AD-776 910

PROCEEDINGS OF THE CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING (18TH) HELD AT ABERDEEN PROVING GROUND, MARYLAND ON 25-27 OCTOBER 1972

Army Research Office
Durham, North Carolina

October 1973

U. S. Army Research Office

Report No. 73-2
October 1973

PROCEEDINGS OF THE EIGHTEENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS

Sponsored by the Army Mathematics Steering Committee

HOST

U. S. Army Test and Evaluation Command

Aberdeen Proving Ground, Maryland

25-27 October 1972

U. S. Army Research Office
Box CM, Duke Station
Durham, North Carolina

FOREWOPD

The Eighteenth Conference on the Design of Experiments in Army Research, Development and Testing was held 25-27 October 1972 at the Aberdeen Proving Ground, Maryland. The U. S. Army Test and Evaluation Command served as its host. This is the second conference in this series to be held at the Proving Ground. The first one, called the Sixth Conference on the Design of Experiments, was held in October 1960 with the Ballistic Research Laboratories serving as the host. The father of these meetings, Professor S. S. Wilks, was in charge of arranging the program, and the undersigned served as the Chairman on Local Arrangements. Having served once in this capacity, one has a better appreciation of the work-load faced this year by the Local Chairman, Mr. Gerard T. Dobrindt. Let me thank Mr. Dobrindt for his excellent handling of the many problems with the physical arrangements as well as the problems presented by the attendees. Thanks are also due to Dr. William McIntosh for his guidance and assistance in many phases of the on-base arrangements.

Professor John Tukey, the first invited speaker, got the conference off to an excellent start with his interesting and informative treatment of the topic "Exploratory Data Analysis". He was followed on the program by one of his colleagues at Princeton University, Progessor G. S. Watson. Dr. Watson discussed some recent developments in the interesting field of "Orientation Analysis". At the Second General Session members of the audience had the pleasure of hearing Professor J. S. Hunter discuss one of his papers on "Sequential Factorial Estimation" and Professor G. E. P. Box present some of his work on "Forecasting and Control". It is interesting to note that both Drs. Box and Hunter served on a panel discussion entitled "Common Pitfalls in the Design and Analysis of Experiments" at the Sixth Design Conference. The fifth invited speaker was Professor Raymond H. Myers who enlightened members of the audience on some recent and important developments in the field of "Dual Response Surface Analysis". The recipient of this years Samuel S. Wilks Memorial Awards was one of the above-mentioned invited speakers, namely Dr. G. E. P. Box. We are pleased to be able to include in these proceedings his acceptance remarks.

The Army Mathematics Steering Committee sponsors these conferences on behalf of the Chief of Research and Development. Members of this committee would like to thank the many Army scientists who contributed to the success of this meeting. Without their dedicated efforts these meetings would not be repeated year after year. Scientists in other government agencies have also lent their talents to the programs. This year we were pleased to have three contributed papers presented by members of the National Bureau of Standards, and also to have Dr. Churchill Eisenhart of the Bureau serve as a member of the Program Committee. The Food and Drug Administration was represented on the agenda by Dr. Clifford Maloney. His services as a member of the Program Committee were also appreciated. One or two Canadian scientists

usually attend and contribute to the discussion at these meetings. This year we were pleased to have one of our Canadian friends, Mr. G. J. McLaughlin, of the Defense Research Establishment Valcartier, present one of the clinical papers.

In addition to the two members of my Program Committee already mentioned, the following individuals served: Robert Bechhofer, Norman Coleman, Gerard Dobrindt, Francis Dressel, Walter Foster, Boyd Harshbarger, William McIntosh, Herbert Soloman, Grace Wahba, and Geoffrey Watson. Those gentlemen and one lady were charged with the responsibility of outlining the general character of the conference, and to select the invited speakers. My thanks to them for preforming this task in a fashion that again led to a successful scientific conference. It seems in order at this time to give special mention to Dr. Walter D. Foster, who is the Chairman of the AMSC Subcommittee on Probability and Statistics. In this capacity Dr. Foster can be looked upon as the one generally responsible for initiating the advanced and overall planning for each conference. He serves in the conduction of many other phases of these meetings; in particular, he serves as the Chairman of the committee that organizes the final form of the agenda. He makes the report to the AMSC on some of the accomplishments of each Design of Experiment Conference. On behalf of all attendees at these meetings, let me express our thanks to Dr. Foster for his dedicated efforts to these scientific conferences.

Finally, we desire to express our sincerest appreciation to Dr. Francis G. Dressel whose many significant contributions make the Army Design of Experiments Conferences a success from year to year.

Frank E. Grubbs
Conference Chairman

TABLE OF CONTENTS

EIGHTEENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS

IN ARMY RESEARCH, DEVELOPMENT, AND TESTING

25-27 October 1972

HQ, U. S. Army Test and Evaluation Command

* * * * * Wednesday, 25 October * * * * *

0830-0900 REGISTRATION - Library Conference Room. Building 330

0900-1130 GENERAL SESSION I - Library Conference Room

CALLING OF CONFERENCE TO ORDER
Gerard Dobrindt, Chairman for Local Arrangements, US Army Test and Evaluation Command

WELCOME TO THE U. S. ARMY TEST AND EVALUATION COMMAND
Benjamin S. Goodwin, Chief Engineer, U. S. Army Test and Evaluation Command

CHAIRMAN OF SESSION I
Dr. Fred Frishman, Physical Sciences & Engineering Division, Office, Chief of Research & Development, Washington, D. C.

EXPLORATORY DATA ANALYSIS
Professor John Tukey, Department of Mathematics, Princeton University, Princeton, New Jersey

ORIENTATION ANALYSIS
Professor G. S. Watson, Princeton University, Department of Statistics, Fine Hall, Princeton, New Jersey

1130-1300 LUNCH - Officers' Open Mess, APG, MD

1300-1445 CLINICAL SESSION A - Library Conference Room, Bldg. 330

CHAIRMAN:
Norman P. Coleman, Jr., HQ, US Army Weapons Command, Rock Island, Illinois

PANELISTS:
A. Clifford Cohen, Institute of Statistics, University of Georgia, Athens, Georgia

Bernard Harris, Mathematics Research Center, The University of Wisconsin, Madison, Wisconsin

Boyd Harshbarger, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia

Herbert Solomon, Department of Statistics, George Washington University, Washington, D. C.

FREQUENCY RESPONSE ON THE M50 ROCKET DUE TO WIND
Bernard F. Engebos, Atmospheric Sciences Laboratory, U. S. Army Electronic Command, White Sands Missile Range, New Mexico

EXPERIMENTAL DESIGNS WITH LARGE NUMBERS OF VARIABLES
Roger L. Brauer and Charles C. Lozar, Architecture Branch, Special Projects Division, U. S. Army Construction Engineering Research Laboratory, Champaign, Illinois

1300-1445 TECHNICAL SESSION 1 - Conference Room B, Building 314

CHAIRMAN:
Jerome R. Johnson, US Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland

ORTHOGONAL ESTIMATES IN WEIGHING DESIGNS
William G. Lese, Jr., US Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland

WEIGHING DESIGNS FOR MASS CALIBRATION
J. M. Cameron and R. C. Raybold, Office of Measurement Services, Institute for Basic Standards, US Department of Commerce, National Bureau of Standards, Washington, D. C.

COMPUTER CONSTRUCTION OF CYCLIC BALANCED INCOMPLETE BLOCK DESIGNS
Malcolm S. Taylor, US Army Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

1300-1445 TECHNICAL SESSION 2 - Conference Room A, Building 314

CHAIRMAN:
Vernon V. Visnaw, Materiel Testing Directorate, Aberdeen Proving Ground, Maryland

WEDNESDAY

REGRESSION ANALYSIS APPROACH TO INTERPOLATION ALGORITHMS
E. L. McDowell, Structural Concepts Branch, Construction Systems Division, US Army Construction Engineering Research Laboratory, Champaign, Illinois

ON SPURIOUS CORRELATIONS FOR PARTIALLY RELATED VARIATES
Oscar M. Essenwanger, Physical Science Directorate, Directorate for Research, Development, Engineering and Missile Systems Laboratory, US Army Missile Command, Redstone Arsenal, Alabama

THE LEAST SQUARES ANALYSIS OF DATA GENERATED BY A "PIECE-WISE" GENERAL LINEAR MODEL
Robert L. Launer, Procurement Research Office, US Army Logistics Management Center, Fort Lee, Virginia

1445-1515 BREAK

1515-1700 CLINICAL SESSION B - Library Conference Room, Building 330

CHAIRMAN:
Paul C. Cox, White Sands Missile Range, New Mexico

PANELISTS:
A. Clifford Cohen, Institute of Statistics, University of Georgia, Athens, Georgia

Bernard Harris, Mathematics Research Center, The University of Wisconsin, Madison, Wisconsin

Boyd Harshbarger, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia

Herbert Solomon, Department of Statistics, George Washington University, Washington, D. C.

EXPERIMENTAL ESTABLISHMENT OF ACCURACY OF RANGE-TO-FUNCTION MEASUREMENT FOR ARTILLERY PROJECTILES
1LT L. Dave Clements, Data Reduction Section, Yuma Proving Ground, Yuma, Arizona

AN IMPROVED METHOD OF ESTIMATING THE CRITICAL VELOCITY OF A PROJECTILE IN PENETRATION BALLISTICS
G. J. McLaughlin, Defence Research Establishment Valcartier, Courcelette, P. Q., Canada

WEDNESDAY

1515-1700 TECHNICAL SESSION 3 - Conference Room A, Building 314

CHAIRMAN:
COL L. Ponder, US Army Test and Evaluation Command, Aberdeen Proving Ground, Maryland

EVALUATING AND SCHEDULING PROTOTYPE REQUIREMENTS FOR SUITABILITY TESTING
Majors Richard B. Cole and William J. Owen, US Army Infantry Board, Fort Benning, Georgia

STOPPING RULES FOR SCHEDULING WITH PARTICULAR REFERENCE TO MISSILE RANGE SCHEDULING
Paul H. Randolph, New Mexico State University, Representing-Instrumentation Directorate, White Sands Missile Range, New Mexico

1515-1700 TECHNICAL SESSION 4 - Conference Room B, Building 314

CHAIRMAN:
John S. Hagan, Materiel Testing Directorate, Aberdeen Proving Ground, Maryland

A ROBUST CONFIDENCE INTERVAL FOR LOCATION
Alan M. Gross, Princeton University, Department of Statistics, Fine Hall, Princeton, New Jersey

APPROXIMATE CONFIDENCE LIMITS FOR $P(X \leq Y)$
J. R. Moore and M. S. Taylor, US Army Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

STATISTICAL EVALUATION OF FLIGHT TEST PERFORMANCE OF THE HELICOPTER LIFT MARGIN SYSTEM (HLMS)
Erwin Biser and Ronald Kurowsky, Avionics Laboratory, US Army Electronics Command, Fort Monmouth, New Jersey

1830- SOCIAL HOUR FOLLOWED BY THE BANQUET. PRESENTATION OF THE SAMUEL S. WILKS MEMORIAL AWARD
Dr. Frank E. Grubbs, Chairman of the Conference

* * * * * Thursday, 26 October * * * * *

0900-1015 TECHNICAL SESSION 5 - Library Conference Room, Building 330

CHAIRMAN:
Royce W. Soanes, Jr., Benet R&E Laboratory, Watervliet Arsenal, Watervliet, New York

THURSDAY

AUTOMATED RADAR DATA PROCESSING AT WHITE SNADS MISSILE RANGE FEATURING ADAPTIVE FILTERING WITH BIAS ESTIMATION
W. A. McCool, Analysis and Computation Division, White Sands Missile Range, New Mexico

ALGORITHM FOR EDITING BIVARIATE DATA FILES WITH RANDOM SPACING IN THE INDEPENDENT VARIABLE
1LT L. Dave Clements, Data Reduction Section, Yuma Proving Ground, Yuma, Arizona

0900-1015 TECHNICAL SESSION 6 - Conference Room B, Building 314

CHAIRMAN:
SP/4 Ray Peterson, Frankford Arsenal, Philadelphia, Pennsylvania

STATISTICAL ANALYSIS OF H. F. OBLIQUE AND VERTICAL INCIDENCE INOSPHERIC DATA APPLICABLE TO FIELD ARMY DISTANCES
Richard J. D'Accardi, US Army Electronics Command, Fort Monmouth, New Jersey
Chris P. Tsokos, University of South Florida, Tampa, Florida

COMPARISON OF THE TRANSMISSION THROUGH FOG OF THE 3-5 AND 8-12 MICRON SPECTRAL REGIONS AS A FUNCTION OF THE VISIBLE TRANSMISSION
James E. Perry and Stuart Layman, Night Vision Laboratory USAECOM, Fort Belvoir, Virginia

0900-1015 TECHNICAL SESSION 7 - Conference Room A, Building 314

CHAIRMAN:
Col. George T. Morris, Jr., US Army Test and Evaluation Command, Aberdeen Proving Ground, Maryland

MAXIMUM LIKELIHOOD ESTIMATION PROCEDURES IN RELIABILITY GROWTH
Larry H. Crow, US Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland

MODIFIED PROPAGATION OF ERRORS WITH APPLICATIONS TO MAINTAINABILITY AND AVAILABILITY
Paul C. Cox, White Sands Missile Range, White Sands, New Mexico

1015-1045 BREAK

10145-1130 TECHNICAL SESSION 8 - Conference Room A, Building 314

THURSDAY

CHAIRMAN:
George L. Kinnett, HQ, US Army Aviation Materiel Laboratories, Fort Eustis, Virginia

WIND TUNNEL MODIFICATION AND EVALUATION
E. G. Peterson, C. E. Sperry and E. Covert, Deseret Test Center, Building 100, Soldiers' Circle, Fort Douglas, Utah

1045-1130 TECHNICAL SESSION 9 - Library Conference Room, Building 330

CHAIRMAN:
Edward Fiske, Product Assurance Director, Edgewood Arsenal, Maryland

TECHNIQUES FOR TAIL LENGTH ANALYSIS
James J. Filliben, Statistical Engineering Laboratory, Institute for Basic Standards, National Bureau of Standards, Washington, D. C.

CRITERIA FOR A BIOCELLULAR MODEL - BIOCELLULAR COMMUNICATION
George I. Lavin, Vulnerability Laboratory, BRL, ARDC, Aberdeen Proving Ground, Maryland

1045-1130 TECHNICAL SESSION 10 - Conference Room B, Building 314

CHAIRMAN:
Boyd Harshbarger, Virginia Polytechnic Institute and State University, Blacksburg, Virginia

EQUATION-OF-STATE AND SHOCK INITIATION EXPERIMENT ON EXPLOSIVES USING PULSED ELECTRON BEAMS
L. Avrami and P. Harris, Picatinny Arsenal, Dover, New Jersey
J. Shea, Physics International Company

AN ANALYSIS OF 5.56MM ALUMINUM CARTRIDGE CASE "BURN-THROUGH" PHENOMENON
Walter H. Squire and Reed E. Donnard, Frankford Arsenal, Philadelphia, Pennsylvania

1130-1300 LUNCH - Officers' Open Mess, APG

1300-1415 TECHNICAL SESSION 11 - Library Conference Room, Building 330

CHAIRMAN:
Harry Greveris, US Army MUCOM, Frankford Arsenal, Philadelphia, Pennsylvania

THURSDAY

STATISTICAL MODELING OF PROPAGATION LOSS DATA
M. Acker, R. D'Accardi, D. Dense, Communications/ADP Laboratory, US Army Electronics Command, Fort Monmouth, New Jersey
C. Tsokos, Department of Mathematics and Statistics, University of South Florida, Tampa, Florida

1300-1415 TECHNICAL SESSION 12 - Conference Room A, Building 314

CHAIRMAN:
A. Clifford Cohen, University of Georgia, Athens, Georgia

GRUBBS' ESTIMATORS TO DATE
Clifford J. Maloney, Bureau of Biologics, Food & Drug Administration, 5600 Fishers Lane, Rockville, Maryland

A SYSTEM FOR POSITION-LOCATION BASED ON RANGES
Richard H. F. Jackson, James A. Lechner, and David J. Sookne, US Department of Commerce, National Bureau of Standards, Washington, D. C.

1300-1415 TECHNICAL SESSION 13 - Conference Room B, Building 314

CHAIRMAN:
COL Luke Vavra, Combat Development Command, Aberdeen Proving Ground, Maryland

APPROXIMATE CONFIDENCE LIMITS ON THE CEP: CASE FOR UNEQUAL VARIANCES
E. Inselmann, Headquarters US Army Materiel Command, Washington, D. C.

ON THE VARIATION IN MECHANICAL PROPERTIES OF LARGE CALIBER GUN TUBE FORGINGS
Peter A. Thornton, Benet Weapons Laboratory, Watervliet Arsenal, Watervliet, New York

1415-1445 BREAK

1445-1700 GENERAL SESSION II - Library Conference Room, Building 330

CHAIRMAN:
Professor Herbert Solomon, George Washington, University, Washington, D. C.

THURSDAY

SEQUENTIAL FACTORIAL ESTIMATION
Professor J. Stuart Hunter, Department of Chemical Engineering, Princeton University, Princeton, New Jersey

FORECASTING & CONTROL
Professor G. E. P. Box, Department of Statistics, University of Statistics, University of Wisconsin, Madison, Wisconsin

* * * * * Friday, 27 October * * * * *

0900-0940 TECHNICAL SESSION 14 - Library Conference Room, Building 330

CHAIRMAN:
Alan W. Benton, Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland

A CONTINUOUS SAMPLING PLAN WITH A PROVISION FOR A REDUCED CLEARANCE NUMBER
Gary L. Aasheim, Mathematical Statistician, US Army Ammunition Procurement & Supply Agency, Joliet, Illinois

0900-0940 TECHNICAL SESSION 15 - Conference Room A, Building 314

CHAIRMAN:
Douglas Tang, Walter Reed Army Institute of Research, Washington, D. C.

SELECTION OF THE MOST MEANINGFUL SUBSET OF RESPONSES IN A MULTIPLE RESPONSE EXPERIMENT
Walter D. Foster, Fort Detrick, Frederick, Maryland

0940-1010 BREAK

1010-1130 GENERAL SESSION III - Library Conference Room, Building 330

CHAIRMAN:
Dr. Frank E. Grubbs, Aberdeen Research & Development Center, Aberdeen Proving Ground, Maryland

OPEN MEETING OF THE AMSC SUBCOMMITTEE ON PROBABILITY AND STATISTICS
Dr. Walter D. Foster, Fort Detrick, Frederick, Maryland

DUAL RESPONSE SURFACE ANALYSIS
Professor Raymond H. Myers, Virginia Polytechnic Institute and State University, Blacksburg, Virginia

* * * * * * * * * * * * *

PAPERS BY TITLE

EXPLORATORY DATA ANALYSIS AS PART OF A LARGER WHOLE*

John W. Tukey

Princeton University, Princeton, New Jersey

(Δ) Most data analysis should be investigative

It is not enough to look for what we anticipate. The greatest gains from data come from surprises. We will usually not be very surprised, but we should try to be.

(ΔΔ) Data analysis is well thought of in three phases

As we come to think over the process of analyzing data, when done well, we can hardly fail to identify the unrealism of the descriptions given or implied in our texts and lectures. The description I am about to give emphasizes three kinds of stages. It is more realistic than the description we are accustomed to but we dare not think it (or anything else) the ultimate in realism.

The first stage is exploratory data analysis, which does not need probability, significance, or confidence, and which, when there is much data, may need to handle only either a portion or a sample of what is available. That there is still much to be said and that there are new simple techniques to be developed is testified to by 3 volumes of a book now in a limited preliminary edition (Tukey 1970-1971) which deals only with the simpler questions, leaving multiple regression and related questions for later treatment.

The second stage is probabilistic. Rough confirmatory data analysis asks, perhaps quite crudely: "With what accuracy are the appearances already found to be believed?" Three answers are reasonable:

- The appearances are so poorly defined that they can be forgotten (at least as evidence though probably not as clues).
- The appearances are marginal (so that crude analysis may not suffice and a more careful analysis is called for).
- The appearances are well-determined (when we may, but more often do not, have grounds for a more careful analysis).

Among the key issues of such a second stage are the issues of multiplicity: How many things might have been looked at? How many had a real chance to be looked at? How should the multiplicity decided upon, in answer to these questions, affect the resulting confidence sets and significance levels? These are important questions; their answers can affect what we think the data has shown.

*At the Eighteenth Conference on the Design of Experiments, Professor Tukey issued this outline of his address. It contains several references for those interested in pursuing this topic.

It will only be after we have become used to dealing with the issues of multiplicity that we will be psychologically ready to deal effectively with correlated estimates, to recognize in particular (a) that the higher the correlation the less the chance -- NOT THE GREATER -- of one or more accidental significances and (b) that correlation of fluctuations need imply nothing as to whether the real effects measured by one calculated quantity will in any way "leak" into other calculated quantities. Leakage of fluctuation and leakage of effect NEED NOT go together, though they sometimes do.

When the result of the second stage is marginal, we need a third stage, in which we wish to muster whatever strength the data before us possesses that bears directly on the question at issue -- and in which we often also want to borrow strength from either other aspects of the same body of data or from other bodies of data. It is at this stage of "mustering and borrowing strength" that we require our best statistical techniques. Medians may be quite good enough for our rough confirmatory analysis, but if we have good robust measures of location they are needed in mustering and borrowing strength.

To argue, as we have implicitly done so often in the past, that -- (1) all data requires mustering and borrowing of strength and (2) this can -- nay should -- be done without any exploratory data analysis -- is surely at least one of the minor heights of unrealism. Trying to make what needs to be data investigation into data processing that really meets our needs involves many new ideas, and ideas come slowly.

(ΔΔΔ) Novel ad hoc analyses need not bar us from confirmatory analysis

To be clear that this is so, we must be prepared to face up to two points:

- questions of multiplicity are not going to be avoided.
- approximate confidence and significance procedures are quite good enough.

Once we do, the jackknife[+] will give us adequate confirmatory assessment; our only struggle will be with assessing degrees of multiplicity.

(Waiting for specific statistical theory for specific analyses is unsound. We have to wait too long, and -- what is worse -- we get theory based on too narrow assumptions.)

+ For an introductory account see:

Mosteller, F. and Tukey, J.W. (1968). Data analysis, including statistics. *Handbook of Social Psychology, 2nd edition,* vol. 2. G. Lindzey and E. Aronson, editors, Addison-Wesley, Reading, Massachusetts, 80-203.

SOME PRINCIPLES

THAT SHOULD GUIDE

EXPLORATORY DATA ANALYSIS

(Δ) Walk first, run later

"It is well to understand what you can do before you learn how to measure how well you seem to have done it".

(ΔΔ) Don't wait for running shoes, start now

(ΔΔΔ) Data analysis should be investigative

"Exploratory data analysis is detective work -- numerical detective work -- or counting detective work -- or graphical detective work".

(ΔΔΔΔ) Resistant techniques should be the usual beginning

A technique is resistant if changing a small part of the data will have only small effects on the result, no matter what is done to the small part. (Means are not resistant, but medians are.)

(ΔΔΔΔΔ) Analyses should come before summaries

Before we summarize, we should analyze, and look at the analysis. Here an analysis is:

> a conversion -- usually a breakdown -- of numerical data into other numbers that is both reversible and relevant.

Reversible means that you can get all the data back IN DETAIL from the results of analysis. Relevant means that some, at least, of the results of analysis illuminate each of the questions most likely to arise. We will come to examples a little later.

(ΔΔΔΔΔΔ) In routine analysis the client should be presented with at least two different versions

If the two versions "agree", fine. Let some summary of one be published.

If the two versions "disagree", the client must think -- very painful, but forcing this may be the best thing the statistician can do.

(ΔΔΔΔΔΔΔ) **"Looking at the data" implies both MORE NUMBERS and BETTER PICTURES**

The unexpected is best brought to our attention by pictures. Failing this, as is always to some extent necessary, more numbers can and do help.

(ΔΔΔΔΔΔΔΔ) **Implicitly defined -- and hence iteratively calculated -- analyses are inevitable**

We have been frightened too long by some mixture of the apparent difficulties of hand calculation and the inaptness of mathematical formulas. Some implicitly defined, iterative calculations are "as easy" and "at least as safe from error" as those that use arithmetic means.

(*********) Not only mathematical statistics, but also data analysis, is going to have to become more like biochemistry. (added March 1972)

The greatest danger of an applied mathematical science is the tacit assumption of

OMNIFERENCE

of the assumption that both users and bystanders, from knowing exactly what is done, will be able to draw --and will, in fact, draw-- the relevant inferences concerning the behavior of every technique at hand. To have direct omniference by as many users as possible about as many techniques as possible is a very good thing. But to avoid a technique, because omniference is hard or impossible, can be very unwise.

We ought all to expect that users are to be told the best information about data analysis techniques that is available --whether or not they can afford the effort to understand how that information was gained-- and whether or not the information is proved, provable, or even subject to confidence statements.

EXPLORATORY DATA ANALYSES

IN RELATION TO

PROBABILITY MODELS

(Δ) Probability models are to give results for guidance.

As statisticians we must take the major share of responsibility here. We ought to make the existence, nature, and details of probability models openly available to all -- encouraging their perusal. But we ought not shove them down throats in the early stages of learning.

Mathematics is the only possible scientific discipline in which responsibility can be completely avoided -- by teaching every student all the proofs, thus making him responsible for the validity of all the mathematics he has thus learned. No one else can avoid responsibility this way. If, as statisticians, we are concerned with the analysis of data, we cannot escape.

Understanding what comes of carefully formulated probability situations is of the essence. Rarely will it be directly and precisely applicable to our problems. Only as we carefully broaden the bases on which it is built will we bring it closer and closer to direct application.

(ΔΔ) In Exploratory Data Analysis, 50% efficiency is plenty

If 50% efficiency will not reveal an effect, 95% will not make it significant.

(ΔΔΔ) Simplicity and flexibility outweigh efficiency

Recall Churchill Eisenhart's definition of the "practical power" of a statistical test:

The product of the probability that the test will be applied and the mathematical power when applied.

(ΔΔΔΔ) As we learn from broader probability models, we will be better guided in Exploratory Data Analysis

TWO-WAY TABLES

OF

RESPONSES

We have analyzed many hundreds of thousands (at least) of two-way tables of responses by fitting something of the form

$$\alpha_i + \beta_j + \text{hash}$$

Almost all of this has been done by explicit arithmetic means.

These are really used as algorithms to meet implicit conditions that certain arithmetic means of residuals and effects are zero.

The results are very NON-resistant, — and hence very NON-robust of efficiency. We can no longer live with them as the only approach.

Implicit medians do quite well, and are not hard to apply.

As an example, let us look at data from page 103 of the Rothamsted Field Experiments of 1969. The analysis by means hardly shows that anything is going on among the residuals. The analysis by implicit medians calls at least two things to our attention.

DATA, weight of sugar beet roots in 0.01 ton

	N1	N3	N5	N7	(Means)
DG	1468	1597	1670	1647	(1596)
ST	980	1237	1379	1449	(1260)
PT	967	1156	1457	1511	(1273)
GM	1304	1411	1470	1444	(1415)
FJ	912	1234	1325	1374	(1211)
FS	963	1234	1351	1367	(1228)
(Means)	(1104)	(1311)	(1442)	(1462)	(1330)

ANALYSIS 1, by explicit or implicit means

DG	94	20	-38	-83	266
ST	-52	14	3	51	-70
PT	-80	-98	72	104	-57
GM	145	15	-57	-105	85
FJ	-73	42	2	29	-119
FS	-39	25	11	2	-102
(eff)	-226	-19	112	132	1330

ANALYSIS 2, by implicit medians

DG	176	22	-21	-67	289
ST	-6	-32	6	41	-17
PT	-53	-147	62	73	17
GM	230	22	-21	-80	103
FJ	-25	6	-5	11	-58
FS	6	-6	7	-11	-41
(eff)	-334	-51	51	84	1337

Note: N1, N3, N5, N7 are four levels of added nitrogen.

OTHER KINDS

OF ANALYSIS

CONSIDERED BRIEFLY

(Δ) When regression is for residuals, as usually in the analysis of covariance, for example, we often need structural regression, rather than predictive regression.

Stripping out an effect can be much more important than minimizing residuals.

(ΔΔ) Almost all applications of spectrum analysis are exploratory in nature

Where spectrum analysis has helped, it has been because of what it has shown to us.

(ΔΔΔ) Numerical classification (numerical taxonomy, cluster analysis, etc.) has been an unrecognized battleground between explanatory and confirmatory data analysis

The techniques of steadily increasing effectiveness pushed onward by W.J. Williams and G.N. Lance are essentially exploratory in nature. The views of N. Jardine and R. Sibson, to pick an antithesis, are basically confirmatory. (Rather than facing the multiplicity problem ((see, for example, Day, N.E. (1969) *Biometrika 56*, 470-473)) the most usual reaction has been one of fear and retreat to axioms and abstract criteria.)

(ΔΔΔΔ) Almost all of multivariate analysis has suffered from an emphasis on confirmatory data analysis, to the concealment of what might have been seen

(The nearly complete book of R. Gnandesikan on multivariate data analysis is a valuable first step forward.) (Canonical analysis, in the sense of M.J.R. Healy et al, is an outstanding example of improved data insight by exploratory methods.)

(ΔΔΔΔΔ) Contingency tables can often be analyzed, not just summarized

It is their analysis that offers an effective foothold for their exploratory data analysis.

(******) Effective local techniques in multivariate analysis seem likely to depend on near-volume indicators. (added March 1972)

Cells and grids are useful in one-dimension from moderate amounts of data up, in two-dimensions from moderately large amounts up, and in three dimensions from quite large amounts of data up. To work in four or more dimensions with any feasible amounts of data, or to work in three or two with lesser amounts of data, we have to do something else. kth nearest volume, for an appropriate shape of neighborhood, seems likely to fill this gap.

EXAMPLES *

OF

BETTER PICTURES

1)	to	4)	Stem-and-leaf
5)	to	7)	Schematic plots
8)	to	9)	Residuals from lines
10)	to	13)	Row-PLUS-column fits
14)			Bar diagrams may need bow legs!
15)	to	19)	Rootograms for amounts or balances
20)	to	22)	Rootograms for counts
23)	to	24)	Rank-size-log plots
25)	to	9)	Counting-in

There needs to be a

GOOD PICTURE

in response to

EVERY

type of question

FREQUENTLY ASKED

* Taken or adapted from John W. Tukey, Exploratory Data Analysis. Limited preliminary edition (3 volumes). Copyright 1970, 1971, Addison-Wesley Publishing Company.

CLOSE

Some would call — out loud, or in their minds — exploratory data analysis "Just descriptive statistics". Those who take this view must believe that "descriptive" statistics is a horrible misnomer. For I hope I have shown that exploratory data analysis is actively incisive rather than passively descriptive, with a real emphasis on the discovery of the unexpected — if necessary by figuratively knocking the analyst's head against the wall until he notices it.

Data analysis should customarily, if not routinely, be investigative. Quantitative detective work has to be a professional responsibility.

Undoubtedly, the swing to exploratory data analysis will go somewhat too far. However:

It is better to ride a damped pendulum

than to be stuck in the mud.

THE STATISTICS OF DIRECTIONS

G. S. Watson
Princeton University

INTRODUCTION. An analysis of directional data is required in many fields of research. The writer's work was stimulated first by studies of the direction of permanent magnetization of ancient rocks (palaeo magnetism -- see Irving (1964)) and then by studies of bird navigation. These yield examples of directions in three and two dimensions respectively. Like so much of statistics, the developement of the required theory and methods owes much to a paper by Fisher (1953). The literature has been recently summarized in a book my Mardia (1972).

The subject not only has practical interest -- it has theoretical interest. The tools of statistics -- means, medians, variances, distribution functions, etc. -- are all fashioned for the real line on which the observations are points. When they are points on a circle or a sphere, one must start afresh -- none of the tools just mentioned make sense any more. Creating a new set, although a simple job, has given the writer more pleasure than any of his other statistical work. Other interest stems from the compact nature of the circle and the sphere which makes things simpler than the line -- but this aspect is not appropriate for further discussion today.

2. Data and its summary descriptions. In two dimensions, a direction may be thought of as an angle, a point on a circle or a unit vector. To display data in angles one could show it on a straight line of length 360° and find the mean, median, variance, etc. But this is surely wrong since it supposes that observations of 1° and 359° are far apart. If all the data is concentrated around 180°, no great harm is done, of course. If it is shown on a circle no such problem arises but new tools are required. If we think of it as

$$\underline{r}_1, \ldots, \underline{r}_N ,$$

a bunch of unit vectors, the new concepts immediately suggest themselves. The direction of the mean vector $\frac{1}{N} \sum_1^N \underline{r}_i$ is suggested as a center of the sample. $\underline{R} = \Sigma \underline{r}_i$ is the vector resultant of the sample and we define the sample mean direction to be a unit vector in the direction of $\underline{R}$. Thus instead of having e.g. the mean of directions 1° and 359° as 180° (!!), it comes out sensibly as 0° = 360°.

The graphical display of directional data is shown in Figure 1 a,b. Figure 1.c shows how undirected lines are plotted — these are called axes and need a different treatment than directions. Figure 1's b and c together indicate the axes of slump folds tend to be parallel with the palaeo current direction.

Given a sample of directions with a single center, how might one describe the scatter or dispersion of a sample of directions? If the bunch of vectors is very tight (i.e. none of $r_1,\ldots,r_N$ make much of an angle with $\underline{R}$) the dispersion is small and R, the length of $\underline{R}$, is almost as large as N. If the $\underline{r}_i$ point in many directions, the dispersion is large and R will be small. Hence

$$\text{Dispersion of sample} = N - R \qquad (1)$$

would be a sensible definition. This needs to be reduced by a factor like N to get it on a per-observation basis so we might define

$$\text{Scatter} = \frac{N - R}{N} = 1 - \frac{R}{N} \qquad (2)$$

It is clear that we have found analogues of $\bar{x}$, $\Sigma(x_i-\bar{x})^2$ and s^2!

If we turn to directions in three dimensions the above arguments and definitions still make sense. To visualize such data, we must look at points on the surface of a sphere. To show them on a two dimensional page, some projection must be used. Different projections are used in different subjects. The Lambert projection projects a hemisphere so that areas are preserved. Thus the density of points is not distorted. Hence it is usually best for statistics. Special paper is available for doing this manually. The point with spherical polar coordinates (θ,ϕ), $0 < \phi < \pi/2$, is made to correspond to a point (ρ,ψ) using planar polars, where $\psi = \phi$ and $\rho = \sqrt{2}C \sin \phi/2$. Thus the upper hemisphere is mapped on a disc of radius C. Figure 2 shows plots of some sets of geological data. Efforts are often made to "contour" the density of points -- a generalization of histogramming. The paper by Watson (1970) gives more details and references and relates all the ways different subjects have used to define a direction in three dimensions.

The definitions of mean direction, dispersion and scatter are only useful for data like that in Figure 2.a. It will be noted that they are related to the center of mass of the points, each of unit mass. By considering the moment of inertia (M. I.) of the set of points we can sort out other configurations e.g. bipolar distributions, girdle distributions. Let the vector $\underline{r}_i$ have components x_i, y_i and z_i so that $x_i^2 + y_i^2 + z_i^2 = 1$. Define

$$M = \sum_{1}^{N} \underline{r}_i \underline{r}_i' = \begin{bmatrix} \Sigma x_i^2 & \Sigma x_i y_i & \Sigma x_i z_i \\ \Sigma y_i x_i & \Sigma y_i^2 & \Sigma y_i z_i \\ \Sigma z_i x_i & \Sigma z_i y_i & \Sigma z_i^2 \end{bmatrix} \qquad (3)$$

The matrix $\underline{M}$ is symmetric and definite. Its eigenvalues, $\lambda_1 > \lambda_2 > \lambda_3$ say, are positive and add to n = trace $\underline{M}$. . They are the stationary values of $\underline{d}'\underline{M}\underline{d}$ where $\underline{d}$ is a unit vector and the eigenvectors are the vectors d yielding these values. The eigenvalues and eigenvectors of $\underline{M}$ may be interpreted. Consider the M. I. of the unit mass at point $\underline{r}_i$ about an axis through the origin parallel to a unit vector $\underline{d}$.

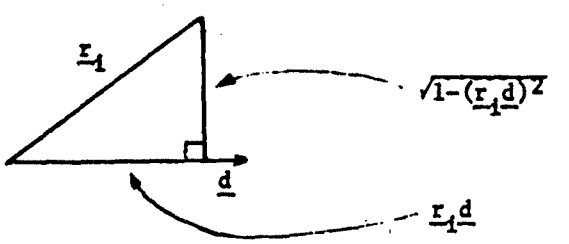


Figure 3

From Figure 3, it is $1 - (\underline{r}_i'\underline{d})^2 = 1 - \underline{d}'\underline{r}_i\underline{r}_i'\underline{d}$ so that the M. I. of all the points is

$$M.I. = N - \sum_{i=1}^{N} \underline{d}'\underline{r}_i\underline{r}_i'\underline{d} = N - \underline{d}'\underline{M}\underline{d}.$$

Suppose now the points are fairly uniformly distributed around a great circle. The direction of greatest M. I. is perpendicular to the great circle. The M. I. is about the same around any orthogonal direction. Thus this distribution corresponds to one small root and two nearly equal larger roots. One large and two small corresponds to a uni- or bi-polar distribution — and these can be distinguished by the length of R (large in one case and small in another). If all three roots are equal, the distribution of points must be uniform on the sphere. And so on.

It is possible to write a simple program to draw Lambert projections of samples and to calculate $\underline{R}$, R , $\underline{M}$ and its eigenvectors and values. These quantities give one a good feel for the data.

3. Parametric distributions and tests. If the observations are symmetrically arranged about a single center with a density falling off as one moves away from the center, one will not go far wrong in assuming

they are a sample drawn a distribution with density proportional to exp k cos θ where $k \geq 0$ is an accuracy parameter and θ is the angle between the center and the observation.

We will now, because of time, stick to three dimensions. There this distribution is called the Fisher distribution. If the center is the unit vector $\underline{\mu}$ and $\underline{r}$ the observation, the density is

$$\frac{k}{4\pi \sinh k} e^{k \underline{r}' \underline{\mu}} \tag{4}$$

With data $\underline{r}_1, \ldots, \underline{r}_N$, it is easily seen that the maximum likelihood (m.l.) estimates of k and $\underline{\mu}$ are

$$\hat{\underline{\mu}} = \text{the direction of } \underline{R} = \underline{R}/R \tag{5}$$

and $\hat{k}$, the solution of

$$\operatorname{Coth} \hat{k} - \frac{1}{\hat{k}k} = \frac{R}{N} \quad . \tag{6}$$

If $\hat{k}$ is greater than 3,

$$\hat{k} \approx \frac{N}{N-R} \tag{7}$$

Now $\hat{k}$ is an "accuracy" parameter, the opposite of scatter, so this matches our intuitive formula (2). This fortifies our belief that (2), (5) and (7) make sense even if (4) is not quite true.

If $\underline{\mu}$ is known, the m.l. estimate of k is

$$N/(N-X), \quad X = \underline{R}' \mu \tag{8}$$

Thus N-X is evidently the dispersion of the sample about μ, just as N-R is the dispersion of the sample about $\hat{\mu}$.

When k = 0, (4) is the uniform distribution. It is often necessary to test whether this is so. For single cluster alternatives, we will naturally reject if R is too large. It is easily shown that

$$\frac{3R^2}{N} \approx \chi_3^2 \tag{9}$$

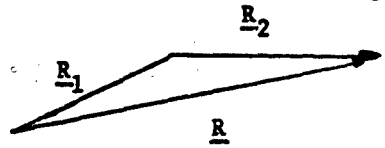
so the test is easy to make. (For more complex alternatives, other tests are appropriate.)

To test whether a sample comes from (4) with a given mean or polar direction, one may consider the analysis of dispersion

so that the test is

$$F_{2,2(N-2)} = \frac{R_1 + R_2 - R}{N - R_1 - R_2} \quad (N - 2) \qquad (17)$$

The logic of (16) is seen from the triangle



It has been shown that these tests are robust against quite severe changes in the parental law of Fisher. The extreme outliers that play havoc on the line cannot occur here.

BIBLIOGRAPHY

Fisher, R. A. (1953) "Dispersion on a Sphere" Proc. Roy. Soc. Lond. A 217 pp 295-305.

Irving E. (1964) "Palaeomagnetism and its applications to Geological and Geophysical problems" John Wiley, New York.

Mardia, K. V. (1972) " Statistics of Directional Data" Academic Press, London - New York.

Watson, G. S. (1970) "Orientation Statistics in the Earth Sciences" Bull. Geol. Instn, Uppsala, N.S. 2:9 pp 73-89.

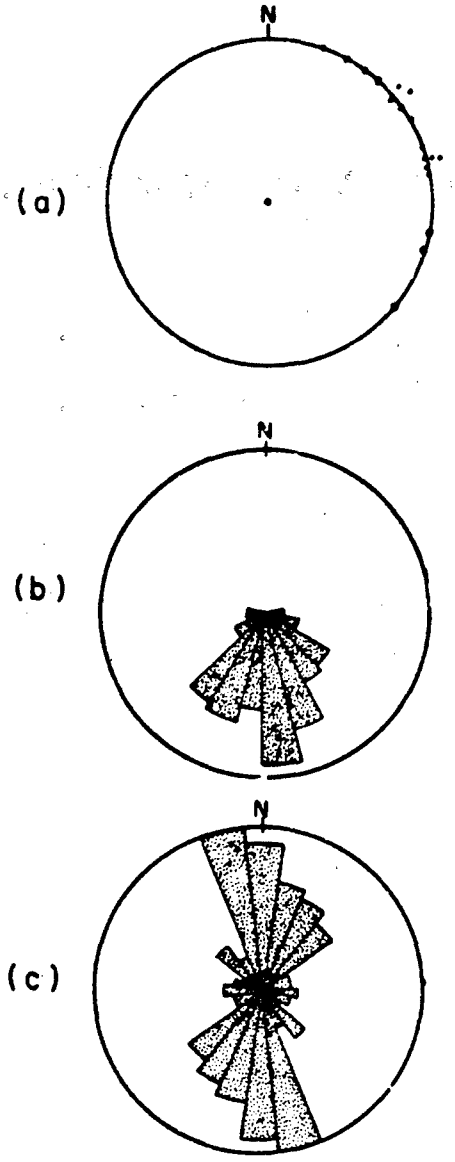


Figure 1 (a) Raw data plotted on circle,

(b) Rose diagram of some palaeocurrent directions,

(c) Rose diagram of some axes of slump folds.

