to their survivability enhancements.

Although Congressional interest focused on the data pertaining to the survivability of the BFV (M2 and M3), OTEA, anticipating the need for similar data relative to other combat vehicles, expanded the scope of the CV-CFOA to include concurrent collection of similar data on the M1 tank, the improved TOW vehicle (ITV), and the M113.

III. The making of the "BIG MAC"

The required data was collected by either instrumented systems or test personnel.

a. Instrumented data and collection systems. Described below, in general terms, are the instrumented systems used to simulate tactical engagements and to collect and record data.

(1) Video system. The video system served as the principal means of data collection. Video tape was the data

source for aiming point, engagement angle, target exposure, firer-target motion, target acquisition, and player identification. Each combat vehicle and hand-held HEAT weapon was equipped with a coaxially mounted video camera and recorder. Approximately two-thirds of the Red force vehicles also had a through-the-sight video camera and recorder installed which recorded what the gunner saw. At the beginning and end of each trial, an image of a boresight panel was recorded. Both the dimensions and the distance to the panel were known. The internal clock of the video system was synchronized with Inter-Range Instrumentation Group-Format B (IRIG-B) time, and the rounds counter was set to zero. At the end of each trial, the video tapes were removed and annotated with the player identification number, weapons systems type, and battle trial number.

(2) Flash box system. The flash box system was used with the video system to provide an additional source for the data required to determine the range between the firer and the target. The flash box system is based on an infrared strobe which flashes once every 3 seconds. The time of the infrared flash is determined from a computer-driven cycle and electronically relayed from the central computer to an infrared strobe located on each combat vehicle. The flash of the target's infrared strobe, which is invisible to the eye, was recorded by video systems on the firing vehicles. The timing of the strobe flash is a 3-second cycle and is unique for each vehicle. This provides a means of correlating the time of the strobe flash with a player identification.

(3) Multiple Integrated Laser Engagement System (MILES)

(a) The need existed for an Real Time Casualty Assessment (RTCA) system to serve as a tactical scenario driver by simulating direct fire weapons engagements and their effects. Emphasis was placed on causing crews and gunners to execute proper gunnery procedures under the most realistic conditions possible. Only two RTCA systems were available--the TCATA Automated Field Instrumentation System (TAFIS) and MILES. MILES was selected as the principal component of the RTCA system because it met the following critical test design considerations:

1. MILES integrates the RTCA play of vehicular weapons systems and individual soldiers.

2. MILES allows the play of fully mobile, hand-held HEAT weapons.

3. MILES requires gunners to aim at the center of mass of a target in order to achieve a hit--a characteristic prerequisite to collection of meaningful data on aiming point distribution.

4. MILES is the standard Army-wide training simulation system which is readily accepted by US soldiers and can be employed with minimal artificiality and no specific training.

(b) The inherent limitations of MILES, such as fixed time of flight for missiles and friendly-only lethality codes, were considered tolerable since the test design did not require the collection of force exchange ratios or casualty counts as data outputs and RTCA system was used only as an exercise driver.

(4) I-MILES. To offset some of the inherent shortcomings of MILES and to provide redundant data collection capabilities, an enhancement to MILES was procured specifically for this test. I-MILES was used as an addition to MILES and was transparent to the soldier and to the basic MILES. It permitted programming of vulnerability and lethality logic into the host MILES, thus enabling the use of specific Pk's, effective ranges, and basic loads for all US and threat weapons systems. Vulnerability was varied by tactical mission exposure. For example, Pk for the BFV fully exposed in the offense differed from the Pk for the BFV in hull defilade in the defense. I-MILES stored all engagement data and provided a time-correlated history of all MILES events. It also provided a method of obtaining the intermediate data necessary to determine the range between the firer and the target.

(5) MILES Laser Detector Decoder System (MLDDS). The MLDDS was used to collect data on the frequency of engagement by all weapons against the BFV in specific quadrants (front, right, left, rear). In addition to the MILES sensor belts installed on all combat vehicles, separate MLDDS sensor belts were installed on the BFV's of one Blue force company in phase I and on the IFV's and HS variants used in phase II. When a MILES laser struck a sensor, the laser message and the firing weapon were recorded and a sequential history of engagements was established. MLDDS recorded small arms engagements. This data was necessary for assessing the effectiveness of reactive tiles.

(6) Position Reporting and Recording System (PRRS). The PRRS was used to collect the position location of each combat vehicle at the time of engagement. Each combat vehicle was equipped with a mobile unit which transmitted a signal at a prescribed time within each second. The signal was received by a tower array and relayed to PRRS central where computers calculated the combat vehicle location in grid coordinates. The grid location and time were recorded on magnetic tape. Recording of position location was necessary as an intermediate step in determining the range between firer and target for each direct fire engagement.

(7) Scanning laser system. During phase II, the scanning laser system was used to determine when line of sight

existed between the defending Red force vehicles and attacking Blue force BFV's. In each medium and long range trial, a scanning laser was installed adjacent to each Red force vehicle. (For safety reasons, scanning lasers were not used in short range trials.) The laser transmitter was mounted on a tripod in such a manner that the transmitter was at the eye level of the track commander when the combat vehicle was in defilade. Line of sight was established when the transmitter laser beam struck the laser detector array system installed on the Blue force vehicle. The laser signal identification, player identification of the Blue force vehicle, and the time were transmitted by the Blue force vehicle to a data collection system where they were stored on a magnetic tape.

b. Manually collected data.

(1) Prior to each trial a player roster was completed for all participants. This form contained the trial identification, unique player identification, and the instrumentation on each player.

(2) The time each trial started and ended and the reason for the ending were recorded. The type of terrain and the weather conditions for each trial and the times that obscurant were used were also recorded.

(3) At the beginning and end of each trial, a controller recorded odometer and engine hour readings for each combat vehicle assigned to the trial. The controller also recorded the number of troops on each IFV, CFV, HS variant, and ASTB surrogate at the start of the trial. During trial execution, the controller noted the times when troops mounted and dismounted. Controllers also reported any administrative situations which might have invalidated a trial.

(4) Indirect fire was played only during phase I trials. Manually collected data on indirect fire engagements included the time of artillery fire: the type of fuse, ammunition, and weapon; the size and number of volleys; which artillery unit should fire; size and location of impact area; and the identification number of the vehicles in the impact area.

(5) Each player completed a demographic questionnaire.

(6) At the end of each trial, each combat vehicle turret crewman and BFV squad member completed questionnaires on the perceived effectiveness of their weapons to include use of the firing port weapons against dismounted troops. The Red force dismounted troops also completed a questionnaire on the perceived effectiveness of Blue force suppressive fire.

(7) At the end of each trial, the subject matter experts

evaluated specific aspects of the trial--maneuver, fire support, intelligence, engineer, communications, logistics, leadership, training, and use of the combat vehicle. They prepared a narrative summary of the tactics and performance of each force for each trial. All record trials were certified by the senior controller (senior subject matter expert) as being doctrinally sound.

(8) During phase II, three Blue force platoons rotated among the HS variants and the ASTB surrogates. Each crew conducted six consecutive trials in each vehicle. At the conclusion of the sixth trail, each crew member completed a questionnaire on the differences in combat performance of these vehicles. The Red force players also completed a questionnaire on the performance of the surrogates.

Data collected by instrumentation was compiled by the test support contractor and test team and forwarded to the data reduction center. Data collected manually was edited at the test site and then forwarded to the data reduction center.

a. The video tape and I-MILES record for a specific trial were issued to a data reducer. When through-the-sight and coaxial video tapes existed for a single vehicle, both tapes were issued to two data reducers for simultaneous processing. The video data reducer reviewed the video tape to determine when the firing system engaged a target. This time of engagement is referred to as a trigger-pull. A burst of 10 rounds from a 25- or 30-millimeter cannon constituted a single engagement. The time of the trigger-pull was indicated by the rounds counter incrementing and by the IRIG-B time on the I-MILES record.

b. The data reducer used a specially fabricated, three-dimensional scale model to determine engagement angles and aiming points for the M1, M2, and M113 type vehicles. The scale model of each combat vehicle is mounted on a pivot so that the model can be rotated through 360° in a horizontal plane. The reducer oriented the model so that it matched the picture shown on the video screen at the time of the trigger-pull. The horizontal engagement angle between the firer and the target was read from the scale and entered on the data reduction form. Discrete angles were recorded for the hull and the turret orientation. A pivoting bar attached to the base of the device and oriented to a vertical scale was used to determine the vertical engagement angle. Front, rear, side, and top view pictures of each combat vehicle with 4-inch grid squares superimposed over the pictures were used to specify the location of the aiming point. From these views the one that matched the video picture was selected and the grid coordinates were recorded.

c. The data reducer determined the target vehicle identification by noting the flash box counter number at the

moment the target vehicle's strobe light flashed and correlating that number to the vehicle identification number. When the pairing report from the I-MILES was available, it was used to validate target identification or establish it when the flash box did not work.

d. Range between the firing and target vehicles at the time of the trigger-pull was estimated by the data reducers using a stadia overlay. The video data reducer positioned the stadia overlay on the video tape and counted the number of stadia lines for the size of the boresight panel. This established a benchmark. At the time of the trigger-pull, the data reducer placed the stadia source over the video picture of the target and counted the number of stadia lines for the target size. A conversion table was used to determine the range based on the number of stadia lines. This stadia range was entered on the data reduction form.

e. Range data was calculated by a computer program which used the PRRS-developed position locations of the firer and the target at the time of the trigger-pull. This method required positive firer and target identifications that were provided by the video tape and/or the I-MILES printouts.

f. If the PRRS and stadia ranges differed by more than 700 meters, a quality review group examined both ranges to assess their accuracy. When the PRRS data was questionable for either the firing or target vehicle, like data on other vehicles of the same platoon was evaluated to verify the range. If the PRRS range could not be verified, the stadia range was used. In those cases where neither the stadia nor the PRRS range could be determined, senior data management assistants viewed the video and estimated the ranges at time of the trigger-pull.

g. Intervisibility data was derived from a computer program that correlated the positions of the players at the time a Blue force vehicle received scanning laser energy. A computer program calculated the duration of the target vehicle exposure and the distance the target vehicle traveled while exposed. The computer program provided an output data set used in the production of the phase II data base.

h. Data collection forms and questionnaires were edited by quality assurance personnel and entered into a computer data file. The data file was structured by player identification, trial number, and time of data event.

IV. "BIG MAC" structure

The data base is composed of two distinct files. The primary file (battle history) is an integrated, chronological record of all the activities occurring during the execution of a trial.

This is called the "BIG MAC". The other file (descriptive data) is an integrated record of the conditions surrounding the execution of the trial and the subjective data collected during the trial.

a. "BIG MAC" file. The primary data file was constructed during an iterative process of data reduction, edit, and quality review.

(1) The initial format was the file of reduced data from the I-MILES and from the video and flash box systems. It reflected the direct fire battle.

(2) When the initial file reached data maturity, a computer program correlated the position location of each firer-target pair. This data resulted in range being added to the file.

(3) Concurrent with the range process, a computer program examined each engagement and calculated ammunition usage for each firer. The same program also determined if squad members were on board when the target was a BFV.

(4) When the above steps were completed, the data was merged with the artillery data and a time-correlated file was output that contained data on all direct and indirect fires.

(5) The integrated firing history was then merged with the data file containing the trial start and finish data on each vehicle. This process continued with a merge program that incorporated the exposure data calculated by the laser system.

(6) The end result was a battle history file for all trials.

b. Descriptive data file. This file contains the information on the conditions of each trial and the subjective data collected. It is organized as subordinate records, each of which is sorted by trial and time.

The mission of the test team was to produce a data base on operational survivability in a timely and accurate manner. The independent evaluator, OTEA, is responsible for data base analysis and determining the specific answers to questions regarding BFV survivability. A copy of the data base was forwarded to OTEA in magnetic tape format. The data base was derived from approximately 6,000 hours of video tape and 1,000 collection forms. The data base contains approximately 26,000 direct fire engagements, 1,200 indirect fire engagements, 3,000 intervisibility segments, and 19,000 descriptive events. The

total data base contains approximately 2 million data elements. The data presented clearly demonstrates the capability of the data base to support analysis of the BFV survivability question.

V. The use of the "BIG MAC"

The basic purpose of the CV-CPOA operational test was to provide realistic force-on-force test data to OTEA. They would use the data in their evaluations to predict the operationally survivability of the BFV and other Army combat vehicles. OTEA provided "BIG MAC" data to the Ballistic Research Laboratory for the use in their models to calculate survivability results. OTEA completed its evaluation and reported its finding to Congress. The findings led to some design changes to the BFV. The "BIG MAC" had served its purpose but the "BIG MAC" is more than a data base with a one time use. The data base is a rich source of force-on-force test data. One can play some mind boggling games with the number of ways the data can be sliced and diced. Due to the large number of different data elements, reports can be prepared with any number of data sets. For example, the following type of reports can be prepared:

a. Direct fire weapons - Frequency of engagements and engagement angles

b. Direct fire weapons - Aiming point distributions

c. Direct fire weapons - Engagement range distributions

d. Ammunition and personnel on board BFV at time of engagement

e. Target exposure and motion of firer and target

f. Indirect fire weapons

g. Combat vehicle intervisibility and engagements

h. Direct fire weapons distribution

In February 1988, the Department of Defense Office of Test and Evaluation (DOTE) asked that an extended evaluation of the CV-CPOA data be made regarding a host of questions asked by DOTE. The questions concern such things as, effect of direct and indirect fire on both the red and blue target vehicles, effects of line of sight on ranges of engagements, how was the ITV engaged on the battlefield, order effects on engagement ranges as the trial progressed, etc. The Deputy under Secretary of the Army (Operations Research) tasked OTEA to respond to these questions and to make the "BIG MAC" available to others. OTEA responded by forming a Study Advisory Group of Army organizations which would have an interest in the data. OTEA also obtained

contractor support to prepare a data dictionary, data base documentation for the "BIG MAC" and also provide assistance in responding to the DOTE questions.

Several DAG members have used data from the "BIG MAC" to assist them in their work. AMSAA compared "BIG MAC" hit distribution data with the currently used cardioid distribution data for tanks and the BFV. The comparison showed that the "BIG MAC" data contained more "Head-on" shots. As a result, the Army has decided to use the "BIG MAC" hit distribution data for future AMSAA modeling work for tanks and BFV. The Infantry School has used the "BIG MAC" data in some of their studies on TOW missile firings.

Upon completion of the contractor's work on the data dictionary and data base documentation, the "BIG MAC" will be available to all Army organizations for official use and can provide realistic force-on-force data to assist them in their work.

MODEL SENSITIVITY IN STRESS-STRENGTH RELIABILITY COMPUTATIONS

Donald M. Neal, William T. Matthews, and Mark G. Vangel

U.S. Army Materials Technology Laboratory
Watertown, Massachusetts 02172-0001

ABSTRACT

There has been a recent interest in determining high statistical reliability in risk assessment of aircraft components. This paper identifies the potential consequences of incorrectly assuming a particular statistical distribution for stress or strength data used in obtaining the high reliability values. The reliability is defined as the probability of the strength being greater than the stress over the range of stress values. This method is often referred to as the stress-strength model.

A sensitivity analysis was performed involving a comparison of reliability results in order to evaluate the effects of assuming specific statistical distributions. Both known population distributions and those that differed slightly from the known were considered. Results show substantial differences in reliability estimates even for almost non-detectable differences in the assumed distributions. These differences represent a potential problem in using the stress-strength model for high reliability computations, since in practice it is impossible to ever know the exact (population) distribution.

An alternative reliability computation procedure is examined involving determination of a lower bound on the reliability values using extreme value distributions. This procedure reduces the possibility of obtaining non-conservative reliability estimates. Results indicated the method can provide conservative bounds when computing high reliability.

Preceding Page Blank

CONTENTS

	Page
INTRODUCTION	1
STRESS-STRENGTH MODEL	2
PROBABILITY DENSITY FUNCTIONS	3
METHODS FOR COMPUTING RELIABILITY R	5
Numerical Integration	6
R Computation from Closed Form Solution	6
Nonparametric Method	7
CONTAMINATED PROBABILITY DENSITY FUNCTIONS	7
LOWER RELIABILITY BOUND	9
GOODNESS-OF-FIT TEST	10
RESULTS AND DISCUSSIONS	
Variance Contamination	10
Location Contamination	14
CONCLUSIONS	19
ACKNOWLEDGMENT	19

INTRODUCTION

There has been an interest in quantitative reliability-based structural design for many years. An early example is the structural reliability development by Freudenthal.[1] Stress-strength reliability computations are a principal consideration in structural reliability design. Reliability methods have been considered for many structural applications including: civil engineering,[2] nuclear reactors,[3] fixed wing aircraft,[4] rotorcraft,[5] and space vehicle propulsion systems.[6] Very high structural reliability is expected to be achieved for most applications. A reliability goal of $0.9_{(9)}$ per flight hour was suggested in 1955 by Lundberg[4] for fixed wing civil aircraft. Recently, Lincoln,[7,8] using reasoning similar to that of Lundberg, cited a reliability goal of $0.9_{(7)}$ per flight for fixed wing military aircraft. The U.S. Army has instituted a new structural fatigue integrity criterion for rotorcraft which has been interpreted[5] as a requirement for a lifetime reliability of $0.9_{(6)}$.

The use of advanced materials whose structural properties are best characterized on a statistical basis appears to be a stimulant for increased interest in statistical-based structural design for airborne structures.

A significant feature associated with predictions of structural reliability is that the consequence of a failure event may be more than reduced system performance or the inconvenience of a system being out of service; structural failure can be catastrophic in terms of loss of life and property. In this context it is imperative to evaluate the sensitivity of structural reliability predictions to uncertainties. It appears that this issue has received little attention except for a brief note by Harris and Soms[9] and a recent presentation by Berens.[10]

There are many issues to be faced in obtaining quantitative structural reliability predictions. Such issues include system complexity (many components, multiple failure modes in each component, and interdependence of component behavior), sample or data set size associated with structural loading spectrum conditions and with mechanical properties, and the basis for characterizing structural qualification tests (the number of duplicate specimens and methods for compensation for untested effects such as the effect of environment).

In addition, when predictions of structural behavior are required in the high reliability range, since sufficiently large data sets are usually not available, it is necessary to use parametric modeling methods. Assumed parametric functions permit extrapolation from available data to determine the probability of failure. Since the probability of failure is extremely small, this will always involve substantial extrapolation from what can be observed experimentally. The estimated reliability will therefore depend strongly on the assumed parametric probability density function (PDF). Slight deviation from the assumed model in tail regions can have dramatic effect on high reliability estimates.

1. FREUDENTHAL, A. M. *The Safety of Structures.* Trans. ASCE, v. 112, 1947.
2. CORNELL, C. A. *A Probability-Based Structural Code.* J. Am. Conc. Inst., v. 66, 1969, p. 974-985.
3. U.S. Nuclear Regulatory Commission. *Reactor Safety Study: An Assessment of Accident Risks in U.S. Nuclear Power Plants.* NRC Report WASH-1400, 1975.
4. LUNDBERG, B. *Fatigue Life of Airplane Structures.* J. of the Aeronautical Sciences, v. 22, no. 6, June 1955, p. 349.
5. ARDEN, R. W., and IMMEN, F. H. *U.S. Army Requirements for Fatigue Integrity.* Proceedings of American Helicopter Society National Technical Specialists Meeting on Advanced Rotorcraft Structures, Williamsburg, VA, October 1988.
6. SHIAO, M. C., and CHAMIS, C. C. *Probability of Failure and Risk Assessment of Propulsion Structural Components.* NASA Technical Memorandum 102323, 1989.
7. LINCOLN, J. W. *Risk Assessment for an Aging Military Aircraft.* J. of Aircraft, v. 22, no. 8, August 1985, p. 687.
8. CORNOG, D. O., and LINCOLN, J. W. *Risk Assessment of the F-16 Wing.* Proceedings of the 1988 Structural Integrity Conference, San Antonio, TX, WRDC-TR-89-8071, Wright-Patterson AFB, Ohio, 1989.
9. HARRIS, B., and SOMS, A. P. *A Note on the Difficulty Inherent in Estimating Reliability from Stress-Strength Relationships.* MRC-2123, Mathematics Research Center, U. of Wisconsin, AD-A073637, October 1980.
10. BERENS, A. *Structural Risk Analysis in Aging Aircraft Fleets.* Proceedings of the 1988 Structural Integrity Conference, San Antonio, TX, WRDC-TR-89-8071, Wright-Patterson AFB, Ohio, 1989.

In fact, one might argue, as does Freudenthal,[11] that because of the extrapolation involved, statistically-based high reliability calculations for complex systems must always be suspect:

"When dealing with probabilities a clear distinction should be made between conditions arising in design of inexpensive mass products in which the probability figures are derived by statistical interpretation of actual observations or measurements (since a sufficiently large number of observations are actually obtainable), and conditions arising in design of structures or complex systems. In the latter, probability figures are used simply as a scale or measure of reliability that permits the comparison of alternative designs. The figures can never be checked by observations or measurement since they are obtained by extrapolations so far beyond any possible range of observation that such extrapolation can no longer be based on statistical arguments but could only be justified by relevant physical reasoning. Under these conditions the absolute probability figures have no real significance ...."

Nonparametric stress-strength procedures do not require specific parametric assumptions, and so it might be hoped that such procedures could circumvent this difficulty. However, Johnson[12] has noted that "The nonparametric approach has one serious drawback. In return for its distribution free property, it is not possible to establish high reliability even with moderate sample sizes." With respect to the use of parametric models, Box[13] has observed "all models are wrong, but some are useful," meaning that no parametric statistical model should be accepted uncritically. Whenever a model is used, it is the obligation of the analyst to investigate the consequences of departures from an assumed model which, though small, are consistent with available data. Harris and Soms[9] has illustrated a "serious problem in the use of stress-strength relationships in estimating reliability." In particular, "stress-strength models in reliability theory are highly sensitive to small perturbations in extreme tails." The perturbations considered may arise from an alternative mode of failure such as the presence of a flaw in a structure. Further, they note that the problem cannot be eliminated unless "astronomically large sample sizes are employed."

In the following, the examination of the sensitivity of structural reliability estimates focuses attention on one of the previously cited issues: the selection of a parametric PDF. The examination of the sensitivity of stress-strength reliability estimates is extended to addition perturbation effects. The sensitivity of reliability estimates to the selection of parametric models is considered with emphasis on graphical representations. The results are evaluated with regard to the usefulness of parametric stress-strength models for application to the high reliability regime of $0.9_{(6)}$ to $0.9_{(7)}$ when the consequence of failure may be catastrophic. An alternative reliability computation procedure is examined involving determination of a lower bound on reliability which can be obtained independently of the assumed PDFs.

STRESS-STRENGTH MODEL

The statistical reliability as referred to in this report is determined in the following manner. Shown in Figure 1 is the stress-strength model where $f_2(s)$ and $f_1(S)$ represent the PDFs for the applied stress s and material strength S.

11. FREUDENTHAL, A. M. *Fatigue Sensitivity and Reliability of Mechanical Systems, Especially Aircraft Structures.* WADD Technical Report 61-53, Wright-Patterson AFB, Ohio, 1961.
12. JOHNSON, R. A. *Stress-Strength Models for Reliability.* Handbook for Statistics, Elsevier Science Publishers, New York, v. 7, 1988, p. 27.
13. BOX, G. E. P. *Robustness in the Strategy of Scientific Model Building.* Robustness in Statistics, R. L. Launer, and G. N. Wilkinson, eds., Academic Press, Inc., New York, 1979.

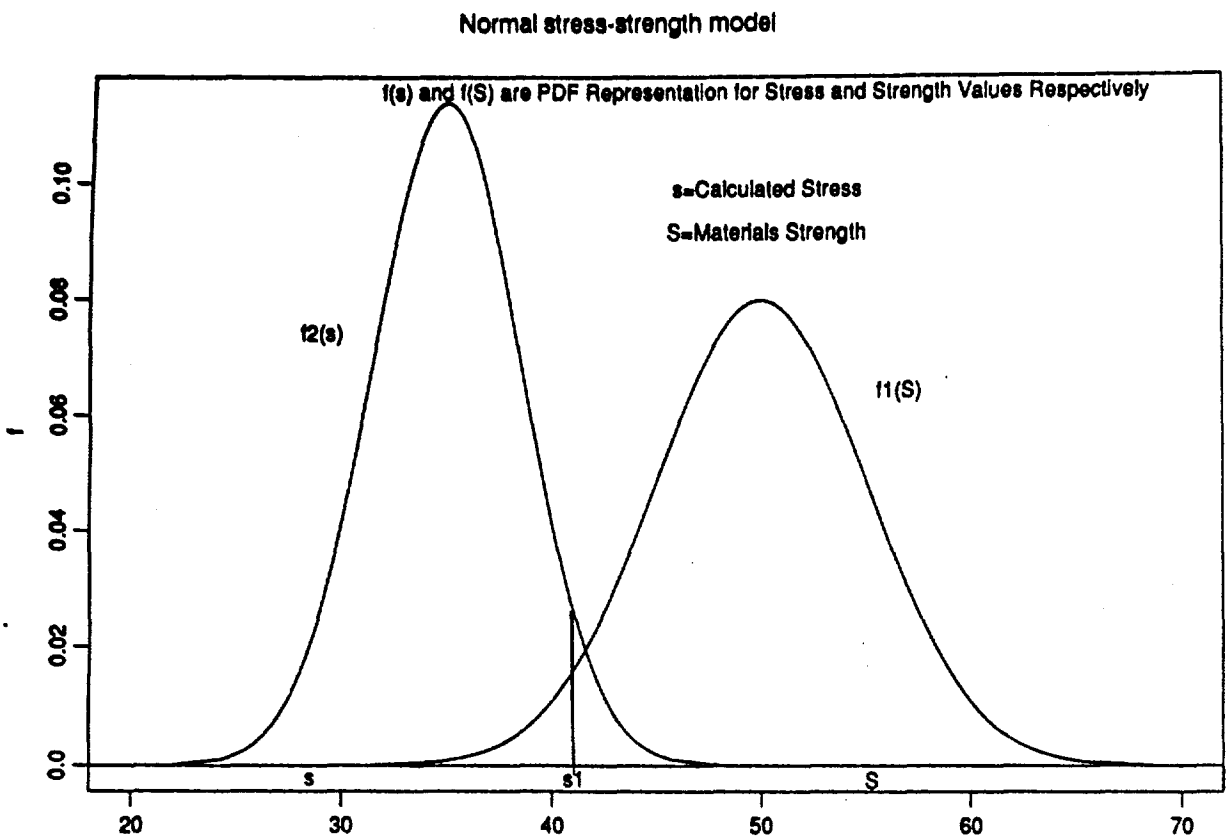


Figure 1. Normal-normal stress-strength model.

Since the joint probability dR for the strength being greater than s_1 can be written as,

$$dR = f_2(s_1)\, ds \int_{s_1}^{\infty} f_1(S)\, dS \tag{1}$$

then the reliability for all s values is

$$R = \int_{-\infty}^{\infty} f_2(s) \left[\int_{s}^{\infty} f_1(S)\, dS \right] ds\,. \tag{2}$$

PROBABILITY DENSITY FUNCTIONS

A wide variety of PDFs may be applied in obtaining R values. Some examples of PDFs are as follows:

The PDF most often used in stress-strength models is the <u>normal distribution</u> (see Figure 1),

$$f_N(S) = N(S, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{S-\mu}{\sigma}\right)^2 \right\}, \tag{3}$$

where $-\infty < S < \infty$, $\mu > 0$, and $\sigma > 0$. The mean of the population is μ, and the standard deviation σ for this model.

A model which is more easily justified on physical grounds is the Weibull PDF,

$$f_W(S) = \frac{\beta}{\alpha}\left(\frac{S}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{S}{\alpha}\right)^{\beta}\right], \tag{4}$$

where $S > 0$, $\alpha > 0$, and $\beta > 0$. Despite the relevance of the Weibull distribution[14] to the strength of brittle materials, it is not often used, possibly because it is more difficult computationally to obtain reliability values than with the normal model.

If S follows the Weibull PDF, then ln(S) will have an extreme value distribution with PDF

$$f_{min}(S) = \frac{1}{b_1}\exp\left[\frac{S-u_1}{b_1} - \exp\left(\frac{S-u_1}{b_1}\right)\right]. \tag{5}$$

The distribution of $-\ln(s)$ is

$$f_{max}(s) = \frac{1}{b_2}\exp\left[-\left(\frac{s-u_2}{b_2}\right) - \exp\left(-\left(\frac{s-u_2}{b_2}\right)\right)\right]. \tag{6}$$

Both of the above formulas are referred to as extreme value distributions. The use of extreme value distributions in a stress-strength model is illustrated in Figure 2. The extreme value distribution parameters are related to the Weibull parameters as follows:

$$b = \frac{1}{\beta} \text{ and } u = -\log\alpha \quad .$$

In order to obtain the population Weibull shape and scale parameters β and α from the known population mean μ and standard deviation σ, the following approximations are suggested:

$$\beta = 1.27\,\mu/\sigma - 0.56 \tag{7}$$

and

$$\alpha = \mu/\Gamma\left(\frac{1}{\beta}+1\right) \quad .$$

The functions defined in Equations 3, 4, 5, and 6 clearly have different shapes and they exhibit dramatically different tail behavior. Since reliability estimates depend strongly on the extreme upper tail of the stress PDF and the extreme lower tail of the strength PDF, the choice of model will typically have a substantial effect on the reliability estimate. For example, R is usually higher when calculated from the normal distribution than when the extreme value model is assumed.

Applying PDFs that are capable of obtaining accurate high reliability estimates (e.g., $0.9_{(6)}$) requires prior knowledge of the functional form of the population PDF in addition to the

14. BURY, K. V. *Statistical Models in Applied Science*. John Wiley and Sons, Inc. (New York, London, Sydney, Toronto), 1975.

availability of large data sets (e.g., 1,000 replicate specimens). For lower reliability values (e.g., 0.9), a goodness-of-fit test for PDF identification with a moderate amount of data is generally adequate. The consequence of incorrect PDF selection and limited sample sizes are discussed later in this report.

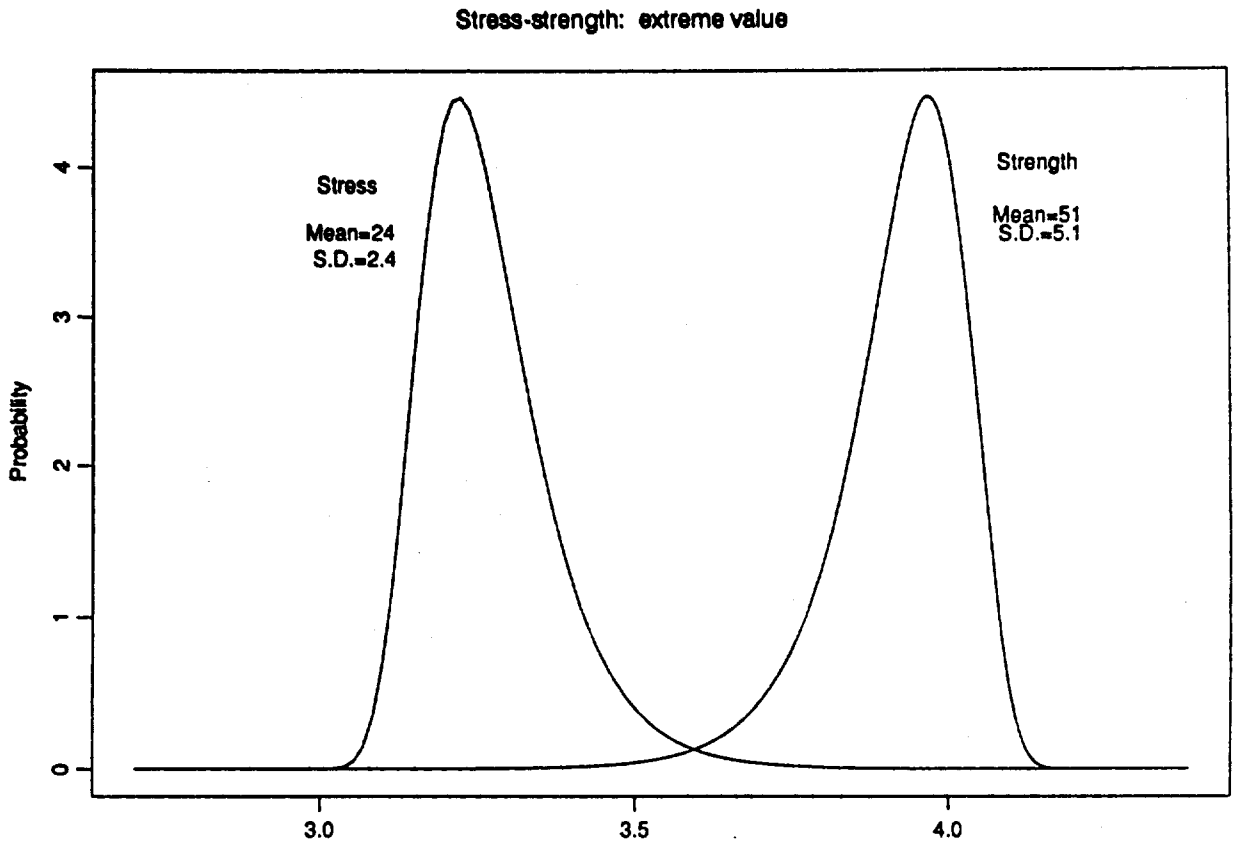


Figure 2. Stress-strength extreme value functions.

METHODS FOR COMPUTING RELIABILITY R

In determining R from Equation 2 it should be noted that the integration process does not determine an area. The area A described by the intersecting functions in Figure 3 does not represent a 1 - R failure probability. The area A is the probability (P) that either $S < T$ or $s > T$, that is,

$$A = P(S < T) + P(s > T) , \tag{8}$$

where T is the point of intersection of the two functions. The area A is obviously not the same as the 1 - R from Equation 2 which determines $P(S > s)$ jointly with $P(s)$.

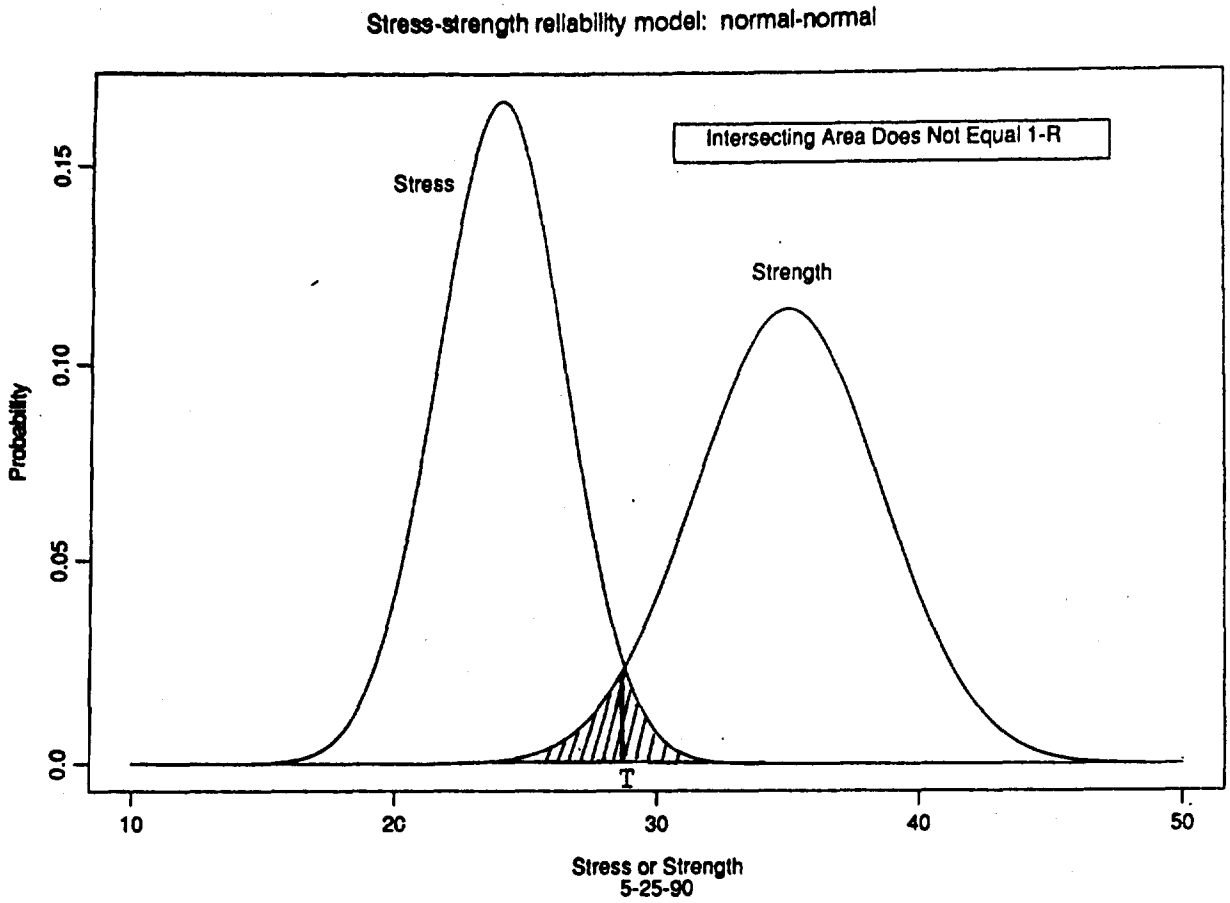


Figure 3. Stress-strength incorrect unreliability region.

Numerical Integration

Numerical integration procedures are usually suggested if a closed form solution of Equation 2 is not available. The numerical integration process involves repeated application of a method such as Simpson's Rule. The inner integral in Equation 2 is evaluated numerically for each ordered s_i value $i = 1, ..., n$ resulting in an $I_1(i)$ array of values. Each of $I_1(i)$ is then multiplied by the corresponding $f_2(s_i)$ forming another array $I_2(i) = f_2(s_i)I_1(i)$. R is obtained from I_2 array by reapplying the numerical integration method. This process will usually provide accurate results for $51 \leq n \leq 101$, where n is the number of mesh points in the integration process. Simulation results showed that the limits of integration can be obtained from ± six standard deviations from the mean.

Closed form solutions are available when the assumed stress and strength PDFs are both normal or both Weibull.

R Computation from Closed Form Solution

If both stress and strength data can be represented by normal PDFs $N(\mu_s, \sigma_s^2)$ and $N(\mu_S, \sigma_S^2)$, respectively, then,

$$R = P(S > s) = \Phi\left(\frac{\mu_S - \mu_s}{\sqrt{\sigma_S^2 + \sigma_s^2}}\right), \tag{9}$$

where (Φ) is the standard N(0, 1) normal cumulative distribution function, μ_s and μ_S are means, and σ_s^2 and σ_S^2 are variances of the stress and strength, respectively.

If both f_1 and f_2 in Equation 2 are Weibull with different scale parameters α_1 and α_2, but with a common shape parameter β, then the integration indicated in Equation 2 gives the following closed form expression[12]

$$R = \frac{1}{1 + \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta}} . \tag{10}$$

The common shape parameter means that both the stress and strength are skewed in the same way, which is a serious limitation. It is much more reasonable to have a stress distribution with a heavy upper tail and a strength distribution with a heavy lower tail, but this is not possible unless the shape parameters can be varied separately.

Nonparametric Method

This method does not assume a PDF for either stress or strength data. It determines reliability from the ordered array of m stress (s) and n strength (S) values, where each of the S values are compared with all s values. R is the proportion of times S > s for the total number of comparisons, that is

$$R = \frac{1}{m\,n} \sum_{j=1}^{m} \sum_{i=1}^{n} \alpha_i, \text{ where } \alpha_i = \begin{cases} 1, & S_i > s_j \\ 0, & s_j \geq S_i \end{cases} . \tag{11}$$

This method is not useful for obtaining high reliability even for relatively large data sets. It is obvious from Equation 11 that for high reliability calculations, mn must be very large; for example, 10^6 would be required in order to obtain R of $0.9_{(6)}$.

The Weibull, normal, or other parametric PDFs can provide estimates of high R values because of their ability to extrapolate beyond the available empirical data. Unfortunately, the amount of extrapolation dependency determines the magnitude of relative error in R.

CONTAMINATED PROBABILITY DENSITY FUNCTIONS

In order to illustrate the sensitivity of high reliability calculations to small deviations from assumed models, we will take the following approach. Consider the situation where with a high probability of $1 - \varepsilon$, specimens are obtained from a primary PDF, while with probability ε, specimens come from a secondary PDF. This probability model is referred to as a contaminated model. The secondary component is called the contamination, and the probability ε is the amount of contamination.

An example may help clarify this idea. Consider the situation where 97% of the time a specimen is obtained from a population of "good" specimens while the remaining 3% of the time consistently lower strength measurements are obtained, either due to manufacturing

defects or to faulty testing. The primary PDF would correspond to the "good" specimens, the contamination would represent the distribution of flawed specimens, and the amount of contamination is $\varepsilon = 0.03$.

The following procedure is introduced in order to examine the effects of computing high reliability values when uncertainties exist in selecting the functions for the stress-strength model. Initially, high reliability values are obtained from the normal stress-strength model (see Equation 9) using known PDFs with different mean values but equal coefficients of variation. The difference in mean values were determined from the required level of high reliability. Another R value is then obtained by applying this known distribution with a small amount of contamination (ε) in order to show an almost undetectable difference graphically between the true and contaminated PDFs. The effects of this difference in the reliability computation is discussed in the following sections in order to examine the sensitivity of the stress-strength model to the assumed PDFs. This procedure provides an effective way of demonstrating the effects of assuming a specific PDF in determining high reliability.

The normal PDF with variance contamination for the strength data is,

$$N_{S_V}(\mu_S, \sigma_V^2) = (1-\varepsilon)\, N(\mu_S, \sigma_S^2) + \varepsilon N(\mu_S, K_1 \sigma_S^2)\,, \tag{12}$$

where μ_S and σ_S^2 are the mean and variance for the uncontaminated normal strength distribution, K_1 is a scaling factor, and 100 ε is the percent contamination.

The strength distribution with location contamination is

$$N_{S_L}(\mu_L, \sigma_S^2) = (1-\varepsilon)\, N(\mu_S, \sigma_S^2) + \varepsilon N(\mu_S \pm K_2 \sigma_S, \sigma_{S*}^2)\,, \tag{13}$$

where K_2 is a scaling factor for the mean μ_S, and the sign determines which tail of the distribution is to be contaminated and σ_{S*}^2 is the variance on $\mu_S \pm K_2\sigma_S$. The location contaminated PDF (see Equation 13) can provide reliability estimates to represent the potential of a secondary failure mode. Contamination of the stress distribution would be similar to that in Equations 12 and 13. It was not necessary to include contaminated distributions for both stress and strength in order to show substantial reduction in the high reliability estimates. The strength PDF contamination was sufficient.

A linear relationship to obtain R for the reliability models when a combination of both contaminated and uncontaminated stress and strength normal PDFs can be written as,

$$R = (1-\varepsilon_1)(1-\varepsilon_2)\, R_{00} + \varepsilon_1 (1-\varepsilon_2)\, R_{10} + \varepsilon_2 (1-\varepsilon_1)\, R_{01} + \varepsilon_1 \varepsilon_2 R_{11} \tag{14}$$

where 100 ε_1 and 100 ε_2 are percent contamination for the stress and strength distribution, and the R_{ij} values are obtained for the case of <u>variance</u> contamination only; that is,

$$R_{ij} = \Phi\left(\frac{\mu_S - \mu_s}{\sqrt{\sigma_{S_j}^2 + \sigma_{s_i}^2}}\right). \tag{15}$$

and for location contamination, R_{K_L} would be

$$R_{K_L} = \Phi\left(\frac{\mu_{S_L} - \mu_{s_K}}{\sqrt{\sigma_S{}^2 + \sigma_s^2}}\right), \tag{16}$$

Equation 14 can be extended to include all combinations of variance and location contamination simultaneously, but it was not necessary for this sensitivity analysis. In Equation 16, if i, j = 0, then there is no contamination; for i, j = 1, then both stress and strength are contaminated. For example, if there is contamination of variance of strength only, then

$$R = (1 - \varepsilon_2)\, R_{00} + \varepsilon_2\, R_{01} \tag{17}$$

where

$$R_{00} = \Phi\left(\frac{\mu_{S_0} - \mu_{s_0}}{\sqrt{\sigma_{S_0}{}^2 + \sigma_{s_0}{}^2}}\right)$$

and

$$R_{01} = \Phi\left(\frac{\mu_{S_0} - \mu_{s_0}}{\sqrt{\sigma_{S_1}{}^2 + \sigma_{s_0}{}^2}}\right)$$

LOWER RELIABILITY BOUND

A conservative lower bound on the reliability is introduced in order to protect against incorrectly identifying statistical functions in determining high R. The bound is obtained from a method proposed by Bolotin,[15] and modified to employ the extreme value PDFs (see Equations 5 and 6). The method provides more conservative bounds than would be obtained from standard methods which are dependent on the assumed PDFs. The selection of the extreme value functions provides additional conservatism because of their heavier tails. The method is simple to use and is not restricted to any specified PDF. The reliability bounds are (see Figure 4),

$$1 - W_1 W_2 > R > (1 - W_1)(1 - W_2) \tag{18}$$

where $(1 - W_1)(1 - W_2)$ represents the probability $s < s_1$ and $S > S_1$, which can be a somewhat conservative estimate.

The lower bound is then,

$$R_L > (1 - W_1)(1 - W_2)\ , \tag{19}$$

where

$$W_2 = \int_{s_1}^{\infty} f_2(s)\, ds \quad \text{and} \quad W_1 = \int_{\infty}^{S_1} f_1(S)\, dS$$

for any choice of $s_1 = S_1$

15. BOLOTIN, V. V. *Statistical Methods in Structural Mechanics.* Holden-Day, Inc., San Francisco, CA, J. J. Brandstaffer, ed., 1969.

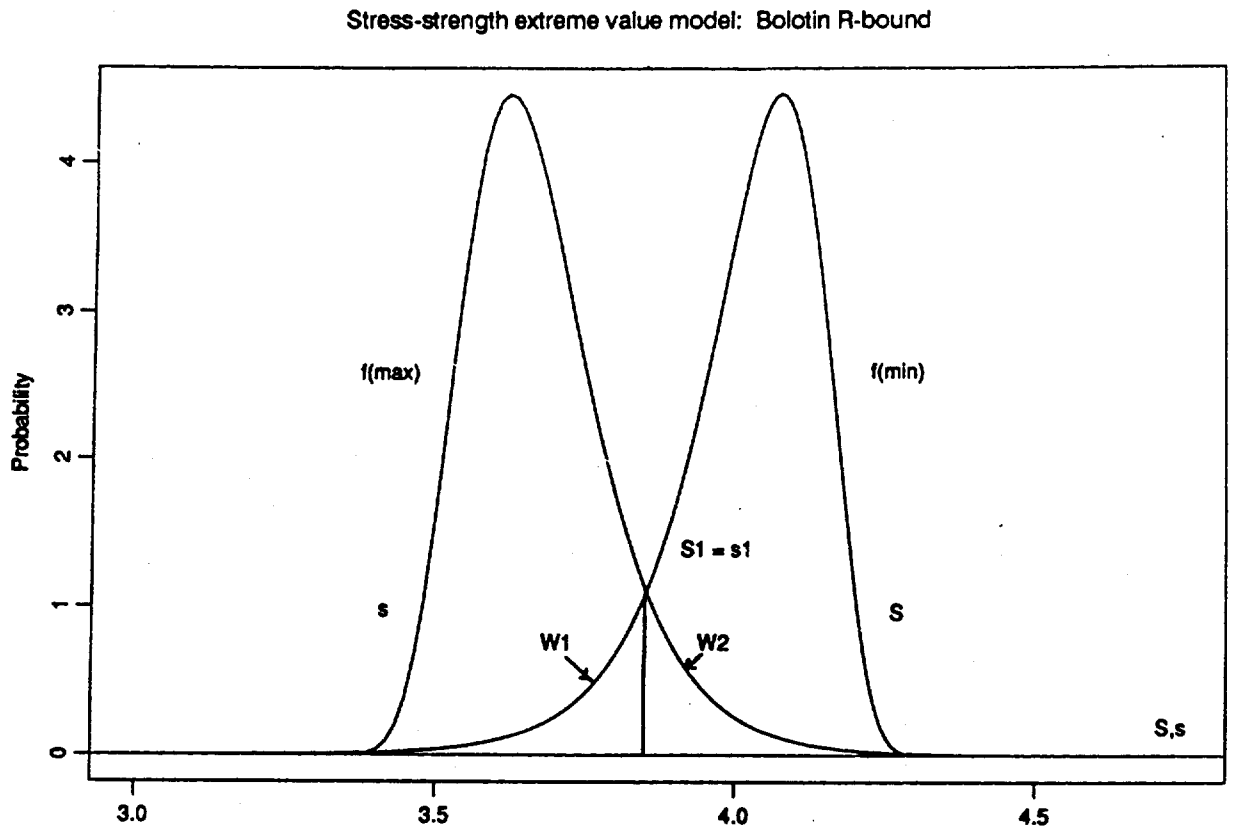


Figure 4. Bolotin reliability bounds using extreme value functions.

GOODNESS-OF-FIT TEST

The capability of determining desired PDFs from empirical data was investigated. The choice of PDF will be shown in the following sections to have a substantial effect on high reliability computations, so it is important to examine model selection procedures. A statistical test[16] of goodness-of-fit was introduced in addition to graphical displays in order to select the desired PDFs. Empirical data used in the investigation was obtained by randomly selecting a relatively large number of values from a known normal PDF. A comparison of known contaminated PDFs and the uncontaminated PDF is made with respect to the empirical values.

RESULTS AND DISCUSSIONS

Variance Contamination

Shown in Figure 5a are reliability computation results and graphical display of a normal/normal stress-strength reliability model, where a 1% (ε = 0.01) variance contamination was introduced and scaled by K_1 = 4. The graphical display was obtained from application of Equations 12

16. ANDERSON, T. W., and DARLING, D. A. *A Test of Goodness-of-fit.* J. Am. Statis. Assoc., v. 49, 1954, p. 765-769.

and 13, where $N(\mu_S, \sigma_S^2)$ is defined in Equation 3. The graph shows an almost undetectable difference between the contaminated and uncontaminated PDFs. This indicates that the choice of ε and K are reasonable with respect to the potential differences between assumed and actual PDFs. However, the reliability values differ substantially ($0.9_{(6)}$ versus 0.998989). This implies that either one failure in a million or 1011 failures in a million is predicted depending on the selection of PDFs which can differ in probability values by less than 0.0005 in the extreme tail regions (see Figure 5b). Using "good" representative PDFs in the stress-strength model in predicting only a single failure will occur in one million operations (e.g., number of flight hours) for $R = 0.9_{(6)}$ can result in a severe anticonservative estimate since for almost identical PDFs, 1011 failures per million could also be predicted.

The accuracy of the high R estimates depends on the level of precision in defining the extreme tail of PDFs. This requires selecting a PDF from a data set that accurately represents the known population function in the extreme tail regions with a probability difference of much less than 0.0005. Unfortunately, this would require an unrealistically large data set. In current practice, if a very large data set is not available, then PDFs are selected from smaller sets with reliance on the functional representation in regions less than first ordered or greater than the largest value.

The stress-strength procedure is quite effective for the range of R values between 0.5 and 0.95 since usually in the extrapolation process, a small difference in the extreme tail probabilities values will not effect the required accuracy in R. Reliability results from uncontaminated and variance contaminated ($\varepsilon = 0.05$ and $K_1 = 5$) PDFs showed no differences for a known $R = 0.95$. Unfortunately, in order to obtain high reliability, extrapolation into the extreme tail of the PDFs is required, thereby increasing the required level of precision necessary to distinguish between, for example, 0.998 and $0.9_{(6)}$.

In order to demonstrate the uncertainties in selecting specific PDFs from empirical data when computing high reliability values, the following displays are shown in Figure 6. In Figure 6a, a plot is shown of the empirical normal cumulative density function (CDF) and the corresponding contaminated and uncontaminated normal distribution functions where the mean is 50 and standard deviation (SD) is 5, with sample size $\mu = 100$. Reliability values are also tabulated from the stress-strength model results using all six candidate functions. For example, R(3, 5) is the reliability obtained from variance contamination of 3% and a scale of 5 for variance. A statistical goodness-of-fit test[16] that measures the relative differences in the tail region of the distributions was applied in addition to visual inspection in order to establish if each function could represent the CDF of the ranked data. Results showed this to be true; see Figure 6b for the tabulated observed significance level (OSL) which show in all cases OSL > 0.05, a requirement for the assumed function to be considered from the same population as the empirical data.

The results show that although each distribution fits the data quite well (see Figure 6b), there is a large relative difference in R values: $0.9_{(6)}$ for R(U.C) and 0.9957 for R(3, 5). In Figure 7, the results are similar to those in Figure 5. The variance contamination was 1% with a scale factor of 6 for both σ_S and σ_s. Again, although the functions are similar, the relative reliabilities differ substantially ($0.9_{(6)}$ versus 0.9977197). As was the case in Figure 5, severe consequences could exist if $R = 0.9_{(6)}$ is assumed and the actual reliability was 0.9977197. Since this could result in a number of premature failures, 2280 in one million, occurring compared to the assumed one failure in a million. The results showed a low level of sensitivity to the selection of the factor K_1.

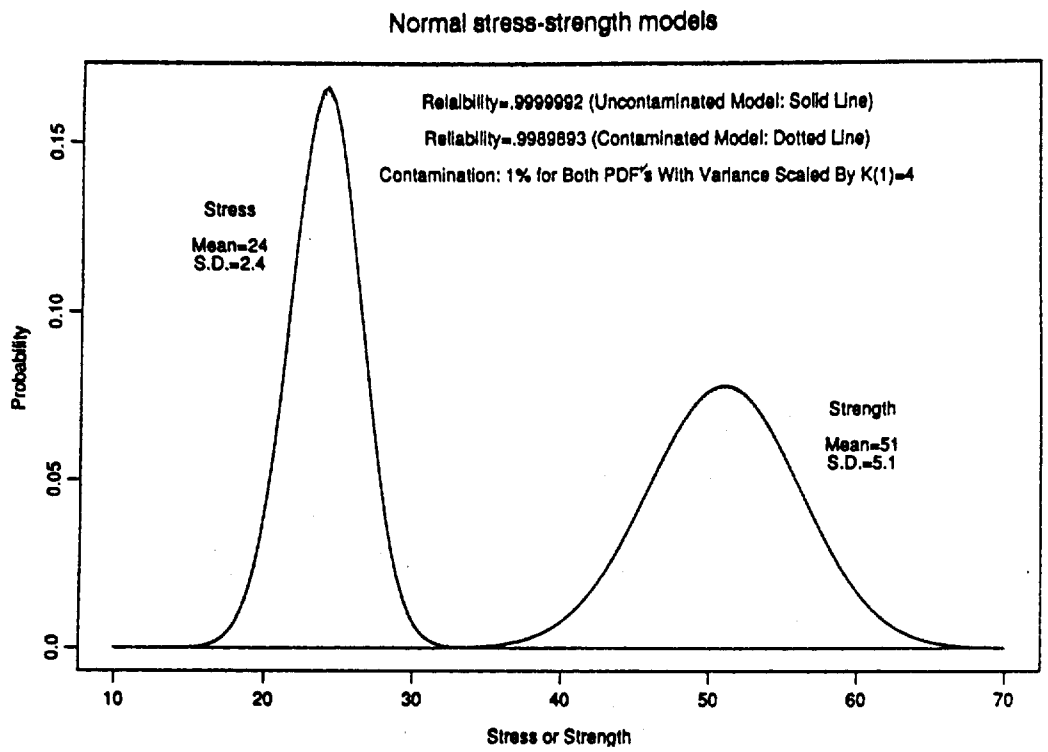


Figure 5a. Stress-strength normal functions with and without variance contamination.

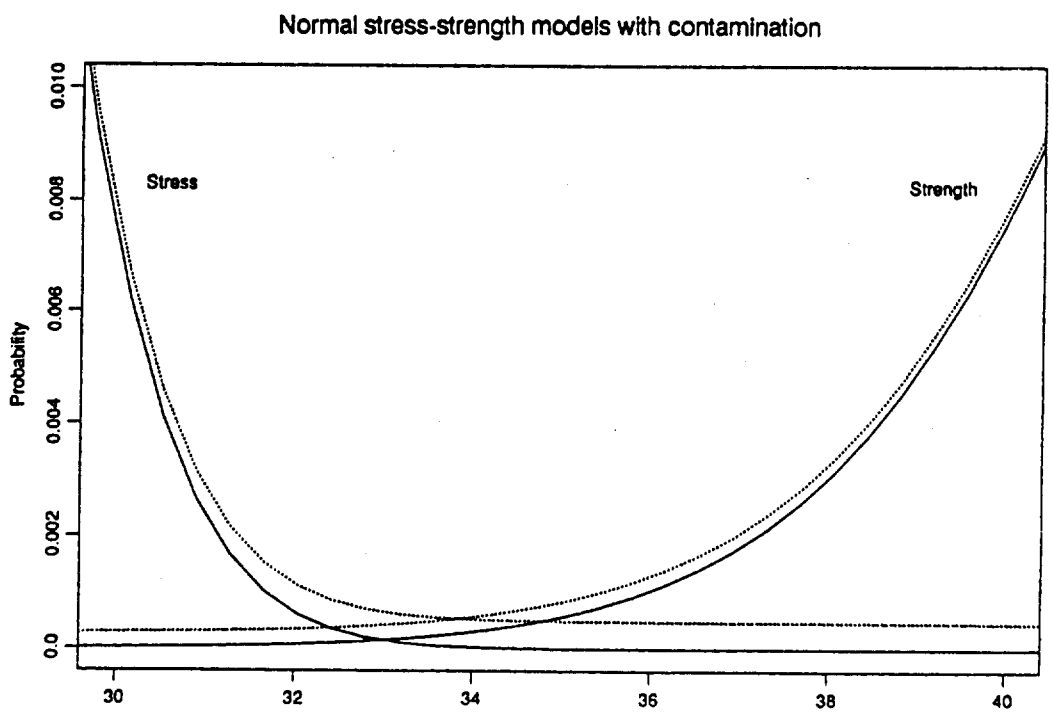


Figure 5b. Detailed region of intersecting functions (see Figure 5a).

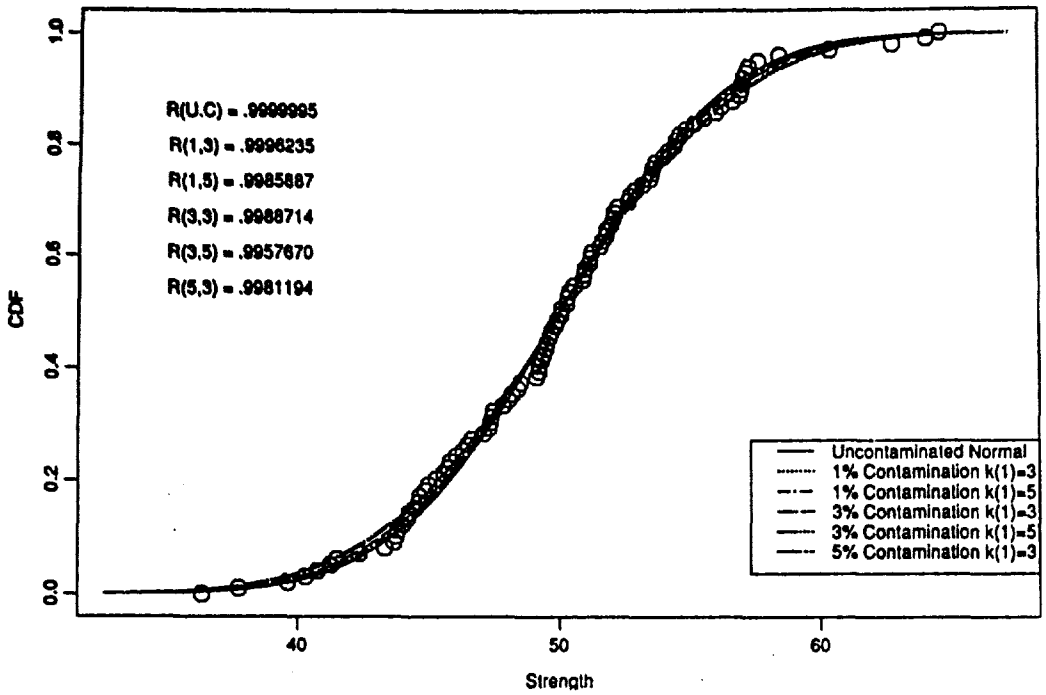


Figure 6a. Goodness-of-fit: empirical versus functional normal CDFs.

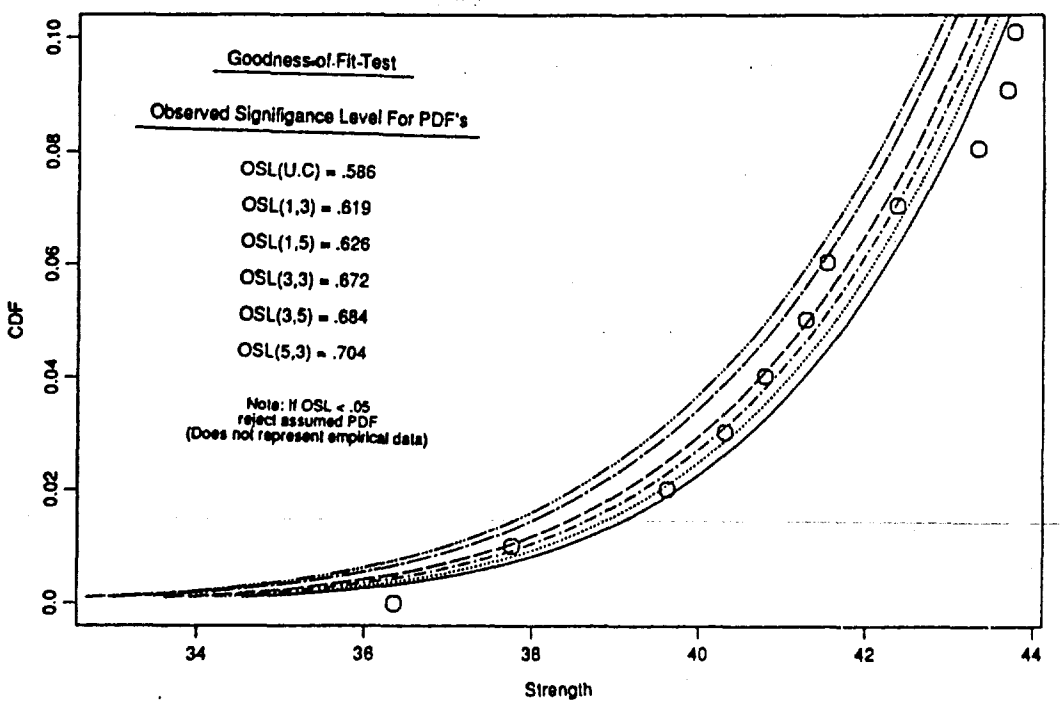


Figure 6b. Lower tail of empirical/normal distributions (see Figure 6a).