(Redrawn from Potter and Pettijohn(1963))

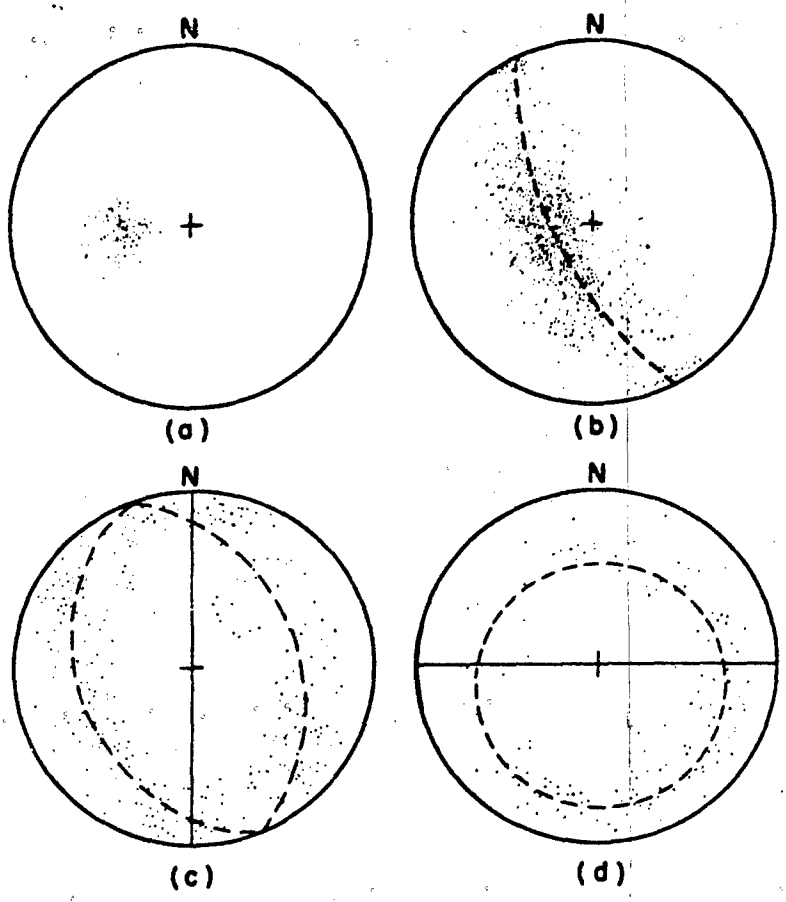


Figure 2: Patterns of preferred orientation (median girdles shown by broken lines). (a) Maximum (symmetric): 150 lineations from Loch Leven, Scottish Highlands. (b) Girdle: 1,000 poles of foliation from Turoka, Kenya. (c) "crossed girdle": 390 [0001] of quartz from quartzite, Barstow, California. (d) Small circle or "cleft" girdle: 140 [0001] of quartz from Orocopia schist, California. (After J. M. Christie; redrawn from Turner and Weiss (1963))

AN INVESTIGATION OF WIND FREQUENCY RESPONSE ON THE M50 ROCKET

Bernard F. Engebos and Abel J. Blanco
Atmospheric Sciences Laboratory, US Army Electronics Command
White Sands Missile Range, New Mexico

ABSTRACT

The wind frequency response for the M50 rocket was studied using quadrant elevation angles of 200, 400, and 800 miles. The wind profile was assumed to be of the form of finite Fourier series with statistically determined amplitudes obtained through a random number generator. Impacts were simulated using a 5-degree-of-freedom trajectory model using 100 randomly generated wind profiles. Correlations between the randomly generated amplitudes and the simulated displacements were then computed. So far the results are inconclusive and improvement is necessary.

INTRODUCTION

The main reason for this study is to delineate the degree of fidelity with which the wind field must be known to achieve acceptable rocketry accuracy. Specifically, how high must a space frequency of wind be before the flight path of the rocket is essentially unaffected by that frequency? An exact answer to this question would involve exhaustive studies into meteorological data collection, data analysis, and the aerodynamics involved in treating wavelengths of wind. Several preliminary studies [1-5] concerning this type problem have been reported. The objective of the study is to find optimal wind layers so that relatively accurate impacts can be achieved. To accomplish this end, a 5-degree-of-freedom ballistic simulation model was used [6].

The Honest John M50 rocket is approximately 8 meters long. The mathematical ballistic model (linear aerodynamics) used to calculate theoretical trajectories assumes a net aerodynamic force acting through the center of pressure of the rocket, which is equivalent to assuming a constant angle of attack over the surface of the rocket, i.e., the wind is invariant along the rocket's length. Thus, it is difficult to speak of the effect of wind oscillations when the wavelength is the same order as the length of the rocket. As a result, the highest space frequency considered corresponds to a wavelength of 16 meters (twice the length of the rocket). For all trajectory simulations considered, all atmospheric and aerodynamic data on the rocket but the wind were held constant.

The remainder of this article was reproduced photographically from the authors' manuscript.

DISCUSSION

The input wind conditions were components of the horizontal wind as functions of the altitude z.

The first approach consisted of letting

$$wx_i(z) = 5 \sum_{j=1}^{m} (A_{ij}\cos\omega_j z + B_{ij}\sin\omega_j z)$$

$$wy_i(z) = 5 \sum_{j=1}^{m} (C_{ij}\cos\omega_j z + D_{ij}\sin\omega_j z)$$

where $wx_i(z)$ is the east-west and $wy_i(z)$ the north-south component of the i-th wind profile (m/sec). ω_j represents m wind space frequencies with wavelengths in intervals of $\frac{1}{16}$ meters up to the burnout altitude of the rocket. It should be noted that m is equal to the greatest integer value of the quotient of the burnout altitude and 16. The various coefficients of the trigonometric functions were generated by a random number generator, assuming a normal distribution of mean zero and standard deviation of one. The multiplier of 5 was chosen to ensure representative wind magnitudes. This yields a finite Fourier series for the wind components.

Associated with the i-th wind profile is the simulated impact point (x_i, y_i). This point was obtained by using the above wind profile during the power on and power off portions of flight. Let (x_o, y_o) be the nowind impact point. Setting

$$Dx_i = x_i - x_o$$

$$Dy_i = y_i - x_o$$

$$TD_i = (Dx_i^2 + Dy_i^2)^{1/2}$$

$$P_{ij} = [(A_{ij}^2 + B_{ij}^2 + C_{ij}^2 + D_{ij}^2)]^{1/2}$$

one then can compute correlation coefficients as follows:

$$\left.\begin{array}{l} Dx_i \text{ vs } A_{ij} \\ Dx_i \text{ vs } B_{ij} \\ Dy_i \text{ vs } C_{ij} \\ Dy_i \text{ vs } D_{ij} \\ TD_i \text{ vs } P_{ij} \end{array}\right\} j=1,2,\ldots,m.$$

Only the latter correlation coefficient is shown, since the others are similar. Figure 1 shows this correlation coefficient versus the wavelength of the space frequency for a quadrant elevation angle of 200 mils, 100 wind profiles, and m equal to 12. One should note here that this correlation coefficient is low in value. This may be caused by too few cases considered and/or by the Doppler effect on the space frequency due to the rocket's changing velocity.

Figures 2 and 3 show similar results for quadrant elevation angles of 400 and 800 mils, respectively.

Another approach involves holding the amplitude of wind constant and varying the space frequency, ω_j; i.e., set

$$wx_j(z) = 5\cos\omega_j z$$

$$wy_j(z) = 5\sin\omega_j z$$

Figure 4 is a plot of total displacement versus the various wavelengths of the space frequency for 200, 400, and 800 mil trajectory simulations. As the wavelengths increase, the total displacement of the simulated impact point from the nominal impact point also increases. This is quite logical since the low frequency wind occurring during the powered portion of the rocket's trajectory (most of the wind weighting effect [7] occurs here) appears more as a trending wind. In the "real world situation," higher frequencies have smaller amplitude and thus can be neglected. Generally frequencies with wavelengths less than

50 meters long can be ignored for the M50 rocket, as can be seen by Figure 4.

Combining several wind space frequencies into a single wind profile seems to bear out the fact that space frequencies are relatively independent of one another in influencing the rocket's flight path.

CONCLUDING REMARKS

Several questions still are unanswered:

(1) Will the statistically derived wind profile technique described herein be successful when the sampling size is increased?

(2) What about wind measurements as a function of horizontal range?

(3) At what altitudes do wind space frequencies most effect the rocket impact accuracy?

(4) Is there a "best" way to determine optimal wind layering to ensure relatively accurate impacts?

A solution to the above questions is extremely difficult. Any possible suggestions on how to solve this overall problem would be greatly appreciated.

LITERATURE CITED

1. Baker and Phillips, 1952, "Loki Wind Response Analysis",Final SIGC Contract DA-36-039 SC52599, North American Instruments,Inc.
2. Grote, H. and Reinhold Marchgraber, 1961, "Wind Dispersion of Free Missiles During the Propulsion Period," U.S. Army Signal Research and Development Laboratory, Fort Monmouth, New Jersey 07703.
3. Matson, Jarem, and Bennet, 1959, "Little John Wind Analysis," Contract DA 36-039 SC78159, Radio Corporation of America.
4. Robinson, 1955, "Response of Rockets to Wind," Douglas Aircraft Corporation, Report SM-18943.
5. Traylor, Larry, 1968,"Preliminary Study of Wind Frequency Response of the Honest John M50 Tactical Rocket," Atmospheric Sciences Laboratory, US Army Electronics Command, White Sands Missile Range, New Mexico 88002.
6. Walter,E.L., 1962,"Six Variable Ballistic Model for a Rocket," MM445, Signal Missile Support Agency, White Sands Missile Range, New Mexico 88002.
7. Duncan, Louis D., and B.F. Engebos, 1970, "A Rapidly Converging Iterative Technique for Computing Wind Compensation Launcher Settings for Unguided Rockets," ARO-D Report 70-1, Transactions of the Fifteenth Conference of Army Mathematicians, Durham, North Carolina.

FIGURE 1

CORRELATION COEFFICIENTS VS WIND PROFILE WAVELENGTHS 200 MILS

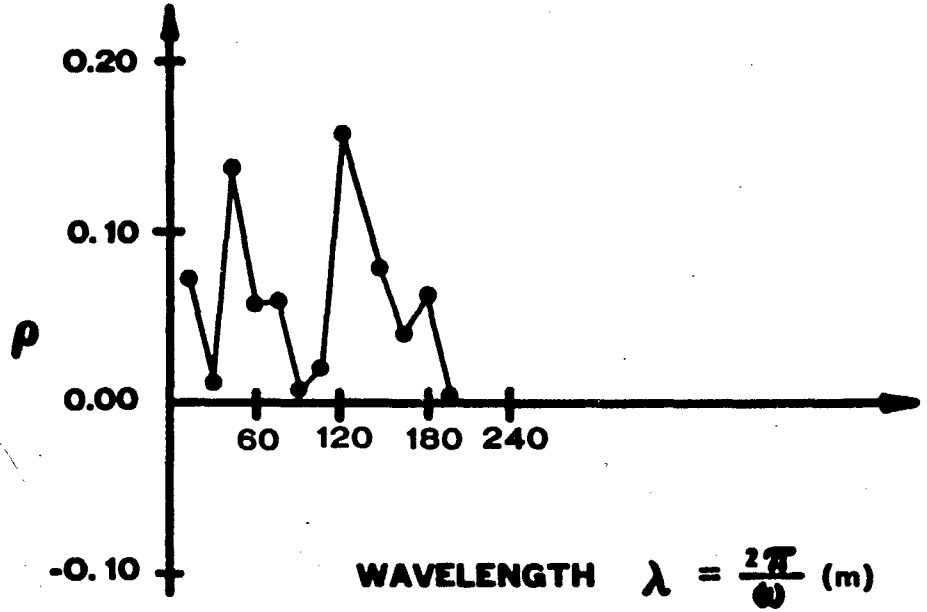
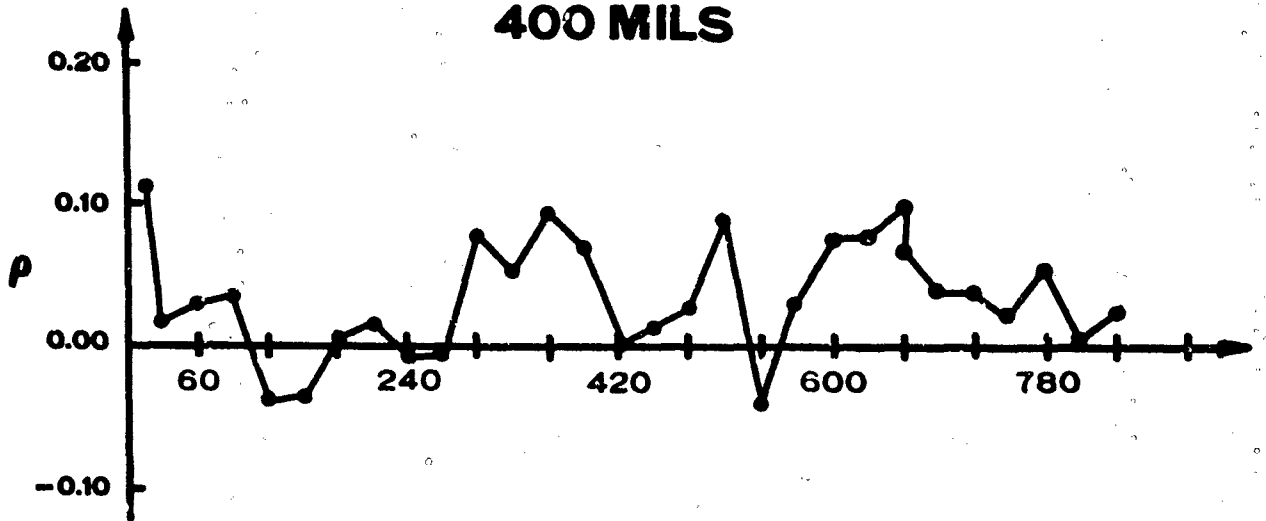


FIGURE 2

CORRELATION COEFFICIENTS
VS
WIND PROFILE WAVELENGTHS

400 MILS



WAVELENGTH $\lambda = \frac{2\pi}{\omega}$ (m)

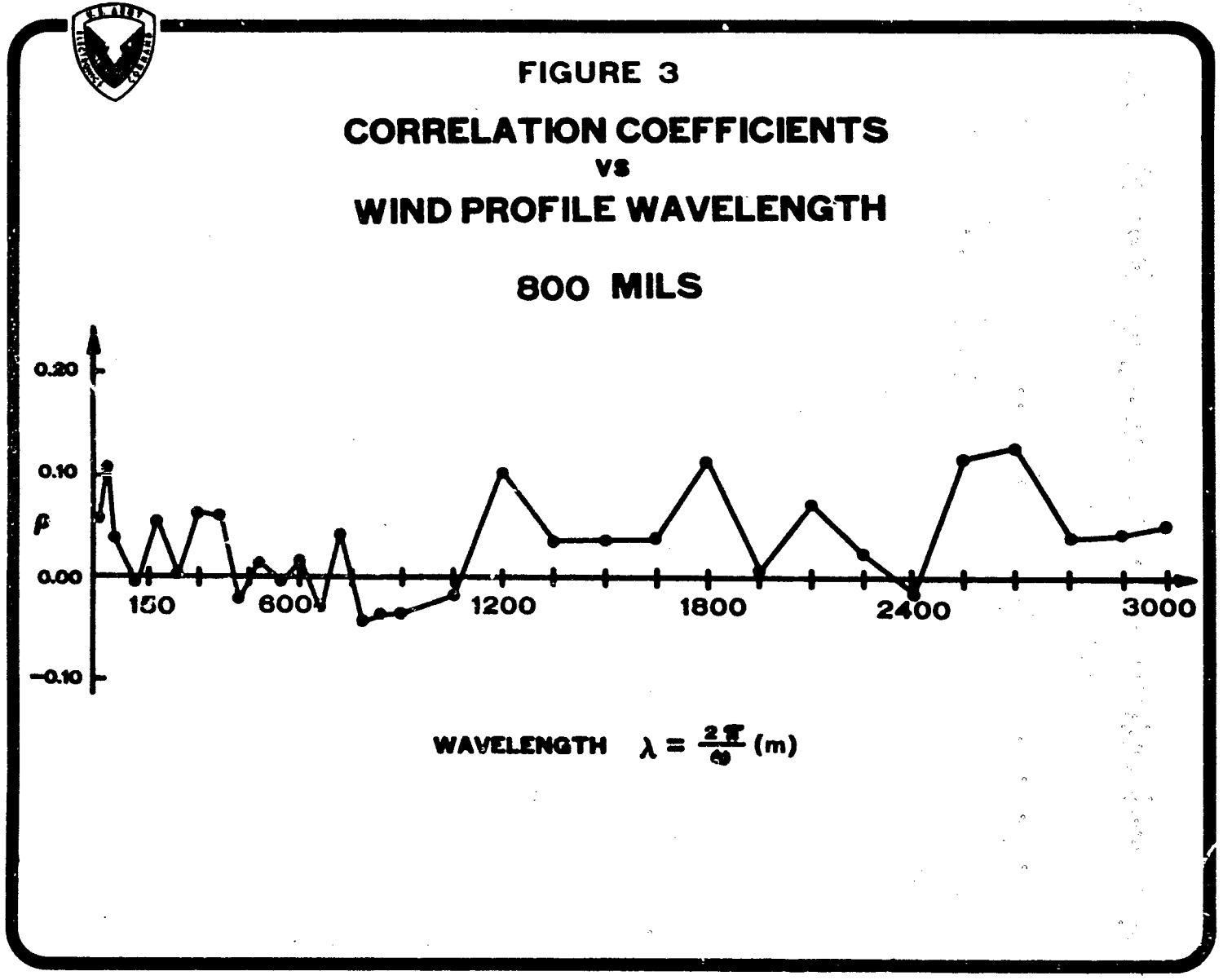


FIGURE 3
CORRELATION COEFFICIENTS
VS
WIND PROFILE WAVELENGTH
800 MILS
0.20
0.10
ρ
0.00
-0.10
150
600
1200
1800
2400
3000
WAVELENGTH λ = 2π/ω (m)

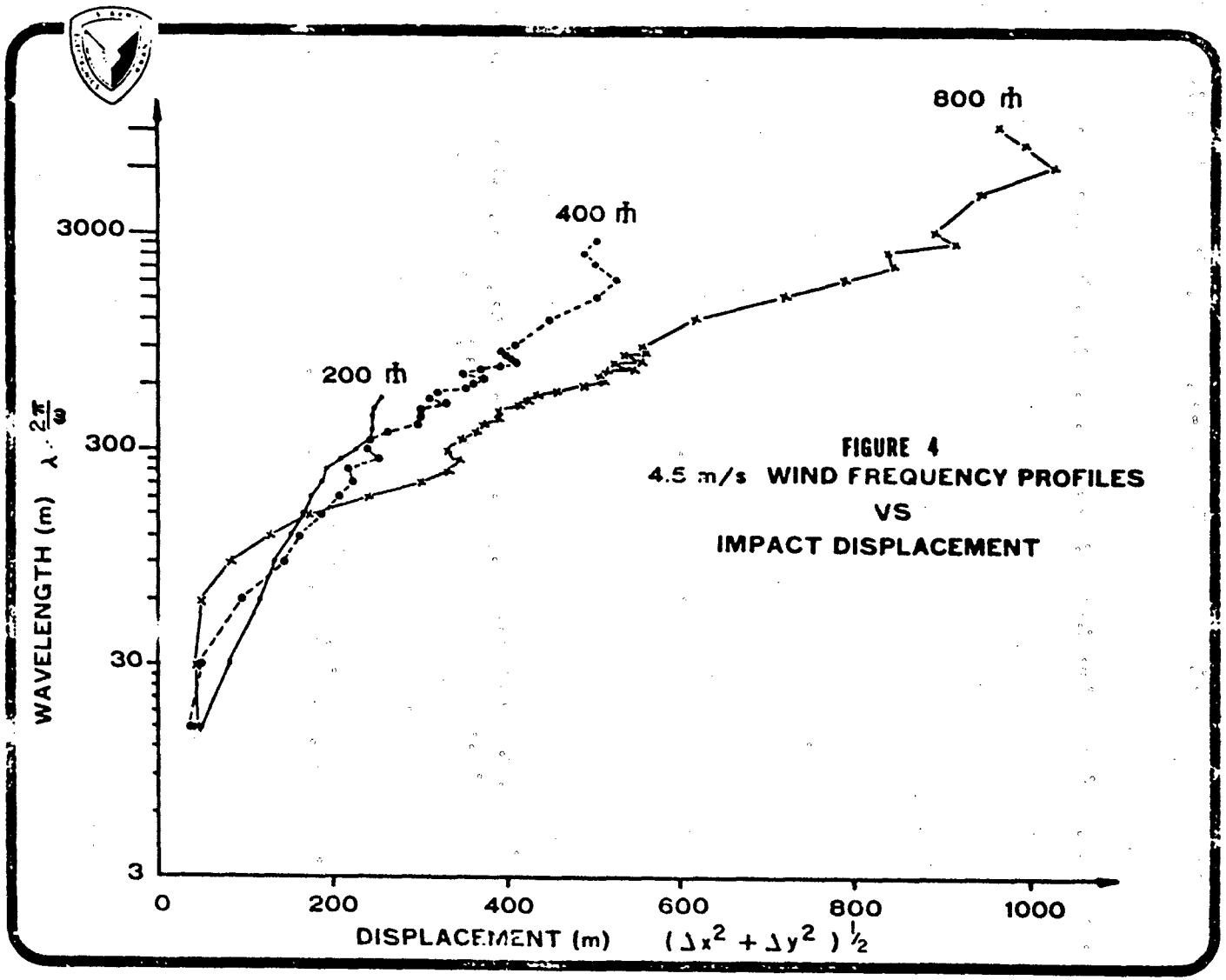


FIGURE 4
4.5 m/s WIND FREQUENCY PROFILES
VS
IMPACT DISPLACEMENT

PROBLEMS IN DESIGNING EXPERIMENTS WITH LARGE NUMBERS OF VARIABLES

Roger L. Brauer and Charles C. Lozar
Special Projects Division Architecture Branch
Construction Engineering Research Laboratory
Champaign, Illinois

The performance of buildings is of primary concern to architects. The designer desires to produce a building for human occupancy that not only achieves high performance of materials of construction but also achieves high quality for the user. The designer wishes to achieve a high degree of user satisfaction and support for user performance.

The problem we wish to present at this clinical session occurs in the evaluation of buildings for users and in measurement of the quality of the constructed space. In completing such evaluations and designing experiments which will measure comparative differences between buildings there is a tremendous number of variables that can contribute to user satisfaction and performance within buildings. We are faced with a problem in measuring and accounting for this wide range of factors and in putting them all together in an experimental design. The results of such experiments are intended to help establish design criteria that will increase the satisfaction of users and meet their needs, as well as satisfy the requirements of management.

Briefly the evaluation of buildings must include the assessment of:

a. Physical conditions,
b. Functionality,
c. Attitudes of users about conditions,
d. Behavior and performance of users,
e. Cost.

The physical conditions include space, heat, light, sound, color, furnishings. Functionality includes such things as traffic flow, productivity of workers, etc. The attitudes of building users can be influenced not only by the physical conditions but also by organizational climate, personal factors, and demographics. The behavior of people within a building can be social or nonsocial and their performance can be in terms of sickness, absenteeism and productivity. Finally, cost is important to determine the cost effectiveness of designs.

Consider the case of family housing as an example of the large number of variables that can influence user performance and satisfaction. Many of these are shown in Figure 1. Everyone has some idea of what

FAMILY HOUSING

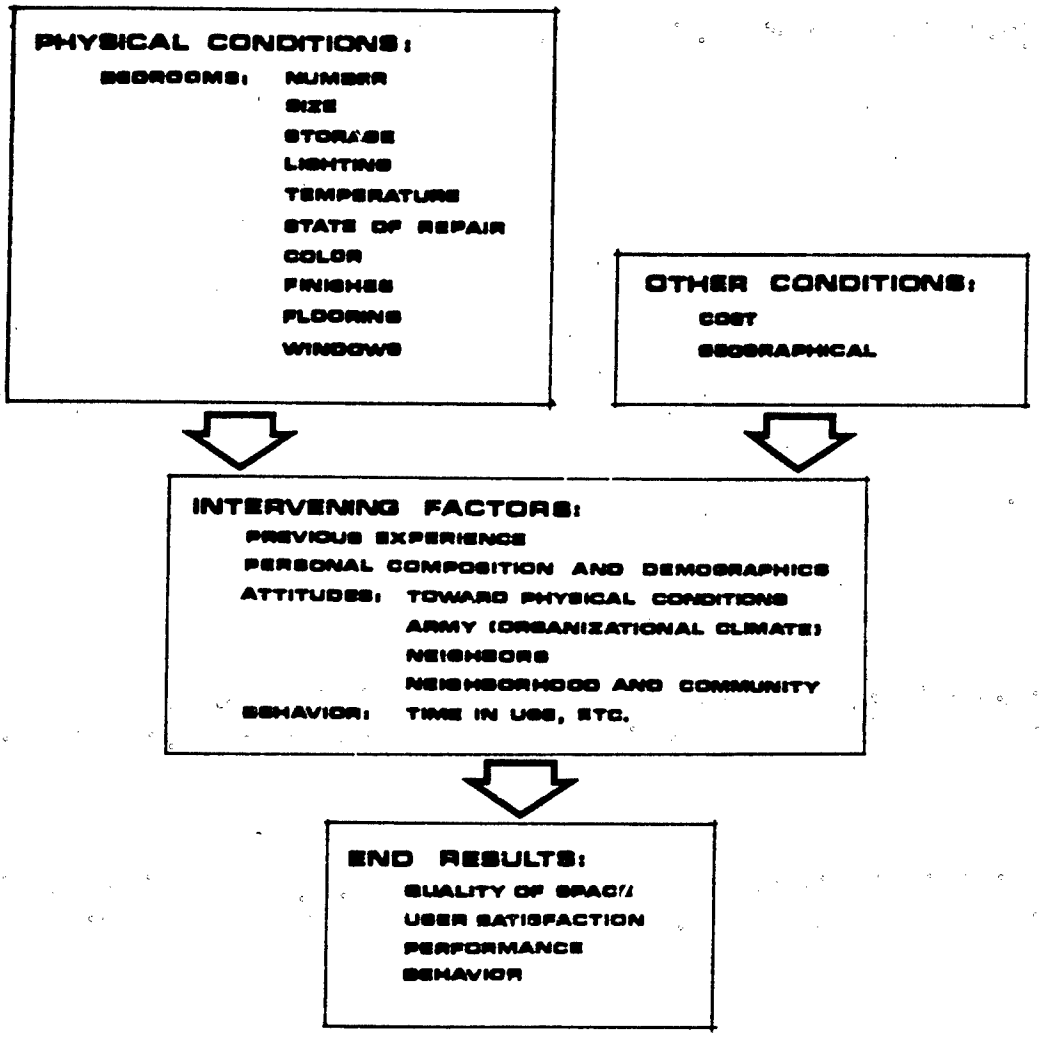


Figure 1. An Example (Family Housing) Showing the Many Variables That Have Some Influence on the Quality of Design

contributes to making a house acceptable, pleasant, desirable, and satisfying to its occupants and how well the design of the house supports the activities that occur there.

In describing the physical condition of the house it is clear that the various rooms in the house individually and collectively contribute to the quality of the house. Each room can be described in several ways. The number of bedrooms would be important, as would the size of the bedrooms, the amount of storage space, the lighting, the temperature, the state of repairs, room color, type of wall and woodwork finish, type of flooring, the kind of windows, the arrangement of one bedroom relative to the others, the arrangement of bedrooms relative to the rest of the house, the distance from the bedrooms to the bathroom, and so on. Each of the other rooms in the house could be described similarly. Each condition in each room contributes in some degree to the overall quality of the house.

Beside physical conditions, other factors can also have an influence on the quality of the house. It could be the overall cost of the house. If a house is too cheap, it may not be durable; if it is too expensive, it may place other constraints on family finances. The geographical location may be important to members of the family. Because the house is located in the wrong part of the country, no house would be good enough to satisfy the user. Geographical location might include distances to shops and stores, convenience to schools and convenience to work. Again each of these factors contributes to some degree to the quality of the house and to the satisfaction of its users.

Furthermore, each of these conditions may not contribute directly to the quality of the space or user satisfaction but are usually affected by intervening factors. The intervening factors could include previous experience with other houses, differences in personal composition--personality factors, age, level of income, social status, and other demographics. Attitudes are also important. There can be attitudes towards the physical conditions themselves, attitudes about one's job, general attitudes towards the Army and attitudes towards neighbors, the neighborhood, the community, the geographical location. In addition, behavior can have an effect on how the physical conditions relate to quality of the space or user satisfaction. If little time is spent in the house, the occupant may not be as critical about conditions. On the other hand, the occupant may prefer to do activities at home for which the design of the house is not very accommodating. The design of the house may directly impact the health and safety of the occupant. All these intervening factors in some way mediate the effect of physical and other conditions on the quality of the space and on user satisfaction and performance.

In designing an experiment which is intended to evaluate the quality of the house and how satisfied the user is with it, all these factors

must be measured and the effect that each one has on the quality of the space or on user satisfaction must be determined. To complicate matters, it is clear that interactions or interrelations between the physical conditions, attitudes, behaviors, previous experience, and personal composition exist. The strength of these interrelationships, as well as their effects must also be determined.

Other buildings could similarly be described and evaluated in terms of user satisfaction and performance as a measure of quality of design. Some buildings need to be evaluated on a more micro-scale basis. Here the problems of all of the other variables mentioned still exist, with one additional parameter attached. In order that the designer be able to make decisions about the physical character of the space he provides, he needs the psycho-social data to be location-specific. He requires that the data be connected to an identifiable physical location in a room, or a specific room in a building. Much of the psychological literature to date ignores this need and therefore is regarded as "unusable theory" by many designers. We find a need to quantify the behavioral success of a building and relate these data to specific locations in the environment. Therefore the problem under examination in this part of the discussion is the quantification of the degree of "fit" between man and his environmental setting and the identification of behavioral "units" of designable environment.

A basic "chunk" or unit of man-environment interaction called a behavior setting has been identified in the psychological literature (Barker, 1968) and applied to such contexts as housing (Bechtel, 1970) and hospitals (LeCompte, 1972). This unit shows great promise for analyzation of micro-scale architectural environments. Barker states that this "chunk," the behavior setting, is characterized by a standing pattern of clearly identifiable behaviors regardless of participants. Examples might be classes of behaviors in restaurants, libraries, and supermarkets. In each of these contexts, patterns of behavior are similar for participants, and are independent of individuals. The behavior setting analysis technology developed by Barker provides a system for identifying chunks of location specific behavior and notating activities and attitudes to these. Just as a language has a vocabulary, syntax, and rules of grammar, the behavior setting has units, qualities, and degrees of independence and interdependence. Now, the problem of using these "chunks" of man-environment interaction becomes more difficult when structured into an experimental situation to provide information usable for the designer.

An example of a military dining hall might serve to develop the concept of behavior and attitude relationships related to specific units of environment. In the normal dining process of food acquisition, respondents experience physical and social environment in a linear sequence (Thiel, 1961). Each of the activities subjects engage in can be differentiated by the nature of the behavior mechanisms employed, i.e., gross motor, manipulative, motion, etc. On the basis of this differentiation,

we can identify some ten, distinctly different key behavior settings such as sign-in desk, silverware pick-up, path to table, etc. Now the behavior in the physical space of one setting will affect the attitudes of respondents at another setting elsewhere in the dining space. We know that respondents will rate the concept of "privacy" lower at the dining table as the number of persons standing in line at the sign-in desk increases (Gibbs, 1972). The behavioral data in this case is location-specific. The designer can make decisions of a physical nature to change the attitude rating. He can shield the check-in desk from view, provide two desks for faster processing, or more the check-in desk elsewhere, all with the intent of increasing the sense of privacy at the dining table. In this example, behavioral data (number of persons in line) and attitude (rating of privacy) have been made location-specific (sign-in desk, table) and the designer can make decisions based upon this information. However it is not quite that simple.

From our previous discussion, we realize the great number of variables interacting in any social setting. Obviously privacy and population at sign-in desk are not independent of other factors in the environment. Not only would the dining "privacy" experience be affected by population movement, but also by physical conditions, noise level, and the management climate. To some degree, all of these affect the rating of privacy. It is the combination of factors that cumulatively make up the concept of privacy and the problem is again one of a large number of variables. It is, of course, possible that the designer could not solve them all, but certainly a knowledge of what amount of the total variance could be accounted for by changes in physical design would be useful in developing a measure of cost effectiveness for changes in the environment.

We have presented two kinds of problems in environmental analysis each involving a great number of variables. The first involved the range of parameters and variables existent in any environmental setting, and the second problem addressed the need for identifiable "chunks" of behavior-environment interaction which the designer might address to begin his architectural translation process, with the intent of improving the performance of the building. From the standpoint of the designer, he realized that behavior and environment interaction is not a one-valued concept, but rather multi-dimensional and interactive. The behavioral scientist would certainly believe that interactions between the many variables suggested in this paper are probably more complicated than simple correlation coefficients would describe, yet he would not suggest a priori a series of interrelated factors. The question is then, what kinds of research designs and analytic techniques will lend themselves to discovering major components of human interaction with environment. Multiple linear regression (Brauer, 1972), factor analysis (Canter, 1972), and cluster analysis (Bechtel, 1972) have been suggested.

Since it is seldom possible to experimentally control variables in building designs and comparative demonstration projects are far too expensive, suggestions for handling a large number of variables in a semi-controlled, real world situation are needed. Are there other techniques more responsive than those suggested to discovering relations and major factors in large data matrices? Can they also relate disparate data from attitude and behavioral investigations to location-specific chunks of environment? Will these techniques be more or less responsive than present ones to investigations in the real world context, and what sort of experimental controls are necessary? These are the questions we ask in order to build a firm scientific basis for the design of buildings that are compatible with human behavior and needs.

REFERENCES

Barker, Roger, *Ecological Psychology*, Stanford University Press, Stanford, California, 1968.

Bechtel, Robert, *Arrowhead Report*, Environmental Research and Development Foundation, Kansas City, Missouri, July 1971.

Bechtel, Robert, Conference with the author, December 4, 1972 at CERL.

Brauer, Roger, *Survey of Soldiers Attitudes to Troop Housing*, U.S. Army Construction Engineering Research Laboratory, Champaign, Illinois, January 1973 (in press).

Canter, David, "Royal Hospital for Sick Children," *The Architect's Journal*, September 6, 1972, pp. 525-564.

Gibbs, Wesley, *Dining Facility User - Attitude Survey and Decor Experiment at Travis AFB*, U.S. Army Construction Engineering Research Laboratory, Champaign, Illinois, January 1973.

LeCompte, "Behavior Settings: The Structure of the Treatment Environment," *EDRA III/AR 8 Proceedings*, UCLA, Los Angeles, California, January 1972, p. 4.2.

ORTHOGONAL ESTIMATES IN WEIGHING DESIGNS

William G. Lese, Jr.
US Army Materiel Systems Analysis Agency, APG, Maryland

and

K. S. Banerjee
University of Delaware, Newark, Delaware

ABSTRACT

A new technique has been developed for modifying all balanced incomplete block designs (BIBD) to provide orthogonal estimates when the modified BIBD are to be used as a weighing design. Previously, K. S. Banerjee developed a method for modifying BIBD to provide orthogonal estimates. However, for a certain class of BIBD, Banerjee's method failed to provide orthogonal estimates. A comparison of the relative efficiencies of the new procedure with that of Banerjee's procedure is also presented. In addition, under the new procedure it is shown that the covariance matrix of the estimators obtained by the least squares procedure is identical to that obtained by the maximum liklihood procedure, even when the design matrix X is not square. Several examples of the utilization of the new technique, along with a historical development of the weighing problem from its origin in a casual example by Yates through the work of Hotelling, Mood, Kempthorne, and Banerjee as relative to the problem of providing orthogonal estimates, is also presented.

CHAPTER I

ORIGIN OF THE WEIGHING PROBLEM

In an article, "Complex Experiments", Yates [10] considered the following problem: A chemist is given the task of determining the weights of seven light objects, and the scale the chemist must use requires a zero correction. The customary technique would be to weigh each of the seven objects individually and then make an eighth weighing with no objects on the scale. This eighth weighing would be used to determine the zero correction factor.

Mathematically, the customary weighing technique for the chemist problem would be as follows: The seven objects will be denoted as a, b, c, d, e, f, and g. The scale bias will be denoted by z.

Weighing Number	Object Weighed	Scale Reading
1	a + z	Y_1
2	b + z	Y_2
3	c + z	Y_3
4	d + z	Y_4
5	e + z	Y_5
6	f + z	Y_6
7	g + z	Y_7
8	z	Y_8

Using this notation the weight of any object can be determined by taking the difference between the scale readings when carrying the object and the scale reading when no object is on the scale. For this example, the weights of the seven objects would be determined as follows:

$$a = Y_1 - Y_8$$

$$b = Y_2 - Y_8$$

$$c = Y_3 - Y_8$$

$$d = Y_4 - Y_8$$

$$e = Y_5 - Y_8$$

$$f = Y_6 - Y_8$$

$$g = Y_7 - Y_8$$

Assuming that systematic errors are non-existent and that the errors are random, the variance of each weighing may be denoted by σ^2 and the standard error by σ. With these assumptions, the variance of the estimated weights using the customary weighing technique is $2\sigma^2$ and the standard error is $\sigma\sqrt{2}$.

For an improvement over the customary weighing technique, Yates suggested that the objects should be weighed in combinations with each other instead of being weighed individually. For example, Figure 1 presents Yates' technique for the determination of the weights of seven objects.

We notice that in Yates method, each object is weighed four times in combination with the other objects. In the four weighings of a given object, every other object is included twice. In the remaining four weighings, i.e. the weighing without the object, every other object is also included twice. Therefore, the weight of any object can be determined by adding the scale readings containing the object, subtracting the scale readings not containing the object and dividing this result by 4. Using this procedure, the weights of the seven objects would be determined as follows:

$$a = \frac{Y_1 + Y_2 + Y_3 + Y_4 - Y_5 - Y_6 - Y_7 - Y_8}{4}$$

$$b = \frac{Y_1 + Y_2 + Y_5 + Y_6 - Y_3 - Y_4 - Y_7 - Y_8}{4}$$

$$c = \frac{Y_1 + Y_3 + Y_5 + Y_7 - Y_2 - Y_4 - Y_6 - Y_8}{4}$$

$$d = \frac{Y_1 + Y_2 + Y_7 + Y_8 - Y_3 - Y_4 - Y_5 - Y_6}{4}$$

$$e = \frac{Y_1 + Y_3 + Y_6 + Y_8 - Y_2 - Y_3 - Y_5 - Y_7}{4}$$

Weighing Number	Objects Weighed	Scale Reading
1	a + b + c + d + e + f + g	Y_1
2	a + b + d	Y_2
3	a + c + e	Y_3
4	a + f + g	Y_4
5	b + c + f	Y_5
6	b + e + g	Y_6
7	c + d + g	Y_7
8	d + e + f	Y_8

Figure 1. Yates Weighing Design

$$f = \frac{Y_1 + Y_4 + Y_5 + Y_8 - Y_2 - Y_3 - Y_6 - Y_7}{4}$$

and

$$g = \frac{Y_1 + Y_4 + Y_6 + Y_7 - Y_2 - Y_3 - Y_5 - Y_8}{4}$$

It should be noted that the bias (i.e. the zero scale correction) cancels out in the above expressions. Since the variance of a sum of independent observations is the equal to the sum of the variances, the variance of an estimated weight determined by Yates technique is $\sigma^2/2$ and the standard error is $\sigma/\sqrt{2}$, whereas the variance and standard error using the customary technique was $2\sigma^2$ and $\sigma\sqrt{2}$, respectively. Therefore, Yates weighing technique improved the precision of the estimated weights as compared to the customary weighing technique without increasing the number of weighing operations (both techniques required eight weighing operations).

This illustration by Yates, whereby the precision of an estimated weight was increased (without additional weighings) by weighing the objects in combinations rather than individually, was the origin of the weighing problem.

CHAPTER II

HOTELLING'S ENUNCIATION OF THE WEIGHING PROBLEM

Hotelling [5] presented a further improvement by suggesting that Yates' procedure be modified by placing in the other pan of the scale those objects not included in each weighing as specified by Yates. For example, in Yates' procedure (see Chapter I) one weighing of the seven objects had the combination of $a + b + d = Y_2$ which represented the objects a, b, and d being weighed on the scale. Hotelling suggested that in addition to objects a, b, and d being placed in one pan for weighing, the remaining objects, i.e. c, e, f, and g, be placed in the other pan. Hotelling's weighing design for the determination of the weights of seven light objects is presented in Figure 2. Using Hotelling's procedure, the estimated weights of each of the seven light objects would be given as follows:

$$a = \frac{W_1 + W_2 + W_3 + W_4 - W_5 - W_6 - W_7 - W_8}{8}$$

$$b = \frac{W_1 + W_2 + W_5 + W_6 - W_3 - W_4 - W_7 - W_8}{8}$$

$$c = \frac{W_1 + W_3 + W_5 + W_7 - W_2 - W_4 - W_6 - W_8}{8}$$

Weighing Number	Objects Weighed	Scale Reading
1	a + b + c + d + e + f + g	W_1
2	a + b - c + d - e - f - g	W_2
3	a - b + c - d + e - f - g	W_3
4	a - b - c - d - e + f + g	W_4
5	- a + b + c - d - e + f - g	W_5
6	- a + b - c - d + e - f + g	W_6
7	- a - b + c + d - e - f + g	W_7
8	- a - b - c + d + e + f - g	W_8

Figure 2. Hotelling's Weighing Design

$$d = \frac{W_1 + W_2 + W_7 + W_8 - W_3 - W_4 - W_5 - W_6}{8}$$

$$e = \frac{W_1 + W_3 + W_6 + W_8 - W_2 - W_4 - W_5 - W_7}{8}$$

$$f = \frac{W_1 + W_4 + W_5 + W_8 - W_2 - W_3 - W_6 - W_7}{8}$$

$$g = \frac{W_1 + W_4 + W_6 + W_7 - W_2 - W_3 - W_5 - W_8}{8}$$

The variance of each unknown weight by Hotelling's method is therefore $\sigma^2/8$ and the standard error is $\sigma/\sqrt{8}$. This standard error is half that of Yates method and a fourth of the value determined by the customary method. Also in Hotelling's method, as in Yates' method, the scale bias (zero correction) gets subtracted out in the equations for the determination of the unknown weights.

The design principle inherent in both Yates' and Hotelling's method may be illustrated even better with reference to a simpler example. Let us suppose that it is required to find the unknown weights of two objects a and b, and that the scale to be used is already corrected for bias. If the two objects are weighed together in one pan of the scale, and also in opposite pans, the equations for the unknown weights will be

$$a + b = Y_1, \quad a - b = Y_2,$$

where Y_1 and Y_2 denote the readings from the scale. From the above,

we get

$$a = \frac{Y_1 + Y_2}{2}$$

and

$$b = \frac{Y_1 - Y_2}{2}$$

If σ^2 is the variance of an individual weighing, the variance of $\underline{a}$ and $\underline{b}$, by this method, is obtained as $\sigma^2/2$. The error for both the estimates, therefore, is $\sigma/\sqrt{2}$. Thus, with only 2 weighing operations, it has been possible to obtain the standard error for both the objects as $\sigma/\sqrt{2}$; whereas if the objects were weighed separately twice each, 4 weighing operations would have been needed in all to obtain this standard error for both. Weighing the objects in combination has, therefore, saved the trouble of making weighing operations by half the number.

The above, therefore, amply illustrates the following quotation due to Hotelling: "When several quantities are to be ascertained there is frequently an opportunity to increase the accuracy and reduce the cost by suitably combining in one experiment what might ordinarily be considered separate operations".

In addition to the improvement in Yates' method, Hotelling also gave a precise formulation of the weighing design problem. This formulation, as later pointed out by Banerjee [1], may be interpretated as follows:

Results of N weighing operations to determine the individual

weights of p light objects fit in to the general linear hypothesis model, $Y = X\beta + \varepsilon$, where Y is an N x 1 random observed vector of the recorded weights; $X = (x_{ij})$, $i = 1, 2, \ldots N$; $j = 1, 2, \ldots, p$, is an N x p matrix of known quantities, with $x_{ij} = +1, -1$ or 0, if, in the i^{th} weighing operation the j^{th} object is placed respectively in the left pan, right pan or in none; β is a p x 1 vector ($p \leq N$) representing the weights of the objects; ε is an N x 1 unobserved random vector such that $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma^2 I_N$.

Consistent with the signs that the elements x_{ij} can take, the record of the i^{th} weighing is taken as positive or negative, according as the balancing weight is placed in the right pan or left.

The matrix X is called the "design matrix". When X is of full rank, that is, when [X'X] is non-singular, the least squares estimates of the weights are given by $\hat{\beta} = [X'X]^{-1} X'Y$, where X' is the transpose of X. The covariance matrix of the estimated weights is given by $COV(\hat{\beta}) = \sigma^2 C$. The i^{th} diagonal element of C, c_{ij}, represents the variance factor for the i^{th} object. The objective of the Weighing Design Problem is to obtain the design matrix, X, such that the c_{ii} are a minimum.

In this connection, Hotelling [5] proved the following Lemma:

Let $A = [X'X] = (a_{ij})$, $i, j = 1, 2, \ldots, p$. Then, if a_{12}, $a_{13}, \ldots, a_{1p}$ ($= a_{21}, a_{31}, \ldots, a_{p1}$ respectively) are free to vary while the other elements of A remain fixed, the maximum value of

$|A|/A_{11}$ is a_{11}, and is attained when and only when $a_{12} = a_{13} = \ldots = a_{1p} = 0$, where A_{11} is the minor of A obtained by deleting the first row and column.

From the above Lemma, it is evident that the variance of $\hat{\beta}_1$, namely $\sigma^2 A_{11}/|A|$, cannot be less than σ^2/a_{11}, and that the variance would reach this value only if the experiment is so arranged that the elements after the first row and column of A are all zero. This minimum value, σ^2/a_{11}, will be attained, when the first column of X is orthogonal to all the others. It will also be clear that the minimum minimorum [5] of the variance will be reached, if the first column of X is not only orthogonal to all the others, but also if it consists entirely of +1's and -1's as its elements, so that $a_{11} = N$. N is the maximum possible value that a_{11} can take. The value of this minimum minimorum will thus be equal to σ^2/N.

It is evident from the Lemma and the above discussion that this minimum minimorum of the variance would be reached in respect of all the estimates $\hat{\beta}_i$, $(i = 1, 2, \ldots, p)$, if the design matrix X is orthogonal in the sense that [X'X] is diagonal with N on the diagonal.

CHAPTER III

TWO TYPES OF WEIGHING PROBLEMS (SPRING AND CHEMICAL BALANCE) AND ILLUSTRATIONS OF EACH

In general, there are two distinct types of weighing problems. These problems have been designated as the spring balance weighing problem and the chemical balance weighing problem. In the spring balance problem (only one pan is used) the design matrix X is composed of elements x_{ij} which can assume only the values of +1 or 0, where +1 denotes that the object is to be placed in the pan and a 0 denotes that the object is not to be placed in the pan. In the chemical balance problem (two pans are used), the design matrix X is composed of elements x_{ij} which can assume values of +1, -1, or 0, where +1 denotes that the object be placed in the left pan, -1 denotes that the object be placed in the right pan and 0 denotes that the object not be placed on either pan. It is evident that the design originally proposed by Yates (see Chapter I) was a spring balance design, and the improved design proposed by Hotelling was really a chemical balance design (see Chapter II).

Example of a Spring Balance Weighing Design

As an example of a spring balance weighing design, consider the problem of determining the weights of three objects (a, b, and c).

A spring balance weighing design matrix would be:

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Where the columns refer to the objects and the rows refer to the weighing operation. For this example the first weighing would have objects a and b on the scale. Object c would not be used in the first weighing operation. The second weighing operation would have objects a and c being weighed together. Object b would not be used in the second weighing operation. Likewise, the third weighing operation would have objects b and c being weighed together. Object a would not be used in the third weighing operation.

The least squares estimates of the unknown weights are given by $\hat{\beta} = [X'X]^{-1}X'Y$ and the covariance of the unknown weights are given by $\text{Cov}(\hat{\beta}) = (X'X)^{-1}\sigma^2$, where X is the weighing design matrix and Y is the matrix of recorded weights of each weighing operation. For this example

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

$$[X'X] = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix},$$

and

$$[X'X]^{-1} = \begin{bmatrix} 3/4 & -1/4 & -1/4 \\ -1/4 & 3/4 & -1/4 \\ -1/4 & -1/4 & 3/4 \end{bmatrix}$$

For the least squares estimates,

$$[X'X]^{-1}X'Y = \begin{bmatrix} 1/2 & 1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 \\ -1/2 & 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

We therefore have, as least squares estimates,

$$\hat{a} = \frac{Y_1 + Y_2 - Y_3}{2}$$

$$\hat{b} = \frac{Y_1 - Y_2 + Y_3}{2}$$

and

$$\hat{c} = \frac{-Y_1 + Y_2 + Y_3}{2}.$$

The variance of each unknown weight is $3/4\ \sigma^2$ (as given by the diagonal elements of $(X'X)^{-1}$).

Example of Chemical Balance Weighing Design

As an example of a chemical balance weighing design consider the problem of determining the weights of four objects (a,b,c, and d). A chemical balance weighing design matrix would be:

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

As in the spring balance example, the columns refer to the objects to be weighed and the rows refer to the weighing operation. In this case, the first weighing operation would have all four objects (a, b, c, and d) placed in the left pan. The second weighing operation would have objects b and d placed in the left pan and objects a and c in the right pan. The third weighing would have objects a and b placed in the right pan and objects c and d in the left pan. Likewise, the fourth operation would have objects b and c placed in the right pan and objects a and d placed in the left pan. Using the same notation as in the spring balance example, we have

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

$$[X'X] = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix},$$

and

$$[X'X]^{-1} = \begin{bmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 1/4 \end{bmatrix}$$

For the least squares estimators,

$$[X'X]^{-1}X'Y = \begin{bmatrix} 1/4 & -1/4 & -1/4 & 1/4 \\ 1/4 & 1/4 & -1/4 & -1/4 \\ 1/4 & -1/4 & 1/4 & -1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix}$$

We therefore have, as least squares estimates,

$$\hat{a} = \frac{Y_1 - Y_2 - Y_3 + Y_4}{4}$$

$$\hat{b} = \frac{Y_1 + Y_2 - Y_3 - Y_4}{4}$$

$$\hat{c} = \frac{Y_1 - Y_2 + Y_3 - Y_4}{4}$$

$$\hat{d} = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4}$$

The variance of each unknown weight is $\sigma^2/4$ (as given by the diagonal elements of $(X'X)^{-1}$).

CHAPTER IV

MOOD AND KEMPTHORNE'S CONTRIBUTIONS

Mood [7] has indicated that if N weighing operations are made to determine the weights of N objects, the minimum variance that the estimated weights could have is σ^2/N; and further, this minimum variance will be reached only when the design matrix is orthogonal (orthogonal in the sense that (X'X) is diagonal) with elements consisting entirely of +1's and -1's. Thus, Mood [7] showed that the problem of finding the best chemical balance design is related to Hadamard matrices and the Hadamard determinant problem.

The theorem that Hadamard proved is as follows:

If the elements x_{ij} of a square matrix X are restricted to the range $-1 \leq x_{ij} \leq 1$, the maximum possible value of the determinant of X is $N^{N/2}$, and when this maximum value is achieved, all $x_{ij} = \pm 1$. The matrix X also will be orthogonal in the sense that (X'X) will be diagonal with all non-zero elements equal to N.

Such matrices are denoted by H_N. If H_N exists for a given N, H_N is the best chemical balance design for N = p. If the number of objects, p, to be weighed is less than N, the best design is one which

is derived from H_N by selecting a number of columns equal to the number of objects to be weighed.

Mood has further pointed out the work of Paley [8] and Williamson [9] and showed that H_{4k} exists for the range of $0 < 4k < 100$ with the possible exception of $4k = 92$. The solution for $4k = 92$ was later found by Baumert, Golomb, and Hall Jr. [3].

In summary, for those chemical balance designs where an H_N exists, the determination of the optimum design is completely solved; i.e. the optimum chemical balance design will be H_N, if H_N exists.

For the remainder of this dissertation we shall devote our attention to the spring balance problem. As mentioned before, the spring balance problem differs from the chemical balance problem in that the elements of the design matrix X can only assume values of +1 or 0, whereas, in the chemical balance design the elements of the design matrix X can assume values of +1, -1, or 0.

We notice that Hadamard's theorem does not directly apply to the spring balance problem since the design matrix, X, can only assume values of +1 and 0. For the spring balance problems, where $N = p$ and $N \equiv 3 \pmod 4$, Mood showed that the best possible spring balance design is determined by H_{N+1}, if H_{N+1} exists. Mood's method of construction of these best possible designs is as follows:

Let K_{N+1} denote a matrix formed from H_{N+1} by adding or

subtracting the elements of the first row of H_{N+1} from the corresponding elements of the other rows in such a way as to make the first element of each of the remaining rows zero. Obviously,

$$|K_{N+1}| = \pm |H_{N+1}| .$$

Except for the first row, the elements of K_{N+1} are 0 and ± 2 with the signs of the non-zero elements being the same for elements in the same row. Let L_N be the matrix obtained by omitting the first row and column of K_{N+1} by changing all non-zero elements to +1, and by permuting two rows, if necessary, to make the determinant of L_N positive. Then

$$|H_{N+1}| = 2^N |L_N|$$

It is clear that, given L_N, one could reverse the procedure and determine an H_{N+1}. In the same manner, there is a correspondence in general between square matrices with elements ± 1 and square matrices of one less order with elements 0 and 1. The ratio of the values of corresponding determinants is always 2^N, if their determinants do not vanish; hence the (0, 1)-determinant will always have its maximum value when its corresponding (+1, -1)-determinant has its maximum possible value. Thus, $|L_N|$ is the maximum value possible for a determinant of 0's and 1's of order N, and the value of $|L_N|$ is

$$|L_N| = \frac{(N+1)^{\frac{1}{2}(N+1)}}{2^N}$$

The variances of the estimated weights will be $a^{ii} = \sigma^2 4N/(N+1)^2$. We knew in advance that a^{ii} would be greater than $1/N$, since an optimum

design cannot exist unless the design matrix has its elements equal ±1 and the spring balance design is restricted to elements +1 or 0.

For the spring balance problems when $N > p$, Mood has presented the following approach for obtaining optimum spring balance designs:

Let P_r be a matrix whose rows are all the arrangements of r ones and $(p - r)$ zeroes $(0 \leq r \leq p)$. (The symbol should also have a subscript p but that is omitted because any specific value for p will always be clear from the context.) The matrix will have p columns and $\binom{p}{r}$ rows. Let X be a matrix made up of matrices P_r arranged in vertical order. Let n_r be the number of times P_r is used in constructing X. The matrix X is then a weighing design for p objects and $N = \sum_r n_r \binom{p}{r}$. Using these notations, Mood has proven the following two theorems giving the best spring balance designs:

Theorem (1): If $p = 2k - 1$, where k is a positive integer, and if N contains the factor $\binom{p}{k}$ then $|a_{ij}|$ (det. $|A|$) will be maximized when $n_k = N/\binom{p}{k}$ and all other $n_r = 0$.

Theorem (2): If $p = 2k$, where k is a positive integer, and if N contains the factors $\binom{p+1}{k+1}$, then $|a_{ij}|$ (det. $|A|$) will be maximized when $n_k = n_{k+1} = N/\binom{p+1}{k+1}$, and all other $n_r = 0$.

When p is odd, Mood observed that P_k is a design which not only minimizes the confidence region for estimating the weights, but also minimizes the individual variance factors. When, however, p is

even, Mood observed that the variance factors may not be the minimum, and makes a surmise that the best design from the point of view of minimum variance factors would be made up largely from P_k and a small proportion from P_{k+1}.

Kempthorne [6] discussed the weighing problem from the point of view of factorial experiments and in [6] has given rules by which the fractional designs may be constructed. Kempthorne has indicated that the fractional designs have the following properties:

(1) The design automatically takes care of any bias in the balance.

(2) The effects or weights may be easily computed.

(3) The effects or weights are uncorrelated.

(4) All the weights are measured with the same precision.

(5) An estimate of the experimental error which is independent of the effects may be computed from the results.

Kempthorne also compared his fractional designs with the designs proposed by Mood and has found that the fractional factorial designs will yield estimates which have a somewhat higher variance than Mood's designs. Kempthorne also indicated that the increase in precision in Mood's designs had been obtained at the expense of having correlated estimates which are subject to any bias that the measuring instrument may have. For these reasons, Kempthorne doubted whether the use of Mood's designs for any practical problem could be justified.

Subsequent to this remark by Kempthorne, Banerjee [1] showed that the L_N designs of Mood are a special class of symmetrical balanced incomplete block designs and that the L_N designs could be easily modified to provide orthogonal estimates as was referred to by Kempthorne. In addition, since the L_N designs are a subset of symmetrical balanced incomplete block designs, Banerjee [2] developed a general method to show how BIBD's in general could be made to provide orthogonal estimates when used as weighing designs. A detailed description of Banerjee's method is presented in the next chapter.

CHAPTER V

BANERJEE'S METHOD OF MODIFYING BIBD TO PROVIDE ORTHOGONAL ESTIMATES IN WEIGHING DESIGNS

As was mentioned in Chapter IV, Kempthorne noted that although the optimum designs for the spring balance problem suggested by Mood furnish somewhat smaller variance than that given by fractional replicates, Mood's designs have the disadvantage that the estimates are correlated, whereas the estimates furnished by fractional replicates are orthogonal. Banerjee [1] has shown that the optimum designs of Mood may also be made to furnish orthogonal estimates when the designs are adjusted to suit estimation in a biased spring balance. Since the optimum designs, L_N, of Mood are a special class of a symmetrical BIBD's a question arises if it would be possible to provide by a similar type of an adjustment, orthogonal estimates when BIBD's are used as spring balance weighing design. A complete detailed procedure indicating how balanced incomplete block designs, in general, may be made to furnish orthogonal estimates in weighing designs was presented by Banerjee [2]. Banerjee's procedure is presented in the following:

Usually v denotes the number of varieties and b denotes the number of blocks in a balanced incomplete block design. However, in weighing designs, v will be used to denote the number of objects to be weighed

and b will denote the number of weighings to be made. The parameter r will be used to denote the number of times each object is weighed, λ will denote the number of times each pair is weighed together and k will denote the number of objects weighed in a combination. Let X be the design matrix where a BIBD is used as a weighing design. The matrix [X'X] will have the form,

$$X'X = \begin{bmatrix} r & \lambda & \lambda & \lambda & \cdots & \lambda \\ \lambda & r & \lambda & \lambda & \cdots & \lambda \\ \lambda & \lambda & r & \lambda & \cdots & \lambda \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \lambda & \lambda & \lambda & \lambda & \cdots & r \end{bmatrix}.$$

[X'X] is of order v x v.

For the determination of the variances and covariances, we need to determine the inverse of [X'X]. The diagonal elements in the inverse matrix, $[X'X]^{-1}$, represent the variance factors and will all be equal to

$$\frac{r + \lambda (v - 2)}{(r - \lambda)[r + \lambda (v - 1)]} \sigma^2.$$

The off-diagonal elements represent the covariance terms and will all be equal to

$$\frac{-\lambda}{(r - \lambda)[r + \lambda (v - 1)]} \sigma^2.$$

Since these off-diagonal elements are not zero, we see that the estimates are correlated.

Banerjee [2] has suggested the following procedure for modifying the BIBD to provide orthogonal estimates. Taking the bias as an additional object to be weighed, a column of ones and only one row of zeroes in that order may be added to the BIBD design matrix to correspond to the bias assumed as an additional object. The modified design will then be suitable for the estimation of the weights. This really means to make one additional weighing to obtain an estimate of the bias and in the subsequent weighings the bias will automatically be included. If, however, a column of ones and t rows of zeroes are added to the design matrix, this implies that t weighings will be devoted to the estimation of the bias. In such a situation, the matrix [X'X] will be of the form,

$$[X'X] = \begin{bmatrix} b+t & r & r & r & \cdots & r \\ r & r & \lambda & \lambda & \cdots & \lambda \\ r & \lambda & r & \lambda & \cdots & \lambda \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ r & \lambda & \lambda & \lambda & \cdots & r \end{bmatrix}.$$

Because of the inclusion of the bias, the order of this matrix is $(v + 1) \times (v + 1)$.

We would like to obtain the inverse of this matrix in the following manner:

The determinant, $|X'X| = (r - \lambda)^{v-1} [(b + t) \{r + \lambda(v - 1)\} - r^2 v]$.

On simplifying we obtain,

$$|X'X| = (r - \lambda)^{v-1} [t\{r + \lambda (v - 1)\} + b\{r + \lambda (v - 1)\} - r^2 v]$$

$$= (r - \lambda)^{v-1} [t\{r + \lambda (v - 1)\} + b\{r + r (k - 1)\} - r^2 v]$$

$$= (r - \lambda)^{v-1} [t\{r + \lambda (v - 1)\} + b k r - r^2 v]$$

$$= (r - \lambda)^{v-1} [t\{r + \lambda (v - 1)\}]$$

$$= t (r - \lambda)^{v-1} \{r + \lambda (v - 1)\} \qquad (1)$$

In a similar fashion, the value of the determinant obtained after suppressing the first row and the first column of $[X'X]$ can be shown to be

$$(r - \lambda)^{v-1} \{r + \lambda (v - 1)\} . \qquad (2)$$

The value of the determinant after suppressing the second row and second column of $[X'X]$ is

$$(r - \lambda)^{v-2} [(b + t)\{r + \lambda (v - 2)\} - r^2 (v - 1)] \qquad (3)$$

The value of the determinant obtained after suppressing the first row and second column of $[X'X]$ is

$$r (r - \lambda)^{v-1} \qquad (4)$$

The value of the determinant obtained after suppressing the second row and the third column of $[X'X]$ is

$$(r - \lambda)^{v-2} \{\lambda(b + t) - r^2\} \qquad (5)$$

For any two estimates to be orthogonal, the off-diagonal

elements of $[X'X]^{-1}$ corresponding to those estimators must be equal to zero. Keeping the bias out of consideration we would like expression (5) to be equal to zero. That is, to obtain orthogonal estimates, the value of t must be chosen such that

$$(r - \lambda)^{v} - 2\{\lambda(b + t) - r^2\} = 0$$

Since r cannot be equal to λ, the expression

$$\lambda(b + t) - r^2 = 0,$$

or

$$t = \frac{r^2}{\lambda} - b$$

Using this value for t, the matrix $[X'X]^{-1}$ becomes:

$$[X'X]^{-1} = \begin{bmatrix} \frac{1}{t} & -\frac{1}{tk} & -\frac{1}{tk} & \cdots & -\frac{1}{tk} \\ -\frac{1}{tk} & \frac{1}{r-\lambda} & 0 & \cdots & 0 \\ -\frac{1}{tk} & 0 & \frac{1}{r-\lambda} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{tk} & 0 & 0 & \cdots & \frac{1}{r-\lambda} \end{bmatrix}.$$

This expression shows that, except for the bias, the other estimates are mutually orthogonal. The estimates given by $\hat{\beta} = [X'X]^{-1} X'Y$ will be given as:

$$\hat{\beta} = [X'X]^{-1} X'Y = \begin{bmatrix} \frac{1}{t} & \frac{1}{t} & \frac{1}{t} & 0 & 0 & 0 & \cdots \\ -\frac{1}{tk} & -\frac{1}{tk} & -\frac{1}{tk} & \frac{1}{r} & \frac{1}{r} & \frac{1}{r} & \cdots \\ -\frac{1}{tk} & -\frac{1}{tk} & \frac{1}{r} & -\frac{1}{tk} & -\frac{1}{tk} & -\frac{1}{tk} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{bmatrix} \begin{bmatrix} Y \end{bmatrix}$$

Given a balanced incomplete block design with parameters v, b, r, k and λ, it is always possible to obtain another balanced incomplete block design with parameters

$$v_0 = v$$
$$b_0 = b$$
$$r_0 = b - r$$
$$k_0 = v - k$$
$$\lambda_0 = b - 2r + \lambda .$$

The two balanced incomplete block designs derived in such a fashion are said to be complementary to one another.

When an integral value of t ($t = \frac{r^2}{\lambda} - b$) is available it is also possible to obtain orthogonal estimates in the complementary design. This is done by adding a column of ones and t rows of ones in that order to the design matrix X. The matrix [X'X] will then have the form,

$$[X'X] = \begin{bmatrix} b_0 + t & r_0 + t & r_0 + t & \cdots & r_0 + t \\ r_0 + t & r_0 + t & \lambda_0 + t & \cdots & \lambda_0 + t \\ r_0 + t & \lambda_0 + t & r_0 + t & \cdots & \lambda_0 + t \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_0 + t & \lambda_0 + t & \lambda_0 + t & \cdots & r_0 + t \end{bmatrix}$$

The value of the determinant after suppressing the second row and the third column of [X'X] in this case will be given by:

$$[(r_0 + t) - (\lambda_0 + t)]^{v-2} \; (\lambda_0 + t)\,(b_0 + t) - (r_0 + t)^2$$

To have orthogonal estimates, for all the weights except for the bias, this quantity must be zero, i.e.

$$[r_0 - \lambda_0]^{v-2} \; (\lambda_0 + t)\,(b_0 + t) - (r_0 + t)^2 = 0$$

or

$$(\lambda_0 b_0 + \lambda_0 t + t b_0 + t^2 - r_0^2 - 2 r_0 t - t^2) = 0 \,,$$

which is satisfied. The matrix $[X'X]^{-1}$ then becomes

$$\begin{bmatrix} \phi & -\frac{1}{r} & -\frac{1}{r} & \cdots & -\frac{1}{r} \\ -\frac{1}{r} & \frac{1}{r-\lambda} & 0 & \cdots & 0 \\ -\frac{1}{r} & 0 & \frac{1}{r-\lambda} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ -\frac{1}{r} & 0 & 0 & \cdots & \frac{1}{r-\lambda} \end{bmatrix}$$

where $\phi = \frac{1}{r}\left\{1 + \frac{t\,k\,(k-1)}{r}\right\}$. The estimated weights of the objects will be given by

$$[X'X]^{-1}\,X'Y = \begin{bmatrix} \phi - \frac{v}{r} & \phi - \frac{v}{r} & \phi - \frac{v}{r} & \phi - \frac{k_0}{r} & \phi - \frac{k_0}{r} & \cdots \\ \frac{1}{tk} & \frac{1}{tk} & \frac{1}{tk} & -\frac{1}{r} & -\frac{1}{r} & \cdots \\ \frac{1}{tk} & \frac{1}{tk} & \frac{1}{tk} & -\frac{1}{r} & \frac{1}{tk} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} Y \end{bmatrix}$$

It may be noticed that the variance factors for the objects in the complementary designs are the same as those of the original design. It may be further noticed that for the designs L_N suggested by Mood, $t = 1$, as was pointed out before. This means that orthogonal estimates may be obtained for L_N, by adding one column of ones and only one row of zeroes to the design matrix. As Banerjee [1] has shown, this modified L_N design is identically the same as given by Kempthorne's fractional replication designs.

The above procedure will fail to furnish orthogonal estimates when r^2/λ is not an integral number, i.e. when t will not be an integral number. For these situations Banerjee has suggested the following procedure for determining orthogonal estimates. Let ξ be the least positive integer such that $(r + \xi)^2$ is divisible by $(\lambda + \xi)$. Then, if a column of ones and ξ rows of ones and η rows of zeroes in that order

are added to the design matrix to suit the estimation of the bias in a biased spring balance, the matrix [X'X] will become:

$$[X'X] = \begin{bmatrix} b+\eta+\xi & r+\xi & r+\xi & \ldots & r+\xi \\ r+\xi & r+\xi & \lambda+\xi & \ldots & \lambda+\xi \\ r+\xi & \lambda+\xi & r+\xi & \ldots & \lambda+\xi \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r+\xi & \lambda+\xi & \lambda+\xi & \ldots & r+\xi \end{bmatrix}$$

The value of the determinant $|X'X|$ is given by

$$|X'X| = (r-\lambda)^{v} (b+\xi+\eta)$$

The value of the determinant obtained after suppressing the first row and first column of $[X'X]$ is given by

$$(r-\lambda)^{v-1} \left\{ r+\lambda (v-1)+\xi v \right\} .$$

The value of the determinant obtained after suppressing the second row and second column of $[X'X]$ is given by

$$(r-\lambda)^{v-1} (b+\xi+\eta) .$$

The value of the determinant obtained after suppressing the first row and second column is given by

$$(r+\xi) (r-\lambda)^{v-1}$$

The value of the determinant $[X'X]$ obtained by suppressing the second row and third column will be equal to

$$[(r + \xi) - (\lambda + \xi)]^{v-2} [(b + \xi + \eta)(\lambda + \xi) - (r + \xi)^2]$$

Setting this expression equal to zero, we obtain

$$(b + \xi + \eta)(\lambda + \xi) = (r + \xi)^2 \quad \text{or}$$

$$(b + \xi + \eta) = \frac{(r + \xi)^2}{(\lambda + \xi)} \quad \text{or}$$

$$\eta = \frac{(r + \xi)^2}{\lambda + \xi} - (b + \xi) .$$

Hence, the value of η is determined.

Using these values the matrix $[X'X]^{-1}$ will reduce to the following form,

$$[X'X]^{-1} = \begin{bmatrix} A & -C & -C & \cdots & -C \\ -C & \frac{1}{r-\lambda} & 0 & \cdots & 0 \\ -C & 0 & \frac{1}{r-\lambda} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -C & 0 & 0 & \cdots & \frac{1}{r-\lambda} \end{bmatrix}$$

where

$$A = \frac{r + \lambda(v - 1) + \xi v}{(b + \xi + \eta)(r - \lambda)}$$

and

$$C = \frac{r + \xi}{(b + \xi + \eta)(r - \lambda)}$$

This procedure shows that a more general class of BIBD designs may be

made to furnish orthogonal estimates as indicated by the off-diagonal elements of $[X'X]^{-1}$ being zero except for the first row and column which corresponds to the bias, and not the objects.

As an illustration of the above development, Banerjee [2] presented the following example. The design matrix X in an L_7 (i.e. a symmetrical balanced incomplete block design which has $v = 7$, $b = 7$, $r = 4$, $k = 4$, and $\lambda = 2$) is given by

$$X = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix},$$

where the rows refer to the weighing operations and the columns refer to the objects $b_1, b_2, \ldots, b_7$ to be weighed.

Seven small copper pieces have been arbitrarily chosen for this illustration. The results of weighings (in grams) of the combinations of objects as given in the above design matrix are:

$$\begin{aligned} 10.76251 &= Y_1 \\ 7.83798 &= Y_2 \\ 6.11380 &= Y_3 \\ 12.07808 &= Y_4 \\ 8.90452 &= Y_5 \end{aligned}$$

$$10.63856 = Y_6$$

and

$$12.80768 = Y_7$$

Using these values, and the normal equations, the estimates of the seven objects are given by:

$$\hat{b}_1 = 5.85763$$
$$\hat{b}_2 = 3.52835$$
$$\hat{b}_3 = 1.78600$$
$$\hat{b}_4 = 1.94650$$
$$\hat{b}_5 = 1.64367$$
$$\hat{b}_6 = 1.04843$$

and

$$\hat{b}_7 = 1.47520$$

For this example the matrix [X'X] is of order 7 x 7 and is given as follows:

$$[X'X] = \begin{bmatrix} 4 & 2 & 2 & 2 & 2 & 2 & 2 \\ 2 & 4 & 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 4 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 4 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 4 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 & 4 & 2 \\ 2 & 2 & 2 & 2 & 2 & 2 & 4 \end{bmatrix}$$

The inverse of this matrix, i.e. $[X'X]^{-1}$ is

$$[X'X]^{-1} = \begin{bmatrix} \frac{7}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & \frac{7}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & -\frac{1}{16} & \frac{7}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & \frac{7}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & \frac{7}{16} & -\frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & \frac{7}{16} & -\frac{1}{16} \\ -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & \frac{7}{16} \end{bmatrix}$$

This matrix shows that the estimates furnished by this design are correlated (i.e. the off-diagonal elements are not zero).

By utilizing Banerjee's technique previously presented the BIBD may be modified to provide orthogonal estimates. For this example $t = \frac{r^2}{\lambda} - b = \frac{16}{2} - 7 = 1$, therefore one column of ones and only one row of zeroes (in that order) must be added to the BIBD design matrix (L_7) to obtain orthogonal estimates.

This modified matrix is given as follows:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

The results corresponding to these eight weighing operations are:

$$Y = \begin{bmatrix} 0.00101 \\ 10.76379 \\ 7.83990 \\ 6.11580 \\ 12.08000 \\ 8.90543 \\ 10.63998 \\ 12.80998 \end{bmatrix}$$

The first reading (0.00101) corresponds to a measurement with no objects on the scale.

The inverse of the matrix [X'X] is given as:

$$[X'X]^{-1} = \begin{bmatrix} 1 & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ -\frac{1}{4} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ -\frac{1}{4} & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ -\frac{1}{4} & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ -\frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

This matrix shows that under the modified BIBD, the estimates are uncorrelated (i.e. the off-diagonal elements are zero), except for the first row and first column which correspond to the scale bias.

The estimates of the scale bias and the seven weights under this modified set-up are:

$\hat{b}_0$ = 0.00101 (scale bias)
$\hat{b}_1$ = 5.85790
$\hat{b}_2$ = 3.52868
$\hat{b}_3$ = 1.78558
$\hat{b}_4$ = 1.94662
$\hat{b}_5$ = 1.64354
$\hat{b}_6$ = 1.04887
$\hat{b}_7$ = 1.47576

This example demonstrates how a BIBD can be modified to produce orthogonal estimates when $t = (\frac{r^2}{\lambda} - b)$ is an integer.

As an example of how a BIBD may be modified in a more general situation (whenever t is not an integer), consider the BIBD where $v = 8$, $b = 14$, $r = 7$, $k = 4$, and $\lambda = 3$. This BIBD will not, in its present form, provide orthogonal estimates. However, orthogonal estimates may be obtained as previously mentioned. For this example $t = \frac{r^2}{\lambda} - b = \frac{7}{3}$ (not an integer). The least integer ξ such that $(r + \xi)^2$ is divisible by $(\lambda + \xi)$ is 1. The value of n, $\frac{r + \xi}{\lambda + \xi} - (b + \xi)$, also equals to 1 for this example. Using these values, the modified weighing design is given as:

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

The results of the 16 weighings as specified by this design are represented by the vector Y and are given as:

$$Y = \begin{bmatrix} 18.91670 \\ 0.00101 \\ 13.12036 \\ 12.49119 \\ 10.32320 \\ 10.32997 \\ 5.79780 \\ 6.42631 \\ 8.59500 \\ 8.58860 \\ 12.07940 \\ 10.76320 \\ 11.08093 \\ 6.83861 \\ 8.15469 \\ 7.83968 \end{bmatrix}$$

The matrix [X'X], corresponding to the modified BIBD, will have the form:

$$[X'X] = \begin{bmatrix} 16 & 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 \\ 8 & 8 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\ 8 & 4 & 8 & 4 & 4 & 4 & 4 & 4 & 4 \\ 8 & 4 & 4 & 8 & 4 & 4 & 4 & 4 & 4 \\ 8 & 4 & 4 & 4 & 8 & 4 & 4 & 4 & 4 \\ 8 & 4 & 4 & 4 & 4 & 8 & 4 & 4 & 4 \\ 8 & 4 & 4 & 4 & 4 & 4 & 8 & 4 & 4 \\ 8 & 4 & 4 & 4 & 4 & 4 & 4 & 8 & 4 \\ 8 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 8 \end{bmatrix}$$

The inverse of this matrix is given as:

$$[X'X]^{-1} = \begin{bmatrix} \frac{9}{16} & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} \\ -\frac{1}{8} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{8} & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{8} & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{8} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ -\frac{1}{8} & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ -\frac{1}{8} & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ -\frac{1}{8} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 \\ -\frac{1}{8} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \end{bmatrix}$$

We notice from this matrix that the estimates furnished by this design will be uncorrelated since the off-diagonal elements are zero (except for the first row and first column which correspond to the scale bias).

Solving the normal equations we get, therefore, the following estimates of the weights of the objects:

$$\hat{b}_0 = 0.00132 \text{ (scale bias)}$$
$$\hat{b}_1 = 5.85791$$
$$\hat{b}_2 = 3.52807$$
$$\hat{b}_3 = 1.78583$$

$$\hat{b}_4 = 1.94731$$

$$\hat{b}_5 = 1.64365$$

$$\hat{b}_6 = 1.04861$$

$$\hat{b}_7 = 1.47471$$

and

$$\hat{b}_8 = 1.62960$$

CHAPTER VI

A GENERALIZED PROCEDURE FOR MODIFYING BIBD TO PROVIDE ORTHOGONAL ESTIMATES IN WEIGHING DESIGNS

For some balanced incomplete block designs the procedure developed by Banerjee [2] fails to provide orthogonal estimates when these modified BIBD are used as weighing designs. For example, in Fisher and Yates Tables [4], the BIBD with reference number 15 has $v = 10$, $b = 30$, $r = 9$, $k = 3$, and $\lambda = 2$. These values produce $n = -7$ For these situations, i.e., whenever a negative value is obtained for n, no procedure yet exists for modifying the corresponding BIBD to provide orthogonal estimates when the BIBD is used as a weighing design.

The remainder of this dissertation will be devoted to obtaining a general procedure for modifying all BIBD designs to provide orthogonal estimates. Several theorems that are directly related to this development will also be presented. In addition, a comparison of the merits of this generalized procedure with those of Banerjee's procedure will be made.

As was noted by Banerjee [2], if a column of ones and t rows of zeroes in that order are added to the BIBD matrix, the matrix [X'X] will be of the form:

$$[X'X] = \begin{bmatrix} b+t & r & r & \cdots & r \\ r & r & \lambda & \cdots & \lambda \\ r & \lambda & r & \cdots & \lambda \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ r & \lambda & \lambda & \cdots & r \end{bmatrix}$$

For this matrix Banerjee obtained the following:

1. $|X'X| = t(r - \lambda)^{v-1}\left\{r + \lambda(v - 1)\right\}$.

2. Value of the determinant obtained after suppressing the first row and the first column of $[X'X]$ is

$$(r - \lambda)^{v-1}\left\{r + \lambda\,(v - 1)\right\}$$

3. The value of the determinant obtained after suppressing the second row and the second column of $[X'X]$ is

$$(r - \lambda)^{v-2}\left[(b + t)\left\{r + \lambda\,(v - 2)\right\} - r^2\,(v - 1)\right].$$

4. The value of the determinant obtained after suppressing the first row and the second column of $[X'X]$ is

$$r(r - \lambda)^{v-1}.$$

5. The value of the determinant obtained after suppressing the second row and the third column of $[X'X]$ is

$$(r - \lambda)^{v-2}\left\{\lambda(b + t) - r^2\right\}.$$

Banerjee suggested that 5 above (which would correspond to the off-diagonal elements in $[X'X]^{-1}$) should be identically zero (except for the

first row and first column) for the estimates to be orthogonal, i.e.

$$(r - \lambda)^{v-2}\left\{\lambda(b + t) - r^2\right\} = 0$$

or

$$t = \frac{r^2}{\lambda} - b$$

Therefore Banerjee suggested adding one column of ones and t rows of zeroes in that order to the BIBD to obtain orthogonal estimates. For this procedure t must always be positive. This will be shown in the following lemma:

Lemma: For balanced incomplete block designs characterized by v, b, r, k, and λ, and the two well known identities $bk = vr$ and $\lambda(v - 1) = r(k - 1)$; the quantity $t = \frac{r^2}{\lambda} - b$ will always be positive.

We wish to show proof that $t = \frac{r^2}{\lambda} - b > 0$, or substituting for r and λ

$$\frac{\left(\frac{bk}{v}\right)^2}{\frac{\left(\frac{bk}{v}\right)(k - 1)}{v - 1}} - b > 0$$

$$\frac{bk\,(v - 1)}{v\,(k - 1)} - b \geq 0$$

$$\frac{bkv - bk - bvk + bv}{v\,(k - 1)} > 0$$

$$\frac{b\,(v - k)}{v\,(k - 1)} > 0$$

The quantity $\frac{b\,(v - k)}{v\,(k - 1)}$ will be greater than zero if $v > k$ and

if $k > 1$. These conditions hold for BIBD's. Therefore the lemma is shown to be valid, i.e. $t = \frac{r^2}{\lambda} - b$ will always be positive.

The above lemma guarantees t to be positive. However t does not necessarily have to be an integer. For these situations Banerjee, as was noted in Chapter V, suggested adding one column of ones, n rows of zeroes, and ξ rows of ones in that order to the BIBD to provide orthogonal estimates, where $n = \frac{(r + \xi)^2}{(\lambda + \xi)} - (b + \xi)$ and ξ is the least positive integer such that

$$\frac{(r + \xi)^2}{(\lambda + \xi)} - (b + \xi)$$

is an integer value. This procedure, however, does not guarantee n to be positive and in fact we have shown at the beginning of this chapter where n is negative. We shall now present a method for modifying all BIBD's to produce orthogonal estimates in weighing designs.

If Banerjee's method is modified by adding one column of ones and only one row where the first element of the row is $\sqrt{t}$ and all other elements are zeroes, we obtain the modified BIBD design as follows:

$$X = \left[\begin{array}{c|ccccc} \sqrt{t} & 0 & 0 & 0 & \cdots & 0 \\ \hline 1 & & & & & \\ 1 & & & & & \\ 1 & & & X_0 & & \\ 1 & & & & & \\ 1 & & & & & \\ 1 & & & & & \end{array}\right]_{(b+1)x(v+1)}$$

where X_0 is the original design matrix given by the balanced incomplete block design. This is consistent with the assumption as made with respect to a spring balance design. We would make only one weighing operation on empty pans and multiply the corresponding reading on the scale by $\sqrt{t}$. It may, however, be remarked here that the structure of the variance-covariance matrix for e will also undergo a corresponding change, i.e. under Banerjee's procedure $E(ee') = \sigma^2 I_n$ whereas under this new procedure $E[ee'] = V\sigma^2$.