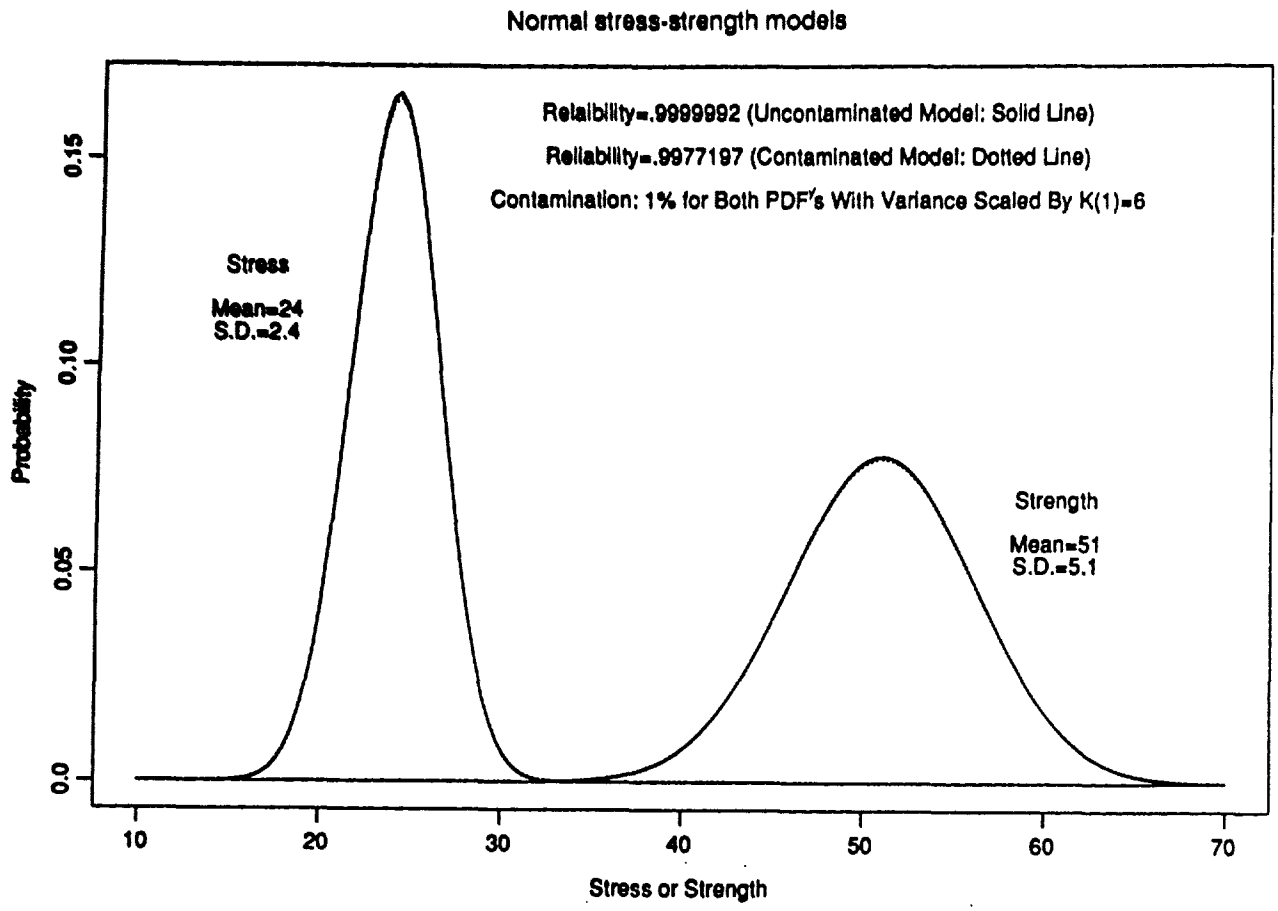


Figure 7. Reliability/normal functions with and without variance contamination.

Location Contamination

In Figure 8, reliability computation results and a graphical display of the stress-strength models are shown. The contaminated functions were obtained from 1% (ε = 0.01) location contamination as defined in Equation 13 where K_2 = 4 and the (-) value is used for strength and (+) value for stress. The contamination in this case represents a secondary failure mode not considered when assuming a specified function from the test results. For example, ignoring the possibility that one in every 100 parts may have a lower strength level, say 4 standard deviations from the mean, can result in the reliabilities tabulated in the figure. That is, for the assumed correct model, R = $0.9_{(6)}$, and the actual case where there was a lower strength level having one chance in 100 of occurring resulted in R = 0.999459. Figure 9 provides similar results to those in Figure 8 except there is a greater difference in reliability values $0.9_{(6)}$ versus 0.991012 due to a greater shift (K_2 = 6) in the mean value for the contaminated PDF. With a 1% contamination this result is predictable since one in a hundred times a failure should occur because μ_S - $K_2\sigma$ is less than the mean of stress value. The above figure shows the consequences of not being able to identify the correct function because of the inability to always detect a flawed component. The result is the determination of an overly optimistic reliability value when the true reliability could actually be orders of magnitude less.

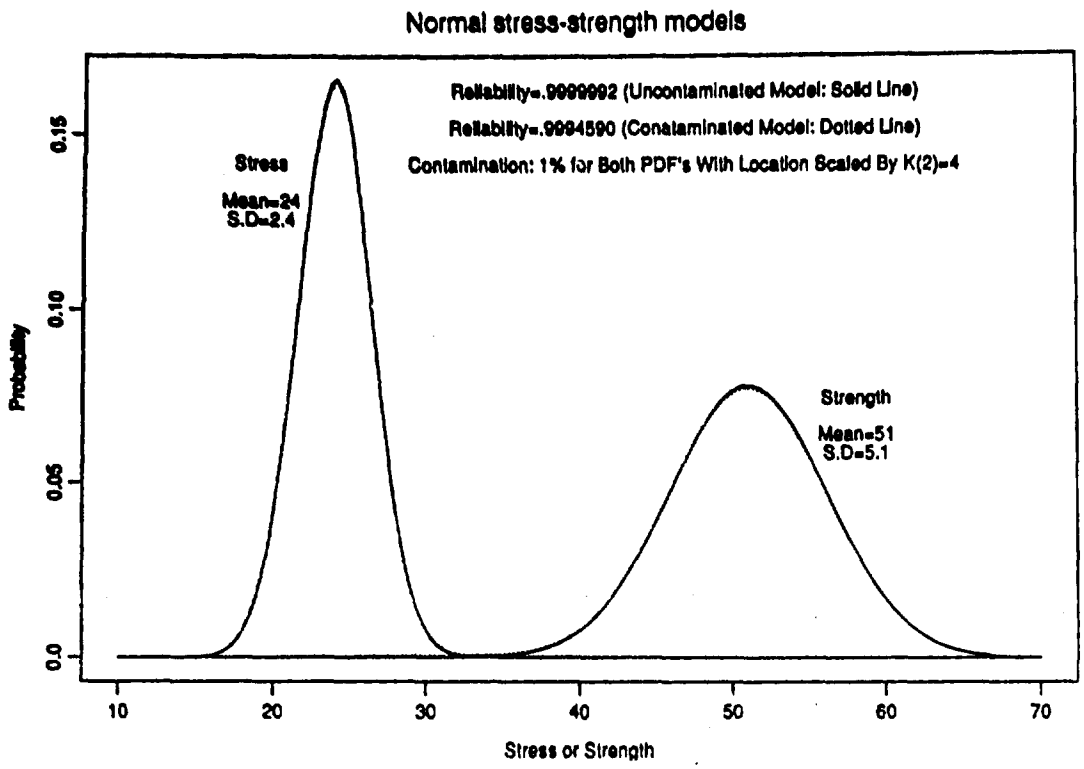


Figure 8. Reliability/normal functions with and without location contamination.

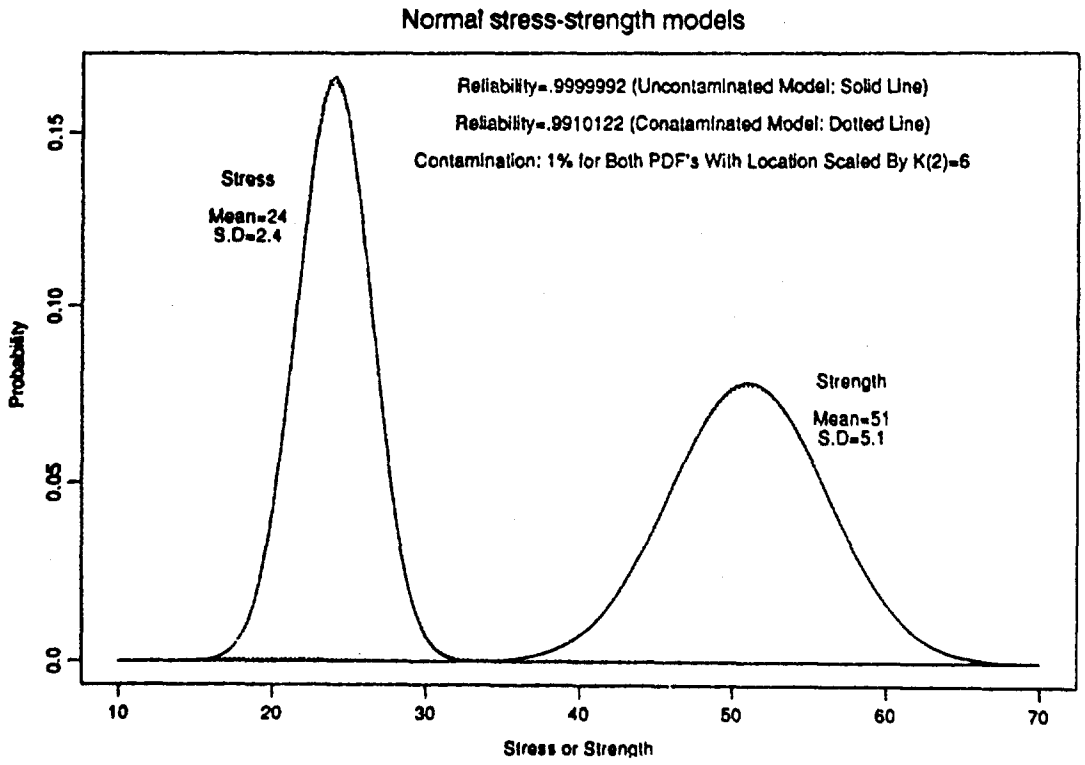


Figure 9. Reliability/normal functions with and without location contamination.

The results in Figure 10 are similar to those in Figure 5 except these were obtained from $\varepsilon = 0.03$ and $K_1 = 3$. If the estimated R = 0.9987350 is obtained from the empirical data and a higher R value is required ($R = 0.9_{(6)}$), then a material with either greater strength or less contamination would be required. In order to obtain the required $0.9_{(6)}$ from the original contaminated model, a mean strength of 87 is required (see Figure 11). The mean of 87 requirement may not be acceptable to the designer, but this situation can occur if there is a substantial amount of dispersion in the strength data resulting in a long-tailed PDF. The above situation shows when a potentially over-design situation could occur because of the inability to identify the correct PDF in the stress-strength model due to inherent sensitivity and lack of information in the tail regions. This could prevent a good design from being accepted if it is required that the assessment of the design be based upon reliability only.

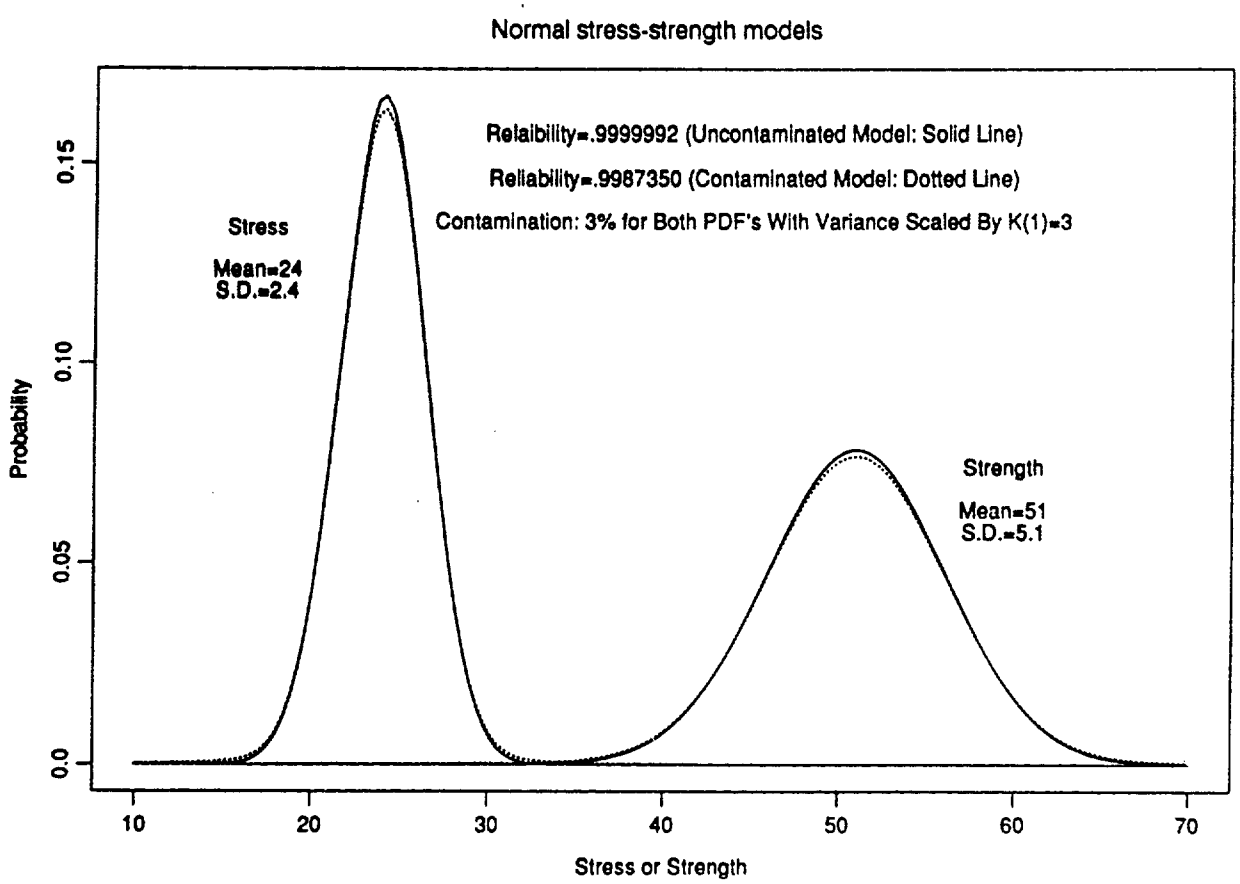


Figure 10. Reliability/stress-strength with and without contamination.

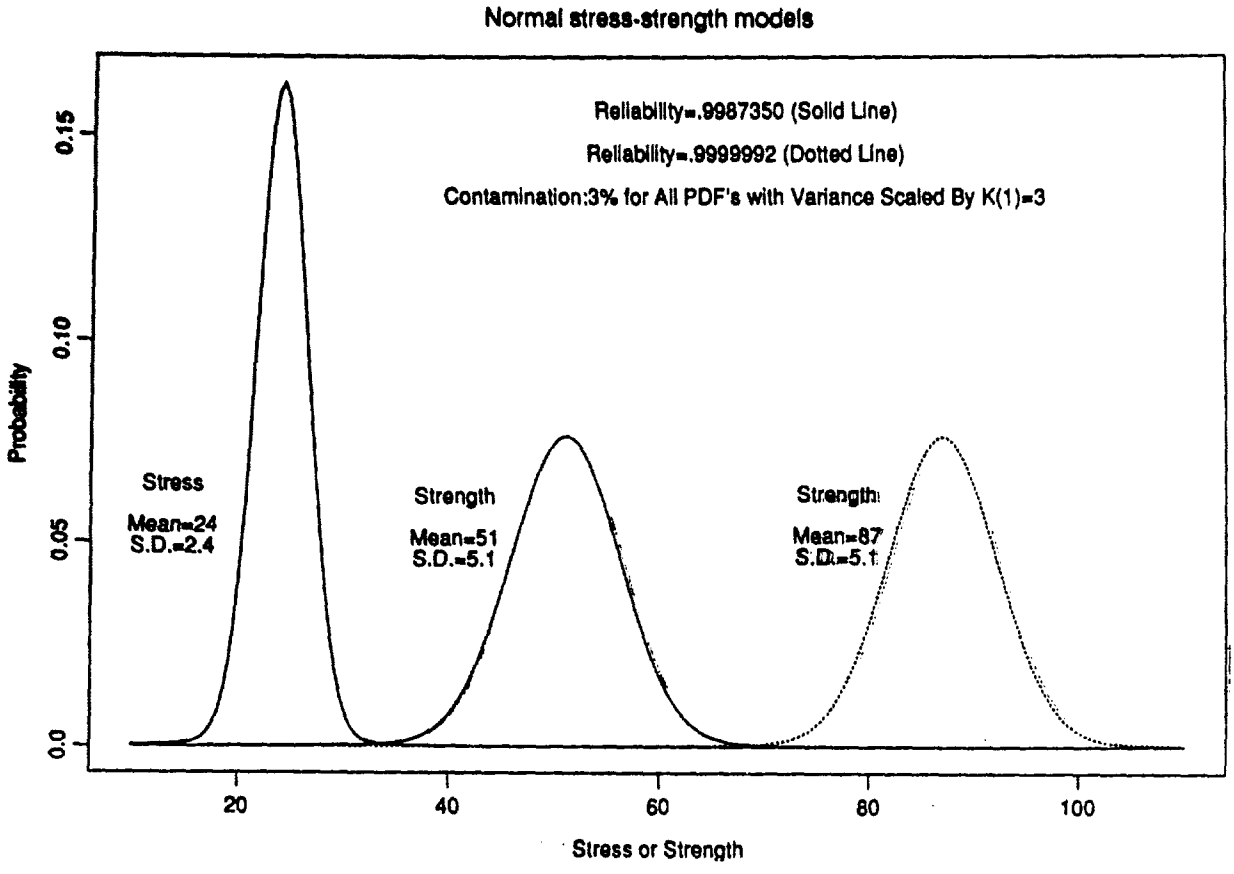


Figure 11. Increased strength requirements for high reliability.

The display in Figure 12 presents four possible reliability values for the case where the means and standard deviation are: stress (24, 2.4) and strength (51, 5.1). The result from the uncontaminated normal is $0.9_{(6)}$. R = 0.995043 was obtained from the contaminated PDF application. Since, as was shown previously, the stress-strength model will often provide either relatively very high or low R values depending upon the chance selection of the PDFs. In order to compensate for the uncertainty in selecting the PDFs for stress and strength data, extreme value distributions are introduced (see Figure 2) in the reliability computation. This resulted in R = 0.999045. Unfortunately, this did not provide a value lower than the contaminated model result of 0.9950428. In order to obtain additional conservatism in the R estimate, a modification of a method by Bolotin is examined involving the determination of lower bound on R (see Figure 4 and Equation 19) in conjunction with the extreme value PDFs. The resultant lower bound estimate of 0.9796063 provides a significantly lower value than that of the contaminated model. This was also true for all contaminated models in this study.

This lower bound estimate could provide some security in estimating R, although results may be excessively conservative for some practical applications. In Table 1, the distribution of R values as a function of the sample size is presented. R values were obtained from repeated application of the uncontaminated stress-strength model of Figure 5 using randomly selected, normally distributed samples. For a sample size of 10, R ranges from $0.9_{(6)}$ to

0.998417 indicating the instability associated with small samples. Higher order quantiles (e.g., 60%) were not included since they were all greater than $0.9_{(6)}$.

Table 1. DISTRIBUTION OF R VERSUS SAMPLE SIZE

Distribution (%)	Reliability R			
	Sample Size			
	10	50	100	1000
0.1	0.998417 (1583)	0.999932 (68)	0.999980 (20)	0.999998 (2)
1	0.999160 (840)	0.999968 (32)	0.999991 (9)	0.999998 (2)
10	0.999943 (57)	0.999994 (6)	0.999998 (2)	0.999999 (1)
25	0.999994 (6)	0.999998 (2)	0.999999 (1)	0.999999 (1)
50	0.999999 (1)	0.999999 (1)	0.999999 (1)	0.999999 (1)

( ) Corresponding Number of Failures Per Million

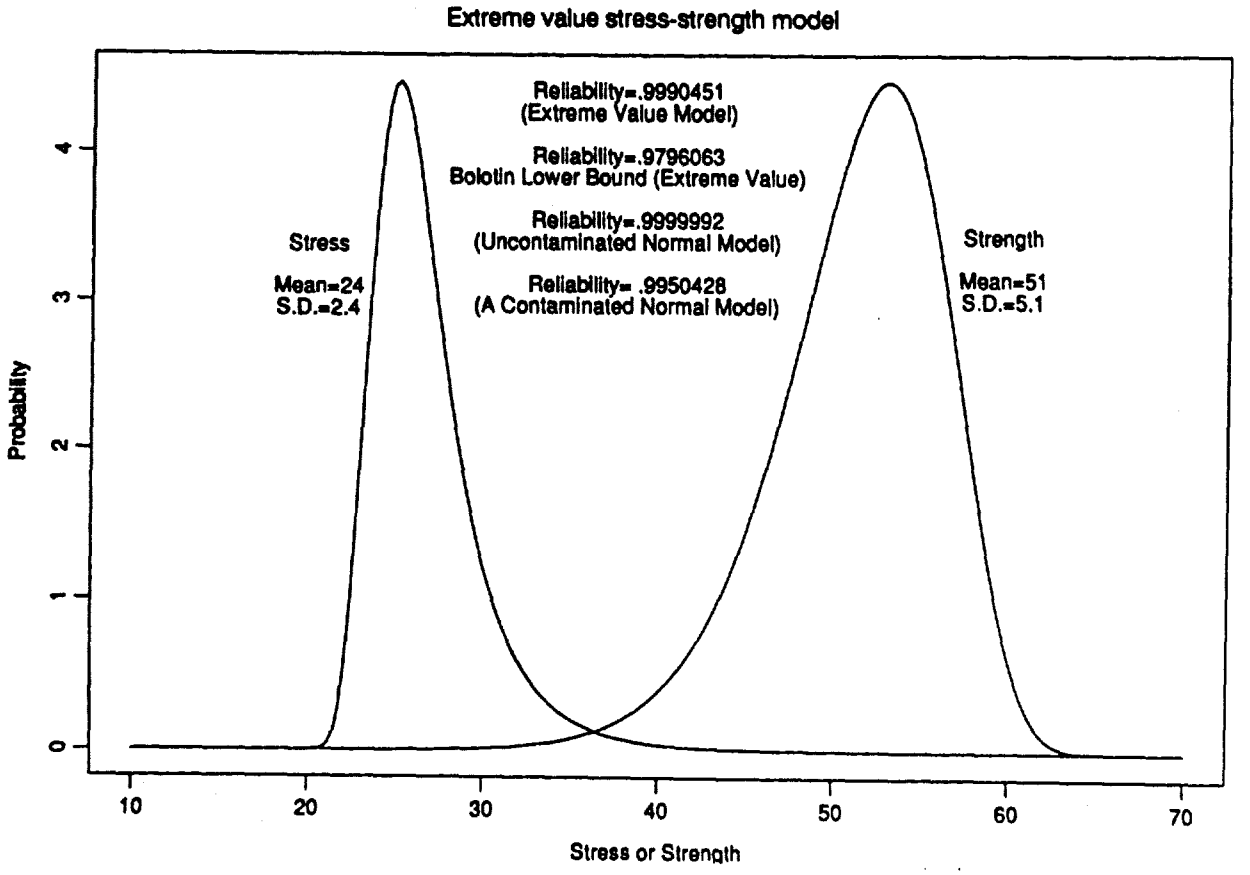


Figure 12. Reliability comparison: PDFs and lower bound.

A sample size of 50 or 100 provides reasonable stability, and a sample of 1000 shows essentially no variability. The results from Table 1 show that for a sample of 1000, an estimate of $R = 0.9_{(6)}$ would be acceptable. This is not necessarily correct since results from the table only address the sample size issue which is independent of the uncertainties in the PDFs selection process. There are two requirements for obtaining accurate high reliability values from the strength model: large samples ($n > 1000$) and knowledge of the population PDF. Reliability estimates of 0.95 are much less sensitive to the PDF assumption. If there is a secondary failure mode due to occasional undetected poor manufacturing of the material or an unusually large load occurs that is not accounted for in the design process, then unknown lower reliability values ($R < 0.95$) can exist.

CONCLUSIONS

High reliability estimates from application of the statistical stress-strength model can vary substantially even for almost undetectable differences in the assumed stress and strength PDFs. Specifying high R values (e.g., $0.9_{(6)}$) for acceptable structural design can result in higher failure rates than anticipated if the assumed PDFs contain shorter tails than actually exist. Over-design situations can also occur when excessively long-tailed PDFs are applied to the stress-strength model. An effective method for identifying this nonrobust behavior involved application of contaminated and uncontaminated PDFs in the determination of reliability values.

A suggested method for obtaining a lower bound on the reliability estimate provided potentially overly conservative results but was effective in determing values that were lower than any of the R values computed for the contaminated models.

The authors' position regarding the computation of high reliability of $0.9_{(6)}$ agrees with Breiman[17] who says "The probability of failure $P_f = 1 \times 10^{-6}$ has an Alice in Wonderland flavor and should be banned from nonfiction literature." It is therefore recommended that if high reliability calculations are absolutely essential, then the results should be subjected to a sensitivity analysis using contaminated distributions. High reliability values are meaningful only when these values are not substantially affected by an amount of contamination (ε) consistent with the sample sizes, and a severity of contamination which is identified by engineering judgement.

ACKNOWLEDGMENT

The authors wish to thank Lucy Ohannesian and the IMA-V group of U.S. Army Materials Technology Laboratory for preparing this manuscript for publication.

17. BREIMAN, L., and STONE, C. *Broad Spectrum Estimates and Confidence Intervals for Tail Quantiles.* Technical Report No. 46, Stat. Dept., University of California, Berkeley, CA, 1985. Also to be published in Journal of Statistical Computation and Simulation.

IMPROVING NONPARAMETRIC TOLERANCE LIMITS FROM POOLED DATA

Donald M. Neal, Mark G. Vangel, and Trevor D. Rudalevige

U.S. Army Materials Technology Laboratory
Watertown, Massachusetts 02172-0001

Abstract

This paper introduces a method for obtaining an improved tolerance limit value for a small sample S of material strength data by pooling with a much larger sample L. This value represents a 'B' basis material property number defined as the 95% lower confidence bound on the tenth percentile of the population distribution.

In the pooling process both data sets are transformed to a common mean value of zero in order to pool samples with significantly different strength levels. Equality of variance is required between S and L. The basis values are obtained from the pooled samples by application of both nonparametric and parametric statistical models. Monte Carlo studies showed that by pooling both data sets S and L, values could be obtained that were less conservative with lower variability than from application of S alone.

Preceding Page Blank

Introduction

The statistically based material property value, or B-basis value, is a statistic which is less than the tenth percentile of the population with probability .95. That is, the B-basis value is a 95% lower tolerance limit for the tenth percentile. In Figures 1a and 1b, a graphical display is shown for the basis value probability density function for sample sizes of n = 10 and 50 from a standard normal population. The dotted vertical lines indicate the 10th percentile ($x_{.10}$) of the population and the probability (basis value $< x_{.10}$) = .95 for the basis value probability density function. The graphical display of the basis value density functions show much less dispersion for n = 50 than for n = 10. Therefore, small samples will usually result in lower basis values. In [1, 2, 3, 4] various procedures are described for determining the statistical property values.

The motivation for the work described in this paper resulted from a need by the aircraft industry to obtain a less conservative statistically based material property value from a small sample of composite material strength data. Here 'conservative' is to be interpreted to mean 'excessively low', which corresponds to a design engineer's use of the word. Statistical conservatism, that is a confidence exceeding the nominal level of .95, need not be present for 'engineering conservatism' to be a problem.

The use of small samples reduces the amount of testing and consequently the manufacturing cost of composite aircraft structures. For example, in order to qualify a composite material to be used in the manufacture of a commercial aircraft, the FAA[5] requires property values for tension, compression, and shear tests subjected to the environmental conditions: hot-wet, cold-dry, and room temperature for three separate batches of material. In the development of

a composite tail section by one of the major aircraft companies the cost of testing was more than 20 million dollars. In addition to the cost, excessively conservative basis values can also result in an overdesign situation, since the value often provides information in determining structural design allowables.

In order to avoid the penalty associated with using small samples in the tolerance limit computation, a procedure is introduced in this paper involving pooling a large sample with a smaller one in order to obtain the property value. This is done in order to reduce the inherent variability that occurs from applying the smaller data set.

In the pooling process the larger data set should be obtained from prior available test results or from less expensive tests. Ideally, both samples should be from the same test and environmental conditioning process. Various combinations of tests and ply-orientations are shown in Figure 2 showing the sources for combining the data sets. In the pooling process it is assumed for a given material (eg., graphite-epoxy) there are similar failure modes and consequently equality of variance[6] between the samples.

In order to avoid the uncertainties involved in identifying a statistical model from a small sample when computing the basis value, two nonparametric methods (Ferguson[7] and the Modified Hanson-Koopmans[8]) are introduced. In applying [7], the larger set represents the prior and the smaller one the empirical data. In [8] an ordered array of strength measurements are obtained from the pooled data sets. The tolerance limit is determined from a specific ordered value multiplied by a factor determined from the sample size of the pooled data. A parametric method[9] was also applied in the study in order to compare the

effectiveness of the various methods. This involved a Weibull analysis, where the shape parameter is obtained from the pooled sample.

A Reduced Ratio method[10] is recommended by the MIL-5 handbook for obtaining basis values from small samples. The procedure involves determining the ratios of paired observations from a prior data set with a known basis value and a new data set for which the value is to be determined. A lower confidence bound is determined for the ratio values. The basis value is then obtained from multiplying the bound by the prior value. This method is often overly nonconservative.

Determination of Basis Values - Nonparametric Bayesian Method

The nonparametric Bayesian basis value is obtained from the following. Let $\{x_i\}_1^n$ represent the current empirical data which the basis value is to represent and $\{t_j\}_1^m$ the prior data obtained from test results or an assumed cumulative distribution function (CDF) F(t).

Posterior Distribution

The posterior distribution using Bayes rule can be written as

$$\hat{F}_n(t \mid x_1, x_2, \ldots x_n) = \frac{\alpha((-\infty, t]) + \sum_i^M \delta_{x_i}((-\infty, t])}{\alpha(R) + M} \qquad (1)$$

where n is the sample size of the current empirical data set and $\alpha(R)$ is the prior data sample size. $\delta_x((-\infty, t])$ represents the number of values from the prior distribution less than or equal to t which are determined from the equality between the corresponding ordered value of the CDF's and the quantile of t.

$\delta_x((-\infty, t])$ represents the number of current empirical data values x less than or equal to t determined from the equality between their corresponding combined ordered values of the CDF's of x and t and the quantile of t where δ_x represents the measure giving the mass of one to the point x.

Equation (1) can be rewritten in the following manner, if

$$F_0(t) = \alpha((- \quad , t])/\alpha(R), \tag{2}$$

represents the prior CDF and

$$F_n(t \; x_1, x_2, \ldots x_n) = \frac{1}{M} \sum \delta_{x_i}((-\infty, t]) \tag{3}$$

represents the current empirical data CDF, then

$$\hat{F}_n(t \; x_1, x_2, \ldots x_n) = P_n \, F_0(t) + (1 - P_n)F_n(t \mid x_1, x_2, \ldots x_n) \tag{4}$$

represents the posterior shown as the weighted ordered array of CDF's, where

$$P_n = \frac{\alpha(R)}{\alpha(R)+M} \tag{5}$$

determines the weight for the prior and current data. By inverting the CDF's in Equation (4), a value of t can be obtained corresponding to a specified quantile.

Illustrations

A few simple examples will show the relative ease in obtaining the posterior distribution $\hat{F}_n$.

Let $t = 1, 2, 3, 4, 5,$ and $F_0 = .2, .4, .6, .8, 1.0,$

$x = 1, 2, 3, 4, 5,$ and $F_n = .2, .4, .6, .8, 1.0$

Since $\hat{F}_n(t \; x_1, x_2, \ldots x_n) = P_n \, F_0 + (1 - P_n)F_n$

Then $\hat{F}_n(1 \; 1, 2, 3, 4, 5) = (.5)(.2) + (.5)(.2) = .2$

This is the estimate for first ordered value for x obtained from a good choice for the prior F_o.

If $t = 1, 2, 3, 4, 5$ and $F_o = .2, .4, .6, .8, 1.0,$

$x = 6, 7, 8, 9, 10$ and $F_n = .2, .4, .6, .8, 1.0$

Then $\hat{F}_n(6\ 6, 7, 8, 9, 10) = (1.0)(.5) + (.2)(.5) = .6$

This is a poor estimate since .2 is the correct result. An unacceptable prior produced this result. The CDF value for t = 1 using the above data is

$$\hat{F}_n(1\ 6, 7, 8, 9, 10) = (.2)(.5) = .10$$

A Bayes estimate of the mean can be obtained as follows,

if μ_o is the mean determined from a prior data set and x_n is the mean of current sample of test results, then the Bayes mean value for x_n is

$$\hat{\mu}_n(x_1, x_2, \ldots x_n) = P_n \mu_o + (1 - P_n)\bar{x}_n, \tag{6}$$

where P_n is defined in Equation (5).

Nonparametric Tolerance Limit on the Bayesian Quantile Estimate

Since a basis value as described previously requires a tolerance limit on the quantile estimates, the following describes the process for obtaining that limit. Initially, a random sample F(y) of size $M = \alpha(R) + n$ is assumed independent of the mixture of the prior and empirical data sets shown in Equation (4). By ordering the sample F(y), the following $Y_{(1)}, Y_{(2)}, \ldots Y_{(M)}$ values are obtained. The probability density function for $Y_{(i)}$, $1 \leq i \leq M$ can be written

$$f_{Z_{(i)}}(z) = \frac{\Gamma(M)\, z^{uM-1} (1-z)^{(1-u)M-1}}{\Gamma(uM)\,\Gamma((1-u)M)} \tag{7}$$

where $z_{(i)} = F(Y_{(i)})$ and $i = uM$ with u representing the CDF value corresponding to i. The tolerance limit y* for y_q is defined as,

$$P(y_q \geq y^*) = 1 - \alpha = P[F(y_q) \geq F(y^*)]$$

where y_q is the $100q^{th}$ percentile of y. Since

$$P(y \leq y_q) = \int_0^q \frac{\Gamma(M)\, z^{uM-1} (1-z)^{(1-u)M-1}}{\Gamma(uM)\,\Gamma((1-u)M)}\, dz \quad (8)$$

from Equation (7), then a $1 - \alpha$ tolerance limit on y_q can be obtained from the following solving for u:

$$\int_0^q f(z)\, dz = 1 - \alpha\,, \quad (9)$$

where f(z) denotes the integrand in Equation (8).

In the case of the 'B' basis computation, $\alpha = .05$ and $q = .10$. Equation (9) can then be written as,

$$\int_0^{.10} \frac{\Gamma(M)\, z^{uM-1} (1-z)^{(1-u)M-1}}{\Gamma(uM)\,\Gamma((1-u)M)}\, dz = .95 \quad (10)$$

See Table I for tabulation of u and M values that satisfy Equation (10).

If, for example, $uM = 1$ and $u = .034$, then

$$\int_0^{.10} f(z) = \left(\frac{M}{M-1}\right) - \frac{M}{(M-1)}(.90)^{M-1} = .95 \quad (11)$$

and $M = 29$, that is $P(Y_{.10} \geq Y_{(1)}) \geq .95$. The tolerance limit is then the first ordered value $Y_{(1)}$ for the sample size of $M = 29$. A 'B' basis value can

now be obtained from the CDF in Equation (4) by the following procedures. Solving for u in Equation (10) determines the lower tolerance limit of the CDF of sample size M and from some distribution F(y) where the i = uM.

Obtaining a lower ordered CDF value from Equation (4) that is approximately equal to u determines the 1 - α tolerance limit of the pth quantile of the posterior CDF for a sample size M.

Example:

If there is only prior data $\{t_j\}_1^{30}$ and a 'B' basis value is required where

$$t_j = 5, 6, 7, 8, 12, 16, 20, 25, \ldots 40$$

$$\text{and} \quad F_0(t) = .033, .066, .099, \ldots 1.0$$

then for M = 30, u = .034 from Table I. The basis value t_j is determined from an approximate solution of $u \approx F(t_j)$ resulting in $t_1 = 5$; therefore, the first ordered value of the prior represents the 'B' basis value, which is the same as the nonparametric quantile sign test[11] result, when the sample size is 30.

Basis Property Values - Nonparametric (HK) Process

A nonparametric procedure (HK)[8] for estimating tolerance limits is introduced for computing the 'B' basis value for any sample size $\geq$ 2. The method is a modification of Hanson-Koopmans[12] process. The modification has reduced the excessive conservatism in computing property values when compared with the original HK method.

The method involves the following. Let $x_1, \ldots x_n$ be the order statistic of an independent and identically distributed sample from a continuous distribution F. Assume that F is log-convex, that is - log F(X) is a convex function. The class of log-convex functions includes a large enough group of distributions so

that following procedure involving log-convex functions can be considered non-parametric.

The lower tolerance limits of the form

$$T_{rs} = K\, x_r + (1 - K)x_s \tag{12}$$

can be obtained where $r < s$ and $K \geq 1$. When T_{rs} is used with positive data values, negative tolerance limit can be obtained which are not valid if the distribution F is zero for any negative values. A practical solution to this problem is to apply the Hanson-Koopmans approach by taking the log of the data x, that is,

$$T^{*}_{rs} = K \log x_r + (1 - K)\log x_s \tag{13}$$

and then obtain by exponentiation the following

$$\tilde{T} = e^{K \log x_r} + e^{(1 - K)\log x_s} = x_j \left(\frac{X_r}{X_s}\right)^K \tag{14}$$

For most distributions of interest, $\tilde{T}_{ij}$ still provides conservative tolerance limits, although technically $\tilde{T}_{ij}$ is valid for a class of distributions smaller than the log-convex class corresponding to T_{ij}.

In order to determine the 'B' basis value, the i, j, and K values are obtained for a given n are shown in Table II.

Weibull Model Property Values - Pooling Data For Shape Estimate

A Weibull method was introduced by[9] in order to obtain 'B' basis material property values for various tests and conditioning processes of composite materials. Although the Weibull scale parameters can differ depending on the type of test or condition, the within sample variances are usually similar. By this assumption, the data from the various sources are pooled in order to determine an approximate population shape parameter α^* for the Weibull distribution. In order to obtain a basis value for some specified smaller data set Y (usually

from tests and conditioning that requires considerable expense), the following relation is applied:

$$'B' = \beta_y[-\ln(0.90)]^{1/\alpha^*}[2n/\chi^2_{.95}(2n)]^{1/\alpha^*} \qquad (15)$$

where the quantity $\chi^2_{.95}$ is the .95 Chi-square quantile with 2n degrees of freedom, and n is the number of Y values. β_y is the Weibull scale parameter estimated from the smaller sample of size n. Equation (15) is not valid unless α^* can be adequately represented by a large enough sample to provide a good approximation to the true α value. The data is obtained from various tests (tension, compression, and shear) and conditionings (hot, cold, hot-wet) in order to obtain the value α^*.

The Pooling Process

In order to effectively apply the previously discussed methods, a moderately large data set (n > 30) is suggested. This is accomplished by pooling a smaller set S (limited available data) of size n_S with a larger set L (from prior testing) of size n_L. In the HK process the objective is to represent S with combined data sets S and L of sample S_p with size $m = n_L + n_S$ obtained from S and L. In the Ferguson's Bayes method the prior is represented by L and the empirical data by S. The Weibull analysis method obtains information for estimating the population shape parameter α by pooling S and L and using S to determine the scale parameter β.

Ideally, if the mean of S and L are equal and their variances are also equal, then pooling could easily be justified. If the mean values differ, then a normalizing procedure combined with an equality of variance test is required.

This involves representing S and L by the distributions S* and L* with means equal to zero, that is, let

$$S^* = \frac{S - \bar{S}}{\bar{S}}$$

and (16)

$$L^* = \frac{L - \bar{L}}{\bar{L}}$$

where S and L are means of the S and L data values. A schematic of this transformation is shown in Appendix A. The variance equality between S and L requirement is stated as,

$$VAR[S^*] = VAR[L^*]. \tag{17}$$

In order to determine if equality of variance exists, the Siegel-Tukey nonparametric rank sum method[13] was applied. The process involved combining S* and L* values into a single ordered array and assigning low ranks to extreme observations and high ranks to central observations. That is, assign rank 1 to lowest number of the sequence, ranks 2 and 3 to the two highest values, then ranks 4 and 5 to the next two lowest. This process is continued until all numbers in the sequence have been assigned a rank R_i. Note, for odd number of values, the middle value is not counted so that the highest rank is even. In the ranking process, knowledge of the assigned values from S* and L* must be retained.

The sum (R_i) of the ranks is obtained for either S* or L*. Usually, the smaller data set is selected in order to reduce the amount of computation. R_S obtained from the S* ordering can be written as,

$$R_S = \sum_{i}^{m} R_i \quad \text{, where i depends on ordering of ranking for S*}$$

The hypothesis test for VAR[S*] = VAR[L*] involves applying the Wilcox Rank Sum (two sample) test[13] as follows. In order for the hypothesis to be accepted, then

$$R_S > R_C, \text{ where } R_C \text{ is} \tag{18}$$

a critical value obtained from tables in Reference (13). The values are tabulated as a function of n_S, n_L, and a specified significance level. For example, a 10% significance level would represent a two sided test with a 5% to 95% interval of acceptance. If $R_C > R_S$, then reject the hypothesis.

An example showing application of the above procedures is as follows,

If sample from S*: 5, 8, 10, 19, 25

and If sample from L*: 4, 9, 12, 14, 16

then

Score	4	5	8	9	10	12	14	16	19	25
Sample	L*	S*	S*	L*	S*	L*	L*	L*	S*	S*
Rank	1	4	5	8	9	10	7	6	3	2

and $R_S = 4 + 5 + 9 + 3 + 2 = 23$

The critical value from the tables is $R_C = 19$ for $n_S = n_L = 5$ with significance level of 10% (a two sided 5% level); therefore, since $R_S > R_C$, there is no significant difference in the variances between S* and L* at the 10% level.

'B' Basis Values for S* and S from Pooled Data

Bayes Solution

If the transformation (Equation (16)) has been applied and the equality of variance is established between S* and L*, then 'B' values can be obtained using the combined data sets from both S* and L*. In the Bayes application let the smaller sample x (newly obtained data) to be represented by the S* values and

the larger sample t (the prior) by L*. The approximate solution to $u \sim F(S_j^*)$ for j determines the jth ordered value representing the 'B' basis value (S_B^*) when u is obtained from the solution in Equation (10).

Nonparametric HK Method

The nonparametric[8] solution for obtaining 'B' basis values involves pooling the values from S* and L* and letting the resultant array of values be x in Equation (14) with sample size $m = n_S + n_L$. Let this value be denoted S_B^*. This method is very simple to apply yet provides accurate results for any sample size greater than 2. The basis values for S* is not sufficient since S and L were the original data sets involved in the analysis. Therefore, the following transformation is required:

$$S_B = S_B^* \bar{S}_{.95} + \bar{S}_{.95} \tag{19}$$

where S_B is the required basis value for the small sample S. The $\bar{S}_{.95}$ values represent the lower 95% confidence value for the mean of the S values. The purpose of applying $\bar{S}_{.95}$ was to prevent a situation where the required 95% coverage was not obtainable due to variability in estimating the mean $\bar{S}$ of S.

S_B values were consistently less than the 10% point, 95% of the time (definition of 'B' values). The results from both simulated and actual test data showed on occasion excessively low S_B values, but the coverage was maintained which was not necessarily so when $\bar{S}$ was substituted for $\bar{S}_{.95}$ in Equation (19).

Weibull Method - Estimating Population Shape Parameter

In Equation (15) the pooled estimate α^* is obtained from the maximum likelihood estimator (MLE) using combined data sets S_β and L_β, where

$$S_\beta = S/\hat{\beta}_S \quad \text{and}$$

$$L_\beta = L/\hat{\beta}_L$$

$\hat{\beta}_S$ and $\hat{\beta}_L$ are the Weibull scale parameter estimates for S and L respectively. If a basis value for S is required, then, after the initial equality of variance between S_β and L_β is established, the value can be obtained from the following:

$$'B'_S = \hat{\beta}_S[-\ln(0.90)]^{1/\alpha^*}\,[2n/\chi^2_{.95}(2n)]^{1/\alpha^*} \qquad (20)$$

In Appendix B a flowchart is shown summarizing the ordered procedures required for applying methods described in this report.

Results and Discussions

In Figure 3A, the 'B' basis value results are shown for the small sample S with or without a contribution in the variance estimate from large sample L. NB is the nonparametric Bayes solution where the prior is represented by L with sample size of $n_L = 30$ and empirical data by S with size $n_S = 6$. The HK values represent the modified Hanson-Koopmans method results where L and S are pooled. W(6) and W(36) represents the property values obtained from the Weibull analysis[2, 3] with a sample of 6 and 36 respectively from S. The HK(36) was determined from the modified Hanson-Koopmans method with 36 values from S only. The data was obtained from randomly selecting samples from Weibull distributions with scale $\beta = 100$, and a randomly selected shape parameter α_L from a uniform distribution. The range of α_L values were from 5 to 25 for L. The S distribution shape parameter α_S was 15. By introducing the variability in α_L,

evaluation of the equality of variance test can be made with respect to acceptability in pooling of L and S and using L as a prior in the methods. The NB, HK, and W(36) 'B' basis values show the required coverage, that is, 95% of the 'B' values are equal to or less than the 10% point of the population distribution. The results are exceptionally good in that they are almost exactly what is required without being either overly conservative or nonconservative in addition to verifying the acceptability of the equality variance test. W(6) and HK(36) are slightly conservative because of the lower estimate measured from the true 10% point. The W(6) = 82.2 is not excessively conservative for $n_S = 6$, although using data from a Weibull model could have been helpful. The Weibull result W(36) should be a good estimate since the 36 values came from a known Weibull model $\alpha = 15$, $\beta = 100$. The fact that NB and HK provided equally good results shows the methods can provide 'B' values for a small sample, which are equivalent to those values obtained from a much larger Weibull sample, when pooling the large and small samples. N(6), basis value obtained from the normal model[1, 2, 3, 4], result is nonconservative with 'B' = 87.5, although it is not excessive for n = 6.

In Figure 3B the range of 'B' value results are shown with HK(36) and W(36) having significantly less variability than the other methods. This is due to the much larger sample size which is available to estimate directly the 'B' value. W(6) and N(6) have an excessively large amount of variability (a range of 45 to 95) which is due to the instability (see Figures 1A and 1B) in estimating the parameters required to compute basis values from the small sample sizes. The NB and HK solution show less variability (a range of 55 to 92) than W(6) and N(6), but is still quite large due to the inability to estimate the mean value from S for 6 data values when the reverse transformation is made in Equation

(19). If the normalization process was not required (L and S from the same population), then variability in NB and HK can be reduced by applying pooled non-normalized data to obtain mean value in the reverse transformation.

The experimental test results used in the analysis showthe following figures were obtained from published data in [14]. In Figure 4A, both S and L are from the same population of short beam shear tests applied to Narmco composite material. The NB and HK basis values were obtained by substituting the $\bar{S}_{.95}$ value in Equation (19) by $\hat{\mu}_n$ from Equation (6) for the reverse transformation described in Equation (16). The results in Figure 4A show excellent agreement between the 10% point and the 95th percentile of the basis values for NB, HK, N(6), and HK(36).

In Figure 4B, variability in the basis value for NB and HK is very small relative to W(6) and N(6) indicating the desirability of the proposed nonparametric methods. With S and L from the same population permitted the application of $\hat{\mu}_n$ in Equation (19). This resulted in a more stable less conservative basis value since the Bayes estimator $\hat{\mu}_n$ is determined from more information than S.

In Figure 5A the 'B' basis values were obtained from 1,000 random samples of short beam shear test results. The material was manufactured by the Hercules Co., Utah. In all cases the coverage was obtained (all 'B' values less than the 10% point), although they were somewhat conservative. NB and HK were approximately 9.2 where the 10% point is 9.8. The W(6) value of 8.6 was exceptionally low. The 6% difference between 10% point and the basis value obtained from NB and HK analysis is much better than the 14% difference determined from W(6).

In Figure 5B the ranges are much smaller for NB, HK, and HK(36) when compared to W(6) and N(6), which again shows the advantage of having more data

available by combining the L and S samples. A low value of 7 and the highest of 9.5 from NB and HK is much better than 4.2 to 9.2 from W(6) method.

The NB and HK methods were very effective as shown in Figure 6A in obtaining a 'B' basis value for crossply tension test results (manufactured by Narmco) when combining with prior available unidirectional tension data. The N(6) estimate also provided desirable coverage. Figure 6B shows less variability in basis value estimates for the NB and HK procedures when compared to the small sample W(6) and N(6) methods. The reduction in the variability in the 'B' value is the primary advantage of using NB and HK methods when compared to the W(6) and N(6) analysis results, although reducing the model identification issue is also desirable.

Figure 7A shows excellent results for the NB and HK by providing 'B' values almost equal to the required 10% point. W(6) basis value is overly conservative, and N(6) and HK(36) do not meet the required coverage rate by having values somewhat greater than the 10% point. The results demonstrate the advantage of the NB and HK methods being a nonparametric procedure in determining basis values for data obtained from relatively expensive (unidirection small sample) test methods by pooling the data with less expensive (short beam shear) coupon test results.

Figure 7B shows the relatively small variability in the 'B' estimate from NB and HK and HK(36) when compared to the small sample procedures W(6) and N(6).

The 'B' basis results from the crossply compression data using additional data obtained from a very simple and inexpensive short beam shear testing procedures are shown in Figure 8A. The NB and HK methods show less conservative 'B' estimates than W(6). The 'B' values from W(6) are somewhat conservative with

N(6) results showing good agreement with the 10% point. The variability in 'B' values in Figure 8B is similar to prior results in Figure 7B.

The crossply compression 'B' values from sample of 6 Narmco material using the additional 30 short beam shear data values from the material made by the Hercules Co. are shown in Figure 9A. The NB and HK results are about 5% below the 10% point which is not excessively low. Although excellent coverage was obtained for N(6), extensive variability in the basis value results is shown in Figure 9B.

The results in Figure 10A show, for large sample of 90 and small sample of 6 obtained from random Weibull samples, excellent agreement between the 10% point and the estimated 'B' basis values. The exception is the SHB method (see Equation (20)) which fails to provide the proper coverage. This result is probably due to the instability in estimating scale parameter $\hat{\beta}$ from a small sample of 6 values. If a lower 95% confidence was obtained for $\hat{\beta}$, then the SHB method could possibly provide satisfactory results. The similar results shown for both W(6) and W(96) show that if the Weibull analysis, even for small samples, is applied to data from the Weibull population of data, accurate 'B' basis values can be obtained that provide the coverage and also are not overly conservative. Results in Figure 10B show a value of 42 for W(6) could be obtained compared to the 85 for the 95 percentile. Although acceptable coverage for W(6) was obtained which was possibly due to the use of known Weibull data, the variability is substantial in the basis value estimate. Since it is impossible to correctly identify models for small samples, it is suggested the nonparametric methods (NB and HK) in this paper be applied since the methods do not require prior knowledge of the assumed distribution. The SHB method shows relatively low variability in estimating the basis value but does require knowledge of distribution (eg., Weibull, normal).

Conclusions

Efficient tolerance limits ('B' values) can be obtained for small samples of composite material strength data when pooled with a much larger data set of similar material but obtained from potentially different mechanical properties. Ideally, both small and large data sets should be obtained from the same test and material in order to maximize the effectiveness of the proposed methods. The proposed methodology resulted in reducing the overconservatism and excessive variability obtained from using only the smaller sample in computing the tolerance limit.

References

1. Military Handbook 17B, Army Materials Technology Laboratory, Polymer Matrix Composites, Volume I, Guidelines, 1988.

2. Neal, D. M., Vangel, M. G., and Todt, F., "Determination of Statistical Based Composite Material Properties" in Engineered Materials Handbook, Composites, C. A. Dostal, ed., American Society of Metals Press, Metals Park, Ohio, Vol. 1, 1987.

3. Neal, D. M. and Vangel, M. G., "Statistical Based Material Properties - A Military Handbook-17 Perspective", MTL TR 90-5, U.S. Army Materials Technology Laboratory, Watertown, Massachusetts 02172-0001, 1990.

4. Neal, D. M. and Spiridigliozzi, L., "An Efficient Method for Determining the 'A' and 'B' Design Allowables", ARO Report 83-2, U.S. Army Laboratory Command, Army Research Office, P.O. Box 12211, Research Triangle Park, North Carolina 27709-2211, 1983.

5. Soderquist, Joseph, National Resource Specialist for Composites (FAA), Private Conversation.

6. Adams, Donald, University of Wyoming, Private Conversation.

7. Ferguson, T. S., "A Bayesian Analysis of Some Nonparametric Problems", Annals of Statistics, Vol. 1, No. 2, 209-230, 1973.

8. Vangel, M. G., "Lower Tolerance Limits for Log-Convex Distributions", to be published.

9. Shyprykevich, P., "The Role of Statistical Reduction in the Development of Design Allowables for Composites", Test Methods for Design Allowables for Fibrous Composites: 2nd Vol., ASTM STP 1003, pp. 111-135, 1989.

10. Metallic Materials and Elements for Aerospace Vehicle Structures, MIL-HDBK-5C, 15 September 1976, pp. 9-14.

11. Conover, W. J., "Practical Nonparametric Statistics", John Wiley and Sons, 1980, p. 111.

12. Hanson, D. L. and Koopmans, L. H., "Tolerance Limits for the Class of Distributions with Increasing Hazard Rates", Annals of Mathematical Statistics, Vol. 35, 1964.

13. Siegel, S. and Tukey, J. W., "A Nonparametric Sum of Ranks Procedure for Relative Spread in Unpaired Samples", Journal of American Statistical Association, September, 1960.

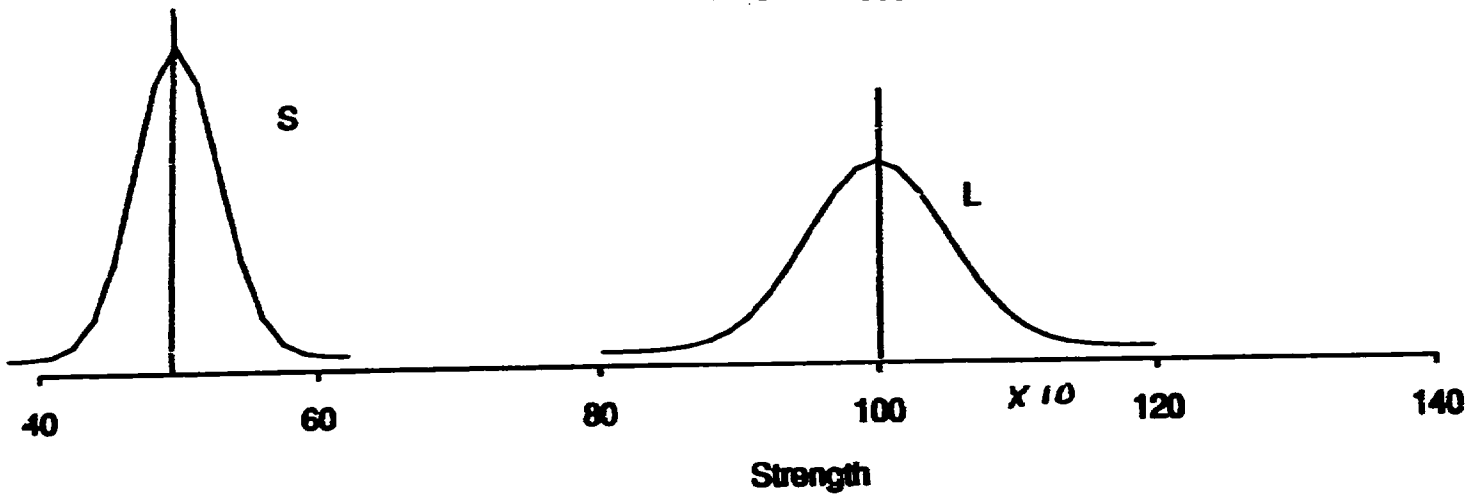
14. Reese, C. and Sorem, J. Jr., "Statistical Distribution of Mechanical Properties for Three Graphite-Epoxy Material Systems", NASA Contract Report No. 165736, 1981.

Acknowledgement

The authors wish to thank Ms. Lucy D. Ohannesian of the Materials Technology Laboratory (MTL) for preparing this paper for publication. Without her effort, this paper would not have been written.