Using this modification, the matrix [X'X] becomes:

$$[X'X] = \begin{bmatrix} b+t & r & r & \cdots & r \\ r & r & \lambda & \cdots & \lambda \\ r & \lambda & r & \cdots & \lambda \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ r & \lambda & \lambda & \cdots & r \end{bmatrix}_{(v+1)x(v+1)}$$

which is identically the same as that when t rows were added instead

of the row, ($\sqrt{t}$ 0 0 0 . . . 0). Since the [X'X] matrix is the same, $[X'X]^{-1}$ will also be the same and therefore will provide orthogonal estimates as before. The advantage of adding only one row ($\sqrt{t}$ 0 0 . . . 0) to the BIBD is that less weighing operations are required and that this method will provide orthogonal estimates for <u>all</u> BIBD's since t does not have to be an integer. For example, the BIBD with reference no. 15 in Fisher and Yates' tables [4] is given by

$$v = 10,\ b = 30,\ r = 9,\ k = 3,\ \text{and}\ \lambda = 2.$$

Previously for this BIBD there was no method for modifying this design to provide orthogonal estimates. Using the new technique we get

$$t = \frac{r^2}{\lambda} - b = \frac{81}{2} - 30 = \frac{21}{2}$$

We would now add a column of ones and a row of ($\sqrt{\frac{21}{2}}$ 0 0 . . . 0) in that order to the original BIBD and obtain the matrix [X'X] as follows:

$$[X'X] = \begin{bmatrix} \frac{81}{2} & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 \\ 9 & 9 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 9 & 2 & 9 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 9 & 2 & 2 & 9 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 9 & 2 & 2 & 2 & 9 & 2 & 2 & 2 & 2 & 2 & 2 \\ 9 & 2 & 2 & 2 & 2 & 9 & 2 & 2 & 2 & 2 & 2 \\ 9 & 2 & 2 & 2 & 2 & 2 & 9 & 2 & 2 & 2 & 2 \\ 9 & 2 & 2 & 2 & 2 & 2 & 2 & 9 & 2 & 2 & 2 \\ 9 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 9 & 2 & 2 \\ 9 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 9 & 2 \\ 9 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 9 \end{bmatrix}$$

The inverse of this matrix will be given as:

$$[X'X]^{-1} = \begin{bmatrix} \frac{2}{21} & -\frac{2}{63} & -\frac{2}{63} & -\frac{2}{63} & -\frac{2}{63} & -\frac{2}{63} & -\frac{2}{63} & -\frac{2}{63} & -\frac{2}{63} & -\frac{2}{63} & -\frac{2}{63} \\ -\frac{2}{63} & \frac{1}{7} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{2}{63} & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{2}{63} & 0 & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{2}{63} & 0 & 0 & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{2}{63} & 0 & 0 & 0 & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 & 0 \\ -\frac{2}{63} & 0 & 0 & 0 & 0 & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 \\ -\frac{2}{63} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{7} & 0 & 0 & 0 \\ -\frac{2}{63} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{7} & 0 & 0 \\ -\frac{2}{63} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{7} & 0 \\ -\frac{2}{53} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{7} \end{bmatrix}$$

This matrix, $[X'X]^{-1}$, shows that under the new procedure, the modified balanced incomplete block designs do provide orthogonal estimates when used as weighing desions (i.e. the estimates will be orthogonal since the off-diagonal elements of $[X'X]^{-1}$, except in the first row and first column which correspond to the estimate of the scale bias, are equal to zero).

It should be noted here that the basic difference between

Banerjee's method and the new method is that Banerjee's method requires t additional weighing operations whenever t is an integer and $\xi + \eta$ additional weighings whenever t is not an integer whereas the new method requires only one additional weighing operation. In addition, the variance-covariance matrix for e will be slightly different for Banerjee's method and the new method. Several questions arise concerning these differences, such as, "What is the relative efficiency of the new procedure as compared to that of Banerjee's procedure?" "What implications arise since the variance-covariance matrices are different?" These questions will be addressed in the following.

Under Banerjee's previous method (adding t rows) the weighing design model had the form $Y = X\beta + e$ with $E[e] = 0$ and $E[ee'] = \sigma^2 I_n$. Under the new method (adding one row, $\sqrt{t}\ 0\ 0\ \ldots\ 0$) the model becomes $Y = X\beta + e$ with $E[e] = 0$ and $E[ee'] = V\sigma^2$ where V has the form:

$$V = \begin{bmatrix} t & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}$$

Since, as noted above, the covariances have different forms, it is appropriate to compare the relative efficiencies between the two methods. Relative efficiency is defined in this case as the ratio of the reciprocals of the variances.

First to aid us in this development, the following theorem

will be proven:

Theorem: In the estimation procedure under consideration, when $E[ee'] = V\sigma^2$, *the covariance of the estimators,* $\mathrm{Cov}(\hat{\beta})$, *obtained by the least squares procedure is identical to the covariance of the estimators obtained by the maximum likelihood procedure even when* X *is not square.*

(We know that for the estimation procedure under consideration, when the design matrix X is square the covariance of the estimators obtained by the least squares procedure is identical to those obtained by the maximum likelihood procedure.)

Proof: We wish to show that L.S. $\mathrm{Cov}(\hat{\beta})$ = M.L. $\mathrm{Cov}(\hat{\beta})$

or

$$I = \text{M.L. Cov}(\hat{\beta})\ [\text{L.S. Cov}(\hat{\beta})]^{-1}$$

We shall now determine the $\mathrm{Cov}(\hat{\beta})$ by the least square method and also by the maximum likelihood method.

1. By Least Squares.

$$\hat{\beta} = (X'X)^{-1} X'Y$$

$$\mathrm{Cov}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$

$$= E\left\{[(X'X)^{-1} X'Y - \beta][(X'X)^{-1} X'Y - \beta]'\right\}$$

$$= E\left\{[(X'X)^{-1} X'(X\beta + e) - \beta][(X'X)^{-1} X'(X\beta + e) - \beta]'\right\}$$

$$= E\left\{[(X'X)^{-1} X'e][(X'X)^{-1} X'e]'\right\}$$

$$= E[(X'X)^{-1} X'e\, e'X\, (X'X)^{-1}]$$

or finally

$$= \sigma^2 [(X'X)^{-1} X'VX\, (X'X)^{-1}]$$

2. Maximum likelihood approach.

The maximum likelihood estimator of β (assuming e is normally distributed) is $\hat{\beta} = (X'V^{-1}X)^{-1} X'V^{-1}Y$

$$Cov(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$

$$= E\left\{[(X'V^{-1}X)^{-1} X'V^{-1} Y - \beta] [(X'V^{-1}X)^{-1} X'V^{-1} Y - \beta]'\right\}$$

$$= E\left\{[(X'V^{-1}X)^{-1} X'V^{-1} (X\beta + e) - \beta]\right.$$
$$\left.[(X'V^{-1}X)^{-1} X'V^{-1} (X\beta + e) - \beta]'\right\}$$

$$= E\left\{[(X'V^{-1}X)^{-1} X'V^{-1} X\beta + (X'V^{-1}X)^{-1} X'V^{-1} e - \beta]\right.$$
$$\left.[(X'V^{-1}X)^{-1} X'V^{-1} X\beta + (X'V^{-1}X)^{-1} X'V^{-1}e - \beta]'\right\}$$

$$= E\left\{[(X'V^{-1}X)^{-1} X'V^{-1}e] [(X'V^{-1}X)^{-1} X'V^{-1}e]'\right\}$$

$$= E\,[(X'V^{-1}X)^{-1} X'V^{-1} ee'V^{-1}X\,(X'V^{-1}X)^{-1}]$$

$$= \sigma^2\,(X^{-1} VX'^{-1})$$

or finally $$= \sigma^2\,(X'V^{-1}X)^{-1}$$

We would now like to show that

$$\sigma^2(X'V^{-1}X)^{-1} = \sigma^2\,[(X'X)^{-1} X'VX(X'X)^{-1}]$$

or

$$I = [(X'X)^{-1} X'VX(X'X)^{-1}]\,[(X'V^{-1}X)]$$

Under the new procedure the matrix X is determined by adding one column of ones and one row, $\sqrt{t}$ 0 0 . . . 0, in that order to the BIBD design where $t = \frac{r^2}{\lambda} - b$. i.e.

$$X = \left[\begin{array}{c|ccc} \sqrt{t} & 0 & 0 \ \ldots & 0 \\ \hline 1 & & & \\ 1 & & & \\ \vdots & & X_0 & \\ 1 & & & \end{array}\right],$$

where X_0 corresponds to the BIBD used as a weighing design. Also

$$X' = \left[\begin{array}{c|ccc} \sqrt{t} & 1 & 1 \ \ldots & 1 \\ \hline 0 & & & \\ 0 & & & \\ \vdots & & X_0' & \\ 0 & & & \end{array}\right]$$

Using this notation, the matrix X'X will have the following form:

$$X'X = \begin{bmatrix} b+t & r & r & \ldots & r \\ r & r & \lambda & \ldots & \lambda \\ r & \lambda & r & \ldots & \lambda \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r & \lambda & \lambda & \ldots & r \end{bmatrix}.$$

The inverse of this matrix will be given as:

$$[X'X]^{-1} = \begin{bmatrix} \frac{1}{t} & -a & -a & \dots & -a \\ -a & b & 0 & \dots & 0 \\ -a & 0 & b & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ -a & 0 & 0 & \dots & b \end{bmatrix}$$

where $a = \frac{1}{tk}$ and $b = \frac{1}{r - \lambda} = \frac{r}{tk\lambda}$.

The matrix V will be as follows:

$$V = \left[\begin{array}{c|c} t & 0 \\ \hline 0 & I \end{array}\right],$$

where 0 and I have the appropriate dimension. Writing V in another form,

$$V = [I + M] = \left[I + \left[\begin{array}{c|c} t - 1 & 0 \\ \hline 0 & 0 \end{array}\right] \right],$$

where

$$M = \left[\begin{array}{c|c} t - 1 & 0 \\ \hline 0 & 0 \end{array}\right].$$

The inverse of V is given as:

$$V^{-1} = \left[\begin{array}{c|c} \frac{1}{t} & 0 \\ \hline 0 & I \end{array}\right]$$

or alternatively,

$$V^{-1} = I + \left[\begin{array}{c|c} \frac{-(t-1)}{t} & 0 \\ \hline 0 & 0 \end{array}\right]$$

$$= I - \frac{1}{t} M .$$

Using this notation, the maximum likelihood covariance, $(X'V^{-1}X)^{-1}$, becomes

$$(X'V^{-1}X)^{-1} = [X'(I - \tfrac{1}{t} M)X]^{-1}$$

or

$$(X'V^{-1}X)^{-1} = (X'X - \tfrac{1}{t} X'MX)^{-1}$$

The least squares covariance becomes

$$(X'X)^{-1} X'VX(X'X)^{-1} = (X'X)^{-1} X'(I + M) X(X'X)^{-1}$$

$$= (X'X)^{-1} (X'X)(X'X)^{-1} + (X'X)^{-1} (X'MX)(X'X)^{-1}$$

and finally,

$$= (X'X)^{-1} + (X'X)^{-1} (X'MX)(X'X)^{-1}$$

To prove equality we now must show that

$$I = [(X'X)^{-1} + (X'X)^{-1} X'MX(X'X)^{-1}] [X'X - \frac{1}{t} X'MX]$$

$$= [(X'X)^{-1} (X'X) - \frac{1}{t} (X'X)^{-1} X'MX + (X'X)^{-1} X'MX(X'X)^{-1}$$

$$(X'X) - \frac{1}{t} (X'X)^{-1} X'MX(X'X)^{-1} X'MX]$$

$$= I + [- \frac{1}{t} (X'X)^{-1} X'MX + (X'X)^{-1} X'MX - \frac{1}{t} (X'X)^{-1} X'MX$$

$$(X'X)^{-1} X'MX].$$

We now need only show that the term in brackets is zero, i.e.

$$- \frac{1}{t} (X'X)^{-1} X'MX + (X'X)^{-1} X'MX - \frac{1}{t} (X'X)^{-1} X'MX (X'X)^{-1} X'MX = 0$$

To simplify this expression, we obtain X'MX as:

$$X'MX = \left[\begin{array}{c|c|c} \sqrt{t} \mid 1 \ldots 1 & t-1 \mid 0 \ldots 0 & \sqrt{t} \mid 0 \ldots 0 \\ \hline 0 & 0 & 1 \\ \vdots \mid X'_0 & \vdots \mid 0 & \vdots \mid X_0 \\ 0 & 0 & 1 \end{array}\right]$$

$$= \left[\begin{array}{c|c|c} \sqrt{t}\,(t-1) \mid 0 \ldots 0 & \sqrt{t} \mid 0 \ldots 0 \\ \hline 0 & 1 \\ \vdots \mid 0 & \vdots \mid X_0 \\ 0 & 1 \end{array}\right],$$

and finally,

$$(X'MX) = \left[\begin{array}{c|c} t(t-1) & 0 \ldots 0 \\ \hline \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} & \begin{matrix} \\ 0 \\ \\ \end{matrix} \end{array}\right]$$

Now,

$$(X'X)^{-1}\, X'MX = \left[\begin{array}{c|ccc} \frac{1}{t} & -a & -a \ \ldots & -a \\ \hline -a & b & 0 \ \ldots & 0 \\ -a & 0 & b \ \ldots & 0 \\ \vdots & \vdots & \vdots \ \ddots & \vdots \\ -a & 0 & 0 \ \ldots & b \end{array}\right] \left[\begin{array}{c|c} t(t-1) & 0 \ldots 0 \\ \hline \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} & \begin{matrix} \\ 0 \\ \\ \end{matrix} \end{array}\right]$$

where $a = \frac{1}{tk}$ and $b = \frac{1}{r - \lambda} = \frac{r}{tk\lambda}$.

Or,

$$(X'X)^{-1}\, X'MX = \left[\begin{array}{c|c} t-1 & 0\ 0 \ldots 0 \\ \hline \begin{matrix} -at(t-1) \\ -at(t-1) \\ \\ \\ -at(t-1) \end{matrix} & \begin{matrix} \\ \\ 0 \\ \\ \end{matrix} \end{array}\right] \qquad (1)$$

Further

$$[(X'X)^{-1}X'MX][(X'X)^{-1}X'MX] = \left[\begin{array}{c|c} (t-1) & 0 \ldots 0 \\ \hline -at(t-1) & \\ -at(t-1) & \\ \vdots & 0 \\ -at(t-1) & \end{array}\right]\left[\begin{array}{c|c} (t-1) & 0 \ldots 0 \\ \hline -at(t-1) & \\ -at(t-1) & \\ \vdots & 0 \\ -at(t-1) & \end{array}\right]$$

or

$$(X'X)^{-1}X'MX(X'X)^{-1}X'MX = \left[\begin{array}{c|c} (t-1)^2 & 0 \ldots 0 \\ \hline -at(t-1)^2 & \\ -at(t-1)^2 & \\ \vdots & 0 \\ -at(t-1)^2 & \end{array}\right], \qquad (2)$$

Substituting equations (1) and (2) into our original expression we get

$$-\frac{1}{t}\left[\begin{array}{c|c} t-1 & 0 \ldots 0 \\ \hline -at(t-1) & \\ -at(t-1) & \\ \vdots & 0 \\ -at(t-1) & \end{array}\right] + \left[\begin{array}{c|c} t-1 & 0 \ldots 0 \\ \hline -at(t-1) & \\ -at(t-1) & \\ \vdots & 0 \\ -at(t-1) & \end{array}\right] - \frac{1}{t}\left[\begin{array}{c|c} (t-1)^2 & 0 \ldots 0 \\ \hline -at(t-1)^2 & \\ -at(t-1)^2 & \\ \vdots & 0 \\ -at(t-1)^2 & \end{array}\right]$$

and we need only to show that this expression reduces to a zero matrix.

On simplification we obtain the following expression

$$\left[\begin{array}{c|c} \frac{-(t-1)}{t} + (t-1) - \frac{(t-1)^2}{t} & 0 \ldots 0 \\ \hline \frac{at(t-1)}{t} - at(t-1) + \frac{at(t-1)^2}{t} & \\ \frac{at(t-1)}{t} - at(t-1) + \frac{at(t-1)^2}{t} & 0 \\ \vdots & \\ \frac{at(t-1)}{t} - at(t-1) + \frac{at(t-1)^2}{t} & \end{array}\right] \qquad (3)$$

Reducing these expressions we get

$$\frac{-(t-1)}{t} + (t-1) - \frac{(t-1)^2}{t} = \frac{-t+1+t^2-t-t^2+2t-1}{t} = \frac{0}{t} = 0$$

and

$$\frac{at(t-1)}{t} - at(t-1) + \frac{at(t-1)^2}{t} = \frac{at(t-1) - at^2(t-1) + at(t-1)^2}{t} =$$

$$= \frac{at(t-1)\,(1-t+t-1)}{t} = \frac{at(t-1)\,(0)}{t} = \frac{0}{t} = 0 \,.$$

Therefore expression (3) in fact reduces to the zero matrix and the theorem is proven.

As an example of this theorem consider the BIBD given by $v = 4$, $b = 6$, $r = 3$, $k = 2$, and $\lambda = 1$. For this example the design matrix X, the transpose of this matrix, X', and the matrices V and V^{-1} are given as follows:

$$X = \begin{bmatrix} \sqrt{3} & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix},$$

$$X' = \begin{bmatrix} \sqrt{3} & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix},$$

$$V = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

and

$$V^{-1} = \begin{bmatrix} \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The covariance obtained by the maximum likelihood method is given by

$$(X'V^{-1}X)^{-1} = \left[\begin{bmatrix} \sqrt{3} & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 0_{1x6} \\ \hline 0_{6x1} & I_{6x6} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix} \right]^{-1}$$

$$= \left[\begin{bmatrix} \frac{\sqrt{3}}{3} & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix} \right]^{-1}$$

$$(X'V^{-1}X)^{-1} = \begin{bmatrix} 7 & 3 & 3 & 3 & 3 \\ 3 & 3 & 1 & 1 & 1 \\ 3 & 1 & 3 & 1 & 1 \\ 3 & 1 & 1 & 3 & 1 \\ 3 & 1 & 1 & 1 & 3 \end{bmatrix}^{-1}$$

Finally the maximum likelihood covariance is given by:

$$(X'V^{-1}X)^{-1} = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{2}{3} & \frac{1}{6} & \frac{1}{5} & \frac{1}{6} \\ -\frac{1}{2} & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & \frac{1}{6} \\ -\frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ -\frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \end{bmatrix} \qquad (4)$$

The least squares covariance is given by $[(X'X)^{-1}X'VX(X'X)^{-1}]$. To evaluate this expression the matrix $(X'X)^{-1}$ is determined as follows:

$$(X'X)^{-1} = \begin{bmatrix} \begin{bmatrix} \sqrt{3} & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} & \begin{bmatrix} \sqrt{3} & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix} \end{bmatrix}^{-1}$$

On simplification we obtain:

$$(X'X)^{-1} = \begin{bmatrix} 9 & 3 & 3 & 3 & 3 \\ 3 & 3 & 1 & 1 & 1 \\ 3 & 1 & 3 & 1 & 1 \\ 3 & 1 & 1 & 3 & 1 \\ 3 & 1 & 1 & 1 & 3 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{2} & 0 & 0 & 0 \\ -\frac{1}{6} & 0 & \frac{1}{2} & 0 & 0 \\ -\frac{1}{6} & 0 & 0 & \frac{1}{2} & 0 \\ -\frac{1}{6} & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

The matrix $X'VX$ is determined as follows:

$$X'VX = \begin{bmatrix} \sqrt{3} & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix},$$

$$X'VX = \begin{bmatrix} 3\sqrt{3} & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix},$$

$$= \begin{bmatrix} 15 & 3 & 3 & 3 & 3 \\ 3 & 3 & 1 & 1 & 1 \\ 3 & 1 & 3 & 1 & 1 \\ 3 & 1 & 1 & 3 & 1 \\ 3 & 1 & 1 & 1 & 3 \end{bmatrix}.$$

Therefore the expression $(X'X)^{-1}X'VX(X'X)^{-1}$ is given as:

$$\begin{bmatrix} \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{2} & 0 & 0 & 0 \\ -\frac{1}{6} & 0 & \frac{1}{2} & 0 & 0 \\ -\frac{1}{6} & 0 & 0 & \frac{1}{2} & 0 \\ -\frac{1}{6} & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 15 & 3 & 3 & 3 & 3 \\ 3 & 3 & 1 & 1 & 1 \\ 3 & 1 & 3 & 1 & 1 \\ 3 & 1 & 1 & 3 & 1 \\ 3 & 1 & 1 & 1 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{2} & 0 & 0 & 0 \\ -\frac{1}{6} & 0 & \frac{1}{2} & 0 & 0 \\ -\frac{1}{6} & 0 & 0 & \frac{1}{2} & 0 \\ -\frac{1}{6} & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}.$$

Or upon multiplication,

$$(X'X)^{-1}X'VX(X'X)^{-1} = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{2}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ -\frac{1}{2} & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & \frac{1}{6} \\ -\frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ -\frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \end{bmatrix} \qquad (5)$$

We can see that expression (5) is identical to expression (4). Therefore for the situation under consideration, the covariance obtained by maximum likelihood method is the same as that obtained by the least squares method even when X is not square.

Having established that the covariance matrix obtained by either the least squares approach or the maximum likelihood approach are identical for BIBD designs modified to be weighing designs to produce orthogonal estimates, we can proceed to compare the relative efficiencies of Banerjee's previous method with the new method. We would like to determine the covariance matrix, $\text{Cov}(\hat{\beta})$, for both Banerjee's method and the new method. $\text{Cov}(\hat{\beta})$ under Banerjee's method was given in Chapter V as

$$\text{Cov}(\hat{\beta}) = \begin{bmatrix} \frac{1}{t} & -\frac{1}{tk} & -\frac{1}{tk} & \cdot & \cdot & \cdot & -\frac{1}{tk} \\ -\frac{1}{tk} & \frac{r}{tk\lambda} & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{tk} & 0 & \frac{r}{tk\lambda} & 0 & 0 & 0 & 0 \\ \cdot & 0 & 0 & \frac{r}{tk\lambda} & 0 & 0 & 0 \\ \cdot & 0 & 0 & 0 & \frac{r}{tk\lambda} & 0 & 0 \\ \cdot & 0 & 0 & 0 & 0 & \frac{r}{tk\lambda} & 0 \\ -\frac{1}{tk} & 0 & 0 & 0 & 0 & 0 & \frac{r}{tk\lambda} \end{bmatrix} \sigma^2$$

for those BIBD's that could be modified, i.e. either t was integer or a positive value for n was obtainable.

For the new procedure (adding on row, $\sqrt{t}$ 0 0 ... 0) the $\text{Cov}(\hat{\beta})$ as given by the method of maximum likelihood (which was previously shown to also be equal to that derived by the method of least squares) is:

$$\text{Cov}(\hat{\beta}) = (X'V^{-1}X)^{-1}\sigma^2 \qquad (6)$$

where V^{-1} has the form

$$V^{-1} = I + \left[\begin{array}{c|c} \frac{-(t-1)}{t} & 0 \\ \hline 0 \quad \cdot & 0 \end{array}\right]$$

Substituting this expression into equation (6) we get

$$\text{Cov}(\hat{\beta}) = \left[X' \left[I + \left[\begin{array}{c|c} \frac{-(t-1)}{t} & 0 \\ \hline 0 & 0 \end{array} \right] \right] X \right]^{-1} \sigma^2$$

or

$$= [X'X + (-\tfrac{1}{t})\, X'MX]^{-1}\sigma^2 ,$$

where

$$M = \left[\begin{array}{c|c} t-1 & 0 \\ \hline 0 & 0 \end{array} \right] .$$

We see that

$$X'MX = \left[\begin{array}{c|ccc} \sqrt{t} & 1 & 1 \ \ldots & 1 \\ \hline 0 & & & \\ 0 & & & \\ \vdots & & X_0' & \\ 0 & & & \end{array} \right] \left[\begin{array}{c|c} (t-1) & 0 \\ \hline 0 & 0 \end{array} \right] \left[\begin{array}{c|ccc} \sqrt{t} & 0 & 0 \ \ldots & 0 \\ \hline 1 & & & \\ 1 & & & \\ \vdots & & X_0 & \\ 1 & & & \end{array} \right] ,$$

where X_0 and X_0' are the original BIBD and transpose of the original BIBD respectively.

Simplifying we obtain

$$X'MX = \left[\begin{array}{c|c} \sqrt{t}(t-1) & 0 \\ \hline 0 & 0 \end{array} \right] \left[\begin{array}{c|c} \sqrt{t} & 0 \\ \hline 1 & X_0 \end{array} \right]$$

or

$$X'MX = \left[\begin{array}{c|c} t(t-1) & 0 \\ \hline 0 & 0 \end{array}\right].$$

Substituting back into the original expression, we get

$$\mathrm{Cov}(\hat{\beta}) = \left[X'X - \left[\begin{array}{c|c} t-1 & 0 \\ \hline 0 & 0 \end{array}\right]\right]^{-1}\sigma^2.$$

Writing this in another form and letting $Z = X'X$, we obtain

$$\mathrm{Cov}(\hat{\beta}) = \left[(X'X) + I\left[\begin{array}{c|c} -(t-1) & 0 \\ \hline 0 & 0 \end{array}\right]\right]^{-1}\sigma^2 = [Z + IW]^{-1}\sigma^2.$$

Using the formula,

$$(Z + UW)^{-1} = Z^{-1} - Z^{-1}U(I_m + WZ^{-1}U)^{-1}WZ^{-1},$$

we have for

$$\mathrm{Cov}(\hat{\beta}) = (X'V^{-1}X)^{-1}\sigma^2$$

$$= (X'X)^{-1} - (X'X)^{-1}\left[I + \left[\begin{array}{c|ccc} \frac{(t-1)}{-t} & a(t-1) & \cdots & a(t-1) \\ \hline 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{array}\right]^{-1}\left[\begin{array}{c|ccc} \frac{(t-1)}{-t} & a(t-1) & \cdots & a(t-1) \\ \hline 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{array}\right]\right],$$

or,

$$= (X'X)^{-1} - (X'X)^{-1} \left[\begin{array}{c|ccc} \frac{1}{t} & a(t-1) & \cdots & a(t-1) \\ \hline 0 & & & \\ \vdots & & I & \\ 0 & & & \end{array}\right]^{-1} \left[\begin{array}{c|ccc} \frac{(t-1)}{-t} & a(t-1) & \cdots & a(t-1) \\ \hline 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots\vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{array}\right].$$

To invert the matrix

$$\left[\begin{array}{c|ccc} \frac{1}{t} & a(t-1) & \cdots & a(t-1) \\ \hline 0 & & & \\ 0 & & I & \\ \vdots & & & \\ 0 & & & \end{array}\right],$$

we use the result that the inverse of a matrix of the form

$$\left[\begin{array}{c|ccccc} d & e & f & g & h & \cdots & l \\ \hline 0 & & & & & & \\ 0 & & & & I & & \\ \vdots & & & & & & \\ 0 & & & & & & \end{array}\right],$$

is given by

$$\left[\begin{array}{c|ccc} \frac{1}{d} & \frac{-e}{d} & \frac{-f}{d} & \frac{-n}{d} \quad \cdots \quad \frac{-1}{d} \\ \hline 0 & & & \\ 0 & & I & \\ \vdots & & & \\ 0 & & & \end{array}\right]$$

to obtain

$$\left[\begin{array}{c|ccc} \frac{1}{t} & a(t-1) & \cdots & a(t-1) \\ \hline 0 & & & \\ 0 & & I & \\ \vdots & & & \\ 0 & & & \end{array}\right]^{-1} =$$

$$\left[\begin{array}{c|ccc} t & -a(t-1)(t) & -a(t-1)(t) & \cdots \quad -a(t-1)(t) \\ \hline 0 & & & \\ 0 & & I & \\ \vdots & & & \\ 0 & & & \end{array}\right].$$

Substituting into the original expression we get for $\mathrm{Cov}(\hat{\beta})$:

$$(X'X)^{-1} - (X'X)^{-1} \left[\begin{array}{c|cc} t & -at(t-1) & -at(t-1) \\ \hline 0 & & \\ \vdots & \multicolumn{2}{c}{I} \\ 0 & & \end{array}\right] \left[\begin{array}{c|ccc} (t-1) & a(t-1) & \cdots & a(t-1) \\ -t & & & \\ \hline 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{array}\right]$$

$$= (X'X)^{-1} - (X'X)^{-1} \left[\begin{array}{c|ccc} -(t-1) & at(t-1) & at(t-1) & \cdots & at(t-1) \\ \hline 0 & 0 & 0 & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & 0 \end{array}\right].$$

To further simplify this expression we obtain $(X'X)^{-1}$ as:

$$(X'X)^{-1} = \left[\begin{array}{c|ccc} \frac{1}{t} & -a & \cdots & -a \\ \hline -a & & & \\ \vdots & \multicolumn{3}{c}{bI} \\ -a & & & \end{array}\right]$$

where $a = \frac{1}{tk}$ and $b = \frac{1}{r - \lambda} = \frac{r}{tk\lambda}$.

Substituting this matrix for $[X'X]^{-1}$, we get $Cov(\hat{\beta}) =$

$$\left[\begin{array}{c|ccc} \frac{1}{t} & -a & \cdots & -a \\ \hline -a & & & \\ \vdots & \multicolumn{3}{c}{bI} \\ -a & & & \end{array}\right] - \left[\begin{array}{c|ccc} \frac{1}{t} & -a & \cdots & -a \\ \hline -a & & & \\ \vdots & \multicolumn{3}{c}{bI} \\ -a & & & \end{array}\right] \left[\begin{array}{c|ccc} -(t-1) & at(t-1) & \cdots & at(t-1) \\ \hline 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 0 & 0 & & 0 \end{array}\right]$$

On further simplification,

$$\mathrm{Cov}(\hat{\beta}) = \left[\begin{array}{c|ccc} \frac{1}{t} & -a & \cdots & -a \\ \hline -a & & & \\ \vdots & & bI & \\ -a & & & \end{array}\right] - \begin{bmatrix} \frac{-(t-1)}{t} & a(t-1) & \cdots & a(t-1) \\ -a(t-1) & -a^2t(t-1) & \cdots & -a^2t(t-1) \\ \vdots & \vdots & \vdots\!\vdots & \vdots \\ -a(t-1) & -a^2t(t-1) & \cdots & -a^2t(t-1) \end{bmatrix}\sigma^2$$

Finally,

$$\mathrm{Cov}(\hat{\beta}) = \begin{bmatrix} 1 & -at & \cdots & -at \\ -at & b+a^2t(t-1) & \cdots & a^2t(t-1) \\ \vdots & \vdots & \vdots\!\vdots & \vdots \\ -at & a^2t(t-1) & \cdots & b+a^2t(t-1) \end{bmatrix}\sigma^2$$

or substituting for a and b we get

$$\mathrm{Cov}(\hat{\beta}) = \begin{bmatrix} 1 & -\frac{1}{k} & \cdots & -\frac{1}{k} \\ -\frac{1}{k} & \frac{rk+\lambda(t-1)}{tk^2\lambda} & \cdots & \frac{t-1}{tk^2} \\ \vdots & \vdots & \vdots\!\vdots & \vdots \\ -\frac{1}{k} & \frac{t-1}{tk^2} & \cdots & \frac{rk+\lambda(t-1)}{tk^2\lambda} \end{bmatrix}\sigma^2 .$$

This expression represents the $\mathrm{Cov}(\hat{\beta})$ for the situation where a BIBD is modified to produce orthogonal estimates in weighing designs by adding one column of ones and one row where the first element in the row is

$\sqrt{t}$ and all other elements are zeroes.

Since the diagonal elements of the covariance matrix represent the variance factors we have for Banerjee's previous method the variances of the estimates for t additional weighings equal to $\frac{r}{tk\lambda}$ whereas in the new method the variance for the estimates for only one additional weighing ($\sqrt{t}$ 0 0 ... 0) is given by $\frac{rk + \lambda(t-1)}{tk^2\lambda}$. Using these values we obtain the relative efficiency (defined as the ratio of the reciprocal of the variancies) as follows:

$$\text{Relative Efficiency} = \frac{\frac{1}{\frac{rk+\lambda(t-1)}{tk^2\lambda}}}{\frac{1}{\frac{r}{tk\lambda}}}$$

$$= \frac{rk}{rk+\lambda(t-1)}$$

Using this definition, the relative efficiencies of the new method as compared to Banerjee's method for those BIBD listed in Fisher and Yates Tables [4] for which t is an integer is given in Table I. For those BIBD were t is not an integer the relative efficiencies are presented in Table II.

TABLE I

RELATIVE EFFICIENCY OF THE NEW METHOD AS COMPARED TO BANERJEE'S METHOD WHEN T IS AN INTEGER

Reference Number	$t^* = \frac{r^2}{\lambda} - b$	Relative Efficiency
3	3	0.857[illegible]
5	2	0.9474
7	14	0.6176
13	14	0.7111
17	3	0.9000
18	24	0.5400
20	2	0.9643
21	18	0.6792
24	12	0.8308
25	30	0.5085
29	18	0.7463
36	22	0.7237
42	15	0.6316
43	6	0.8276
44	3	0.9231
46	4	0.9000
49	3	0.9600
61	7	0.7447
64	5	0.8596
66	14	0.7636
67	12	0.8167
71	20	0.4412
72	6	0.7500
75	5	0.8182
77	10	0.7692
78	3	0.9375
81	2	0.9783
82	6	0.9091
83	5	0.9375
85	1	0.9615

*t indicates the number of additional weighings required by Banerjee's method. The new method requires only one additional weighing.

TABLE II

RELATIVE EFFICIENCY OF THE NEW METHOD AS COMPARED TO BANERJEE'S METHOD WHEN t IS NOT AN INTEGER

Reference Number	$t = \frac{r^2}{\lambda} - b$	η^*	ξ^{**}	$(\eta+\xi)^{***}$	Relative Efficiency
1	2.50	1	1	2	0.8333
2	1.50	0	1	1	0.9615
6	2.33	1	1	2	0.8753
9	1.33	0	1	1	0.9802
11	3.33	0	2	2	0.8207
16	2.25	1	1	2	0.9000
19	1.25	0	1	1	0.9878

* η denotes the number of additional weighings with no objects on the scale.

** ξ denotes the number of additional weighings with all the objects on the scale.

*** $\eta+\xi$ indicate the total number of additional weighings required to provide orthogonal estimates by Banerjee's method.

Tables I and II indicate the following:

a. Whenever t is an integer Banerjee's method produces smaller variances than the new method. (In all situations in Table I the relative efficiency was less than 1.0.)

b. Whenever t is not an integer and ξ along with a positive η could be found, Banerjee's method again provided smaller variances. (In all situations in Table II the relative efficiencies were less than 1.0.)

However, whenever Banerjee's method produces a negative η, e.g. reference numbers 15, 22, 23, 26, and 27 in [4], or whenever a value for ξ could not be determined e.g. the design given as $v = 4$, $b = 4$, $r = 3$, $k = 3$, and $\lambda = 2$, the new method is the only available method for modifying BIBD to provide orthogonal estimates in weighing designs.

It is of interest to note that in the design given as $v = 4$, $b = 4$, $r = 3$, $k = 3$, and $\lambda = 2$, the value of t is 0.5. For these situations, i.e. whenever t is less than 1.0, the relative efficiency, when compared to Banerjee's method, is greater than 1. That is for those designs when t is less than 1.0, the new method provides smaller variances than Banerjee's method.

Before we compared the relative efficiency of the new method to Banerjee's method we had shown that the covariance matrix, for our estimation procedure, obtained by the least squares approach was identical to that obtained by the maximum likelihood approach. Since these covariance matrices for the least squares and maximum likelihood

approach are the same, one may wonder if the estimators of β for the least squares and maximum likelihood approach are identical, i.e. does $(X'X)^{-1}X'Y$ (least squares estimator) = $(X'V^{-1}X)^{-1}X'V^{-1}Y$ (maximum likelihood estimator)? This will be shown in the following theorem.

Theorem: In the estimation procedure under consideration, with $E(ee') = V\sigma^2$, *the estimators,* $\hat{\beta}$, *obtained by the least squares procedure is identical to the estimators obtained by the maximum likelihood procedure.*

Proof: We wish to show that $(X'X)^{-1}X'Y$ (the least squares estimator of $\hat{\beta}$) = $(X'V^{-1}X)^{-1}X'V^{-1}Y$ (the maximum likelihood estimator of $\hat{\beta}$), i.e. we wish to show that,

$$(X'X)^{-1}X'Y = (X'V^{-1}X)^{-1}X'V^{-1}Y,$$

or

$$(X'X)^{-1}X' = (X'V^{-1}X)X'V^{-1} \tag{1}$$

The method of proving equality is to show that the left and right sides of expression (1) reduce to the same expression. The left side of expression (1) reduces to

$$(X'X)^{-1}X' = \left[\begin{array}{c|ccc} \frac{1}{t} & -a & -a & \cdots & -a \\ \hline -a & c & 0 & \cdots & 0 \\ -a & 0 & c & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots & \vdots \\ -a & 0 & 0 & \cdots & c \end{array}\right] \left[\begin{array}{c|ccc} \sqrt{t} & 1 & 1 & \cdots & 1 \\ \hline 0 & & & & \\ 0 & & & & \\ \vdots & & X'_0 & & \\ 0 & & & & \end{array}\right]$$

where $a = \frac{1}{tk}$, $c = \frac{r}{tk\lambda}$, and X_0 is the original BIBD. On multiplication

we get,

$$(X'X)^{-1}X' = \begin{bmatrix} \sqrt{t}/t & \frac{1}{t} - ka & \frac{1}{t} - ka & \cdots & \frac{1}{t} - ka \\ & \overbrace{\text{r of these terms*}} & \overbrace{\text{b-r of these terms*}} & & \\ -a\sqrt{t} & c-a \ \ldots \ c-a & -a \ \ldots \ -a & & \\ \vdots & \vdots \ \vdots \ \vdots & \vdots \ \vdots \ \vdots & & \\ & \overbrace{\text{r of these terms*}} & \overbrace{\text{b-r of these terms*}} & & \\ -a\sqrt{t} & c-a \ \ldots \ c-a & -a \ \ldots \ -a & & \end{bmatrix} \qquad (2)$$

*The arrangement of these terms is determined from the X'_0 matrix i.e. for every 0 and 1 in all rows of X'_0 there will respectively be (c-a) and (-a) in all rows of (2).

Substituting for a and c we get

$$(X'X)^{-1}X' = \begin{bmatrix} \frac{\sqrt{t}}{t} & 0 & 0 & 0 \ . \ \ldots & 0 \\ & \overbrace{\text{r of these terms}} & & \overbrace{\text{b-r of these terms}} & \\ -\frac{\sqrt{t}}{tk} & \frac{r-\lambda}{tk\lambda} \ \cdots \ \frac{r-\lambda}{tk\lambda} & & -\frac{1}{tk} \ \cdots \ -\frac{1}{tk} & \\ \vdots & \vdots \ \vdots \ \vdots & & \vdots \ \vdots \ \vdots & \\ & \overbrace{\text{r of these terms}} & & \overbrace{\text{b-r of these terms}} & \\ -\frac{\sqrt{t}}{tk} & \frac{r-\lambda}{tk\lambda} \quad \frac{r-\lambda}{tk\lambda} & & -\frac{1}{tk} \ \cdots \ -\frac{1}{tk} & \end{bmatrix} \qquad (3)$$

The right side of expression (1) reduces to

$$(X'V^{-1}X)^{-1} \qquad X' \qquad V^{-1}$$

$$(X'V^{-1}X)^{-1}X'V^{-1} = \left[\begin{array}{c|ccc|c|ccc|c|ccc} 1 & -\frac{1}{k} & -\frac{1}{k} & \cdots & -\frac{1}{k} & \sqrt{t} & 1 & 1 \cdots & 1 & \frac{1}{t} & 0 & 0 \cdots & 0 \\ \hline -\frac{1}{k} & e & d & \cdots & d & 0 & & & & 0 & & & \\ -\frac{1}{k} & d & e & \cdots & d & 0 & & X'_0 & & 0 & & I & \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots & & & & \vdots & & & \\ -\frac{1}{k} & d & d & \cdots & e & 0 & & & & 0 & & & \end{array}\right]$$

where $e = \frac{rk+\lambda(t-1)}{tk^2\lambda}$, $d = \frac{t-1}{tk^2}$, and X_0 is the original BIBD. On multiplication, we get

$$(X'V^{-1}X)^{-1}X'V^{-1} = \left[\begin{array}{c|ccc|c|ccc} 1 & -\frac{1}{k} & -\frac{1}{k} & \cdots & -\frac{1}{k} & \frac{\sqrt{t}}{t} & 1 & 1 \cdots & 1 \\ \hline -\frac{1}{k} & e & d & \cdots & d & 0 & & & \\ -\frac{1}{k} & d & e & \cdots & d & 0 & & X'_0 & \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots & & & \\ -\frac{1}{k} & d & d & \cdots & e & 0 & & & \end{array}\right]$$

$$\begin{bmatrix} \frac{\sqrt{t}}{t} & 1-\frac{k}{k} & 1-\frac{k}{k} & \cdots & 1-\frac{k}{k} \\ & \overbrace{\qquad\qquad}^{\text{r of these terms*}} & & & \overbrace{\qquad\qquad}^{\text{b-r of these terms*}} \\ -\frac{\sqrt{t}}{tk} & -\frac{1}{k}+e+(k-1)d & \cdots & -\frac{1}{k}+e+(k-1)d & -\frac{1}{k}+kd \quad \cdots \quad -\frac{1}{k}+kd \\ \vdots & \vdots & & \vdots & \vdots \\ & \overbrace{\qquad\qquad}^{\text{r of these terms*}} & & & \overbrace{\qquad\qquad}^{\text{b-r of these terms*}} \\ -\frac{\sqrt{t}}{tk} & -\frac{1}{k}+e+(k-1)d & \cdots & -\frac{1}{k}+e+(k-1)d & -\frac{1}{k}+kd \quad \cdots \quad -\frac{1}{k}+kd \end{bmatrix}$$

*The arrangement of these terms is again determined from the X_0' matrix. Therefore the arrangement of these terms will be identical to those in the matrix given in expression (2).