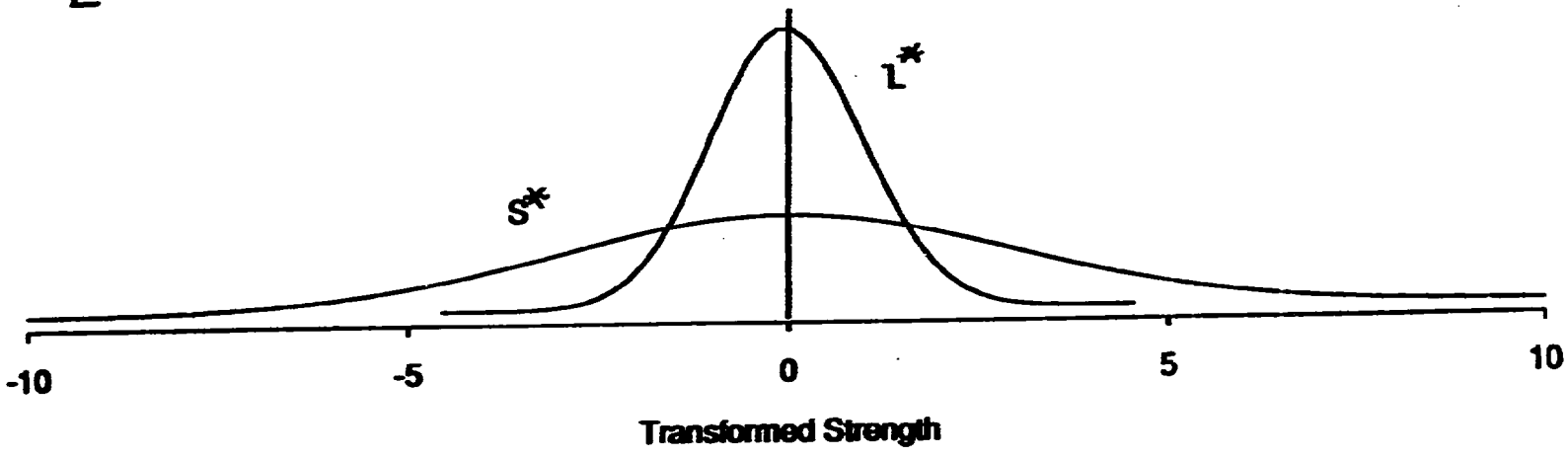
Appendix A: PDF's of Small Dataset (S) and Large Dataset (L)

Transformation



$S^* = TS(i) = (S(i) - \bar{S})/\bar{S}$

$L^* = TL(i) = (L(i) - \bar{L})/\bar{L}$



Appendix B: Criteria for Combining Data S and L

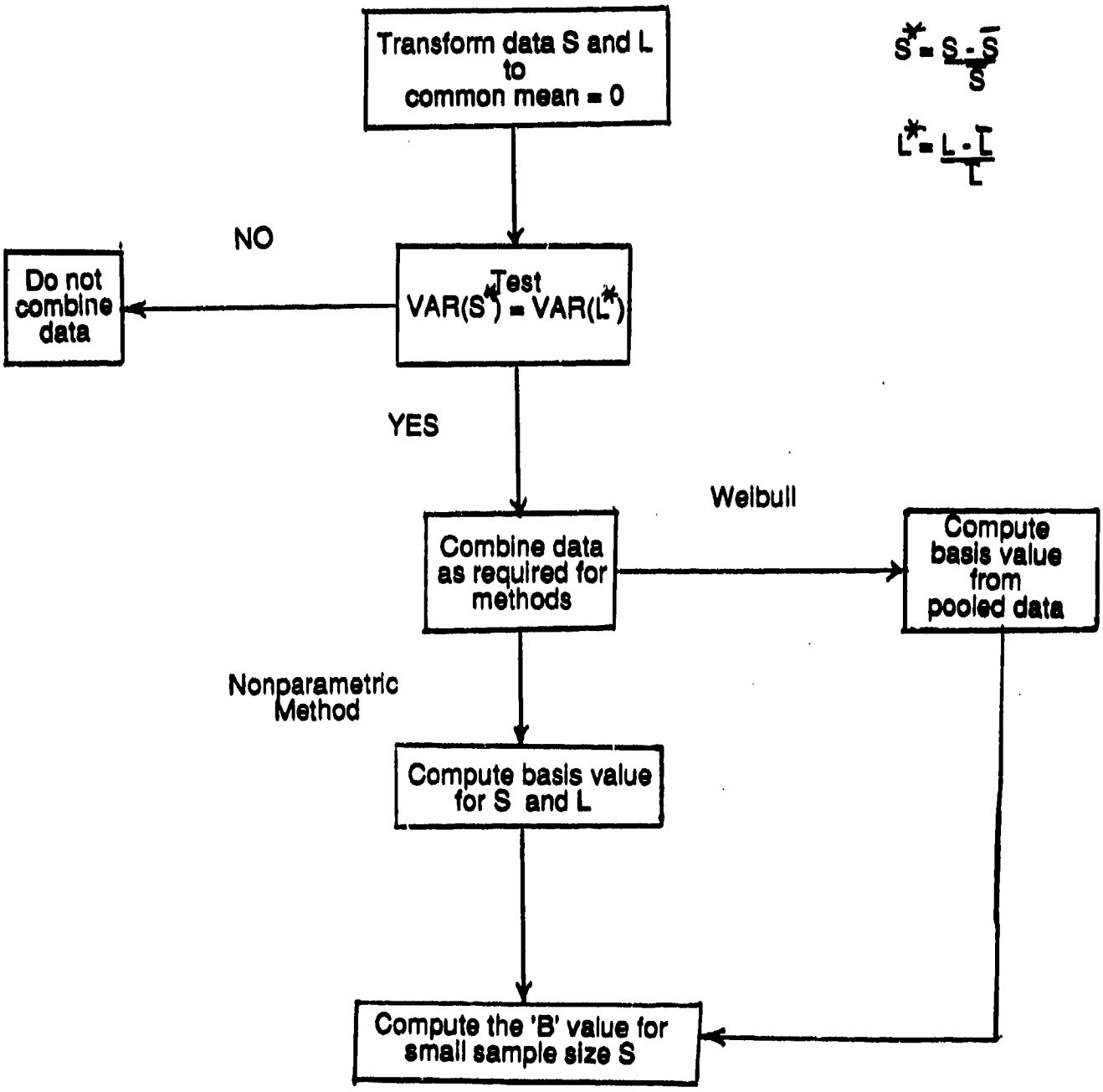
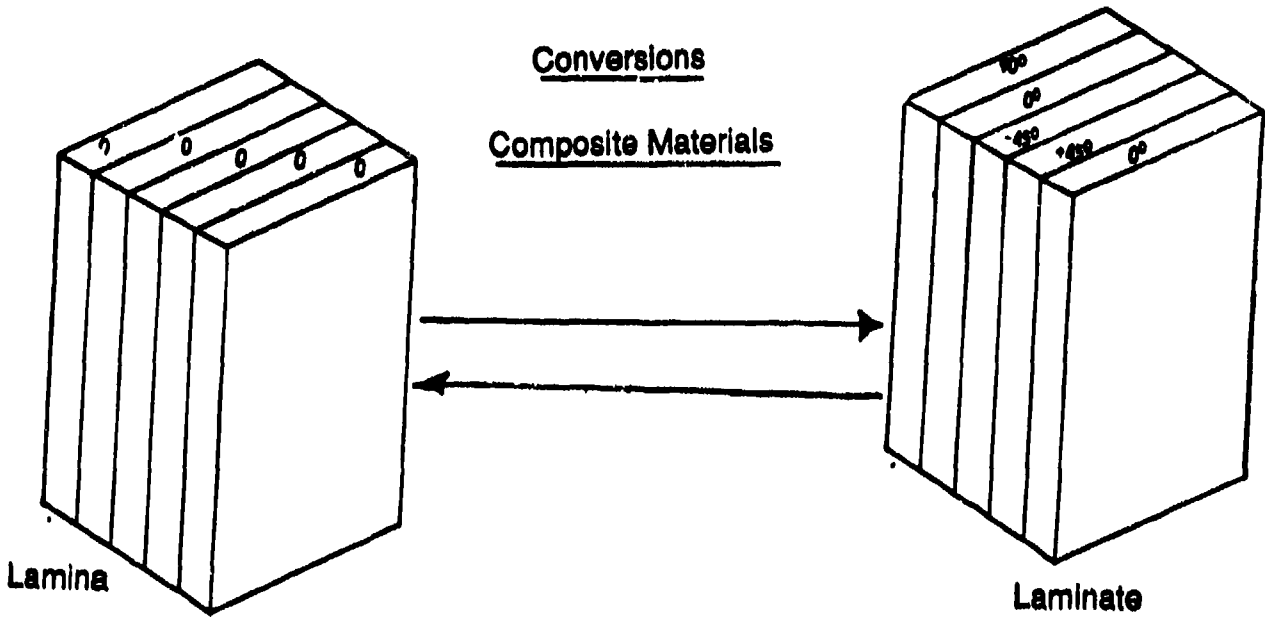


Table I: M and u Values for Bayesian Basis Value Computation

M	u	M	u	M	u	M	u
1	0.021953	51	0.044804	101	0.057686	151	0.064302
2	0.017855	52	0.045192	102	0.057856	152	0.064395
3	0.016529	53	0.045565	103	0.058023	153	0.064514
4	0.016140	54	0.045937	104	0.058188	154	0.064609
5	0.016199	55	0.046301	105	0.058352	155	0.064717
6	0.016516	56	0.046648	106	0.058517	156	0.064814
7	0.016997	57	0.046996	107	0.058670	157	0.064912
8	0.017590	58	0.047339	108	0.058837	158	0.065010
9	0.018264	59	0.047673	109	0.059006	159	0.065099
10	0.018996	60	0.048011	110	0.059156	160	0.065193
11	0.019769	61	0.048318	111	0.059313	161	0.065273
12	0.020570	62	0.048642	112	0.059454	162	0.065382
13	0.021391	63	0.048945	113	0.059619	163	0.065462
14	0.022223	64	0.049255	114	0.059761	164	0.065555
15	0.023060	65	0.049563	115	0.059914	165	0.065658
16	0.023897	66	0.049848	116	0.060051	166	0.065734
17	0.024729	67	0.050144	117	0.060192	167	0.065822
18	0.025554	68	0.050421	118	0.060344	168	0.065910
19	0.026368	69	0.050695	119	0.060480	169	0.065996
20	0.027171	70	0.050968	120	0.060628	170	0.066108
21	0.027959	71	0.051238	121	0.060754	171	0.066192
22	0.028734	72	0.051506	122	0.060883	172	0.066277
23	0.029491	73	0.051771	123	0.061031	173	0.066384
24	0.030233	74	0.052034	124	0.061162	174	0.066449
25	0.030959	75	0.052284	125	0.061292	175	0.066530
26	0.031666	76	0.052530	126	0.061420	176	0.066613
27	0.032361	77	0.052773	127	0.061547	177	0.066705
28	0.033033	78	0.053017	128	0.061679	178	0.066789
29	0.033695	79	0.053244	129	0.061802	179	0.066872
30	0.034339	80	0.053479	130	0.061933	180	0.066934
31	0.034967	81	0.053702	131	0.062065	181	0.067007
32	0.035577	82	0.053932	132	0.062179	182	0.067098
33	0.036172	83	0.054160	133	0.062293	183	0.067176
34	0.036754	84	0.054375	134	0.062430	184	0.067258
35	0.037328	85	0.054600	135	0.062553	185	0.067333
36	0.037884	86	0.054808	136	0.062667	186	0.067418
37	0.038420	87	0.055017	137	0.062784	187	0.067486
38	0.038952	88	0.055221	138	0.062894	188	0.067569
39	0.039461	89	0.055435	139	0.063010	189	0.067628
40	0.039964	90	0.055634	140	0.063128	190	0.067720
41	0.040459	91	0.055831	141	0.063245	191	0.067794
42	0.040944	92	0.056024	142	0.063344	192	0.067871
43	0.041409	93	0.056215	143	0.063459	193	0.067952
44	0.041864	94	0.056417	144	0.063550	194	0.068022
45	0.042314	95	0.056599	145	0.063666	195	0.068103
46	0.042751	96	0.056781	146	0.063763	196	0.068178
47	0.043182	97	0.056960	147	0.063899	197	0.068237
48	0.043596	98	0.057153	148	0.063985	198	0.068315
49	0.044009	99	0.057332	149	0.064101	199	0.068388
50	0.044413	100	0.057502	150	0.064197	200	0.068459

Table II: Modified Hanson-Koopmans Constants for Basis Value

n	r	s	k
2	1	2	35.177
3	1	3	7.859
4	1	4	4.505
5	1	4	4.101
6	1	5	3.064
7	1	5	2.858
8	1	6	2.382
9	1	6	2.253
10	1	6	2.137
11	1	7	1.897
12	1	7	1.814
13	1	7	1.738
14	1	8	1.599
15	1	8	1.540
16	1	8	1.485
17	1	8	1.434
18	1	9	1.354
19	1	9	1.311
20	1	10	1.253
21	1	10	1.218
22	1	10	1.184
23	1	11	1.143
24	1	11	1.114
25	1	11	1.087
26	1	11	1.060
27	1	11	1.035
28	1	12	1.010
29	1	--	1
30	2	12	1.373
31	2	12	1.344
32	2	12	1.315
33	2	13	1.270
34	2	13	1.245
35	2	13	1.221
36	2	13	1.197
37	2	13	1.174
38	2	13	1.151
39	2	13	1.129
40	2	13	1.108
41	2	14	1.083
42	2	14	1.064
43	2	14	1.045
44	2	14	1.027
45	2	14	1.009
46	2	--	1



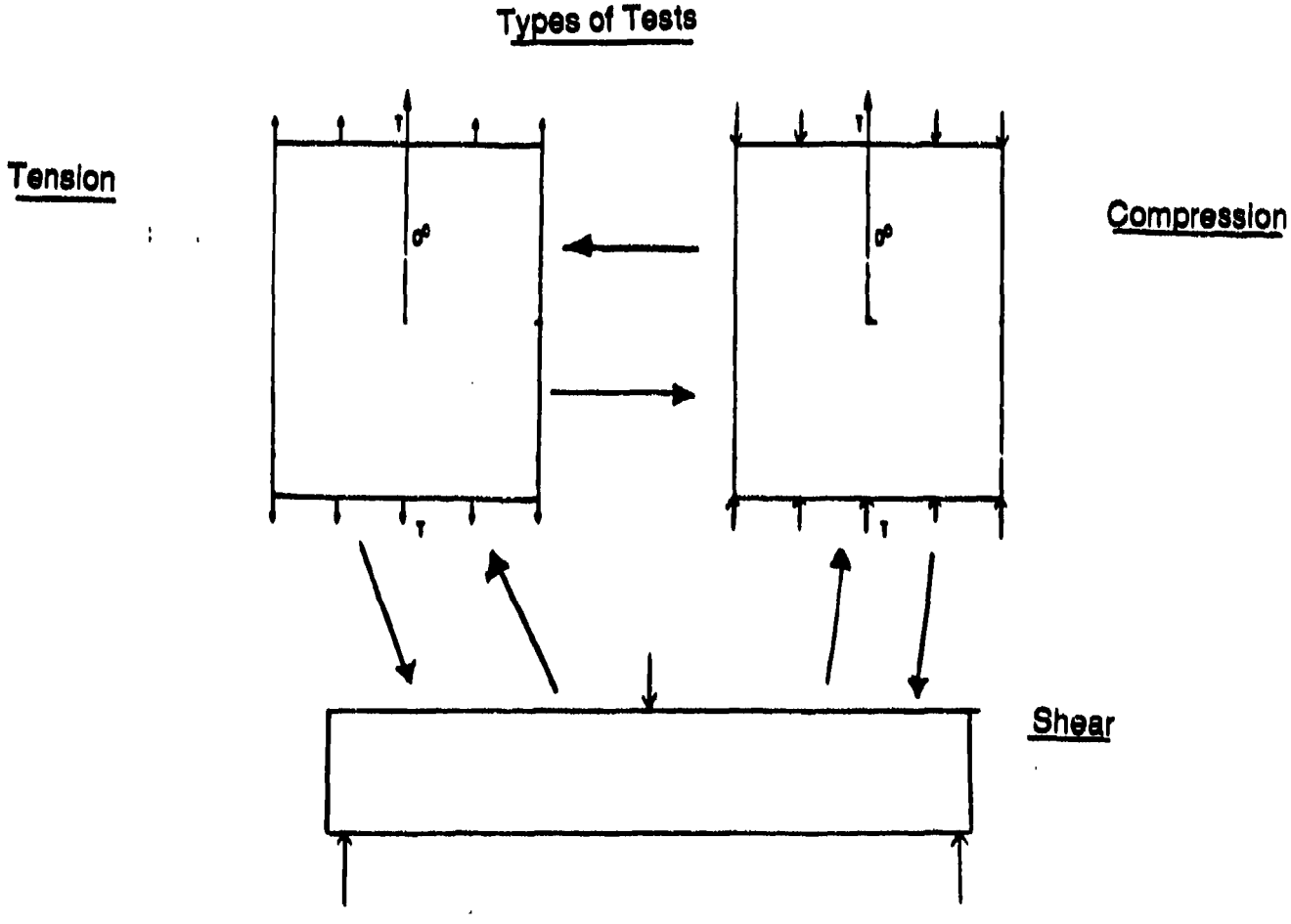


Figure 2: Composite Material Tests/

Lamina to Laminate Conversions

Tolerance Limits for a N(0,1) Population

Figure 1A

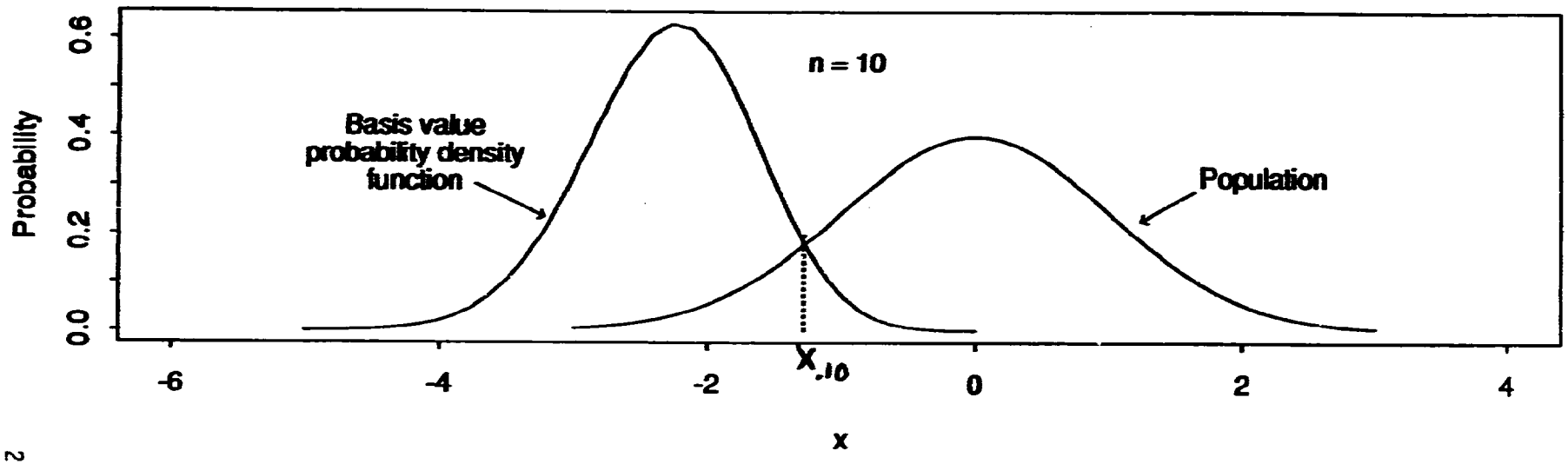
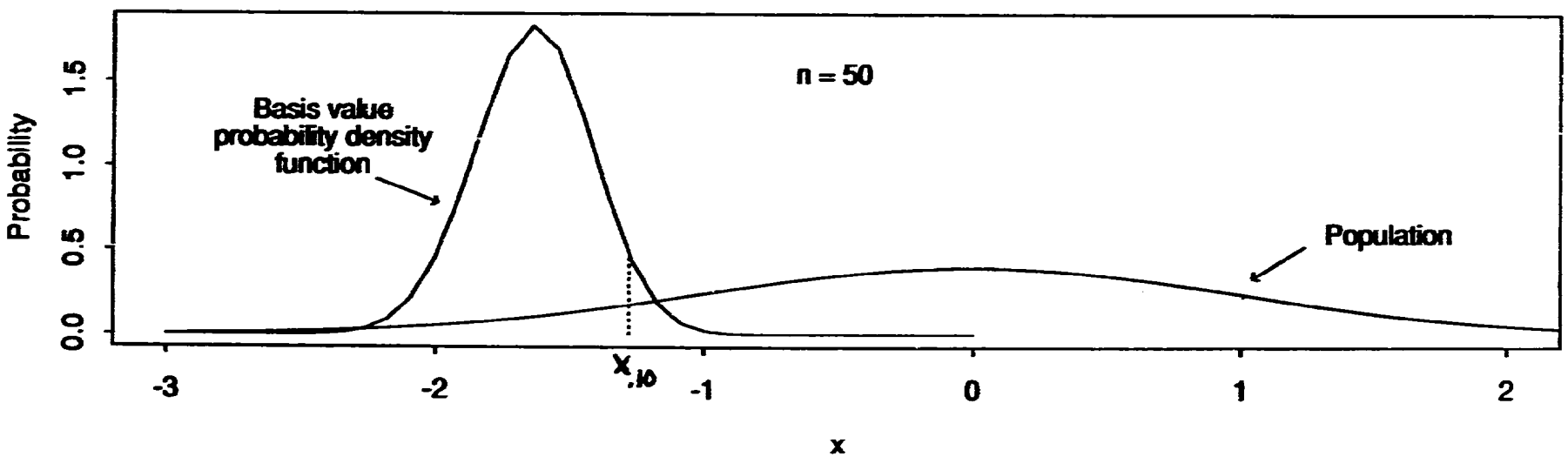


Figure 1B



Monte Carlo Study - Weibull Data

L = 30
S = 6

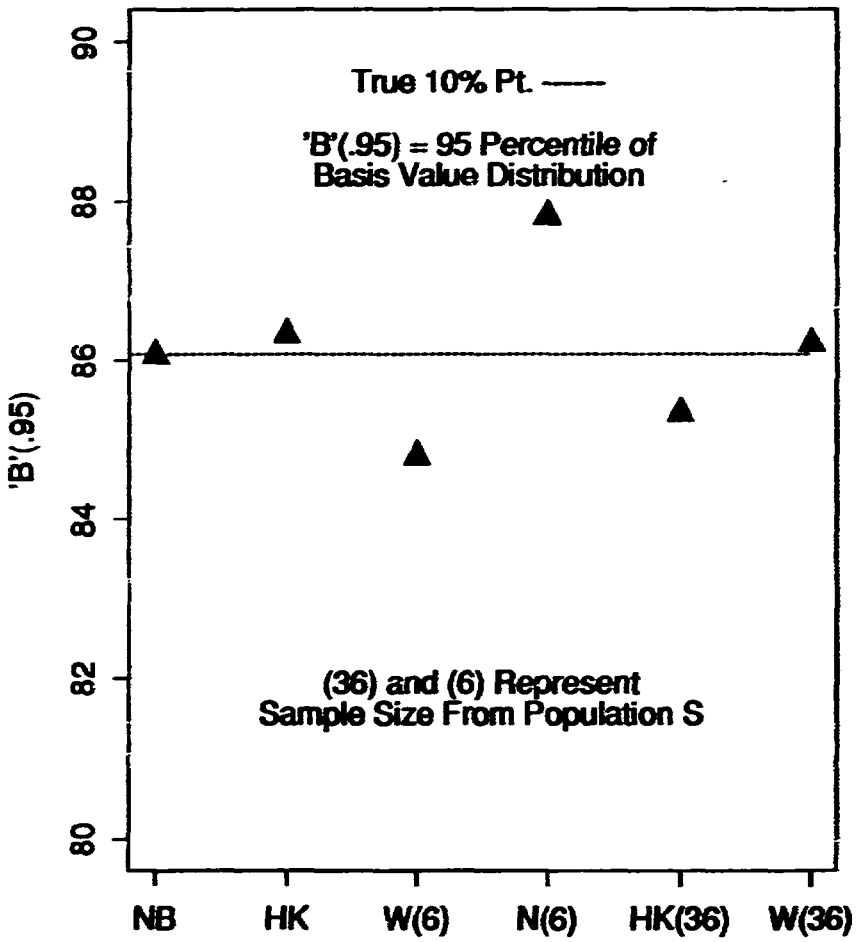


Figure 3A : Coverage Evaluation

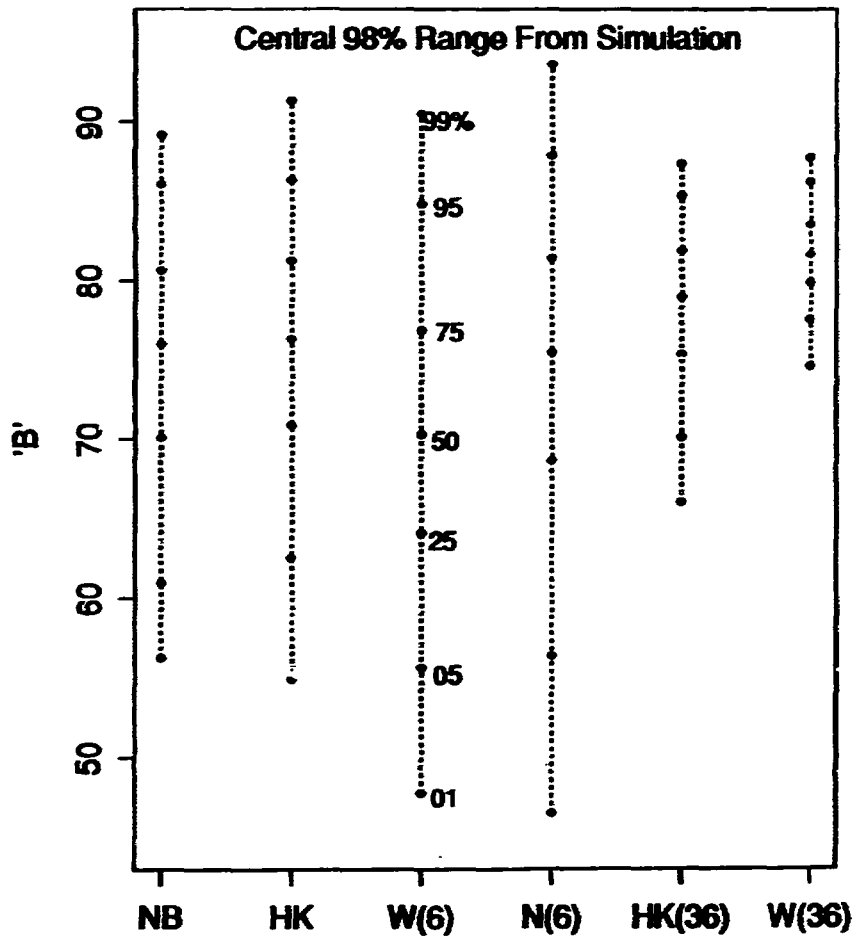


Figure 3B : The Range of 'B' Values

Monte Carlo Study - Narmco Short Beam Shear Data

L and S From Same Population
L = 30
S = 6

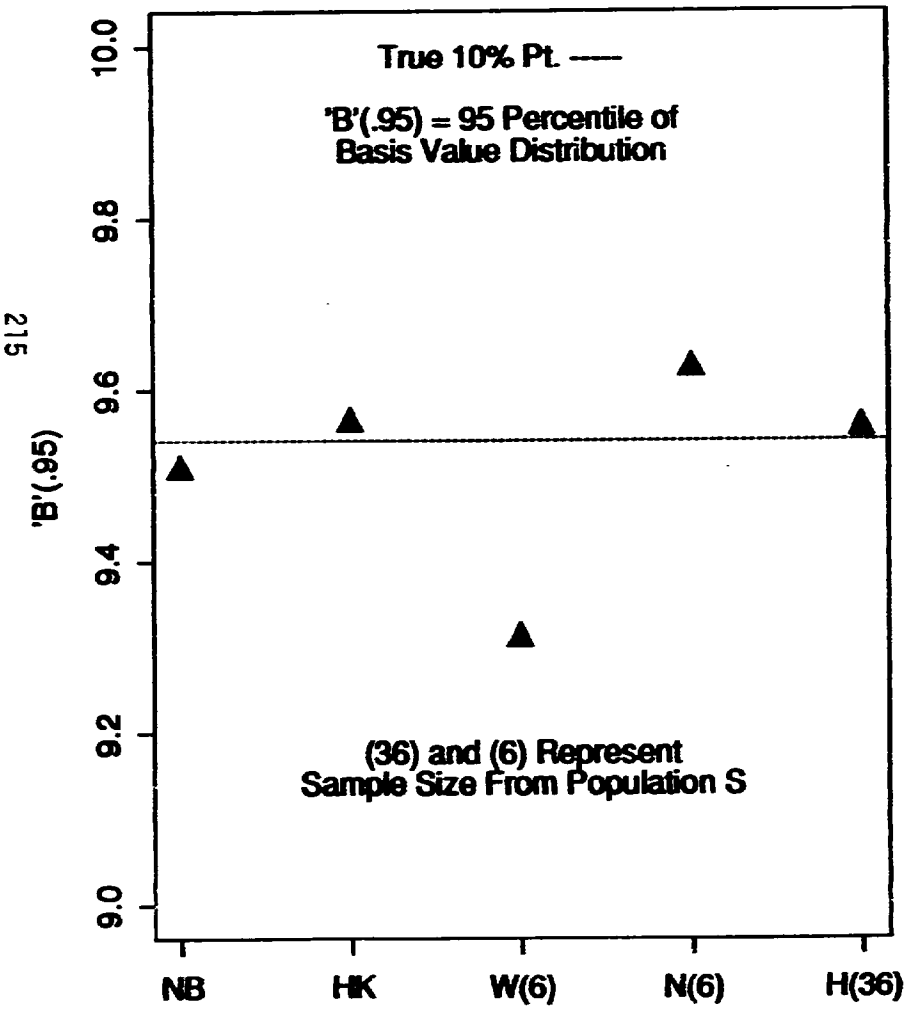


Figure 4A.: Coverage Evaluation

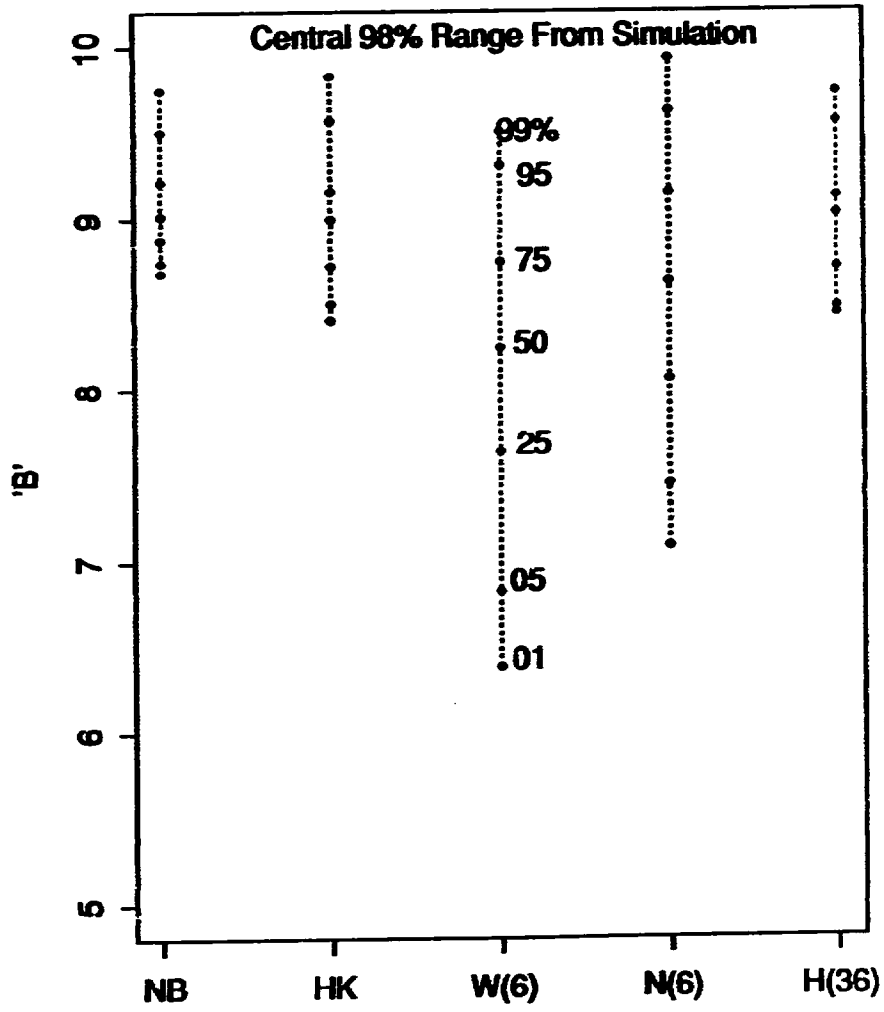


Figure 4B : The Range of 'B' Values

Monte Carlo Study - Hercules Short Beam Shear Data

L and S From Same Population
L = 30
S = 6

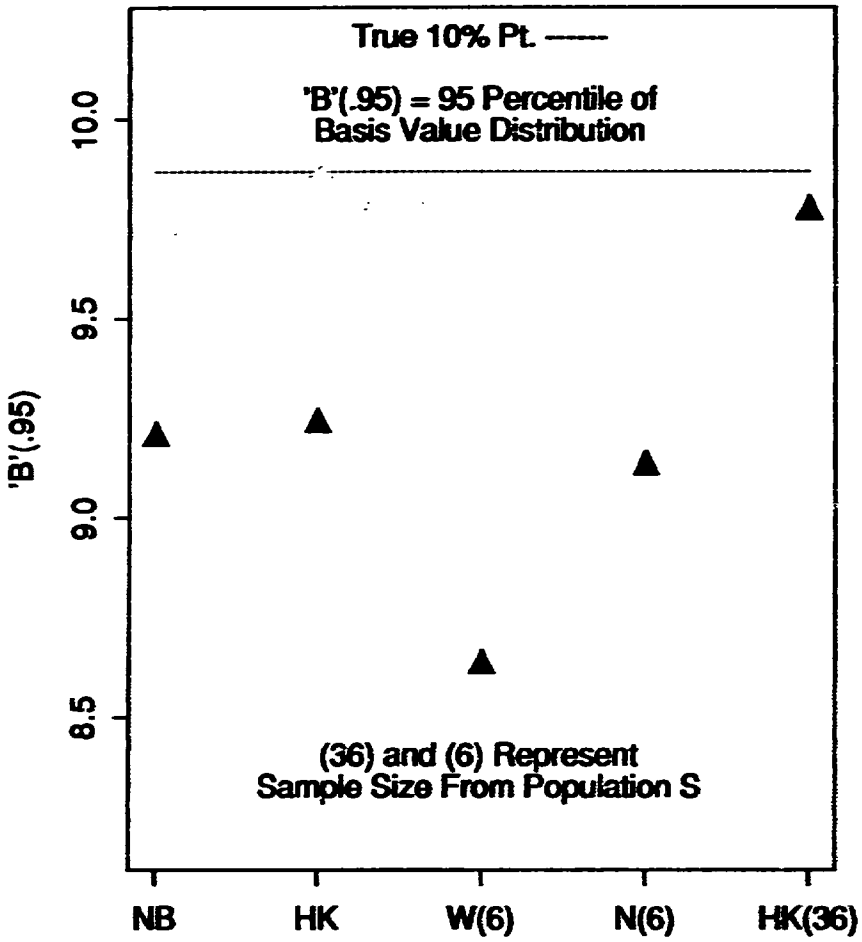


Figure 5A : Coverage Evaluation

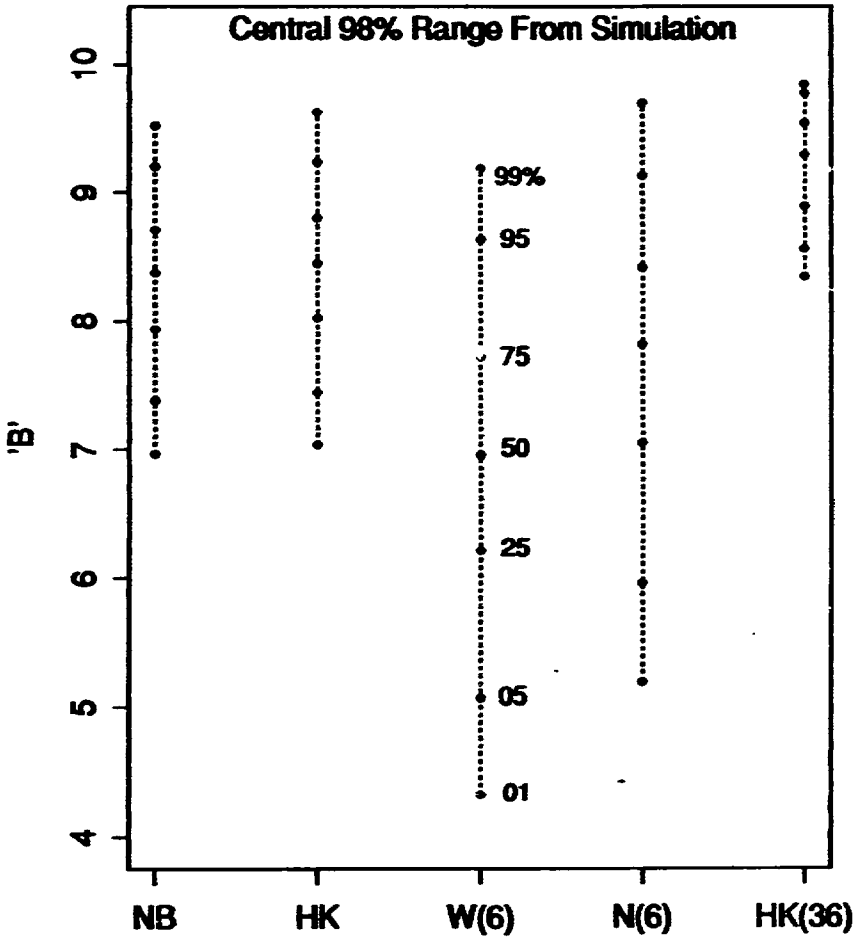


Figure 5B : The Range of 'B' Values

Monte Carlo Study - Narmco Data

L : Uni-Directional Tension Results = 30
S : Crossply Tension Results = 6

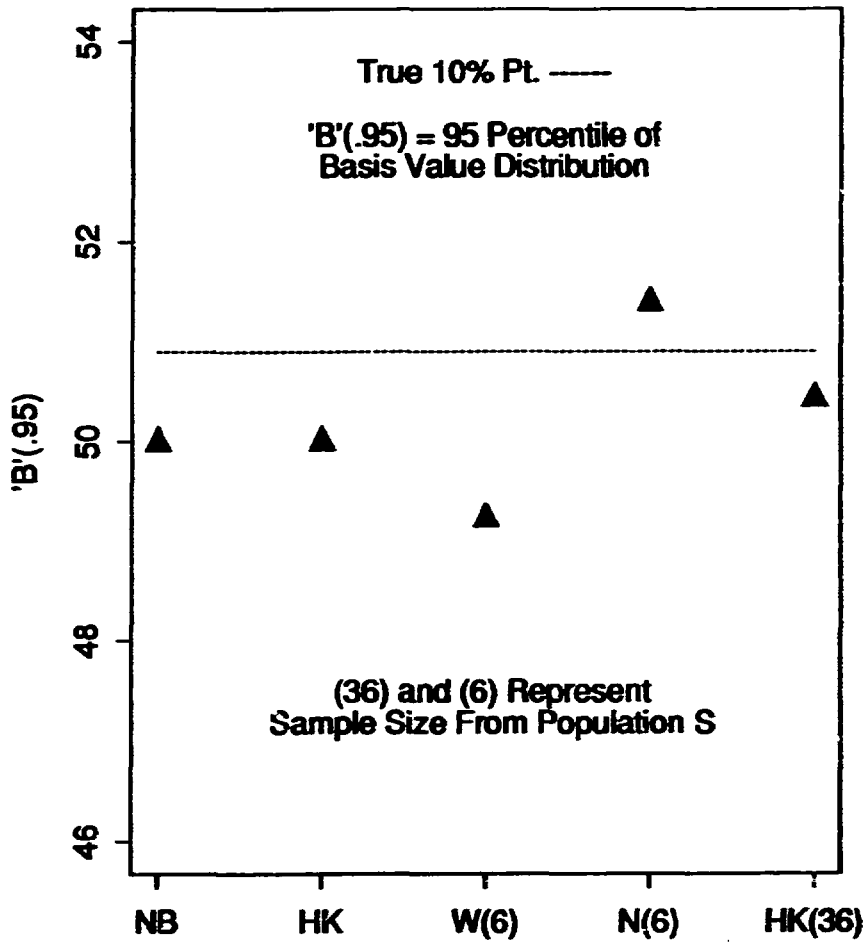


Figure 6A: Coverage Evaluation

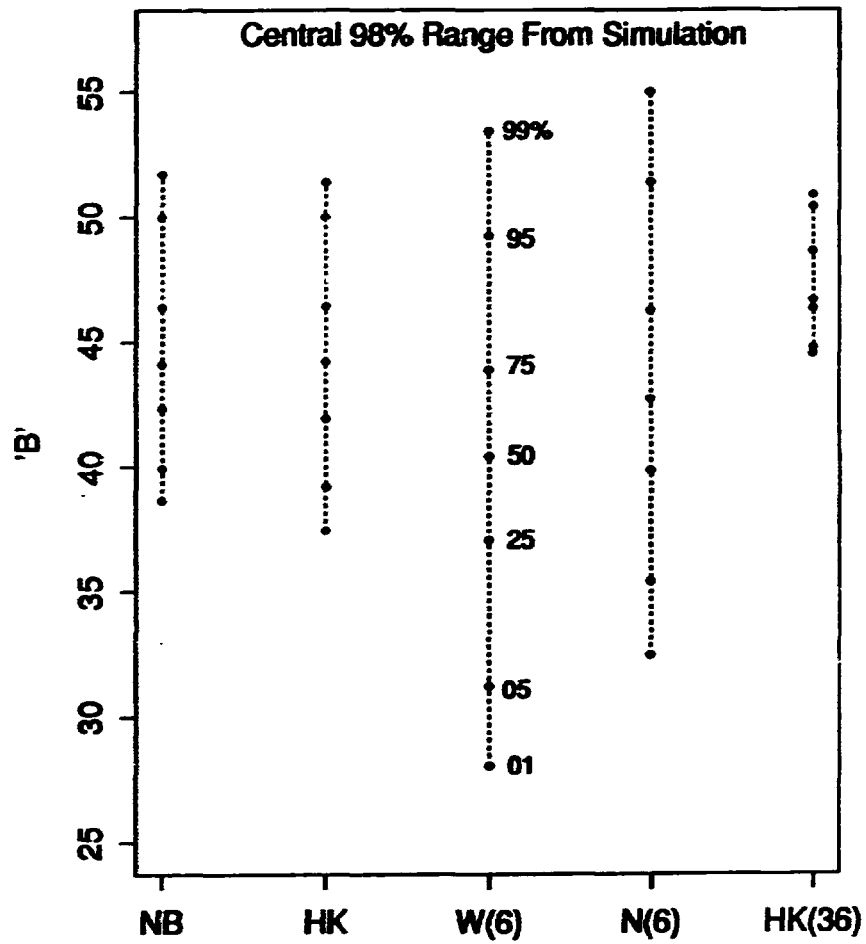


Figure 6B : The Range of 'B' Values

Monte Carlo Study - Narmco Data

L : Short Beam Shear Results = 30
S : Uni-Directional (12 Ply) Tension Results = 6

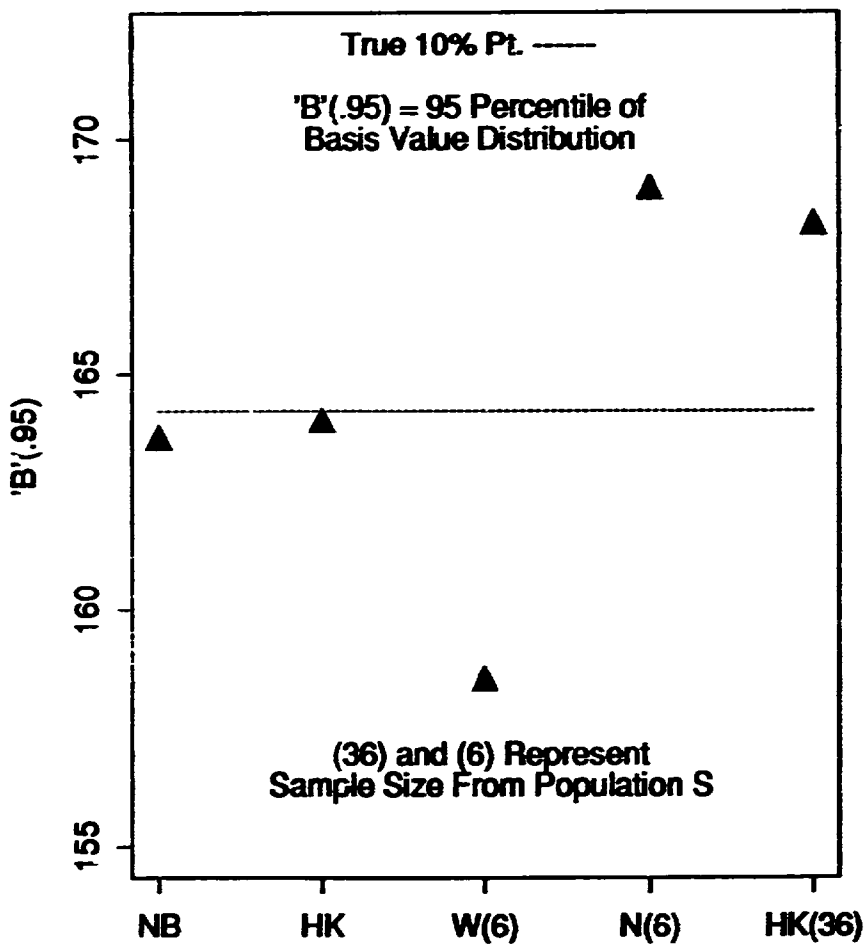


Figure 7A: Coverage Evaluation

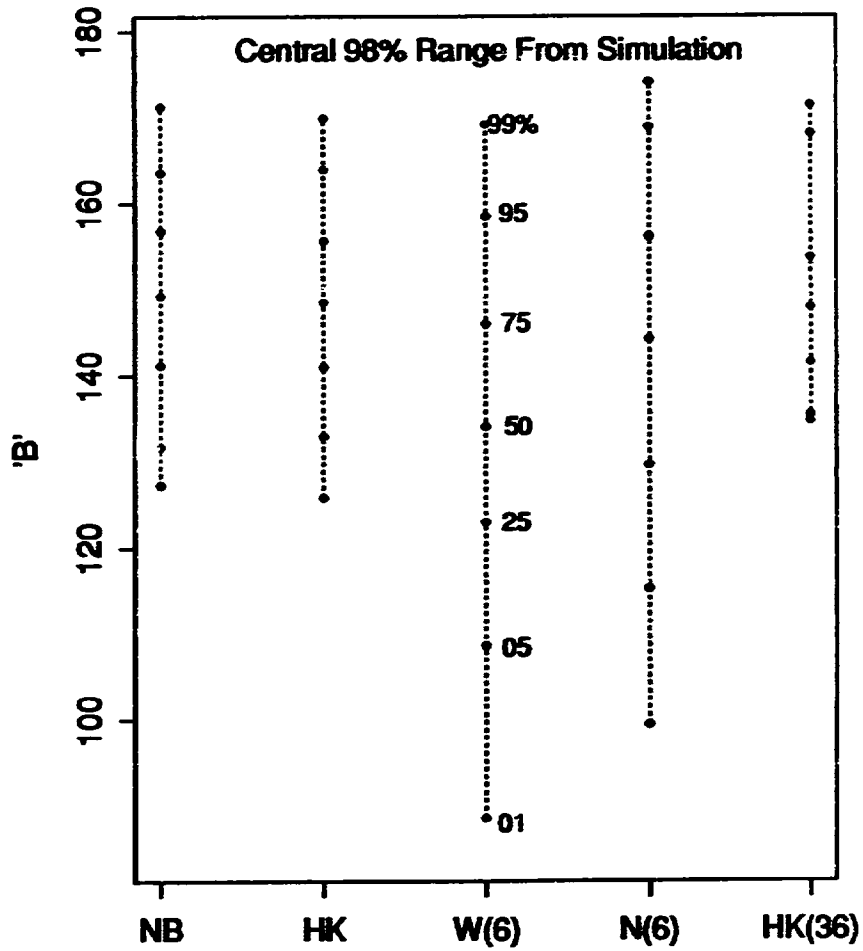


Figure 7B: The Range of 'B' Values

Monte Carlo Study - Narmco Data

L : Short Beam Shear Results = 30
S : Crossply Compression Results = 6

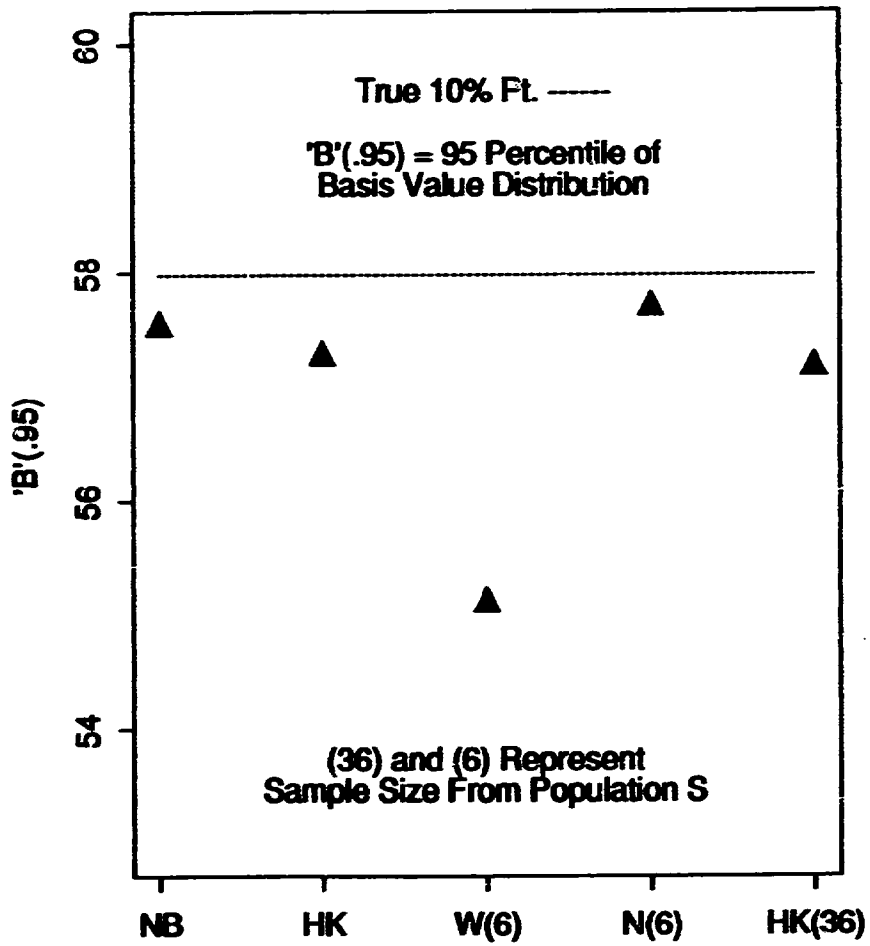


Figure 8A: Coverage Evaluation

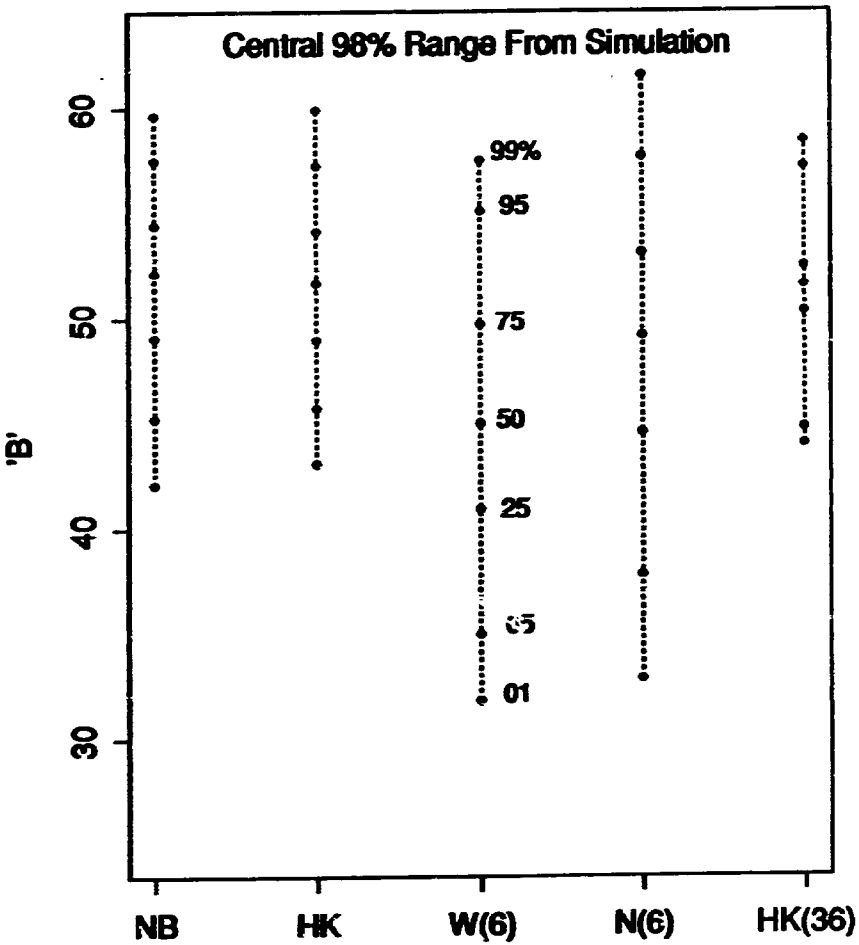


Figure 8B: The Range of 'B' Values

Monte Carlo Study - Different Companies

L : Short Beam Shear Results (Hercules) = 30
S : Crossply Compression Results (Narmco)= 6

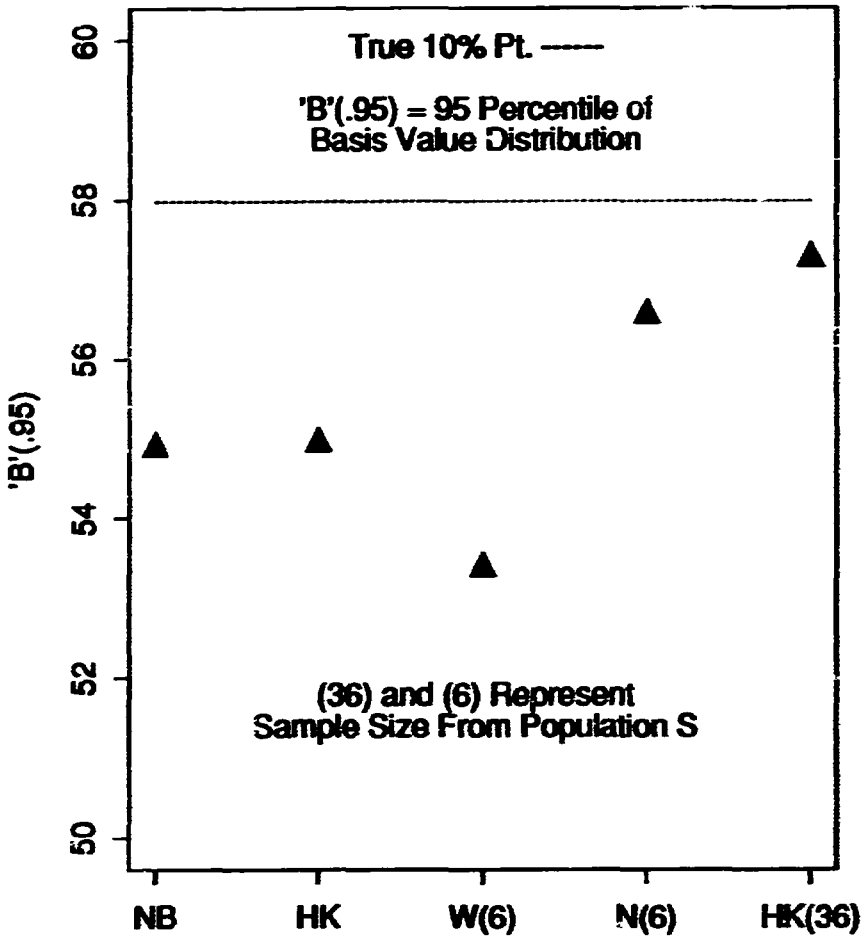


Figure 9A: Coverage Evaluation

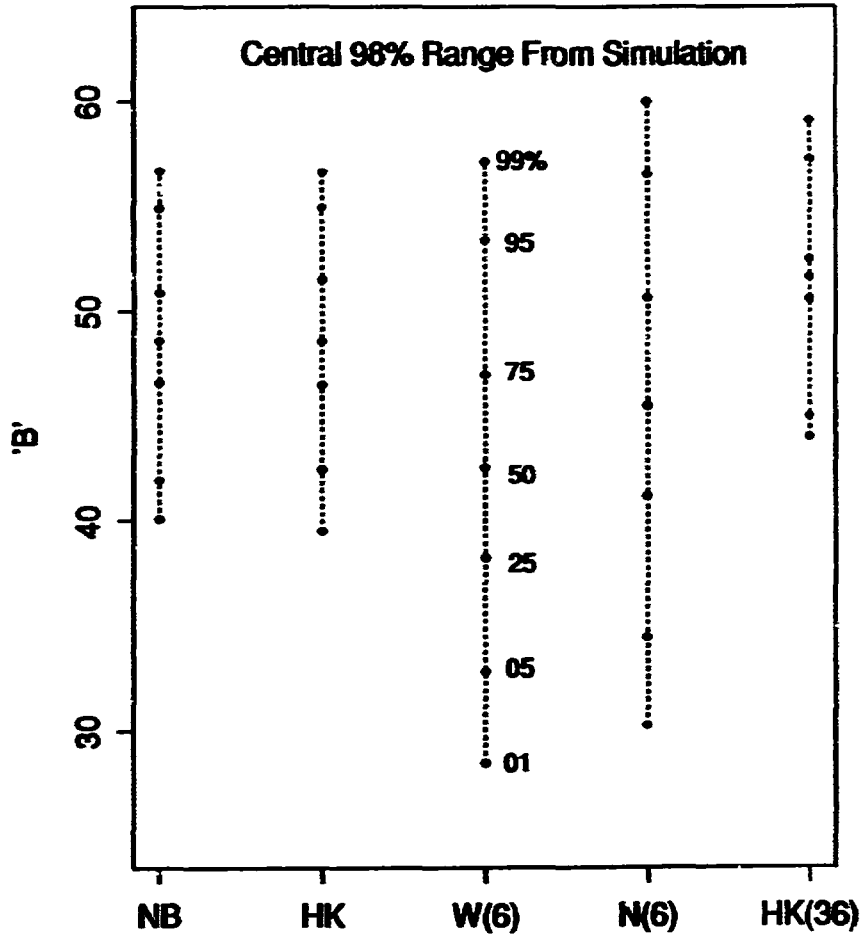


Figure 9B: The Range of 'B' Values

Monte Carlo Study - Weibull Data

L = 90
S = 6

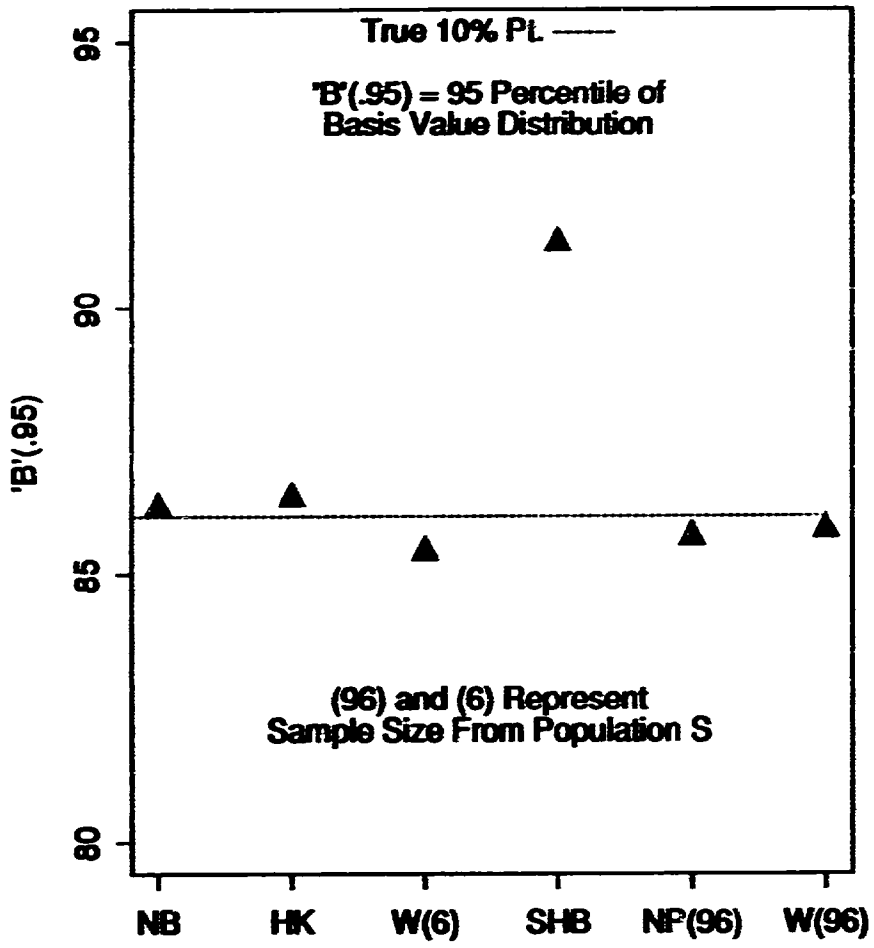


Figure 10A: Coverage Evaluation

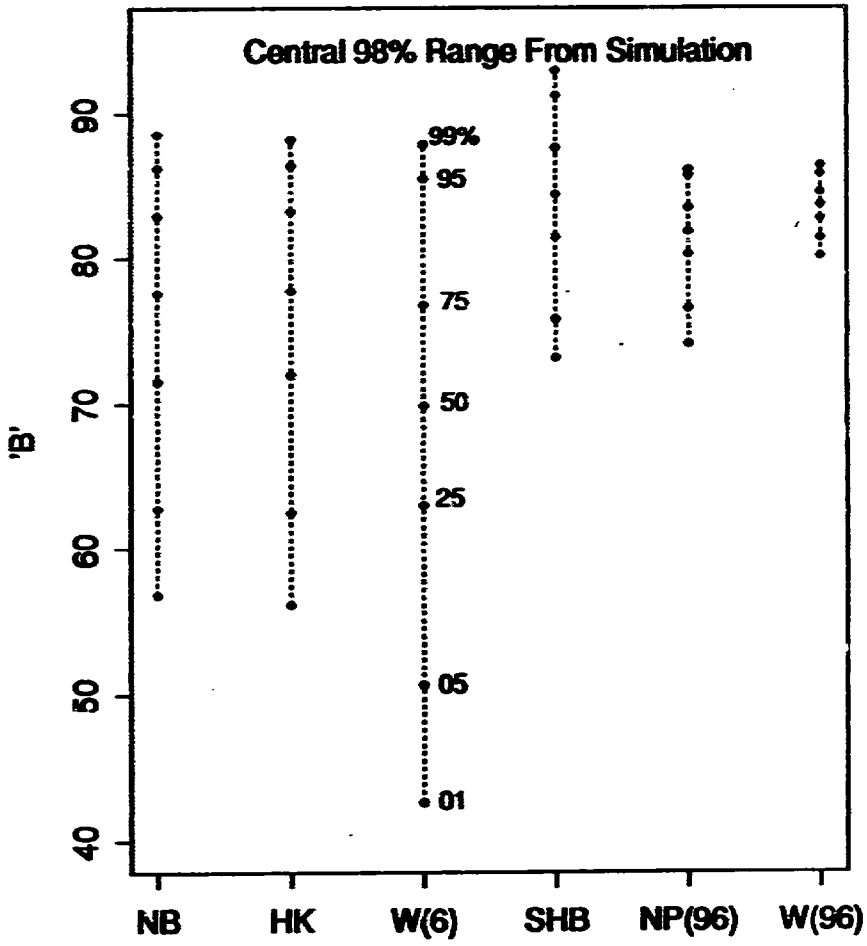


Figure 10B: The Range of 'B' Values

TARGET PRIORITIZATION TO OPTIMIZE EXPECTED UTILITY BASED ON RANDOM FIRE

Ann E.M. Brodeen

Director
U.S. Army Ballistic Research Laboratory
ATTN: SLCBR-SE-W
Aberdeen Proving Ground, MD 21005-5066
(301) 278-8947, AV 298-8947

Douglas H. Frank

Indiana University of Pennsylvania
208 Stright Hall
Dept. of Mathematics
Indiana, PA 15705-1072
(412) 357-3793

Abstract

The U.S. Army Ballistic Research Laboratory has been researching innovative methodologies directly applicable to the problem of target value analysis (TVA), *i.e.*, assigning values to targets for the purpose of developing an optimal target engagement ordering. The problems associated with assigning target values as an aid in target selection have been examined by many approaches, by both other government agencies and contractors. Here the question is examined from the standpoint of optimizing some utility function. Expected values of utility functions are derived in terms of the two chosen research parameters, rate of fire and probability of kill, using stochastic techniques. In the case of two enemy targets, an optimal ordering for all utilities investigated is obtained. Some special cases of utilities are considered and the results generalized for an arbitrary mix of targets.

Preceding Page Blank

1. Introduction

The problem of assigning values to targets as an aid in target selection has been examined for over a decade by many approaches, *e.g.*, Fire Support Mission Area Analysis, US Army Material Systems Analysis Activity military worth study, classification tree methodology (another approach being pursued by the US Army Ballistic Research Laboratory (BRL)). During the summer of 1988, Dr. Douglas H. Frank, Associate Professor, Department of Mathematics, Indiana University of Pennsylvania, worked with Ann E.M. Brodeen, one of the BRL's principal target value analysis (TVA) investigators, to consider a probabilistic approach to the problem. (Dr. Frank was assigned to the BRL under the sponsorship of the US Army Summer Faculty Research and Engineering Program.)

Target values are assessments keyed to the enemy's perception of the functions of its assets; TVA is the methodology which identifies potential high value target sets, *i.e.*, assets which the enemy threat commander requires for the successful completion of his mission, within the given tactical scenario. These targets, if successfully countered, can provide the friendly force with a tactical opportunity. Although the TVA process may include complex algorithms, it should be simple enough for the user, *i.e.*, the soldier, to understand. Simply put, he must be able to influence the process in order to meet the specific needs of his commander. For the field artillery to remain responsive, the soldier must be able to change target priorities as quickly as the tactical situation changes and be able to interpret the overall impact that such changes may have on the outcome of the operation.

Although TVA is a very subjective issue, the intent of this research was to show that assigned target values can be based on mathematical models. Two objectives were defined for the proposed study: 1) define a value for each enemy target in a target array such that a target engagement sequence can be determined, and 2) evaluate the target engagement sequence from the standpoint of optimizing an expected utility function based on a desired tactical outcome.

Details of the BRL's probabilistic approach to target value analysis based on a random battle scenario are outlined in this paper.

2. The Battle

In our earlier research, we considered target engagement orderings to maximize the outcomes of a simple battle in which both the friendly fire unit and the enemy targets fired simultaneously. Here, battle outcome probabilities and optimal engagement orderings are considered in a manner similar to that of the simple battle [1,2].

The random battle is between a single friendly fire unit and a group of enemy targets. Each enemy target as well as the friendly fire unit fire independently and at a rate of fire that is exponentially distributed with different mean rates and removal, *i.e.*, kill, probabilities. The kill probabilities are constant from volley to volley for both the fire unit and each enemy target.

2.1 Parameter Selection

Subject matter discussions held with MAJ William T. Dougherty, Field Artillery Coordination Officer assigned to the BRL when the TVA probabilistic approach was initiated, led to the selection of target vulnerability and target threat as the parameters of interest. When considering an enemy target's value, it is natural to characterize this value by the ability of the friendly fire unit to remove the enemy target within some time frame, *i.e.*, target vulnerability, as well as by the ability of the enemy target to achieve its objective within that same time frame, *i.e.*, threat. Removal of the enemy target is considered to be either its complete destruction or the infliction of a level of damage severe enough to abate the target's contribution to the enemy force given some particular tactical scenario. The objective of the enemy target might also be either the destruction of the friendly fire unit or the infliction of a severe level of damage upon it. (It should be noted that the definitions of the parameters developed by the principal investigators are in the interest of the research and may not be in accordance with those of the field artillery community.)

2.2 Two Target Battle

We first derive results for a battle with two enemy targets and then extend the results to T targets.

Consider the following parameters for targets $i = 1, 2$:

P_{B_i} = probability of friendly fire unit removing target i

P_{R_i} = probability of friendly fire unit being removed by target i

β_i = mean rate of fire of friendly fire unit against target i

ρ_i = mean rate of fire of target i against friendly fire unit

$B_i = \beta_i P_{B_i}$ = friendly fire unit firepower against target i
(vulnerability of target i)

$R_i = \rho_i P_{R_i}$ = target i firepower against friendly fire unit
(threat of target i)

We first show that the results of the classic stochastic duel can be extended to a battle between a single friendly fire unit and two enemy targets [3,4]. We assume target 1 is engaged until it is removed before target 2 is engaged. The battle concludes when either the friendly fire unit is removed or it has removed both targets. For our purposes we define a victory as the removal of both targets regardless of whether or not the fire unit survives.

If a target w has a mean rate of fire r and a kill probability p, then the probability of not killing the target in some time h is

$$Q_w(h) = \exp^{-rhp} . \tag{1}$$

Eq. (1) will hereafter be referred to as **Lemma 1.** The following proof is offered.

Let X be the number of volleys fired during some time h. Then X has a Poisson distribution with mean rh.

$$\begin{aligned} Q_w(h) &= \sum_{x=0}^{\infty} \frac{e^{-rh}(rh)^x}{x!}(1-p)^x \\ &= e^{-rh} e^{rh(1-p)} \\ &= e^{-rhp} . \end{aligned} \tag{2}$$

Our **Lemma 2** states that in a battle with two enemy targets, the probability of the friendly fire unit removing target 1 before being removed is

$$P[1 \mid NOT] = \frac{B_1}{B_1 + R_1 + R_2} . \tag{3}$$

Lemma 2 is proved as follows. Divide the time of battle into units of length h and consider the absorbing Markov process formed [5]. From this process the probability of the event occurring is

$$\begin{aligned} P_h[1 \mid NOT] &= \left([1-Q_{B_1}(h)]\, Q_{R_1}(h)\, Q_{R_2}(h)\right) \sum_{s=0}^{\infty} \left(Q_{B_1}(h)\, Q_{R_1}(h)\, Q_{R_2}(h)\right)^s \\ &= \frac{[1-Q_{B_1}(h)]\, Q_{R_1}(h)\, Q_{R_2}(h)}{1-(Q_{B_1}(h)\, Q_{R_1}(h)\, Q_{R_2}(h))} . \end{aligned} \tag{4}$$

Applying **Lemma 1**, we obtain

$$P_h[1 \mid NOT] = \frac{e^{-h(\rho_1 P_{R_1} + \rho_2 P_{R_2})} - e^{-h(\beta_1 P_{B_1} + \rho_1 P_{R_1} + \rho_2 P_{R_2})}}{1 - e^{-h(\beta_1 P_{B_1} + \rho_1 P_{R_1} + \rho_2 P_{R_2})}} . \tag{5}$$

$$P[1 \mid NOT] = \lim_{h \to 0} P_h[1 \mid NOT] = \frac{\beta_1 P_{B_1}}{\beta_1 P_{B_1} + \rho_1 P_{R_1} + \rho_2 P_{R_2}} = \frac{B_1}{B_1 + R_1 + R_2} \tag{6}$$

In a battle between the friendly fire unit and target 2, the probability of destroying the target is

$$P[2] = \frac{B_2}{B_2 + R_2} . \tag{7}$$

Lemma 3 is a well-known result [4].

Theorem 1 defines w, the number of targets destroyed. The density of w is

$$f(0) = \frac{R_1 + R_2}{B_1 + R_1 + R_2} , \quad f(1) = \frac{B_1 R_2}{(B_1 + R_1 + R_2)(B_2 + R_2)} ,$$

$$f(2) = \frac{B_1 B_2}{(B_1 + R_1 + R_2)(B_2 + R_2)} \tag{8}$$

The event $w = 0$ is the complement of the event stated as **Lemma 1.** Because the battle is a Markov process we have

$$f(1) = P[1 \mid NOT] \cdot (1 - P[2]) \quad \text{and} \quad f(2) = P[1 \mid NOT] \cdot P[2] . \tag{9}$$

The result follows from **Lemmas 2** and **3**.