Substituting for e and d, we get

$$(X'V^{-1}X)^{-1}X'V^{-1} = \begin{bmatrix} \frac{\sqrt{t}}{t} & 0 & 0 & 0 & \cdots & 0 \\ & \overbrace{\qquad\qquad}^{\text{r of these terms}} & & & \overbrace{\qquad\qquad}^{\text{b-r of these terms}} & \\ -\frac{\sqrt{t}}{tk} & \frac{r-\lambda}{tk\lambda} & \cdots & \frac{r-\lambda}{tk\lambda} & -\frac{1}{tk} \quad \cdots \quad -\frac{1}{tk} & \\ \vdots & \vdots & \vdots & \vdots & \vdots \quad \vdots \quad \vdots & \\ & \overbrace{\qquad\qquad}^{\text{r of these terms}} & & & \overbrace{\qquad\qquad}^{\text{b-r of these terms}} & \\ -\frac{\sqrt{t}}{tk} & \frac{r-\lambda}{tk\lambda} & \cdots & \frac{r-\lambda}{tk\lambda} & -\frac{1}{tk} \quad \cdots \quad -\frac{1}{tk} & \end{bmatrix} \qquad (4)$$

We see that expressions (3) and (4) are identical, therefore for our estimation procedure $(X'X)^{-1}X' = (X'V^{-1}X)^{-1}X'V^{-1}$ and the theorem is proven.

As an example of the above theorem, consider the BIBD given as $v = 4$, $b = 6$, $r = 3$, $k = 2$, and $\lambda = 1$.

The least squares estimator of β is given as:

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{2} & 0 & 0 & 0 \\ -\frac{1}{6} & 0 & \frac{1}{2} & 0 & 0 \\ -\frac{1}{6} & 0 & 0 & \frac{1}{2} & 0 \\ -\frac{1}{6} & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} Y \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sqrt{3}}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\sqrt{3}}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{\sqrt{3}}{6} & \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{6} \\ -\frac{\sqrt{3}}{6} & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} & \frac{1}{3} \\ -\frac{\sqrt{3}}{6} & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} Y \end{bmatrix} \qquad (1)$$

The maximum likelihood estimator of β is given as:

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{2}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ -\frac{1}{2} & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & \frac{1}{6} \\ -\frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ -\frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 111111 \\ 0 & 111000 \\ 0 & 100110 \\ 0 & 010101 \\ 0 & 001011 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 000000 \\ 0 & 100000 \\ 0 & 010000 \\ 0 & 001000 \\ 0 & 000100 \\ 0 & 000010 \\ 0 & 000001 \end{bmatrix} [Y]$$

$$= \begin{bmatrix} \sqrt{3} & 0 & 0 & 0 & 0 & 0 \\ -\frac{\sqrt{3}}{2} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{\sqrt{3}}{2} & \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{6} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} & \frac{1}{3} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} [Y]$$

$$= \begin{bmatrix} \frac{\sqrt{3}}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\sqrt{3}}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{\sqrt{3}}{6} & \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{6} \\ -\frac{\sqrt{3}}{6} & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} & \frac{1}{3} \\ -\frac{\sqrt{3}}{6} & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} [Y] \qquad (2)$$

We see that expressions (1) and (2), i.e. the least squares and maximum likelihood estimators for β, are identical.

REFERENCES

[1] Banerjee, K. S. (1949). A note on weighing designs. Ann. Math. Statist. 20, 300-304.

[2] Banerjee, K. S. (1950). How balanced incomplete block designs may be made to furnish orhtogonal estimates in weighing designs. Biometrika. 37, 50-58.

[3] Baumert, L. D., Golomb, S. W. and Hall, Jr. M. (1962). Discovery of a Hadamard matrix of order 92. Bull. Amer. Math. Soc. 68, 237-238.

[4] Fisher, R. A. and Yates, F. (1963). Statistical tables for biological, agricultural and medical research. Sixth Edition.

[5] Hotelling, H. (1944). Some improvements in weighing and other experimental techniques. Ann. Math. Statist. 15, 297-300.

[6] Kempthorne, O. (1949). The factorial approach to the weighing problem. Ann. Math. Statist. 19, 238-248.

[7] Mood, A. M. (1946). On Hotelling's weighing problem. Ann. Math. Statist. 17, 432-446.

[8] Paley, R.E.A.C. (1933). On orthogonal matrices. Journal of Mathematics and Physics, Massachusetts Institute of Technology. 12, 311-326.

[9] Williamson, J. (1944). Hadamard's determinant theorem and the sum of four squares. Duke Math. Jour. 11, 65-81.

[10] Yates, F. (1935). Complex experiments. J. R. Statist. Soc. Suppl. 2, 181-247.

COMPUTER CONSTRUCTION OF BALANCED INCOMPLETE BLOCK DESIGNS

Malcolm S. Taylor

US Army Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

ABSTRACT. Numerous papers on construction of BIBD's (Balanced Incomplete Block Designs) appearing in the literature consist of derivations of a set of base blocks with symmetrically repeated differences. The inherent properties of the algebraic or geometric structures that are employed lead to sharply constrained values of the design parameters. There would appear, therefore, to be some interest in a more efficient computational scheme to discriminate readily between a set of blocks which have symmetrically repeated differences and those which do not. This is the topic of this paper, and although the values of the parameters are limited in magnitude by computational considerations, no restrictive parameter relationships are involved, and a number of new design configurations are presented.

INTRODUCTION. A BIBD is an arrangement of v distinct elements (varieties) into b sets (blocks) of exactly k distinct members, each element occurs in exactly r different blocks, and every pair of distinct elements occurs together in exactly λ blocks. The parameters v, b, r, k, λ which characterize a BIBD satisfy the fundamental relations

$$bk = vr \tag{1.1}$$

and

$$r(k-1) = \lambda(v-1) \tag{1.2}$$

For example,

$$\begin{bmatrix} 0, 1, 3 \\ 1, 2, 4 \\ 2, 3, 5 \\ 3, 4, 6 \\ 4, 5, 0 \\ 5, 6, 1 \\ 6, 0, 2 \end{bmatrix}$$

is a BIBD with parameters $v = b = 7$, $r = k = 3$, $\lambda = 1$.
Notice that if the block [0, 1, 3] was specified, the entire design could be generated by successively adding (modulo v) the non-zero residue classes to each element of the specified (or base) block.

Bose [1] presented a technique for generating a BIBD directly from a set of blocks, called base blocks, when a sufficient condition known as symmetrically repeated differences is satisfied. Within the scope of this investigation symmetrically repeated differences means that the totality of the inner-block differences of the base block elements modulo v results in the occurence of every non-zero element exactly λ times.

The $k(k\text{-}1)$ modular differences of [0, 1, 3] are

$$0 - 1 \equiv 6 \qquad 1 - 0 \equiv 1 \qquad 3 - 0 \equiv 3$$
$$0 - 3 \equiv 4 \qquad 1 - 3 \equiv 5 \qquad 3 - 1 \equiv 2$$

and so [0, 1, 3] is a base block for a BIBD with $\lambda = 1$.

Since each block gives rise to $k(k\text{-}1)$ modular differences, it is desirable to discriminate readily between a set of blocks which have symmetrically repeated differences and those which do not.

One can characterize a block of a BIBD with parameters v, b, r, k, λ as a vector of dimension v with elements 0 or 1, where the presence of a variety i is indicated by a 1 in the i+1th position, and zero otherwise. For example, the [0, 3, 4, 7] in a design with $v = 9$, $k = 4$ is uniquely represented by the vector $x = (1, 0, 0, 1, 1, 0, 0, 1, 0)$. Notice that the varieties are represented by the residue classes modulo v.

A NECESSARY AND SUFFICIENT CONDITION FOR SYMMETRICALLY REPEATED DIFFERENCES

If we consider the v x v matrix

$$M_1 = \begin{bmatrix} 1 & 1 & 0 & \dots & & & 0 \\ 0 & 1 & 1 & 0 & \dots & & 0 \\ 0 & 0 & 1 & 1 & & \dots & 0 \\ & & & & \dots & 1 & 1 \\ 1 & 0 & 0 & \dots & & & 1 \end{bmatrix} \qquad (2.1)$$

and form the product $M_1 x = b$, to every pair of elements a_i, a_j in the block represented by x whose difference $a_i - a_j \equiv 1 \pmod{v}$ there occurs a 2 in the resultant vector b. Converseley, the number of 2's in the vector b is precisely the number of differences congruent to $1 \pmod{v}$ that would occur if the totality of differences modulo v of the elements of the block were computed.

Similarly the matrix

$$M_2 = \begin{bmatrix} 1 & 0 & 1 & 0 & \dots & & 0 \\ 0 & 1 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & 1 & \dots & \\ & & & & & & \\ 0 & 1 & 0 & \dots & & & 1 \end{bmatrix} \qquad (2.2)$$

upon multiplying the vector x will cause a 2 to occur in the resultant b for every difference of elements of the block $a_i - a_j \equiv 2 \pmod{v}$.

It is now apparent how the construction should proceed. We form successively the products $M_1 x$, $M_2 x$, ..., $M_{[v/2]}\, x$, and if exactly λ twos occur in b at each stage, then the block differences are symmetrically repeated. This constitutes a necessary and sufficient condition for a block to have symmetrically repeated differences.

One need never proceed further than $[v/2]$ steps, since the existence of a pair of elements a_i, a_j, $a_i - a_j \equiv n \pmod{m}$ implies $a_j - a_i \equiv m - n$. The generalization from a single block to a set of blocks consists merely of replacing the vector x in the product Mx, where $X = (x_1, x_2 \dots, x_{b_o})$ denotes a partitioned matrix, each column of which is the representative of a block.

We proceed to summarize this observation as

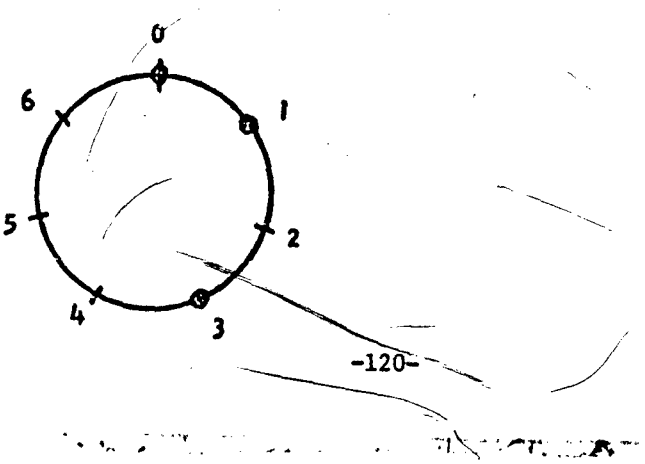
THEOREM 2.1 Consider a block $B_i = [a_{i1}, a_{i2}, \ldots, a_{ik}]$, with elements represented by the congruence classes modulo v, and its representative x_i. A necessary and sufficient condition that B_i satisfy the property of symmetrically repeated differences is that the products $M_j X_i$; $j = 1, 2, \ldots, [v/2]$ each contain the same number, λ, of twos, in the resultant vector.

Generalizing from a single block B_i to a set of blocks $\{B_i\}$ we obtain the immediate

COROLLARY 2.1 A necessary and sufficient condition that a set of blocks $\{B_i\}$ constitute a base for a (v, b, r, k, λ)-design is that the products $M_1 X, M_2 X, \ldots, M_{[v/2]} X$ contain exactly λ twos in each of the resultant matrices.

Observe that the matrices $M_2, M_3, \ldots, M_{[v/2]}$ may be obtained from M_1 in the following fashion: row i of matrix M_m, $2 \leq m \leq [v/2]$, is precisely the vector sum, modulo 2, of rows $i, i+1, \ldots, i+m-1$ (mod v) of matrix M_1. This observation is quite useful when implementing this technique for automatic computation, since matrices $M_2, M_3, \ldots, M_{[v/2]}$ can be generated internally.

If one considers a block as a set of k beads on a necklace of v positions we have, for example, for the block [0, 1, 3] the following representation:



Notice that in computing modular differences what is important is not the labels we have attached to the varieties, but their positions relative to each other. In other words, blocks [0, 1, 3] ≡ [1, 2, 4] ≡ ... ≡ [6, 0, 2]] are all equivalent since each gives rise to the same set of modular differences, and as such each could serve as a base block for the BIBD. These blocks are all members of the same equivalence class, or orbit.

Clearly, when constructing designs from base blocks, we want to consider as candidates only blocks in distinct orbits. Toward this end a convenient way to characterize an orbit is to notice the 1-1 correspondence between an orbit and the circular partition of an integer v; e. g.,

$$[0, 1, 3] \sim 1 + 2 + 4$$

where the summands are simply the "distance" between adjacent beads (varieties). If one generates distinct circular partitions, a task to which the computer is well suited, we are equivalently generating representatives of distinct equivalence classes.

A COMPUTATIONAL ALGORITHM

If we denote the ith row of the matrix X as $a_i = (a_{i1}, a_{i2}, \ldots, a_{ib_o})$, then X can be represented as

$$X = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_v \end{bmatrix}$$

In this compact notation

$$M_i X = \begin{bmatrix} a_1 + a_{i+1} \\ a_2 + a_{i+2} \\ \cdot \\ \cdot \\ \cdot \\ a_v + a_{i+v} \end{bmatrix} \qquad (3.1)$$

where $a_i + a_{i+n}$ represents the usual component-wise addition of the row vectors a_i, a_{i+n}. Notice that if the subscripts $i + n > v$ in (3.1) then $i + n = i + n \pmod{v}$. We compute successively $M_i X$, $i = 1, 2, \ldots, [v/2]$ and terminate the procedure as soon as the required number, λ, of twos fail to appear. Otherwise, completion of the process, indicated here by i taking its maximum value $[v/2]$ without rejection, is sufficient to establish the blocks represented by X as base blocks, generating a BIBD.

Since addition can be performed much more rapidly than multiplication by the computer, in practice we compute $M_i X$ additively as expressed in the right-hand side of (3.1) rather than performing the actual matrix multiplication indicated in Section 2. By a judicious selection of candidates for base blocks, it may be possible to determine a set of blocks generating a (v, b, r, k, λ) - design. Some solutions for BIBD's with large replicates determined in this manner are presented ~~in Table 1~~.

TABLE I

Parameters (v, b, r, k, λ)	Base Blocks
(17, 68, 16, 4, 3)	[0, 1, 2, 4], [0, 1, 7, 10], [0, 2, 6, 11], [0, 3, 7, 12]
(19, 57, 18, 6, 5)	[0, 1, 2, 3, 5, 10], [0, 1, 3, 7, 11, 14], [0, 1, 5, 7, 11, 14]
(11, 44, 20, 5, 8)	[0, 1, 2, 3, 5], [0, 1, 2, 4, 7], [0, 1, 3, 6, 7], [0, 1, 4, 6, 8]
(11, 55, 20, 4, 6)	[0, 1, 2, 3], [0, 1, 3, 6], [0, 1, 4, 7], [0, 1, 5, 7], [0, 2, 4, 7]
(17, 68, 20, 5, 5)	[0, 1, 2, 3, 6], [0, 1, 3, 8, 11], [0, 1, 5, 9, 12], [0, 2, 6, 10, 12].

REFERENCES

Bose, R. (1939), On the Construction of Balanced Incomplete Block Designs, *Annals of Eugenics 9, 353-399.*

Coleman, M. W. and Taylor, M. S. (1971), Circular Integer Partitioning, *Communications of the ACM 14.*

Taylor, M. S. (1972), Computer Construction of Cyclic BIBD's, *J. Statist. Comput. Simul. 1, 65-70.*

ON SPURIOUS CORRELATIONS FOR PARTIALLY RELATED VARIATES

Oskar M. Essenwanger

Physical Sciences Directorate
US Army Missile Research, Development
and Engineering Laboratory
US Army Missile Command
Redstone Arsenal, Alabama

ABSTRACT. The spurious correlation coefficient between related objects such as $a(a \pm b)$ is well known, although sometimes overlooked. Mostly unnoticed goes the spurious correlation, however, when only a subset of the material variates is identical.

The respective formulae for the spurious correlation coefficient are being developed in the case of correlation between wind profile characteristics of the lower tropospheric layers and the atmosphere up to 25 km. Significance testing of the (linear) correlation coefficient against these spurious correlations is described and demonstrated by the wind profile analysis for four typical climatic zones.

Although the method has been developed primarily for the analysis of wind profiles and its physical interpretation, the statistical methodology has general validity.

INTRODUCTION. It is well known that a spurious correlation coefficient is produced when the correlated variates are related such as $x_1 = a$ and $x_2 = a \pm b$, where x_1 and x_2 represent the first and second data sets, respectively. It is overlooked sometimes that a spurious correlation also appears when a subset of the data is related from which x_1 and x_2 are formed. We may call this case a "partial" spurious correlation.

The particular instance arises in the correlation between characteristic coefficients of wind profiles from overlapping layers when x_1 is taken, e.g., for the surface to 3 km profile and x_2 would be computed for the surface to 25 km layer. Although the variates may apparently not be directly related in the sense of an $a(a \pm b)$ multiplication, the parameters are based on material data of which one part is a subset of the total. The appropriate equations will be developed in the following sections.

It should be pointed out that the partial spurious correlation coefficient is not automatically useless for practical applications. Its merit depends on the problem to be solved. When significance of the coefficient is tested against the hypothesis of zero correlation one must be aware that significance may be caused by the interrelation between the two variables

and not by physical cause. When the independent additional correlation is the question, the test basis should be the spurious correlation coefficient rather than the zero value. This may not preclude the utilization of the spurious relationship for inter- or extra-polation, prediction, etc., but the user should be aware that probably no new information is gained in addition to the one already available from the lower layer. This can be checked as discussed later.

The following pages have been reproduced photographically from the author's manuscript.

II. PARTIAL SPURIOUS CORRELATION

1. Correlation Between Means.

This partial spurious correlation is not restricted to the application in the wind profile analysis where it originated. Correlations between parameters which are mean values rather than individual observations can be found in the literature frequently. When one mean is calculated from a subset of the data we have a case of partial relation.

In the parameterization of the wind profile (see Essenwanger 1970, 1971) the characteristic coefficient for the representation of the windspeed v_h as function of the altitude (surface to 25 km) is the mean speed A_o. Assume the characteristic coefficient for the surface to 3 km layer windspeed profile is the mean of that layer, too. Then we have

$$x = \left(\sum_{1}^{h_1} v_h\right)/n_x = v_1 = A_o \qquad (1)$$

and

$$y = \left(\sum_{1}^{h_2} v_h\right)/n_y = \left[\sum_{1}^{h_1} v_h + \sum_{h_1+1}^{h_2} v_h\right]/n_y = B_o \qquad (2)$$

It is evident that $h_1 < h_2$. Further, n_x and n_y are the respective number of points in the computation of the means, $n_x < n_y$.

We have now the choice of notation for the term $\sum_{n_1+1}^{h} v_h$. We may select either one of the subsequent expressions eqn. 3a or b.

First we state

$$\sum_{h_1+1}^{h_2} v_h/n_y = V_2 \qquad (3a)$$

which provides

$$y = v_1 n_x/n_y + V_2 \qquad (2b)$$

or

$$y = w_1 v_1 + V_2 \qquad (2c)$$

Then we define

$$\sum_{h_1+1}^{h_2} v_h/(n_y - n_x) = v_2 \qquad (3b)$$

where v_2 is the regular average of the top layer. Accordingly

$$y = w_1 v_1 + (n_y - n_x)v_2/n_y \qquad (2d)$$

or

$$y = w_1 v_1 + w_2 v_2 \qquad (2c)$$

We have denoted

$$w_1 = n_x/n_y \text{ and } w_2 = (n_y - n_x)/n_y.$$

The two forms (eqn. 3a and 3b) are equivalent, but the subsequent development has been written for the v_2. The equations can be readily brought into the first form by setting $w_2 = 1$.

The definition of the x and y can be expanded to $\bar{x}$ and $\bar{y}$, the mean values, with

$$\bar{y} = w_1 \bar{v}_1 + w_2 \bar{v}_2 \qquad (4a)$$

or

$$\bar{y} = w_1 \bar{v}_1 + \bar{V}_2 \qquad (4b)$$

$$\bar{x} = \bar{A}_o \qquad (4c)$$

As customary, the linear correlation coefficient can be written

$$r_{xy} = Cov_{xy}/\sigma_x\sigma_y = r_{A_oB_o} \tag{5}$$

with

$$Cov_{xy} = \left[\Sigma(x - \bar{x})(y - \bar{y})\right]/N \tag{6a}$$

$$\sigma_x^2 = \left[\Sigma(x - \bar{x})^2\right]/N \tag{6b}$$

$$\sigma_y^2 = \left[\Sigma(y - \bar{y})^2\right]/N. \tag{6c}$$

Now

$$\sigma_x^2 = \Sigma(v_1 - \bar{v}_1)^2/N = \sigma_1^2 = \sigma_{A_o}^2 \tag{6d}$$

$$\sigma_y^2 = \Sigma\left[w_1(v_1 - \bar{v}_1) + w_2(v_2 - \bar{v}_2)\right]^2/N$$
$$= w_1^2\sigma_1^2 + 2\,w_1w_2r_{12}\sigma_1\sigma_2 + w_2^2\sigma_2^2 \tag{6e}$$

where r_{12} is the true correlation between v_1 and v_2 and σ_2^2 the variance of the layer h_{1+1} through h_2, namely $\sigma_{v_2}^2$.

$$NCov_{xy} = \Sigma(v_1 - \bar{v}_1)\left[w_1(v_1 - \bar{v}_1) + w_2(v_2 - \bar{v}_2)\right]$$
$$= w_1\sigma_1^2 + w_2\sigma_1\sigma_2r_{12} \tag{6f}$$

Finally, the linear correlation coefficient becomes

$$r = \left[w_1\sigma_1^2 + w_2\sigma_1\sigma_2r_{12}\right] / \left[\sigma_1(w_1^2\sigma_1^2 + 2\,w_1w_2r_{12}\sigma_1\sigma_2 + w_2^2\sigma_2^2)^{\frac{1}{2}}\right] \tag{5a}$$

In order to derive the spurious correlation we assume that there is no correlation between v_1 and v_2, i.e., $r_{12} = 0$. Then

$$r_{spl} = w_1\sigma_1/(w_1^2\sigma_1^2 + w_2^2\sigma_2^2)^{\frac{1}{2}} \tag{5b}$$

is the spurious correlation. Since the weight w_1 and standard deviation σ_1 are positive this coefficient will be positive, too.

It may be confusing to think in the first moment that for $w_2 = 1$ we have the second form of the spurious coefficient, and the expression looks different. It should be noted, however, that now $\sigma_2^2 = \sigma_{V_2}^2$ and this compensating change makes the new form identical with the first expression.

The $\sigma_{v_2}^2$ or $\sigma^2{}_{V_2}$ is usually not known when the correlation of A_o and B_o is calculated. When σ_v must be obtained from the data there may be no need for the computation of the spurious correlation since probably the coefficient $r_{A_o v_2}$ can be calculated at the same time. When data sets of A_o and B_o are available with σ_{A_o}, σ_{B_o} and $r_{A_o B_o}$, the establishment of v_2 (or V_2) and subsequent determination of $r_{A_o v_2}$ may be costly or time consuming. Then the examination of the spurious correlation would be advantageous. In order to replace $\sigma_{v_2}^2$ in eqn. 5b, we go back to eqn. 2. After some arithmetic we find

$$w_2^2\sigma_{v_2}^2 = (\sigma_{B_o}^2 + w_1^2\sigma_{A_o}^2 - 2w_1\sigma_{A_o}\sigma_{B_o}r_{A_oB_o}) = w_2^2\sigma_2^2 \quad (6g)$$

The second form appears with $w_2 = 1$, from which we can deduct that the same terms replace the second term in the denominator of eqn. 5b. Consequently, the spurious correlation may be written as

$$r_{spl} = w_1\sigma_{A_o}/(2w_1^2\sigma_{A_o}^2 + \sigma_{B_o}^2 - 2w_1\sigma_{A_o}\sigma_{B_o}r_{A_oB_o})^{\frac{1}{2}} \quad (5c)$$

It should be pointed out that $r_{spl} \leq r_{A_oB_o}$.

If we take the equal sign it can be shown that

$$2cr^3 - r^2(2c^2 + 1) + c^2 = 0 \qquad (7)$$

where $c = w_1\sigma_A/\sigma_B$.

This equation can be used to study the behavior between spurious correlation, σ_A and σ_B.

We find the two solutions

$$c_1 = r \qquad (7a)$$

or $$r = w_1\sigma_A/\sigma_B \qquad (7b)$$

and $$c_1 = -r/(1 - 2r^2) \qquad (7c)$$

The latter must be discarded since $0 < r_{spl} \leq 1.0$ and $c_1 > 0$.

Equation (7a) represents the maximum value the spurious correlation could assume when all the correlation between A_o and B_o would be spurious. Since the empirical correlation comprises the spurious and the added part, we must state

$$r_{A_oB_o} = r_{sp} + r_I. \qquad (7d)$$

The independent contribution in the empirical coefficient is not known a priori. It should be added that eqn. 7d is symbolic and a linear addition of the two parts is not applicable, i.e. the r_I is not identical with the $r_{A_ov_2}$ (see eqn. 8 and 9 later).

We may also interpret this spurious correlation coefficient for partially related variables as a weighted correlation. When $w_1 = w_2$ the weights cancel out in eqn. 5b and the notation of the familiar spurious correlation remains.

We may test the significance of the difference between two correlation coefficients by the transformation to Fisher's z (e.g. see Hald, 1952 Brownlee, 1960, etc.). We find

$$z_r = 0.5 \cdot \ln[(1 + r)/(1 - r)] \tag{8}$$

and

$$\Delta z = z_{r_1} - z_{r_2} \tag{9}$$

The z is approximately Gaussian distributed and the standard error of z can be written as

$$\epsilon_z \sim (N - 3)^{-0.5} \tag{10}$$

The null hypothesis that two correlation coefficients came from the same population can be tested by

$$\epsilon_{\Delta z} = [1/(n_1 - 3) + 1/(n_2 - 3)]^{-0.5} \tag{11}$$

which is again normally distributed.

In our case $n_1 = n_2$ as both samples have the same number of observations. Hence

$$\epsilon_{\Delta z} = \sqrt{2}\ (N - 3)^{-0.5} = \sqrt{2}\ \epsilon_z \tag{11a}$$

At the 95% level of significance we accept the null hypothesis when

$$|\Delta z| \leq |1.96\ \epsilon_{\Delta z}| \tag{11b}$$

Table 1 illustrates the testing procedure with the surface to 5 km versus surface to 25 km windspeed profiles and the 3 km versus 25 km system. The headings should be self explanatory.

Table 1. Correlation Between Lower and Upper Systems (A_o versus B_o)

a) Surface to 5 km versus surface to 25 km

	Winter			95%		Spring			95%	
	r	r_{sp}	Δz	$\epsilon_{\Delta z}$	η	r	r_{sp}	Δz	$\epsilon_{\Delta z}$	η
Albrook	.204*	.179	-.026	.092	.26	.206*	.203	-.002	.091	.27
Montg.	.619	.197	-.524	.109	.62	.774	.174	-.855	.096	.79
Chat.	.675	.201	-.615	.098	.67	.759	.213	-.776	.087	.76
Thule	.351*	.174	-.191	.142	.42	.496	.216	-.325	.065	.53
	Summer					Fall				
Albrook	.223*	.204	-.019	.113	.27	.204*	.191	-.014	.096	.23
Montg.	.494	.218	-.320	.084	.49	.745	.176	-.784	.087	.75
Chat.	.785	.218	-.837	.089	.79	.765	.205	-.799	.092	.76
Thule	.698	.290	-.575	.057	.70	.496	.198	-.343	.087	.50

b) Surface to 3 km versus surface to 25 km

	Winter					Spring				
Albrook	.186*	.144	-.043	.092	.21	.199*	.163	-.037	.091	.23
Montg.	.446	.105	-.375	.109	.45	.613	.085	-.628	.096	.61
Chat.	.553	.119	-.502	.098	.56	.587	.118	-.554	.087	.59
Thule	.295	.102	-.201	.142	.35	.396	.115	-.304	.065	.42
	Summer					Fall				
Albrook	.136*	.133	-.002	.075	.18	.143*	.136	-.008	.096	.20
Montg.	.370	.132	-.255	.084	.37	.588	.091	-.583	.087	.60
Chat.	.6195	.119	-.605	.089	.62	.620	.119	-.605	.092	.62
Thule	.504	.172	-.381	.057	.50	.277	.116	-.167	.087	.28

*Not significantly different from spurious correlation coefficient.

It can be concluded from Table 1 that in the tropics close to the equator (Albrook 9°N) the correlation between the two coefficients has the same magnitude as the spurious correlation, but for all other regions the linear relationship goes beyond the one expected merely by a spurious relationship. Since the expectation in zero correlation is the mean value, we could interprete the tropic region result as a justification that the most likely windspeed profile above the lower layer in the tropical region is the mean profile. When the (spurious) correlation coefficient is utilized for extrapolation from the lower layer to the 25 km altitude then only for reason of continuity of the speed profile at the top of the lower layer. There is no apparent physical cause to associate the lower and upper layer. This conclusion agrees with the present facts about the general circulation in the tropical zones.

In contrast to the tropical region midlatitudes and subtropics appear with one closed system from surface to the upper layers. The upper boundary of this system reaches far into the stratosphere but cannot be exactly determined from this wind profile analysis. The top of 25 km was chosen by other considerations and should not be taken literally for the dominance of the wind regime up to that altitude. The question about the upper boundary would

have to be investigated by different techniques. From other correlation analysis (see Stewart and Essenwanger, 1968) 20 km appears most likely. In this article the point of interest is a correlation beyond the spurious relationship.

The columns next to the 95% value of $\epsilon_{\Delta z}$ display the correlation ratio η (see Mills, 1955, etc.) which includes a non-linear relationship. As illustrated by this last column in the parts of Table 1, there is virtually no addition to the linear correlation.

2. Correlation Between Slope and Means.

In the development of the characteristic coefficient of the windspeed profile for the layers surface to 3 km and surface to 5 km the slope is a better representation than the mean. Hence, correlation coefficients between A_1 (slope) of the lower and B_o (mean) of the total altitude range have been computed. Since again the windspeed in the lower layers is the same for both profiles the question arises how much of it is spurious contribution.

We assume

$$x = (\sum_{1}^{h_1} v_h \, \phi_{1h})/n_\phi = A_1 \qquad (12)$$

where ϕ_{1h} is a linear polynomial term and n_ϕ the respective deviser. The y remains the same as defined by eqn's (2). Let us introduce instead of v_1 in eqn (2c) the A_o, then we can write

$$y = w_1 A_o + w_2 v_2 \qquad (13)$$

(The w_1 and w_2 were defined previously.)

With $\bar{x}$ and $\bar{y}$ as customary we derive

$$N\sigma_x^2 = \Sigma(A_1 - \bar{A}_1)^2 = N\sigma_{A_1}^2 \qquad (14a)$$

$$\sigma_y^2 = w_1^2\sigma_{A_o}^2 + 2\, w_1 w_2 r_{A_o v_2} \sigma_{A_o}\sigma_{v_2} + w_2^2\sigma_{v_2}^2 \qquad (14b)$$

and $NCov_{xy} = \Sigma(A_1 - \bar{A}_1)[w_1(A_o - \bar{A}_o) + w_2(v_2 - \bar{v}_2)]$

with $Cov_{xy} = w_1 r_{A_1A_o}\sigma_{A_1}\sigma_{A_o} + w_2 r_{A_1v_2}\sigma_{A_1}\sigma_{v_2}$ (14c)

We denote $\rho_1 = r_{A_ov_2}$, $\rho_2 = r_{A_1v_2}$ and $R_1 = r_{A_1A_o}$. Then the correlation coefficient follows as

$$r = [w_1R_1\sigma_{A_1}\sigma_{A_o} + w_2\rho_2\sigma_{A_1}\sigma_{v_2}]/[\sigma_{A_1}(w_1^2\sigma_{A_o}^2 + 2w_1w_2\rho_1\sigma_{A_o}\sigma_{v_2} + w_2^2\sigma_{v_2}^2)^{\frac{1}{2}}] \quad (15a)$$

We can now again require for the determination of the spurious correlation that wind profiles in the lower layer are independent from the upper layer. This postulation makes ρ_1 and ρ_2 zero. The third correlation coefficient R_1 stems from the lower layer alone. Thus the spurious correlation is

$$r_{sp2} = w_1R_1\,\sigma_{A_o}/(w_1^2\sigma_{A_o}^2 + w_2\sigma_{v_2}^2)^{\frac{1}{2}} \quad (15b)$$

We find by comparison with eqn. (5b) that now the spurious correlation comprises the same terms except the added factor R_1, i.e.

$$r_{sp2} = r_{sp1} \cdot R_1 \quad (15c)$$

It is evident that the spurious correlation is produced again based on the existence of the variances alone as in r_{sp1} which will always be positive. The change is now

the multiplication by the second factor, R_1. It modifies the size and determines the sign of this spurious coefficient. One could call r_{sp2} a spurious correlation coefficient of second kind. It depends on the relationship between the A_o and A_1 coefficient in the lower layer, or in other words the mean and slope and their association. When these two characteristics are independent the R_1 is zero and the spurious relationship, too, becomes zero.

Table 2 lists the actual correlations, the spurious correlation and the correlation between A_o and A_1 of the lower layer (R_1). The checking procedure was the same as for Table 1 and is not repeated here.

We notice that the spurious correlation in all cases is virtually zero. The remaining correlation should, therefore, be due to the relationship between the lower layers and the one above. One would now expect that the correlation coefficient between A_1 and B_o has been adjusted by excluding the spurious part from the $r_{A_oB_o}$ and thus the new $r_{A_1B_o}$ would include this correction. This leaves the optimum correlation between lower and upper layer. This can be confirmed by inspection of Table 2, especially for Albrook, where $r \sim 0$.

Table 2. Correlations Between Slope (A_1) of Lower Layer Windspeed Profile and Mean (B_o) for Surface to 25 km.

a) 5 km versus 25 km system

	Winter			Spring			Summer			Fall		
	r	r_{sp}	R_1	r	r_{sp}	R_1	r	r_{sp}	R_1	r	r_{sp}	R_1
Albrook	-.014*	.006	.110	-.043*	.003	.060	.230	.017	.380	.102*	.010	.230
Montg.	.684	.023	.610	.787	.023	.710	.428	.018	.430	.721	.016	.510
Chat.	.587	.019	.610	.690	.024	.640	.747	.024	.660	.668	.020	.620
Thule	.321	.026	.740	.484	.030	.700	.607	.024	.470	.516	.023	.560

b) 3 km versus 25 km system

	r	r_{sp}	R_1	r	r_{sp}	R_1	r	r_{sp}	R_1	r	r_{sp}	R_1
Albrook	.000*	.003	.120	-.032*	-.003	-.110	.136	.004	.180	.007*	.003	.120
Montg.	.536	.010	.650	.592	.007	.620	.249	.009	.420	.546	.007	.470
Chat.	.490	.007	.610	.533	.008	.560	.597	.009	.560	.589	.008	.630
Thule	.268	.006	.430	.431	.009	.560	.333	.005	.230	.302	.005	.300

*Not significantly different from spurious correlation.

Since the correlation in the other regions is based on physical cause, however, the reduction is not very large. In fact, some of the correlations have even increased slightly. This result is not contradictory as we are dealing with a different parameter, the slope.

Replacement of $\sigma^2_{v_2}$ from eqn. 15b follows by eqn. 6g as discussed.

3. CORRELATIONS BETWEEN SLOPES.

In this phase we treat the problem that both x and y represent a slope of windspeed profiles. The determination of the spurious correlation becomes more difficult than in the previous cases.

$$x = \sum_{1}^{h_1} v_h \phi_{1h}/n_\phi = A_1 \tag{16a}$$

$$y = \sum_{1}^{h_2} v_h \psi_{1h}/n_\psi = B_1 \tag{16b}$$

where ϕ_{1h} and ψ_{1h} are linear polynomials of range h_1 and h_2 with dividers $n_\phi < n_\psi$. We split y again into a lower and upper layer part

$$y = (\sum_{1}^{h_1} v_h \psi_{1h} + \sum_{h+1}^{h_2} v_h \psi_{1h})/n_\psi . \tag{16c}$$

In order to formulate the correlation in terms of lower and full layer let us replace the v_h in the first term of y by an orthogonal polynomial expression

$$v_h = A_o + A_1\phi_{1h} + A_2\phi_{2h} + \ldots \tag{17}$$

This substitution is merely a representation of the lower layer windspeed profile by polynomials and can be expanded to higher order if necessary. Then

$$y = \left[\sum_{1}^{h_1} (A_o + A_1\phi_{1h} + A_2\phi_{2h})\psi_{1h}/n_\psi\right] + V_2 \tag{16d}$$

with the abbreviation $V_2 = \sum_{h_1+1}^{h_2} v_h \psi_{1h}/n_\psi$ (16e)

The y can be further developed into

$$y = A_o \sum_1^{h_1} \psi_{1h}/n_\psi + A_1 \sum_1^{h_1} \phi_{1h}\psi_{1h}/n_\psi + A_2 \sum_1^{h_1} \phi_{2h}\psi_{1h}/n_\psi + V_2$$

or $$y = a_o A_o + a_1 A_1 + a_2 A_2 + \ldots + V_2 \qquad (16f)$$

with $$a_o = \sum_1^{h_1} \psi_{1h}/n_\psi \qquad (18a)$$

$$a_1 = \sum_1^{h_1} \phi_{1h}\psi_{1h}/n_\psi \qquad (18b)$$

$$a_2 = \sum_1^{h_1} \phi_{2h}\psi_{1h}/n_\psi \qquad (18c)$$

where the summation is carried out over the range of the lower layer only.

By analogy we derive $\bar{x}$ and $\bar{y}$ by the usual operation.

Finally we calculate

$$\sigma_x^2 = \sigma_{A_1}^2 \qquad (19a)$$

$$N\sigma_y^2 = \Sigma[a_o(A_o - \bar{A}_o) + a_1(A_1 - \bar{A}_1) + a_2(A_2 - \bar{A}_2) + \ldots(V_2 - \bar{V}_2)]^2$$

$$\sigma_y^2 = a_o^2\sigma_{A_o}^2 + a_1^2\sigma_{A_1}^2 + a_2^2\sigma_{A_2}^2 + \ldots + \sigma_{V_2}^2 + 2a_o a_1 R_1 \sigma_{A_o}\sigma_{A_1} + 2a_o a_2 R_2 \sigma_{A_o}\sigma_{A_2} + 2a_1 a_2 R_3 \sigma_{A_1}\sigma_{A_2} + \ldots$$
$$+ 2a_o\rho_1\sigma_{A_o}\sigma_{V_2} + 2a_1\rho_2\sigma_{A_1}\sigma_{V_2} + 2a_2\rho_3\sigma_{A_2}\sigma_{V_2} \qquad (19b)$$

where the ρ_i are correlation coefficients between lower and upper layer.

The covariance becomes

$$NCov = a_o\Sigma(A_o-\bar{A}_o)(A_1-\bar{A}_1) + a_1\Sigma(A_1-\bar{A}_1)^2 + a_2\Sigma(A_1-A_1)(A_2-\bar{A}_2) + \ldots + \Sigma(V_2-\bar{V}_2)(A_1-\bar{A}_1)$$

$$Cov = a_o R_1 \sigma_{A_o}\sigma_{A_1} + a_1\sigma^2_{A_1} + a_2 R_3 \sigma_{A_1}\sigma_{A_2} + \ldots + \rho_1\sigma_{A_1}\sigma_{V_2} \tag{19c}$$

Again we derive the correlation coefficient

$$r = (a_1\sigma^2_{A_1} + a_o R_1\sigma_{A_o}\sigma_{A_1} + a_2 R_3\sigma_{A_1}\sigma_{A_2} + \ldots + \rho_1\sigma_{A_1}\sigma_{V_2})/\sigma_{A_1}\sigma_y \tag{20}$$

In the spurious correlation coefficient we assume that no relationship between lower and upper layer exists. Hence, $\rho_1 = \rho_2 = \rho_3 = 0$.

$$r_{sp3} = (a_1\sigma^2_A + a_o R_1\sigma_{A_o}\sigma_{A_1} + a_2 R_3\sigma_{A_1}\sigma_{A_2} + \ldots)/\sigma_{A_1}\cdot\sigma'_y \tag{20a}$$

$$\text{with } \sigma'_y = [a^2_o\sigma^2_{A_o} + a^2_1\sigma^2_{A_1} + a^2_2\sigma^2_{A_2} + \ldots + 2a_o a_1 R_1\sigma_{A_o}\sigma_{A_1} + 2a_o a_2 R_2\sigma_{A_o}\sigma_{A_2} + 2a_1 a_2 R_3\sigma_{A_1}\sigma_{A_2} + \ldots + \sigma^2_{V_2}]^{\frac{1}{2}} \tag{20b}$$

We determine $\sigma^2_{V_2}$ from

$$\sigma^2_{V_2} = \sigma^2_{B_1} + (a^2_o\sigma^2_{A_o} + a^2_1\sigma^2_{A_1} + a^2_2\sigma^2_{A_2} + 2a_o a_1\sigma_{A_o}\sigma_{A_1}R_1 + 2a_o a_2\sigma_{A_o}\sigma_{A_2}R_2 + 2a_1 a_2\sigma_{A_1}\sigma_{A_2}R_3) - 2(a_o\sigma_{A_o}\sigma_{B_1}r_{A_oB_1} + a_1\sigma_{A_1}\sigma_{B_1}r_{A_1B_1} + a_2\sigma_{A_2}\sigma_{B_1}r_{A_2B_1}) \tag{20c}$$

$$\sigma'_y = [2a^2_o\sigma^2_{A_o} + 2a^2_1\sigma^2_{A_1} + 2a^2_2\sigma^2_{A_2} + \sigma^2_{B_1} + 2(2a_o a_1\sigma_{A_o}\sigma_{A_1}R_1 + 2a_o a_2\sigma_{A_o}\sigma_{A_2}R_2 + 2a_1 a_2\sigma_{A_1}\sigma_{A_2}R_3) - 2(a_o\sigma_{A_o}\sigma_{B_1}r_{A_oB_1} + a_1\sigma_{A_1}\sigma_{B_1}r_{A_1B_1} + a_2\sigma_{A_2B_1}r_{A_2B_1})]^{\frac{1}{2}} \tag{20d}$$

When no correlation between the coefficients of the lower layer exists, i.e. $R_1 = R_2 = R_3 = \ldots R_i = 0$, then

$$r'_{sp3} = a_1\sigma_{A_1}/(a_o^2\sigma_{A_o}^2 + a_1^2\sigma_{A_1}^2 + a_2^2\sigma_{A_2}^2 + \ldots + \sigma_{V_2}^2) \quad (20e)$$

or with $\sigma_{V_2}^2$ replaced

$$r'_{sp3} = a_1\sigma_{A_1}/[2a_o^2\sigma_{A_o}^2 + 2a_1^2\sigma_{A_1}^2 + 2a_2^2\sigma_{A_2}^2 + \sigma_{B_1}^2 - 2(a_o\sigma_{A_o}\sigma_{B_1}r_{A_oB_1} + a_1\sigma_{A_1}\sigma_{B_1}r_{A_1B_1} + a_2\sigma_{A_2}\sigma_{B_1}r_{A_2B_1})]^{\frac{1}{2}} \quad (20f)$$

This is the logically expanded form of r_{sp1} and r_{sp2}. Usually some correlations R_1, R_2 or R_3 are not zero, and the computation must be based on (20a).

The actual r_{sp3} can again assume positive and negative values depending on the a_1.

The analysis results of the surface to 3 km and surface to 5 km wind profile system (slopes) is illustrated in Table 3. This table depicts the correlation between the slopes and the spurious correlation in the upper part. The correlation between A_o, A_2 of the lower layer and the slope B_1 of the entire layer up to 5 km is added in the middle part, while the lower part contains the intercorrelations of the surface to 3 km layer.

It proves again that the spurious correlation is close to zero in the subtropics, midlatitudes and polar regions, while it is significantly different from zero in the tropical region. In contrast

to the previous results, however, the actual (linear) correlation coefficient displays significant difference from the spurious correlation at the 95% level. Although the possibility exists that the additional information between spurious and actual correlation in the tropics could be attributed to physical cause, the suspicion of identity between spurious and actual correlation remains. Two factors may contribute to produce significance.

Since $N > 1000$ in all cases, already smaller differences between spurious and actual correlations render significant dissimilarity. Since eqn. (17) is an approximation, the inclusion of a third order term may bring both coefficients closer together. Whether the actual spurious correlation is underestimated, however, cannot be readily predicted. There is no doubt that the correlations in the other climatic regions are real.

Again, the correlation ratio did not prove of any practical value beyond the linear relationship and has been omitted from publication.

Table 3. Correlation Between Slope (A_1) of 3 km Windspeed Profile and Slope (B_1) of 5 km Windspeed Profile.

	Winter		Spring		Summer		Fall	
	r	r_{sp}	r	r_{sp}	r	r_{sp}	r	r_{sp}
Albrook	.46	.27	.61	.42	.48	.23	.54	.28
Montg.	.78	-.06	.77	-.06	.62	.05	.78	.01
Chat.	.72	-.13	.73	-.05	.78	-.04	.75	-.14
Thule	.67	.09	.78	-.01	.66	.11	.68	.09

Correlations Between A_o, A_2 and B_1

	$r_{A_oB_1}$	$r_{A_2B_1}$	$r_{A_oB_1}$	$r_{A_2B_1}$	$r_{A_oB_1}$	$r_{A_2B_1}$	$r_{A_oB_1}$	$r_{A_2B_1}$
Albrook	-.43	.39	-.50	.46	-.12	.31	-.26	.38
Montg.	.42	.39	.53	.33	.12	.28	.41	.28
Chat.	.33	.17	.28	.31	.37	.28	.33	.24
Thule	.15	.33	.42	.28	.07	.30	.17	.34

Intercorrelations, surface to 3 km Windspeed Profiles.

	$r_{A_oA_1}$	$r_{A_oA_2}$	$r_{A_1A_2}$	$r_{A_oA_1}$	$r_{A_oA_2}$	$r_{A_1A_2}$	$r_{A_oA_1}$	$r_{A_oA_2}$	$r_{A_1A_2}$	$r_{A_oA_1}$	$r_{A_oA_2}$	$r_{A_1A_2}$
Albrook	.12	-.55	.05	-.11	-.40	.12	.18	-.21	-.09	.11	-.36	.00
Montg.	.65	-.37	.15	.62	-.38	.11	.42	-.47	.00	.47	-.49	.15
Chat.	.61	-.53	-.17	.56	-.45	.04	.56	-.43	.07	.63	-.54	-.05
Thule	.33	-.43	-.02	.56	-.27	.00	.23	-.37	-.02	.30	-.31	.03

CONCLUSION. It has been demonstrated that the linear correlation coefficient may be spurious when the data are only partially related. The investigated case in this article deals with the particular problem in which one parameter is computed from a subset of the total data.

Three cases were examined. The first case was based on the condition where two mean values are computed, and one mean is calculated from a subset of the data. As the example for windspeed profiles in four different climatic regions shows (Table 1), a spurious correlation different from zero emerges in all four zones, but only in the tropical zone does it appear that the actual correlation is identical with the spurious one as tested at the 95% level of significance.

The study is expanded to examine the spurious correlation between slope in the lower layer and mean of the entire altitude range. The spurious correlation must be modified by the inclusion of a correlation term. The example for windspeed profiles from four climatic zones displays that this time the spurious correlation is approximately zero. Adjustments to reflect this reduction of the spurious correlation appear in the correlation coefficients, especially in the tropical zone.

The last case deals with the problem of two slopes. The spurious correlation assumes a more intricate form containing various correlation terms. The empirical example for windspeed profiles exhibits spurious correlation different from zero only in the tropical zone. Present tests indicate, however, that in this zone spurious and actual correlation coefficients are different, too, as tested at the 95% level of significance.

The spurious correlation has been largely developed in this study to check correlation coefficients between characteristic coefficients of windspeed profiles which were established in the analysis of wind data for missile design and operational purposes. The consequences go far beyond this limit, however, and the method has application in various fields of statistical analysis. One of the main goals was further to illustrate in this article that spurious correlations can arise in cases where only parts of the data are related. It is, therefore, advisable to examine the parameters to be correlated for any source of possible spurious relationship before conclusions are drawn about physical causes of any existing correlation.

ACKNOWLEDGEMENT. The author wishes to express his gratitude to Mrs. H. Boyd who supported the work by establishing the computer programs for correlation coefficients and correlation ratio. Without her aid the manuscript could not have been established in time. Dr. D. A. Stewart may be thanked for her critical review of the manuscript. Mrs. C. Brooks must be commended for her assiduous effort in typing the manuscript in record time.

REFERENCES CITED:

Brownlee, K. A., 1960. Statistical Theory and Methodology in Science and Engineering. Wiley & Sons, New York, pp 570.

Essenwanger, O. M., 1970. On the Wind Profile Representation for Missile Design and Environment Analysis. Proceedings of the 4th National Conference on Aerospace Meteorology (Publ. by Am. Meteor. Soc., Boston, May 1970) p. 246-253.

Essenwanger, O. M., 1971. Characteristic Coefficients, Probability and Classification of Wind Profiles (Surface to 25 km). Proceedings of the 16th Conference on the Design of Experiments ARO-D Report 71-3, p 163-184.

Hald, A., 1952. Statistical Theory with Engineering Application. Wiley & Sons, New York, pp 783.

Mills, F. C., 1924. Statistical Methods. Holt and Co., New York, Third Edition 1955, pp 842.

Stewart, D. A. and Essenwanger, O. M., 1968. Zur Interpretation Aerologischer Korrelationen und deren Zusammenhang mit Speziellen Schichten. Meteor. Rundsch., Vol 21, No 5, p 134-138.

THE LEAST SQUARES ANALYSIS OF DATA GENERATED BY A "PIECE-WISE" GENERAL LINEAR MODEL

Robert L. Launer

Army Procurement Research Office
Fort Lee, Virginia

This study was motivated by the frequent appearance of economic data which can be described as piece-wise linear with certain "end-point" or "cross-sectional" constraints. (Figure 1 depicts several examples of this.) The study is intended mainly for the field analyst who is confronted with this type of data and insufficient time to work out more than the barest details.

All of the models in figure 1 can be expressed as straight lines within each of several intervals with a linear constraint on the parameters. For example, the broken line can be represented as follows:

$$y = \begin{cases} a_1 + b_1 x & , \quad x \leq x^* \\ a_2 + b_2 x & , \quad x > x^* \end{cases} ,$$

subject to the constraint:

$$a_1 + b_1 x^* + d = a_2 + b_2 x^* \quad .$$

The "bent line" model is just the "broken line" with $d = o$, and the third example in figure 1 has the same form of representation except that the auxilliary condition is just $b_1 = b_2$.

EXAMPLES OF PIECE-WISE LINEAR MODELS

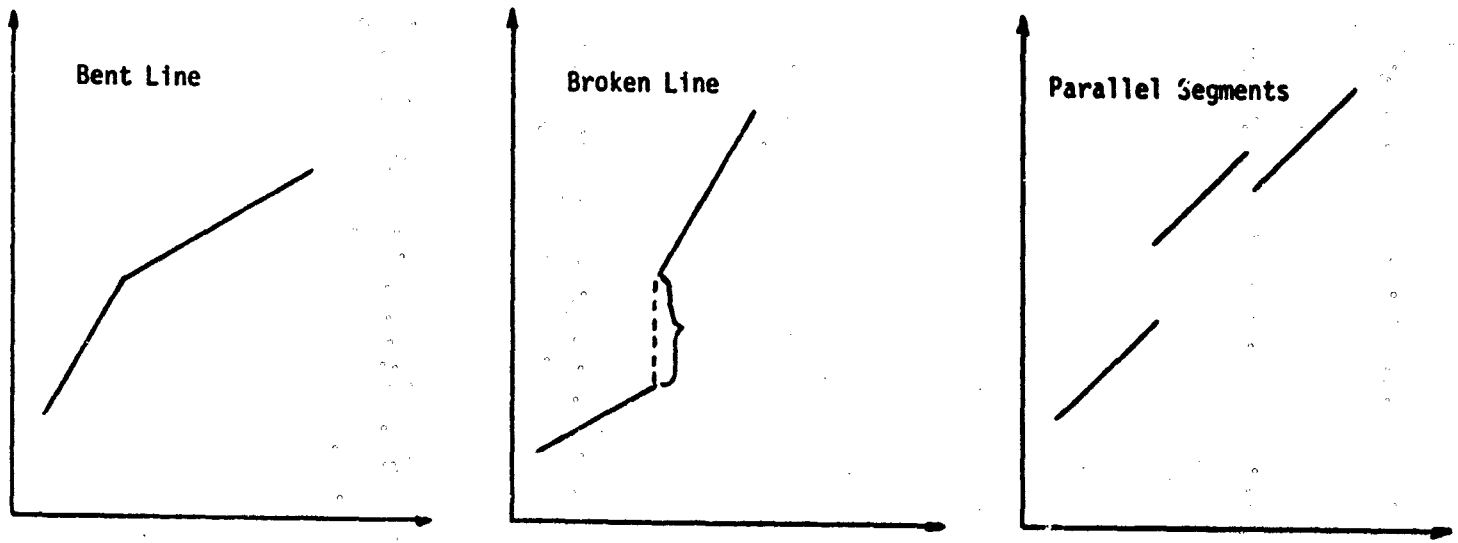


FIGURE 1

Examples of models which follow this broken line pattern, but which are not linear within each interval, also exist. For example, a corporation's stable economic growth might be suddenly interrupted by an external factor such as a war or merger, after which its growth is again stable but progressing at a different rate. If the growth equation contains only linear parameters and the exact time of the change in growth is known, then this model can be analyzed with the methods outlined in this paper.

The theory of least squares and regression analysis subject to parametric constraints has been treated extensively. In this paper, the theory subject to linear parametric constraints is presented and cast in a form which allows easy adaptation to "piece-wise" general linear models. A test for linearity of data is proposed and finally, the theory is illustrated with data obtained from US Army cost-incentive contracts. A general familiarity with the theory of linear hypotheses will be assumed.

1. Regression Analysis with Linear Constraints.

Let it be required to estimate the elements of $\underline{\beta}' = (\beta_1, \beta_2, \ldots, \beta_p)$, from the n observations $\underline{y}' = (y_1, y_2, \ldots, y_n)$ which are generated by the general linear model $\underline{y} = X\underline{\beta} + \underline{\epsilon}$. If $\underline{\epsilon}' = (\epsilon_1, \ldots, \epsilon_n)$ is the error vector and the matrix X is known, of dimension nxp and of rank $p \leq n$, then the least squares estimate of $\underline{\beta}$ is

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}. \qquad (1.1)$$

Suppose now that the elements of $\underline{\beta}$ are constrained by the $k<p$ known linearly independent relationships:

$$\begin{array}{l} t_{11}\ \beta_1+t_{12}\beta_2+\ldots+t_{1p}\ \beta_p=d_1 \\ \vdots \qquad\qquad \vdots \\ t_{k1}\ \beta_1+t_{k2}\beta_2+\ldots+t_{kp}\beta_p=d_k \end{array} \tag{1.2}$$

If $T=(t_{ij})$ and $\underline{d}'=(d_1,d_2,\ldots,d_k)$, equations (1.2) may be written

$$T\,\underline{\beta} = \underline{d}. \tag{1.3}$$

Note that T is of full rank $k< p$.

In order to minimize $(\underline{y}-X\underline{\beta})'(\underline{y}-x\underline{\beta})$ with respect to $\underline{\beta}$ subject to the constraint (1.2), the method of LaGrangian multipliers immediately suggests itself. Let $\underline{\lambda}'=[\lambda_1,\ldots,\lambda_k]$. It is necessary then to find the extreme value of $(\underline{y}-x\underline{\beta})'(\underline{y}-x\underline{\beta})+\underline{\lambda}[T\underline{\beta}-\underline{d}]$ (1.4) with respect to $\underline{\beta}$ and $\underline{\lambda}$. Differentiation of (1.4) with respect to $\underline{\beta}$ and $\underline{\lambda}$ yields the "constrained" normal equations:

$$\begin{array}{l} x'x\tilde{\beta} + \frac{1}{2}\,T'\,\underline{\lambda} = x'\underline{y} \\ \qquad T\,\tilde{\underline{\beta}} = \underline{d} \end{array} \tag{1.5}$$

Solving (1.5) for $\tilde{\underline{\beta}}$ yields:

$$\tilde{\underline{\beta}} = (x'x)^{-1}x'\underline{y} - \frac{1}{2}\,(x'x)^{-1}T'\,\underline{\lambda} \tag{1.6}$$

where,

$$\frac{1}{2}\underline{\lambda} = [T(x'x)^{-1}T']^{-1}[T(x'x)^{-1}x'\underline{y} - \underline{d}] \tag{1.7}$$

Thus from (1.1):

$$\tilde{\underline{\beta}}=\hat{\underline{\beta}} - (x'x)^{-1}T'[T(x'x)^{-1}T'] [T \hat{\underline{\beta}} -\underline{d}] \qquad (1.8)$$

Notice that $\tilde{\underline{\beta}}$ is a linear combination of the components of $\hat{\underline{\beta}}$ and therefore, is unbiased for $\underline{\beta}$ (as it must be.) Furthermore, $\tilde{\underline{\beta}}$ is normally distributed with

$$\text{cov } (\tilde{\underline{\beta}})= \sigma^2\{(x'x)^{-1}-(x'x)^{-1}T'[T(x'x)^{-1}T']^{-1} T(x'x)^{-1}\} , \qquad (1.9)$$

and the rank of cov $(\tilde{\underline{\beta}})$ is $p - k$.

The remaining distributional properties of $\tilde{\underline{\beta}}$ and the tests of linear hypotheses are not difficult to establish. First, recall that

$$\frac{1}{\sigma^2} (\underline{y}-x\underline{\beta})'(\underline{y}-x\underline{\beta}) \sim \chi^2_n . \qquad (1.10)$$

Furthermore, if M is an nxn matrix, then

$$(\underline{y}-x\underline{\beta})'(\underline{y}-x\underline{\beta})=(\underline{y}-x\underline{\beta})'(I-M)(\underline{y}-x\underline{\beta})+(\underline{y}-x\underline{\beta})'M(\underline{y}-x\underline{\beta}). \qquad (1.11)$$

Finally, from 1.3 it may be observed that

$(\tilde{\underline{\beta}}-\underline{\beta})x'(\underline{y}-x\tilde{\underline{\beta}})=(T\underline{\beta}-\underline{d})' [T(x'x)^{-1}T']^{-1} [T\hat{\underline{\beta}}-\underline{d}] = 0$ and it follows that

$$(\underline{y}-x\underline{\beta})'(\underline{y}-x\underline{\beta})=(\underline{y}-x\tilde{\underline{\beta}})'(\underline{y}-x\tilde{\underline{\beta}}) + (\tilde{\underline{\beta}}-\underline{\beta})'x'x(\tilde{\underline{\beta}}-\underline{\beta}). \qquad (1.12)$$

If

$$M = x(x'x)^{-1}x'-x (x'x)^{-1}T'[T(x'x)^{-1}T']^{-1}T(x'x)^{-1}x' \qquad (1.13)$$

it also follows that

$$(\tilde{\underline{\beta}}-\underline{\beta})'x'x\,(\tilde{\underline{\beta}}-\underline{\beta}) = (\underline{y}-x\underline{\beta})'M(\underline{y}-x\underline{\beta}). \qquad (1.14)$$

Notice that M is idempotent of rank p - k. Therefore (1.14) is distributed as $\sigma^2\chi^2_{p-k}$ so long as $\underline{y} \sim N(X\underline{\beta}, \sigma^2 I)$. Finally, from (1.12), (1.14) and Cochrans Theorem, the distribution of the residual or error estimate $\tilde{\sigma}^2$ is:

$$\frac{1}{\sigma^2}(\underline{y}-x\tilde{\underline{\beta}})'(\underline{y}-x\tilde{\underline{\beta}}) \sim \chi^2_{n-(p-k)}. \qquad (1.15)$$

The error sum of squares (1.15) is related to the "unconstrained" error sum of squares which can be shown by direct substitution of (1.8) into (1.15):

$$(\underline{y}-x\tilde{\underline{\beta}})'(\underline{y}-x\tilde{\underline{\beta}}) = (\underline{y}-x\hat{\underline{\beta}})'(\underline{y}-x\hat{\underline{\beta}})$$
$$+[T\hat{\underline{\beta}}-\underline{d}]'\,[T(x'x)^{-1}T']^{-1}[T\hat{\underline{\beta}}-\underline{d}].$$

This formula leads to computational efficiencies later.

Tests of hypotheses regarding the elements of $\underline{\beta}$ must be conducted with slightly more care than is usual. Evidently, any test of hypothesis on $\underline{\beta}$ subject to the given constraints $T\underline{\beta}=\underline{d}$ may be regarded as a test of hypothesis on a second set of constraints, $T_o\underline{\beta} = \underline{d}_o$, given the first set. Since $\underline{\beta}$ contains p elements and the rank of T is $k \leq p$, then the test of hypothesis may be expressed as no more than p-k linearly independent equations. In other words, the rank of T_o may be no more than p-k.

Suppose that the matrix T_o referred to in the preceding paragraph is m x p of full rank m, and that the rows of T_o are linearly independent of the rows of T. Then define the augmented matrix T* of dimension (k+m)xp and rank (k+m) and the augmented vector $\underline{d}^*$ of dimension k+m as follows:

$$T^* = \begin{bmatrix} T \\ \cdots \\ T_o \end{bmatrix} \qquad \underline{d}^* = \begin{bmatrix} \underline{d} \\ \cdots \\ \underline{d}_o \end{bmatrix}$$

The null hypothesis, Ho, is $T_o\,\underline{\beta}=\underline{d}_o$ and the alternative is $T_o\underline{\beta}\neq\underline{d}_o$. Then the sum of squares due to $\underline{\beta}$ subject to T*, SS* ($\underline{\beta}$), is given by (1.13) with T replaced by T*.

$$SS^*(\underline{\beta}) = (\underline{y}-x\underline{\beta})' M^*(\underline{y}-x\underline{\beta}) \qquad (1.16)$$

$$M^*=x(x'x)^{-1}x' - x(x'x)^{-1}T^{*\prime}\,[T^*(x'x)^{-1}T^{*\prime}]^{-1}T^*(x'x)^{-1}x' \qquad (1.17)$$

If Ho is true then,

$$SS^*(\underline{\beta}) \sim \sigma^2 \cdot \chi^2_{p-k-m} \qquad (1.18)$$

Note that M* is idempotent of rank p-k-m. The "corrected" sum of squares (Graybill) to test Ho is the difference (1.14) and (1.16)

$$(\underline{y}-x\underline{\beta})' (M-M^*) (\underline{y}-x\underline{\beta}) \qquad (1.19)$$

For computational purposes, this may be written as

$$[T^*\hat{\underline{\beta}}-\underline{d}^*]'[T^*(x'x)^{-1}T^{*\prime}]^{-1}[T^*\hat{\underline{\beta}}-\underline{d}^*] - [T\hat{\underline{\beta}}-\underline{d}]'[T(x'x)^{-1}T']^{-1}[T\hat{\underline{\beta}}-\underline{d}] \qquad (1.20)$$

In order to investigate the properties of (1.20), introduce the notation:

$$A=T(x'x)^{-1}x'$$

$$A_0=T_0(x'x)^{-1}x'$$

$$A^*=T^*(x'x)^{-1}x'=\begin{bmatrix}A\\ \vdots\\ A_0\end{bmatrix} \qquad (1.21)$$

Then the sums of squares (1.15) (through (1.13), (1.14), and (1.11)) and (1.19) can be written respectively as:

$$(\underline{y}-x\underline{\beta})'[I-x(x'x)^{-1}x'-A'[AA']^{-1}\ (\underline{y}-x\underline{\beta}) \qquad (1.22)$$

and

$$(\underline{y}-x\underline{\beta})'[A^{*\prime}[A^*A^{*\prime}]^{-1}A^*-A'[AA']^{-1}A]\ (\underline{y}-x\underline{\beta}) \qquad (1.23)$$

Now, notice that both $A'[AA']^{-1}A$ and $A^*[A^*A^{*\prime}]^{-1}A^*$ are idempotent. Furthermore, $x(x'x)^{-1}x'A'=A'$ and $x(x'x)^{-1}x'A^{*\prime}=A^{*\prime}$. Since the inverse of $[A^*A^{*\prime}]$ is

$$\begin{bmatrix}[AA'-AA_0'(A_0A_0')^{-1}A_0A']^{-1} & -(AA')^{-1}AA_0'[A_0A_0'-A_0A'(AA')^{-1}AA_0']^{-1}\\ -[A_0A_0']^{-1}A_0A'[AA'-AA_0'(A_0A_0')^{-1}A_0A']^{-1} & [A_0A_0'-A_0A'(AA')^{-1}AA_0']^{-1}\end{bmatrix} \qquad (1.24)$$

then $A^{*\prime}[A^*A^{*\prime}]^{-1}A^*$ is

$$[A'-A_0'(A_0A_0')^{-1}A_0A'][AA'-AA_0'(A_0A_0')^{-1}A_0A']^{-1}A+$$
$$+A_0'[A_0A_0'-A_0A'(AA')^{-1}AA_0']^{-1}[A_0-A_0A'(AA')^{-1}A] \qquad (1.25)$$

The matrix (1.24) is symmetric, from which fact one can show that each of the two terms of (1.25) is symmetric. Therefore,

$$A^{*\prime}[A^*A^{*\prime}]^{-1}A^*A'[AA']^{-1}A=A'[AA']^{-1}A \qquad (1.26)$$

These facts together with the transpose relationship derived from (1.26) allow one to show that the matrices of the two quadratic forms (1.22) and (1.23) are idempotent and that the two quadratic forms are independent of ranks m and n-p+k, respectively.

Therefore, to test the hypothesis

Ho: $T_o \underline{\beta} = \underline{d}_o$ (given $T \underline{\beta} = \underline{d}$)

use the upper tail F statistic $F=F_{m,\, n-p+k,\, \lambda}$

$$F=\frac{[T^*\hat{\underline{\beta}}-\underline{d}^*]'[A^*A^{*\prime}]^{-1}[T^*\hat{\underline{\beta}}-\underline{d}^*]-[T\hat{\underline{\beta}}-\underline{d}]'[AA']^{-1}[T\hat{\underline{\beta}}-\underline{d}]}{[\underline{y}-x\hat{\underline{\beta}}]'[\underline{y}-x\hat{\underline{\beta}}]+[T\hat{\underline{\beta}}-\underline{d}]'[AA']^{-1}[T\hat{\underline{\beta}}-\underline{d}]} \cdot \frac{(n-p+k)}{m} \quad (1.27)$$

where,

$$\lambda = \frac{1}{2}\ \underline{\Delta}'[A_oA_o'-A_oA'[AA']^{-1}AA_o']\ \underline{\Delta} \quad \text{for } T_o \underline{\beta} = \underline{d}_o + \underline{\Delta}.$$

2. Sectional Models;

Suppose that $\underline{y} = f(\underline{x};\underline{\beta})$ is a "piece-wise" general linear model described earlier. Since the functional form itself depends on the independent variable, one may write:

$$y = \begin{cases} f_1(x;\underline{\beta}_1) & x \le x_1^* \\ f_2(x;\underline{\beta}_2) & x_1^* \le x \le x_2^* \\ \quad\vdots & \\ f_q(x;\underline{\beta}_q) & x_{q-1}^* \le x \end{cases} \quad (2.1)$$

subject to the constraints:

$$\begin{aligned} T_1 \underline{\beta} &= \underline{d}_1 \\ T_2 \underline{\beta} &= \underline{d}_2 \\ &\vdots \\ T_r \underline{\beta} &= \underline{d}_r \end{aligned} \qquad (2.2)$$

Equation (2.1) may be written

$$y \begin{cases} x_1\beta_1 & x_1 \leq x_1^* \\ x_2\beta_2 & x_1^* < x_2 \leq x_2^* \\ \vdots & \\ x_q\beta_q & x_{q-1}^* \leq x_q \end{cases}$$

subject to the constraints:

$$T \underline{\beta} = \underline{d}.$$

It is clear that x, x'x, and $(x'x)^{-1}$ may be written as partitioned matrices with zero in every non-diagonal block.

$$X = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ 0 & 0 & & \\ & & & \\ 0 & 0 & & X_q \end{bmatrix} \qquad (2.3)$$

and similarly for x'x and $(x'x)^{-1}$.

A glance at equations (1.8), (1.16) and (1.20) indicate that the calculation involved in obtaining $\underline{\tilde{\beta}}$ and the F statistic are simplified since $\underline{\tilde{\beta}}$ is a linear combination of the $\underline{\hat{\beta}}_j$, and $\underline{\hat{\beta}}_j = (x_j'x_j)^{-1}x_j'\ \underline{y}$.

3. The Broken Line Example.

Suppose it is necessary to estimate the parameters in the broken line model

$$y_i = \begin{cases} a_1+b_1(x_i-\bar{x}_1) & , x_i \leq x^*,\ i=1,2,\ldots,m_1 \\ a_2+b_2(x_i-\bar{x}_2) & , x^* < x_i,\ i=m_1+1,\ldots,n \end{cases} \qquad (3.1)$$

where $a_1+b_1(x^*-\bar{x}_1)+d=a_2+b_2(x^*-\bar{x}_2)$, (3.2)

$\bar{x}_1 = \frac{1}{m_1}\sum_{x\leq x^*} x_i$ and $\bar{x}_2 = \frac{1}{n-m_1}\sum_{x>x^*} x_i$.

The subscripts 1 and 2 refer to the elements of the model and their estimates which lie to the left and right hand side, respectively, of the discontinuity point x^*. x^* and d are known, and $m_2=n-m_1$.

Let

$$X_1 = \begin{bmatrix} 1 & x_1-\bar{x}_1 \\ 1 & x_2-\bar{x}_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_{m_1}-\bar{x}_1 \end{bmatrix}$$

and similarly for X_2. Then

$$X=\begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix} \qquad (3.3)$$

$T=[1, x^*-\bar{x}_1, -1, -(x^*-\bar{x}_2)]$ and (3.4)

$\underline{\beta}' = [a_1, b_1, a_2, b_2]$, so that the broken line model can be written in matric notation,

$\underline{y}=X\underline{\beta}+\underline{\epsilon}$ subject to $T\underline{\beta}=\underline{d}$.

The matrix $(x'x)$ is

$$(x'x)=\begin{bmatrix} m_1 & 0 & 0 & 0 \\ 0 & s_{xx1} & 0 & 0 \\ 0 & 0 & m_2 & 0 \\ 0 & 0 & 0 & s_{xx2} \end{bmatrix} \quad \text{and}$$

$$T[x'x]^{-1}T' = \frac{1}{m_1} + \frac{(x^*-\bar{x}_1)^2}{s_{xx1}} + \frac{1}{m_2} + \frac{(x^*-\bar{x}_2)^2}{s_{xx2}}$$

The results of section 1 give:

$$\hat{\lambda} = \frac{\bar{y}_1+(x^*-\bar{x}_1)\dfrac{s_{xy1}}{s_{xx1}} - \bar{y}_2-(x^*-\bar{x}_1)\dfrac{s_{xy2}}{s_{xx2}} + d}{\dfrac{1}{m_1} + \dfrac{(x^*-\bar{x}_1)^2}{s_{xx1}} + \dfrac{1}{m_2} + \dfrac{(x^*-\bar{x}_2)^2}{s_{xx2}}},$$

$$\tilde{a}_1 = \bar{y}_1 - \frac{\lambda}{2m_1}$$

$$\tilde{b}_1 = \frac{S_{xy1}}{S_{xx1}} - \frac{\lambda}{2}\,\frac{(x^* - \bar{x}_1)}{S_{xx1}} \qquad (3.6)$$

$$\tilde{a}_2 = \bar{y}_2 + \frac{\lambda}{2m_2}$$

$$\tilde{b}_2 = \frac{S_{xy2}}{S_{xx2}} + \frac{\lambda}{2}\,\frac{(x^* - \bar{x}_2)}{S_{xx2}}.$$

It is interesting to note that, if

$$\hat{y}_1^* = \hat{a}_1 + \hat{b}_1(x^* - \bar{x}_1), \quad \hat{y}_2^* = \hat{a}_2 + \hat{b}_2(x^* - \bar{x}_2)$$

then

$$\frac{\lambda}{2} = \frac{y_1^* + d - y_2^*}{[\mathrm{Var}(y_1^*) + \mathrm{Var}(y_2^*)]/\sigma^2}.$$

It is important to note, that if d is not known, then (from (1.4)) it follows that $\lambda = 0$ and the least squares estimate of β is merely $\underline{\hat{\beta}} = \underline{\tilde{\beta}}$. This means that the constraint is unknown and must be estimated with the unconstrained estimate $\underline{\hat{\beta}}$. If $\underline{d}$ contains at least one known element and at least one unknown element then this is not always true.

A test of hypothesis which seems potentially very useful is a test for "linearity" of data; i.e., a test to determine whether or not the data is generated by a single straight line model over the entire range of the independent variable. The test is easily conceptualized.

First, the bent line model, eq. (3.1) and eq. (3.2) with d=o, is assumed. The "bend" point (x^*, y_1^*) is then one point in common to the straight line segments which compose the model. Then the hypothesis is either, Hol, that the two slopes in question are equal or, Ho2, that a point different from the bend point is common to the line segments or their extensions.

Formally, the two hypotheses are:

$$\begin{aligned} &\text{Hol: } b_1 = b_2 \ , \\ &\text{Ho2: } a_1 + b_1(x^o - \bar{x}_1) = a_2 + b_2(x^o - \bar{x}_2) \end{aligned} \qquad (3.7)$$

The statistics for both tests are presented here. (Ho2 involves $x^o=o$). Unfortunately, both tests involve unwieldly formulae and can be recommended only by the possible savings in degrees of freedom. The F-statistics are of the form (1.27). To facilitate writing, let $B=(T\underline{\beta}-\underline{d})' [AA']^{-1}(T\underline{\beta}-\underline{d})$ and $C_i=[T^*\underline{\beta}-\underline{d}^*]' [A^*A^{*'}]^{-1}[T^*\underline{\beta}-\underline{d}^*]$ where C_1 and C_2 refer to Hol and Ho2, respectively, and let SS_ε refer to the denominator of the F-test(1.27). Then,

$$B = \frac{\bar{y}_1^2 + \left(\frac{S_{xy1}}{S_{xx1}}\right)^2 (x^*-\bar{x}_1)^2 + \bar{y}_2^2 + \left(\frac{S_{xy2}}{S_{xx2}}\right)^2 (x^*-\bar{x}_2)^2}{\left(\frac{1}{m_1}+\frac{1}{m_2}\right) + \frac{(x^*-\bar{x}_1)^2}{S_{xx1}} + \frac{(x^*-\bar{x}_2)^2}{S_{xx2}}} \ ,$$

On setting

$$H_i = \frac{Sxyi}{Sxxi}\ ,\quad K_i = \frac{1}{Sxxi}\quad ,\quad i = 1,2,$$

C_1 and C_2 can be expressed as follows

$$C_1 = \frac{[\ \bar{y}_1 - \bar{y}_2 - H_1(\bar{x}_1 - \bar{x}_2)\ ]^2 K_2 + [\ \bar{y}_1 - \bar{y}_2 - H_2(\bar{x}_1 - \bar{x}_2)\ ] K_1 + [\ H_1 + H_2\ ]^2 [\ \frac{1}{m_1} + \frac{1}{m_2}\ ]}{[\ \frac{1}{m_1} + \frac{1}{m_2}\ ]\ [K_1 + K_2] \ +\ K_1K_2\ (\bar{x}_1 - \bar{x}_2)^2}$$

$$C_2 = \frac{[\ (\bar{y}_1 - \bar{y}_2) + H_1^2 + H_2^2\ ]\ [\ K_1 + K_2\ ]\ -\ 2\ H_1H_2\ [\ \frac{1}{m_1} + \frac{1}{m_2}\ ]}{[\ \frac{1}{m_1} + \frac{1}{m_2}\ ]\ [\ K_1 + K_2\ ]\ +\ K_1K_2\ (\bar{x}_1 - \bar{x}_2)^2}$$

Then,

$$F_i = \frac{C_i - B}{SS_e} \cdot \frac{n - 3}{1} \sim F_{1,\ n-3}\ ,\quad \text{for } i = 1,2$$

4. A Procurement Example.

By law, a corporation which is desirous of selling goods or services to the Department of Defense, must become a party to a contract with the US Government. The contract includes, among other things, an agreed upon price (cost plus profit) which the Government is obligated to pay for the product. When the cost of satisfying the contract is uncertain or technical uncertainity is high, an incentive feature and a "target" cost and target profit are introduced. If the contractor's cost is lower than the target cost, then a (previously established) percentage of the savings are returned to the contractor as an increased profit. If, however, the actual cost exceeds the target cost, then a percentage of this cost growth is subtracted from the contractors' profit.

This study concerns only the Cost Plus Incentive Fee (CPIF) type contracts. The CPIF contract type must always state a maximum fee and usually a minimum fee while the Government pays for all allowable costs. Figure 2 illustrates the relationship between cost and profit for this contract type. Notice that the sum of cost and profit is called the price (or total price) so that a broken line relationship also exists between price and cost.

The following symbols will be used throughout:

π_T = target profit

π_M = maximum profit

π_m = minimum profit

C_T = target cost

C_U = greatest lower bound of all costs which should yield profit π_m

C_L = least upper bound of all costs which should yield profit π_M

RELATIONSHIP BETWEEN COST AND
PROFIT FOR A CPIF CONTRACT

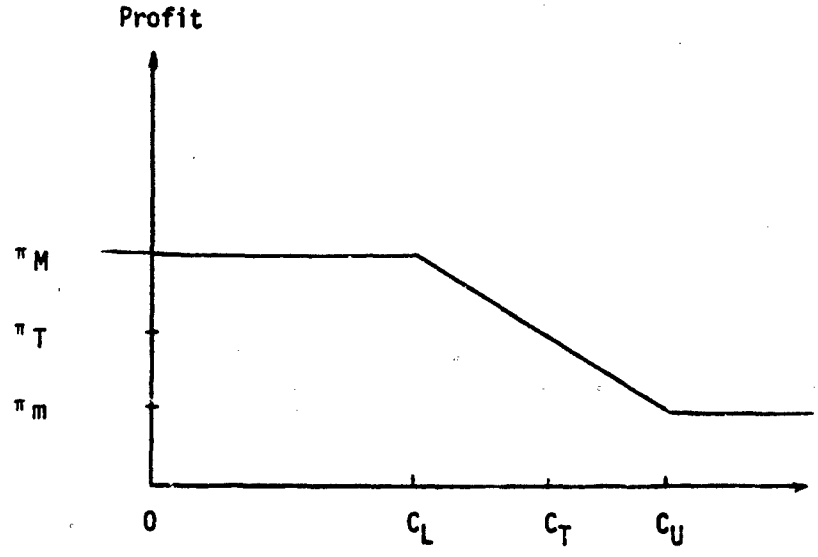


FIGURE 2

The datum set consists of 29 randomly selected CPIF contracts with target price $375,000 or more which were definitized after 1963 and completed prior to September 1971. The data were normalized to unit target cost and target profit and unit difference between the maximum profit and target profit and between the target profit and minimum profit. The costs were similarly transformed. These transformations are linear or strictly piecewise linear if the original relationship is not symmetric with respect to the point (C_T, π_T). The normalized data is presented in Figure 3.

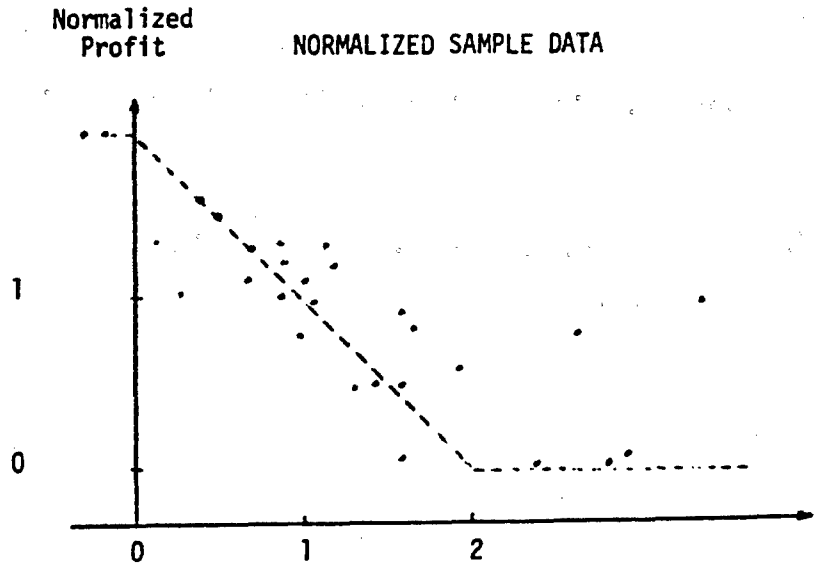


FIGURE 3

Each data point should (theoretically) lie on the dashed line, but for various reasons, variation is introduced into the system. If the contract incentive feature is to properly motivate the contractor then at least the expected normalized profit would coincide with the dashed line. To test this hypothesis, the piecewise general linear model procedure outlined in the preceding sections was used.

Notice that the two points in figure 3 which lie to the left of the origin exhibit no variation. There is (apparently) reason to believe that this will always be the case for points to the left of the vertical axis and to the right of the vertical line at $X=2$ in figure 3.

There is however, considerable variation for the points to the right. Expert advise could not resolve this issue. Therefore, the two leftmost points were discarded in this analysis.

The model used in this section then, is:

$$y=\begin{cases} a_1 + b_1 (x-\bar{x}_1) + \epsilon & , \ x \leq 2 \\ a_2 + b_2 (x-\bar{x}_2) + \epsilon & , \ x > 2 \end{cases}$$

and $a_1 + b_1 (2-\bar{x}_1) = a_2 + b_2 (2 - \bar{x}_2)$;

where y = normalized profit.

x = normalized cost.

The null hypothesis is:

$$H_o: \begin{cases} b_1 = -1 \\ a_2 = 0 \\ b_2 = 0 \end{cases}$$

In the notation of the previous section the notation is:

n=29, p=4, k=1, m=3 and, $\underline{y} = X \underline{\beta} + \underline{\epsilon}$ subject to $T \underline{\beta} = \underline{d}$

where T= (1, 1.12, -1, .65), $\underline{d}'$=(0) and

$$T^* = \left[\begin{array}{cccc} 1, & 1.12, & -1, & .65 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] ; \ \underline{d}^* = \left[\begin{array}{c} 0 \\ \hline -1 \\ 0 \\ 0 \end{array}\right]$$

The data yields the following: $\hat{a}_1=1.09$, $\hat{b}_1=-.84$, $\hat{a}_2=.35$, $\hat{b}_2=.6$,

$\tilde{a}_1=1.08$, $\tilde{b}_1=-.88$, $\tilde{a}_2=.38$, $\tilde{b}_2=.42$, $\frac{\lambda}{2}=.13375$. Also: $T(x'x)^{-1}T'=1.445$ and $(T\underline{\beta}-\underline{d})'(AA')^{-1}(T\underline{\beta}-\underline{d})=.036$ and

$$T^*(x'x)^{-1}T^{*\prime}=\begin{bmatrix} 1.45 & .29 & -.2 & 1.35 \\ .29 & .26 & 0 & 0 \\ -.2 & 0 & .2 & 0 \\ 1.352 & 0 & 0 & 2.08 \end{bmatrix}$$

$$[T^*(x'x)^{-1}T^{*1}]^{-1}=\begin{bmatrix} 22.5 & -25.2 & 22.6 & -14.6 \\ -25.2 & 32.2 & -25.4 & 16.4 \\ 22.6 & -25.4 & 27.7 & -14.7 \\ -14.6 & 16.4 & -14.7 & 10.0 \end{bmatrix}$$

Finally $[T^*\underline{\beta}-\underline{d}^*]'$ = [.189, .12, .35, .6] and $F_{3,24}$ = 3.68**

EXPERIMENTAL ESTABLISHMENT OF ACCURACY OF RANGE-TO-FUNCTION MEASUREMENTS FOR ARTILLERY PROJECTILES

1LT L.D. Clements
Data Reduction Section
Yuma Proving Ground
Yuma, Arizona

FIELD MEASUREMENT OF AIRBURST LOCATION. An accurate means of determining the range-to-function (slant range) is a necessity in testing of artillery fuzing mechanisms. Precise location of ground impacts is not particularly difficult, but exact measurement of airbursts is often a definite problem. Various means are available for locating airbursts directly. The most common is the use of observers and some form of transit to locate the smoke signature by triangulation (digital transit or cinetheodolite being the most accurate). Another direct location method which has been proposed is the use of acoustic sensors to locate a point sound source but the reliability of the acoustic method is questionable.