2.3 Multi Target Battle

Suppose there are T targets which are engaged in numerical order. We extend the definitions of B_i and R_i for $i = 1, 2, ..., T$.

Define $$S_i = \sum_{j=1}^{T} R_j, \quad S_0 = 0, \quad \text{and} \quad B_0 = 1. \tag{10}$$

The proof of Theorem 1 can be easily generalized. If w is the number of targets removed in a battle with T targets, Theorem 2 states that the density of w is

$$f(w) = \begin{cases} \left(\prod_{i=0}^{w} \frac{B_i}{B_i + S_i} \right) \cdot \frac{S_{w+1}}{B_{w+1} + S_{w+1}}, & \text{if } w < T \\ \prod_{i=1}^{T} \frac{B_i}{B_i + S_i}, & \text{if } w = T \end{cases} \tag{11}$$

3. Evaluation Criteria

In general, decision makers such as gamblers, baseball managers, insurance companies, and others engage in what is colloquially refered to as "playing the percentages", characterized by a preference for the optimal act that yields the greatest long-run average profit. That is, the optimal act is the one that would result in the largest long-run average profit if the same decision were to be made repeatedly under the same conditions; as the number of repetitions becomes large, the observed average payoff approaches the theoretical expected payoff. However, many important decisions are made under unique sets of conditions, and it may not be realistic to think in terms of many repetitions of the same decision situation. Indeed, many of the field artillery commander's most important decisons are unique, high-risk situations, whereas less important, routine decisions are ones that may be delegated to subordinates. Therefore, it is useful to have an apparatus for dealing with one-time decision making.

Utility theory provides such an apparatus, as well as providing a logical method for repetitive decision making. The term "utility" as conceived by Von Neumann and Morgenstern [1947] is a measure of value used in the assessment of situations involving risk, which provides a basis for decision making. Different sets of axioms that imply the existence of utilities with the property that expected utility is an appropriate guide for consistent decision making are presented in Von Neumann and Morgenstern [1947], Savage [1954], Luce and Raiffa [1957], Pratt, Raiffa, and Schlaifer [1965], and Fishburn [1970].

3.1 Construction of Utility Functions

The different algorithms for determining target values do not always yield the same target engagement ordering. This poses the obvious question of which approach should be used. The desired approach would be the one whose target ordering provides the "best" result. However, if "best" is interpreted as "total victory", and if T, the total number of enemy targets, is large, then "victory" for a single friendly fire unit would quite likely be a rare event. Thus, additional criteria shall be considered for assessing "best" results.

Recall that the overall objective is to assign a value to each enemy target to determine the order in which to engage the targets. This order should be chosen to maximize some desired result of the battle. Therefore, consider a utility function, U, of the number of targets removed, W, by the friendly fire unit during the battle. This function should depend on the battlefield scenario as well as the desired battle objective of the friendly fire unit. Assume that U(W) will be non-decreasing, U(0) = 0 and U(T) = 1.

Generally, U(H) is assigned over a continuous range of possibilities; however, special liberty has been taken in the analysis of the utility functions discussed below. Since each of these utility functions is based on the mathematical model's assumption that an enemy target either survives or is completely removed from the battle, these functions are evaluated only at discrete points.

3.2 Utility Based on Total Victory

If the goal is to remove all enemy targets, then we wish to maximize f(T). This is an extreme example of a convex funtion.

$$U1(W) = \begin{cases} 0, & \text{if } W < T \\ 1, & \text{if } W = T. \end{cases} \tag{12}$$

Note that $f(T) = \dfrac{B_1 \cdots B_T}{(B_1 + S_1) \cdots (B_T + S_T)}$.

Since the numerator is the same for all permutations, we must minimize the quantity,

$$\prod_{i=1}^{T} (B_i + S_i).$$

3.3 Utility Based on the Number of Hits

If hitting all targets is not essential, and all targets are equally important, we may wish to maximize the number of targets removed.

$$U2(W) = \frac{W}{T}. \tag{13}$$

For utility two we wish to maximize $\sum_{w=1}^{T} \frac{W}{T}$, for $f(w)$.

3.4 Utility Based on a Reduction in Threat

Lemma 2 shows that the combined threat of an array of targets acting as a single target is the sum of their individual threats. Thus, in removing w targets, we reduce the overall threat from S_1 to S_{w+1}

$$U3(W) = 1 - \frac{S_{W+1}}{S_1}, \tag{14}$$

where we define $S_{T+1} = 0$. Note that $U3(W)$ depends on prioritizing the targets according to VAL 3, whereas utilities 1, 2 and 4 are not dependent on any specific target value algorithm.

3.5 Inflective Shaped Utility

In many battles the enemy can be halted when it loses only a small proportion of its forces. In these cases an inflective utility function seems appropriate. For convenience we consider an extreme example of an inflective type function.

$$U4(W) = \begin{cases} 0, & \text{if } W \leq .3T \\ 1, & \text{if } W > .3T \end{cases} \tag{15}$$

4. Values Based on Two Targets

For two enemy targets there are only two possible permutations: 1,2 or 2,1. The utility U will be determined by the utility for target 1, that is, U(1). We will derive a generic target value algorithm for all utility functions and examine the four special cases from Section 3.

4.1 General Optimal Results

Suppose $U(1) = c$ is the same for either target permutation. Let $E_1[U]$ and $E_2[U]$ be the expected utilities for the orderings 1,2 and 2,1, respectively. From Theorem 1

$$E_1[U] = \frac{B_1(cR_2 + B_2)}{(B_1 + R_1 + R_2)(B_2 + R_2)} \quad \text{and}$$

$$E_2[U] = \frac{B_2(cR_1 + B_1)}{(B_2 + R_1 + R_2)(B_1 + R_1)} . \tag{16}$$

Setting $E_1[U] > E_2[U]$ and simplifying, we obtain an inequality concerning the threat and vulnerability for each target.

We now state Theorem 3. $E_1[U] > E_2[U]$ if and only if

$$\frac{B_1(B_1 + R_1)}{(B_1 + S)(cR_1 + B_1)} > \frac{B_2(B_2 + R_2)}{(B_2 + S)(cR_2 + B_2)} , \tag{17}$$

where $S = R_1 + R_2$ is a constant representing total enemy firepower.

Theorem 3 gives us a generic target value algorithm for utility c. The value of a target with threat R and vulnerability B relative to a utility c is

$$\frac{B(B + R)}{(B + S)(cR + B)} . \tag{18}$$

The shortcomings of this definition are the presence of S which depends on the entire array of targets and the limitation to utilitites which are independent of target orderings. One approach is replace S by 2R (where S is based upon knowledge about target 1 only). We do not recommend this since optimality is not guaranteed. We will see in some of the special cases that an equivalent value without S can be obtained.

To overcome the second objection, if the value of U(1) for an ordering can be expressed in terms of the second target, we can redefine value by interpreting c for a target as its utility value when it is second.

4.2 Value for Utility One

For utility 1, c = 0, therefore, the generic value algorithm reduces to $\frac{B+R}{B+S}$,

since $\frac{B_1+R_1}{B_1+S} > \frac{B_2+R_2}{B_2+S}$ if and only if $R_1(B_1+R_1) > R_2(B_2+R_2)$.

Therefore, the definition of a value for utility one is

$$VAL1 = R(B+R). \tag{19}$$

In the two target value, the order of battle based on VAL 1 will maximize E [U 1].

4.3 Values for Utilities Two and Three

For utility 2, c = 1/2. No simplified value can be found for this case.

$$VAL2 = \frac{B(B+R)}{(B+S)(1/2R+B)}. \tag{20}$$

Orders based on VAL 2 will maximize E [U 2]. VAL 2 can also be written as

$\frac{B}{B+S}\left(1+\frac{R}{R+2B}\right)$ or $\frac{2B}{R+2B}\left(\frac{1}{1+\frac{S-R}{B+R}}\right)$ but neither of these forms seems to lead toward the elimination of S.

In the case of utility 3, the value of $c = (1-R)/S$, therefore, the generic target value can be written as the following definition

$$VAL3 = \frac{BS(B+R)}{(B+S)(R^2-R+BS)}. \tag{21}$$

No equivalent form without S is apparent.

4.4 Value for Utility Four

Given two targets for utility four, $c = 0$. Thus, we are interested in minimizing the probability of no hits. This occurs when $B_1 > B_2$.

$$VAL4 = B. \tag{22}$$

5. Concluding Remarks

Each of the values examined has interesting features. If the desired battle objective is to remove as many targets as possible, then VAL 1 appears to be best. If the goal is to inflict as much damage as possible on the enemy, as measured by U3, then VAL 3 seems most appropriate. Unfortunately, VAL 4, which almost always gives optimal results when considering a complete victory, does not perform well for other considerations.

One of the obvious needs is a method for acquiring accurate values for the vulnerability (P) and threat (R) parameters. These values not only depend on inherent target characteristics, but also the battlefield conditions and the missions assigned to the friendly fire unit. Initially, the literature could be perused for probabilities of hit and kill. One promising statistical approach would be to utilize the CART software, with input in the form of experimental data, simulated data, and officers judgments [5].

Additional conditions for optimality of U1, U2, and U3, as well as other utility functions, should be developed.

The battle scenario is rather simplistic. Indeed, the battle may be criticized since it assumes the friendly fire unit has only one weapon whose removal terminates the battle. More sophisticated simulations should be developed and the results from all models should be compared.

The values and evaluation criteria presented in this paper may be used, but should be regarded only as an interim step in the development of optimal target engagement orderings.

REFERENCES

[1] Frank, Douglas H., "Target Value and Payoff", DAAL03-86-D-001, Battelle Memorial Institute, Columbus, OH, August 1988.

[2] Frank, Dr. Douglas H. Frank and Ann E.M. Brodeen, "Target Prioritization to Optimize Expected Utility for a Simple Battle Scenario", U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, BRL Memorandum Report in preparation.

[3] Kemeny, John G. and Snell, J.L., Finite Markov Chains , Princeton, New Jersey: D. Van Nostrand Company, Inc., 1960.

[4] U.S. Army Materiel Development and Readiness Command, Engineering Design Handbook, Army Weapons Systems Analysis, Part One , DARCOM-P 706-101, November 1977.

[5] Brodeen, Ann E.M. and Wendy A. Winner, "Classification Tree Methodology: Another Approach to the Allocation and Distribution of 155mm Howitzer Fire", BRL-MR-3682, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, July 1988.

A Performance Model for a System Using Range and Angle of Arrival Information

Andrew Anderson Thompson III
Ballistic Research Laboratory

Abstract

This paper develops a performance model for a system using AoA and range information to estimate the location of a target. The ideas used to derive and validate the performance model are presented, and then a procedure for evaluating specific sensor systems is discussed.

1. Introduction

Many sensor systems have been developed using angle of arrival (AoA) information for two separate receivers. As a rule of thumb the error associated with these systems is a function of range to the target over the separation of the two receivers. As this ratio becomes large, the estimate loses its value. H. Bruce Wallace of the Ballistic Research Laboratory (BRL) has proposed that range information also be utilized in order to have a worthwhile target location estimate. There is a large class of potential sensors fitting this description. These can be classified according to the type of electronic processing they use and the parameters that are important for that type of processing. The goal here is to find a general model that allows for various levels of detail in the performance evaluation of this class of sensor systems.

This paper develops a performance model for a system using AoA and range information to estimate the location of a target. The ideas used to derive and validate the performance model are presented, and then a procedure for evaluating specific sensor systems is discussed.

First the two dimensional case is examined in detail. Then the results of applying the same ideas to the three dimensional case are presented. Both models were validated through the use of a simulation. The procedure used to validate the three dimensional model is presented. The next two sections present some electronic models from the literature associated with the accuracy of angle of arrival and range measurements. Finally, a method for the evaluation of a particular system is given.

2. XY Covariance Model

In this section a performance model for a system processing one range measurement and one AoA measurement is presented. The measurements are from a polar coordinate system but the system performance is to be in the Cartesian coordinate system. The focus of the discussion is on how the measurement errors effect target estimation in the Cartesian plane. The relationship between an xy location and an AoA range coordinate is straight forward.

$$x = R\cos\Theta + x_0 \qquad (1)$$

$$y = R\sin\Theta + y_0$$

when R is the range to the target and Θ is the AoA to the target. For the rest of the discussion we assume that $x_0 = 0$ and $y_0 = 0$.

A common approach for relating measurement errors to the x-y domain is through the partial derivatives of the location with respect to the measured quantities (Reference 1 and 2). In the x direction the changes caused by range measurement are $\Delta R \cos \Theta$. Changes caused by increasing the angle move x closer to the origin by the amount $R \sin(\Delta\Theta) \sin(\Theta)$. In deriving the following model several assumptions will be made. Afterwards it will be shown that the assumptions are statistically reasonable. Perturbations in x due to measurement errors are described by

$$x + \Delta x = (R + \Delta R) \cos(\Theta + \Delta\Theta).$$

After expanding the cosine term we have:

$$(R + \Delta R) \cos(\Theta + \Delta\Theta) = R \cos\Theta \cos\Delta\Theta - R \sin\Theta \sin\Delta\Theta +$$
$$\Delta R \cos\Theta \cos\Delta\Theta - \Delta R \sin\Theta \sin\Delta\Theta.$$

Since $\Delta\Theta$ is small we will assume $\cos\Delta\Theta = 1$ also since $\Delta\Theta$ is small we assume $\sin\Delta\Theta = \Delta\Theta$ and finally we assume $\Delta\Theta\,\Delta R = 0$. So we now have:

$$(R + \Delta R) \cos(\Theta + \Delta\Theta) \approx R \cos\Theta + \Delta R \cos\Theta - R\Delta\Theta \sin\Theta.$$

Recalling that $x = R\cos\Theta$ we have

$$\Delta x \approx \Delta R \cos\Theta - R\,\Delta\Theta \sin\Theta.$$

By a similar argument it can be shown that

$$\Delta y \approx \Delta R \sin\Theta + R\Delta\Theta \cos\Theta.$$

Note taking partials of both the x and y values results in the same set of equations, and thus those terms we have ignored correspond to higher order differentials. We will assume the measurement errors have the following properties

$$\begin{aligned} E(\Delta R) &= 0 \\ E(\Delta\Theta) &= 0 \\ E(\Delta R^2) &= \sigma^2_R \\ E(\Delta\Theta^2) &= \sigma^2_\Theta \\ E(\Delta R\,\Delta\Theta) &= 0. \end{aligned} \tag{2}$$

Perturbations of x and y, can be expressed in terms of measured quantities as

$$\begin{aligned} \Delta x &= \Delta R \cos\Theta - \Delta\Theta R \sin\Theta \\ \Delta y &= \Delta R \sin\Theta + \Delta\Theta R \cos\Theta \\ (\Delta x)^2 &= (\Delta R)^2 \cos^2\Theta + (\Delta\Theta)^2 R^2 \sin^2\Theta - 2\,\Delta R\,\Delta\Theta R \sin\Theta \cos\Theta \\ (\Delta y)^2 &= (\Delta R)^2 \sin^2\Theta + (\Delta\Theta)^2 R^2 \cos^2\Theta + 2\,\Delta R\,\Delta\Theta R \sin\Theta \cos\Theta \\ \Delta x\,\Delta y &= \Delta R^2 \sin\Theta \cos\Theta - \Delta\Theta^2 R^2 \sin\Theta \cos\Theta - \Delta R\,\Delta\Theta R \sin^2\Theta + \Delta R\,\Delta\Theta R \cos^2\Theta \end{aligned} \tag{3}$$

Combining Equations 2 and 3, we get the following expressions for the variance and covariance of the target location.

$$E(\Delta x) = 0$$

$$
\begin{aligned}
E(\Delta y) &= 0 \\
E(\Delta x^2) &= \sigma^2_R \cos^2\Theta + \sigma^2_\Theta R^2 \sin^2\Theta \\
E(\Delta y^2) &= \sigma^2_R \sin^2\Theta + \sigma^2_\Theta R^2 \cos^2\Theta \\
E(\Delta x\,\Delta y) &= .5\left(\sigma^2_R \sin 2\Theta - \sigma^2_\Theta R^2 \sin 2\Theta\right)
\end{aligned}
\tag{4}
$$

These equations describe the error matrix for the system in terms of the XY coordinate system. In matrix notation the covariance matrix of the estimated position would be

$$
\begin{bmatrix} E(\Delta x^2) & E(\Delta x\,\Delta y) \\ E(\Delta x\,\Delta y) & E(\Delta y^2) \end{bmatrix} = \begin{bmatrix} \sigma^2_x & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma^2_y \end{bmatrix}
$$

By rotating the coordinates through the angle α defined by

$$
\tan^{-1} 2\alpha = \frac{2\,E(\Delta x\,\Delta y)}{E(\Delta x^2) - E(\Delta y^2)}
$$

we decouple the system and can determine the major and minor axis. (This is derived in Appendix A).

From Appendix A we have

$$
\text{Major axis} = \frac{1}{2}\left\{E(\Delta x^2) + E(\Delta y^2) + ((E(\Delta x^2) - E(\Delta y)^2)^2 + 4\,E(\Delta x\,\Delta y)^2)^{1/2}\right\}
$$

$$
\text{Minor axis} = \frac{1}{2}\left\{E(\Delta x^2) + E(\Delta y^2) - ((E(\Delta x^2) - E(\Delta y)^2)^2 + 4\,E(\Delta x\,\Delta y)^2)^{1/2}\right\} \tag{5}
$$

The following equalities were used to rewrite Equation 5

$$
\begin{aligned}
E(\Delta x^2) + E(\Delta y^2) &= \sigma^2_R \sin^2\Theta + \sigma^2_\Theta R^2 \cos^2\Theta + \sigma^2_R \cos^2\Theta + \sigma^2_\Theta R^2 \sin^2\Theta \\
&= \sigma^2_R (\cos^2\Theta + \sin^2\Theta) + \sigma^2_\Theta R^2 (\cos^2\Theta + \sin^2\Theta) \\
&= \sigma^2_R + \sigma^2_\Theta R^2 \\
E(\Delta x^2) - E(\Delta y^2) &= \sigma^2_R (\cos^2\Theta - \sin^2\Theta) + \sigma^2_\Theta R^2 (\sin^2\Theta - \cos^2\Theta) \\
&= \sigma^2_R \cos 2\Theta - \sigma^2_\Theta R^2 \cos 2\Theta \\
E(\Delta x\,\Delta y) &= .5\,\sigma^2_R \sin 2\Theta - \sigma^2_\Theta R^2 \sin 2\Theta
\end{aligned}
$$

Thus the major axis and minor axis are defined by

$$
\begin{aligned}
&\frac{1}{2}\left((\sigma^2_R + R^2\sigma^2_\Theta) \pm ((-\sigma^2_R + \sigma^2_\Theta R^2)^2 \cos^2 2\Theta + (\sigma^2_R - \sigma^2_\Theta R^2)^2 \sin^2 2\Theta)^{1/2}\right) \\
&= \frac{1}{2}\left((\sigma^2_R + R^2\sigma^2_\Theta) \pm (\sigma^2_R - R^2\sigma^2_\Theta)\right)
\end{aligned}
$$

The final simplification of the above equation results in principle components of either σ^2_R and $R^2\sigma^2_\Theta$. In the case of uncorrelated measurment errors, the axis describing the covariance structure are directly related to the measurement errors.

For the two dimensional case, performance can be discussed in terms of x and y or in terms of principle components. Although the ideas are the same in three dimensions the relationships are more complex and it is difficult to give a purely trigonometric explanation.

3. XYZ Covariance Model

In three dimensions a covariance model can be derived from range, azimuth angle, and elevation measurements. The elevation angle Θ is measured from the positive z axis and the azimuth angle, ϕ, is measured from the x axis toward the positive y axis. Using the same simplification arguments as for the XY Covariance Model the following set of measurement error transformation equations is obtained.

$$\Delta x = \Delta R \sin\Theta \cos\phi + \Delta\Theta R \cos\Theta \cos\phi - \Delta\phi R \sin\Theta \sin\phi$$

$$\Delta y = \Delta R \sin\Theta \sin\phi + \Delta\Theta R \cos\Theta \sin\phi + \Delta\phi R \sin\Theta \cos\phi$$

$$\Delta z = \Delta R \cos\Theta - \Delta\Theta R \sin\Theta$$

From these equations the covariance model can be found, the quantities of interest are as follows.

$$\sigma^2_X = \sigma^2_R \sin^2\Theta \cos^2\phi + \sigma^2_\Theta R^2 \cos t^2\Theta \cos^2\phi + \sigma^2_\phi R^2 \sin^2\Theta \sin^2\phi$$

$$\sigma^2_Y = \sigma^2_R \sin^2\Theta \sin^2\phi + \sigma^2_\Theta R^2 \cos^2\Theta \sin^2\phi + \sigma^2_\phi R^2 \sin^2\Theta \cos^2\phi$$

$$\sigma^2_Z = \sigma^2_R \cos^2\Theta + \sigma^2_\Theta R^2 \sin^2\Theta$$

$$\sigma_{XY} = \sigma^2_R \sin^2\Theta \sin\phi \cos\phi + \sigma^2_\Theta R^2 \cos^2\Theta \sin\phi \cos\phi - \sigma^2_\phi R^2 \sin^2\Theta \sin\phi \cos\phi$$

$$\sigma_{XZ} = \sigma^2_R \sin\Theta \cos\Theta \cos\phi - \sigma^2_\Theta R^2 \sin\Theta \cos\Theta \cos\phi$$

$$\sigma_{YZ} = \sigma^2_R \sin\Theta \cos\Theta \sin\phi - \sigma^2_\Theta R^2 \sin\Theta \cos\Theta \sin\phi$$

By finding the eigenvalues and eigenvectors associated with this matrix, the principle components and orientation can be found.

4. Model Validation Effort

Both the XY and the XYZ models were compared with simulation data to verify their performance. Using a Gaussian random number generator, errors for range and azimuth or for range, azimuth, and elevation were generated. These errors were added to the true values and then the position was calculated from the corrupted values. Using ten thousand such points, the covariance of the target position was calculated and then compared to the covariance predicted by the model. The code designed to perform the simulation is included as Appendix B.

As a first test, the determinants were compared to see if they were in agreement. A test based on the asymptotic distribution of the sample covariance was used for this. It is

$$\sqrt{n}\left(\frac{\det(S)}{\det(\Sigma)} - 1\right)$$

is distributed as N(0,2p)

where

p is the dimension of the matrix
n is the degrees of freedom

In this case, an acceptance based on a determinant is not conclusive. These tests have no sensitivity to the orientation of the covariance structure, and will not detect certain differences in the magnitude of the major components of the matrix. For example, a diagonal matrix with components (1 ,1) has the same determinant as the diagonal matrix with components (.1, 10) or the matrix with row one elements (5, 3) and row two elements (3, 2).

The statistical properties of a covariance structure are described by the Wishart distribution. The use of this distribution in the three dimensional case leads to six independent tests. Each one checking a separate component of the covariance matrix. Note that the off-diagonal terms are based on estimates of the diagonal terms and thus have more uncertainty associated with them. The procedure used was as follows:

1. Calculate the error structure from the model.
2. Calculate the error structure from the simulated data.
3. Find the normalizing transformation based on the model.
4. Apply this to the result of step 2.
5. Test the resulting matrix to see if it is statistically equivalent to the identity matrix.

In each of the cases investigated the model and the simulated data produced covariance structures that were statistically the same.

5. AoA Errors

The two models included are based entirely on thermal noise and should be used as the best case situations. When additional sources of error are modeled it is usually correct to take their root mean square with the error due to thermal noise. Dr. Alexander in (2) gives the following two equations for relating electronic parameters to σ_Θ.

For pulsed AoA processing the thermal noise of a phase interferometer is given by

$$\sigma_{th} = \frac{(360/2\Pi)C}{2\,\Pi\, f\, d \cos\Theta\, 4\,(S/N)_{INTG}}$$

For an amplitude monopulse the thermal noise is

$$\sigma_{th} = \frac{25.4\,\Theta_B}{(S/N_{INTG})^{1/2}}$$

where,

c is the propagation velocity (M/S)
f is the RF carrier (H_Z)
d is the spacing between the receiving antennas.
Θ_B is the antenna half power beam width (deg).

S/N_{INTG} is the integrator output of signal to noise ratio.

6. Range Errors

Errors in range depend on the ability to measure the time of arrival of a given pulse. The range error is $\sigma_r = C/2\,\sigma_t$ where σ_t is the time error. The time error is dependent on pulse type, the following are suggested by Skolnik(3):

For a rectangular pulse

$$\sigma_t \approx \left(\frac{\tau}{4\,\beta\,E/N_0}\right)^{1/2}$$

where τ is the pulse width
β is the bandwidth
E is the received signal energy
N_0 is the noise per unit bandwidth

For a trapezoidal pulse

$$\sigma_t = \left(\frac{T_2^{\,2} - 3\,T_1\,T_2}{6\,E/N_0}\right)^{1/2}$$

T_2 is the rise and fall time of the pulse
T_1 is the direction of the top of the pulse

For a triangular pulse

$$\sigma_2 = \frac{2\,T_2}{\sqrt{12}\,(2\,E/N_0)^{1/2}}$$

For a Gaussian pulse of the form $s(t) = \exp\left(-\frac{1.384t^2}{\tau^2}\right)$

$$\sigma_t = \frac{1.18}{\pi\,\beta\,(2\,E/N_0)^{1/2}}$$

If the pulse has the form $\frac{\sin(\pi\,\beta\,\tau)}{\pi\,\beta\,\tau}$

then

$$\sigma_t = \frac{\sqrt{3}}{\pi\,\beta\,(2\,E/N_0)^{1/2}}$$

Continuous Wave Error

$$\sigma_R = \frac{C}{4\,\pi\,\Delta f\,(2\,E/N_0)^{1/2}}$$

where Δf is the difference between the two frequencies.

The specific electronic model will vary depending on the method used to extract

information from the signal. The models included above only represent some of the common techniques. The important thing is that they relate specific electronic or geometric parameters to the measurement errors associated with either range or angle of arrival.

7. Analysis Method

In using these models to analyze the performance of a system, the following steps must be taken.

1. Choose a AoA error model and assign values to the parameter.
2. Choose a range error model and assign values to the parameters.
3. Use the range and AoA errors as input to the XYZ location model.

By following this procedure at a number of different points a system's performance can be presented as a function of target location.

8. Conclusion

The models presented herein can be used to evaluate the performance of many range - angle sensor systems. The performance can be based on specific electronic parameters such as frequency or pulse shape or on more general specifications such as three degree angle with five percent range errors. This work could be continued by designing a software package that includes the selection of the possible options.

This application is typical of the error analysis approach used in many engineering studies. In this case , the extra step of checking the statistical validity of the model was included. The dominant feature of this approach to system analysis is to start with the measurement errors and follow them as they propagate through the system and degrade the ideal system performance.

APPENDIX A

We wish to find the angle α of rotation in the coordinate system to decouple the system.

$$\Delta x = \Delta x \cos\alpha + \Delta y \sin\alpha$$

$$\Delta y = -\Delta x \sin\alpha + \Delta y \cos\alpha$$

In terms of the new coordinate system we have

$$E(\Delta x) = 0$$

$$E(\Delta y) = 0$$

$$\begin{aligned} E(\Delta x) &= E(\Delta x^2)\cos^2\alpha + 2E(x)(y)\sin\alpha\cos\alpha + E(\Delta y^2)\sin^2\alpha \\ &= E(\Delta x^2) + E(\Delta y^2) - E(\Delta x^2)\sin^2\alpha - E(\Delta y^2)\cos^2\alpha \\ &\quad + 2E(\Delta x\,\Delta y)\sin\alpha\cos\alpha \end{aligned}$$

Recall that

$$\sin\alpha\cos\alpha = \frac{1}{2}\sin 2\alpha$$

$$\sin^2\alpha = \frac{1 - \cos 2\alpha}{2}$$

$$\cos^2\alpha = \frac{1 + \cos 2\alpha}{2}$$

then

$$E(\Delta x^2) = \frac{1}{2}(E(\Delta x^2) + E(\Delta y^2)) + \frac{1}{2}[E(\Delta x^2 - E(\Delta y^2)]\cos 2\alpha$$

$$+ E(\Delta x\,\Delta y)\sin 2\alpha$$

Similarly it can be shown that

$$E(\Delta y^2) = \frac{1}{2}\left[E(\Delta x^2) + E(\Delta y^2)\right] - \frac{1}{2}(E(\Delta y^2) - E(\Delta y^2))\cos 2\alpha$$

$$- E(\Delta x\,\Delta y)\sin 2\alpha$$

$$\begin{aligned} E(\Delta x\,\Delta y) &= -E(\Delta x^2)\sin\alpha\cos\alpha + E(\Delta x\,\Delta y)(\cos^2\alpha - \sin^2\alpha) \\ &\quad + E(\Delta y^2)\sin\alpha\cos\alpha \\ &= \frac{1}{2}\left[E(\Delta y^2) - E(\Delta x^2)\right]\sin 2\alpha + E(\Delta x\,\Delta y)\cos 2\alpha \end{aligned}$$

The covariance term will be zero if

$$\tan 2\alpha = \frac{2E(\Delta x\,\Delta y)}{E(\Delta x^2) - E(\Delta y^2)}$$

Note that for this angle

$$\sin 2\alpha = \frac{2\,E\,(\Delta x\,\Delta y)}{(4\,E\,(\Delta x\,\Delta y)^2 + (E\,(\Delta x^2 - E\,(\delta y^2)^2)^{1/2}}$$

$$\cos 2\alpha = \frac{E\,(\Delta x^2) - E\,(\Delta y^2)}{(4\,E\,(\Delta x\,\Delta y)^2 + (E\,(\Delta x^2) - E(\Delta y^2))^2)^{1/2}}$$

Using these relations it can be shown that the major and minor axis are defined by

$$1/2\,(E\,(\Delta x^2) + E\,(\Delta y^2) \pm (E\,(\Delta x^2) - E(\Delta y^2) + 4\,E\,(\Delta x\,\Delta y)^2)^{1/2})$$

APPENDIX B

```
#include <math.h>
#include "ranvar.h"
#include "stat.h"

main()
{
 float range,r,r_sd,r_var,r_sq;
 float theta,th,theta_sd,theta_var;
 float psi,p,psi_sd,psi_var;
 float x,y,z,x_sum,y_sum,z_sum,x_var,y_var,z_var;
 float xsq,ysq,zsq,xy,xz,yz;
 float r_sim,th_sim,psi_sim;
 float sinth,costh,sinpsi,cospsi,sin2th,cos2psi,cos2th,sin2psi;
 float cov_xy,cov_xz,cov_yz;
 float x2_mod,y2_mod,z2_mod,xy_mod,xz_mod,yz_mod;
 float data_det,model_det,minus,plus,z_test,prob;

 int i,n,seed;

 /* initialize the variables for this run */
 range=50;
 r_sd=.05*range;        /* five percent range */
 r_var=r_sd*r_sd;
 r_sq=range*range;
 theta=M_PI/4;
 theta_sd=5*2*M_PI/360;   /*five degree error elevation */
 theta_var=theta_sd*theta_sd;
 psi=M_PI/4;
 psi_sd=5*2*M_PI/360;   /* five degree error azimuth */
 psi_var=psi_sd*psi_sd;

 n=10000;            /* number of replications */
 seed=23719;          /* random number seed */

 xsq=0;
 ysq=0;
 zsq=0;
 x_sum=0;
 y_sum=0;
 z_sum=0;
 xy=0;
 xz=0;
 yz=0;

 for (i=0;i<n;i++)
```

```
{
/* rv_gauss is a gaussian random number generator
   the first two sections find the measured value
   and calculate the position
*/
 th = theta+theta_sd*rv_gauss(seed);
 r=range+r_sd*rv_gauss(seed);
 p=psi+psi_sd*rv_gauss(seed);

 z=r*cos(th);
 x=r*sin(th)*cos(p);
 y=r*sin(th)*sin(p);

 x_sum+=x;
 y_sum+=y;
 z_sum+=z;

 xsq+=x*x;
 ysq+=y*y;
 zsq+=z*z;
 xy+=x*y;
 xz+=x*z;
 yz+=y*z;
} /* end of replication loop */
x=x_sum/n;
y=y_sum/n;
z=z_sum/n;

/* the following values are the covariance elements based on
   the simulations data
*/
x_var=(xsq-x*x_sum)/(n-1);
y_var=(ysq-y*y_sum)/(n-1);
z_var=(zsq-z*z_sum)/(n-1);
cov_xy=(xy-x*y_sum)/(n-1);
cov_xz=(xz-x*z_sum)/(n-1);
cov_yz=(yz-y*z_sum)/(n-1);

/* the next section uses the model to find the predicted covariance
   structure
*/
sinth=sin(theta);
sin2th=sinth*sinth;
costh=cos(theta);
cos2th=costh*costh;
sinpsi=sin(psi);
sin2psi=sinpsi*sinpsi;
```

```
cospsi=cos(psi);
cos2psi=cospsi*cospsi;

x2_mod=r_var*sin2th*cos2psi;
x2_mod+=theta_var*r_sq*cos2th*cos2psi;
x2_mod+=psi_var*r_sq*sin2th*sin2psi;

y2_mod=r_var*sin2th*sin2psi+theta_var*r_sq*cos2th*sin2psi;
y2_mod+=psi_var*r_sq*sin2th*cos2psi;

z2_mod=r_var*cos2th+theta_var*r_sq*sin2th;

xy_mod=r_var*sin2th*cospsi*sinpsi+theta_var*r_sq*cos2th*cospsi*sinpsi;
xy_mod-=psi_var*r_sq*sin2th*cospsi*sinpsi;

xz_mod=r_var*sinth*costh*cospsi-theta_var*r_sq*sinth*costh*cospsi;
yz_mod=r_var*sinth*costh*sinpsi-theta_var*r_sq*sinth*costh*sinpsi;

/*perform determinate test
*/
plus = x_var*y_var*z_var;
plus += cov_xy*cov_yz*cov_xz*2;
minus = cov_xz*cov_xz*y_var;
minus += cov_xy*cov_xy*z_var;
minus += cov_yz*cov_yz*x_var;
data_det=plus-minus;

plus = x2_mod*y2_mod*z2_mod;
plus += 2*xy_mod*xz_mod*yz_mod;
minus = xy_mod*xy_mod*z2_mod;
minus += y2_mod*xz_mod*xz_mod;
minus += x2_mod*yz_mod*yz_mod;
model_det=plus-minus;

z_test = sqrt(n-1) * (data_det/model_det - 1);
z_test /= sqrt(2*3);
prob = cnf(z_test);

printf("0);
printf("****************-----------------------------*******************");
printf(" Range : %f  theta : %f  psi : %f",range, theta,psi);
printf("0d(range)= %f sd(theta)= %f sd(psi)= %f",r_sd,theta_sd,psi_sd);
printf("0imulated Var(X)= %f model value was %f",x_var,x2_mod);
printf("0imulated Var(Y)= %f model value was %f",y_var,y2_mod);
printf("0imulated Var(Z)= %f model value was %f",z_var,z2_mod);
printf("0imulated Cov(XY)= %f model value was %f",cov_xy,xy_mod);
printf("0imulated Cov(XZ)= %f model value was %f",cov_xz,xz_mod);
```

```
printf("0imulated Cov(YZ) = %f model value was %f",cov_yz,yz_mod);
printf("0 value of %f with probability of %f",z_test,prob);
}
```

RELIABILITY DESIGN PROCEDURES FOR FLEXIBLE PAVEMENTS

Yu T. Chou

Research Civil Engineer, Pavement Systems Division
Geotechnical Laboratory
US Army Engineer Waterways Experiment Station
Vicksburg, MS 39180-6199

ABSTRACT: A procedure has been developed to analyze layered elastic flexible pavement systems in terms of reliability. A computer program RELIBISA was prepared to carry out the computations. The Rosenblueth's method, instead of the conventional Taylor series expansion, is used to estimate the expected value and variance of the strains (dependent parameters) based on the input mean values of independent parameters, i.e., aircraft load, layer thicknesses, and material moduli. The relationships between the reliability level and the allowable strain repetition of the designed system, which is established with results computed using RELIBISA, provide a decision-making tool for engineers to design pavements at a desired reliability level. The design can be optimized by selecting thicknesses of the bituminous concrete and the base layer so that the pavement fails in fatigue cracking of

Preceding Page Blank

the bituminous concrete and in subgrade at nearly the same traffic level for a given reliability level. The reliability-strain repetition curves have a steeper slope with the bituminous concrete strain failure criterion than with the subgrade strain failure criterion. This steepness indicates that for flexible pavements designed using the Corps of Engineer's failure criteria, the design has a greater degree of uncertainty in preventing subgrade failure than fatigue cracking of the bituminous concrete surface course. However, this may not be true in real cases because the bituminous concrete failure criteria are determined based on controlled laboratory test data which do not consider the uncertainties existing in laboratory-to-field correlations. The actual performance of the pavement with respect to fatigue cracking will be more uncertain than is considered in the design. The significance of the failure criteria employed in the analysis on the derived conclusions is discussed and illustrated.

It was found that the performance of a conventional flexible pavement is sensitive, in descending order, to variations of gear load P, the thickness of the granular base h_2, the subgrade modulus E_3, the thickness and modulus of the bituminous concrete surface course h_1, and E_1, respectively, for the subgrade strain failure criterion, and to variations of P, h_1, E_1, E_2, h_2, and E_3 for the bituminous concrete strain failure

250

criteria. Although pavement performance is more sensitive to the variation of layer thickness than to variation of the elastic modulus of the layer, the actual variation of material moduli in the field is known to be much larger than the variation of layer thickness. Strict control during construction is recommended to reduce the degree of material variability and thus to lessen the degree of uncertainty and to increase the confidence level of the designed pavement.

INTRODUCTION

The design of flexible airfield pavements in the US Army Corps of Engineers (USACE) is currently based on two methods: (1) the California Bearing Ratio (CBR) equation that is empirical in nature and yields a design thickness for a given design condition, and (2) the multilayered elastic method that is analytical in nature and yields stresses, strains, and deflections in the pavement system for a particular loading condition and pavement geometry which in turn are compared with established failure criteria to determine the performance of the given pavement. Both design approaches are deterministic, i.e., a unique pavement system is designed for the specific set of input variables necessary to solve the problem. The input variables are unique. In the CBR method, a pavement thickness is determined from given values of subgrade CBR, gear load and configuration, tire

contact area, and design coverage level. The effect of material variability on pavement performance is considered in the designer's selection of the subgrade CBR value, and the design safety factor is implicitly contained within construction specifications such as compaction requirements.

A design methodology that has the capability of considering design parameter variability in the USACE design procedure for flexible airfield pavements using the elastic layered method is presented in this paper. The design procedure is expressed in probabilistic and reliability terms, i.e., the design pavement thicknesses at different performance levels are computed for a range of reliability levels. The designer can select the pavement thickness and in some cases develop an overlay design scheme based on the desired reliability level. The design procedure is incorporated in a computer program. By using the procedure, the partial effect of the variability of each design parameter on pavement performance can also be investigated, and its effects on the final design can be quantified. Emphasis can be placed on the crucial parameters to be tightly controlled in the construction phases and/or the crucial loading parameters dictated by the intended use of the pavement.

PREVIOUS WORKS

Witczak, Uzan, and Johnson (1) of the University of Maryland, under a contract from the Waterways Experiment Station (WES), developed design methodologies for rigid airfield pavements in terms of

probability and reliability. The work involved design procedures based on the Westergaard free edge stress slab theory (2) and the multilayer elastic theory (3). Taylor series expansion (4) was used in the probabilistic analysis. The development of probabilistic and reliability design methodologies for flexible airfield pavements has been conducted at WES. Effort was first conducted on the CBR design equation (5). The work is briefly described in the next paragraph.

The original and the new CBR equation for flexible airfield pavements are shown below as Equations 1 and 2, respectively.

$$t = \alpha\sqrt{\frac{P}{8.1\ CBR} - \frac{A}{\pi}} \tag{1}$$

$$t = \alpha\left\{\sqrt{A}\left[-0.0481 - 1.1562\left(\log\frac{CBR \cdot A}{P}\right) - 0.6414\left(\log\frac{CBR \cdot A}{P}\right)^2 - 0.473\left(\log\frac{CBR \cdot A}{P}\right)^3\right]\right\} \tag{2}$$

where

t = pavement thickness, in.

α = a traffic factor

P = single-wheel load (or the equivalent single-wheel load (ESWL) in the case of the multiple-wheel loads), lb

CBR = California Bearing Ratio of the subgrade soil

A = tire contact area, square inches

Equation 1 was formulated in the 1950's, and Equation 2 was formulated in the early 1970's based on additional test data (6). The design parameters considered were the load P (or the ESWL), the subgrade CBR, the tire contact area A, and the pavement total thickness h. The expected value and the variance of the dependent variable traffic factor α were estimated using the Taylor series expansion and the Rosenblueth method (7). Differences in computed results between the two methods were found to be small, although the derivation of the expressions for Taylor series expansion is very complicated. A computer program was developed to estimate the reliability of the designed pavement system based on known variabilities of design parameters. Results of the reliability analysis indicate that prediction of pavement performance is most influenced by variations of pavement total thickness and is least influenced by variations of tire contact area A. The effects of variations of wheel load P and subgrade CBR are identical. The relative sensitivity normalized to thickness t for parameters t, CBR, P, and A, in general cases, are approximately 1, 0.34, 0.34, and 0.01, respectively. It was thus concluded that in the future analysis of pavements involving input parameter variabilities, the effect of the variation of wheel contact area may be neglected. It was also recommended that strict quality control be exercised during construction to reduce variations of pavement thickness and subgrade CBR, and that the Rosenblueth method (7) be used because of its simplicity and accuracy in the probabilistic analysis of layered elastic system.

PROBABILISTIC AND RELIABILITY APPROACH

In analyzing a pavement structure in probabilistic and reliability terms, the expected value and variance of a function (such as the computed stresses, strain, or load repetition) should first be determined, and the reliability of the design can then be evaluated. The Taylor series expansion (4) and the Rosenblueth (7) procedure are generally used. These methods are presented below.

Taylor Series Expansion. The Taylor formula or the expansion of a function $f(x)$, which has N continuous derivatives, about the function's mean μ is

$$f(x) = f(\mu) + f'(\mu)(x - \mu) + \frac{f''(\mu)}{2}(x - \mu)^2$$

$$+ \ldots \text{ higher order terms} + \text{remainder} \tag{3}$$

Since the expected value of $(x - \mu)$ is zero and the expected value of $(x - \mu)^2$ is the variance of x, i.e., $E(x - \mu) = 0$ and $E(x - \mu)^2 = \sigma_x^2$, the expected value of $f(x)$ becomes

$$E[f(x)] = f(\mu) + \frac{1}{2} f''(\mu)\sigma_x^2 \tag{4}$$

The variance of a function $f(x)$ is

$$V[f(x)] = E[f^2(x)] - E[f(x)]^2 \tag{5}$$

and

$$V[f(x)] \approx [f'(\mu)]^2\sigma_x^2 - \frac{1}{4}[f''(\mu)]^2\sigma_x^4 \tag{6}$$

In Equations 4 and 6, if the random variables can be assumed normally distributed, the second-order terms may be neglected. Equation 6 also assumes that the variables are uncorrelated.

Rosenblueth Method--Equations 4 and 6 are obtained from the truncated Taylor series expansion of the function about the expectations of the random variables. This method requires the existence and continuity of the first and second derivatives of the function. Rosenblueth used the point estimates of the function, and the expressions for the expected value are:

$$E[\varepsilon^N] = \frac{1}{2}\left(\varepsilon_+^N + \varepsilon_-^N\right) \text{ for one variable} \tag{7}$$

$$E[\varepsilon^N] = \frac{1}{2^2}\left(\varepsilon_{++}^N + \varepsilon_{+-}^N + \varepsilon_{-+}^N + \varepsilon_{--}^N\right) \text{ for two variables} \tag{8}$$

$$E[\varepsilon^N] = \frac{1}{2^3}\left(\varepsilon_{+++}^N + \varepsilon_{++-}^N + \varepsilon_{+--}^N + \varepsilon_{+-+}^N + \varepsilon_{-++}^N + \varepsilon_{-+-}^N + \varepsilon_{--+}^N + \varepsilon_{---}^N\right) \text{ for three variables} \tag{9}$$

$$E[\varepsilon^N] = \frac{1}{2^4}\Big(\varepsilon^N_{++++} + \varepsilon^N_{+++-} + \varepsilon^N_{++-+} + \varepsilon^N_{++--} + \varepsilon^N_{+-++}$$
$$+ \varepsilon^N_{+-+-} + \varepsilon^N_{+--+} + \varepsilon^N_{+---} + \varepsilon^N_{-+++} + \varepsilon^N_{-++-}$$
$$+ \varepsilon^N_{-+-+} + \varepsilon^N_{-+--} + \varepsilon^N_{--++} + \varepsilon^N_{--+-} \tag{10}$$
$$+ \varepsilon^N_{---+} + \varepsilon^N_{----}\Big) \text{ for four variables}$$

$$E[\varepsilon^N] = \frac{1}{2^M}\Big(\varepsilon^N_{\underbrace{+++++++}_{M}} + \cdots + \varepsilon^N_{\underbrace{-------}_{M}}\Big) \text{ for } M \text{ variables} \tag{11}$$

Note that the number of total terms to calculate the expected value of a function ε (strain computed using the elastic layered BISAR program (8)) which has M variables is 2^M, and N has a value of either 1 or 2 as shown in Equation 5, i.e., ε is represented by $f(x)$, and N is the power of the function. ε_+ and ε_- in Equation 7 are the strain values evaluated at the mean plus one standard deviation of the variable and the mean minus one standard deviation of the variable, respectively. ε_{+-} in Equation 8 is the strain value evaluated at the mean plus one standard deviation of the first variable and the mean minus one standard deviation of the second variable. Similar reasoning holds true for the other terms.

To reduce the number of variables in elastic layered method computations, variations of Poisson's ratio of pavement materials are neglected, as it has insignificant effect on pavement response to loads. The variation of tire contact area can also be neglected in this computation, based on the conclusion of the previous study of CBR design method for airfield pavements (5). To illustrate the use of the Rosenblueth method, the computation of the expected value of the

strain ε for a two-layer flexible pavement is presented. The independent parameters considered but uncertain are the wheel load P, the elastic modulus and thickness of the top layer E_1 and h_1, respectively, and the elastic modulus of the subgrade E_2. Other parameters (wheel contact area and Poisson's ratio) are precisely known, i.e., the standard deviations are zero. For a four-parameter problem, Equation 10 is used to determine the expected value of the strain ε. Assuming that the standard deviations of the parameters are σ_P, σ_{E1}, σ_{E2}, and σ_{h1} and that the parameters are arranged in the order of P, E_1, E_2, and h_1 (i.e., the order of the symbols ++++, +++-, ···, etc.), each term in Equation 10 is computed using the BISAR program. Once the mean values for parameters $\bar{P}$, $\bar{E}_1$, $\bar{E}_2$, and $\bar{h}_1$ and their standard deviations σ_P, σ_{E1}, σ_{E2}, and σ_{h1} are specified, the expected value of ε can be determined from Equation 10, and the variance of ε is computed using Equation 5.

Reliability Analysis--as soon as the expected value and the variance of a function (such as the strain values computed in an elastic layered pavement system or the α factor in Equation 2 representing the traffic performance level) are determined, the reliability level of the function can be computed. Reliability is defined as the probability that the pavement system will perform its intended function over its design life (or time) and under the conditions (or environment) encountered during operation (9).

Two failure criteria are used (10). The criteria for allowable strain repetitions N for the bituminous concrete is mathematically expressed as

$$N_{Allowable_{(AC)}} = 10^{-A} \tag{12}$$

where

$$A = 5 \text{ Log}_{10} \varepsilon + 2.665 \text{ Log}_{10}\left(\frac{E_{AC}}{14.22}\right) + 0.392$$

ε = maximum horizontal tensile strain at the bottom of the asphaltic concrete layer

E_{AC} = elastic modulus of the asphaltic concrete, psi

The criteria for allowable strain repetition N for the subgrade is expressed as

$$N_{Allowable_{(subgrade)}} = 200{,}000\left(\frac{A}{\varepsilon_{subg}}\right)^{B} \tag{13}$$

where

$$A = 0.000247 + 0.000245 \text{ Log}_{10} (E_{subg})$$

E_{subg} = subgrade modulus, psi

ε_{subg} = subgrade strain, dimensionless

$$B = 0.0658 \, (E_{subg})^{0.559}$$

The bituminous concrete strain criteria in Equation 12 are derived based on laboratory fatigue data on the breaking stress and strain of bituminous base-course materials with about 5 percent air voids. The subgrade strain criteria in Equation 13 are derived based on the full-scale accelerated traffic test data.

The strain value ϵ for each term of the Rosenblueth expression (Equations 7 to 11) is computed using the BISAR program. The performance (strain repetitions to failure) of the pavement may be estimated from the failure criteria shown in the equations for two different failure modes.

With the strain value ϵ assumed normally distributed, the number of allowable strain repetitions corresponding to $\epsilon + \epsilon\sigma_\epsilon$ (or $\epsilon[1 + C \cdot CV(\epsilon)]$) can be determined from Equations 12 and 13, and the probability of $\epsilon \leq \epsilon[1 + C \cdot CV(\epsilon)]$ is taken from the normal distribution. $CV(\epsilon)$ is the coefficient of variation of ϵ, which is the ratio of the standard deviation of ϵ to a mean of ϵ, (i.e., $\sigma_\epsilon/\bar{\epsilon}$), and C is the selected number varying from -3 to +3. The selection of C values less than -3 and greater than +3 are not necessary because the areas under a normal distribution curve beyond -3 and +3 standard deviations are negligible. The computations of the reliabilities, and allowable strain repetitions are carried out in the RELIBISA Computer program. The program logic of RELIBISA is presented in Reference 10. Although the program is prepared based on failure criteria presented in Equations 12 and 13, other failure criteria such

as those for highway pavements and other design conditions for airfield pavements can also be used in the programs.

ANALYSIS OF THE FLEXIBLE PAVEMENT

In the analysis of an airfield pavement in terms of probability and reliability, the RELIBISA computer program is used to calculate the allowable load repetitions for a given pavement section for various reliability levels. The input parameter variations are CV (coefficient of variations) defined to be the ratio of the standard deviation to the mean value of the parameter. For instance, if the mean gear load is 178,000 lb and the CV of the gear load is assumed to be 10 percent, the standard deviation of the gear load will be 17,800 lb, i.e., 68.3 percent of the time the gear load would lie between 160,200 and 195,800 lb (which is plus and minus one standard deviation for a normally-distributed variate).

Relationships between reliability level and the corresponding allowable load repetitions are established for many pavement sections with various input parameter variabilities. The reliability of the Corps of Engineers flexible pavement design model is 0.5, (.11), but the design method has higher reliability value because of the design safety factor implicitly contained within construction specification such as compaction requirements and the selection of subgrade CBR value.

Elastic Moduli of Pavement Layers--for conventional flexible pavements, the critical period during the year is the colder winter months for the surface bituminous layer as the bituminous concrete becomes more brittle and the warmer summer months for the subgrade as the surface layer becomes less stiffer. Accordingly, a modulus value of 1,000,000 psi and a Poisson's ratio of 0.3 are used in the bituminous concrete strain criterion (winter), and a modulus value of 200,000 psi and a Poisson's ratio of 0.5 are used in the subgrade strain criterion (summer). The modulus value of the granular base material is assumed to be 55,000 and 32,000 psi for the bituminous concrete strain criterion and the subgrade strain criterion, respectively. The selection of these values is explained in Reference 10.

Analysis of a Three-layer Flexible Pavement--a gear assembly load of 178,000-lb B-747 aircraft was used in the computation. The aircraft has twin-tandem gear assemblies and the wheels are spaced 44 by 58 in. Each wheel load is 44,530-lb, and the tire contact pressure is 182 psi.

Two series of computations were made using the RELIBISA computer program to analyze flexible airfield pavement parameters in terms of probability and reliability. In the first series, the thickness of the bituminous concrete surface course h_1 is held 9 in., and a range of granular base layer h_2 is assumed. In the other series, the thickness of the base layer h_2 is held constant 30 in., and a range of surface layer h_1 is assumed. In both computations, the CV of

the input parameters, gear load P, moduli of the bituminous concrete E_1, the granular base course E_2 and the subgrade E_3, and the thicknesses of the surface course h_1 and the base course h_2 were assumed to be 0.1, 0.15, 0.2, 0.25, 0.1, and 0.15, respectively. The computed results are plotted in Figures 1 and 2 for the subgrade strain criterion (summer temperatures) and the bituminous concrete strain criterion (winter temperatures), respectively.

The curves shown in Figure 1 for the subgrade strain criterion are generally parallel to each other, except in the area where the reliability is close to one and zero. For a given bituminous concrete thickness h_1, increasing the thickness of granular layer h_2 can increase the allowable strain repetition of the pavement. This is also true if this procedure is reversed. Figure 2 shows that for the bituminous concrete strain criterion, the performance of the pavement can certainly be improved with the increase of the thickness of the bituminous concrete surface layer h_1 (for a given thickness of the granular layer). This is also true if the thickness of the surface layer h_1 is held a constant and the thickness of the granular layer h_2 is varied, but the benefit reduces rapidly for very thick granular layer, which is demonstrated by the closely spaced curves at greater h_2 values presented in the lower part of Figure 2. The significance of the curves is that for a given pavement thickness, the allowable strain repetition to failure varies with its reliability level. For a 9-in. bituminous concrete surface layer h_1 and a 20-in. granular layer, the allowable strain repetition against subgrade failure (Figure 1) at a reliability level of 0.5 is 8,500 strain repetitions,

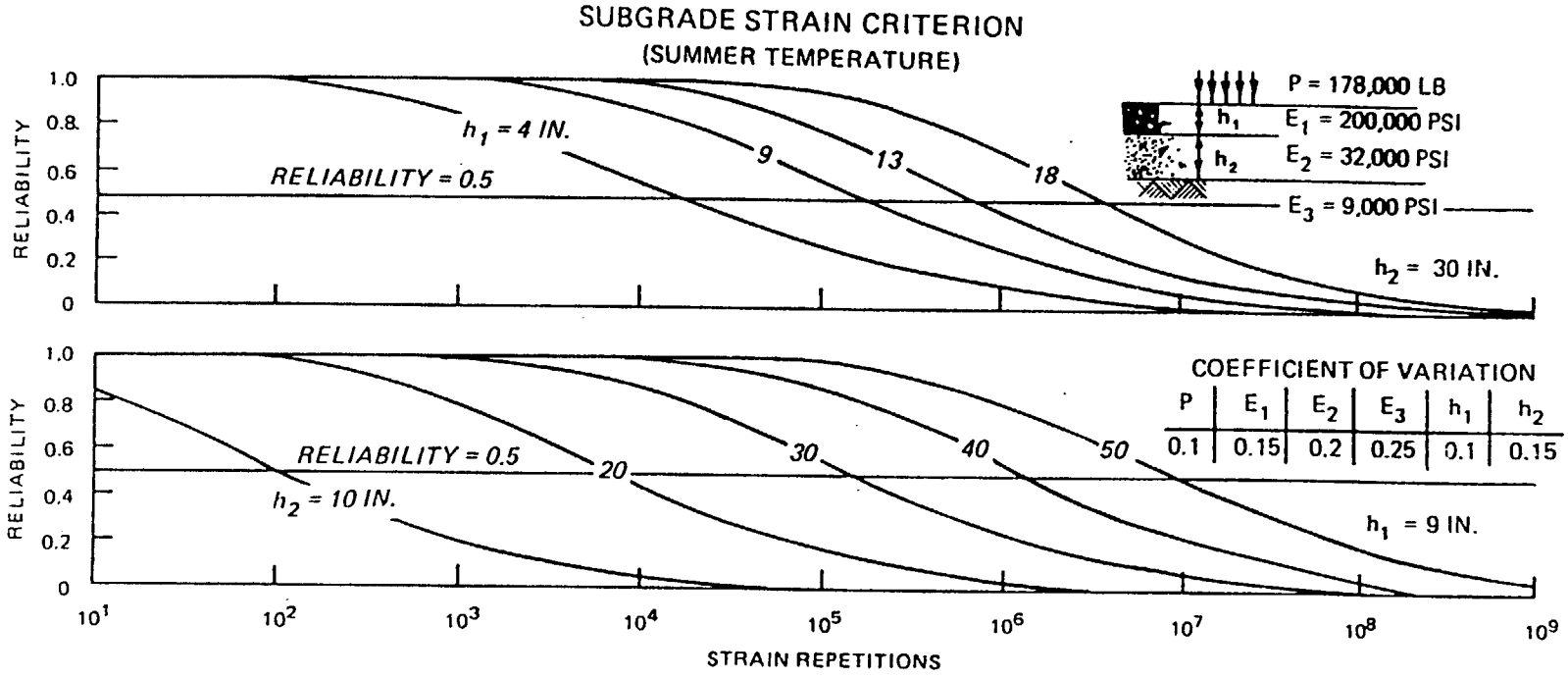


Figure 1. Relationships between reliability and allowable strain repetition for flexible airfield pavements with subgrade strain criterion

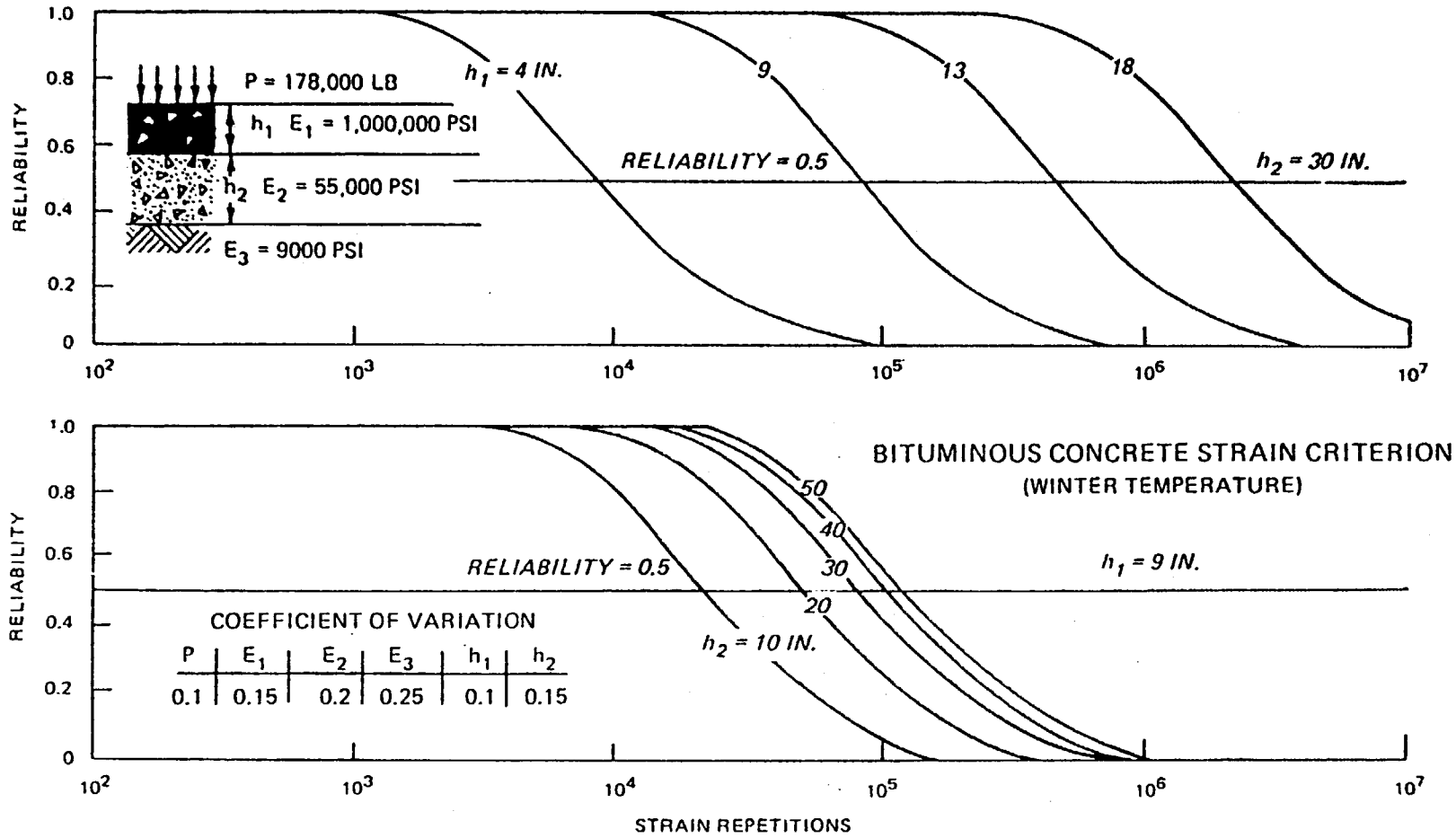


Figure 2. Relationships between reliability and allowable strain repetition for three-layer flexible airfield pavements with bituminous concrete strain criterion

i.e., the chance of success of the design of this pavement to sustain 8,500 strain repetitions of the B-747 aircraft load before failure is 50 percent. Figure 1 shows that the chance of success can be increased to 80 percent if the design strain repetition is reduced to 1,000. If the 8,500 repetitions are considered as a 20-year design period and the 1,000 repetitions are thus equivalent to 2.4 years, there is an 80 percent chance that the pavement can last 2.4 years without failure (with routine maintenance), but there is only a 50 percent chance that the pavement can last a full 20-year design period.

It is to be noted that the flatter the slope of the curves in Figures 1 and 2, the greater are the uncertainties involved in the design. However, the shapes of the curves are influenced by the failure criteria (Equations 12 and 13) employed in the computations. This will be discussed later in this paper. The curves in Figure 2 for the bituminous concrete strain criterion have steeper slopes than those in Figure 1 for the subgrade strain criterion, indicating that, using the Corps of Engineers' failure criteria, the designed pavement may have a greater degree of uncertainty in preventing subgrade failure than in preventing fatigue cracking of bituminous concrete surface course.

For a given design strain repetition, the relationships between reliability and pavement thickness can also be obtained from Figures 1 and 2. Engineers can choose the pavement thickness suitable for the selected reliability level of the design. This point can best be

266

demonstrated from the curves presented in Figure 3 which is plotted from Figures 1 and 2. Figure 3 shows the relationships between the strain repetition and layer thicknesses h_1 and h_2 at different reliability levels. The relationships in Figure 3 can be helpful to designers in selecting the allowable strain repetitions of a given pavement section for a desired reliability level or to vary the layer thicknesses h_1 and h_2 that will be suitable for a specific design performance level of the pavement for a given reliability level. The slopes of the curves indicate the rate of change of allowable strain repetition due to change of layer thicknesses h_1 or h_2. Obviously the steeper the slope of the curve, the better the design for that particular failure mode would be.

The curves in Figure 3b in the region where the granular layer thickness h_2 is less than 30 in. have generally the same slope as those in Figure 3a. Since the unit cost of granular base course is much less than the bituminous concrete surface course, it is economically more beneficial to increase the thickness of the granular base course (h_2) to prevent the pavement from subgrade failure. This is logical from the structural viewpoint, as the base course is placed directly on the subgrade. However, the slopes in Figure 3b drop slightly at greater h_2 thicknesses, indicating the significance of granular layer thickness increase also drops slightly at greater h_2 thickness.

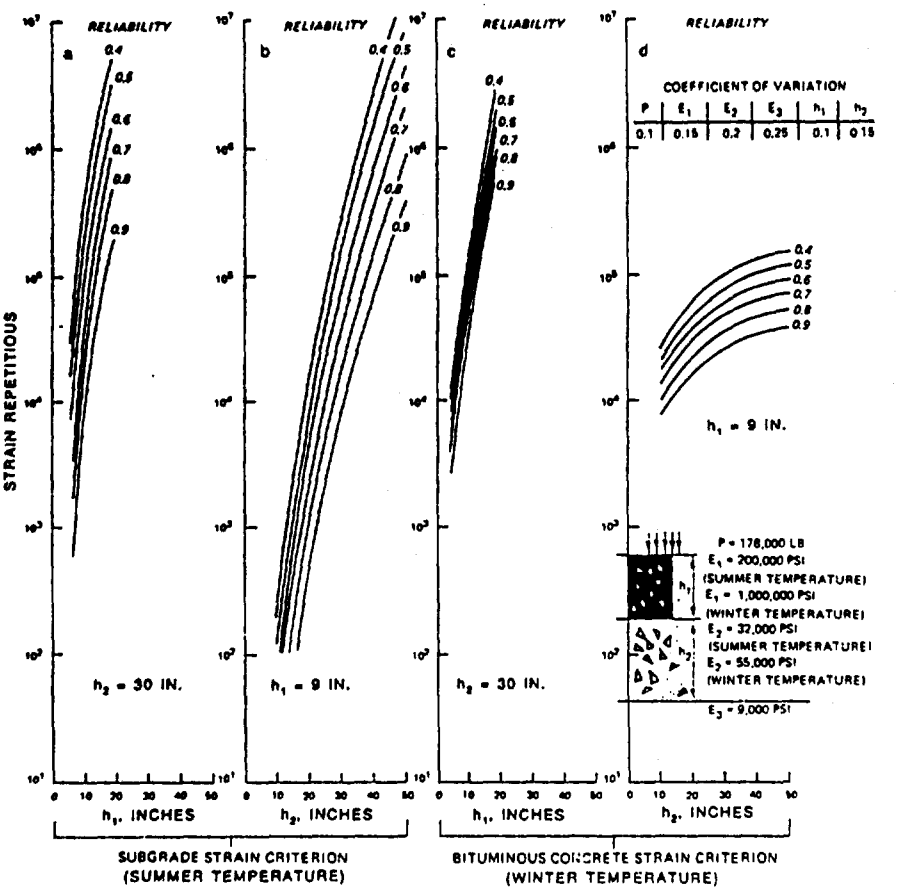


Figure 3. Relationships between pavement layer thicknesses and strain repetition (from Figures 1 and 2) with flexible pavements

Figures 3c and 3d show curves plotted for the bituminous concrete strain criterion. The steep slopes in Figure 3c indicate the structural benefit of increasing bituminous concrete thickness to prevent the pavement from failure of surface cracking. The flatter slopes of the curves in Figure 3d indicate relatively little significance of base course support in a flexible pavement in the limitation of fatigue cracking of the surface course. Figure 3d also shows that the significance of base course reduces rapidly as its thickness continues to increase.

The curves presented in Figure 3 provide engineers with a tool to vary the layer thicknesses (h_1 or h_2) to be suitable for the specific design performance level of the pavement and for a given reliability level of the design. An optimum design may be made to select the thicknesses of the bituminous concrete and the base layers so that the pavement fails in fatigue cracking and in subgrade at nearly the same traffic level for the same reliability level.

The conclusions drawn from Figures 1 to 3 are based on a subgrade modulus of 9,000 psi. Questions arise as to whether a stronger subgrade support of the pavement would reverse the observed trend. Computations similar to those presented in Figures 1 and 2 were made for a subgrade modulus of 25,000 psi. It was found that the relationships between the reliability and the strain repetition are very similar to those shown in Figures 1 and 2, except that the curves shift to higher strain repetition values because of stronger subgrade support. This is more predominate in the subgrade strain failure mode than in the

bituminous concrete failure mode because stronger subgrade support has a greater effect on pavement performance with respect to subgrade failure than with respect to bituminous concrete failure.

Figures 1 and 2 show the relationships between the reliability and strain repetition for a given set of CV's. Figure 4 shows results of the effect of each individual parameter on the performance of flexible pavements. Computations were made to vary only one parameter each time while variations of the other five parameters were zero. The results presented in Figure 4 are for the CV's of 0.1. The figure shows that the variabilities of different input parameters have different degree of impact on pavement performance. For both subgrade and bituminous failure criteria, the pavement performance (allowable strain repetition) is most sensitive to the variation of the aircraft load P. For the subgrade strain criteria, the pavement performance is more sensitive to variations of the thickness of the granular base layer h_2 (which is placed directly on the subgrade) and the modulus of the subgrade E_3. The pavement performance is less sensitive to the thickness and modulus of the bituminous concrete surface course h_1 and E_1, respectively, and the modulus of the base course E_2. For the bituminous concrete strain criterion, the pavement performance is more sensitive to variations of the thickness and modulus of the bituminous concrete layer h_1 and E_1, respectively, and is less sensitive to variations of the thickness and modulus of the base layer h_2 and E_2, respectively, and the modulus of the subgrade E_3. It is interesting to note that the pavement performance is more sensitive to variations of layer thickness than the elastic modulus of the layer

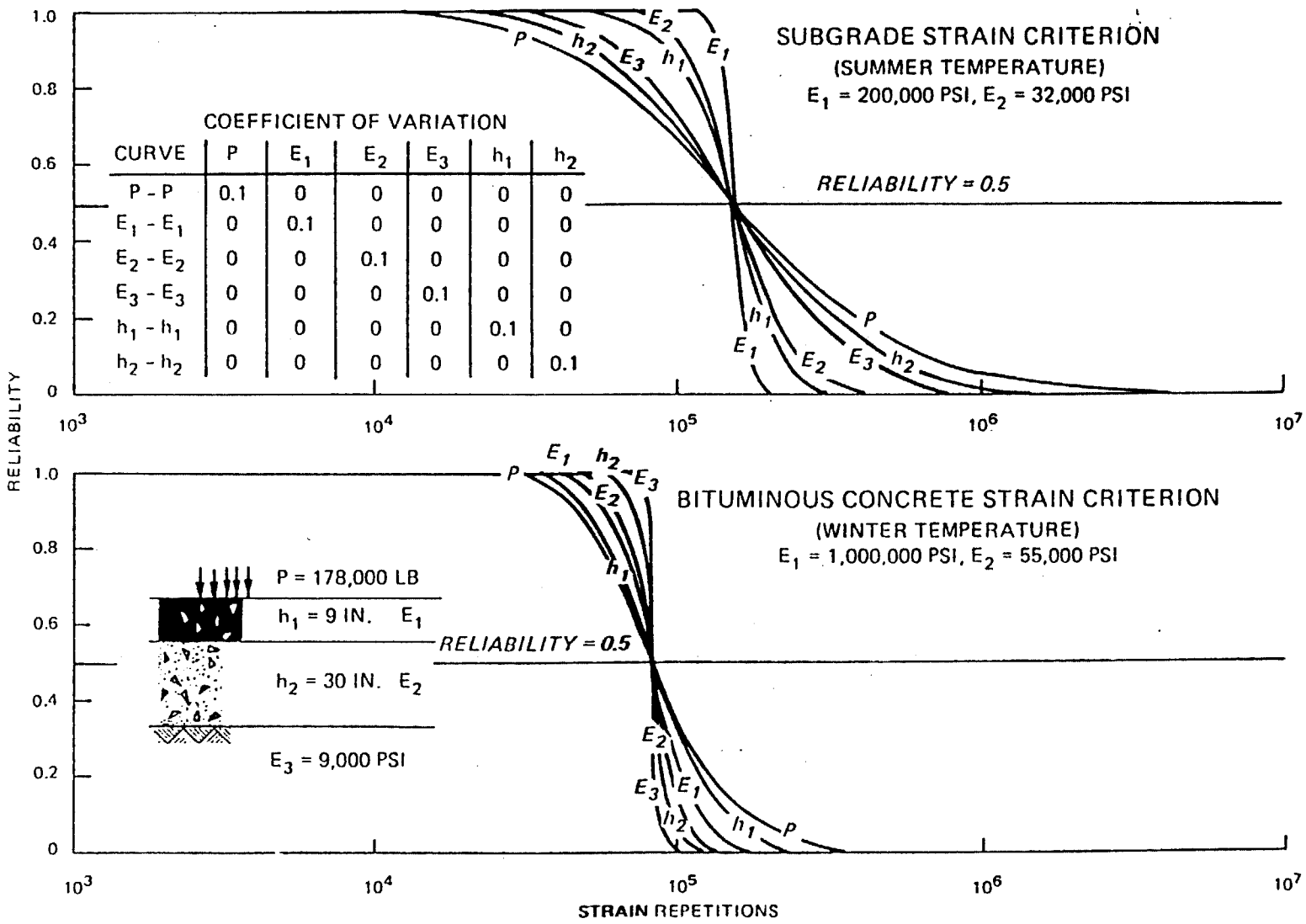


Figure 4. Relationships between variability and strain repetition for flexible pavements with varying CV

material in both failure criteria, and it is least sensitive to the variation of modulus of the bituminous concrete surface layer E_1 in the subgrade strain criterion and to the variation of subgrade modulus E_3 in the bituminous concrete strain criterion. Thus, the performance of a flexible pavement is sensitive in the descending order, to variations of P, h_2, E_3, h_1, E_2, and E_1 in the subgrade strain criterion, and to variations of P, h_1, E_1, E_2, h_2, and E_3 in the bituminous concrete strain criterion.

The significance of the results presented in Figure 4 may also be explained from another viewpoint by using the values listed in Table 1. Table 1 shows the ranges of computed allowable strain repetitions within +1 and -1 standard deviation of the subgrade strain value for six different cases. In each case, the CV of one parameter is equal to 0.1, and the CV's of the other five parameters are set at zero. The subgrade strain value computed for the pavements is 0.0009404 in./in. Table 1 shows that the standard deviation of the subgrade strain is the largest for the load P and is the smallest for the modulus of bituminous concrete surface course E_1. When only the variation of the load P is accounted for, there is a 68.3 percent chance (i.e., the area within +1 and -1 standard deviation under a normal distribution curve) that the predicted pavement performance falls within the range of 56,160 to 479,000 strain repetitions. If only the variation of the modulus of bituminous concrete E_1 is accounted for, the predicted performance for the same percent of chance narrows down to a range from 142,500 to 169,670 strain

Table 1. Flexible Pavement Performance Variations as Functions of Variations of Input Parameters (Subgrade Strain Criterion) Flexible Pavement

Coefficient of Variation						Standard Deviation of Strain	Strain Repetitions for	
E_p	E_1	E_2	h_3	h_1	2	σ_ε*	$\varepsilon + \sigma_\varepsilon$	$\varepsilon - \sigma_\varepsilon$
0.1	0.0	0.0	0.0	0.0	0.0	0.0000940	56,160	479,000
0.0	0.1	0.0	0.0	0.0	0.0	0.0000077	142,500	169,670
0.0	0.0	0.1	0.0	0.0	0.0	0.0000249	117,540	207,120
0.0	0.0	0.0	0.1	0.0	0.0	0.0000618	78,790	321,140
0.0	0.0	0.0	0.0	0.1	0.0	0.0000330	107,580	227,570
0.0	0.0	0.0	0.0	0.0	0.1	0.0000724	70,360	365,940

* The subgrade strain computed for the pavement is 0.0009494 in./in.

repetitions, indicating a smaller variation and thus a design with less uncertainty.

VARIABILITY OF INPUT PARAMETERS OF FLEXIBLE AIRFIELD

The results presented in Figure 4 are based on the analysis assuming that the input parameters have the same coefficient of variation. In reality, some parameters have larger variations than others. Although the pavement performance is more sensitive to the variations of layer thickness than the modulus of the layer, it has been found that the thickness variations in actual field constructions are not very large; the average CV's are generally near 10 percent. (Nevertheless, efforts should be made to reduce pavement thickness variation during construction as much as possible.) The actual variations of moduli of layer materials in the field are known to be very

large. The CV of moduli can be as much as 50 percent or more. More efficient construction methods and equipments should be used, and strict compaction and quality controls should be exercised in construction to reduce material modulus variations. The control of load variation is beyond the jurisdiction of pavement engineers. Since the variation of aircraft load has a large effect on pavement performance, the airfield operators should be informed and advised to limit aircraft overload cases.

SIGNIFICANCE OF FAILURE CRITERIA

All the reliability-strain repetition curves shown in Figures 1 to 4 have steeper slopes for the bituminous concrete strain criterion than for the subgrade strain criterion. It seems that flexible pavements designed using the Corps of Engineer's failure criteria will have a greater degree of uncertainty in preventing subgrade failure than preventing fatigue cracking of bituminous concrete surface course. However, this may not be true for pavements in the field. It is extremely important to point out that the subgrade failure criteria (Equation 13) are based on traffic test data while the bituminous concrete strain criteria (Equation 12) are derived based on laboratory fatigue data. Since failure criteria derived from laboratory tests do not consider the uncertainties existing in the laboratory-to-field correlations, the actual performance of the pavement will be more uncertain than is considered in the design. Even though the slopes of

the reliability-strain repetition curves are steeper in the bituminous concrete strain criterion than in the subgrade strain criterion, it is not necessarily true that pavements designed using the Corps of Engineer's failure criteria will have lesser degree of uncertainty in preventing fatigue cracking of bituminous concrete surface course than preventing subgrade failure. There is a need to incorporate the field uncertainties into the laboratory determined failure criteria. More discussion on the significance of failure criteria on the reliability-strain repetition curves (Figures 1 to 4) is presented in Reference 10.

CONCLUSIONS

Based on the analysis of RELIBISA computer program which is a layered elastic pavement design approach in terms of probability and reliability, the following conclusions can be drawn for flexible airfield pavement design using the layered elastic method.

The relationships between reliability and strain repetition (such as Figures 1 and 2) can be used to design a pavement in terms of probability and reliability. For a desired reliability level, the thickness of the bituminous concrete surface course or the thickness of the granular base can be varied to agree with the designed strain repetition, or the allowable strain repetition can be modified for a given pavement structure.

For the subgrade strain failure criterion, equal changes in the thickness of bituminous concrete surface course or granular base course result in equal changes in allowable strain. Since the unit cost of granular layer is less than the bituminous concrete course, it is economically beneficial to increase the thickness of the granular base to prevent the pavement from subgrade failure. The support from base course has relatively lesser significance in preventing fatigue cracking of the bituminous concrete surface course than increasing the thickness of the bituminous concrete layer itself, and the significance reduces rapidly as the thickness of the granular base continues to increase (Figure 3d).

The performance of a flexible pavement is sensitive for the subgrade strain criterion to variations of the following input parameters (in the descending order) the gear load P, the thickness of the granular base h_2, the subgrade modulus E_3, the thickness of the bituminous concrete surface course h_1, the granular base modulus E_2, and the modulus of the bituminous concrete surface course E_1. For the bituminous concrete strain criterion pavement performance is sensitive to variations of P, h_1, E_1, E_2, h_2, and E_3. Although pavement performance is generally more sensitive to the variation of layer thickness than to that of the elastic modulus of the layer material, actual variations of layer thickness in the field are known to be lesser than variations of material moduli.

The results of the reliability analysis are very much influenced by the nature of the failure criteria employed. There is a need to

incorporate the field uncertainties into the laboratory determined failure criteria.

The relationships between reliability and strain repetition developed for flexible pavements (Figures 1 to 4) can be used to optimize the design. The thicknesses of the bituminous concrete and the granular layers can be selected so that the pavement is failed in fatigue cracking of the bituminous concrete and subgrade failure at nearly the same traffic level and the same reliability level.

ACKNOWLEDGEMENTS

This research was sponsored by the US Army Engineer Waterways Experiment Station (WES). The opinions expressed in this paper are the views of the author and do not represent the views of the Department of the Army. Permission was granted by the Chief of Engineers to publish this information.

REFERENCES

1. Witczak, M. W., Uzan, J., and Johnson, M. 1983 (Dec). "Development of Probabilistic Rigid Pavement Design Methodologies for Military Airfields," Technical Report GL-83-18, US Army Engineer Waterways Experiment Station, Vicksburg, Miss.

2. Westergaard, H. M. 1948. "New Formulas for Stresses in Concrete Pavements of Airfields," Transactions, American Society of Civil Engineers, Vol 113, p 425-439.

3. Parker, F., Jr., Barker, W. R., Gunkel, R. C., and Odom, E. C. 1979. "Development of a Structural Design Procedure for Rigid Airport Pavements," Technical Report GL-79-4, US Army Engineer Waterways Experiment Station, Vicksburg, Miss.

4. Benjamin, J. R., and Cornell, C. A. 1970. Probability, Statistics, and Decision for Civil Engineers, McGraw-Hill, New York.

5. Chou, Y. T. 1986. "Probabilistic and Reliability Analysis of the California Bearing Ratio (CBR) Design Method for Flexible Airfield Pavements," Technical Report GL-86-15, US Army Engineer Waterways Experiment Station, Vicksburg, Miss.

6. Hammitt, G. M., II, Hutchington, R. L., Rice, J. L., Thompson, O. O., and Brown, D. N. 1971 (Nov). "Multiple-Wheel Heavy Gear Load Pavement Tests; Volume IV, Analysis of Behavior Under Traffic," Technical Report S-71-17, US Army Engineer Waterways Experiment Station, Vicksburg, Miss.

7. Rosenblueth, Emilio. 1975 (Oct). "Point Estimates for Probability Moments," Proceedings, National Academy of Sciences, USA, Mathematics, Vol 72, No. 10, pp 3812-3814.

8. Koninklijke/Shell Laboratorium. 1972. "BISAR Users Manual; Layered System Under Normal and Tangential Loads," Amsterdam, Holland.

9. Darter, M. I., and Hudson, W. R. 1973 (May). "Probabilistic Design Concepts Applied to Flexible Pavement System Design," Report 123-18, Center for Highway Research, University of Texas, Austin, Tex.

10. Chou, Y. T. 1987 (Jul). "Probabilistic and Reliability Design Procedures for Flexible Airfield Pavements, Elastic Layered Method," Technical Report GL-87-24, US Army Engineer Waterways Experiment Station, Vicksburg, Miss.

11. Potter, J. 1985 (Sep). "Reliability of the Flexible Pavement Design Model," Miscellaneous Paper GL-85-27, US Army Engineer Waterways Experiment Station, Vicksburg, Miss.

Rick Nordheim University of Wisconsin October 1990

Nimble Consultant's Articles of Advice

Field of 64

Before first visit ar early in consulting relationship

1 Learn what you can about the subject matter in client's field.
2 Learn what you can about client and client's group.
3 At the beginning of your relationship with client, agree on expectations; what will be done, by when, and at what cost.

Particularly during first consulting session

4 Find out what client's major objective is.
5 Recognize that client will often have difficulty articulating major objective.
6 Find out specifically what data are available, how they were obtained, and what are their units.
7 Determine relation between particular issue under discussion and overall research goals.
8 Find out about previous studies similar to the one under discussion.
9 Find out what statistical techniques are used in client's field.
10 Learn about the limits and constraints on client's problem.
11 Try to determine what client knows.
12 Try to determine client's view of statistics and statisticians.
13 Formulate goals as precisely as possible.
14 Determine who real decision-maker is.
15 Make client feel comfortable; avoid scolding and excess criticism.
16 Be prepared to spend some time on pleasantries.
17 Reassure a frightened client.
18 Avoid any adversarial relationship with client.
19 Do not let client lead you up the garden path.
20 Do not feel embarrassed when you do not know something.
21 Encourage client to expound on relevant substantive issues.
22 Try to evolve a relationship as collaborative equals.
23 Have a clear understanding of consulting service philosophy.
24 Be careful of the 'five-minute' question.

Preceding Page Blank

During all consulting sessions

25 Where possible, make client feel like an expert.
26 Rephrase in your own words the major points made by client. (Do this often.)
27 Encourage client to rephrase major points.
28 At end of session, clarify in writing major decisions and subsequent goals for consultant and client.
29 In suggesting solutions, present options and tradeoffs.
30 Encourage interchange among multiple clients.
31 Listen carefully to the throwaway line.
32 Be aware of body language.
33 Be friendly and patient.
34 Take good notes.
35 Have your homework done for every meeting.
36 Ask lots of questions.
37 Make sure the consulting environment is pleasant.
38 Be careful with client-consultant seating arrangements.
39 Avoid excessive interruption.
40 Avoid excessive lecturing.
41 Avoid appearing too theoretical.
42 Be aware of 'the politics' of a study (if such exists).
43 Make the statistical procedure fit the problem rather than the other way around.
44 Encourage pilot studies.
45 Probe into underlying assumptions.
46 Emphasize exploratory plots; make some during session if appropriate.
47 Tailor your suggestions to the capabilities of the client.
48 Realize a major role of consultant is helping client clarify understanding.
49 Do not hesitate to make general recommendations to general experimental effort.
50 Realize you can always make a contribution.
51 Encourage subsequent visits by client at the earliest possible stage of the project.
52 Visit the client's home turf if possible.

While working on problem under discussion

53 Pay attention to details of design and experiment management.
54 Make sure data and computer output are carefully and continually scrutinized.
55 Examine carefully time-order of data.
56 Think carefully about what is (are) the experimental unit(s).
57 Enumerate sources of error.
58 Dig hard on diagnostics.
59 Dig into literature, when necessary, in statistics or client's field.
60 Find the simplest solution that does the job.
61 Be aware of political reality.
62 Be prepared to make some compromises with the real world.
63 Meet all deadlines; approximate solutions on time usually better than definitive solution late.
64 Make written reports clear with major points emphasized, with lots of plots, and with minimum of excessive technical detail.

THE SURVIVAL PROBABILITY FUNCTION OF A TARGET MOVING ALONG A STRAIGHT LINE IN A RANDOM FIELD OF OBSCURING ELEMENTS

S. Zacks and M. Yadin

State University of New York at Binghamton

Binghamton, NY 13901

ABSTRACT. A target is moving along a straight line path. Random portions of the path might be invisible to the hunter (in shadow). Shooting trials are conducted only along the visible segments of the path. An algorithm for the numerical determination of the survival probability of the target is developed. This algorithm is based on the distribution of shadow length which is also developed.

Key Words: *Lines of sight*; *visibility probabilities*; *distributions of shadows*; *survival probability.*

Acknowledgement: Research supported by the U.S. Army Research Office, Contract DAAL03-89-K-0129 with the State University of New York at Binghamton.

1. INTRODUCTION.

The present study is focused on the problem of determining the survival probability of a moving target, which is under attack by a hunter. The target (vehicle, tank, etc.) is moving along a straight line path, which is partially obscured from the hunter by randomly distributed objects (trees, clouds, terrain objects, etc.). The target can be destroyed by the hunter only along the visible segments of the path. Visibility contact between the hunter and the target is needed for τ_0 time units for a shooting trial to occur. In any given shooting trial the probability that the target is destroyed is fixed. If the target survives a shooting trial, another identical trial may be attempted if continuous visibility for τ_0 time units is possible. If the target enters an obscured segment of the path, the shooting trials terminate, until visibility contact is reestablished. Under the above assumptions, if the target has to cross a visible segment of length L, its survival probability can be approximated by the negative exponential function $\exp\{-qL\}$, for suitably chosen constant q, $0 < q < \infty$. The problem is that the number of visible segments on the moving path, between two specified points $\mathbf{P}_L$ and $\mathbf{P}_U$, and their lengths are random variables, whose distributions depend on the characteristics of the random field.

The present study is based on the model of a random Poisson field of obscuring elements. This model is presented in Section 2. Under the assumptions of this model, it is relatively

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

simple to derive the conditional distribution of the length of a visible segment on the right hand side (r.h.s.) of a point, $\mathbf{P}_x$, on the path, given that the point $\mathbf{P}_x$ is visible. This distribution is given in Section 3.1. On the other hand, it is more complicated to determine the distribution of the length of a segment which is obscured (in shadow). In Section 3.2 we present the methodology for determining the distribution of the length of shadows. This methodology is based on a theory given by Chernoff and Daly (1957). For a given point $\mathbf{P}_x$ on the moving path, Chernoff and Daly (C-D) define the functional $T(x)$, which is the right hand limit of the shadow to the r.h.s. of P_x, cast by obscuring elements in the field which intersect the ray $\mathbf{R}_x$ from the origin $\mathbf{O}$ through $\mathbf{P}_x$. Employing functions $K_{\pm}(x,y)$, which are defined in Section 2 and derived explicitly for the standard-uniform case in the Appendix, we express the cumulative distribution function (c.d.f.) of $T(x)$ explicitly. The right hand limit of a shadow to the r.h.s. of $\mathbf{P}_x$, is $U(x) = \lim_{n\to\infty} T^n(x)$, where $T^{n+1}(x) = T(T^n(x))$, for $n = 0,1,\cdots$, $T^0(x) = x$. The relationship between the c.d.f. of $T^{n+1}(x)$ to that of $T^n(x)$, $n = 0,1,\cdots$ is discussed in Section 3.2. The distribution of $U(x)$ is obtained as a limit of that of $T^n(x)$. From the distribution of $U(x)$ we obtain the conditional distribution of the right end limit of a shadow to the r.h.s. of $\mathbf{P}_x$, given that $\mathbf{P}_x$ is the first point in the shadow.

In Section 4 we employ the results of Section 3 to approximate the survival probability function $S(x,y)$ along the moving path between the points $\mathbf{P}_x$ and $\mathbf{P}_y$, $x < y$. The function $S(x,y)$ is given by the integral equation

$$S(x,y) = A(x,y) + \int_x^y B(x,w)S(w,y)dw \tag{1.1}$$

where $A(x,y)$ and $B(x,y)$ are defined in terms of the distributions of the lengths of visible and non-visible random segments, as shown in Section 4. An algorithm for the discrete approximation of the solution of (1.1) is given in Section 5. Numerical solutions based on this algorithm are provided there too. A Quick Basic program (version 4.5) for computations can be obtained upon request.

In a previous Technical Report [7] we approximated the survival probabilities by deriving lower and upper bounds to the distribution of the number of shooting trials, N, along the path. The present study provides the method of computing the survival probability function $S(x,y)$, which is required for various applications. With the new algorithms for determining distributions of shadows and survival functions we can tackle problems like the Hunter-Escort problem, which will be discussed in another paper.

2. THE RANDOM FIELD MODEL AND THE DETERMINATION OF VISIBILITY PROBABILITIES.

In the present paper we consider a two dimensional physical model. A generalization to a three dimensional model can be done in a similar fashion to that of Yadin and Zacks [5]. The moving path of the target is a straight line $\mathcal{C}$. The Hunter is located at a point $\mathbf{O}$ (the origin), at distance r from $\mathcal{C}$. Let $\mathcal{U}$ and $\mathcal{W}$ be two straight lines parallel to $\mathcal{C}$, located between $\mathbf{O}$ and $\mathcal{C}$, at distances u and w from $\mathbf{O}$, respectively; $0 < u < w < r$. The obscuring objects are modeled by a countable number of disks of random size, which are centered at random points in the strip $\mathcal{S}$, bounded by $\mathcal{U}$ and $\mathcal{W}$. We consider a cartesian coordinate system in which the y-axis is a straight line through $\mathbf{O}$, perpendicular to $\mathcal{C}$, which intersects $\mathcal{U}$, $\mathcal{W}$ and $\mathcal{C}$ at the point $(0,u)$, $(0,w)$ and $(0,r)$, respectively. A point $\mathbf{P}_x$ on $\mathcal{C}$ has coordinates (x,r).

A random disk is represented by the random vector (X,Y,Z), where (X,Y) are the random coordinates of the center of the disk, and Z is its random radius. Without loss of generality, assume that the sample space of (X,Y,Z) is $\mathcal{S}^* = \mathcal{S} \times [a,b]$, where $0 < a < b < \infty$. Let $\mathcal{B}^*$ be the Borel σ-field on $\mathcal{S}^*$. Let $\{(X_i,Y_i,Z_i),\ i = 1,2,\cdots\}$ represent a sequence of countable random disks measurable w.r.t. the same space $(\mathcal{S}^*,\mathcal{B}^*,P)$. It is assumed that the random vectors are independent and identically distributed (i.i.d.), and have a common distribution $H(x,y,z)$. Let $F(z \mid x,y)$ denote the conditional c.d.f. of the radius Z, given the center (X,Y) is at (x,y). Let $h(x,y)$ be the joint p.d.f. of (X,Y), such that $h(x,y) = 0$ for all $(x,y) \notin \mathcal{S}$. We further assume that the probability that a random disk intersects either $\mathbf{O}$ or $\mathcal{C}$ is zero. Let B be any Borel set in $\mathcal{B}^*$. Let $N\{B\}$ designate the number of random disks with coordinates in B.

If $\{B_1,\cdots,B_m\}$ is any finite partition of $\mathcal{S}^*$, $m = 1,2,\cdots$, it is assumed that the random variables $N\{B_i\}$, $i = 1,\cdots,m$ are independent, having Poisson distributions with expected values

$$\mu\{B_i\} = \lambda \iiint_{B_i} dH(x,y,z), \qquad i = 1,\cdots,m, \tag{2.1}$$

$0 < \lambda < \infty$. Such a random field is called a *Poisson field.* The Poisson field is called *standard-uniform* if $dH(x,y,z) = hI_C(x,y)f(z)dxdydz$, where $0 < h < \infty$, C is a subset of $\mathcal{S}$ which represents the field of view of the Hunter, and $I_C(x,y)$ is the indicator function of C. A point $\mathbf{P}_x$ on $\mathcal{C}$ is said to be *visible* from $\mathbf{O}$, if the ray $\mathbf{R}_x$ from $\mathbf{O}$ through $\mathbf{P}_x$ is not intersected by random disks. In a similar manner we can define the notion of simultaneous visibility of several points on $\mathcal{C}$. In our previous papers [2,3,4] we have introduced the functions $K_+(x,t)$ and $K_-(x,t)$ for $0 < t < \infty$; where $\lambda K_\pm(x,t)$ is the expected number

of disks centered in S between the rays $\mathbb{R}_x$ and $\mathbb{R}_{x\pm t}$, which *do not* intersect $\mathbb{R}_x$. Explicit formulae for $K_\pm(x,t)$, for the standard-uniform case, with a uniform distribution of radii on $[a,b]$, $0 < a < b$, is given in the appendix.

Let $[L,U]$ be an interval of the x-coordinates of the point on C belonging to a segment of interest. Let $L^* < L$ and $U^* > U$ be properly chosen, and C^* the set in S (trapez) between the rays $\mathbb{R}_{L^*}$ and $\mathbb{R}_{U^*}$. One can verify that the probability that $\mathbf{P}_x$ is visible, for some $L < x < U$, is

$$\psi(x) = \exp\{-[\mu\{C^*\} - \lambda K_-(x, x-L^*) - \lambda K_+(x, U^*-x)]\}. \tag{2.2}$$

For a formula of the simultaneous visibility of n points in $[L,U]$, see Yadin and Zacks [4].

3. DISTRIBUTIONS OF LENGTH OF VISIBLE AND OF SHADOWED SEGMENTS.

3.1. DISTRIBUTIONS OF THE LENGTH OF VISIBLE SEGMENTS.

In the present section we derive a formula for the conditional c.d.f. of the length of a visible segment to the r.h.s. of $\mathbf{P}_x$, given that $\mathbf{P}_x$ is visible.

Let $I(x)$ be an indicator function which assumes the value 1 if $\mathbf{P}_x$ is visible, and the value zero otherwise.

Let $L(x)$ be the length of the visible segment of C to the r.h.s. of $\mathbf{P}_x$, i.e.,

$$L(x) = \inf\{y : y \geq x, \prod_{x<u\leq y} I(u) = 1\} - x. \tag{3.1}$$

We derive here the formula for

$$\begin{aligned} V(l \mid x) &= P\{L(x) \leq l \mid I(x) = 1\}. \\ &= 1 - P\{L(x) > l \mid I(x) = 1\}. \end{aligned} \tag{3.2}$$

Let C^* be the set of (x,y) points in S, which was defined in the previous section. We derive the formula of $V(l \mid x)$, for $L < x < U$, and $0 \leq l \leq u - x$.

Let $C_-(x)$ be the set bounded by $\mathcal{U}$, $\mathcal{W}$ and the rays $\mathbb{R}_{L^*}$ and $\mathbb{R}_x$. Let $C(x,l)$ be the set bounded by $\mathcal{U}$, $\mathcal{W}$ and the rays $\mathbb{R}_x$, $\mathbb{R}_{x+l}$; and $C_+(l+x)$ the set bounded by $\mathcal{U}$, $\mathcal{W}$, $\mathbb{R}_{l+x}$ and $\mathbb{R}_{U^*}$. Notice that $C^* = C_-(x) \cup C(x,l) \cup C_+(l+x)$. As before, we denote by $\mu\{C\}$ the expected number of disks having centers at the set C, as given by (2.1). Accordingly,

$$\begin{aligned} P\{L(x) > l, I(x) = 1\} &= \exp\{-[\mu\{C_-(x)\} - \lambda K_-(x, x-L^*)] - \mu\{C(x,l)\} \\ &\quad - [\mu\{C_+(l+x)\} - \lambda K_+(l+x, U^*-l-x)]\} \\ &= \exp\{-\mu\{C^*\} + \lambda[K_-(x, x-L^*) + K_+(l+x, U^*-l-x)]\}. \end{aligned} \tag{3.3}$$

Dividing (3.3) by (2.2) we obtain

$$P\{L(x) > l \mid I(x) = 1\} = \exp\{-\lambda[K_+(x, U^* - x) - K_+(l + x, U^* - l - x)]\}. \tag{3.4}$$

3.2. THE DISTRIBUTION OF SHADOW LENGTH.

We have denoted by $U(x)$ the right hand limit of the shadow on C to the r.h.s. of $\mathbf{P}_x$. Let $D(u \mid x)$ denote the conditional c.d.f. of $U(x)$, given that the shadow starts at $\mathbf{P}_x$.

Consider the rays $\mathbb{R}_x$ and $\mathbb{R}_y$ for $y > x$. Let $N(x, y)$ denote the number of disks centered in S, which intersect both $\mathbb{R}_x$ and $\mathbb{R}_y$. Define the functional

$$T(x) = \sup\{y : N(x, y) \geq 1\}. \tag{3.5}$$

Furthermore, let $T^{i+1}(x) = T(T^i(x))$, $i = 0, 1, \cdots$ where $T^0(x) = x$. Obviously, $T^{i+1}(x) \geq T^i(x)$, for all $i \geq 0$, and therefore $U(x) = \lim_{i \to \infty} T^i(x)$. $U(x) - x$ is the length of the shadow to the r.h.s. of $\mathbf{P}_x$. We derive first the c.d.f. of $T(x)$. Clearly, $\{T(x) > t\} = \{N(x, t) \geq 1\}$. Thus,

$$P\{T(x) \leq t\} = P\{N(x, t) = 0\} = \exp\{-\mu(x, t)\}, \tag{3.6}$$

where $\mu(x, t) = E\{N(x, t)\}$. Furthermore,

$$\mu(x, t) = \mu\{C^*\} - \lambda K_+(x, U^* - x) - \lambda K_-(t, t - L^*) + \lambda K_+(x, \tilde{t} - x) + \lambda K_-(t, t - \tilde{t}), \tag{3.7}$$

where $\tilde{t}$ is the coordinate of the bisector between $\mathbb{R}_x$ and $\mathbb{R}_t$. Notice that, since $K_+(x, 0) = K_-(x, 0) = 0$ for all x,

$$\begin{aligned} \mu(x, x) &= \lim_{t \downarrow x} \mu(x, t) \\ &= \mu\{C^*\} - \lambda K_+(x, u^* - x) - \lambda K_-(x, x - L^*). \end{aligned} \tag{3.8}$$

Hence,

$$\lim_{t \downarrow x} P\{T(x) \leq t\} = \psi(x), \tag{3.9}$$

which is the probability that $\mathbf{P}_x$ is visible. Thus, the c.d.f. of $T(x)$, $H(t; x)$ is zero for $t < x$, it has a jump point at x, $H(x; x) = \psi(x)$, and is absolutely continuous at $t > x$. This property is inherited by the c.d.f. of $T^n(x)$, $H_n(t; x)$. We provide now the recursive relationship between $H_n(t; x)$ and $H_{n-1}(t; x)$. Introduce the bivariate distribution

$$G_n(t_1, t_2; x) = P\{T^{n-1}(x) \leq t_1, T^n(x) \leq t_2\}.$$

Since $\{T^n(x) \le t\} \subset \{T^{n-1}(x) \le t\}$,

$$\begin{aligned} H_n(t;x) &= P\{T^n(x) \le t\} \\ &= G_n(t^*, t; x), \quad \text{all } t^* \ge t. \end{aligned} \tag{3.10}$$

For $x < z < y < t$,

$$P\{T^n(x) \le t \mid T^{n-2}(x) = z, T^{n-1}(x) = y\} = \exp\{-[\mu(y,t) - \mu(z,t)]\}. \tag{3.11}$$

Indeed, given that $\{T^{n-2}(x) = z, T^{n-1}(x) = y\}$, $\{T^n(x) > t\}$ if and only if, there exists at least one disk which intersects $\mathbb{R}_y$ and $\mathbb{R}_t$, but *does not* intersect $\mathbb{R}_z$. Hence,

$$G_n(t_1, t_2; x) = \int_x^{t_1} \int_z^{t_1} \exp\{-[\mu(u,t_2) - \mu(z,t_2)]\} dG_{n-1}(z,u;x). \tag{3.12}$$

These bivariate c.d.f. can be determined recursively, starting with $G_1(t_1, t_2; x) = H(t_2; x)$ for all $t_1 \le t_2$. Moreover,

$$G_2(t_1, t_2; x) = \int_x^{t_1} e^{-\mu(u,t_2)} \left(\int_x^u e^{\mu(z,t_2)} dz \right) dH(u;x). \tag{3.13}$$

Finally, since $H_{n+1}(t;x) \le H_n(t;x)$ for each $t \ge x$ and all $n = 1, 2, \cdots$ the c.d.f. of $U(x)$ is

$$P\{U(x) \le t\} = \lim_{n\to\infty} H_n(t;x). \tag{3.14}$$

Thus, $P\{U(x) \le t\} = 0$ for all $t < x$, and $\lim_{t \downarrow x} P\{U(x) \le t\} = \psi(x)$. The conditional c.d.f. of $U(x)$, given $\{I(x) = 0\}$ is

$$P\{U(x) \le t \mid I(x) = 0\} = \begin{cases} \dfrac{P\{U(x) \le t\} - \psi(x)}{1 - \psi(x)}, & \text{for } t \ge x \\ 0, & \text{for } t < x. \end{cases} \tag{3.15}$$

We are interested, however, in the conditional c.d.f. $D(u \mid x)$, where $\mathbf{P}_x$ is the first point (the left hand limit) of the random segment in shadow.

Simple geometric considerations yield that the length of a random shadow cast by a *single* disk, having left hand limit at $\mathbf{P}_x$, with center on a line parallel to $\mathcal{U}$ at distance h from $\mathbf{O}$, and disk radius Z, is

$$\tilde{U}(x, h, Z) = r \tan\left(2\sin^{-1}\left(\frac{Z}{\sqrt{h^2 + x_c^2}}\right) + \tan^{-1}\left(\frac{x}{r}\right)\right) - x, \tag{3.16}$$

where (x_c, h) are the coordinates of the center of the disk, with

$$x_c = \frac{x}{r}\, h + Z\left(1 + \left(\frac{x}{r}\right)^2\right)^{1/2}. \tag{3.17}$$

Thus, if $a \le Z \le b$ w.p. 1, the minimal length of shadow starting at x is

$$\tilde{u}_m(x) = r\ \tan\left(2\ \sin^{-1}(B(x)) + \tan^{-1}(\frac{x}{r})\right) - x \tag{3.18}$$

where

$$B(x) = \frac{a}{w}\left[1 + \left(\frac{x}{r} + \frac{a}{w}\left(1 + \left(\frac{x}{r}\right)^2\right)^{1/2}\right)^2\right]^{1/2}. \tag{3.19}$$

In Section 5 we provide computing algorithms for the numerical determination of the c.d.f.'s $V(l \mid x)$ and $D(u \mid x)$, and illustrate them with a numerical example.

4. THE SURVIVAL PROBABILITY FUNCTION.

In the present section we establish the integral equation (1.1). Let $\mathbf{P}_x$ and $\mathbf{P}_y$ be a visible point on C and a point to its right, $L^* < x \le y < U^*$. The Hunter starts shooting trials when the target is at $\mathbf{P}_x$. The attack terminates when the target reaches $\mathbf{P}_y$, if it has not been destroyed before. Let $S(x, y)$ designate the survival probability function. We recognize three exclusive and exhaustive events.

(i) The visible segment to the r.h.s. of $\mathbf{P}_x$ terminates at a point to the right of $\mathbf{P}_y$;

(ii) The visible segment on the r.h.s. of $\mathbf{P}_x$ terminates at a point $\mathbf{P}_t$, $t < y$, and the length of the shadow starting at $\mathbf{P}_t$ is greater than $y - t$.

(iii) The visible segment on the r.h.s. of $\mathbf{P}_x$ terminates at a point $\mathbf{P}_t$, $t < y$, and the length of the shadow to the r.h.s. of $\mathbf{P}_t$ is smaller than $y - t$.

As mentioned in Section 1, the conditional survival probability of a target moving on a visible segment of length L is $\exp\{-q, L\}$, for some $0 < q < \infty$. Accordingly,

$$\begin{aligned} S(x,y) = {} & e^{-q(y-x)}(1 - V(y - x \mid x)) \\ & + \int_x^y e^{-q(t-x)}(1 - D(y \mid t))dV(t - x \mid x) \\ & + \int_x^y e^{-q(t-x)} \left\{ \int_{\tilde{u}_m(t)+t}^y S(z,y)D'(z \mid t)dz \right\} dV(t - x \mid x), \end{aligned} \tag{4.1}$$

where $D'(z \mid t) = \dfrac{\partial}{\partial z} D(z \mid t)$ is the p.d.f. of $D(z \mid t)$. Notice that $D'(z \mid t) = 0$ for all $t \le z \le \tilde{u}_m(t) + t$. Let $z_m(t) = \tilde{u}_m(t) + t$. $z_m(t)$ is the first term on the r.h.s. of (3.18)

with $x = t$. Let $t_m(z)$ be the inverse of $z_m(t)$ then, by changing the order of integration we obtain

$$\begin{aligned}\int_x^y e^{-q(t-x)} \left\{ \int_{\ddot{u}_m(t)+t}^y S(z,y) D'(z \mid t) dz \right\} dV(t-x \mid x) \\ = \int_x^y S(z,y) \left\{ \int_0^{t_m(x)-x} e^{-qt} D'(z \mid t+x) dV(t \mid x) \right\} dz.\end{aligned} \tag{4.2}$$

Accordingly, define

$$A(x,y) = e^{-q(y-x)}(1 - V(y-x \mid x)) + \int_0^{y-x} e^{-qt}(1 - D(y \mid t+x)) dV(t \mid x), \tag{4.3}$$

and

$$B(x,z) = \int_0^{t_m(x)-x)} e^{-qt} D'(z \mid t+x) dV(t \mid x). \tag{4.4}$$

Thus, the integral equation (4.1) can be written as in (1.1).

5. ALGORITHMS FOR DISCRETE APPROXIMATIONS AND NUMERICAL EXAMPLES.

In the present section we consider discrete approximations to the functions $H_n(t;x)$, $G_n(t_1,t_2;x)$, $n \geq 2$ and $S(x,y)$.

For a given integer, N, partition the interval (x,y) to N subintervals. Accordingly, let $\delta = (y-x)/N$, $t_0 = x$ and $t_j = t_0 + j\delta$, $j = 0,1,\cdots,N$.

For $i = 0,\cdots,N$, let

$$\hat{H}_1(i) = H(t_i;t_0) = \exp\{-\mu(t_0,t_i)\}. \tag{5.1}$$

For $i = 0,\cdots,N$ and $j = i,\cdots,N$, let

$$\hat{G}_2(i,j) = \sum_{k=0}^{i} \exp\{-(\mu(t_k,t_j) - \mu(t_0,t_j))\} \cdot [\hat{H}_1(k) - \hat{H}_1(k-1)], \tag{5.2}$$

where $\hat{H}_1(-1) = 0$. This is an approximation to (3.13). Notice that $\hat{G}_2(0,j) = \hat{H}_1(0)$ for all $j = 0,1,\cdots,N$; and for $i \geq 1$, $j \geq i$

$$\hat{G}_2(i,j) = \hat{H}_1(0) + \exp\{\mu(t_0,t_j)\} \cdot \sum_{k=1}^{i} \exp\{-\mu(t_k,t_j)\}[\hat{H}_1(k) - \hat{H}_1(k-1)]. \tag{5.3}$$

Moreover,

$$\hat{G}_2(i,j) = \hat{G}_2(j,j) \text{ for all } i > j. \tag{5.4}$$

We compute afterwards recursively, for every $n \geq 3$, $i = 1, \cdots, N$, $j = i, \cdots, N$

$$\hat{G}_n(i,j) = \sum_{k=0}^{i} \sum_{l=k}^{i} \exp\{-(\mu(t_l, t_j) - \mu(t_k, t_j))\}[\hat{G}_{n-1}(k,l) - \hat{G}_{n-1}(k-1,l) \\ - \hat{G}_{n-1}(k,l-1) + \hat{G}_{n-1}(k-1,l-1)], \tag{5.5}$$

and for $i > j$

$$\hat{G}_n(i,j) = \hat{G}_n(j,j). \tag{5.6}$$

For $i = 0$, $\hat{G}_n(0,j) = \hat{G}_{n-1}(0,j) = \hat{H}_1(0)$, $j = 0, \cdots, N$. After computing these functions we determine $\hat{H}_n(j) = \hat{G}_n(j,j)$, $j = 0, 1, \cdots, N$. $\hat{H}_n(j)$ is the discrete approximation to the c.d.f. of $T^n(x)$, namely $H_n(t;x)$; i.e., $H_n(t_j;x) \approx \hat{H}_n(j)$.

In Table 5.1 we present numerical results obtained by applying this algorithm to the following special case.

We consider a standard-uniform Poisson field, with uniform distribution for the disk radius on the interval (a,b). In the appendix we present the functions $K_{\pm}(x,t)$, $t \geq 0$, for this case. We compute the numerical example for Table 5.1 with the following geometrical parameters: $r = 100[m]$, $u = 40[m]$, $w = 60[m]$, $a = 1[m]$, $b = 2.5[m]$, $x = 10[m]$, $L^* = -100[m]$, $U^* = 100[m]$. We present in the tables the values of $\hat{H}_n(j)$, $n = 1,2,3$, $j = 0, \cdots, 20$, when $\delta = 1[m]$.

As seen in Table 5.1, the convergence of $\hat{H}_n(j)$ to the c.d.f. of $U(x)$ is quite rapid. We have therefore approximated the c.d.f. $D(u \mid x)$ by the sequence $\hat{D}(j \mid i) = D(t_j \mid t_i)$, $i = 0, 1, \cdots, N$, $j = i, i+1, \cdots$. The function $A(x,y)$ was computed for the arguments t_i, t_j, by the approximation

$$\begin{aligned} &\hat{A}(N,N) = 1 \\ &\hat{A}(N-1,N) = e^{-q\delta}(1 - \hat{V}(1 \mid N-1)) + (1 - \tilde{D}(N \mid N-1)) \\ &\hat{A}(N-j,N) = e^{-jq\delta}(1 - \hat{V}(j \mid N-j)) \\ &+ \sum_{l=1}^{j} e^{-q(l-\frac{1}{2})\delta}(\hat{V}(l \mid N-j) - \hat{V}(l-1 \mid N-j)) \cdot \\ &\cdot (1 - \tilde{D}(N \mid N-j+l)), \quad j = 2, \cdots, N \end{aligned} \tag{5.7}$$

where

$$\tilde{D}(N \mid N-i) = \frac{1}{2}(\hat{D}(N \mid N-i) + \hat{D}(N \mid N-i+1)), \tag{5.8}$$

for all $i = 1, 2, \cdots, N$. Recall that $\hat{D}(N \mid N) = 0 = \hat{D}(N \mid N+1)$.

Table 5.1. *Values of $\hat{H}_n(j)$ for two values of λ.*

		$\lambda = 0.02\ [1/m^2]$			$\lambda = 0.2\ [1/m^2]$	
j	$\hat{H}_1(j)$	$\hat{H}_2(j)$	$\hat{H}_3(j)$	$\hat{H}_1(j)$	$\hat{H}_2(j)$	$\hat{H}_3(j)$
0	0.4914	0.4914	0.4914	0.0000	0.0000	0.0000
1	0.5434	0.5434	0.5434	0.0000	0.0000	0.0000
2	0.6009	0.6008	0.6008	0.0000	0.0000	0.0000
3	0.6645	0.6642	0.6442	0.0003	0.0003	0.0003
4	0.7342	0.7336	0.7336	0.0021	0.0020	0.0020
5	0.8039	0.8030	0.8030	0.0130	0.0119	0.0119
6	0.8664	0.8652	0.8652	0.0576	0.0522	0.0522
7	0.9182	0.9170	0.9170	0.1828	0.1663	0.1663
8	0.9569	0.9559	0.9559	0.4164	0.3843	0.3842
9	0.9811	0.9804	0.9804	0.6842	0.6484	0.6482
10	0.9934	0.9930	0.9930	0.8763	0.8507	0.8505
11	0.9985	0.9984	0.9984	0.9707	0.9590	0.9588
12	0.9999	0.9999	0.9999	0.9981	0.9958	0.9957
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Similarly, we define

$$\begin{aligned} \hat{B}(N,N) &= 0 \\ \hat{B}(N-1,N) &= \frac{1}{2}e^{-q\delta/2}\hat{V}(1 \mid N-1)\hat{D}(N \mid N-1), \end{aligned} \tag{5.9}$$

and for $j = 2, \cdots, N$, $l = 1, \cdots, j$ we compute

$$\begin{aligned} \hat{B}(N-j, N-j+l) = \sum_{i=1}^{l} e^{-q\delta(i-\frac{1}{2})}[\hat{V}(i \mid N-j) - \hat{V}(i-1 \mid N-j)]\cdot \\ \cdot [\hat{D}(N-j+l \mid N-j+i) - \hat{D}(N-j+l-1 \mid N-j+i)], \end{aligned} \tag{5.10}$$

where $\hat{D}(N-j+l-1 \mid N-j+l) = 0$.

Table 5.2. *Survival Probabilities* $\hat{S}(N-j,N)$, *for* $\lambda = 0.02(0.01)0.05$, $N = 20$; $a = 2$, $b = 3.5$, $r = 100$, $u = 40$, $w = 60$.

$j \backslash \lambda$	0.02	0.03	0.04	0.05
0	1.00000	1.00000	1.00000	1.00000
1	0.81711	0.82447	0.83112	0.83715
2	0.69730	0.72040	0.74052	0.75808
3	0.61877	0.65867	0.69187	0.71965
4	0.56729	0.62203	0.66572	0.70094
5	0.53352	0.60025	0.65163	0.69179
6	0.51135	0.58728	0.64401	0.68729
7	0.49677	0.57954	0.63987	0.68505
8	0.48718	0.57489	0.63759	0.68390
9	0.47460	0.56551	0.63027	0.67817
10	0.45820	0.55111	0.61788	0.66784
11	0.43776	0.53154	0.60001	0.65204
12	0.41376	0.50726	0.57671	0.63029
13	0.38732	0.47959	0.54917	0.60348
14	0.36009	0.45064	0.51979	0.57418
15	0.33332	0.42283	0.49155	0.54589
16	0.30991	0.39803	0.46678	0.52137
17	0.28902	0.37709	0.44652	0.50189
18	0.27122	0.35999	0.43073	0.48741
19	0.25614	0.34607	0.41851	0.47687
20	0.24309	0.33427	0.40849	0.46864

Using these sequences, we compute

$$\hat{S}(N,N) = 1$$

$$\hat{S}(N-1,N) = \hat{A}(N-1,N)$$

and, for $j = 2, \cdots, N$

$$\hat{S}(N-j,N) = \hat{A}(N-j,N) + \hat{B}(N-j,N) + \sum_{i=N-j+1}^{N-1} \hat{B}(N-j,i)\hat{S}(i,N). \qquad (5.11)$$

The function $S(x,y)$ is approximated by $\hat{S}(0,N)$. In Table 5.2 we present the values of $\hat{S}(N-j,N)$, for the geometrical parameters of Table 5.1, with $a = 2[m]$, $b = 3.5[m]$ and

several values of λ. Also here $\delta = 1[m]$. The value of q is $-ln(0.8)$. This corresponds to the situation in which one shooting trial takes as long as the target travels $1[m]$, and the probability of destroying the target in one trial is 0.8.

6. REFERENCES.

Chernoff, H. and Daly, J.F. (1957) The Distribution of Shadows, *J. Of Math. and Mechanics* **6**: 567-584.

Yadin, M. and Zacks, S. (1982) Random Coverage of A Circle with Applications to a Shadowing Problem, *J. App. Prob.*, **19**: 562-577.

Yadin, M. and Zacks, S. (1984) The Distribution Of The Random Lighted Portion Of A Curve In A Plane Shadowed By A Poisson Random Field Of Obstacles, *Statistical Signal Processing*, E.J. Wegman and J.G. Smith, Ed. Marcell Dekker, New York.

Yadin, M. and Zacks, S. (1985) The Visibility of Stationary and Moving Targets In The Plane Subject To A Poisson Field of Shadowing Elements, *J. App. Prob.*, **22**: 776-786.

Yadin, M. and Zacks, S. (1988) Visibility Probabilities on Line Segments in Three-Dimensional Spaces Subject to Random Poisson Fields of Obscuring Spheres. *Naval Res. Logist. Quarterly*, **35**: 555-569.

Yadin, M. and Zacks, S. (1986) Discretization of A Semi-Markov Shadowing Process, Tech. Report No. 2, ARO Contract DAAG29-84-K-0191, SUNY at Binghamton, NY.

Zacks, S. and Yadin, M. (1984) Approximating The Probabilities of Detecting And Hitting Targets And The Probability Distribution of The Number of Trials Along The Visible Portions Of Curves In The Plane Subject To A Poisson Shadowing Process, Tech. Report No. 3, ARO Contract DAAG29-83-K-0176, SUNY at Binghamton, NY.

Appendix: *The Functions $K_{\pm}(x,t)$ in the Standard-Uniform Case, With Uniform Distribution of radii on (a,b).*

Let $K_+(x,t,z)$ denote the area of the set bounded by the line $\mathcal{L}_x^+$, the ray $\mathbb{R}_{x+t}$, $t \geq 0$, and the lines $\mathcal{U}$ and $\mathcal{W}$; $\mathcal{L}_x^+$ is the line parallel to $\mathbb{R}_x$, on its r.h.s., of distance z from it. This is the set of all disk centers between $\mathbb{R}_x$ and $\mathbb{R}_{x+t}$, of radius $Z = z$, which do not intersect $\mathbb{R}_x$. In order to simplify notation, we assume that $w = r$. In actual computations we substitute xw/r and tw/r for x and t in the formulae given below. Let $d = (x^2 + w^2)^{1/2}$. Simple geometrical considerations yield:

$$\begin{aligned} K_+(x,t,z) = {} & I\left\{\frac{zd}{t} < u\right\}\left[\frac{w^2 - u^2}{2w}t - 2\,\frac{zd}{w+u}\right] \\ & + I\left\{u \leq \frac{zd}{t} < w\right\}\left[\frac{1}{2tw}(tw - zd)^2\right] \end{aligned} \tag{A.1}$$

where $I\{A\}$ is the indicator set function, which assumes the value 1 if A is true, and the value 0 otherwise.

Notice that $K_+(x,t,z)$ depends on x only via x^2. Symmetry implies that $K_-(-x,t,z) = K_+(x,t,z) = K_+(-x,t,z)$ for all $-\infty < x < \infty$. Hence, $K_+(x,t) = K_-(x,t)$ and we delete the $\pm$ subscript of K. Finally, $K(x,t) = E\{K(x,t,Z)\}$ with respect to the uniform distribution of Z over (a,b). Let $x_1 = tu/d$ and $x_2 = tw/d$. The function $K(x,t)$ assumes the following forms:

(i) If $b < x_1$,

$$K(x,t) = \frac{w^2-u^2}{2w}\left(t - \frac{d}{u+w}(a+b)\right). \tag{A.2}$$

(ii) If $a < x_1 < b \le x_2$

$$\begin{aligned} K(x,t) = {} & \frac{w^2-u^2}{2w}\left(t\cdot\frac{x_1-a}{b-a} - \frac{d}{u+w}\cdot\frac{1}{b-a}(x_1^2-a^2)\right) \\ & + \frac{1}{2tw}\left(t^2w^2\frac{b-x_1}{b-a} - tw\frac{d}{b-a}(b^2-x_1^2) + \frac{d^2}{3(b-a)}(b^3-x_1^3)\right). \end{aligned} \tag{A.3}$$

(iii) If $a < x_1 < x_2 \le b$

$$\begin{aligned} K(x,t) = {} & \frac{w^2-u^2}{2w}\left(t\frac{x_1-a}{b-a} - \frac{d}{u+w}\frac{1}{b-a}(x_1^2-a^2)\right) \\ & + \frac{1}{2tw}\left(t^2w^2\frac{x_2-x_1}{b-a} - tw\frac{d}{b-a}(x_2^2-x_1^2) + \frac{d^2}{3(b-a)}(x_2^3-x_1^3)\right). \end{aligned} \tag{A.4}$$

(iv) If $x_1 \le a < b \le x_2$,

$$K(x,t) = \frac{tw}{2} - \frac{d}{2}(a+b) + \frac{d^2(a^2+ab+b^2)}{6tw}. \tag{A.5}$$

(v) If $x_1 \le a < x_2 \le b$

$$K(x,t) = \frac{tw}{2}\frac{x_2-a}{b-a} - \frac{d}{2(b-a)}(x_2^2-a^2) + \frac{d^2}{6tw(b-a)}(x_2^3-a^3). \tag{A.6}$$

(vi) If $x_2 < a$

$$K(x,t) = 0. \tag{A.7}$$

IMSE WORKING PAPER 90-133

Graphical Methods for Experiment Design
A Tutorial

Russell R. Barton

The Pennsylvania State University

Industrial and Management Systems Engineering (IMSE)
Working Paper Series

The editors of these Proceedings would like to thank Dr. Russell R. Barton for his permission to reproduce the notes he prepared for his two day tutorial delivered just before the start of the Thirty-Sixth Conference on the Design of Experiments.

Preceding Page Blank

GRAPHICAL METHODS FOR EXPERIMENT DESIGN

A TUTORIAL

Russell R. Barton

Department of Industrial and Management Systems Engineering, Penn State University

COURSE ABSTRACT

This tutorial presents several methods that are useful for creating experiment designs. Five steps in experiment design are outlined, and graphical tools are shown for three of these steps. These techniques can be applied in any experimental situation, but experimental runs of simulation models are discussed frequently as examples. Exercises are included, some of which depend on design problems that the student brings to the class.

copyright May 1990 Russell R. Barton

All rights reserved

GRAPHICAL METHODS FOR EXPERIMENT DESIGN
A TUTORIAL
for
U.S. Army Design of Experiments Conference
University of Delaware
by

Russell R. Barton
Department of Industrial and Management Systems Engineering, Penn State University

COURSE OUTLINE

I. INTRODUCTION
- A. The Role of Experiment Design in Statistical Methodology
- B. Main Steps in the Design of an Experiment
- C. Overview of Tutorial
- D. Introduction of "Main Example"

II. GRAPHICAL METHODS FOR IDENTIFYING AND CLASSIFYING VARIABLES
- A. Andrews Diagrams
- B. IDEF Diagrams
- C. Ishikawa Diagrams
- D. Discussion of Main Example

III. MULTIDIMENSIONAL POINT PLOTS
- A. Review of Design Terminology: confounding, balance, factorial, mixture
- B. Examples of Graphical Design: distance, projections, distribution, decomposition
- C. Construction for Main Example: constraints, goals, construction strategy
- D. Displaying Results on Graphical Designs

IV. EXERCISE REVIEW
- A. IDEF and Ishikawa Presentations and Discussion
- B. Fractional Factorial Presentations and Discussion

V. VALIDATION
- A. Importance of the Model: examples
- B. The General Linear Model and the Design Matrix
- C. Measures of Design Goodness
- D. Checking X^TX and X^TX^{-1}: examples with SAS
- E. Graphical Views of X^TX and X^TX^{-1}

VI. SUMMARY

VII. FURTHER TOPICS
- A. Graphical Design for Many Factors, Many Levels
- B. Mixture Experiments: models and graphs
- C. Incomplete Block Designs
- D. Network Representations
- E. Nomograms

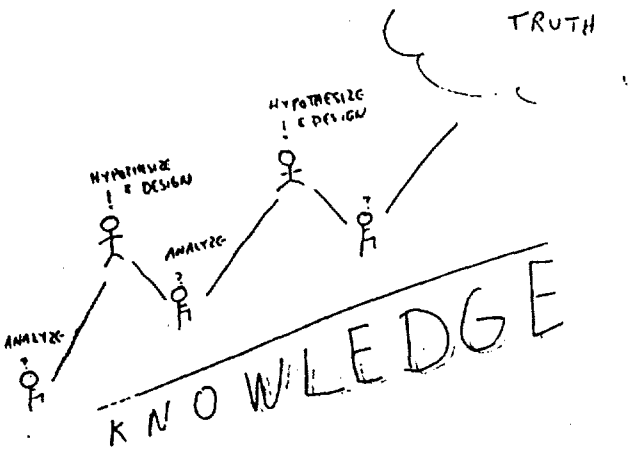
VIII. REFERENCES

GRAPHICAL METHODS FOR EXPERIMENT DESIGN
A TUTORIAL

TABLE OF CONTENTS

PAGE	SECTION	
1	I.	INTRODUCTION
12	II.	GRAPHICAL METHODS FOR IDENTIFYING AND CLASSIFYING VARIABLES
20	III.	MULTIDIMENSIONAL POINT PLOTS
45	IV.	EXERCISE REVIEW
46	V.	VALIDATION
54	VI.	SUMMARY
56	VII.	FURTHER TOPICS
64	VIII.	REFERENCES

copyright May 1990 Russell R. Barton



I. INTRODUCTION

A. The Role of Experiment Design in Statistical Methodology

Before we begin to study graphical methods for experiment design, let's review the role of this activity in the overall process of scientific investigation. How does experiment design fit in? This figure gives Horace Andrews' view of the pursuit of knowledge as a repeating cycle of forming hypotheses, *designing experiments*, collecting data, analyzing the data, leading to new or revised hypotheses. Think of the steepness of this slope as the rate of knowledge increase. (Source: Andrews, 1964).

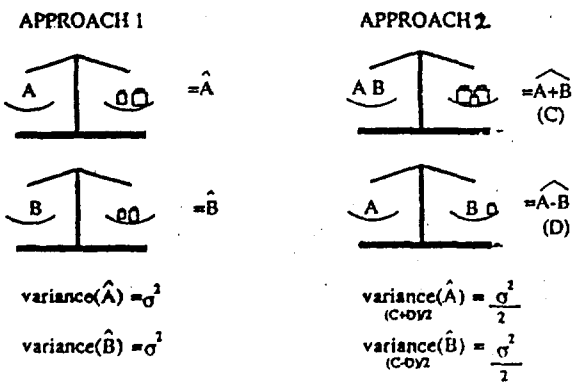
The experiment design plays a key role in determining the slope of the line. A well-designed experiment will provide maximum information for a given level of effort. The amount of information provided by the experiment or data collection process can be measured in several ways. Our concern as statisticians will focus on three measures:

Variance
Confounding
Bias

We'll discuss these concepts in some detail on the following pages. For now, we'll say that **variance** measures the precision of our information, **confounding** the ability to make assertions about one hypothesis independent of the verity of another, and **bias** the degree to which our estimates measure the things we think they measure.

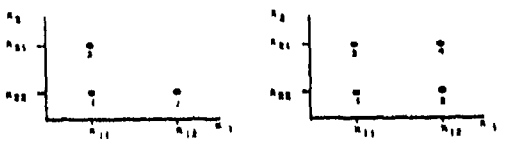
A SIMPLE STOCHASTIC EXAMPLE:

WEIGH TWO ITEMS (A and B)
SCALE HAS ERROR WITH variance = σ^2



Let's illustrate the importance of experiment design for controlling variance. Most of our data collection efforts will have to deal with random perturbations in the values. Reduced variance of our estimates means greater precision: greater information for the same amount of work. Let's weigh two items, A and B with a scale whose output contains a random component with mean zero and variance σ^2. The first picture shows one design to estimate the weight of A and B in two weighing sessions. The resulting variance of the weights is σ^2. A less obvious weighing design is shown as Approach 2. This design also requires only two weighing sessions, yet the variance of the estimates has been reduced by a factor of two!