An indirect means of obtaining the slant range to function is to use the time-velocity record from a Doppler velocimeter and numerically integrate to get slant range. Although the numerical techniques involved are explained more fully in Brittain (1966) and in Clements (1973), briefly the process is this. During the time interval when the Doppler is locked onto the round, successive radial velocity readings are averaged and multiplied by the time interval between readings. The resulting distances traveled in each time increment are summed up to give an estimate of the distance traveled during the locked-on period. The distance the shell traveled before the Doppler locked on is estimated using the muzzle velocity, the first Doppler measured velocity, and the time interval between tube exit and lock-on. Since the Doppler break-track coincides with shell function, the sum of the distances traveled before lock-on and from lock-on to function is the slant range to function. This direct numerical integration is quite good at low gun elevations and a mathematical routine to calculate actual shell tangential velocities for use in integration has been developed.

At Yuma Proving Ground use of observers is the most common means of acquiring slant range data, with use of the Doppler enjoying an increasing interest. Unfortunately, with both acquisition methods, the precision of measurement is known but the actual accuracy of measurement is unknown. The observers only occasionally are able to catch the flash of light accompanying the function, and more generally are sighting on the tell-tale puff of smoke. The relation of the event measured to the actual fuze function is not known. Similarly, the function point on the Doppler record is evidenced by a relatively sudden loss of track. Again, the relation of this break-track point to the actual function point is unknown.

PHOTOGRAPHIC AIRBURST REGISTRATION. In order to obtain an accurate standard against which the other slant range determinations could be compared, photographic techniques were employed. The method used was to emplace a bank of cameras at a point down range such that each camera could look at a rectangular window along the line of fire (see Figure 1). The nominal center of bursts was obtained by plotting data from previous tests and locating the ground range where most of the functions took place. The cameras were located along a line parallel to the nominal line of fire, at a distance of 750 meters from the line of fire. A bank of four cameras was used. Each camera, a Milliken 35mm framing camera with a 40 inch lens, was aimed normal to the line of fire. The cameras were spaced at 17 meter intervals to insure overlap in the fields of view and the aimpoints were stairstepped upwards to follow the trajectory. The "windows" described by the cameras appeared as shown in Figure 2. Time correlation among the several acquisition media was provided by the Proving Ground range timing facility.

Actual data collection was extremely simple. As indicated above, the cameras were prepositioned based upon previous experimental data so no major adjustments were possible. The round-by-round collection sequence consisted only of listening for the sound of the firing over the intercom system, delaying for an appropriate time, and starting the cameras. The cameras were allowed to run for two to three seconds after the sound of the burst was noted. For maximum contrast high speed color film was used. Physical operating limitations consisted of a need for clear skies, preferably with the sun in a position to provide back lighting, and minimal winds.

Data from the cameras were reduced by locating the smoke puff on a frame and backing off until either the puff was no longer visible, or the flash from the fuze function was observed. The location of the function point relative to the center of optics was calculated by ratio and proportion (see Figure 3). Then, knowing the location of the burst along the nominal line of fire and the deflection of the round (from observers), the coordinates of the burst can be calculated. Observed slant range was then calculated from the burst coordinates and gun coordinates.

DATA ANALYSIS. Sample data comparing observer and camera values are given in Table 1. "Extrapolated" data are those events where the actual burst was not within the view field, but was close enough to be estimated from the smoke pattern. The mean error for each of the quantities gives some idea of the overall accuracy of the measurement, while the unbiased estimate of the standard deviation can be taken to indicate the precision.

Similar statistics for the Doppler data may be developed to provide some basis for comparison between the two methods, perhaps through an F-test. Also, as more data are acquired, the mean errors and estimated standard deviations may be refined and the basis for comparison between acquisition methods strengthened.

SUMMARY. This short clinical paper is intended to show an approach to fulfilling an ever present need in testing operations, that of assigning reliable accuracies to experimental data. The example cited is a real one. The analysis of the data, though simple, allows realistic bounds to be placed on data accuracy requirements and comparison of redundant data acquisition methods for future applications.

REFERENCES

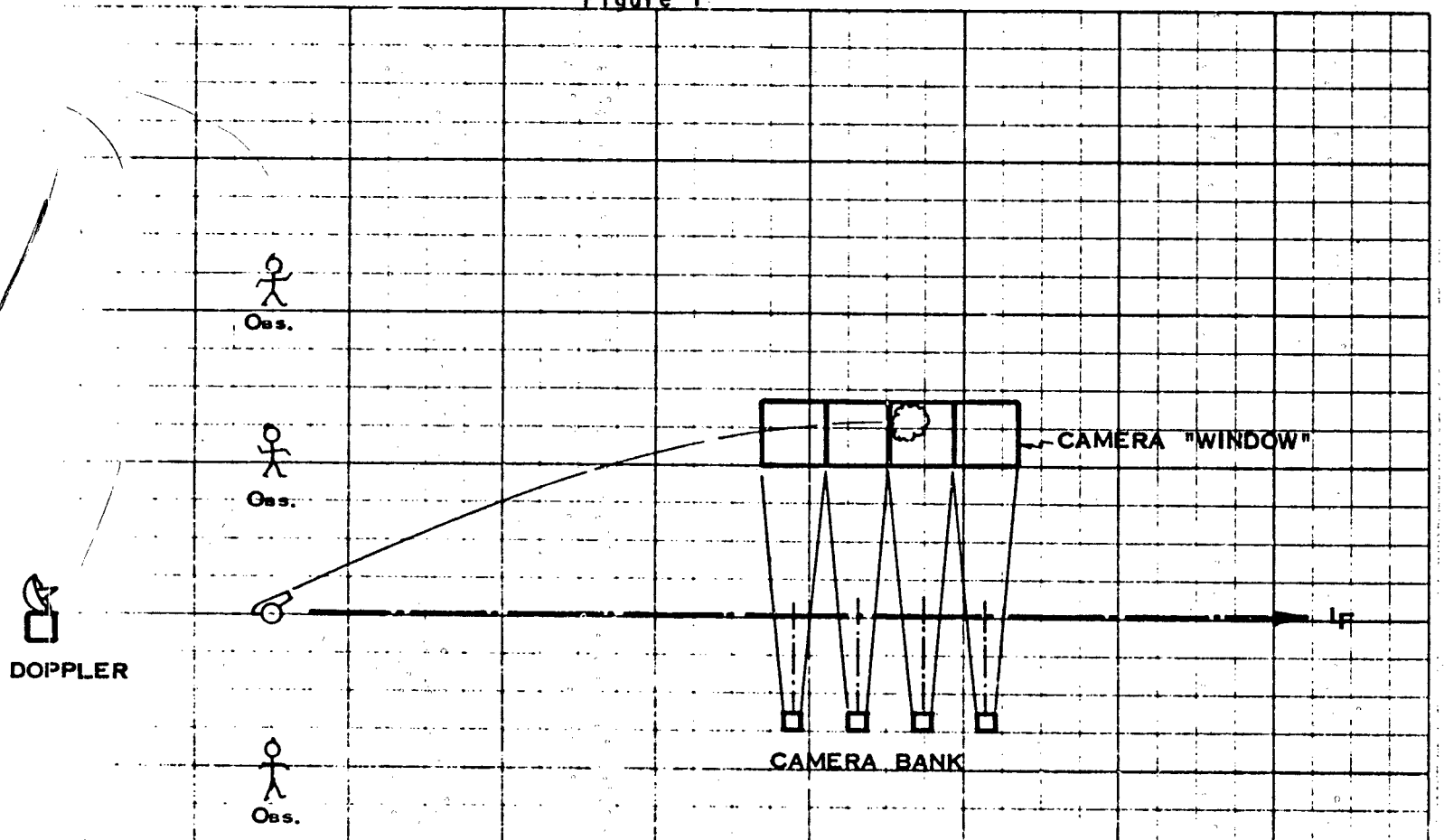
1. Brittain, J.E., "Derivation of Trajectory Data from Records of Doppler Radar and Cameras", USATECOM Report No. DPS-2066, 1966.

2. Clements, L.D., "Doppler Guide", Yuma Proving Ground Technical Memorandum to be published, 1973.

Table 1
Summary of Results for All Rounds

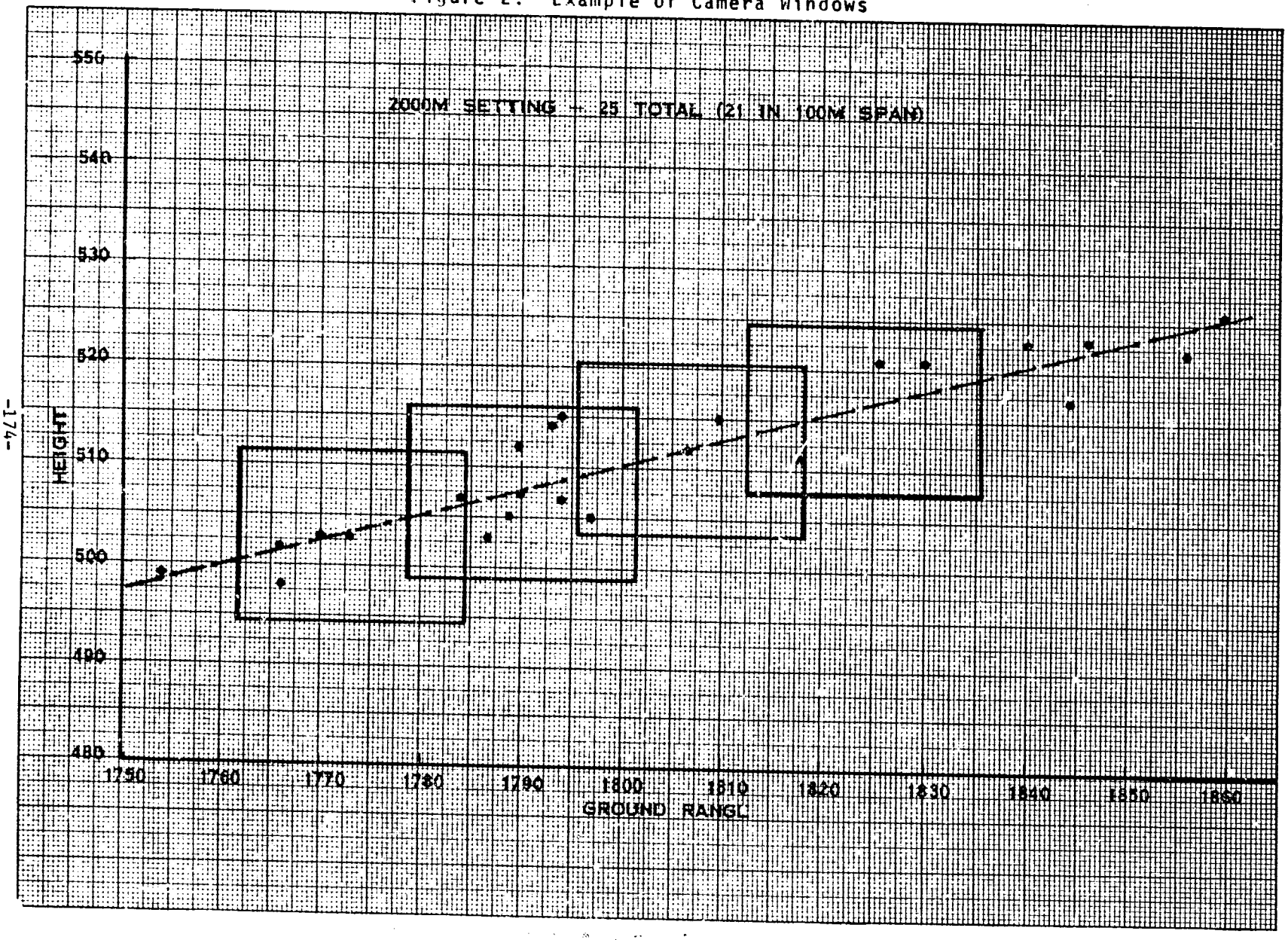
	Ground Range(O-C)			Height(O-C)			Slant Range(O-C)		
Rd.No.	Camera	Obs	Diff	Camera	Obs	Diff	Camera	Obs	Diff
190	2799	2804	+ 5	764	763	- 1	2901	2906	- 5
191	2799	2796	- 3	762	755	- 7	2901	2896	- 5
193	2796	2795	- 1	761	756	- 5	2898	2895	- 3
228	1804	1784	-20	514	495	- 9	1876	1851	-25
231	1804	1785	-19	515	495	-20	1876	1852	-24
236	1791	1807	+16	512	501	+11	1863	1875	+12
241	1820	1835	+15	521	524	+ 3	1893	1908	+15
245	1819	1837	+18	521	518	- 3	1892	1909	+17
247	1788	1807	+19	512	496	-16	1860	1874	+14
250	1820	1838	+18	523	515	- 8	1894	1909	+15
254	1823	1833	+10	526	514	-12	1897	1904	+ 7
256	1792	1786	- 6	514	507	- 7	1864	1857	- 7
Extrapolated									
226	1771	1770	- 1	508	499	- 9	1842	1839	- 3
234	1817	1837	+20	517	516	- 1	1889	1908	+19
235	1768	1778	+10	505	499	- 6	1839	1847	+ 8
	Mean Error=+5.4m			Mean Error=-6.0m			Mean Error=+2.3m		
	$\hat{\sigma}_{GR}$ = 12.87m			$\hat{\sigma}_{H}$ = 7.24m			$\hat{\sigma}_{SR}$ = 13.64m		

Figure 1



SCHEMATIC DIAGRAM OF TEST ARRANGEMENT FOR SLANT RANGE ACCURACY TEST

Figure 2: Example of Camera Windows



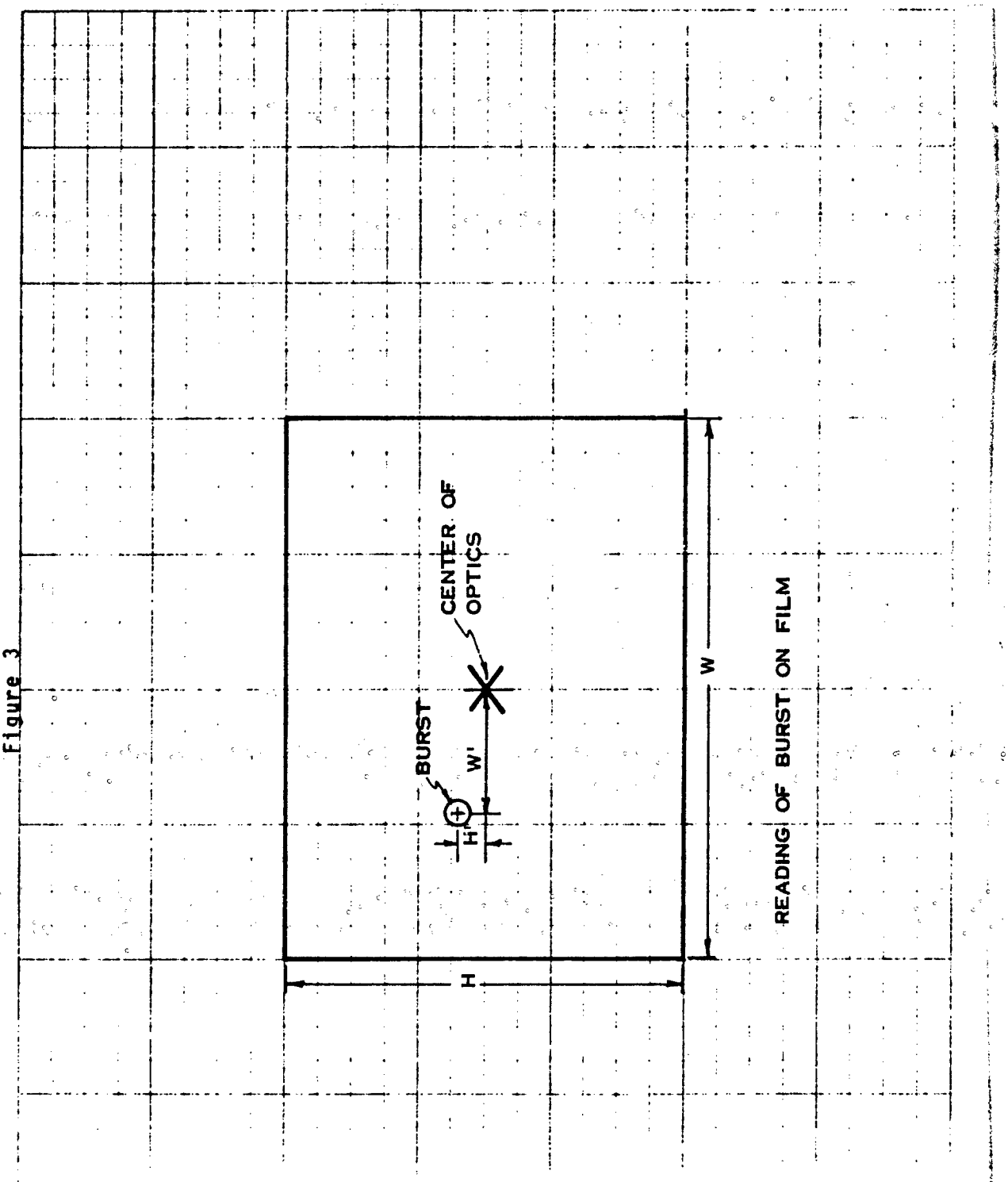


Figure 3

AN IMPROVED METHOD OF ESTIMATING THE CRITICAL VELOCITY OF A PROJECTILE IN PENETRATION BALLISTICS

G. J. McLaughlin
DEFENCE RESEARCH ESTABLISHMENT
VALCARTIER, CANADA

ABSTRACT. In recent years, many studies have been done on the relative merits of several methods of fitting the logistic and specially the normal distribution functions as dosage response curves. Most assessments have been made only for sensitivity experiments where the stimulus had no random fluctuations around the chosen test levels. The purpose of this study is to assess the relative efficiency of some of the methods based on the 'up and down' sampling technique in an experiment where the stimulus has random variations around some fixed levels. One of those methods has been found more efficient than the one currently used to determine the critical velocity of a projectile.

NOTATION.

V: The dosage or stimulus in general; the striking velocity in the case of tests to determine ballistic limits.
σ: A parameter measuring the spread of tolerances in the response curve, usually called standard deviation.
D: Step by which the stimulus (velocity) is increased or decreased depending on whether the previous trial was a failure or a success (in units of σ).
K: Error of estimation for the starting value of V (in units of σ).
S: Standard deviation of the stimulus at each level (in units of σ).
N: Minimum number of observations required for one determination of the 50% point, V50.
NR: Number of determinations of the 50% point for a given set of conditions.
R: Allowable spread for N/2 successes and N/2 failures according to Method B (in units of σ).
RMS: Root mean square error of the NR determinations of the 50% point.
G: A random value from a normal distribution with mean 0 and variance 1.
U: A random value from a uniform distribution between 0 and 1.
$\bar{N}_o$: Average number of observations used to approach V50 in Method A, but not included in sample of N.

1.0 INTRODUCTION. In the last twenty years, there has been much discussion of the relative merits of several methods of fitting the normal integral or the logistic integral response curves to sensitivity data, i.e. to data obtained from experiments in which an increasing proportion of items either fail, explode or die as the severity of the test is increased. In such an experiment the severity of test which would barely produce a failure cannot be measured exactly; one can only observe whether an applied severity

produces a failure or not. Sensitivity data involve responses which can be either positive or negative, and which are observed at different levels of some variable of interest. The response is said to be "quantal" because it is measured not in terms of a continuous scale such as weight or length but in terms of the observed proportion that is positive. The main purpose in analyzing sensitivity data is to estimate the 50% point, i.e. the level of the variable for which the positive and negative responses are equally likely.

In some methods of sensitivity testing, like the Probit, Normit or Logit methods, the experimenter chooses in advance the stimulus levels to be applied and the number of observations at each level. In other methods, like those based on the 'Up and Down'sampling technique, the choices are made sequentially, as the experiment progresses. Up to now, most methods have been applied and their efficiency tested only in the case where the stimulus levels are free of random errors. The object of this study is to assess the relative efficiency of the various methods of sensitivity testing when the dosage or stimulus is subject to random errors around the levels chosen.

After a brief description of the general sensitivity problem, two methods of estimating the 50% point of the stimulus variables are described together with a Monte Carlo simulation procedure used to assess their relative accuracy. The two methods described here use the 'Up and Down' sampling technique to gather the data, but they involve different estimation procedures of the 50% point.

2.0 THE PROBLEM OF SENSITIVITY TESTING. The cumulative normal distribution function has been used extensively in bioassay and in other sensitivity experiments because the probability of some all-or-none response is a monotonic non-decreasing function of a quantity V which measures the potency of the agent producing the response. The occurrence or non-occurrence of the response in a particular individual depends on whether or not the dose exceeds the tolerance value for that individual. Individual tolerances are assumed to have a normal frequency distribution in the population. Therefore the probability that a subject chosen at random from the population will respond to a dose V is given by

$$P = \int_{-\infty}^{W} (2\pi)^{-1/2} \exp(-t^2/2)\, dt. \qquad (1)$$

where $W = (V-V50)/\sigma$, V50 is the value of V corresponding to $P = .5$ and σ is the standard deviation of tolerances in the response function. In a sensitivity experiment, a two-category response is observed to determine the effect of different levels of the dose or stimulus V. Each experiment has the goal of estimating the value of the variable for which the two responses occur with equal probability, i.e. the V50 which is also the mean in a normal distribution. The estimation of the 50% point is more desirable than that of other percentage points for two main reasons. It can be more accurately determined with a reasonable number of observations. Furthermore, it provides the most satisfactory basis for comparison, because for two distribution curves with the same mean but with different values of σ, the only percentage point in common is the 50% one.

3.0 PURPOSE OF THE STUDY. In recent years, many studies have been done on the relative merits of several methods of fitting the logistic and especially the normal distribution functions as dosage response curves. The first standard techniques of analysis were those called Logit or Probit methods depending on whether the assumed response curve was the logistic or the normal distribution functions. They are fully described in References 1 and 2 respectively. Several alternative methods for estimating the 50% point were later suggested, either because they involve less computation or because their validity may depend to a lesser extent on the choice of response curve. Most of these recent methods are based on the 'Up and Down' sampling technique. They are purely arithmetical processes that use the observed responses independently of what the true functional form of the response curve, P, may be. Their merits for any given set of data depend upon the particular form of P that applies. These alternative methods have been described and evaluated in References 3 to 6. Unfortunately all assessments have been made only for sensitivity experiments where the stimulus had no random fluctuations around the chosen test levels.

The purpose of this study is to generalize two of the methods mentioned previously and to assess their efficiency in an experiment where the stimulus has random variations around some fixed levels. This is the case in tests to determine the critical velocity of a projectile to defeat a target since sampling variations in velocity occur for a fixed weight of propellant. Two methods to determine the critical velocity of a projectile or equivalently the ballistic limits of its corresponding armoured target are described herein and their efficiency is assessed for various combinations of the parameters involved.

As far as the authors are aware, Reference 7 is the only existing study on the efficiency of sensitivity testing methods for obtaining the critical velocity. However most of the methods suggested in it are a subset of Method B given in the present study and were evaluated for some particular cases only.

4.0 THE METHODS OF SENSITIVITY TESTING USED FOR DETERMINING CRITICAL VELOCITIES

4.1 General. In experiments to estimate the sensitivity of armour plate to projectile velocity, a common procedure is to fire a given type of projectile at various velocities against a given armour plate. Obviously, there are velocities at which some projectiles will perforate the armour and others will not. It is assumed that those which do not defeat the plate would do so were the projectiles fired with a sufficiently larger velocity. It is therefore assumed that there is a critical velocity, V50, over which a success (defeat of the plate) is more likely and under which a failure is more likely. This critical velocity is the velocity corresponding to 50% successes and 50% failures. On account of the symmetry of the normal distribution, the median velocity, V50, is the same as the mean of the normal integral response function of Section 2. In the case of tests to determine the critical velocity, the parameter σ in the response function measures the spread of tolerances of a type of armour with respect to the striking velocities of a given type of projectile. It should not be confused with the parameter S which measures the spread in striking velocity corresponding to a fixed weight of propellant.

It is assumed that the probability of response (defeat of target), P, to the stimulus (striking velocity), V, is given by an integrated normal curve with parameters V50 and σ. According to Reference 7, this assumption is supported by results of tests involving the firing of a considerable number of rounds at one target.

4.2 DESCRIPTION OF THE METHODS. The purpose of the two methods studied here is to estimate V50 under the assumption that the normal response function is valid.

For the two methods, an attempt is made, by adjusting the weight of propellant, to fire the first round at a velocity which is the best estimate of the V50 known to the experimenter, say VT_1. Each subsequent round is fired according to the 'Up and Down' firing technique, increasing the velocity by D (in units of σ) for any round following a failure, and decreasing it by D for any round following a success. The series of velocities used in this 'Up and Down' experiment form a stochastic process, whose main feature is that the velocities tend to have a distribution concentrated around the V50. Method A uses the fact that an initial run of responses of the same sign is an indication that the first velocity was badly chosen. If the initial run of constant sign contains N_o+1 rounds, another N-1 rounds are fired. In this case, the estimator is

$$\overline{VA} = \frac{1}{N+1}\left[\sum_{i=N_o+1}^{N_o+N} V_i + V_{N_o+N} \pm D\right] \qquad (2)$$

where the sign associated with D is positive if the last round was a failure and negative if it was a success. Method B requires that firing should continue until N/2 successes and N/2 failures are achieved within a range of velocities of R units. The estimator $\overline{VB}$ suggested by this method is the arithmetic mean of the N velocities corresponding to the N/2 successes and N/2 failures.

4.3 SIMULATION OF THE METHODS. No attempt was made to assess the relative merits of those methods from actual firing data because such a procedure would have involved a tremendous number of rounds besides loosing its generality through its association with specific weapons. The error inherent in each method of estimating the 50% point was evaluated using Monte Carlo techniques. A program was written to simulate on a computer the complete firing procedure for the two estimation methods.

Without any loss of generality for the present study, the variables were scaled in such a way that V50 = 0 and σ = 1 in the response function. The velocity of the j^{th} round for the two methods can therefore be simulated using the following equation:

$$V_j = K + G_j S + D \sum_{i=1}^{j} C_i \quad \text{for } j=1,2,\ldots(N+N_o) \qquad (3)$$

where K = error of estimation for the starting value of V

G_j = a random value from a normal distribution with zero mean and unit standard deviation

C_j = 0 for $j = 1$

-1 for $U_{j-1} \leq P_{j-1}$ and $j = 2,\ldots(N+N_o)$

$+1$ for $U_{j-1} > P_{j-1}$ and $j = 2,\ldots(N+N_o)$

U_j = a uniformly distributed random real number between 0 and 1

P_j = value of response function when $V = V_j$

Each group of $(N+N_o)$ velocities thus calculated yields one estimate of V50 which is obtained by averaging according to the appropriate formula of Section 4.2. This process is repeated as many times as required to allow the computation of a RMS error for each method and for each combination of parameters. A complete description of the simulation programs is given in Appendix A of Reference 8.

4.4 RANGE OF PARAMETERS AND TYPE OF RESPONSE FUNCTION. A RMS error based on 3500 determinations of the 50% point has been computed for each method and for all conbinations of the following values of the parameters: D=.5, 1., 2., S=0, .5, 1., K=0, 1,2,4. The values of N were 6, 9, 12, 15 for method A. In the case of method B, which is based on the first X successes and X failures within a velocity spread of R, the RMS error has been evaluated for the following pairs of (X,R) values: (2,2), (3,2), (5,2), (2,3), (3,3), and (5,3).

The most frequently used functional forms for the probability of response to a stimulus are the normal and the logistic distribution functions. Since the curves corresponding to those functions are almost identical, only the normal integral has been used as response function throughout this study.

The RMS error calculated with the simulation program has been expressed in units of σ as were D, S, K, and R. The RMS error of each method is given in the first two Tables for the combinations of parameters mentioned previously.

The RMS errors of method B for D=1, (X,R)=(5,2) and (3,3) check with those of Figure 10, Reference 7, when interpolating over S in Table B1.

5.0 DATA REDUCTION AND ANALYSIS.

5.1 WEIGHTING OVER THE PARAMETERS. Since the purpose of this study is to find the best method of sensitivity testing for critical velocity determination, the RMS errors have been averaged over the various parameters using the weighting system which appeared the most realistic in critical velocity estimation problems.

The right portion of Tables A1 and B1 gives for each method the RMS error averaged over K according to normal distributions with a common mean $\bar{K}=0$ but different standard deviations σ_K=1.5, 2.5 and 4.0. This means that the initial velocity estimate is assumed to follow a normal distribution centered on the true V50 with standard deviations of 1.5, 2.5 and 4.0 times the basic parameter σ. The value of σ_K=2.5 is believed to be more appropriate for most applications, unless a preliminary "feeler" round is fired to improve the accuracy of the initial estimate of V50, in which case the value

σ_K=1.5 appears more realistic. This is in agreement with the set of starting velocities used in Reference 7, which corresponds to values of σ_K ranging between 1 and 2 in units of σ, with one "feeler" round to improve the starting velocity.

The average number of observations used in approaching V50 for Method A, but not included in the sample of N, is designated $\bar{N}_O$ and given in Table A1.

The average sample size, $N+\bar{N}_O$, on which each RMS error of Method B is based, is given in Table B1.

The best distribution for S to cover the range of weapons is not known, but since it is definitively heavily concentrated around 0.5, a triangular distribution between 0 and 1 and centered at 0.5 has been assumed realistic for S. Therefore the RMS values of Tables A1 and B1 already weighted over K according to normal distributions with σ_K=1.5,2.5 and 4.0 have been averaged over S according to the set of weights corresponding to a triangular distribution of S between 0 and 1 with center at 0.5, that is 0.125, 0.75, 0.125, for S=0, 0.5, 1., respectively. The resulting RMS errors weighted over K and S are given in Tables A2 and B2.

5.2 SELECTION OF THE OPTIMUM STEP LEVEL. On account of the variable sample size, it is not obvious from Tables A2 and B2 which level of D is optimum for the two methods. A graphical comparison of the RMS error associated with each level of D is made in Figures 1 to 6. The curves indicate that the optimum level of D is 1 for Method A and Method B, whether σ_K is 2.5 or 4.0. From Figures 1 and 2 of Reference 8, this is also true for Method A when σ_K is 1.5, but not for Method B where D=0.5 is better. Therefore a value of one σ for D should be aimed at, since it is associated with a greater accuracy for both methods over the values of σ_K likely to be met in practice.

5.3 COMPARISON OF THE METHODS. The curves plotted in Figure 7 indicate clearly that Method A is superior to Method B with R equals to either 2 or 3 when σ_K=2.5 and the step level D is one. The same conclusion can be drawn from Figures 8 and 9 for σ_K=4.0 and 1.5 respectively. Therefore Method A is definitively more accurate than Method B for σ_K=1.5, 2.5 and 4.0 when the step level D is at its optimum value of one σ.

In a critical velocity test, unfortunately, the parameter σ is not known in advance. It is therefore necessary to use for σ a reasonable estimate, $\hat{\sigma}$ based upon experience. If no such estimate is available, the value of 50 ft/s which is recommended in Reference 7, appears to be realistic. It is assumed here that $\hat{\sigma}$ has a normal distribution with mean σ and standard deviation $\sigma/4$. Therefore, the step D being taken equal to $\hat{\sigma}$ it also has a normal distribution with mean σ and standard deviation $\sigma/4$. The RMS error values of Tables A2 and B2 were averaged over D using the set of weighting factors corresponding to this distribution, and the resulting RMS values are given in Tables A3 and B3. The weights were .16, .82 and .02 for D equal to 0.5, 1.0 and 2.0 respectively. The confidence that the V50 estimate is between the true V25 and V75 is also given in Tables A3 and B3, for Methods A and B respectively. The RMS error and the confidence in the V50 estimate are plotted in Figure 10, for both methods when σ_K is 2.5 and D is normally distributed with mean 1 and standard deviation .25. The curves on Figure 10 illustrate again the superiority of Method A over Method B.

5.4 ACCURACY vs SAMPLE SIZE FOR METHOD A. The authors are not aware of any agreed level of accuracy required from a sensitivity testing method. However, a level of accuracy such that one is 90% confident that a V50 estimate is between V25 and V75 is considered desirable and realistic. Assuming that the error, K, in the initial value of V is normally distributed with mean 0 and standard deviation σ_K=2.5 (an acceptable assumption when no preliminary "feeler" round is used), a confidence level of 90% would require a total sample size of 16.48 on the average with Method A. This total sample is made up of $\bar{N}_o$+1 = 2.48, which is the average length of the initial run of identical responses, and N-1 = 14, which is the fixed number of observations after the run of $\bar{N}_o$+1. In this case the average number of observations used in approaching V50, but not included in the sample of N, is $\bar{N}_o$ = 1.48 and the subsequent sample on which V50 is based has size N = 15. Such a sample would yield a confidence of 90.9% according to Table A3.

An interesting feature of Method A is that it ignores any initial run of identical responses (an indication that the initial V was badly chosen) and therefore produces an estimate of V50 which has a guaranteed accuracy independent of the error K in the initial V. Of course, the greater K or σ_K for a fixed N, the longer the initial run of rejected values N_o and therefore the greater the total sample size, $N+N_o$, required to achieve a given accuracy.

6.0 CONCLUSIONS AND RECOMMENDATIONS

6.1 CONCLUSIONS. Two methods A and B to evaluate the 50% point in a sensitivity test when the stimulus has random variations have been assessed by Monte Carlo simulation, and Method A has been found superior to the other over a realistic range of error in the starting value of the stimulus. It is more accurate and therefore more economical than Method B which is currently used in critical velocity determination.

The optimum step level D by which the stimulus is increased or decreased was determined to be around 1σ. However the accuracy provided by Method A is not too sensitive to variations up to 50% in the size of this step level. Therefore the performance of the estimate of V50 provided by Method A is not sensitive to errors in the guessed value of σ.

Method A requires on the average 16.48 observations to insure a 90% confidence that the estimate of V50 lies between V25 and V75. This number is made up of an average of 1.48 observations that are rejected, followed by a sequence of observations with a predetermined length of 15.

6.2 RECOMMENDATIONS. It is recommended that Method A be used to evaluate the critical velocity required from a given projectile to defeat a target, since it is more accurate than Method B and also can be handled more quickly and more easily than Method B during a field trial to determine the V50. In an experiment using Method A, the steps are:

a) Select from past experience an estimate, $\hat{\sigma}$, of the parameter σ in the response function. Otherwise, use $\hat{\sigma}$ equal to 50 ft/s as an estimate since the procedure requires that σ be known within rough limits.
b) Choose N in advance. A value of N=15 will yield a 90.9% confidence that the V50 estimate is within V25 and V75.

c) Fire the first shot at a velocity as close as possible to an initial guess of V50.

d) Carry out a series of trials, increasing the velocity by $\hat{\sigma}$ ft/s following a failure and decreasing it by $\hat{\sigma}$ ft/s following a success. This is done by regulating carefully the weight of propellant for each step.

e) Continue firing until the chosen nominal sample size N is reached. If N_o+1 responses are alike at the beginning, the total number of trials is N_o+N.

f) Use as an estimate of V50 the average,

$$\overline{VA} = \frac{1}{N+1}\left[\sum_{i=N_o+1}^{N_o+N} V_i + V_{N_o+N} \pm \hat{\sigma}\right]$$

where the plus sign is associated with a failure in the last trial and the minus one with a success.

REFERENCES

1. Finney, D.J., "The Estimation of the ED 50 for a Logistic Response Curve", Sankhya, Vol. 12, Parts 1 and 2, 1952.

2. Finney, D. J., "Probit Analysis - A Statistical Treatment of the Sigmoid Response Curve", Cambridge University Press, 1952.

3. Dixon, W.J., Mood, A.M., "A Method for Obtaining and Analyzing Sensitivity Data", Journal of the American Statistical Association, Vol. 43, 1948.

4. Brownlee, K.A., Hodges, J.L., Rosenblatt, M., "The "Up and Down" Method with Small Samples", Journal of the American Statistical Association, Vol. 48, 1953.

5. Tsutakawa, R.K., "Random Walk Design in Bio-Assay", Journal of the American Statistical Association, Vol. 57, 1962.

6. Brown, B.W., "Some Properties of the Spearman Estimator in Bio-Assay", Biometrika, Vol. 48, Parts 3 and 4, 1961.

7. Feroli, J.A., "The Accuracy and Reproducibility of Several Methods for Obtaining Ballistic Limits of Armor", Aberdeen Proving Ground, Development and Proof Services, Report No. DPS/TB4-10/9, July 1957. (UNCLASS)

8. McLaughlin, G.J., and Labbé, J.C., "An Improved Method of Estimating the Critical Velocity of a Projectile in Penetration Ballistics", DREV Report 668/72, August 1972, UNCLASSIFIED.