EXAMPLE.



$$y = a_0 + a_1 x_1 + a_2 x_2$$

$\hat{a}_1 = (y_2 - y_1) / (x_{12} - x_{11}) = a_1$ $\qquad$ $\hat{a}_1 = ((y_2 + y_4) - (y_1 + y_3)) / 2(x_{12} - x_{11}) = a_1$

$\hat{a}_2 = (y_3 - y_1) / (x_{22} - x_{21}) = a_2$ $\qquad$ $\hat{a}_2 = ((y_3 + y_4) - (y_1 + y_2)) / 2(x_{22} - x_{21}) = a_2$

BUT, suppose model was:

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2$$

THEN

$\hat{a}_1 = a_1 + a_3 x_{21}$ $\qquad$ $\hat{a}_1 = a_1$

In this example, we'll see how experiment design can be used to reduce bias. Consider a set of experiments to estimate the first order polynomial coefficients for a Monte-Carlo simulation model output, y, which is a function of two parameters, x_1 and x_2.

The first design requires the minimum number of runs. The estimated values for the coefficients are random variables, since the simulation model output has random variation. What is their expected value? If the simulation response truly has no nonlinear component, then the expected value (i.e. long run average value) of the estimated parameters will be the true values.

However, if the true response has some nonlinear component, as illustrated at left, the first design will provide *biased* estimates for the linear coefficients. That is, the expected value of the coefficient estimates will not be the true values. They will be offset by an amount that depends on the nonlinear terms.

The second design requires an additional run, but the resulting estimates are not biased by the presence of the interaction term if it is present. Planning designs to minimize bias may conflict with planning a design to minimize variance. We will return to this issue later in our discussions.

"WE EMPHASIZE THAT THE SELECTION OF THE MATRIX OF EXPERIMENTAL POINTS REPRESENTS ONLY THE PROVERBIAL TIP OF THE ICEBERG. THEREFORE, WE STRESS SUCH MATTERS AS THE NEED FOR CLEARLY DEFINING THE GOAL OF THE TEST PROGRAM, ENUMERATING ALL POSSIBLE VARIABLES, AND HOW TO HANDLE THEM."

-Hahn

"WHAT IS THE OBJECTIVE OF THIS INVESTIGATION ?"

-J.S. Hunter

B. Main Steps in the Design of an Experiment

This concludes our brief motivation for the importance of experiment design in the scientific process. Now we need to review the steps that must be taken to produce an effective experiment design.

Of course, our whole design strategy depends on what we hope to learn from our efforts. The quotes at the left summarize these issues, in the words of two well-known applied statisticians.

Our specific interest is in computer simulation models. The table on the next page is a classification of common goals that engineers and scientists use simulation models to achieve. The goals are arranged in order of increasing computational burden. The last three experimental goals apply in design and policymaking settings.

GOAL	DESCRIPTION	MATHEMATICAL REPRESENTATION
1	nominal analysis, what-if comparisons	$F(x)$, $F(x)$ vs $F(y)$
2	sensitivity analysis, "insight", screening	$F(x) \cong F(x_0)+F'(x_0)(x-x_0)+ \ldots$
3	optimization (constrained or unconstrained)	$\min g(F(x))$, s.t. $x \in S$, $F(x) \in T$
4	tolerance analysis	$\mathrm{Prob}(F(x) \in T)$ and related statistics
5	tolerance design	$\max \mathrm{Prob}(F(x) \in T)$, s.t. $E(x) \in S$

Goal 1: Nominal analysis is used to test a theory or to validate a design. Often there is interest in comparing a small number of alternatives. For example, a pair of discrete event simulation runs might be used to compare the performance of two alternative digital communications protocols. Here the vector function $F(x)$ will provide various measures of communications performance (e.g. average delay, maximum delay, delay by message class, etc.). Schroer and Tseng (1987) use a DEDS simulation to perform "what-if" analyses on alternative manufacturing system parameters.

Goal 2: Sensitivity analysis serves three purposes. First, it presents a local model of the system response surface (e.g. a Taylor series or least squares polynomial approximation) that can be used to study the operating behavior of the true system. Second, it provides an opportunity to screen out unimportant factors before moving on to more detailed experiments (goals 3-5). Third, it identifies highly influential variables that require careful control for process stability or that require accurate estimation from empirical data. These roles are interrelated.

Goal 3: One might want to design a particular amplifier circuit to minimize power dissipation subject to constraints on frequency response, output power, component costs, and environmental controls. This optimization could be quite difficult if the decision vector x included the kind of circuit components used (e.g. high or normal efficiency transformers) and if the response included random perturbations. It would then constitute a discrete factor stochastic optimization problem with implicitly defined constraints (and objective function). The satisfactory general solution of such problems is beyond the scope of current methodology. On the other hand, cases where all variables are continuous and deterministic have been addressed in a number of studies (e.g. Freeman, et. al., 1988).

Goal 4: Tolerance analysis identifies the multivariate distribution of the performance vector, $F(x)$, that will occur for a particular (multivariate) distribution of the control parameters represented by x. For example, if the holes in an electron gun grid have x,y location errors with a multivariate normal distribution $N\left(\binom{0}{0}, \sigma^2 I\right)$, what will be the distribution of the spot size vertical and horizontal diameter (D_V, D_H)?

Goal 5: Tolerance design involves trade-offs between incompatible objectives. The 'optimal' design in terms of circuit performance may result in a design that is difficult to manufacture. Component value variations can easily move the operating point outside some set of constraints, resulting in low yield. A more manufacturable design might be the point would allow for some variation in component values and still maintain acceptable (if not optimal) performance. These design problems are beyond the scope of this course.

EXPERIMENTAL DESIGNS

CLASSIFICATION:

PILOT / SCREENING / EXPLANATORY / CONFIRMATORY

SEQUENTIAL / SIMULATANEOUS (run specs)

CONTINUOUS / DISCRETE (factors)

Experiment design goals are classified in another dimension: the stage of the investigation; that is, how far along the path of knowledge we have progressed. At the earliest stage, a **pilot experiment** is designed. Its main purposes are to debug the experiment running and data collection procedure, and to verify the feasible ranges for the independent variables (see more below). **Screening experiments** may then be run to eliminate variables with little effect on system performance. This allows us to reduce the size of future designs to study the important effects in detail. The first detailed experiments may still be **exploratory**, in that the nature of the key variables has not yet been established. Findings from such exploratory designs should be checked with additional data from a **confirmatory experiment**.

Exercise 1: Identify the purpose for an experiment you are considering in your work. Give the purpose in words, and identify which if any goals in the previous table correspond to your interests. Determine whether you are at the pilot, screening, exploratory, or confirmatory stage.

STEPS IN EXPERIMENTAL DESIGN

2 IDENTIFY

INDEPENDENT
DEPENDENT
NUISANCE
INTERMEDIATE
} VARIABLES

The second step in the experiment design process is to identify variables or *factors* that will concern us in the course of this investigation. This includes things that we can adjust or control (independent variables), resulting system outputs or performance measures (dependent variables), things that we can't control but that we know affect the system (nuisance variables).

These three kinds of variables are usually all that statisticians identify in the design process, but there is a fourth important class: **intermediate variables.** Often a scientist or engineer will understand the effect of one factor on a dependent variable, but will not be able to control that factor directly.

For example, the strength of a composite material may depend on the average size of gas bubbles in the material, but this quantity cannot be controlled directly. Rather, bubble size is affected by mixing rates, chemical composition, curing temperatures, etc. One of the most difficult tasks in the initial stages of experiment design is to distinguish the intermediate variables from the truly independent, controllable variables.

The third step in designing an experiment is where modeling comes into play. Usually we narrow the scope of the investigation by deciding to hold some independent variables fixed at particular values. For the factors that will be varied, ranges must be established.

3 CLASSIFY INDEPENDENT VARIABLES

WHICH VARIED, WHICH FIXED
WHICH QUANTITATIVE, QUALITATIVE
WHICH EFFECTS LINEAR, NONLINEAR

We usually entertain one or more *models* of the system response as a function of the independent variables of interest. Often this will be a regression model. We must also determine which variables will be treated as *quantitative* (e.g. maximum vehicle speed, total number of troops), and which as *qualitative* (e.g. kinds of vehicle, type of weather). The quantitative factors may affect the response variable(s) in *linear* or *nonlinear* ways.

Systems with all qualitative factors are often analyzed with *Analysis of Variance* (ANOVA) models. Systems with some quantitative and some qualitative independent variables are often analyzed with *Analysis of Covariance* (ANOCOVA) models. Systems with all quantitative variables are often analyzed with *Multivariate Regression* models.

Exercise 2: Classify the variables for an experiment that you propose to conduct:

Independent Variables:

Quantitative:

-Linear: _____

_____

-Nonlinear: _____

_____

Qualitative: _____

_____

Which of the above variables will be held fixed for the proposed experiment? Which will be varied?

FIXED: at what values?

variable	value

VARIED: over what range?

variable	lower limit	upper limit

Dependent Variables:

Quantitative (preferred): _____

Qualitative (if you must): _____

Nuisance/Noise Variables: _____

_____

Intermediate Variables: _____

_____

13.1 A LIST OF CONSTRUCTION METHODS

The following methods of constructing factorial designs literature:

(i) Orthogonal arrays.
(ii) Balanced arrays.
(iii) Latin squares and orthogonal Latin squares.
(iv) Hadamard matrices.
(v) Finite geometries.
(vi) Confounding.
(vii) Group theory.
(viii) Algebraic decomposition.
(ix) Combinatorial topology.
(x) Foldover.
(xi) Collapsing of levels.
(xii) Composition (direct product and direct sum)
(xiii) Codes.
(xiv) Block designs.
(xv) F-squares.
(xvi) Weighing designs.
(xvii) Lattice designs.
(xviii) Finite graphs.
(xix) One-at-a-time.
(xx) Trial and error.

We are now ready to identify the set of conditions that will be used for our experiment. This is the fourth step in the design process, and the main focus of this report. There are many ways to choose a *design* for the experiment. The list at the left covers many of these, both graphical and non-graphical (Raktoe, Heydayat and Federer, 1981).

Furthermore, in addition to the considerations imposed by the purpose, the number and kind of variables, and the model, the design choice will also depend on whether we want to use a *simultaneous* or sequential strategy. For a simultaneous strategy we establish a number of experimental conditions *a-priori* and collect data for all conditions before beginning analysis. This has advantages where parallel runs are possible, such as in agricultural experiments, or for certain parallel computing applications of simulation models.

A sequential procedure changes the design based on the available information before each new data point is collected. The settings for the next run are not determined until the current run's results have been analyzed.

In this course we well focus on graphical methods for choosing the design points. We will devote the greatest attention to simultaneous designs, but evolutionary operation (EVOP) sequential designs will be discussed also.

4 CHOOSE A DESIGN:

COOKBOOK

MATHEMATICAL (COED, E-CHIP, ACED)

GRAPHICAL

While the list above is long, the realistic strategies available to engineers for developing an experiment design are simple: a) choose a design from a book, b) use a computer program to generate a design based on your input specifications, or c) generate your design graphically.

Once a design has been selected, its mathematical properties should be checked to verify that it will provide information which meets the goals for low variance, bias, and confounding. Fortunately and surprisingly, it is usually possible to assess the quality of the information that the design promises before we get the results! We'll discuss the mathematics of this procedure for *the General Linear Model* in a later section.

GRAPHICAL METHODS FOR
EXPERIMENT DESIGN

DESIGN STEPS:

1 PURPOSE
2 IDENTIFY VARIABLES
3 CLASSIFY VARIABLES
4 CHOOSE OR CREATE A DESIGN
5 VALIDATE THE DESIGN

These five steps of experiment design are summarized in the table at the left. Only after completing all five are we ready to carry out the experiment, collect data, and analyze the results.

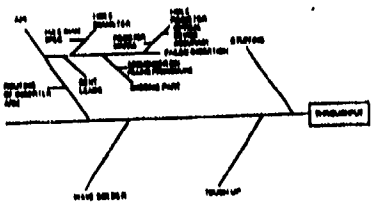
WHY GRAPHICAL METHODS?

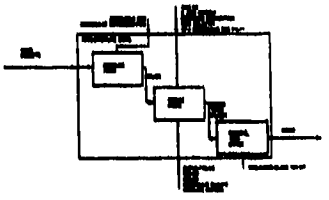
STANDARD DESIGNS WON'T WORK

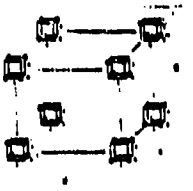
EASIER TO CREATE NEW DESIGNS WITH RIGHT BRAIN TOOLS

INTERPRETATIONS SIMPLER

CAN BE USED IN DATA ANALYSIS







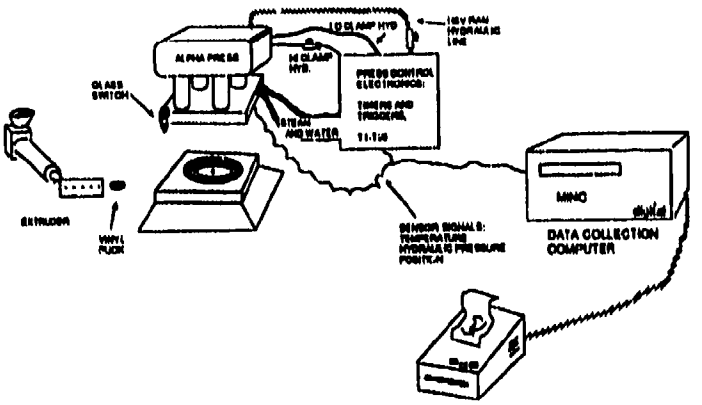
C. Overview of this Tutorial

In this course we will focus on steps two, four, and five. For each of these activities we will present graphical tools to make these tasks easier and more fun! Only step 5 will we deal in some detail with the statistical models and analyses that will eventually be performed.

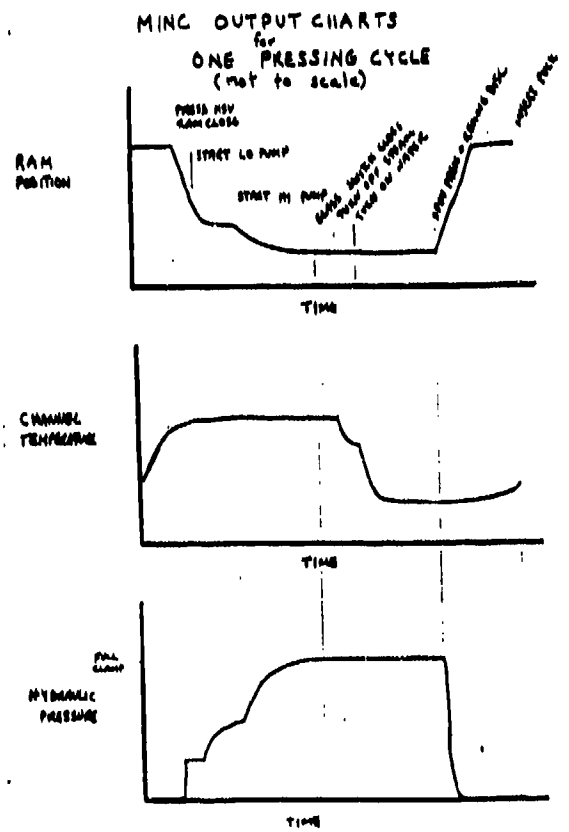
Each new technique will be illustrated with one or two examples. We introduce the main example in the next section.

There are exercises included for you to try to apply these tools to a design problem that you have brought with you.

By the end of the course you will be understanding and using graphical tools like those at the left. And remember, an experiment that is well designed will provide you with maximum information, and will usually be easier to analyze!



Alpha press for videodisc manufacture



D. Introduction of Main Example

It will be helpful to illustrate these methods by considering examples along the way. We will return often to one particular example and follow it through from beginning to end. It is based on a real study of a disk pressing operation.

MAIN EXAMPLE: Disk Pressing Operation

We want to understand a vinyl disc pressing operation, and through this understanding improve the quality of the product while reducing the manufacturing costs. The disk manufacturing operation consists of a number of steps.

First, vinyl pellets are melted and formed into a 3" puck by the extruder shown at the left. The puck is removed from the extruder by a mechanical arm and placed in a large press. The vinyl is compressed between two steam-heated molds under hydraulic pressure. When the molds nearly touch, a switch triggers the end of the steam feed to the molds and, after a short delay, cooling water is pumped though the same internal channels in the molds. The duration of this cooling cycle is controlled by a timer. After this time, the hydraulic pressure is released, the press is opened, and the disk is removed.

One problem with disks is warp. Warp is measured as peak-to-peak variations in the height of the outside edge of a disk as it is spun on a turntable. Too much warp (more than 15 mils, say) is unacceptable. We expect that warp is caused by internal stresses in the disk, which may in turn be caused by the rate at which the vinyl is compressed to form the disk and the cooling process. We also expect that thicker disks will be less susceptible to warp.

The figures at the left show typical temperature, pressure, and thickness profiles over a single pressing cycle.

PRESSING GOALS:

LOW WARP

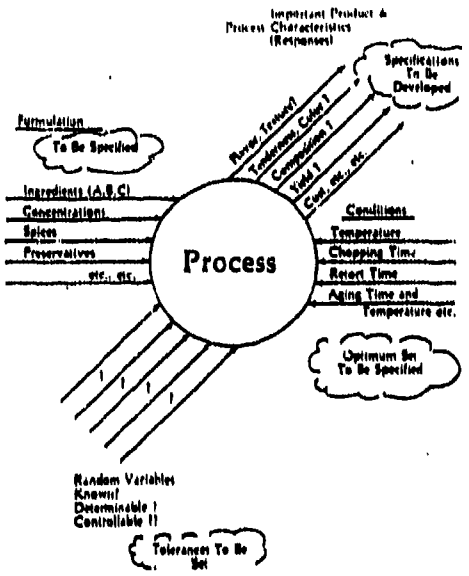
LOW DISK WEIGHT (MATERIAL COST)

LOW CYCLE TIME (HIGH THROUGHPUT)

Unfortunately, warp is not our only concern. We also must worry about disk weight. Thicker, heavier disks contain more material, and thus cost more. We also know that changing timers and pressures on the press will change the time it takes to make a single disk. Since the machines which make the disks are expensive, a long cycle time will not be acceptable. As in many experimental situations, we have multiple competing objectives. Increasing the disk thickness may decrease warp but it will increase disk weight and consequently material costs will be higher.

We will return to this example to design an experiment to learn how the pressing parameters affect warp, disk weight, and cycle time. We will want to predict warp, weight, and cycle time as polynomial functions of the independent variables. We will need to identify the key variables, choose models, choose a design, and validate the design.

(We can think of our experiments as carried out on an actual disk press or by running a complex computer simulation of the disk pressing operation.)



II. GRAPHICAL METHODS FOR IDENTIFYING AND CLASSIFYING VARIABLES

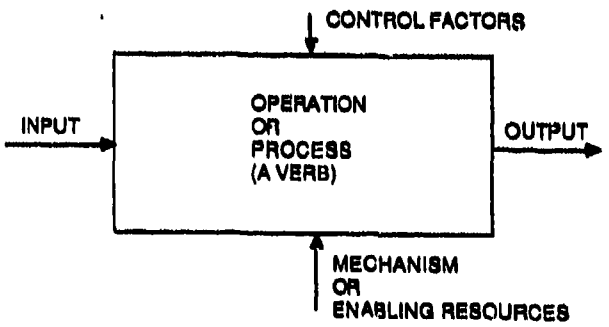
How do we identify and classify variables? Without tools, this process can be long and confusing. Exercise 2 was intended to provide a worksheet to help in this process, but it can only serve as a list *checker*, not a list *generator*.

A. Andrews Diagrams

One graphical tool for representing the factors of an experiment is the *Andrews Diagram* shown at the left (Andrews, 1964). For this example, the goal is to study the process of making SPAM! The incoming arrows represent independent and nuisance factors, the outgoing arrows represent the dependent variables of interest: flavor, texture, etc.

This graphical tool is a start, but it fails to help us in two ways:
1) to understand the distinction between independent and intermediate variables, and
2) to represent complex causal relations involving intermediate variables, where a response variable depends, through a chain of intermediate variables, on one or more independent variables.

The Andrews diagram leaves us with the uncomfortable feeling that we haven't identified ALL of the important factors for a particular experimental situation. The next two tools address these shortcomings.

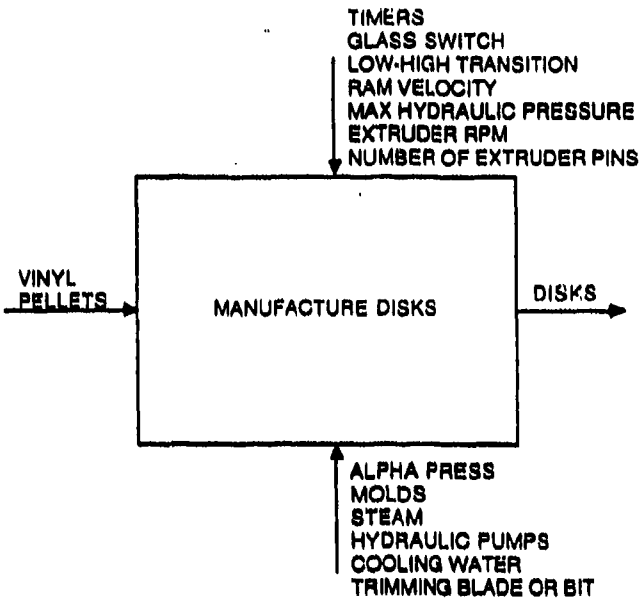


IDEF DIAGRAM: Basic Structure

B. IDEF Diagrams

These causal diagrams were first presented by Douglas Ross in 1977. These diagrams can be used to describe the operation of any complex system. They will help us to identify all the important factors in the system we are trying to model, and will enable us to separate independent, nuisance, intermediate, and dependent variables easily.

IDEF diagrams (also known as SADT) are block diagrams with a formal process for drawing and labeling incoming and outgoing arrows. Arrows indicate input quantities from the left, control parameters from the top, enabling resources from the bottom (mechanism) and outputs to the right. The action taking place is described in the box. Thus the arrows are labeled with nouns (factors) and the box with verbs (model).

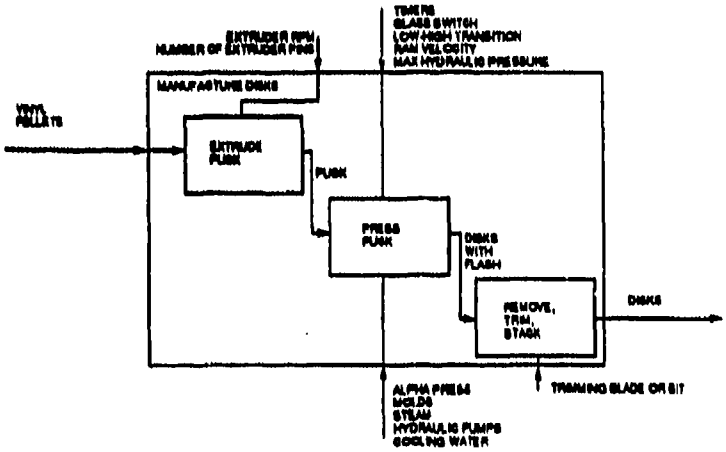


These figures illustrate the use of IDEF diagrams for the disk pressing example. Notice that

a) the diagrams are hierarchical: it is possible to break down the activities in a high level box to lower level boxes. Outputs from one lower level box which are inputs to another lower level box are *intermediate* variables.

b) a pattern of boxes from upper right to lower left indicates a dominance/dependence relation.

The lower figure shows a more detailed breakdown of the pressing activities. This allows one to identify dependency relationships among variables, and to uncover variables previously overlooked. For example, we must now wonder whether there really are no control factors for the disk trimming operation.

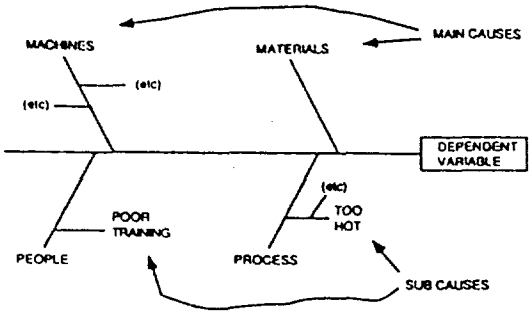


Below are some general tips on using these diagrams.

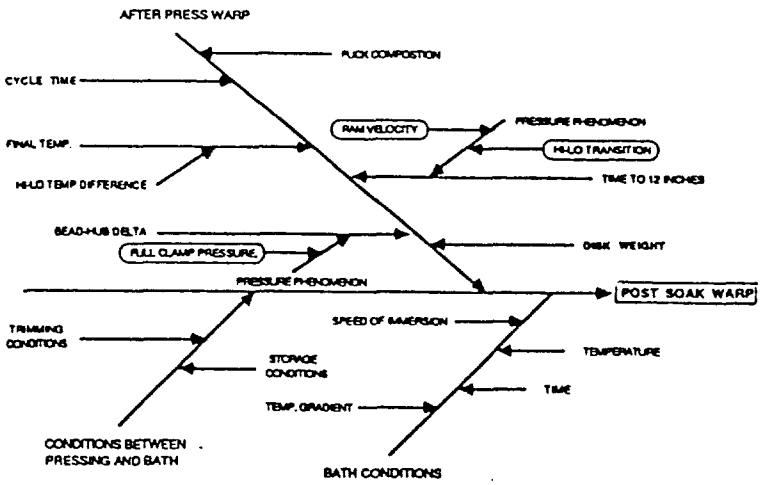
1) Break down to the bottom level of detail to identify all important variables and classify them as independent, intermediate, etc.

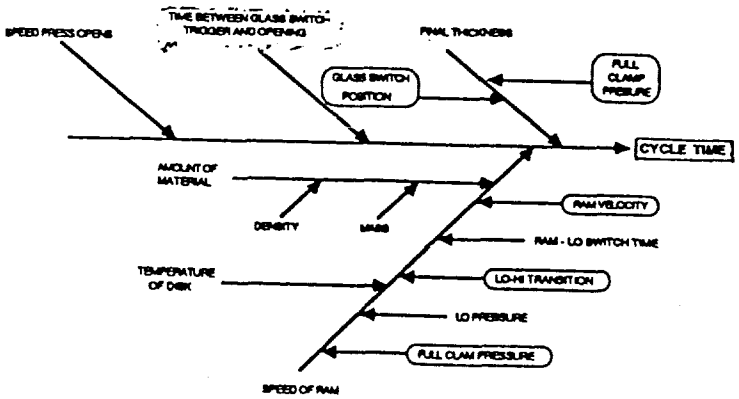
2) Control variables (top) are the 'independent' or 'nuisance' variables, if they come from outside the main box, otherwise they are 'intermediate' variables. The dependent variables exit the main box at the right. Distinguish independent variables from nuisance variables by whether they are controllable (the former) or not (nuisance). Controllable means there is a 'knob' to adjust the parameter.

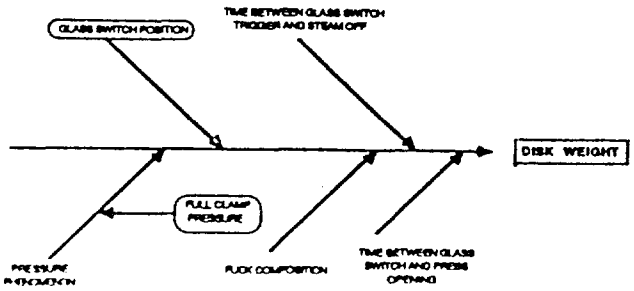
Exercise 3: Draw an IDEF diagram for a process or simulation model on which you plan experimental studies. Carry the diagram to at least two hierarchical levels of boxes, three levels if necessary. Create a list of independent, intermediate, nuisance, and dependent variables from the diagram.



ISHIKAWA FISHBONE DIAGRAM: Basic Structure







C. Ishikawa Fishbone Diagrams

The relation between independent, intermediate, and dependent variables is made clearer in an IDEF diagram. Another graphical tool that makes this relationship clear is the Ishikawa (1982) Fishbone Diagram.

These diagrams must be created separately for each dependent variable. That variable is named in a box at the right edge of the paper. A horizontal line is extended to the left, and diagonal lines representing 'causes' of the dependent variable are attached. Ishikawa suggests four main cause lines: people, machines, material, and process. Of course you are free to choose other main causes.

The diagram at the left illustrates a fishbone diagram for the dependent variable POST-SOAK WARP for the disk pressing system. The diagrams below shows a similar fishbones for cycle time and disk weight. The latter two both occur as branches of the warp fishbone as well (but are not elaborated there). This fishbone interaction is common for multi-objective studies.

As with the IDEF diagrams, it is possible to read off the independent, dependent, etc. variables from the diagram.

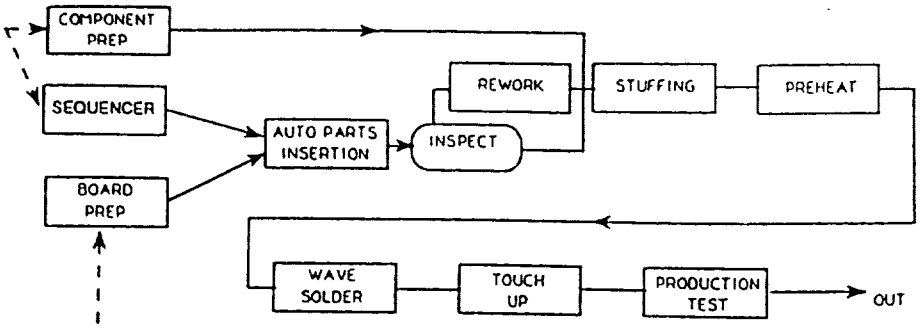
1) Independent and nuisance parameters are labeled causes or subcauses which themselves have no further subcauses (impinging lines). If you think of the diagram as a tree, these are like the leaves.

2) Intermediate and dependent variables are causes or subcauses which are in turn caused by other things. They are branches with twigs or leaves attached. There is at least one dependent variable, the box at the right of the diagram.

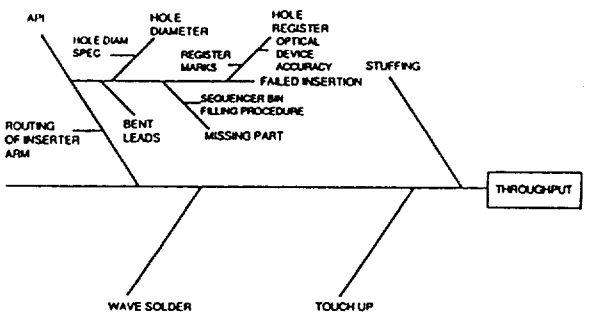
3) Independent and nuisance variables are distinguished by whether a 'knob' exists. Intermediate and dependent variables are distinguished by whether or not you are <u>directly</u> concerned about the parameter value (dependent), or only about its consequential effect on other measures of system performance (intermediate).

The circled parameters indicate the set that has been chosen to vary in a hypothetical experiment. These figures are from Young, et. al. (1987).

Exercise 4: Draw a fishbone diagram for one dependent variable for a process or simulation model on which you plan experimental studies. Carry the diagram to at least two hierarchical levels of subcauses, more if necessary. Create a list of independent, intermediate, nuisance, and dependent variables from the diagram.



The block diagram (not IDEF!) above illustrates the key steps in a printed circuit board manufacturing operation. Assume we have built a discrete event computer simulation model of this operation which we will use to optimize the throughput.



ISHIKAWA FISHBONE DIAGRAM: Circuit Board Manufacturing

This diagram shows a partially completed fishbone diagram for the printed circuit board process. The main causes have been chosen to correspond to processes. The variables that have been identified so far are:

Independent:
hole diameter specification
sequencer bin filling procedure
insertion arm routing instructions
('register marks' needs more specification before we have the appropriate 'knobs')

Dependent:
throughput

Intermediate:
% failed insertions
# missing part occurrences
hole diameter

Nuisance:
optical device accuracy
bent leads

D. Discussion of Main Example

The IDEF and Fishbone diagrams have helped us to perform Step 2 in the experiment design process. For our disk pressing example, we have identified the variables shown at the left. This list has been shortened and simplified for this course presentation.

VARIABLES FOR THE DISK PRESSING EXAMPLE:

INDEPENDENT VARIABLES:

- Timers T-6, T-12, T-13, T-22
- Full Clamp Pressure
- Lo-Hi Transition Pressure Threshold
- Ram Velocity
- Glass Switch Setting
- Vinyl Composition
- Soak Bath Temperature
- Soak Bath Time

INTERMEDIATE VARIABLES (some of them):

- Time to Reach 12" Diameter
- Bead-Hub Delta
- Disk Temperature at Release From Press

NUISANCE VARIABLES (some of them):

- Room Temperature
- Maximum Steam Pressure

DEPENDENT VARIABLES:

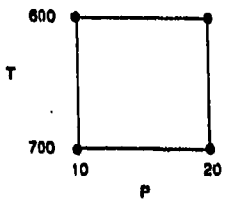
- Post-Soak Warp
- Cycle Time
- Disk Weight

Step 2 in experiment design involves identifying the key variables. In Step 3 these variables are classified as fixed or varied, linear or nonlinear, quantitative or qualitative. All of these decisions are intimately related to the mathematical model(s) that will be fitted to the system response(s).

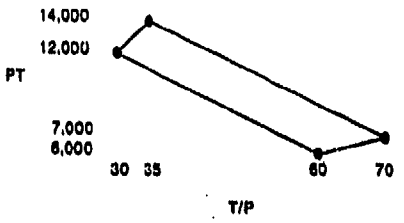
OPERATIONAL/MODELING DECISIONS:
STEP 3 IN EXPERIMENT DESIGN

- SELECT A SET OF DEPENDENT VARIABLES FOR STUDY
- CLASSIFY THE INDEPENDENT VARIABLES:

WHICH VARIED IN EXPERIMENT (SPECIFY RANGE)	WHICH FIXED (AT WHAT VALUE)
WHICH QUANTITATIVE	WHICH QUALITATIVE
WHICH LINEAR	WHICH NONLINEAR



$y = a_0 + a_1 f_1(T) + a_2 f_2(P) + \varepsilon \Rightarrow$ BALANCED



$y = a_0 + a_1(PT) + a_2(T/P) + \varepsilon \Rightarrow$ UNBALANCED

Note that we are considering the model form BEFORE designing the experiment. The example here illustrates the importance of this step. Imagine an experiment to estimate a reaction rate that depends on temperature and pressure. Let y represent the rate and let P and T represent pressure and temperature, respectively. Here a particular design is *a full factorial* for a model of the first sort. This means that each value T run in the experiment is run in combination with each value of P run in the experiment.

If our model were based on the same two variables but in a different way, say **PT** and **T/P**, then the same design takes on a decidedly poorer appearance, and is no longer a full factorial!

Source: Satterthwaite (1959).

DESIGN GOODNESS DEPENDS ON THE MODEL

In our modeling work we will often fit polynomial response models to simulation model outputs as a function of simulation model inputs. These will take the general form:

$$y_i = \sum_j \beta_j x_{ij} + \varepsilon_i, \quad \varepsilon_i \text{ are i.i.d.} \sim N(0, \sigma^2)$$

Usually $x_0 \equiv 1$, and the x's may be different independent variables, powers of independent variables, or products of independent variables. For our temperature/pressure example, we may have $x_{i1} = T$, $x_{i2} = P$, $x_{i3} = T^2$, $x_{i4} = P^2$, $x_{i5} = TP$, and so forth. The ε_i are random variations, which are usually assumed to be independent, indentically distributed Gaussian random variables. We will return to a more detailed discussion of this 'General Linear Model' in Section V, where we also mention some alternatives for metamodels of simulation model outputs.

DEFINITIONS

Definition: An **experiment** is a set of one or more runs (of a simulation model) made to meet a particular set of objectives.

Definition: An **independent variable** is a parameter of the (simulation) system that can be explicitly adjusted by the experimenter.

Definition: A **design factor** is an independent variable that will have its value changed during the course of an experiment.

Definition: A **design frame** is a specification of
- which independent variables will be held fixed(& value)
- design factors(& ranges)
- what system outputs will be measured.

Definition: An **experiment design** is a set of specifications of design factors for an experiment, along with a single specification (vector) for the independent variables that are not design factors.

Definition: A **design matrix** depends on the model to be fitted as well as on the experimental conditions. There is a column in the matrix for each term in the model to be fitted, and a row for each (simulation model) run. Each row of the matrix holds the values of the model terms for the corresponding run.

TERMS AND TOPICS IN EXPERIMENTAL DESIGN

Nuisance variables:

Examples: random # seed, starting conditions, time

Blocking- nature of influence understood (treat like any other design factor)

Randomization- nature of influence not understood

Factorial Designs:

An experiment run for each possible combination of factor levels: if all factors at two levels, then runs correspond to vertices of an n-dimensional cube.

Fractional Factorial Designs- not all vertices:

Full Factorial
2^3 (3 factors)

Fractional Factorial
2^{3-1} (3 factors)

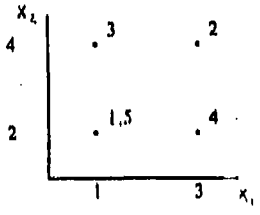
III. MULTIDIMENSIONAL POINT PLOTS

A. Review of Design Terminology

Our main focus in this section will be on graphical methods for developing designs. First we will review some basic concepts and terminology from the field of experiment design. We will use many of these terms in our discussions later with no further explanation.

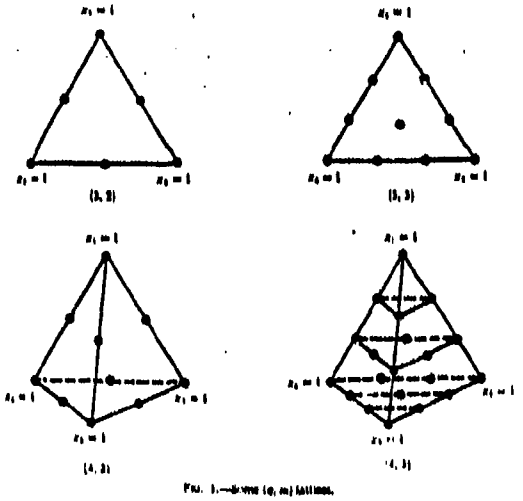
EXAMPLE:

Model: $y = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2$



$$\begin{array}{cccc} 1 & x_1 & x_2 & x_1x_2 \end{array}$$
$$\begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 3 & 4 & 12 \\ 1 & 1 & 4 & 4 \\ 1 & 3 & 2 & 6 \\ 1 & 1 & 2 & 2 \end{bmatrix}$$

This illustration of the design matrix also gives us our first graphical design. It is a 'factorial' design.



Another important class of designs are 'mixture' designs. In chemical formulation problems, one often has to have the components add to 100%. Thus arbitrary combinations are not possible. This kind of experimentation can be important for strategic models as well, where the total resource pool is fixed, and the choice is how to deploy the resources or what to purchase with them.

Source: Scheffé (1958)

TERMS AND TOPICS IN EXPERIMENTAL DESIGN

Confounding:
Design won't allow independent estimation of two or more model parameters.

Example: change two variables from previous run, and simulation output improves. Which change (or both) caused the improvement?

Resolution:
A mathematical measure of the nature of confounding for various fractional factorial designs.

Bias:
What if postulated model is wrong? What impact on parameter estimates?

Optimal Design:
Optimal in a limited mathematical sense. Definitions of optimality based on properties of the matrix X^TX, (X is the design matrix)

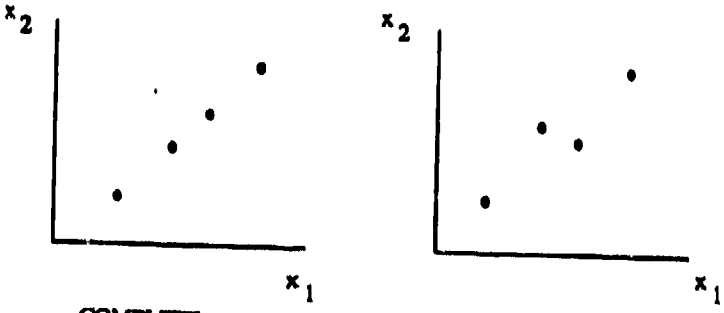
Example: D-Optimality <==> $\max_{X} \text{determinant}(X^TX)$

(for a fixed number of rows in X)

Optimality depends on the particular model.

For practical design problems, 'optimal' designs often do not make sense, because there may be considerations or constraints that cannot be described mathematically.

These figures remind you of the issues of bias and confounding. For the first pair, the design at the right guards against bias in the first order terms caused by a nonzero interaction (x_1x_2) term. The second pair illustrate full and partial confounding of two effects.



COMPLETE CONFOUNDING

PARTIAL CONFOUNDING

"Definition 13.1: A txn matrix A with entries from a set S of s symbols is called an orthogonal array of size n, t constraints, s levels, strength d, and index λ if any dxn submatrix of A contains all s^d possible dx1 column vectors based on s symbols of S with the same frequency λ."

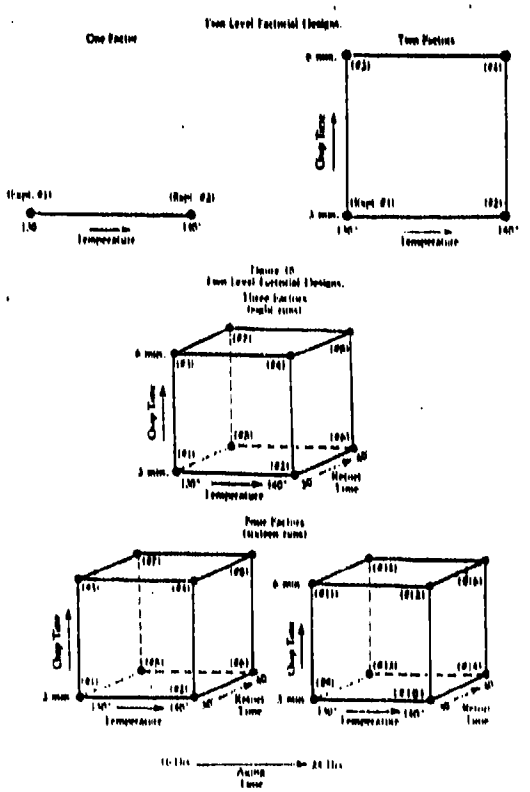
-Raktoe, Heydayat, and Federer

Recall that your main strategies for choosing a design are cookbook, software, and graphical. Each approach has advantages, and none should be used exclusive of the others. One advantage of graphical design is the clarity of the presentation. This makes it easy to communicate your study to others in your organization or to your clients. Compare the clarity of idea with the mathematical description at the left.

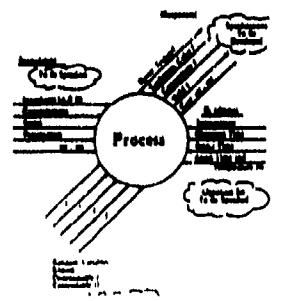
Historically, graphical methods for experiment design have not been recognized as an entity. A computerized literature search of scientific journals gave zero entries with keywords *graphical* and *experiment design* in the last ten years. Yet graphical methods have been used by outstanding statisticians to develop well known designs, including Box's central composite design. Let's look at some of these designs and see what we can learn about creating our own.

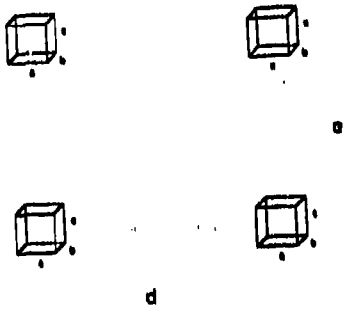
B. Examples of Graphical Design

Lets begin with the simplest and most frequently used class of designs: factorial designs. These are applicable when we have quantitative variables with simple upper and lower bounds on reasonable values.



Here we return to the Andrews (1964) SPAM study. The graphical representation here is very simple, with designs for experiments involving one, two, three, or four independent variables (factors). Note that the four-dimensional case is represented as a *pair* of cubes. We will use this approach to generate even higher dimensional designs.





A 2^5 Factorial Design

(at 4-block layout: factors is factors, blocks is blocks!)

Here we see a template for a 2^5 factorial design. For experiments with many factors, we will usually choose to run only a fraction of these points. This diagram, then can be thought of as a template for identifying conditions that will be run. At these vertices, we will draw a circle, square, triangle, or some other marker.

How do we decide which corners to choose? Of course, one could use a standard fraction, choose a defining relation, or use a computer package to do this. We will illustrate here how to do it *graphically*.

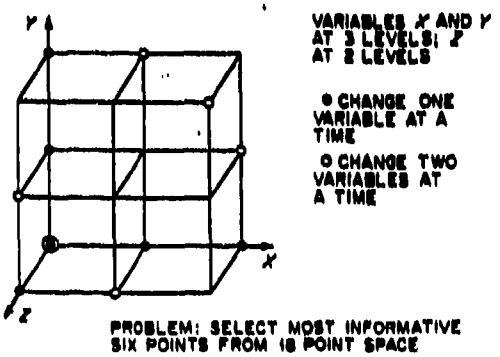


FIGURE 4—Comparison of test beds for more informative study of three variables.

Consider these two designs described in two ways by Youden (1962, 1972). The design representation at the top makes it clear that the one-at a time strategy is inferior in terms of covering the design space to the two variables at a time design. The preferred design is shown in tabular form in the lower figure. Its advantages are no longer apparent. It is interesting to note that Youden presented the clearer representation ten years after the tabular form. Do you think he created the design using the table or the graph? Which do you think would be easier?

What advantages do you see in the two-factors at a time design?

TABLE 4.

Program for three variables, two with three choices, one with two choices.

Variables

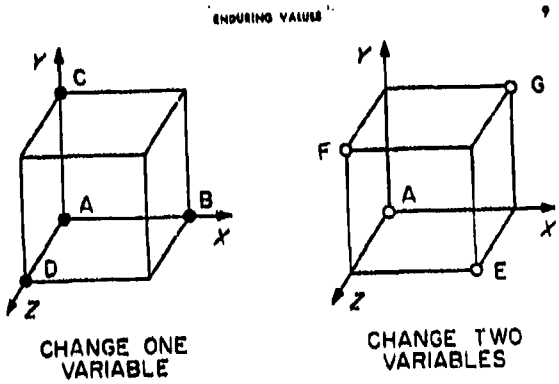
z	x	X
y	y	Y
z		Z

Six sets

1	2	3	4	5	6
z	z	z	x	X	X
y	y	y	Y	y	Y
z	Z	Z	z	z	Z

	z	x	X
y	1 2	Z −2	
y	Z 1		z −1
Y		z −1	Z 1

Above coefficients are weighing factors to estimate $z - Z$



Consider these two designs for estimating the effects of three independent variables, X, Y, and Z on a dependent variable, say W.

Source: Youden (1972)

Q: Why might you say that the design on the right is 'better' than the design on the left?

A: Lots of ways to answer. Jot down yours:

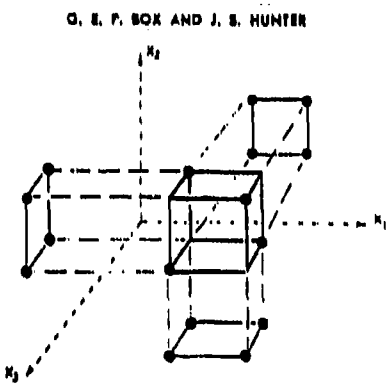


FIGURE 2—Projection of 2^{3-1}_{III} into three 2^2 factorials.

This figure illustrates that each projection is a full factorial. That means if *any* one of the three factors has insignificant effect on W, we will have gotten a full factorial design on the other two factors for free! Box and Meyer (1986) call this 'exploiting effect sparsity'. That is, when we study many factors, we expect that most of them will have little effect on the dependent variable (here called W).

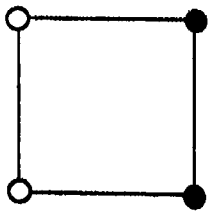
Q: How do we check that we will be able to exploit effect sparsity?

A: Look at the projections of the design. Do they yield full factorials or at least good fractions?

Q: What does a 'good fractional design' look like?

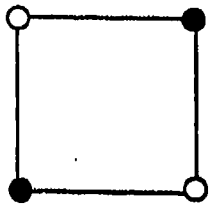
A: Check the **confounding patterns**

Source: Box and Hunter (1961)



● – ○ -

MAIN EFFECT



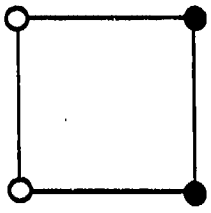
● – ○ -

TWO-FACTOR INTERACTION

These diagrams can be viewed in two ways. First, consider an experiment with two independent variables, X and Y. We will study their effect on a dependent variable W. To compute the main (linear) effect of X, we would subtract the average of the W values measured at the light circles from the average W values at the dark circles on the leftmost figure. The XY interaction effect on W would be measured by subtracting the average W's at the light circles from the average at the dark circles for the rightmost figure.

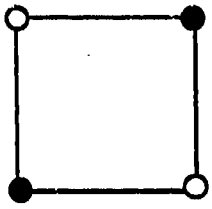
(To see this, label the low and high X and Y values as 0 and 1 respectively, substitute the values in the model a+bX+cY+dXY to get a formula for each of the four points, and then perform the subtraction)

NOW view the figure in a different light. Suppose that the DARK circles correspond to a HIGH value of a third independent factor, Z. How do we measure the main effect (linear coefficient) for Z? By subtracting the average of the light circles from the average of the dark circles. For the design at the left, the same quantity is also used to estimate the main effect for X. Thus for the design at left, we will not be able to separate the effects of X and Z. We say that the main effect of X is completely confounded with the main effect of Z.



● – ○ -

MAIN EFFECT
AND
Z



● – ○ -

TWO-FACTOR INTERACTION
AND
Z

Q: In this new light, which of these designs is to be preferred?

A: The design on the right.

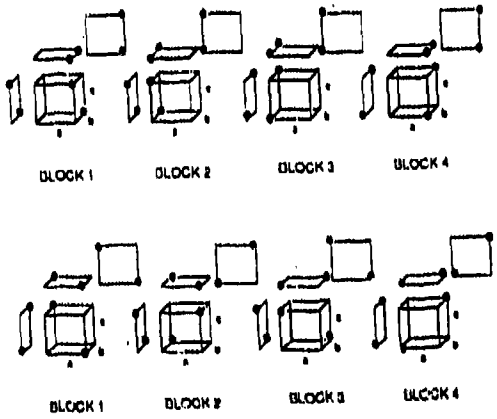
Q: Why?

A: Statisticians assume (sometimes wrongly) that second order effects (nonlinear) will be less important than first order (linear) effects. Thus the second design confounds the coefficient of Z with the XY interaction coefficient, and the XY interaction coefficient is assumed to be smaller and less important than any of the main effects for X, Y, or Z.

Block		
+ + − + − +	I	+ − − − + +
− + − − − +	II	− + − + − +
− − − − + +	III	+ + − − − +
+ − − + + +	IV	− − − + + +

Undesirable vs Better Incomplete Block Design

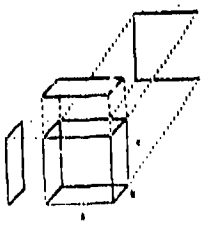
From Box, Hunter, and Hunter, pp.339-341



BLOCKED DESIGNS FROM

BOX, HUNTER & HUNTER,

(pp339-341)



We have learned several graphical lessons from these simple designs:

- check projections for goodness
- look for patterns of confounding
- make the design points 'cover' the design space
- make the design points as far from each other as possible

Let's try to apply this information to the 'Bad Block' example from Box, Hunter, and Hunter (1978).

Consider an experiment to estimate the effect of three independent factors, a, b, and c, on a dependent variable e. Furthermore, there is a nuisance variable, d, that forces us to block our experiment to try and provide equal values of d within a block. Unfortunately, the homogeneous block capacity is only two, so we can't run all 8 combinations of the other three factors in one block. How should we assign runs to blocks?

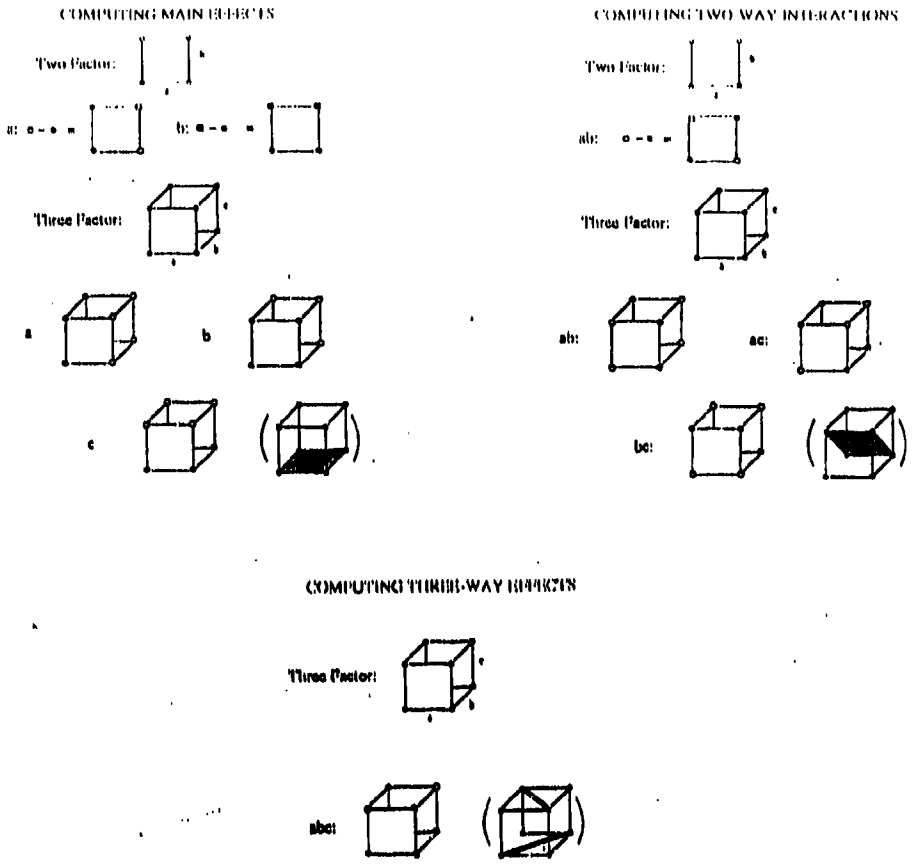
The figure at the top of the page shows two designs represented in the typical tabular fashion. The three columns on the left correspond to one design, the three columns on the right to another. For example, the first design has a high, b high, and c low for the first point in block one, and a and c high and b low in the second. Can you tell which of these two incomplete block designs is defective? Can you tell why?

The figure on the lower half of the page shows graphical representations of the same two designs, along with the three projections found by dropping either a, b, or c.

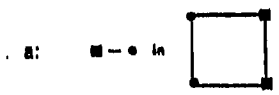
Q: What are the blocks confounded with in the top design? In the bottom?

Q: As a statistician, which design would you prefer, making the normal assumptions about main effect vs second order confounding?

Notice also that the top design violates our other design principles: the points on a cube are not as far apart as possible, nor do they cover the space as well as possible.

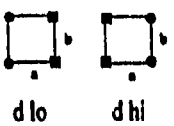


CONFOUNDING OF EFFECT
OR
PARITY OF EFFECT



Confounding of effect c with main effect a if

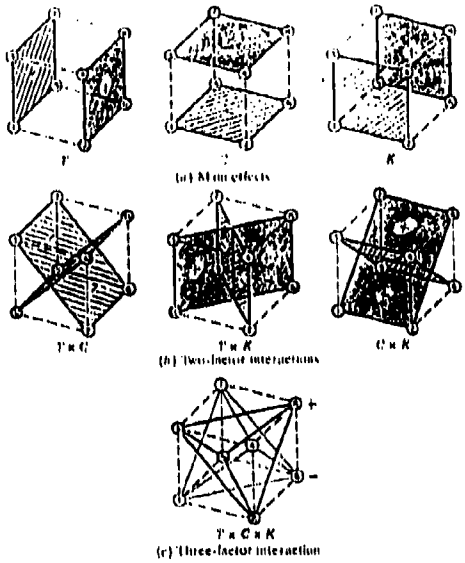
■ means c=hi
● means c=lo



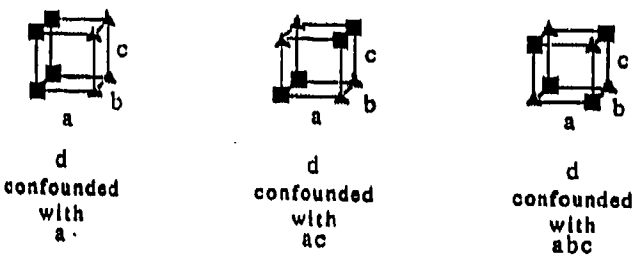
Confounding of parity of c-a confounding
with
main effect d in a 4-factor design

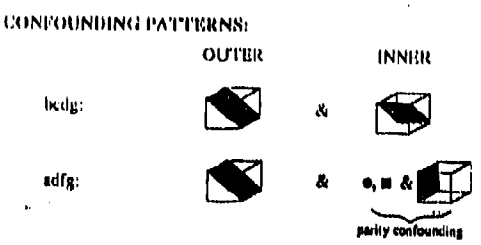
The graphical patterns of confounding are based on the graphical patterns for computing main effects and higher order interactions. There are two graphical operations that induce confounding. The first, which we'll call *confounding of effect*, superimposes the graphical pattern for one effect in the same fashion as another. In the first example at the left, the hi-lo pattern for main effect c is the same as the hi-lo pattern for main effect a or an ac confounding.

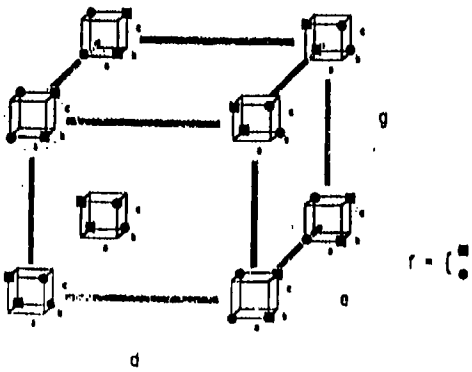
The second figure illustrates another way confounding patterns may appear graphically, which we call *confounding of parity*. Here we have four factors, with factor c coded by squares and circles. The c main effect is confounded with a again within a single level of d, but the sense of the confounding (parity) reverses between the lo and hi values for d. This gives an acd confounding term in the defining relation for the design.



We can evaluate even larger designs graphically, but first we need to look at higher order confounding patterns. The figure at the left is from Box, Hunter, and Hunter (1978) and shows the interaction patterns for 3-factor experiments. Thus if we color code a fourth factor on a three-factor plot, the color pattern will convey which effect the fourth factor is confounded with. In fact, we can carry this tool to higher dimensions by looking at some (not all) of the confounding patterns.

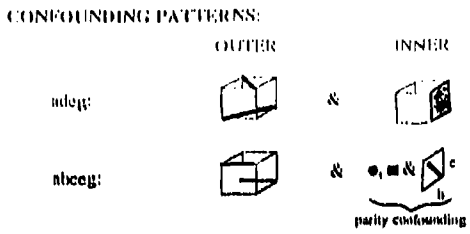


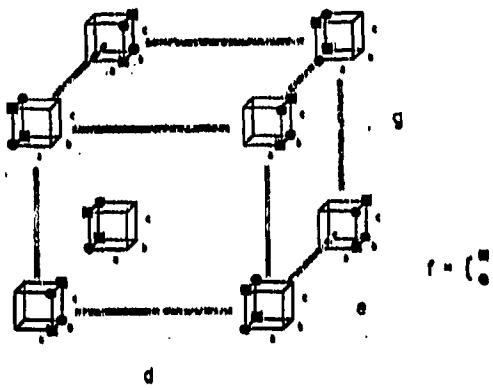




Here is a fractional factorial design from Fries and Hunter (1980). Can you pick out interaction patterns graphically?

The bc interaction pattern on the small cubes has been circled. This pattern falls on the dg interaction pattern on the large cube (the alternate dg vertices have the other bc interaction pair). Thus bc is confounded with dg, as shown in the second term of the defining relation. Similarly on these cubes, the main effect of a is confounded with the main effect of f. The opposite pattern occurs on the complementary dg interaction nodes. Thus dg is confounded with af, the third term in the defining relation. Finally, the a-f main effect confounding switches sense for opposite sets of the bc interaction (we already knew this since bc is confounded with dg, but we can see it graphically directly). Thus af is confounded with bc, the first term in the defining relation.



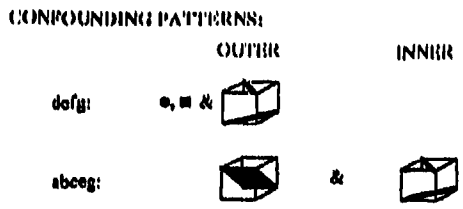


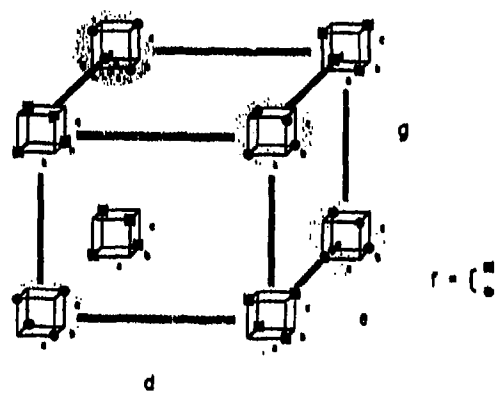
I=abcf=adeg=bcdefg

Q: What should be our goal in developing a better design graphically?

A: Lay out our points in patterns that correspond to high order interactions, not main effect patterns.

In this figure, we are able to generate a pattern that corresponds to three-way interaction on the large cube vertices, but the small cube vertices still show a main effect (a) pattern which is confounded with the large cube three way interaction (deg). Thus we have an adeg term in the defining relation.



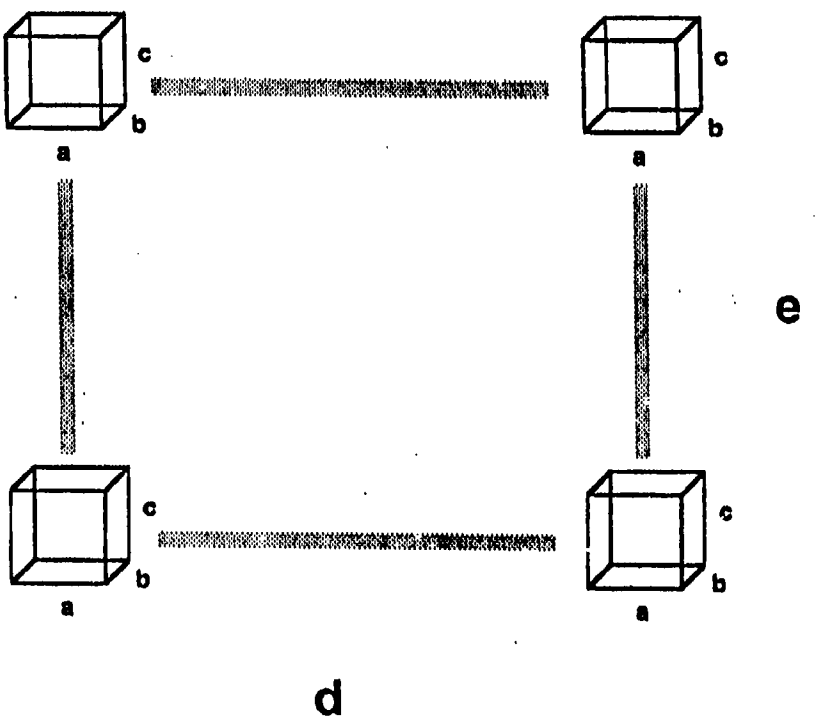


I=abcdf=abceg=defg

This third attempt is deceiving! It appears that we have done it, getting three-way interaction patterns on both the small and large cube vertices. But we have forgotten the f main effect (square vs circle). It is completely confounded with the three factor interaction (deg) on the large cube vertices. Thus we still have one four-letter word for this fractional design! At least it's repeatable: defg.

Note: these designs appeared in the original article, but the graphical representations did not. The last design was cited by the authors as the 'minimum aberration' design because it minimized the number of words in the defining relation that had minimum length.

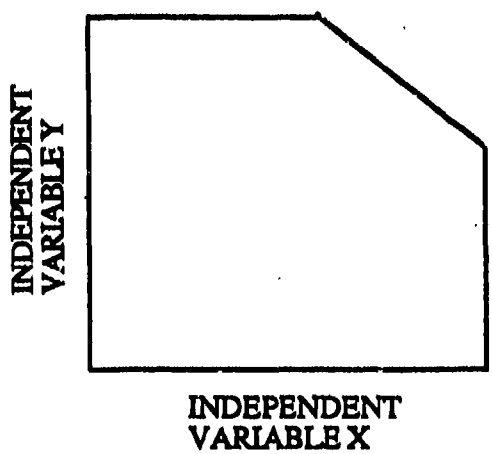
Exercise 5: Using the template below, fill in points for a 'good' 1/2 fraction design. (Hint: check projections and interaction patterns as you go).



Exercise 6: Can you 'fix' this design by moving just one point?

A	B	C	D	E
−	−	−	−	−
+	−	+	−	−
+	+	−	−	−
−	+	+	−	−
−	−	+	−	+
+	−	−	−	+
−	+	−	−	+
+	+	+	−	+
+	−	−	+	−
−	−	+	+	−
−	+	−	+	−
+	+	+	+	−
+	−	−	+	+
+	−	+	+	+
−	+	+	+	+
+	+	−	+	+

Exercise 7: You have a model for a response W as a function of two parameters, X, and Y. You expect the response to be nonlinear, and so a two-level factorial is not satisfactory. You have enough resources to run 9 experiments, so a 3^2 factorial design is possible. Unfortunately, your available design space is not rectangular (see below). Lay out your 9 runs within the design space provided below:



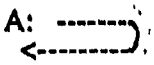
Box represents allowable range for values of X and Y

"... convenient to regard designs as built up from a number of component sets of points, each set having its points equidistant from the origin ..."

"... form the vertices of a regular polygon, polyhedron, or polytope..."

- Box and Hunter (1957)

Q: How do we choose the design points in complicated situations when the factors can take on three or more levels??

A: 

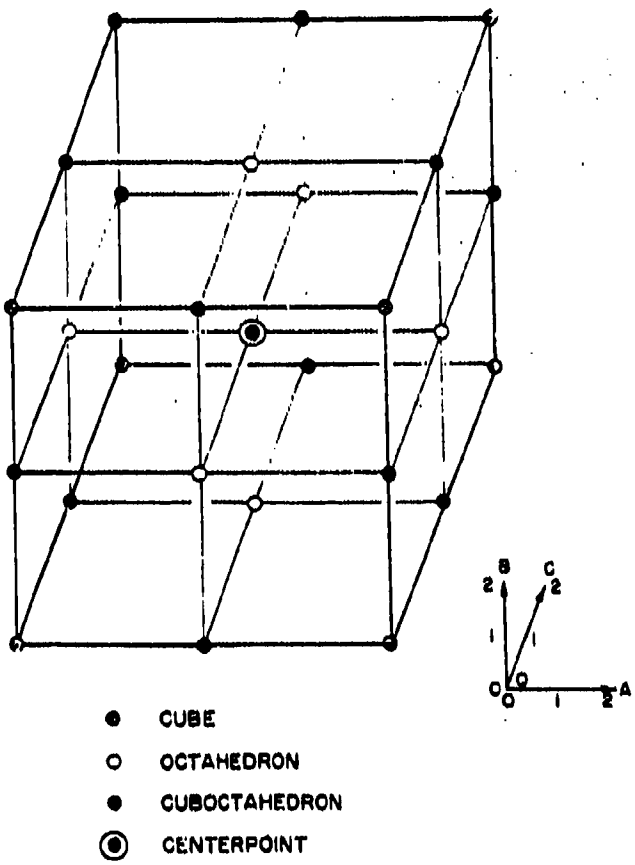


FIGURE 1

This figure shows a graphical design from the first paper in the first issue of Technometrics (DeBaun, 1959). This design is broken down into graphical subcomponents to make it easier to understand.

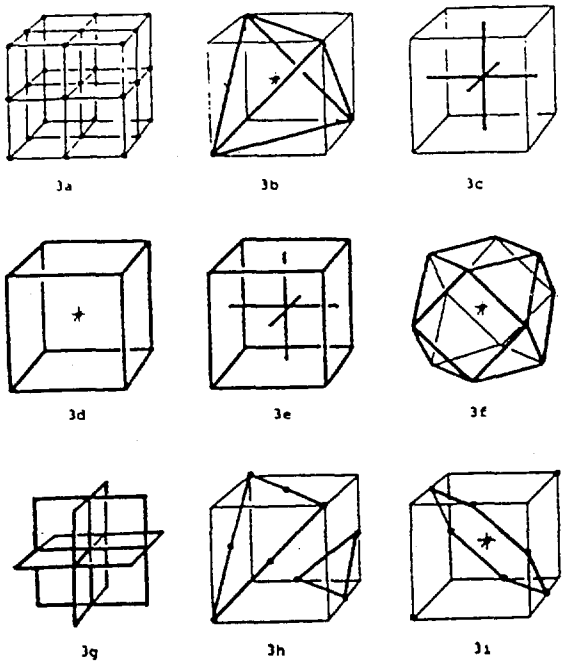


FIGURE 3. (3a) The Twenty-Seven Points Comprising the 3^3 Factorial; (3b) the 2^{3-1} Fractional Factorial, the Tetrahedron, With or Without Center Point; (3c) The Measure Polytope, Bipyramid or "Star" Design; (3d) The Cube or 2^3 Factorial, With or Without Center Point; (3e) The Central Composite Design; (3f and 3g) Two Views of the Cuboctahedron or Box-Behnken Design; (3h and 3i) Two 3 × 3 × 3 Latin Squares.

Hunter (1985) illustrates many common designs graphically as the vertices or edge points of regular polyhedra.

Let's summarize our discoveries about how to generate good designs graphically. These ideas are not new, as these quotes illustrate.

Remember, it is easier to view a complicated design as being made up of simpler graphical components.

Source: Box and Hunter (1957)

".. convenient to regard designs as built up from a number of component sets of points, each set having its points equidistant from the origin ..."

"... form the vertices of a regular polygon, polyhedron, or polytope..."

"Choose new points to MAXIMIZE the minimum distance from all existing design points.."

Also, keep design points as far apart as possible: span the design space.

Source: Kennard and Stone (1969)

"...It is proved (Appendix 1) that if a polynomial of *any* degree d_1 is fitted by the method of least squares over *any* region of interest R in the k variables, when the true function is a polynomial of any degree $d_2 > d_1$, then the bias averaged over R is minimized for all values of the coefficients of the neglected terms, by making the moments of order d_1+d_2 and less of the design points equal to the corresponding moments of a uniform distribution over R."

Unfortunately, the Box and Draper goal is in conflict with measures taken to minimize bias. To minimize bias, the distribution of points in space should appear to come from a uniform distribution. A uniform distribution does not put points at the extreme locations, but spreads them evenly over the design space. Thus there will be a tension between the second and third goals when you design experiments.

Source: Box and Draper (1959)

SOME USEFUL CONCEPTS
for generating
GOOD DESIGNS
from
MULTIDIMENSIONAL POINT PLOTS

- COVER THE DESIGN SPACE UNIFORMLY
- CHECK PROJECTIONS TO PLANES AND LINES
- SPAN THE WHOLE DESIGN SPACE: MAKE ADDED DESIGN POINTS FAR FROM EXISTING POINTS TO MINIMIZE VARIANCE FOR FIRST ORDER EFFECTS
- DECOMPOSE COMPLICATED DESIGNS INTO GRAPHICAL SUBCOMPONENTS

This table presents our findings on generating graphical designs. Let's apply these tools to a real design problem.

We return to our discussion of the disk pressing problem. Suppose we have narrowed the list of factors to vary to the list shown at the left. Upper and lower limits are specified for each independent variable, and whether they appear to have linear or quadratic effect on disk warp values.

REDUCED LIST OF INDEPENDENT VARIABLES
FOR THE
DISK PRESSING STUDY

VARIABLE	LOWER LIM	UPPER LIM	NONLIN?
GLASS SWITCH	4.900"	4.925"	LINEAR
RAM VELOCITY	LOW(-1)	HIGH(+1)	LINEAR
COOLING TIME (T-12)	12 SEC.	15 SEC.	LINEAR
LOW-HIGH TRANSITION	200PSI	600PSI	NONLIN
FULL CLAMP PRESSURE	1500PSI	2000PSI	NONLIN

Before we design the main 'exploratory' experiment, we'll try to design a small **'pilot'** experiment. The purpose of the pilot experiment is to verify that the proposed ranges for the independent variables are feasible (we don't break the press) and to debug the experiment running and data collection process. It is analogous to **kicking the tires** and checking the brakes before beginning a long trip.

PILOT EXPERIMENTS

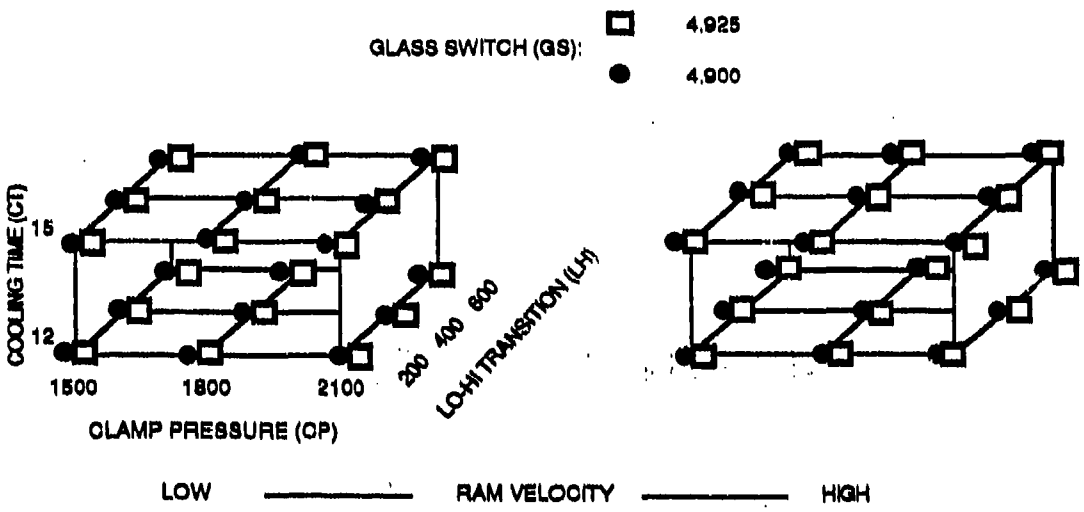
WHAT ARE THEY?

WHAT ARE THEY FOR?

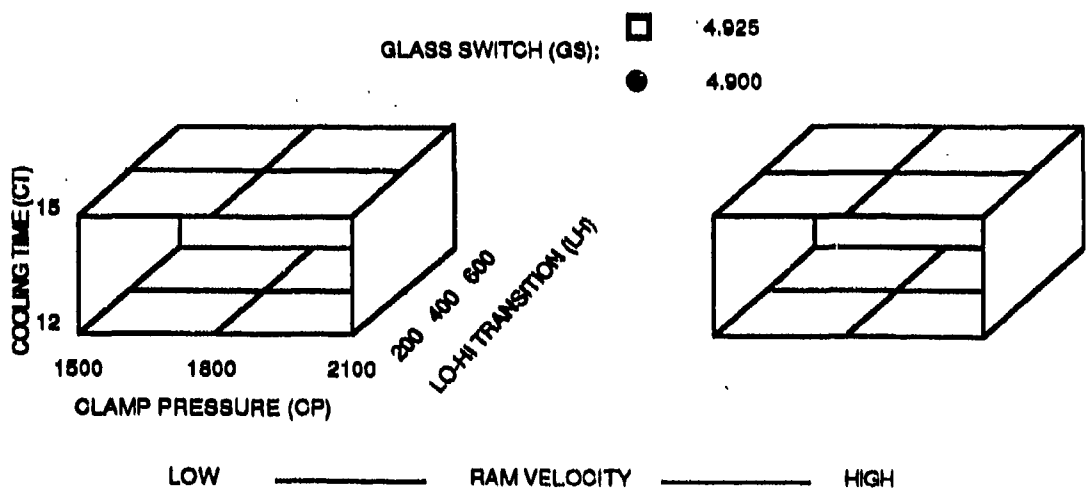
HOW IMPORTANT ARE THEY?

Exercise 8: You have a total of six runs available for a pilot experiment. Identify the six runs you choose to make *graphically* and state why you chose them. Below is a filled-out full-factorial design requiring 72 runs, followed by a template for you to use to place your design points.
(Spend 5 minutes individually, then 5 minutes in groups)

FULL FACTORIAL DESIGN



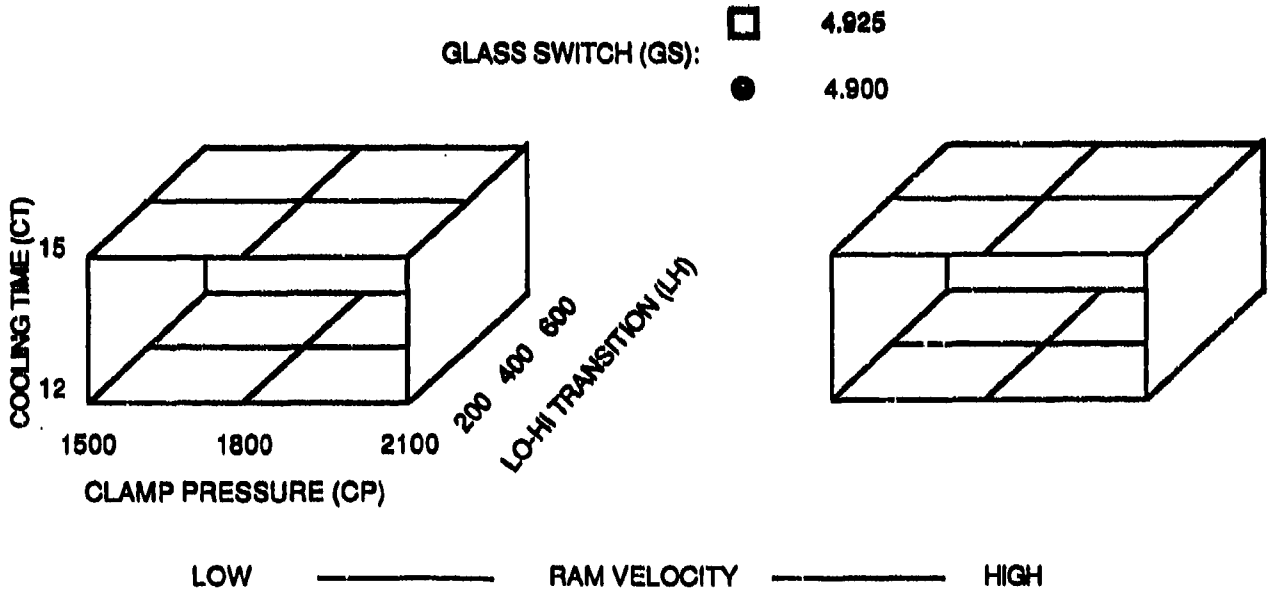
DESIGN TEMPLATE



Exercise 9: You have a total of twenty-one runs available for an exploratory experiment. You will want to fit a model that is linear in three terms (glass switch, ram velocity, and cooling time) and quadratic in two variables (low-high transition pressure and full clamp pressure). Identify the twenty-one runs you choose to make *graphically* and state why you chose them.

(Spend 10 minutes individually, then 10 minutes in groups)

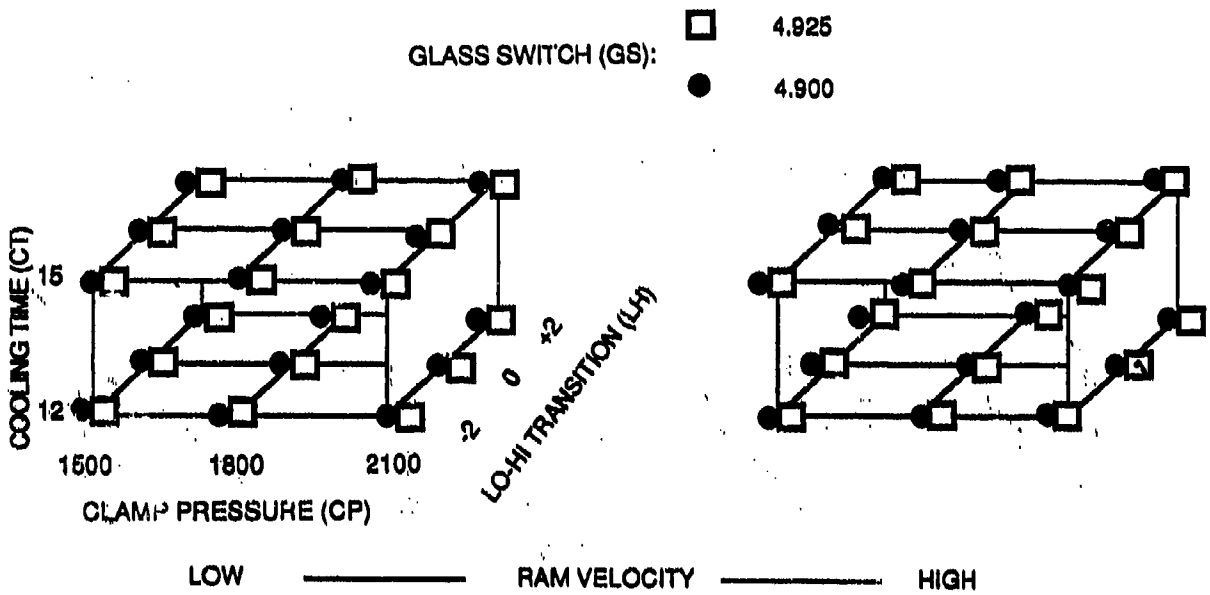
DESIGN TEMPLATE



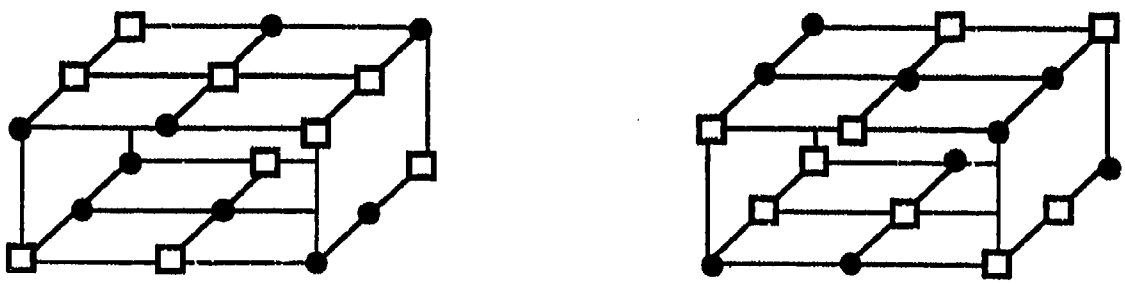
C. Constructing a Design for the Main Example

This figure shows the full 72-run factorial design and a 36-run 1/2 fraction. How was the 1/2 fraction chosen? The breakdown into graphical subcomponents is made clear on the following page.

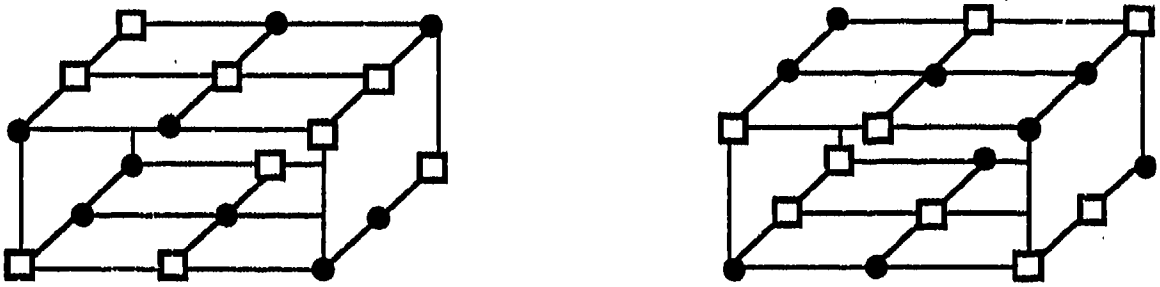
FULL FACTORIAL DESIGN



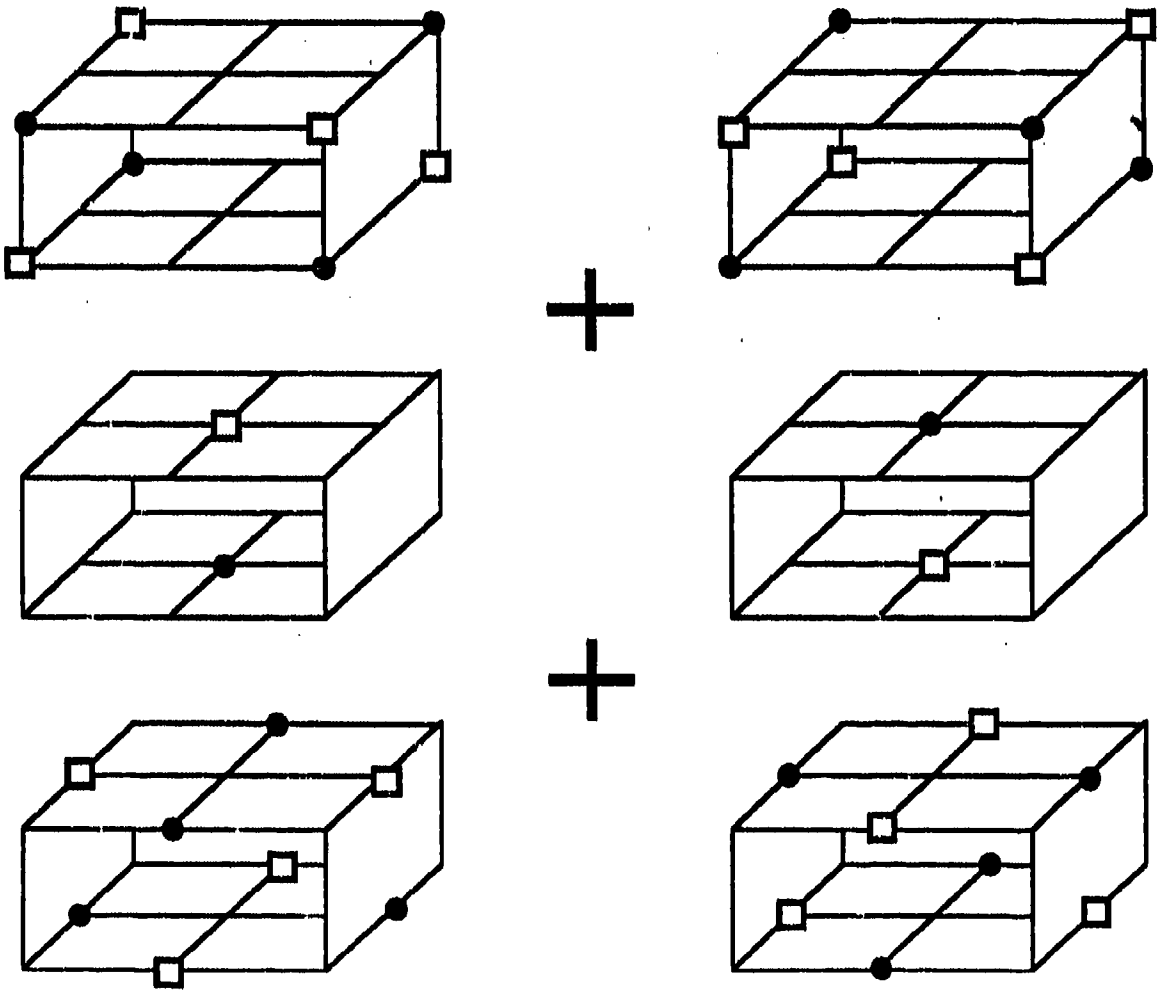
FRACTIONAL DESIGN



FRACTIONAL DESIGN



AS EASY-TO-RECOGNIZE COMPONENTS
(easy to choose run order by labeling points)



For 3-level designs such as this one, confounding properties are much harder to detect, and will require the mathematical checks discussed in Section V. The figure below shows a modified design with improved confounding patterns. This was generated by Master of Engineering candidates after a one hour tutorial on these graphical techniques (Young, Moore, and Girard, 1987).

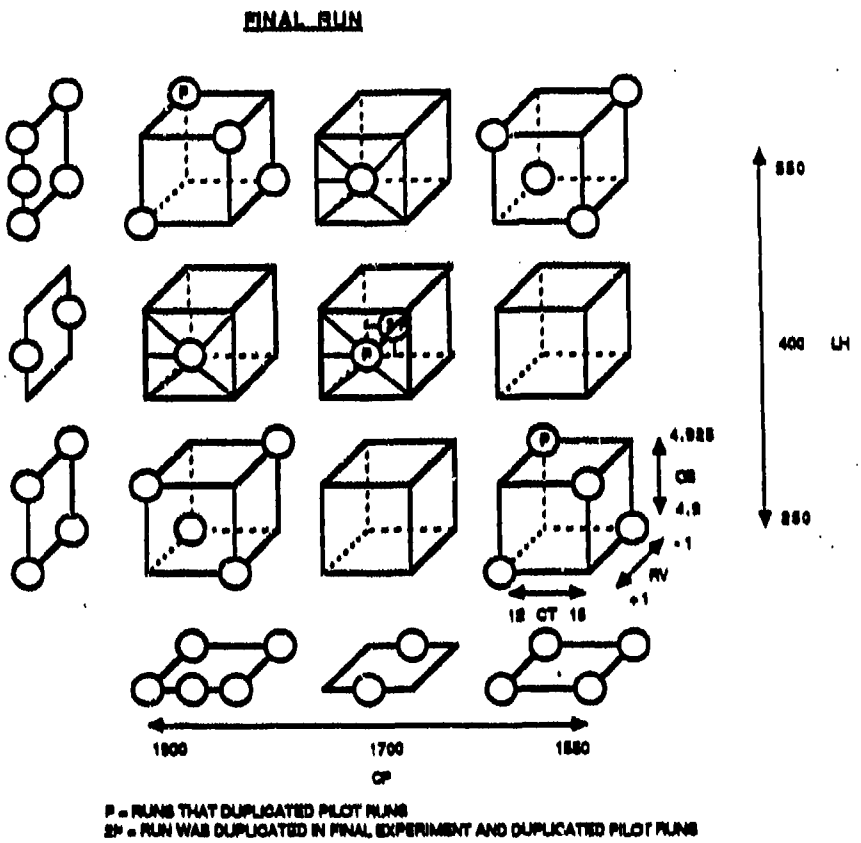
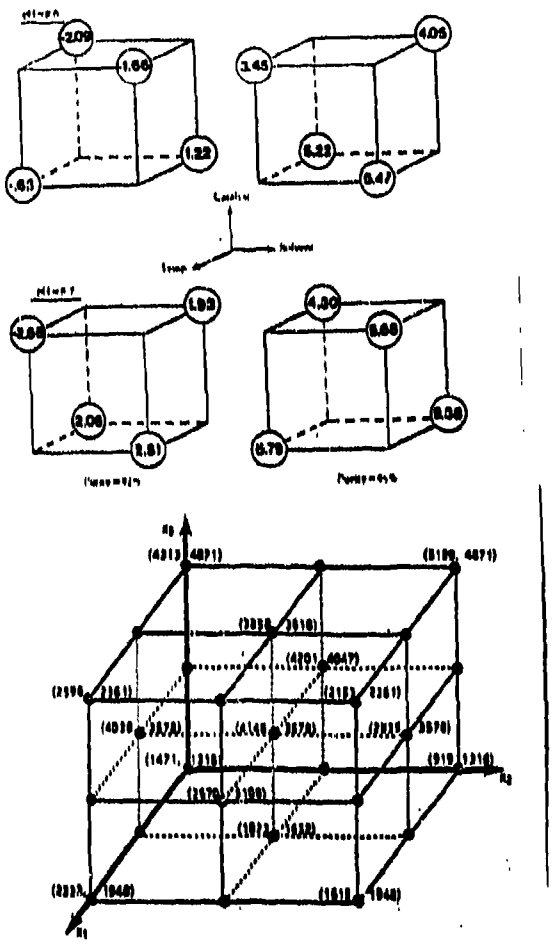


Figure 12

Exercise 10: Use graphical methods to design an experiment to fit a response metamodel to a simulation system of your choice. The design should be for at least three independent variables, at least one of which is at more than two levels. At least ten runs should be planned.

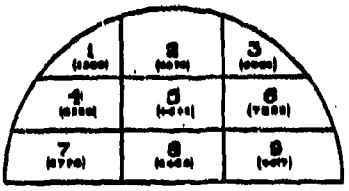


D. Displaying Results on Graphical Designs

Graphical designs provide another advantage: they provide a frame on which to visualize the system responses. This allows one to interpret the findings more easily and to propose appropriate models for the response function more easily.

The simplest examples of such a display are shown here. The first is from Snee (1985a). Pearson (1934) draws simple examples of brick strength vs. kiln location.

DIAGRAMMATIC SECTION OF KILN



FIGURES IN BRACKETS ARE AVERAGE STRENGTHS OF BRICKS IN LB.PER.SQ.IN. FROM THE 9 REGIONS.

This figure is from Bates (1989). Elaborate physical design models are illustrated below and on the following page.

Sources:
below: Box and Wilson (1951)
next page: Neyman (1935)

Box's physical model of the original central composite design was less complicated than the figure below. A simple bare wire structure, it was augmented to display response values using insulated wire. Each vertex was wrapped with a piece of insulated wire. The color of the insulation was keyed to the level of response, from red to blue to white. Design points with intermediate values of the response function were wrapped with the two adjacent colors (red&blue or blue&white).

(Model on loan from J. Stuart Hunter)

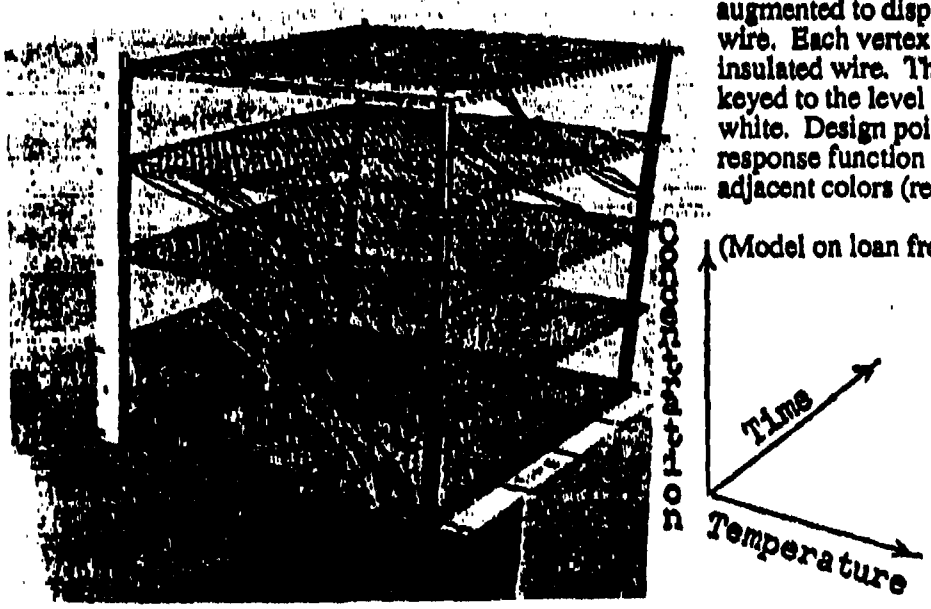


FIGURE 15. PHOTOGRAPH OF A THREE DIMENSIONAL MODEL SHOWING THE CONTOURS OF THE APPROXIMATE PLANE STATIONARY RIDGE SYSTEM CONSTRUCTED ON THE BASIS OF TRIALS RUN AT THE POINTS INDICATED BY THE MARBLES. A DESCRIPTION OF THIS PARTICULAR EXPERIMENT IS FOUND IN SECTION 8. THE TYPE OF EXPERIMENTAL DESIGN USED IS THAT ILLUSTRATED IN FIGURE 8.

FIG. 1.

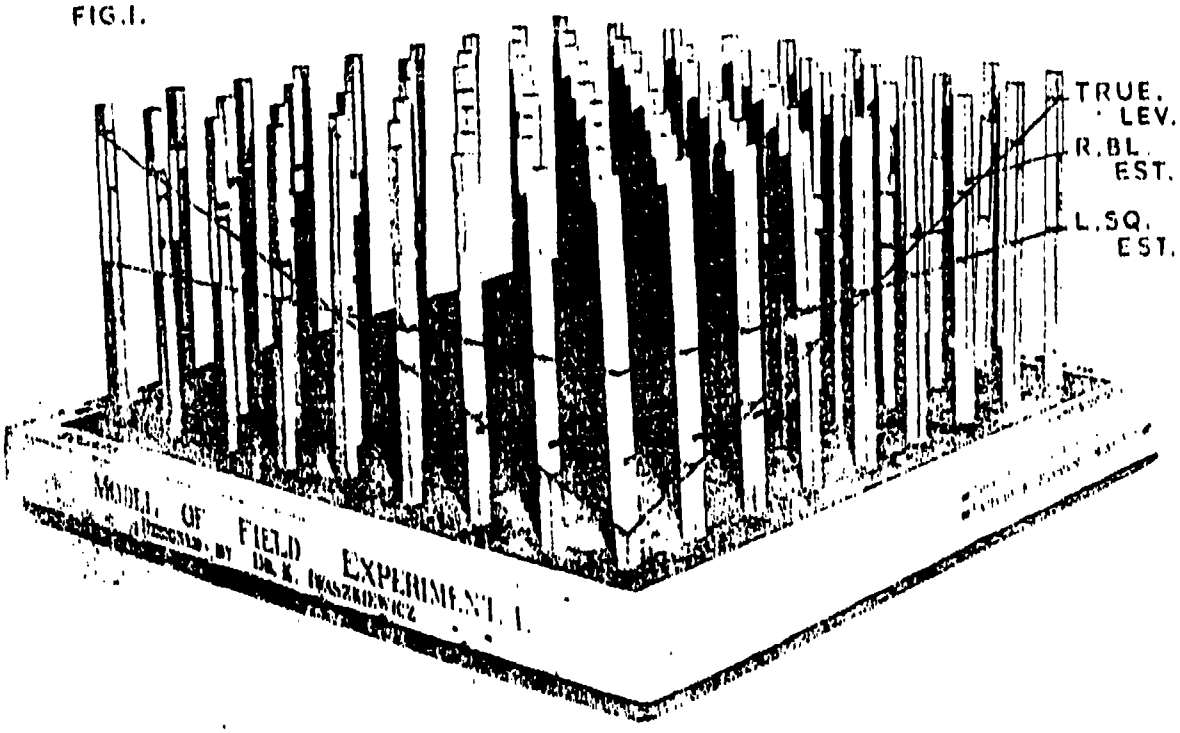
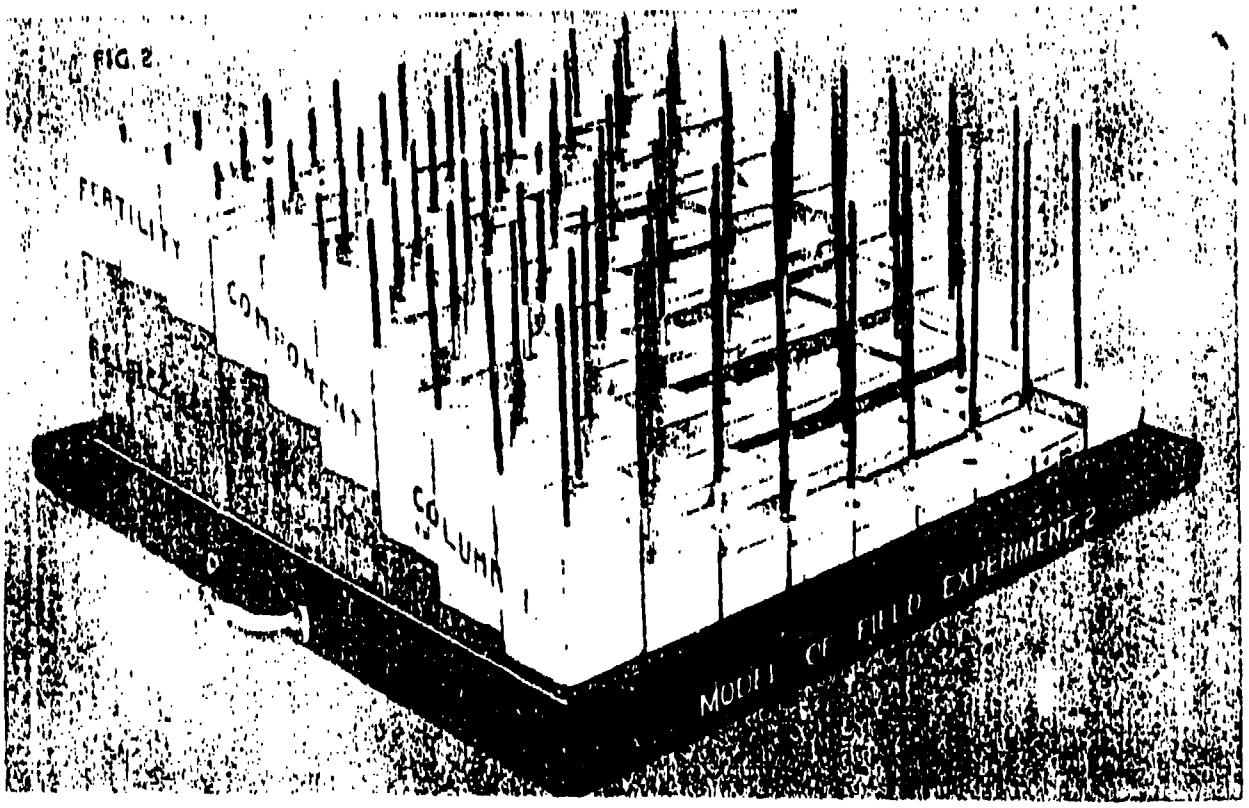

TRUE. LEV.
R.BL. EST.
L.SQ. EST.
MODEL OF FIELD EXPERIMENT 1.

FIG. 2.


FERTILITY
COMPONENT
COLUMN
MODEL OF FIELD EXPERIMENT 2

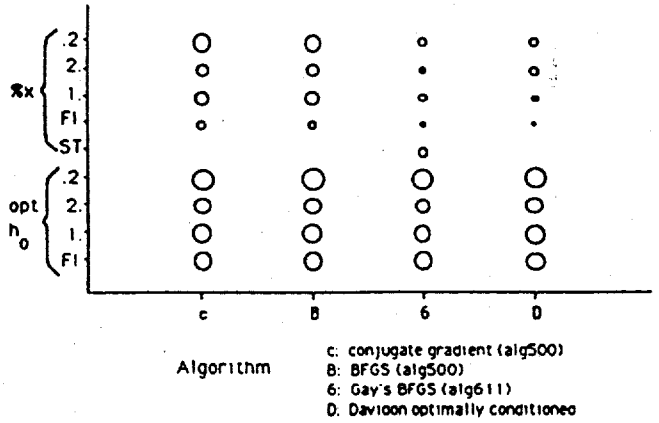
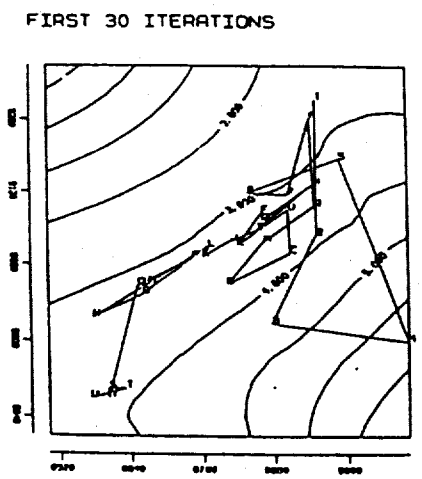
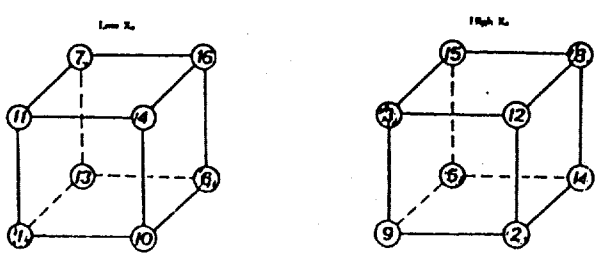


Figure 9. Diameter of circles represents average rank of function reduction for each finite difference method on 18 test functions, with test function arguments shifted by 100. Rank comparisons over all optimization codes. Function accuracy 2 decimal digits, shifted starting points.

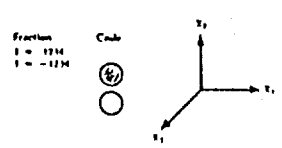
Another way to display the level of response is with the size of the ball at the vertex. For two-factor experiments at two or more levels, this display is called a bubble chart. This example is for an experiment comparing the performance of several types of optimization algorithms with several strategies for computing finite difference derivatives (Barton, 1990). A larger circle means that the effectiveness was higher.



This figure from Barton (1985) shows the contours of the fitted response function superimposed with the sequential simplex (Nelder-Mead) design points.



It is also possible to label the design points with their run order, to look for confounding problems. The figure at the left is from Snee (1985a). A better approach would be to shade the circles corresponding to run order, or to size them from small circles to large based on run order. This would help to make confounding patterns more obvious.



IV. EXERCISE REVIEW:

A. Fishbone/IDEF Diagrams

B. Multidimensional Point Plot

VALIDATING DESIGNS THAT HAVE BEEN CREATED GRAPHICALLY

GRAPHICAL TOOLS EXPLOIT THE POWER OF THE "RIGHT BRAIN"
- CREATIVITY
- INSIGHT

BUT THESE TECHNIQUES ARE QUALITATIVE AND PRONE TO ERROR

WE NEED MATHEMATICAL CHECKS:
- CAN WE ESTIMATE THE PARAMETERS?
- HOW GOOD WILL OUR ESTIMATES BE?
- HOW BADLY WILL THE ESTIMATES BE CONFOUNDED?
- HOW WELL CAN WE EXPECT THE MODEL TO PREDICT?

A GENERAL LINEAR MODEL

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 (x_1)^2 + \text{error}$$

$$\left[y^i = a_0 + a_1 x_1^i + a_2 x_2^i + a_3 x_3^i + a_4 (x_1^i)^2 + \text{error}^i \right]$$

$$\text{error}^i \quad iid \sim N(0, \sigma^2)$$

$$\vec{y} = X \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} + \vec{\text{error}}$$

DESIGN MATRIX

$$X = \begin{bmatrix} x_1^1 & x_2^1 & x_3^1 & (x_1^1)^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1^k & x_2^k & x_3^k & (x_1^k)^2 \end{bmatrix}$$

$$\Sigma = \text{variance-covariance of} \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \end{pmatrix}$$

$$\Sigma_{ij} = \text{cov}(\hat{a}_i, \hat{a}_j)$$

$$\Sigma = \sigma^2 (X^T X)^{-1}$$

V. VALIDATION OF EXPERIMENT DESIGN

These techniques allow us to use the creative human graphical processing resources to create new designs and to gain insight on design properties. Graphical methods are necessarily qualitative, however. We need to back up our graphical work with precise mathematical measures of design goodness.

A. Importance of the Model

As we discussed earlier, the properties of an experiment design depend closely on the kind of model the experimenter has chosen. What properties do we care about? These include

- estimability of parameters
- the variance of the parameter estimates
- the covariance/correlation of "
- the mean square prediction error over the design space

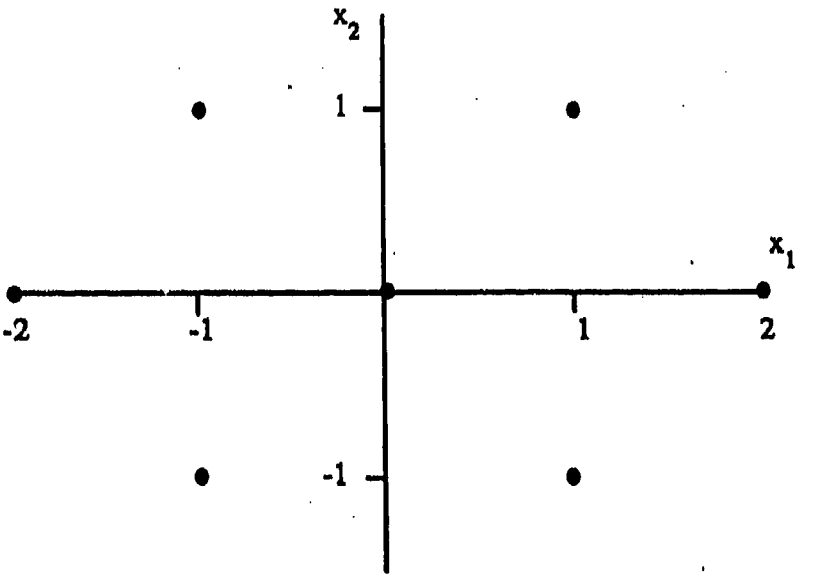
The most common applications in simulation metamodeling will be General Linear Models: ANOVA, ANOCOVA, or Regression models with iid Gaussian error. There are other interesting choices for metamodels, particularly for deterministic computer simulation output (Sacks, Schiller, and Welch, 1989). At present these models require computer-aided selection of design points.

B. The General Linear Model and the Design Matrix

We will often represent a single dependent variable (say WARP) as a y, and the independent variables as x's. In this formulation the unknown parameters are the a's and the value of σ^2. Note that the a's appear LINEARLY in the formulation. The x's do not have to be linear, as illustrated by the fourth factor. This could also be written as x_4 rather than $(x_1)^2$.

The observations are indexed by a letter (i), and the independent variable coefficients are indexed by another letter (j). The x's are indexed by both letters. The representation is often x_{ij}, rather than as shown at left. This model can be written in matrix form. The matrix X is called the design matrix.

Exercise 11: Construct the design matrix (don't forget the column of 1's for the intercept) for the following design



a) when the model is

$y = a_0 + a_1x_1 + a_2x_2 + \text{error}$

b) when the model is

$y = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + \text{error}$

C. Measures of Design Goodness

OPTIMAL DESIGN:

D-Optimality

maximize $\det(X^TX)$

corresponds to minimum volume confidence ellipsoid for model parameters, i.e. tightest estimates

E-Optimality

maximize minimum eigenvalue of X^TX

corresponds to minimizing maximum prediction error over a unit sphere

Measures for the goodness of a design will be difficult to generate for arbitrary models of system response. For the general linear model, though, we have a simple situation: the goodness of the design for fitting a general linear model will depend only on the properties of the X matrix. In particular, the estimate for the variance-covariance matrix for the parameter estimates (except the estimate for σ^2) depends on $(X^TX)^{-1}$. Since the x's for a candidate design are known before the results are collected (only the y's are unknown) we can assess the quality of the design before the experiment is run!

The list at the left presents two of the traditional definitions of good or 'optimal' designs. In general we would like the diagonal entries of Σ to be small. This corresponds to small variances for our parameter estimates. We would like the off-diagonal elements to be even smaller. This corresponds to low confounding of effects, that is, small covariance of parameter estimates.

SAS CODE to check Σ

```
CMS FILEDEF DES1 DISK DESIGN DATA A;

DATA DESIGN;
   INFILE DES1;

   INPUT X1 X2 X3;
      X1X1 = X1*X1;
      X1X2 = X1*X2;
         ⋮
         ETC

 PROC MATRIX;

    FETCH X DATA = DESIGN;
    SIGFAC = INV((X')*X);
    PRINT X SIGFAC;
```

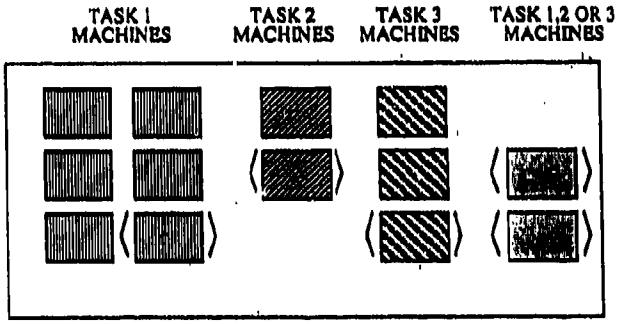
D. Checking (X^TX) and $(X^TX)^{-1}$

The code at the left is for SAS. We enter the design and can augment it if we have nonlinear x terms in the general linear model. Don't forget to include a column of 1's if you have an intercept term. We can study the resulting matrix which is (except for the factor σ^2) the variance covariance matrix Σ.

Other software such as MATLAB ACED, COED, etc. can be used to generate mathematical properties of design goodness for general linear models.

Exercise 12: Construct the design matrix (don't forget the column of 1's for the intercept) for your own problem. Use SAS, MATLAB, ACED, or any other software to evaluate $(X^TX)^{-1}$.

MANUFACTURING SYSTEM LAYOUT DESIGN PROBLEM



E. Graphical Views of (X^TX) and $(X^TX)^{-1}$

While it is easy to generate these matrices, it is difficult to understand the design properties for designs with more than two or three parameters.

Instead we'll look for way to represent these matrices graphically! Understand that these graphics are for *analysis* rather than *generation* of new experiment designs. They are used for checking rather than creating.

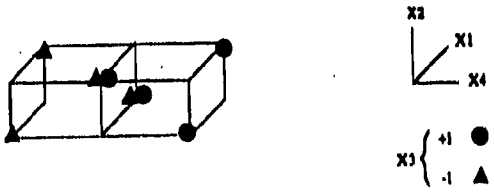
We'll begin by introducing a new example. We'll try to design an experiment where the factors are quantitative but integer, and the realistic values correspond to small integers. An analogy might be a battle deployment plan where the number of divisions to assign to an area must be selected.

In our example the simulation model represents the performance of a small manufacturing operation. The goal of our study is actually optimization, but it was pursued by Kleijnen and Standridge (1987) through fitting a polynomial approximation for sensitivity analysis and insight (goal 2).

The figure at the top of this page illustrates the manufacturing system layout problem. The factory produces a single product, which requires three tasks to be performed in sequence. Special machines are available to perform each of these tasks. In addition, flexible manufacturing machines are available (at higher cost) which can perform all three machining operations. To meet throughput requirements on task 1, 5.2 machines are required. For task 2, 1.3 machines are needed. For task 3, 2.6 machines are needed. We cannot purchase a fractional machine, and so without any flexible equipment we will need six machines for task 1, two machines for task 2, and three machines for task 3. On the other hand, the purchase of one flexible machine could reduce the demand on one or more other tasks, and so reduce the number of machines needed. If only five machines for task 1, one for task 2, and two for task 3 are purchased, two flexible machines will be needed to meet throughput requirements.

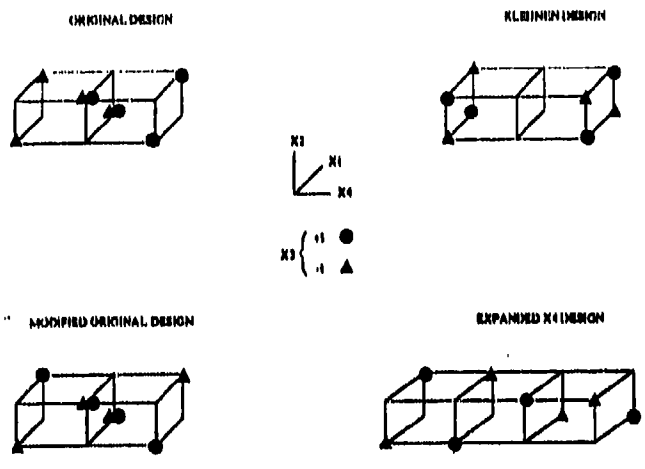
The goal of this design is to estimate the throughput as a function of the number of machines of each type that are purchased. This is expected to give some insight on the most economical choice. Of course, there are only 2x2x2x3=24 possible configurations, so the exact answer can be had for the cost of 24 simulation runs. We assume that we have a budget of only 8 runs from which we must make a decision.

ORIGINAL DESIGN



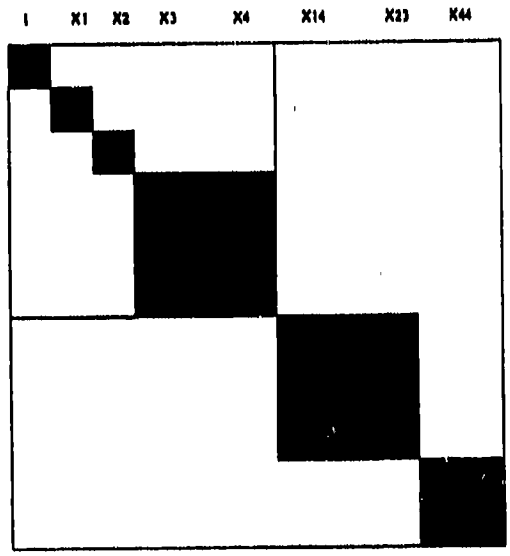
The original design had a serious problem. Can you see it graphically? The effect of the number of flexible machines (x_4) is confounded with the number of machines of type 3 (x_3-- here re-coded as ±1 rather than 2 or 3). Since the confounding is only partial, it is hard to eyeball how severe the problem is. We need to look at the covariance and correlation matrices.

EXPERIMENT DESIGNS
FOR A
MANUFACTURING DESIGN PROBLEM



This figure shows several designs along with the original design. Which is better? In what way? These are difficult questions to answer, even when the variance-covariance matrix is available. For only 8 runs, the matrix with all main effects, two factor interactions, and a quadratic term for the number of flexible machines IS NOT INVERTIBLE. To study the covariance matrix the model must be limited to a few terms.

ORIGINAL DESIGN



The second figure illustrates the entries of the variance-covariance factor $(X^TX)^{-1}$ for the original design. The edge length of each diagonal block is proportional to the square root of the diagonal entry in $(X^TX)^{-1}$. Thus the size of the diagonal blocks are proportional to standard deviations of coefficient estimates and the areas are proportional to their variances. This structure determines the size of the off-diagonal blocks. Each block's shading is proportional to the absolute value of the correlation coefficient of the two parameters corresponding to that row and that column. The on-diagonal blocks are shaded 100% (black) because a term's correlation with itself is 1.

Q: How was this graphic created?

A: Using Canvas™ for the Macintosh. The software allows one to see the edge dimension of a square as it is created. It also provides a numerical scale (0-100) for shading boxes. Matlab™ was used to generate the covariance/correlation data. The code for one of these is shown at the left.

Q: What will a good design look like?

A: The overall size of the box is determined by the sum of the standard deviations of all estimated parameters. The smaller the better. If we are particularly interested in a subset of the parameters, this sub-box should be small. Furthermore, the off-diagonal shading should be as light as possible. It should be white in the sub-box corresponding to our most important parameters (no confounding), and it should be pale in the rows and columns which overlap the sub-box but are outside it.

Matlab Code:

```
A3 = [ -1 -1 -1 -1
 1 -1 1 -1
-1 1 1 -1
 1 1 -1 -1
-1 -1 1 1
 1 -1 -1 1
-1 1 -1 1
 1 1 1 1]

B = [ A3(:,1) .* A3(:,2)]
C = [ A3(:,1) .* A3(:,3)]
D = [ A3(:,1) .* A3(:,4)]
E = [ A3(:,2) .* A3(:,3)]
F = [ A3(:,2) .* A3(:,4)]
G = [ A3(:,3) .* A3(:,4)]
H = [ A3(:,4) .* A3(:,4)]

H = H - 1

AA3 = [ A3 B C D E F G H ]

XTX = AA3'*AA3

corr(AA3)

%XTXI = inv(XTX)
```

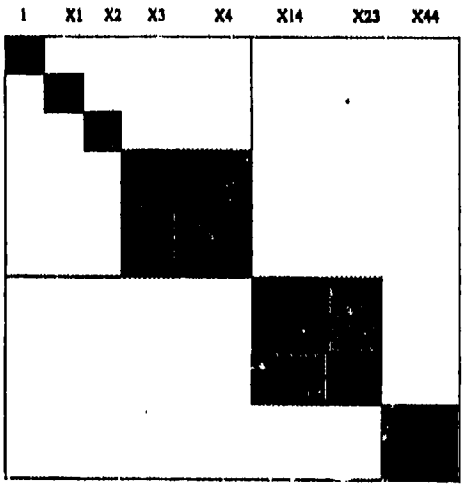
The figures on this page show the graphical representation of the correlation/variance structure for all candidate designs. The important effects are the constant (1), the quadratic term (x_{44}) and the linear terms x_1, x_2, x_3, and x_4.

Note particularly:

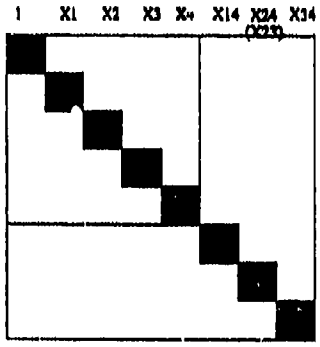
- the original design has substantial confounding
- the overall box for the original design is bigger than the designs at the right
- the modified original design eliminates confounding from the sub-box of interest while providing the same accuracy on parameter estimates
- the Kleijnen box is smaller overall and smaller for the key effects that it does estimate
- the Kleijnen design cannot estimate one of the key parameters (x_{44})
- the expanded x_4 design provides both a smaller box and estimability of (x_{44})

EXPERIMENT DESIGNS
FOR A
MANUFACTURING DESIGN PROBLEM

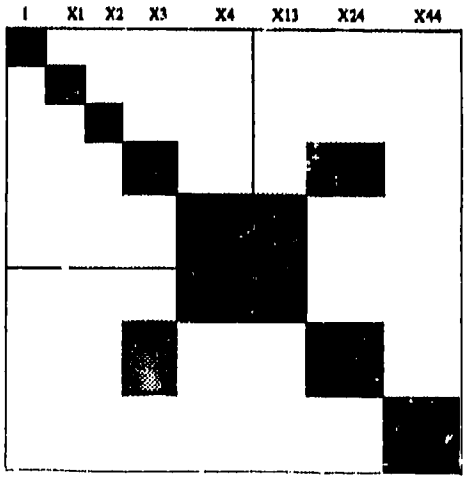
ORIGINAL DESIGN



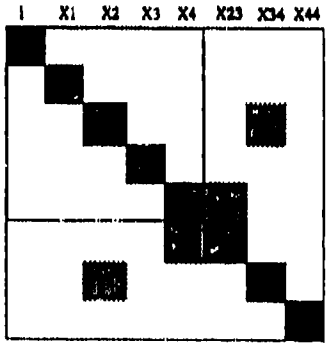
KLEIJNEN DESIGN



MODIFIED ORIGINAL DESIGN



EXPANDED X4 DESIGN



This graphical tool for analysis of correlation and covariance needs further revision. It is not clear from the picture that the Kleijnen design cannot estimate the quadratic effect. This is important, though; it is like throwing the baby out with the bathwater.

It is interesting that covariance patterns can be traced from dark cells to the corresponding variable labels on the boundary. If the matrix could somehow include all model terms (not just an estimable subset) one could construct Taguchi linear graphs by reading the dark boxes as edges between variables. An example of Taguchi linear graphs is shown here.

This is also a good reminder that the Taguchi Linear Graphs, just like the graphical correlation/covariance matrices, are useful for analysis (or cookbook look-up) but not for creating new designs.

Source: Pignatiello and Ramberg (1985)

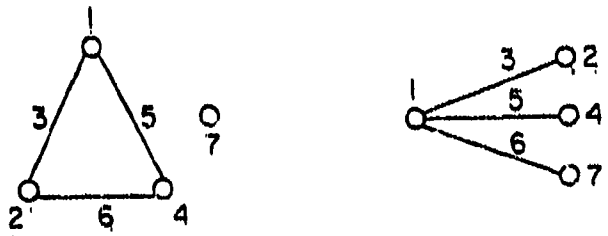


FIGURE 1. Linear Graphs for the L_8 Orthogonal Array.

1 DEFINE GOALS

2 IDENTIFY VARIABLES

3 CLASSIFY VARIABLES

4 DESIGN EXPERIMENT

5 VALIDATE DESIGN

VI. SUMMARY

We have examined graphical methods for three of the five steps in designing experiments. These methods are easy to learn, fun to do, and provide a new dimension of insight to the qualities and consequences of a particular design.

Graphical designs are not a replacement for other design techniques. Like other techniques, they have their own advantages and disadvantages. But until recently, they have not been publicized.

STRENGTHS AND WEAKNESSES
OF
GRAPHICAL METHODS

+ FLEXIBLE
  - MAKE TRADEOFFS VISUALLY
  - INCORPORATE CONSTRAINTS GRAPHICALLY
+ ROBUST (to a degree; for 1st order models & interactions)
+ USES POWERFUL COMPUTING DEVICE
  (human vision system and right brain)
+ EASY TO USE AND EASY TO REMEMBER
+ EASY TO TEACH

- NOT QUANTITATIVE
- DIMENSIONAL LIMITATIONS

As this quote from George Box (1984) suggests, graphical methods are perhaps more important for design than for analysis, since design is a highly creative activity. In this age of computerized statistical packages, errors in analysis can always be redone, but an error in design cannot be fixed by a clever analysis.

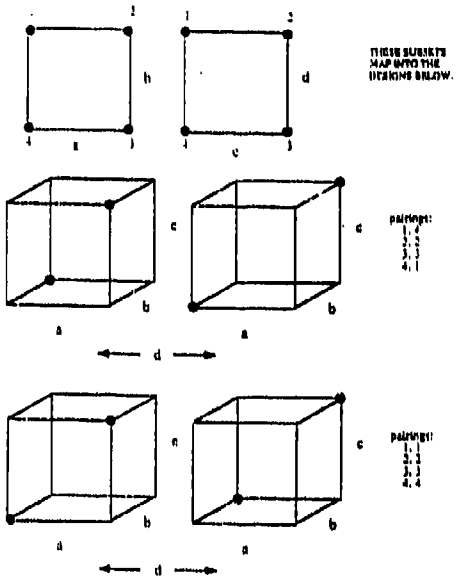
It is well-known that while the left brain plays a conscious and dominant role, one may be quite unaware of the working of the less assertive right brain. For example, the apparently instinctive knowledge of what to do and how to do it enjoyed by an experienced tennis player comes from the right brain. It is significant that this skill may be temporarily lost if we invite the tennis player to explain how he does it, and thus call the left brain into a dominant and interfering mode.

In this context we see the data analyst's insistence on "letting the data speak to us" by plots and displays as an instinctive understanding of the need to encourage and to stimulate the pattern recognition and model generating capability of the right brain. Also, it expresses his concern that we not allow our pushy deductive left brain to take over too quickly and perhaps forcibly produce unwarranted conclusions based on an inadequate model.

While the accomplishment of the right brain in finding patterns in data and residuals is of enormous consequence to scientific discovery, some check is obviously needed on its pattern-seeking ability, for common experience shows that some pattern or other can be seen in almost any set of data or facts. A check

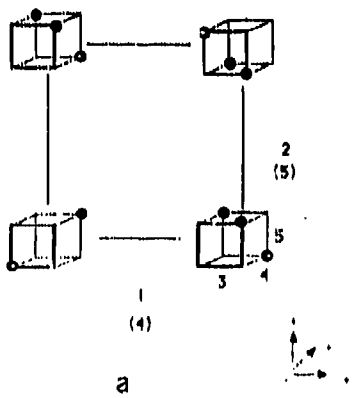
VII. FURTHER TOPICS

A. Graphical Designs for Many Factors and/or Many Levels



Graphical methods for experiment design are inherently low-dimensional. How can they be used to help design experiments with many factors?

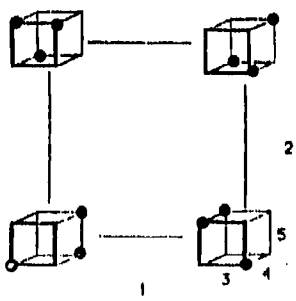
It may be possible to develop low-dimensional designs on a subset of the factors and 'paste' these subset designs together to form the complete design. Pasting here means a 1-1 pairing between the design points for one subset of factors and the design points for the other subset. The goodness of the design will depend not only on the (graphical) quality of the two subset designs, but on the way in which the subset design points are paired. This is illustrated in the figure at the left. Two full-factorial 2^2 subset designs (a,b and c,d) are paired in two different ways, yielding two different 2^{4-2} designs.



a

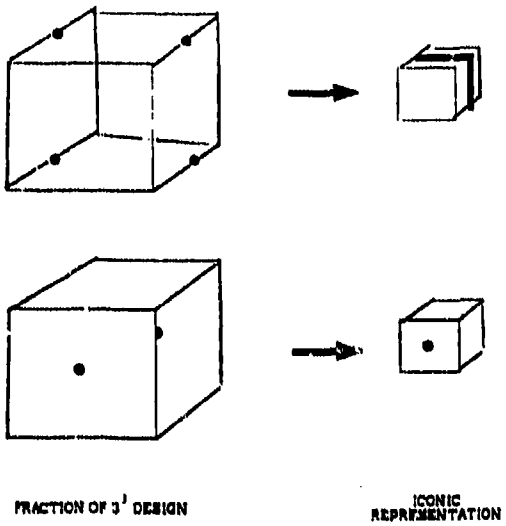
Some designs can be projected onto a lower-dimensional space. Draper (1985) found that a Plackett-Burman (1946) design for 11 factors had only two possible projections* in 5 dimensional space. They are illustrated graphically at the left.

Projecting designs into lower-dimensional space can help in design analysis, but its value for design synthesis remains unclear.



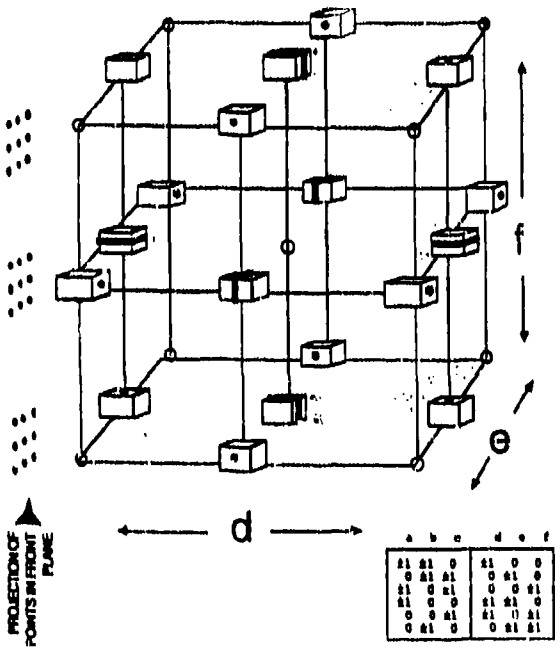
b

*allowing sign changes and permutations of axes



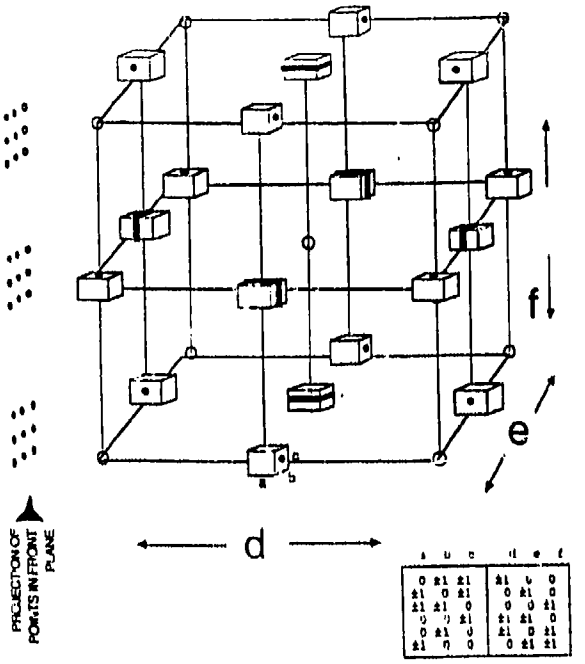
Even with relatively few dimensions, say six, it can be difficult to visualize and develop a design if the factors each occur at three or more levels. In this case the nested cube approach described earlier becomes visually distracting rather than helpful.

One way to simplify the representation is to substitute icons for the smaller cubes. The figures here show icons for small cubes for a Box-Behnken (1960) design. Each small cube has 27 possible sites for an experiment, based on the 3^3 plan for these three factors.



Box - Behnken 3^6 Fractional Design

Using these icons gives this representation for the Box-Behnken design for 6 factors at three levels. Some characteristics of the design are immediately apparent from its graphical representation. First, none of the extreme corners of the design space are included in the design. Second, the center point is not included.



Bad - Barion 3^6 Fractional Design

These icons also make it easier to synthesize new designs by manipulating the icon patterns. This design has poorer properties than the Box-Behnken design. Reasons become clear when one looks at the design projections. Those for the front face of the large cube (projecting out d [illegible]) are shown at the left of the figure.

Again, these weaknesses are more clear by looking at the design and its projections than by trying to read and compare the tabular representations at the lower right of each figure.

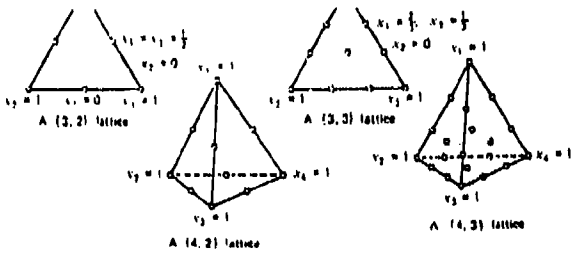
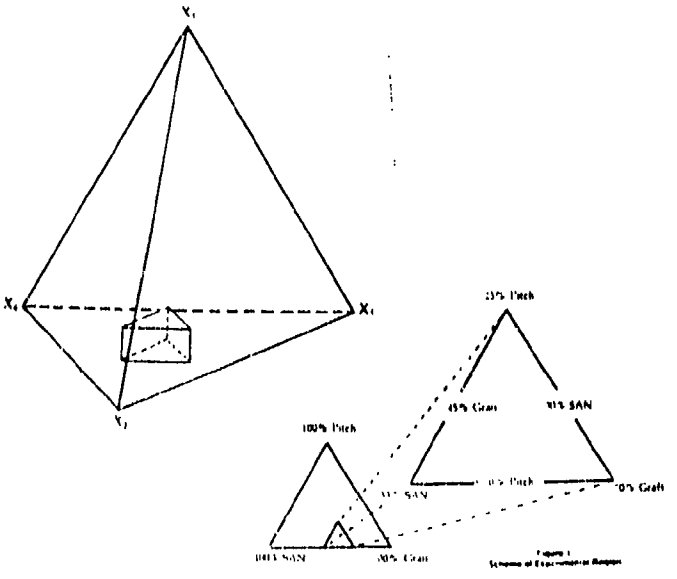


FIGURE 2.1. Some (3, m) and (4, m) simplex-lattice arrangements, m = 2 and m = 3.

Figure 2. Experimental Region for Instant Soup Thickener Study.



B. Mixture Experiments

While mixture experiments are often associated with chemical and petroleum industries, they are useful in many resource allocation problems as well. They can be used for military strategy, medical treatment, and corporate strategic planning.

Mixture experiments involve factors whose sum must be a constant. For example, the constituents of a chemical compound must sum to 100%. The capital expenditures for a large corporation must sum to the amount allocated for this purpose (if we include a dummy category: reserves). Optimal weapon mix for strategic forces might be studied through mixture experiments, especially as budget constraints become ever tighter.

Design points for a mixture experiment are represented graphically as points on or in a simplex (triangle, tetrahedron, etc.), in contrast with the cubes and rectangles used for factorial designs. The figures at the left show designs for three and four factors at two and three levels (Cornell, 1981).

In addition to the usual mixture constraint, most practical designs have additional requirements that limit the design space. These constraints may be to focus the study on reasonable mixtures, as in the figures from Koons and Wilt (1985) and Hare (1985) at the left, or they may be due to real limits, as shown in the figure from Snee (1981) on the following page.

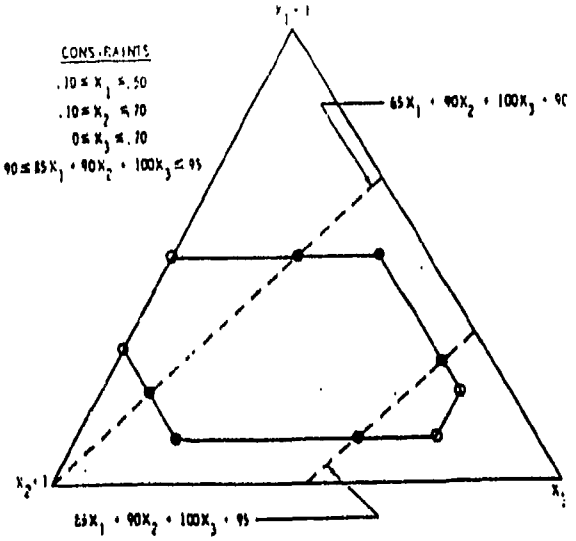


Figure 4. *Three-Component Mixture System With Single Component and Multiple Component Constraints*

This design is limited by 5 bounding constraints on the three components in the mixture, and by two multiple-component constraints. The design points selected for the experiment are indicated by solid dots.

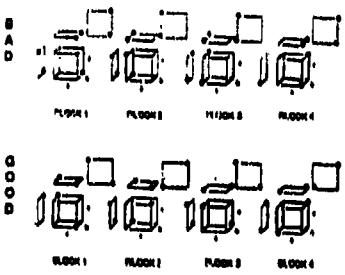
The computerized design software XVERT (Snee, 1981) uses geometric concepts of edges, vertices, and face centers to select good design points. This approach to design can be thought of as *mental graphics*, since a graphical image is used but is not actually drawn.

C. Incomplete Block Designs

A graphical representation for an incomplete block design was presented on page 27. A blocking variable is thought of much as any other factor, except that a) it is qualitative, not quantitative, and b) there are usually more than two values (blocks).

The blocking factor is usually qualitative, not quantitative. This makes checking for confounding patterns more difficult. The graphical subcomponents corresponding to each block must be shuffled around, mentally at least, to identify confounding patterns. The confounding on page 27 is seen by pairing the second and third blocks against the 1-4 pair.

'Mental graphics' yields several general classes of partially balanced incomplete block designs. These are described at the left by Kempthorne (1952).



27.6.1 Geometrical Configurations

Several designs arise by translation of geometrical configurations: e.g., the Desargue configuration of 2 triangles in perspective gives rise to a design for 10 treatments in blocks of 3. If we take any regular polyhedron and regard the points as treatments and the faces as blocks, we get a partially balanced incomplete block design. For example, with a cube (Figure 27) we get the design

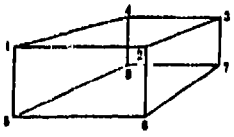
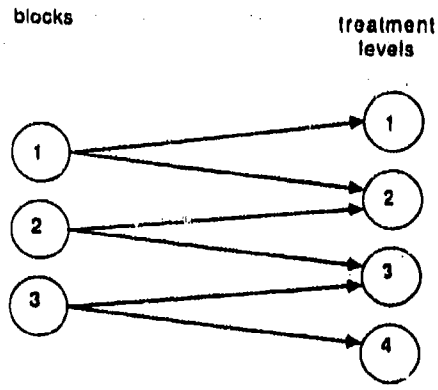


Figure 27.

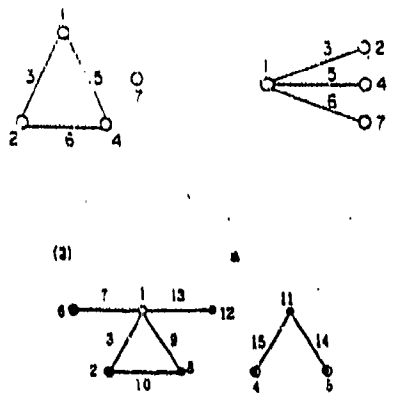
D. Network Representations

The edges of a multidimensional point plot, along with the design points, make up a network that directly represents the design. There are also *indirect* network representations that can be useful in creating and analyzing designs.



This network represents the allocation of factor levels to blocks. Butz (1982) relates connectivity in such networks to *estimable contrasts* of factor levels. These plots can be used to develop and analyze designs for use with ANOVA models.

ALL TREATMENT CONTRASTS ESTIMABLE ⟷ BIPARTITE GRAPH IS CONNECTED



As mentioned on page 53, Taguchi's 'linear graphs' provide analytical insight, but do not really help generate designs. They illustrate confounding patterns graphically by labeling nodes and edges with effects. For example, the triangle at the upper left indicates that the main effect of factor 5 will be confounded with the 1-4 interaction effect in that design.

The linear graphs make cookbook design selection a more effective approach. A catalog of these network representations makes it possible to select a design with confounding patterns that match the particular needs at hand.

Sources: Taguchi (1980) and Pignatiello and Ramberg (1985)

The defining contrast is $+ABC$. The confounding pattern can be summarized as $I+ABC$. Written out in more detail, the three contrasts measure:

$$A + BC$$
$$B + AC$$
$$C + AB, \quad \text{respectively.}$$

The six points of the lattice of Figure 1a designate the six effects just listed. The tie-lines, first used in reference [5], show the pairs that are confounded, that is, whose sum is measured by the corresponding contrast. This over-simple

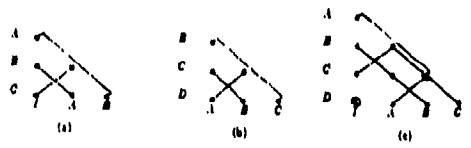
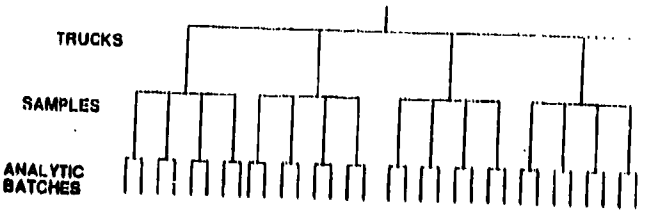


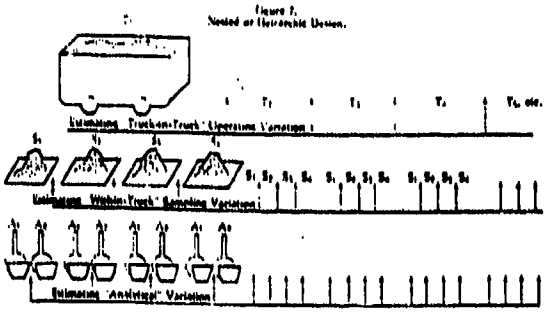
FIG. 1. Confounding patterns of $2I_V$ for 2^{3-1} and 2^{4-1}.

Cuthbert Daniel (1962) also used networks to represent confounding patterns in fractional factorial designs. These are harder to decipher. Their usefulness seems to be for design analysis, not design synthesis.

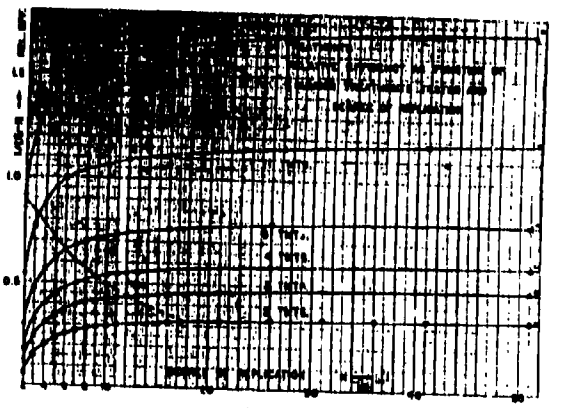


SCHEMATIC for ANDREWS' HIERARCHIC DESIGN

Graphical representations are helpful in developing nested designs for random and mixed effects models. The simple schematic at right makes the sampling pattern clear. Leone, et. al. (1968) use these simple figures to present their hierarchical designs.

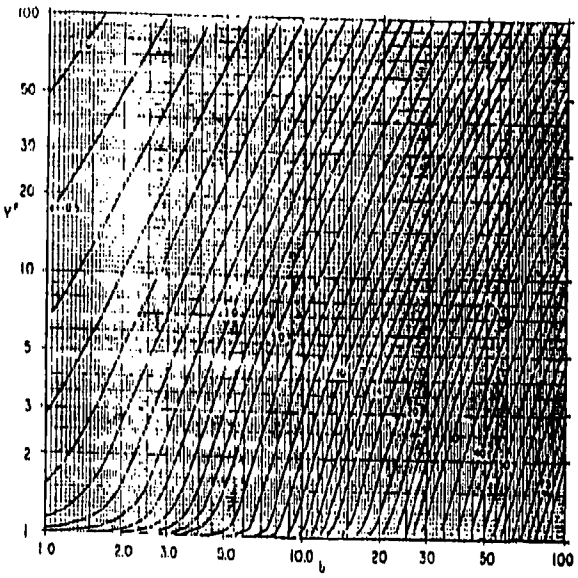


Andrews used more elaborate representations for nested designs. The figures take more time to draw, but they can prompt the experimenter to think about important procedural or design issues. This elaborate figures may well have merit over its simpler counterpart above.

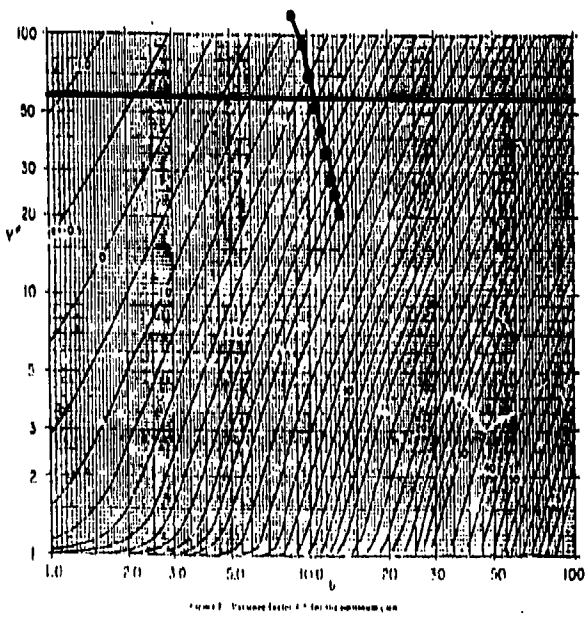


E. Nomograms

Nomograms are graphical aids for *computation*. In experiment design they can be used for design (choosing sample sizes, operating conditions) or for analysis (identifying design properties such as variance of the estimated parameters). In the example at the left, Villars (1951) shows the relative efficiency of ANOVA designs as a function of the number of levels of the treatment factor and the number of replications.



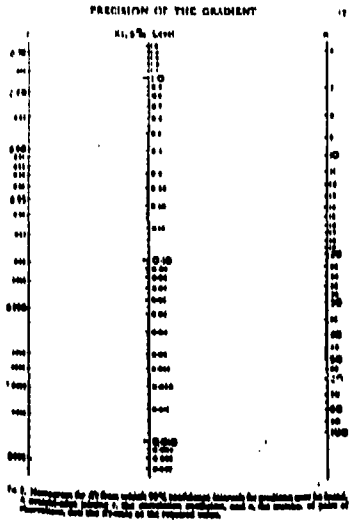
This nomogram from Nelson and Kielpinski (1975) is for high temperature accelerated lifetests. It allows one to identify the variance of a predicted response as a function of test temperatures and sample proportions.



$a^* \cong 2.4$
$b^* \cong 10.3$

Nomograms can be modified graphically to provide alternative or additional information. This allows one to optimize other properties (than prediction variance) by adding graphical constructs to the figure. This example allows the experimenter to choose test conditions that will yield the required accuracy while minimizing the maximum test temperature. This is important because high test temperatures can introduce failure mechanisms that are not active in the normal operating range.

Source: Barton (1987)



There are many nomograms aid in sample size determination. The example at the left is from Beech (1961) for the selection of sample size to estimate the regression coefficient to a desired precision.

Andrews, H.P. (1964), "The Role of Statistics in Setting Food Specifications," *Proceedings of the Sixteenth Research Conference of the Research Council of the American Meat Institute*, 43-56. Reprinted in *Experiments in Industry: Design, Analysis, and Interpretation of Results*, eds. R.D. Snee, L.B. Hare, and J.R. Trout, Milwaukee: American Society for Quality Control, 1985.

Barton, R.R. (1982). RCA Internal Correspondence to I. Chen.

Barton, R.R. (1985), RCA Internal Report on CONTURIT.

Barton, R.R. (1990), "Experiments in Computing Finite Difference Derivatives When Optimizing Low Accuracy Functions", Technical Report No. 884, School of Operations Research and Industrial Engineering, Cornell University.

Barton, R.R., (1987), "Optimum Accelerated Lifetest Plans which Minimize the Maximum Test Stress," Technical Report No. 7 , School of Operations Research and Industrial Engineering, Cornell University (accepted for publication in *IEEE Transactions on Reliability*).

Bates, C.B. (1989), "Application of a Composite Design to Test a Combat Simulation Model", presented at the 35th Conference on Design of Experiments in Army Research, Development, and Testing, Monterey, California, October 18-24.

Beech, D.G. (1961), "Some Notes on the Precision of the Gradient of an Estimated Straight Line," *Applied Statistics, 10, 14-31.*

Box, G.E.P. (1954), "The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples," *Biometrics*, 10, 16-60.

Box, G.E.P. (1984), "The Importance of Practice in the Development of Statistics," *Technometrics*, 26, 1-8.

Box, G.E.P., and Behnken, D.W. (1960), "Some New Three Level Designs for the Study of Quantitative Variables," *Technometrics*, 2, 455-475.

Box, G.E.P., and Draper, N.R. (1959), "A Basis for the Selection of a Response Surface Design," *Journal of the American Statistical Association*, 54, 622-653.

Box, G.E.P., and Hunter, J.S. (1957), "Multifactor Experimental Designs for Exploring Response Surfaces," *Annals of Mathematical Statistics*, 28,195-241.

Box, G.E.P., and Hunter, J.S. (1959), "Condensed Calculations for Evolutionary Operation Programs," *Technometrics*, 1, 77-95.

Box, G.E.P., and Hunter, J.S. (1961), "The 2^{k-p} Fractional Factorial Designs Part I," *Technometrics*, 3, 311-351.

Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978), *Statistics for Experimenters*, New York: Wiley.

Box, G.E.P., and Lucas, H.L. (1959), "Design of Experiments in Nonlinear Situations," *Biometrika*, 46, 77-90.

Box, G.E.P., and Meyer, R.D. (1986), "An Analysis for Unreplicated Fractional Factorials," *Technometrics*, 28, 11-21.

Box, G.E.P., and Wilson, K.B. (1951), "On the Experimental Attainment of Optimal Conditions," *Journal of the Royal Statistical Society series B*, 13, 1-38.

Butz, L. (1982), "Connectivity in Multi-Factor Designs: A Combinatorial Approach," *Research and Education in Mathematics 3*, Berlin: Helderman Verlag.

Cornell, J.A. (1981), *Experiments with Mixtures*, New York: Wiley.

Daniel, C. (1962), "Sequences of Fractional Replicates in the 2^{p-q} Series,"*Journal of the American Statistical Association*, 58, 403-429.

DeBaun, R.M. (1959), "Response Surface Designs for Three Factors at Three Levels," *Technometrics*, 1, 1-8.

Draper, N.R. (1985), "Small Composite Designs," *Technometrics*, 27, 173-180.

Feder, P.I. (1974), "Graphical Techniques in Statistical Data Analysis," *Technometrics* , 16, 287-319.

Fries, A. and Hunter, W.G. (1980), "Minimum Aberration 2^{k-p} Designs," *Technometrics*, 22, 601-608.

Freeman, G.H., Mark, J.W., and Blake, I.F. (1988), "Trellis source codes designed by conjugate gradient optimization," *IEEE Transactions on Communications*, 36, 1-12.

Fry, K.H. (1961), "Finding New Fractions of Factorial Experimental Designs," *Technometrics*, 3, 359-370.

Hahn, G.J. (1984), "Experimental Design in the Complex World," *Technometrics*, 26, 19-31.

Hahn, G.J., Bemesderfer, J. and Olsson, D.M. (1986), "Explaining Experimental Design Fundamentals to Engineers: A Modern Approach," unpublished.

Hardy, R.L. (1971), "Multiquadric Equations of Topography and Other Irregular Surfaces," *Journal of Geophysical Research*, 76, 1905-1915.

Hare, L.B. (1985), "Graphical Display of the Results of Mixture Experiments," in *Experiments in Industry: Design, Analysis, and Interpretation of Results*, eds. R.D. Snee, L.B. Hare, and J.R. Trout, Milwaukee: American Society for Quality Control.

Hunter, J.Stuart (1985), "Statistical Design Applied to Product Design," *Journal of Quality Technology*, 17, 210-221.

Kempthorne, O. (1952), *The Design and Analysis of Experiments*, New York: Wiley.

Ishikawa, K. (1982), *Guide to Quality Control*, Tokyo: Asian Productivity Organization, (second edition).

John, J.A. and Mitchell, T.J. (1977), "Optimal Incomplete Block Designs," *Journal of the Royal Statistical Society, Series B*, 39, 39-43.

Kleijnen, J.P.C., and Standridge, C.R. (1987), "Experimental Design and Regression Analysis in Simulation: an FMS Case Study," Report FEW-212, Department of Economics, Tilburg University, the Netherlands.

Kennard, R.W., and Stone, L.A. (1969), "Computer Aided Design of Experiments," *Technometrics*, 11, 137-148.

Kinzer, G.R. (1985), "Application of Two-Cubed Factorial Designs to Process Studies," in *Experiments in Industry: Design, Analysis, and Interpretation of Results*, eds. R.D. Snee, L.B. Hare, and J.R. Trout, Milwaukee: American Society for Quality Control.

Koons, G.F., and Wilt, M.H. (1985), "Design and Analysis of an ABS Pipe Compound Experiment," in *Experiments in Industry: Design, Analysis, and Interpretation of Results*, eds. R.D. Snee, L.B. Hare, and J.R. Trout, Milwaukee: American Society for Quality Control.

Leone, F.C., Nelson, L.S., and Johnson, N.L. (1968), "Sampling Distributions of Variance Components: Empirical Studies of Unbalanced Nested Designs," *Technometrics*, 10, 719-726.

Mitchell, T.J. (1974), "An Algorithm for the Construction of D-optimal Experimental Designs," *Technometrics*, 16, 203-210.

Myers, G.C. Jr. (1985), "Use of Response Surface Methodology in Clinical Chemistry," in *Experiments in Industry: Design, Analysis, and Interpretation of Results*, eds. R.D. Snee, L.B. Hare, and J.R. Trout, Milwaukee: American Society for Quality Control.

Nelson, W.B., and Kielpinski, T.J. (1975), "Optimum Accelerated Lifetests for the Normal and Lognormal Life Distributions," *IEEE Transactions on Reliability*, 24, 310-320.

Neyman, J., Iwaszkiewicz, K., and Kotodziejczyk, S. (1935), "Statistical Problems in Agricultural Experimentation," *Journal of the Royal Statistical Society Series B*, 2, 107-154.

Pearson, E.S. (1934), "Sampling Problems in Industry," *Journal of the Royal Statistical Society Series B*, 1, 107-136.

Pignatiello, J.J., and Ramberg, J.S. (1985), "Discussion of 'Offline Quality Control, Parameter Design, and the Taguchi Method'," *Journal of Quality Technology*, 17, 198-206.

Plackett, R.L., and Burman, J.P. (1946), "Design of Optimal Multifactorial Experiments," *Biometrika*, 33, 305-325.

Raktoe, B.L., Heydayat, A., and Federer, W.T. (1981), *Factorial Designs*, New York: Wiley.

Ross, D.T. (1977), "Structured Analysis (SA): A Language for Communicating Ideas," *IEEE Transactions on Software Engineering*, SE-3, 16-34.

Sacks, J., Schiller, S.B., and Welch, W.J. (1989), "Designs for Computer Experiments," *Technometrics*, 31, 41-47.

Satterthwaite, F.E. (1959), "Random Balance Experimentation," *Technometrics*, 1, 111-137.

Scheffé, H. (1958), "Experiments with Mixtures," *Journal of the Royal Statistical Society Series B*, 20, 344-360.

Schroer, B.J., and Tseng, F.T. (1987), "Modeling complex manufacturing systems using simulation," *Proceedings of the 1987 Winter Simulation Conference*, A. Thesen, H. Grant, W. David Kelton (eds.), 677-682.

Snee, R.D. (1981), "Developing Blending Models for Gasoline and Other Mixtures," *Technometrics*, 23, 119-130.

Snee, R.D. (1985a), "Experimenting with a large Number of Variables," in *Experiments in Industry: Design, Analysis, and Interpretation of Results*, eds. R.D. Snee, L.B. Hare, and J.R. Trout, Milwaukee: American Society for Quality Control.

Snee, R.D. (1985b), "Computer-Aided Design of Experiments-- Some Practical Experiences," *Journal of Quality Technology*, 17, 222-236.

Taguchi, G., and Wu, Y. (1980), *Introduction to Off-Line Quality Control*, The Central Japanese Quality Control Association (available from American Supplier Institute, 32100 Detroit Industrial Expressway, Romulus, MI 48174.

Villars, D.S. (1951), *Statistical Design and Analysis of Experiments for Development Research*, San Francisco: Wm.C. Brown.

Welch, W.J. (1984), "Computer-Aided Design of Experiments for Response Estimation," *Technometrics*, 26, 217-224.

Youden, W.J. (1962), "Systematic Errors in Physical Constants," *Technometrics*, 4, 111-123.

Youden, W.J. (1972), "Enduring Values," *Technometrics*, 14, 1-10.

Young, A., Moore, M., and Girard, T. (1987), "A Warped Disk," OR516: Case Studies, Project Report, School of Operations Research and Industrial Engineering, Cornell University.

WORKING PAPERS
IN THE
DEPARTMENT OF INDUSTRIAL & MANAGEMENT SYSTEMS ENGINEERING

Below is a listing of current working papers from January 1990 to Present. Prior listings from January 1987 through December 1989 are available upon request.

90-101 Kim, S.H. and K. Knott. "The Effect of Standard Limits and Fits on the Productivity of Assembly Robots." (January 1990).

90-102 Goldberg, J.H., V. Parthasarathy, and D.J. Murphy. "Orientation Perception in Farm Tractor Overturn: A Modeling and Simulation Approach." (January 1990).

90-103 Kim, Y.W., M.W. Lee, and A. Freivalds. "EYES - An Expert System for the Development of High Touch Consumer Electronic Products." (January 1990).

90-104 Davidoff, N. and A. Freivalds. "Computer-Aided Design Modeling of the Human Hand." (January 1990).

90-105 Triantaphyllou, E., S.R.T. Kumara, and A.L. Soyster. "Generating Logical Expressions from Positive and Negative Examples via a Branch-and-Bound Approach." (February 1990).

90-106 Triantaphyllou, E., S.R.T. Kumara, and A.L. Soyster. "On the Equivalence of Rule-Based Systems." (February 1990).

90-107 Kumara, S.R.T. and J.D. Lee. "Qualitative Reasoning in Continuous Process Diagnostics and Control." (March 1990).

90-108 Harmonosky, C.M. and S.F. Robohn. "Real-Time Scheduling in Computer Integrated Manufacturing: A Review of Recent Research." (March 1990).

90-109 Harmonosky, C.M. and J.S. Smith. "A Scheduling Heuristic for Work Station Control with Explicit Material Handling Considerations." (March 1990).

90-110 Chen, J.M. and B.J. Melloy. "Determining An Optimal Policy for WIP Storage Allocation Problem in Serial Production Lines." (May 1990).

90-111 Shin, Y.C. and C.A. Betts. "Control of Chips in the Turning of 4150 Steel by Using an Obstruction Type Chip Breaker." (May 1990).

90-112 Shin, Y.C. and Y.S. Joo. "Optimization of Machining Conditions with Practical Constraints." (May 1990).

90-113 Elanayar V.T., S., Y.C. Shin and S. Kumara. "Machining Condition Monitoring for Automation Using Neural Networks." (May 1990).

90-114 Shin, Y.C. and K.W. Wang. "Design of an Optimal Damper to Minimize the Vibration of Machine Tool Structures Subject to Random Excitation." (May 1990).

90-115 Cavalier, T.M. and B.J. Melloy. "A Mathematical Programming Solution to the Euclidean Regression Model." (May 1990).

90-116 Joshi, S., R.A. Wysk and A. Jones. "A Scaleable Architecture for CIM Shop Floor Control." (May 1990).

90-117 Ventura, J.A. and C-H Wu. "Computer Vision Inspection of Elliptical Profiles." (May 1990).

90-118 Ventura, J.A., F.F. Chen and C-H Wu. "Grouping Parts and Tools in Flexible Manufacturing Systems Production Planning." (May 1990).

90-119 Shih, S.C. and I. Ham. "An Integrated Expert Database Systems Approach to Intelligent Manufacturing Information Retrieval for Effective CIM Implementation." (June 1990).

90-120 Melloy, B.J. "Determining the Optimal Process Mean and Screening Limits for Packages Subject to Compliance Testing." (June 1990).

90-121 Soyster, A.L. and C. Qi. "Another Approach to the Optic Star Problem." (July 1990).

90-122 Cheraghi, S.H., E.A. Lehtihet and P.J. Egbelu, "Development of a Vision Assisted Optimal Part-to-Pad Placement Technique for Printed Circuit Board Assembly." (July 1990).

90-123 Melloy, B.J., J.I. McCool and M. Rosenshine, "Production Lot Sizing for Master Components." (July 1990).

90-124 Conway, Jr., J.C., P.H. Cohen, B.R. Tittmann, W.R. Drawl, M. Meloncelli, D.F. Poeth and R. Kropiewnicki, "Evaluation of Adhesive Strength of CVD Polycrystalline Diamond Coatings Deposited on Sialon Substrates." (July 1990).

90-125 Cohen, P.H., S. Joshi, S. Hsieh, M.S. Lu, R. Goddard, and R. Stimpson, "Attribute-Based Group Technology System." (August 1990)

90-126 Smith, J.S. and P.J. Egbelu, "Offline Robot Programming Using Natural Language Interface." (August 1990).

90-127 Seo, Yoonho and P.J. Egbelu, "Configuration and Operation of A Pull Type Flexible Manufacturing System." (August 1990).

90-128 Egbelu, P.J., "Cost Reduction in Multi-State Manufacturing System with Unit Load Handling." (August 1990).

90-129 Kamarthi, S.V., S.R.T. Kumara, F.T.S. Yu, and I. Ham, "Neural Networks and Their Applications in Component Design Data Retrieval." (August 1990).

90-130 Anand, S., S.V. Kamarthi, P.J. Egbelu, and S.R.T. Kumara, "Intelligent Robotic Spray Painting Using Machine Vision." (September 1990).

90-131 Supinsky, M.R., E.A. Lehtihet, and P.J. Egbelu, "Automatic Generation of Robotic Code for Printed Circuit Board Assembly." (September 1990).

90-132 Lehtihet, E.A. and B.L. Dindelli, "Computer-Aided Simulation of Manufacturing Tolerances." (September 1990).

90-133 Barton, R.R., "Graphical Methods for Experiment Design. A Tutorial." (September 1990).

90-134 Serger, C.M. and A. Freivalds, "How Much Time Do Educators Actually Spend Teaching? (The Measurement of Faculty Productivity)." (September 1990).

90-135 Shin, Y.C., S.J. Oh and C.O. Ruud, "Interrogation of Residual Stresses of Machined Surface by An X-Ray Diffraction Technique." (September 1990).

90-136 Ventura, J. and K. Tofang-Sazi, "An Efficient Switching Strategy for Flexible Workstations." (September 1990).

90-137 Voigt, R.C. and R.L. Rosmait, "Development of As-Cast Steels." (September 1990).

90-138 Lehtihet, El-Amine and S. Joshi, "A Framework for Dynamic Tolerance Control in Discrete Part Manufacturing." (September 1990).

90-139 Harmonosky, C.M., "Simulation for Real-Time Control: Advantages, Potential Pitfalls, Opportunities." (September 1990).

90-140 Harmonosky, C.M., "Implementation Issues Using Simulation for Real-Time Scheduling, Control, and Monitoring." (September 1990).

90-141 Lacksonen, T.A. and C.M. Harmonosky, "Evaluation of Different Simulation Approaches for Modeling a Custom Build Assembly Line." (September 1990).

90-142 Kozaczek, K., C.O. Ruud, J.C. Conway, Jr. and C.J. Yu, "Development of a System for In-Process Control of Material Properties." (September 1990).

90-143 Ventura, J.A., "Computational Development of a Lagrangian Dual Approach for Quadratic Networks." (September 1990).

90-144 Hong, J-D., T.M. Cavalier and J.C. Hayya, "On the (t, S_t) Policy in an Integrated Production/Inventory Model with Time-Proportional Demand." (October 1990).

90-145 Voigt, R.C., "Fracture of Cast Irons." (October 1990).

90-146 Schall, S.O. and J. Chandra, "An Analytical Model to Evaluate Alternative Tool Combinations in a Flexible Manufacturing System." (November 1990).

90-147 M'Hallah, Rym and B.J. Melloy, "An Investigation of Process Setting Procedures." (November 1990).

90-148 Cheraghi, S.H., E.A. Lehtihet and P.J. Egbelu, "Development of a Vision Assisted Optimal Part-To-Pad Placement Technique for Printed Circuit Board Assembly." (November 1990).

90-149 Brevick, J.R., J.W. Davis, S. Joshi and S. Puskuri, "Computer Aided Engineering of Tooling for Castings." (November 1990).

WORKING PAPERS IN THE DEPARTMENT OF INDUSTRIAL & MANAGEMENT SYSTEMS ENGINEERING

Below is a listing of current working papers from January 1989 to Present. Prior listings from January 1986 through December 1988 are available upon request.

89-101 Brevick, J. R. and J. Davis. "Thermal Analysis for the Control of Thin Section Gray Iron Microstructures." (January 1989).

89-102 Micalizzi, J. and J. H. Goldberg. "Decision-Making in Visual Inspection: Effects of Knowledge of Results." (March 1989).

89-103 Egbelu, P. J. "Lot Sizing with Backorder Consideration in a Multi-Stage Manufacturing System with Sequence Dependent Setup and Processing Costs. (March 1989).

89-104 Egbelu, P. J. "Establishing of Economic Production Rate, Production Batch Size, and Production Sequence in Manufacturing Systems with Flexible Routing." (March 1989).

89-105 Egbelu, P. J. "Framework for Dynamic Positioning of Storage/Retrieval Machines in an Automated Storage/ Retrieval System." (March 1989).

89-106 Geril, N., P. J. Egbelu and R. A. Wysk. "Manufacturing Simulation for Precision Optical Fabrication of Large Mirror: an end-to-end Process Simulation." (March 1989).

89-107 Tonkay, G. L. and K. Knott. "Intelligent Process Specification for Robotic Arc Welding." (March 1989).

89-108 Gunasena, N. U. and E. A. Lehtihet. "Automatic Program Generation for PC-based Sequence Control Problems." (March 1989).

89-109 Dietz, D. C. and C. M. Harmonosky. "Application of a Control Variate Technique to Simulation Analysis of Aircraft Sortie Generation." (April 1989).

89-110 Shin, Y. C. "An Improved Model for Machining Simulation in Face Milling Operation." (April 1989).

89-111 Shin, Y. C. "Bearing Non-Linearity and Stability Analysis in High Speed Machining." (April 1989).

89-112 Chandra, M. J. and B. J. Melloy. "The Use of Screening Limits to Minimize the Effect of Measurement Error on Outgoing Quality." (April 1989).

89-113 Tothero, G. and C. M. Harmonosky. "A Multi-Factor Plant Layout Methodology." (April 1989).

89-114 Traband, M. T. and D. J. Medeiros. "Part Measurement Techniques Using Point Sampling: Current Status and Future Directions." (April 1989).

89-115 Yang, B. N. and R. A. Wysk. "A Detection Procedure for Resolution of System Deadlocks in Real Time Control of Flexible Manufacturing Systems." (April 1989).

89-116 Goldberg, J. H. and S. A. O'Rourke. "Prediction of Skill Retention and Retraining from Initial Training." (May 1989).

89-117 Shin, Y. C. "Recognition of Cutting State Via Compensated Piezoelectric Force Dynamometer." (May 1989).

89-118 Shin, Y. C. "System Identification of Multivariate Systems with Feedback." (May 1989)

89-119 Freivalds, A. "Quantifying the European Lifting Guidelines." (May 1989).

89-120 Nann, S. R., A. Ray and S. Kumara. "A Decision Support System for Real-Time Control and Monitoring of Dynamical Processes." (May 1989).

89-121 Kumara, S., I. Ham, M. Al-Hamando and K. Goodnow. "Causal Reasoning and Data Abstraction in Component Design." (May 1989).

89-122 Triantaphyllou, E. and R. T. Kumara. "Learning Rules from Examples in Rule-Based Systems Via an Integer Programming Approach." (May 1989).

89-123 Gunasena, U., S. R. T. Kumara and A. L. Soyster. "A Knowledge Based System for Course Scheduling." (May 1989).

89-124 Goldberg, J. H., V. Parthasarathy and D. J. Murphy. "Farm Tractor Rollover and Pitchover Simulator for Modeling Instability Perception and Response." (May 1989).

89-125 Gadea, N. A. S. and M. J. Chandra. "A Generalized Deterministic Inventory Model Without Shortages." (May 1989).

89-126 Shin, Y. C. "Static and Dynamic Characteristics of a Two Stage Pilot Relief Valve." (May 1989).

89-127 Chandra, M. J. and S. Palkovics. "Optimal Design of a Pre-Control Chart." (June 1989).

89-128 Ruud, C. O., J. C. Conway, K. Kozaczek and C-J. Yu. "Development of a System for Nondestructive In-Process Characterization of Copper Alloy Strip." (June 1989).

89-129 Goldberg, J. H. and V. Parthasarathy. "Specifications and Design Computations for a Farm Tractor Motion-Based Simulator." (June 1989).

89-130 Jung, E. S. and A. Freivalds. "Development of an Expert System for Designing Workplace in Manual Materials Handling Jobs." (July 1989).

89-131 Goldberg, J. H. and K. A. Latorella. "An Integrated Visual Search and Memory Retrieval Model of Inspection." (July 1989).

89-132 Ruud, C. O. "Residual Stresses and Their Measurement." (July 1989).

89-133 Melloy, B. J. "Determining the Optimal Process Mean and Screening Limits for a Canning Problem Under the Assumption of Measurement Error." (July 1989).

CANCELED

89-134 Cavalier, T. M., P. M. Pardalos and A. L. Soyster. "Modeling and Integer Programming Techniques Applied to Propositional Calculus." (July 1989).

89-135 Sagar, K. V., S. R. T. Kumara, F. T. S. Yu and I. Ham. "Neural Networks and Their Applications in Component Design Data Retrieval." (July 1989).

89-136 Mettala, E. G., S. B. Joshi and R. A. Wysk. "Computer Integrated Manufacturing Control Software Generation." (August 1989).

89-137 Anand, S., K. Knott, J. Rubinowitz and M. Firth. "A Practical Method for Structuring the Bill of Materials for the Shoe and the Garment Industry - A Case Study." (August 1989).

89-138 Brevick, J. R. and J. Davis. " Thermal Analysis for the Control of Thin Section Gray Iron Microstructures." (August 1989).

89-139 Kajita, T. and J. W. Davis. "A Statistical Analysis of Gate Size for Sound Casting in Investment Casting Process." (August 1989).

89-140 Witter, J. A. and J.W. Davis. "Impregnation for Vacuum Sealed Castings." (August 1989).

89-141 Lin, C-W. R. and R. A. Wysk. "Generating Alternative Computer Aided Process Plans." (September 1989).

89-142 Fellows, G. L. and A. Freivalds. "The Use of Force Sensing Resistors in Ergonomic Tool Design." (September 1989).

89-143 Freivalds, A. "Current Research Thrusts in Ergonomics." (September 1989).

89-144 Goldberg, J. H. and K. A. Latorella. "Processing of Multi-Attribute Defects in Visual Inspection." (October 1989).

89-145 Ventura, J. A. "Algorithms for Quadratic Transportation Networks." (October 1989).

89-146 Cohen, P. H., B. J. Melloy and R. A. Wysk. "The Impact of the SME Foundation on Manufacturing Engineering Education at Penn State." (October 1989).

89-147 Egbelu, P. J. "Batch Production with Unit Load Design and Scheduling Consideration." (October 1989).

89-148 Egbelu, P. J. "Batch Production Time in a Multi-State System with Material Handling Consideration." (October 1989).

89-149 Ventura, J. A. and K. Tofang-Sazi. "Models of Production Lines with Flexible Workstations." (October 1989).

89-150 Mittal, R. O., P. C. Cohen, and B. J. Gilmore. "Dynamic Modeling of the Fixture-Workpiece System." (November 1989).

89-151 Goldberg, J. H., . Parthasarathy, and D. J. Murphy. "A Farm Tractor Overturn Simulator." (November 1989).

89-152 Shin, Y. C., K. W. Wang, and C. H. Chen. "Dynamic Analysis and Modeling of a High Speed Spindle System." (November 1989).

89-153 Brevick, J. R. and J. Davis. "Applying Delta T Measurements to Gray Iron Inoculation. (November 1989).

89-154 Davis, J. and J. R. Brevick. "Effects of Sand and Carbon Fiber Additions on Plaster Mold and Casting Properties." (November 1989).

89-155 Medeiros, D. J. and M. T. Traband. "Tolerance Estimation for Linear Features: A Statistical Approach." (December 1989).

89-156 Medeiros, D. J. and B. Emamizadeh. "Optimal Container Location in Miniload AS/R Systems." (December 1989).

89-157 Chandra, M. J. and C. Harmonosky. "Economic Aspects of Advanced Production and Manufacturing Systems." (December 1989).

89-158 Lehtihet, E. A., U. N. Gunasena, submitted by Mr. R. Weills "Statistical Models for the Relationship Between Production Errors and the Position Tolerance of a Hole." (December 1989).

WORKING PAPERS
IN THE
DEPARTMENT OF INDUSTRIAL & MANAGEMENT SYSTEMS ENGINEERING

Below is a listing of current working papers from January 1988 to Present. Prior listings from January 1985 to January 1988 are available upon request.

88-102 Goncalves, E. V. and I. Ham. "Pattern Recognition Approach to Group Technology Implementation." (January 1988).

88-103 Egbelu, P. J. and A. Lehtihet. "Operation Routing With Lot Sizing Consideration in a Manufacturing System." (January 1988).

88-104 Goldberg, J. H. and J. W. Alred. "Prediction of Physical Workload in Reduced Gravity." (January 1988).

88-105 Freivalds, A. and J. H. Goldberg. "A Methodology For Assigning Variable Relaxation Allowances: Manual Work and Environmental Conditions." (January 1988).

88-106 Goldberg, J. H. and A. Freivalds. "A Methodology for Assigning Variable Relaxation Allowances: Visual Strain, Illumination, and Mental Strain." (January 1988).

88-107 Goetz, W. G., Jr., and P. J. Egbelu. "Guide Path Design and Location of Load Pickup/Dropoff Points for Automated Guided Vehicle System." (February 1988).

88-108 Watson, E. F. and P. J. Egbelu. "Scheduling and Machining of Jobs Through Parallel Nonidentical Machine Cells." (February 1988).

88-109 Jung, K. H. and T. M. Cavalier. "Locating a Traveling M/G/1 Server Using Euclidean Distances to a Finite Set of Demand Points." (February 1988).

88-110 Manivannan, S., A. Lehtihet and P. J. Egbelu. "A Knowledge-Based System for the Specification of Manufacturing Tolerances (ROSCAT)." (February 1988).

88-111 Harmonosky, C. M. and R. P. Sadowski. "The System Generator for FMS Research and Analysis." (February 1988).

88-112 Harmonosky, C. M. "Methodology of Determining a Classification Scheme for Automated Manufacturing Systems." (April 1988).

88-113 Melloy, B. J., A. L. Soyster and M. J. Chandra. "Buffer Allocation in a Flow Line with a Common Buffer." (April 1988). This paper was revised and renamed "An Observation on Storage Allocation in a Flow with A Common Buffer," (December 1988).

88-114 Cohen, P. H. and B. Bidanda. "Automated Fixture Selection for Rotational Parts." (May 1988).

88-115 Mackowiak, R. M., P. H. Cohen, R. A. Wysk and C. Goss. "Development of a Group Technology Workstation." (May 1988).

88-116 Lin, R. C., E. A. Lehtihet. "Microcomputer Aided Tolerance Chart for Squareness and Concentricity." (May 1988).

88-117 Hong, J. D. and M. Jeya Chandra. "A Heuristic Algorithm for an Airlift Crew Allocation Problem." (May 1988).

88-118 Rubinovitz, J. and R. A. Wysk. "Task Planning and Optimization for Robotic Arc Welding -- An Algorithmic Approach. (May 1988).

88-119 Cavalier, T. M. and K. H. Jung. "A Fixed Priority Queue Minisum Location in Euclidean Space. (May 1988).

88-120 Melloy, B. J. and T. M. Cavalier. "A Procedure for Bounding the Probability of Failure Resulting from Stress/Strength Interference. (May 1988).

88-121 Irani, S. A., R. O. Mittal and E. A. Lehtihet. "Optimal Tolerance Charting Within a CAPP Framework." (May 1988).

88-122 Lehtihet, E. A. and B. A. Dindelli. "TOLCON: Microcomputer Based Module for Simulation of Tolerances." (June 1988).

88-123 Mettala, E. G. and P. J. Egbelu. "Alternative Approaches to Sequencing Robot Moves for PCB Assembly." (June 1988).

88-124 Schall, S. O. and M. J. Chandra. "Optimal Tool Inspection Intervals Using a Process Control Approach." (June 1988).

88-125 Manivannan, S., C. D. Pegden and P. J. Egbelu. "Recent Trends Manufacturing Simulation Systems and an Expert Systems Approach for Modeling Just-In-Time Systems." (June 1988).

88-126 Goldberg, J. H. and V. Parthasarathy. "Farm Tractor Overturn Recognition and Response: Operator Limitations." (June 1988).

88-127 Liao, C. J., C. D. Pegden and M. Rosenshine. "Planning Timely Arrivals to a Stochastic Production or Service System." (June 1988).

88-128 Lehtihet, E. A. and U. N. Gunasena. "Models for the Influence of Production errors on the Position Tolerance of a Single Hole Regardless of Feature Size." (June 1988).

88-129 Lehtihet, E. A. and U. N. Gunasena. "Models for the Position and Size Tolerance of a Single Hole." (June 1988).

88-130 Mettala, E. G. and P. J. Egbelu. "Sequencing Robot Moves for PCB Assembly." (June 1988).

88-131 Harmonosky, C. M. and D. C. Barrick. "Simulation in a CIM Environment: Structure for Analysis and Real-Time Control." (July 1988).

88-132 Freivalds, A. and J. H. Goldberg. "Instructional Use of Personal Computers in Human Factors Labs." (July 1988).

88-133 Goldberg, J. H. and M. J. Telek. "Prediction of Visual Inspection Times from Predetermined Time Systems and Visual Search Models." (July 1988).

88-134 Irani, S. A., R. O. Mittal and E. A. Lehtihet. "Tolerance Chart Optimization." (July 1988).

88-135 Goldberg, J. H. and B. J. Melloy. "Courseware for Statistical Inference." (July 1988).

88-136 Micalizi, J. and J. H. Goldberg. "Knowledge of Results in Visual Inspection: Implications for Training." (September 1988).

88-137 Medeiros, D. J. and E. E. Enscore. "Analysis of Operation Rules for WIP Miniload Systems." (September 1988).

88-138 Goldberg, J. H. and S. A. O'Rourke. "Prediction of Skill Retention and Retraining from Initial Skill Training." (September 1988).

88-139 Davis, D. A. and D. J. Medeiros. "A Comparison of In-Transit AGV Routing Policies." (September 1988).

88-140 Sanvido, V., S. Kumara and I. Ham. "A Top Down Approach to Integrating the Building Process." (September 1988).

88-141 Egbelu, P. J. "Route Selection and Flow Control in a Multi-Stage Manufacturing System With Heterogeneous Machines Within Stage." (September 1988).

88-142 Egbelu, P.J. "Machining and Material Flow System Design for Minimum Cost Production." (September 1988).

88-143 Egbelu, P. J. "Reduction of Manufacturing Lead Time Through Selection of Machining Rate and Unit Handling Size." (September 1988).

88-144 Gunasena, U., S. R. T. Kumara and A. L. Soyster. "A Knowledge Based System for Course Scheduling." (October 1988).

88-145 Fox, J. J., S. R. T. Kumara, S. S. Iyengar and J. Rubinovitz. "HAL: A New Robotic Programming Language." (October 1988).

88-146 Goldberg, J. H. "The Development of Visual Search Models as a Basis for Predicting Visual Inspection Time and Accuracy." (October 1988).

88-147 Kim, S. H. and K. Knott. "An Analysis on the Effect of Position Tolerance on the Performance of a Robot for a Loose Fastener Assembly." (October 1988).

88-148 Cohen, P. H. and R. O. Mittal. "A Methodology for Fixturing and Machining of Prismatic Components." (October 1988).

88-149 Tosirisuk, P. and M. Jeya Chandra. "Multiple Finite Source Queueing Model with Dynamic Priority Scheduling." (October 1988).

88-150 Egbelu, P. J. "Integration of Operation Routing and Economic Production Quantity Decisions in Batch Manufacturing System." (October 1988).

88-151 Freivalds and Y. J. Kim. "Blade Size and Weight Effects in Shovel Design." (December 1988).

ARMY DESIGN OF EXPERIMENTS CONFERENCE – MAILING LIST

Gerald R. Andersen
Army Research Office (ARO)
P.O. Box 12211
RTP, NC 27709-02211

Miguel Andriolo
Ballistic Research Lab.
ATTN: SLCBR-LF-T
Aberdeen Proving Ground,
MD 21005-5066
(301) 278-3978

George Anitole
U.S. Army Belvoir Research
Development and Eng. Ctr.

Sanjay Arora
Howard University

Dr. John C. Bailar
Office of Disease Prevention
and Health Promotion
Dept. of Health and
Human Services

William E. Baker
USABRL
Aberdeen Proving Grounds,
MD 21005
(301) 278-6658

Russell Barton
Dept of Industrial and
Management Systems Eng.
207 Hammond Bldg.
Penn State University
State College, PA 16802

Carl B. Bates
USACCAA
8120 Woodmont Avenue
Bethesda, MD 20814-2797
(301) 295-0163

Dr. Robert Bechhofer
College of Engineering
Cornell University
Ithaca, NY 14853
(607) 255-9138

Dr. Barnard H. Bissinger
Hershey Foods Corp.
281 W. Main Street
Middletown, PA 17057
(717) 944-0649

Abel J. Blanco
U.S. Army Atmospheric Sciences Lab
Atmospheric Effects Division
Attn: SLACAS-AE-F
White Sands Missile Range,
NM 88002-5501
(505) 678-3924

Dr. Barry A. Bodt
BRL
Aberdeen Proving Grounds,
MD 21005
(301) 278-6659

John R. Boyarski
Navy Ships Parts Control Ctr.
Code 041
5450 Carlisle Pike
P.O. Box 2020
Mechanicsburg, PA 17055-0788
(717) 790-4882

Dennis L. Brandon
USAE Waterways Experiment Station
3909 Halls Ferry Road
Vicksburg, MS 39180-6199
(601) 634-2807

Ann E. M. Brodeen
U.S. Army BRL
Aberdeen Proving Ground,
MD 21005-5066
(301) 278-8947

Mel Brown
Army Research Office
Research Triangle Park,
NC 27709-2211
(919) 549-0641

Richard M. Brugger
U.S. Army Armament
Munitions and Chemical Command
Rock Island, IL 61299-6000
(309) 782-6419

Dr. Marshall Brunden
Biomathematics – The UpJohn Co.
24170 Oak Lane
Mattawan, MI 49071
(616) 385-7957

Robert Burge
Walter Reed
7134 Smooth Path
Columbia, MD 21045
(301) 596-2815

Rina S. Carter
BRL
SLCBR-SE-P
Aberdeen Proving Grounds,
MD 21005-5066
(301) 278-6646

Dr. Jagdish Chandra
Mathematical and Computer Scienc
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park,
NC 27709-2211
(919) 549-0641 X254

Y. T. Chou
U.S. Army Engineer Waterways
Experiment Station
Vicksburg, MS 39180
(601) 634-3423

Harry E. Clark
315 Riblett Lane
Wilmington, Delaware 19808
(302) 994-0712

Dr. Aivars Clemins
BRL
Aberdeen Proving Ground,
MD 21005-5066
(301) 278-6986

Joseph Collins
U.S. Army BRL
SLCBR-SE-P
Aberdeen Proving Ground,
MD 21005-5066
(301) 278-6832

W. J. Conover
Texas Tech University
College of Business Adm.
Lubbock, TX 79409
(806) 742-1546

Paul Cooper
DynCorp
Test and Eval. Services Div.
Electromagnetic Environmental
Test Facility, P.O. Box 0850
Sierra Vista, AZ 85636-0850

Terence M. Cronin
U.S. Army CECOM
Center for Signals Warfare

Dr. Francis Dressel
Army Research Office (ARO)
P.O. Box 12211
Research Triangle Park,
NC 27707
(919) 549-0641

Eugene Dutoit
U.S. Army Infantry School
ATSH-CDC-O
Ft. Benning, GA 31905-5400
(404) 545-3165/3166

Preceding Page Blank

Dr. James N. Elele
US Army Electronic Proving Ground
STEEP-CT-E
Fort Huachuca, AZ 85613

Douglas H. Frank
I.U.P.
208 Stright Hall
I.U.P. Indiana, PA 15705
(412) 357-3793

Dr. Fred Grimes
TEXCOM Combined Arms Test Ctr.
ATCT-AEH
Fort Hood, Texas 76544-5065
(817) 288-9614

Dr. Robert G. Hinkle
Office, Deputy Under
Secretary of the Army
(Operations Research)
Room 1E643, Pentagon
Washington, DC 20310-0102
(202) 697-1175

Larry Hotchkiss
Academic Computing Support
University of Delaware
Newark, DE 19716
(302) 451-1989

Richard M. Jones
Philip Morris Research Center
P.O. Box 20583
Richmond, VA 23261
(804) 274-3450

A. A. Khan
U.S. Army Concepts Analysis Agency
8120 Woodmont Ave.
Bethesda, MD 20814-270
(301) 295-5566

Elizabeth A. Laurie
US Army Ballistic Research Lab
Aberdeen Proving Ground, MD 21005
(301) 278-6646

William T. Matthews
U.S. Army Materials Technology

Donald M. Neal
U.S. Army Materials Technology Lab

Dr. Steven Fast
CDR USAEPG
Ft Huachuca, AZ 85613-7110

Jitu Ganju
Dept. of Mathematical Sciences
University of Delaware
Newark, DE 19716
(303) 451-6628

Jock Grynovicki
U.S. Army Human Engineering Lab

Joyce Hiros
Test and Evaluation Command
Aberdeen Proving Ground,
MD 21005-5055

Ronald L. Johnson
U.S. Army Belvoir Research
Development and Engineering Ctr.

Huey Ju
Dept. of Mathematical Sciences
University of Delaware

Rickey A. Kolb
Dept. of Mathematical Sciences
West Point, NY 10996
(914) 938-3200

Parvatha Manikkalingan
Howard University

Devan Mehrotra
Dept. of Mathematical Sciences
University of Delaware
Newark, DE 19716
(302) 451-6628

Duc Minh Nguyen
C2NVEO
AMSEL-RD-NV-ISP
Fort Belvoir, VA 22060-5677
(703) 664-5635

Sangita Fatnani
Dept of Mathematical Sciences
University of Delaware
Newark, DE 19716
(302) 451-6628

Robert Gorman
Compu-Val Investments
1702 Lovering Avenue
Wilmington, DE 19806
(302) 652-6767

Joe Harmon
USA Airborne Board

Arthur Hoerl
Dept. of Mathematical Sciences
University of Delaware
Newark, DE 19716

William P. Johnson
Ballistic Research Laboratory
Aberdeen Proving Ground,
MD 21005
(301) 278-6673

Robert J. Kelly
Bendix Communications Division

Mark Kuebler
Dept. of Mathematical Sciences
University of Delaware
Newark, DE 19716
(302) 451-8067

Dr. Thomas Mathew
Department of Mathematics
University of Maryland
Baltimore County
Baltimore, MD 21228
(301) 455-2418

Linda L. C. Moss
Ballistic Research Laboratory
SLCBR-SE-P
Aberdeen Proving Ground,
MD 21005-5066
(301) 278-6658

Erik V. Nordheim
University of Wisconsin

Emanuel Parzen
Texas A & M University
Department of Statistics
College Station, TX 77843-3143
(409) 845-3188

John Riemenschneider
TEXCOM Combined Arms Test Center

Carl Russell
Operational Test & Eval Agency

Dr. J. Schuenemeyer
Dept. of Mathematical Sciences
University of Delaware
Newark, DE 19716
(302) 451-1883

Dr. Ivar Stakgold
Dept. of Mathematical Sciences
University of Delaware
Newark, DE 19716

Douglas B. Tang
Walter Reed Army
Inst. of Research

Deloris Testerman
Yuma Proving Ground
(STEYP-MT-MR)
Yuma, Arizona 85365-9103
(602) 328-6749

Dr. Henry B. Tingey
Dept. of Mathematical Sciences
University of Delaware
Newark, DE 19716
(302) 451-8034

Albert VanHorn
USATACOM, ATTN: AMNSTA-VSM
US. Army Tank-Automotive Command
Warren, MI 48397-5000
(313) 574-8701

Larry West
Test and Evaluation Command
Aberdeen Proving Ground,
MD 21005-5055

David Payne
1525 Meadow Crest Court
Aberdeen, MD 21001
(301) 278-6867

Dr. Lydia Rejto
Dept. of Mathematical Sciences
University of Delaware
Newark, DE 19716
(301) 451-1888

Dr. Richard S. Sacher
Academic Computing Support
University of Delaware
Newark, DE 18716
(302) 451-1466

Dr. Bimal K. Sinha
Dept. of Math & Stat
University of Maryland
Baltimore County
Baltimore, MD 21228
(301) 455-2347

Dr. Doug Stalker
Department of Philosophy
University of Delaware
Newark, DE 19716
(302) 451-8208

Dr. Howard Taylor
Dept. of Mathematical Sciences
University of Delaware
Newark, DE 19716
(302 451-1874

Jerry Thomas
Ballistic Research Lab
Aberdeen Proving Ground,
MD 21005
(301) 278-6728

Paul A. Tukey
Statistical Research Group
Bell Communication Research

Todd Walton
WESCD
P.O. Box 631
Vicksburg, MS 39180

Sarah Wilson
Operational Test & Evaluation Agency
ATTN: CSTE-CCA
4501 Ford Avenue
Alexandria, VA 22302-1458
(703) 756-2166

Michael Prather
Test and Evaluation Command
Aberdeen Proving Ground,
MD 21005-5055

David Roselle, President
University of Delaware
Newark, DE 19716

Dr. Jayaram Sethuraman
Department of Statistics, B-167
Florida State University
Tallahassee, FL 32306
(904) 644-2010

Jill H. Smith
USA Ballistic Research Lab
SLCBR-BL-V
Aberdeen Proving Ground,
MD 21005-5066
(301) 278-6714

Dr. Robert M. Stark
Dept. of Mathematical Sciences
University of Delaware
Newark, DE 19716
(302) 451-2846

Dr. Malcolm S. Taylor
USA Ballistic Resch. Lab.
Aberdeen Proving Ground, MD 2
(301) 278-6638

Andrew Thompson
BRL, SLCBR-SE-P
Aberdeen Proving Grounds,
MD 21005
(301) 278-6832

Mark G. Vangel
U.S. Army Materials Technology

David W. Webb
Ballistic Research Laboratory
ATTN: SLCBR-SE-P
Aberdeen Proving Ground,
MD 21005-5066
(301) 278-6646

Wendy A. Winner
USA BRL ATTN: SLCBR-VL-V
Aberdeen Proving Grounds,
MD 21005-5066
(301) 278-6655

Micha Yadin
Israel Institute of Technology

Prof. Shelemyahu Zacks
Dept. of Mathematical Sciences
State University of New York
at Binghamton
Binghamton, NY 13902-6000
(607) 777-6035

Lorraine Zank
Dept. of Mathematical Sciences
University of Delaware
Newark, DE 19716
(302) 451-6628

Xiaosha Zhang
Dept. of Mathematical Sciences
University of Delaware
Newark, DE 19716
(302) 451-8067

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188
Exp. Date Jun 30, 1986

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release: distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ARO REPORT 91-2	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Army Research Office 6b. OFFICE SYMBOL (If applicable) SLCRO-ARO	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) P.O. Box 12211 Research Triangle Park, NC 27709	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AMSC on behalf of ASO (ADA) 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	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-Sixth Conference on the Design of Experiments in Army Research, Development and Testing

12. PERSONAL AUTHOR(S)

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM Jan 90 TO Feb 91	1991 August	384

16. SUPPLEMENTARY NOTATION

17. COSATI 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-Sixth 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 ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ 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 SLCO-MA

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

END