TABLE A1

RMS ERROR AND $\bar{N}_o$ OF METHOD A

D	S	N	K=0		K=±1		K=±2		K=±4		σ_K=1.5		σ_K=2.5		σ_K=4.0	
			RMS	$\bar{N}_o$	RMS	$\bar{N}_o$	RMS	$\bar{N}_o$	RMS	$\bar{N}_o$	RMS	$\bar{N}_o$	RMS	$\bar{N}_o$	RMS	$\bar{N}_o$
0.5	0.0	6	0.371	0.38	0.517	0.98	0.677	2.55	0.720	6.52	0.546	1.52	0.605	2.66	0.646	3.82
0.5	0.0	9	0.329	0.35	0.434	0.99	0.554	2.58	0.586	6.53	0.455	1.53	0.499	2.67	0.530	3.83
0.5	0.0	12	0.303	0.34	0.376	1.00	0.461	2.57	0.495	6.56	0.391	1.53	0.424	2.68	0.449	3.84
0.5	0.0	15	0.292	0.36	0.338	0.98	0.406	2.60	0.431	6.49	0.352	1.53	0.377	2.66	0.398	3.82
1.0	0.0	6	0.464	0.17	0.530	0.49	0.617	1.28	0.654	3.26	0.546	0.76	0.580	1.33	0.605	1.91
1.0	0.0	9	0.414	0.17	0.438	0.49	0.503	1.29	0.521	3.27	0.455	0.76	0.476	1.34	0.491	1.92
1.0	0.0	12	0.363	0.17	0.385	0.49	0.418	1.29	0.429	3.29	0.391	0.76	0.403	1.34	0.412	1.92
1.0	0.0	15	0.336	0.17	0.359	0.50	0.376	1.31	0.397	3.26	0.360	0.77	0.370	1.34	0.379	1.92
2.0	0.0	6	0.558	0.03	0.620	0.13	0.683	0.49	0.700	1.47	0.627	0.27	0.651	0.54	0.668	0.82
2.0	0.0	9	0.509	0.03	0.507	0.13	0.537	0.51	0.565	1.47	0.519	0.27	0.531	0.54	0.542	0.82
2.0	0.0	12	0.439	0.02	0.465	0.13	0.473	0.50	0.518	1.48	0.463	0.27	0.476	0.54	0.489	0.82
2.0	0.0	15	0.405	0.02	0.398	0.14	0.427	0.49	0.457	1.49	0.409	0.27	0.421	0.54	0.433	0.83
0.5	0.5	6	0.390	0.33	0.554	0.94	0.758	2.45	0.848	6.27	0.600	1.47	0.680	2.56	0.739	3.68
0.5	0.5	9	0.358	0.41	0.463	0.97	0.599	2.34	0.668	6.28	0.492	1.47	0.548	2.56	0.590	3.67
0.5	0.5	12	0.325	0.39	0.403	0.97	0.489	2.44	0.562	6.25	0.420	1.48	0.462	2.56	0.496	3.67
0.5	0.5	15	0.304	0.40	0.363	0.95	0.431	2.38	0.477	6.23	0.375	1.46	0.406	2.53	0.430	3.64
1.0	0.5	6	0.484	0.20	0.554	0.49	0.654	1.24	0.732	3.19	0.576	0.76	0.622	1.31	0.659	1.87
1.0	0.5	9	0.435	0.19	0.473	0.50	0.540	1.25	0.566	3.17	0.488	0.76	0.512	1.30	0.530	1.87
1.0	0.5	12	0.377	0.19	0.418	0.48	0.443	1.23	0.478	3.19	0.418	0.74	0.434	1.30	0.449	1.87
1.0	0.5	15	0.352	0.18	0.370	0.50	0.397	1.22	0.427	3.19	0.376	0.74	0.390	1.30	0.402	1.86
2.0	0.5	6	0.593	0.04	0.640	0.15	0.704	0.51	0.750	1.46	0.652	0.28	0.680	0.55	0.704	0.82
2.0	0.5	9	0.521	0.04	0.529	0.15	0.570	0.50	0.598	1.46	0.542	0.28	0.557	0.54	0.571	0.82
2.0	0.5	12	0.459	0.03	0.473	0.14	0.503	0.50	0.520	1.46	0.480	0.28	0.492	0.54	0.501	0.82
2.0	0.5	15	0.414	0.03	0.426	0.15	0.443	0.51	0.453	1.46	0.429	0.28	0.436	0.55	0.442	0.82
0.5	1.0	6	0.502	0.48	0.653	0.87	0.890	2.03	1.094	5.65	0.720	1.32	0.832	2.28	0.922	3.28
0.5	1.0	9	0.425	0.45	0.532	0.90	0.729	2.02	0.909	5.56	0.592	1.32	0.686	2.26	0.763	3.24
0.5	1.0	12	0.379	0.45	0.463	0.86	0.612	2.05	0.735	5.67	0.505	1.32	0.574	2.28	0.629	3.29
0.5	1.0	15	0.344	0.49	0.421	0.84	0.526	2.05	0.622	5.63	0.446	1.31	0.498	2.27	0.540	3.27
1.0	1.0	6	0.506	0.28	0.637	0.52	0.770	1.14	0.894	2.99	0.672	0.75	0.736	1.25	0.789	1.77
1.0	1.0	9	0.475	0.27	0.530	0.50	0.613	1.14	0.692	3.01	0.550	0.74	0.590	1.25	0.624	1.77
1.0	1.0	12	0.430	0.26	0.466	0.50	0.516	1.14	0.586	2.98	0.478	0.73	0.507	1.24	0.532	1.76
1.0	1.0	15	0.391	0.26	0.416	0.50	0.469	1.15	0.497	2.98	0.429	0.74	0.449	1.24	0.465	1.76
2.0	1.0	6	0.677	0.08	0.695	0.19	0.791	0.49	0.856	1.41	0.726	0.30	0.763	0.55	0.794	0.80
2.0	1.0	9	0.577	0.08	0.584	0.20	0.647	0.50	0.673	1.42	0.605	0.31	0.625	0.56	0.641	0.81
2.0	1.0	12	0.499	0.09	0.523	0.18	0.544	0.50	0.505	1.40	0.526	0.30	0.541	0.55	0.555	0.80
2.0	1.0	15	0.460	0.08	0.462	0.19	0.494	0.50	0.521	1.42	0.473	0.30	0.486	0.55	0.497	0.81

TABLE B1

RMS ERROR AND $(N+\bar{N}_o)$ OF METHOD B

N/2	R	D	S	K=0		K=±1		K=±2		K=±4		σ_K=1.5		σ_K=2.5		σ_K=4.0	
				RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$
2	2	0.5	0.0	0.456	4.68	0.789	6.23	0.812	8.19	0.626	12.24	0.717	6.66	0.718	7.99	0.698	9.29
2	2	0.5	0.5	0.534	4.85	0.683	6.34	0.695	8.33	0.652	12.32	0.649	6.79	0.658	8.11	0.658	9.40
2	2	0.5	1.0	0.592	6.26	0.629	7.26	0.646	9.10	0.644	13.00	0.625	7.79	0.632	9.01	0.637	10.22
2	2	1.0	0.0	0.642	4.59	0.743	5.52	0.692	6.48	0.682	8.47	0.701	5.68	0.697	6.36	0.693	7.00
2	2	1.0	0.5	0.633	5.41	0.686	6.35	0.678	7.33	0.666	9.35	0.670	6.52	0.671	7.20	0.670	7.86
2	2	1.0	1.0	0.627	6.98	0.650	7.69	0.660	8.66	0.651	10.70	0.647	7.92	0.650	8.58	0.651	9.23
2	2	2.0	0.0	0.894	5.28	0.696	5.66	0.846	6.27	0.842	7.29	0.800	5.81	0.813	6.17	0.823	6.51
2	2	2.0	0.5	0.668	8.02	0.654	8.53	0.658	8.93	0.658	9.96	0.659	8.58	0.658	8.90	0.658	9.23
2	2	2.0	1.0	0.669	9.14	0.657	9.67	0.639	10.14	0.650	11.17	0.655	9.73	0.652	10.08	0.650	10.41
3	2	0.5	0.0	0.422	6.71	0.658	8.45	0.593	10.48	0.531	14.45	0.581	8.85	0.576	10.20	0.564	11.50
3	2	0.5	0.5	0.473	7.21	0.553	8.82	0.546	10.85	0.537	14.89	0.531	9.25	0.535	10.61	0.536	11.92
3	2	0.5	1.0	0.500	9.53	0.515	10.72	0.518	12.47	0.537	16.45	0.513	11.17	0.519	12.41	0.524	13.63
3	2	1.0	0.0	0.555	6.77	0.582	7.71	0.589	8.74	0.546	10.72	0.575	7.89	0.572	8.58	0.565	9.24
3	2	1.0	0.5	0.535	8.33	0.553	9.29	0.531	10.35	0.529	12.34	0.541	9.48	0.538	10.18	0.535	10.84
3	2	1.0	1.0	0.518	10.73	0.523	11.57	0.536	12.65	0.526	14.54	0.525	11.79	0.527	12.46	0.527	13.10
3	2	2.0	0.0	0.787	8.36	0.474	8.28	0.726	9.35	0.727	10.38	0.651	8.69	0.674	9.12	0.692	9.51
3	2	2.0	0.5	0.540	12.34	0.537	13.23	0.550	13.39	0.532	14.38	0.541	13.10	0.540	13.40	0.539	13.70
3	2	2.0	1.0	0.533	14.17	0.534	14.70	0.528	15.14	0.539	16.20	0.532	14.76	0.533	15.10	0.535	15.44
5	2	0.5	0.0	0.364	10.77	0.487	12.57	0.428	14.63	0.410	18.67	0.438	12.96	0.433	14.34	0.426	15.67
5	2	0.5	0.5	0.392	11.92	0.424	13.77	0.409	15.74	0.411	19.78	0.411	14.12	0.411	15.49	0.411	16.80
5	2	0.5	1.0	0.404	16.20	0.411	17.39	0.417	19.24	0.410	23.17	0.411	17.87	0.412	19.12	0.411	20.35
5	2	1.0	0.0	0.414	11.28	0.446	12.32	0.423	13.32	0.426	15.32	0.431	12.47	0.429	13.16	0.428	13.82
5	2	1.0	0.5	0.412	14.41	0.414	15.42	0.411	16.43	0.412	18.44	0.413	15.58	0.412	16.28	0.412	16.94
5	2	1.0	1.0	0.402	18.57	0.410	19.36	0.405	20.25	0.415	22.38	0.407	19.55	0.408	20.21	0.410	20.87
5	2	2.0	0.0	0.674	15.05	0.336	13.42	0.596	16.11	0.593	17.08	0.528	14.75	0.548	15.38	0.563	15.93
5	2	2.0	0.5	0.410	21.12	0.419	22.55	0.418	22.25	0.413	23.20	0.416	22.13	0.416	22.37	0.415	22.62
5	2	2.0	1.0	0.422	24.36	0.418	25.01	0.410	25.28	0.413	26.41	0.416	24.98	0.415	25.31	0.414	25.64

TABLE B1 (Cont'd)

RMS ERROR AND $(N+\bar{N}_o)$ OF METHOD B

N/2	R	D	S	K=0		K=±1		K=±2		K=±4		σ_K=1.5		σ_K=2.5		σ_K=4.0	
				RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$
2	3	0.5	0.0	0.444	4.64	0.773	6.22	1.180	8.23	0.653	12.25	0.836	6.65	0.865	8.00	0.829	9.30
2	3	0.5	0.5	0.496	4.71	0.760	6.22	0.957	8.22	0.654	12.20	0.752	6.67	0.766	7.99	0.745	9.28
2	3	0.5	1.0	0.622	5.11	0.715	6.26	0.774	8.08	0.702	11.95	0.708	6.74	0.718	7.97	0.716	9.18
2	3	1.0	0.0	0.583	4.56	0.811	5.47	0.940	6.45	0.665	8.46	0.791	5.65	0.794	6.33	0.767	6.98
2	3	1.0	0.5	0.665	4.68	0.781	5.54	0.795	6.57	0.685	8.55	0.752	5.74	0.747	6.42	0.732	7.07
2	3	1.0	1.0	0.699	5.33	0.728	6.13	0.741	7.08	0.714	9.08	0.723	6.32	0.724	6.98	0.722	7.62
2	3	2.0	0.0	0.897	5.28	0.674	5.57	0.838	6.29	0.841	7.29	0.791	5.82	0.805	6.18	0.817	6.52
2	3	2.0	0.5	0.827	5.57	0.736	6.02	0.793	6.56	0.807	7.58	0.779	6.13	0.786	6.48	0.792	6.82
2	3	2.0	1.0	0.753	6.59	0.749	6.97	0.743	7.53	0.743	8.59	0.748	7.10	0.746	7.46	0.745	7.81
3	3	0.5	0.0	0.418	6.71	0.637	8.47	0.989	10.49	0.520	14.48	0.703	8.86	0.722	10.22	0.686	11.53
3	3	0.5	0.5	0.458	6.78	0.640	8.48	0.746	10.50	0.526	14.54	0.625	8.89	0.628	10.25	0.606	11.56
3	3	0.5	1.0	0.535	7.49	0.606	8.80	0.596	10.62	0.561	14.70	0.583	9.24	0.582	10.53	0.576	11.80
3	3	1.0	0.0	0.518	6.61	0.646	7.49	0.679	8.51	0.561	10.49	0.621	7.68	0.620	8.36	0.606	9.02
3	3	1.0	0.5	0.553	6.90	0.610	7.81	0.618	8.79	0.580	10.81	0.596	7.99	0.597	8.67	0.593	9.32
3	3	1.0	1.0	0.585	8.06	0.589	8.85	0.603	9.78	0.579	11.79	0.591	9.04	0.590	9.69	0.587	10.33
3	3	2.0	0.0	0.791	8.39	0.490	8.33	0.721	9.38	0.721	10.39	0.656	8.73	0.675	9.15	0.691	9.54
3	3	2.0	0.5	0.686	8.71	0.567	8.99	0.670	9.74	0.663	10.75	0.633	9.21	0.643	9.59	0.650	9.95
3	3	2.0	1.0	0.605	10.08	0.606	10.49	0.606	11.04	0.599	12.08	0.605	10.61	0.604	10.97	0.603	11.31
5	3	0.5	0.0	0.354	10.75	0.492	12.52	0.710	14.60	0.396	18.58	0.527	12.92	0.536	14.30	0.510	15.61
5	3	0.5	0.5	0.393	10.89	0.484	12.67	0.607	14.75	0.418	18.76	0.465	13.07	0.464	14.45	0.453	15.77
5	3	0.5	1.0	0.439	12.46	0.457	13.78	0.443	15.64	0.441	19.68	0.448	14.23	0.446	15.52	0.444	16.79
5	3	1.0	0.0	0.423	10.64	0.503	11.53	0.481	12.57	0.436	14.55	0.474	11.73	0.469	12.42	0.460	13.07
5	3	1.0	0.5	0.448	11.46	0.462	12.35	0.453	13.38	0.447	15.42	0.455	12.56	0.454	13.26	0.452	13.91
5	3	1.0	1.0	0.458	13.65	0.457	14.40	0.452	15.34	0.454	17.38	0.456	14.61	0.455	15.27	0.454	15.91
5	3	2.0	0.0	0.665	15.13	0.326	13.39	0.592	16.08	0.595	17.10	0.503	14.59	0.506	15.04	0.561	15.93
5	3	2.0	0.5	0.546	15.33	0.408	14.70	0.528	16.32	0.526	17.33	0.486	15.43	0.499	15.92	0.508	16.37
5	3	2.0	1.0	0.459	17.23	0.456	17.66	0.463	18.26	0.455	19.35	0.459	17.79	0.459	18.17	0.458	18.53

TABLE A2

RMS ERROR AND $\bar{N}_o$ OF METHOD A

D	N	K=0		K=±1		K=±2		K=±4		σ_K=1.5		σ_K=2.5		σ_K=4.0	
		RMS	$\bar{N}_o$	RMS	$\bar{N}_o$	RMS	$\bar{N}_o$	RMS	$\bar{N}_o$	RMS	$\bar{N}_o$	RMS	$\bar{N}_o$	RMS	$\bar{N}_o$
0.5	6	0.403	0.40	0.563	0.94	0.766	2.41	0.868	6.22	0.610	1.46	0.692	2.54	0.754	3.65
0.5	9	0.364	0.41	0.469	0.97	0.611	2.37	0.693	6.22	0.502	1.46	0.561	2.53	0.607	3.64
0.5	12	0.330	0.39	0.408	0.96	0.503	2.41	0.579	6.21	0.428	1.47	0.473	2.54	0.509	3.64
0.5	15	0.308	0.40	0.368	0.94	0.441	2.36	0.492	6.18	0.382	1.45	0.415	2.52	0.441	3.62
1.0	6	0.492	0.21	0.582	0.49	0.665	1.23	0.745	3.18	0.586	0.75	0.632	1.30	0.670	1.86
1.0	9	0.438	0.20	0.476	0.50	0.545	1.24	0.578	3.16	0.492	0.75	0.518	1.30	0.538	1.86
1.0	12	0.382	0.20	0.420	0.48	0.450	1.23	0.487	3.18	0.423	0.74	0.440	1.30	0.456	1.86
1.0	15	0.355	0.19	0.375	0.50	0.404	1.22	0.433	3.17	0.381	0.75	0.395	1.29	0.408	1.86
2.0	6	0.600	0.04	0.645	0.15	0.713	0.50	0.758	1.45	0.659	0.28	0.688	0.54	0.711	0.82
2.0	9	0.527	0.04	0.534	0.15	0.576	0.50	0.604	1.46	0.547	0.28	0.563	0.55	0.576	0.82
2.0	12	0.462	0.04	0.479	0.14	0.505	0.50	0.528	1.45	0.484	0.28	0.496	0.54	0.507	0.82
2.0	15	0.419	0.04	0.427	0.15	0.448	0.50	0.463	1.46	0.432	0.28	0.441	0.55	0.448	0.82

TABLE B2

RMS ERROR AND $(N+\bar{N}_o)$ OF METHOD B

N/2	R	D	K=0		K=±1		K=±2		K=±4		σ_K=1.5		σ_K=2.5		σ_K=4.0	
			RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$	RMS	$N+\bar{N}_o$
2	2	0.5	0.532	5.01	0.691	6.44	0.705	8.41	0.648	12.40	0.655	6.90	0.663	8.21	0.661	9.49
2	2	1.0	0.634	5.50	0.689	6.41	0.678	7.39	0.666	9.41	0.671	6.59	0.672	7.27	0.671	7.92
2	2	2.0	0.700	7.61	0.660	8.31	0.682	8.75	0.683	9.78	0.678	8.37	0.679	8.71	0.680	9.04
3	2	0.5	0.470	7.44	0.563	9.01	0.549	11.00	0.536	15.03	0.535	9.44	0.538	10.78	0.539	12.08
3	2	1.0	0.535	8.44	0.553	9.37	0.539	10.43	0.531	12.41	0.544	9.57	0.541	10.26	0.538	10.92
3	2	2.0	0.576	12.07	0.529	12.80	0.572	13.10	0.561	14.11	0.555	12.76	0.558	13.08	0.560	13.39
5	2	0.5	0.390	12.31	0.431	14.07	0.412	16.04	0.411	20.07	0.415	14.45	0.414	15.80	0.413	17.10
5	2	1.0	0.411	14.54	0.417	15.53	0.412	16.52	0.414	18.54	0.414	15.69	0.414	16.38	0.414	17.04
5	2	2.0	0.453	20.77	0.409	21.72	0.443	21.86	0.440	22.83	0.432	21.56	0.434	21.87	0.436	22.16
2	3	0.5	0.507	4.75	0.756	6.23	0.953	8.20	0.660	12.17	0.758	6.67	0.773	7.99	0.753	9.27
2	3	1.0	0.660	4.74	0.778	5.60	0.808	6.62	0.686	8.60	0.753	5.80	0.750	6.48	0.735	7.13
2	3	2.0	0.827	5.66	0.730	6.09	0.792	6.85	0.804	7.67	0.777	6.21	0.783	6.57	0.790	6.90
3	3	0.5	0.464	6.83	0.636	8.52	0.764	10.51	0.530	14.56	0.630	8.93	0.635	10.28	0.613	11.59
3	3	1.0	0.553	7.01	0.612	7.90	0.624	8.88	0.578	10.90	0.599	8.08	0.599	8.76	0.594	9.41
3	3	2.0	0.690	8.85	0.563	9.09	0.669	9.86	0.663	10.87	0.632	9.32	0.642	9.71	0.650	10.07
5	3	0.5	0.394	11.07	0.482	12.79	0.529	14.84	0.418	18.85	0.471	13.20	0.471	14.57	0.460	15.88
5	3	1.0	0.446	11.63	0.467	12.50	0.456	13.52	0.447	15.56	0.458	12.71	0.456	13.41	0.453	14.06
5	3	2.0	0.552	15.54	0.405	14.91	0.529	16.53	0.527	17.55	0.485	15.62	0.495	16.09	0.509	16.58

TABLE A3

RMS, $\bar{N}_o$ AND CONFIDENCE IN V50 ESTIMATES OF METHOD A

N	σ_K=1.5			σ_K=2.5			σ_K=4.0		
	RMS	$\bar{N}_o$	CONF	RMS	$\bar{N}_o$	CONF	RMS	$\bar{N}_o$	CONF
6	0.591	0.86	74.5	0.643	1.48	70.6	0.685	2.13	70.6
9	0.495	0.86	82.7	0.526	1.48	80.0	0.551	2.12	80.0
12	0.425	0.85	88.8	0.447	1.48	86.9	0.466	2.13	86.9
15	0.382	0.85	92.2	0.400	1.48	90.9	0.414	2.12	90.9

TABLE B3

RMS ERROR, $(N+\bar{N}_o)$, AND CONFIDENCE IN V50 ESTIMATES OF METHOD B

N/2	R	σ_K=1.5			σ_K=2.5			σ_K=4.0		
		RMS	$N+\bar{N}_o$	CONF	RMS	$N+\bar{N}_o$	CONF	RMS	$N+\bar{N}_o$	CONF
2	2	0.668	6.67	68.7	0.670	7.45	68.6	0.669	8.20	68.6
3	2	0.543	9.61	78.6	0.541	10.40	78.8	0.538	11.16	79.0
5	2	0.415	15.61	89.6	0.414	16.40	89.6	0.414	17.16	89.7
2	3	0.755	5.95	62.9	0.755	6.72	62.9	0.739	7.47	63.9
3	3	0.605	8.24	73.5	0.606	9.02	73.4	0.598	9.77	74.0
5	3	0.460	12.85	85.7	0.459	13.65	85.8	0.455	14.40	86.2

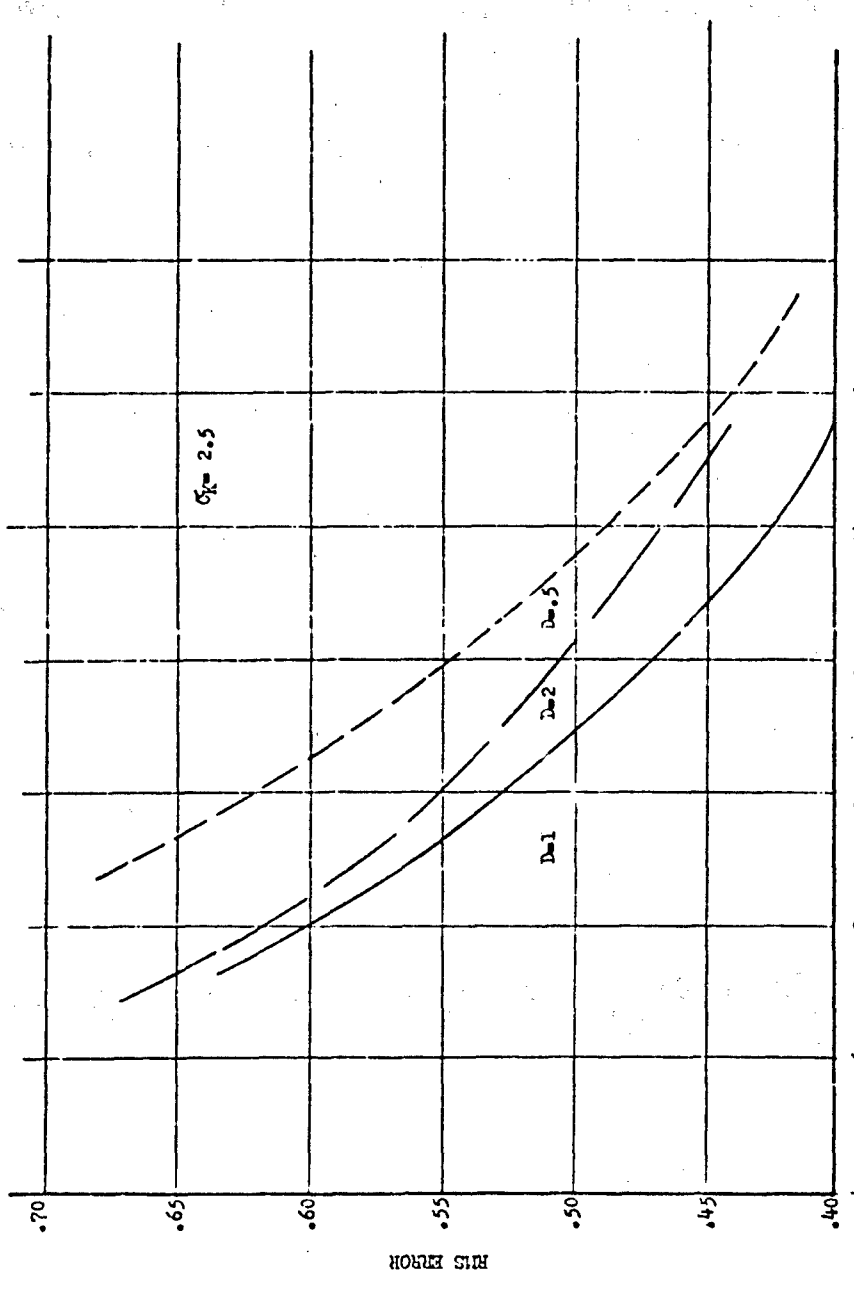


FIGURE 1 - RMS ERROR VS $(N+\bar{N}_o)$ FOR METHOD A

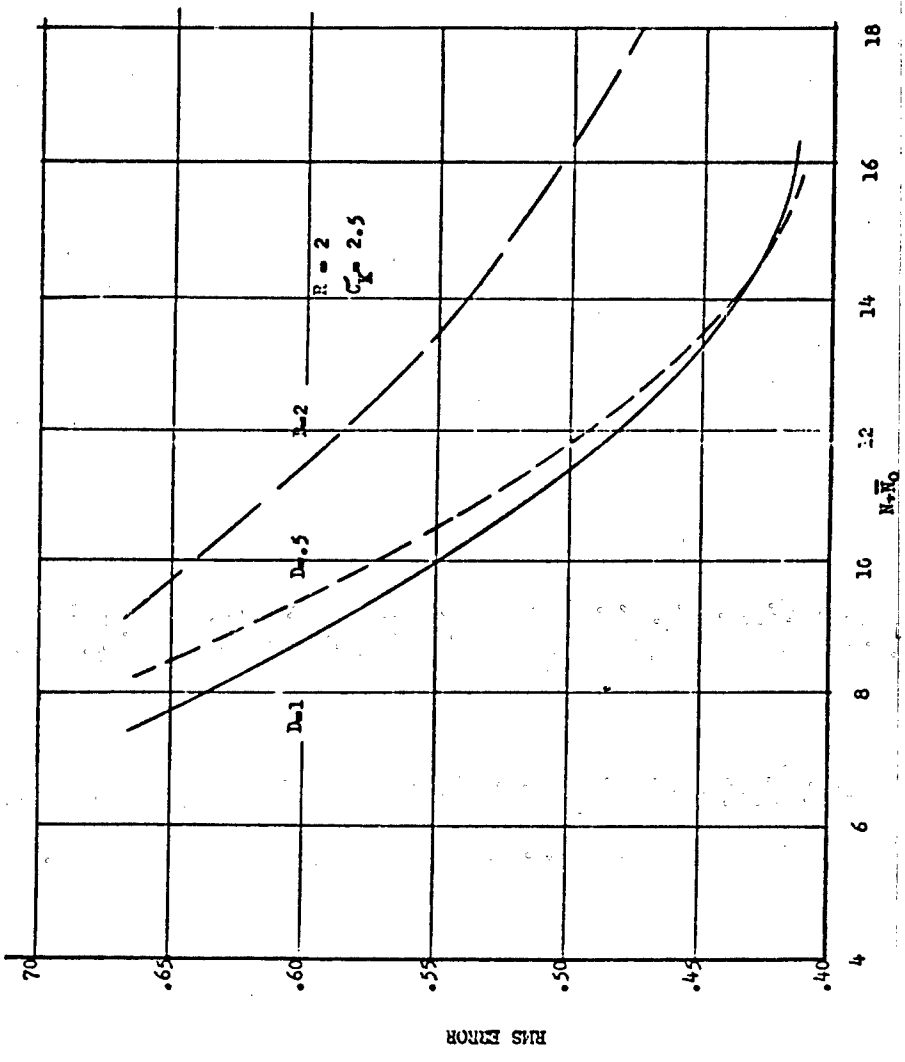


FIGURE 2 - RMS ERROR VS ($N+\bar{N}_0$) FOR METHOD B

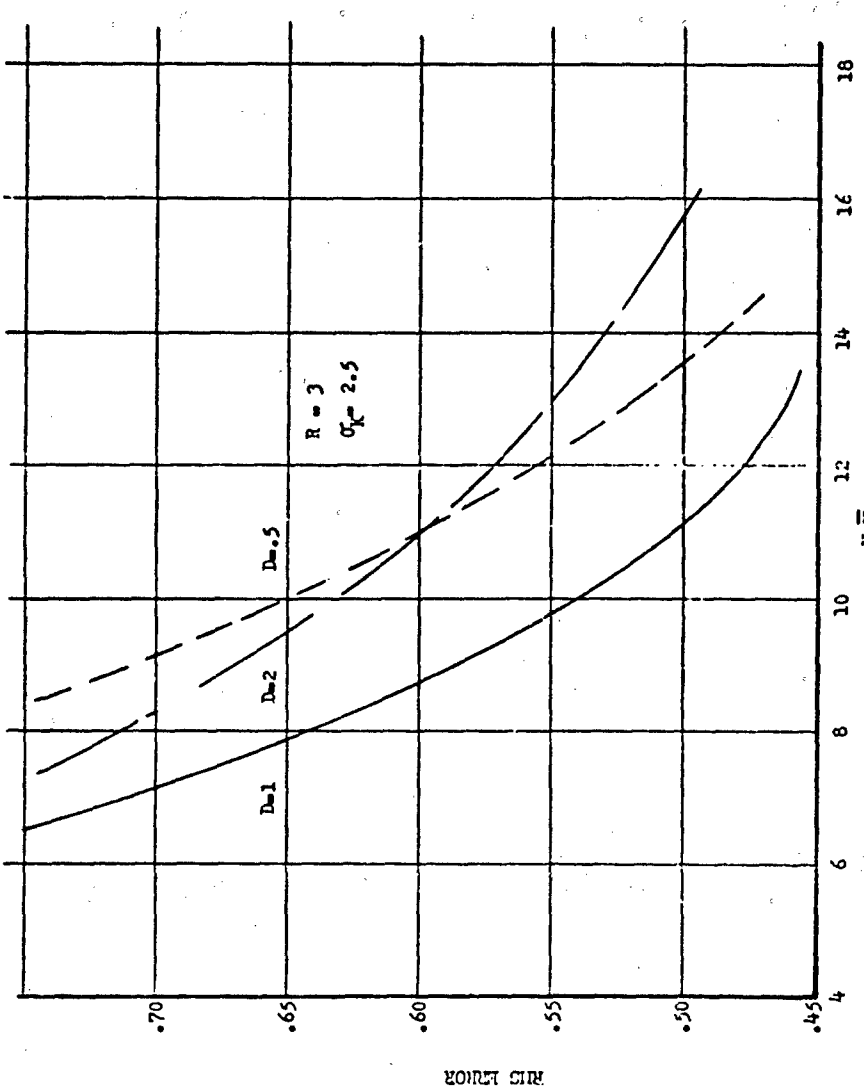


FIGURE 3 - RMS ERROR VS $(M+\bar{N}_0)$ FOR METHOD B

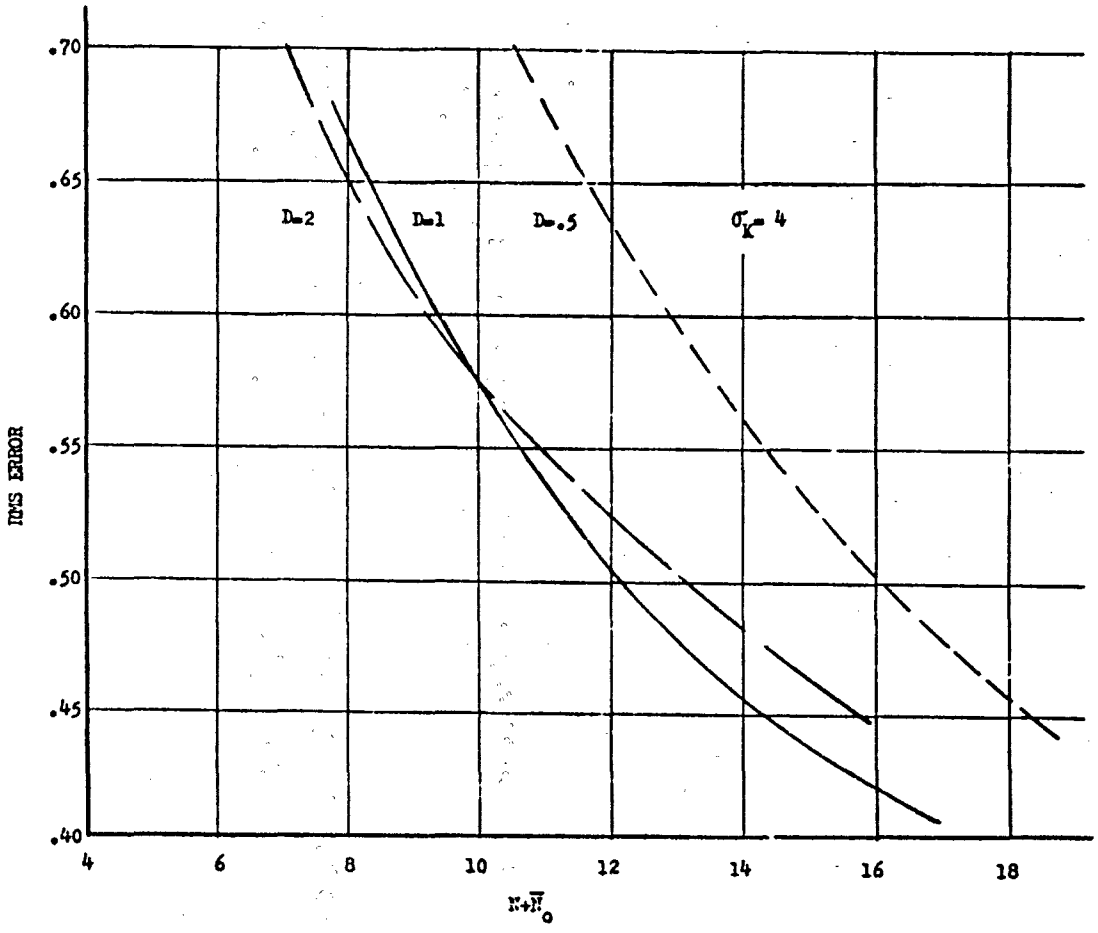


FIGURE 4 - RMS ERROR VS $(N+\bar{N}_o)$ FOR METHOD A

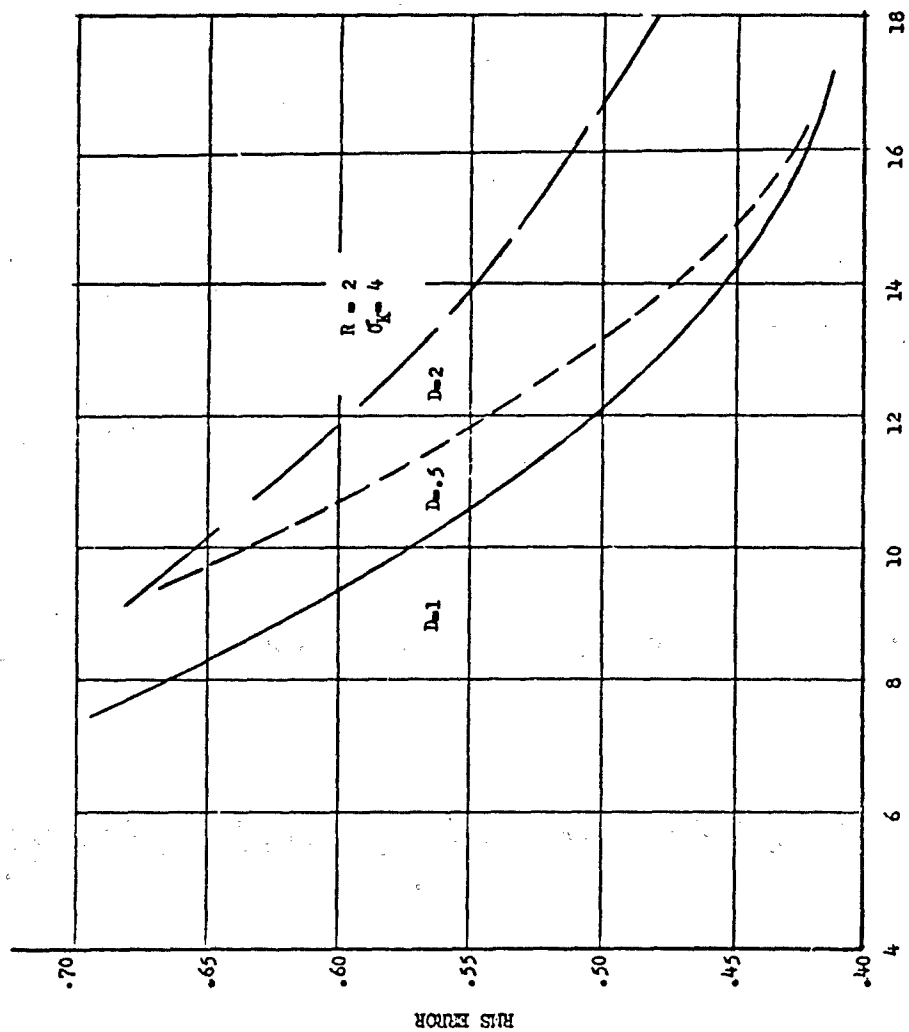


FIGURE 5 - RMS ERROR VS $(H+\bar{H}_c)$ FOR METHOD B

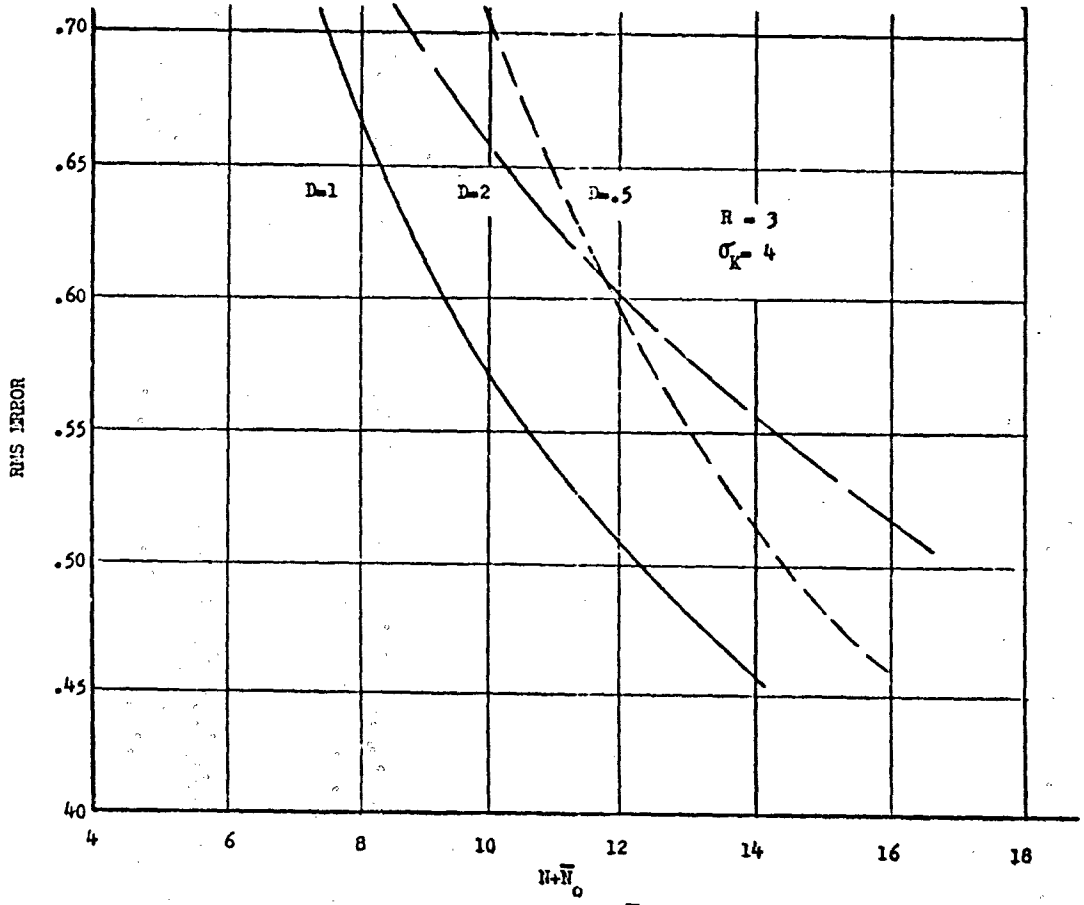


FIGURE 6 - RMS ERROR VS. $(N+\bar{N}_o)$ FOR METHOD B

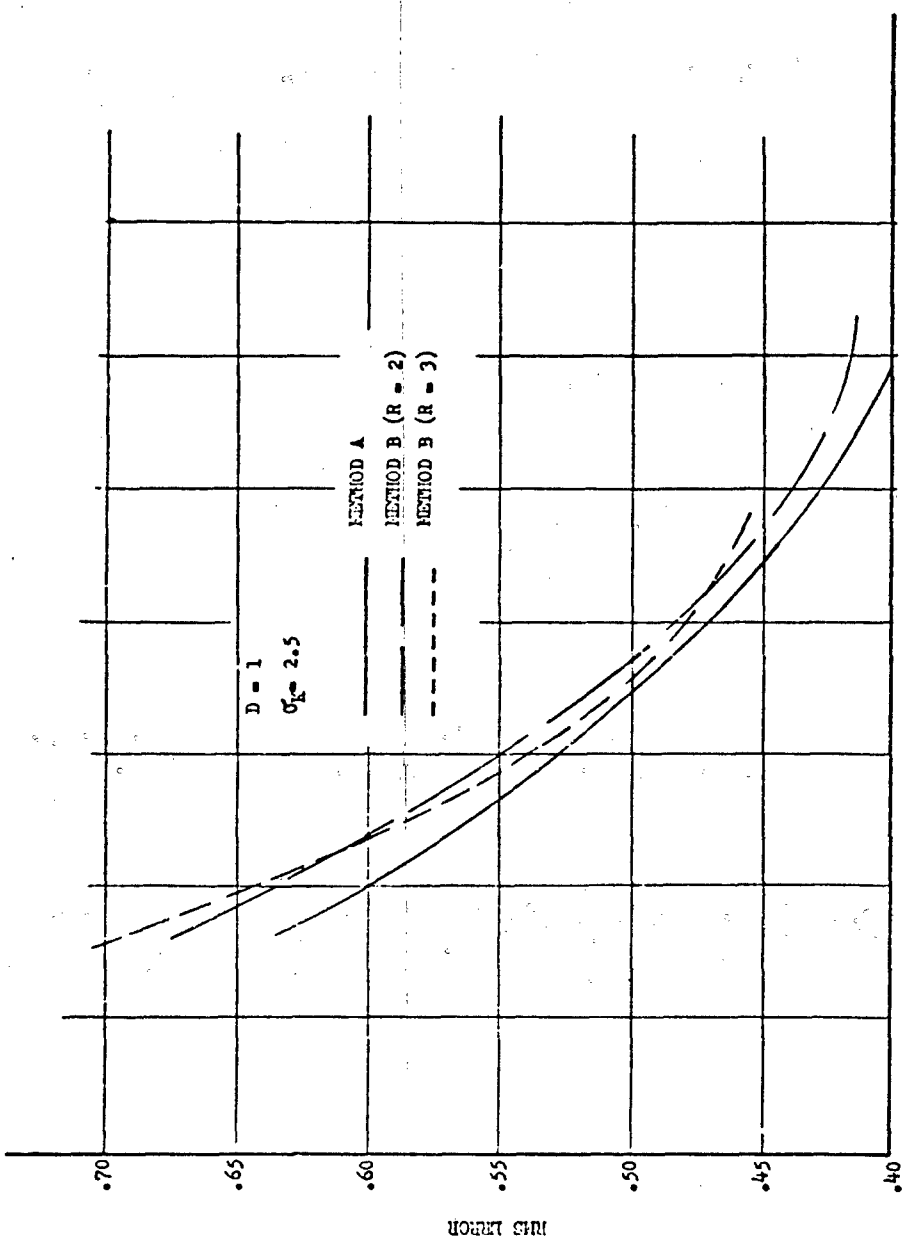


FIGURE 7 - RMS ERROR VS ($N+\bar{T}_o$) FOR METHOD A & B

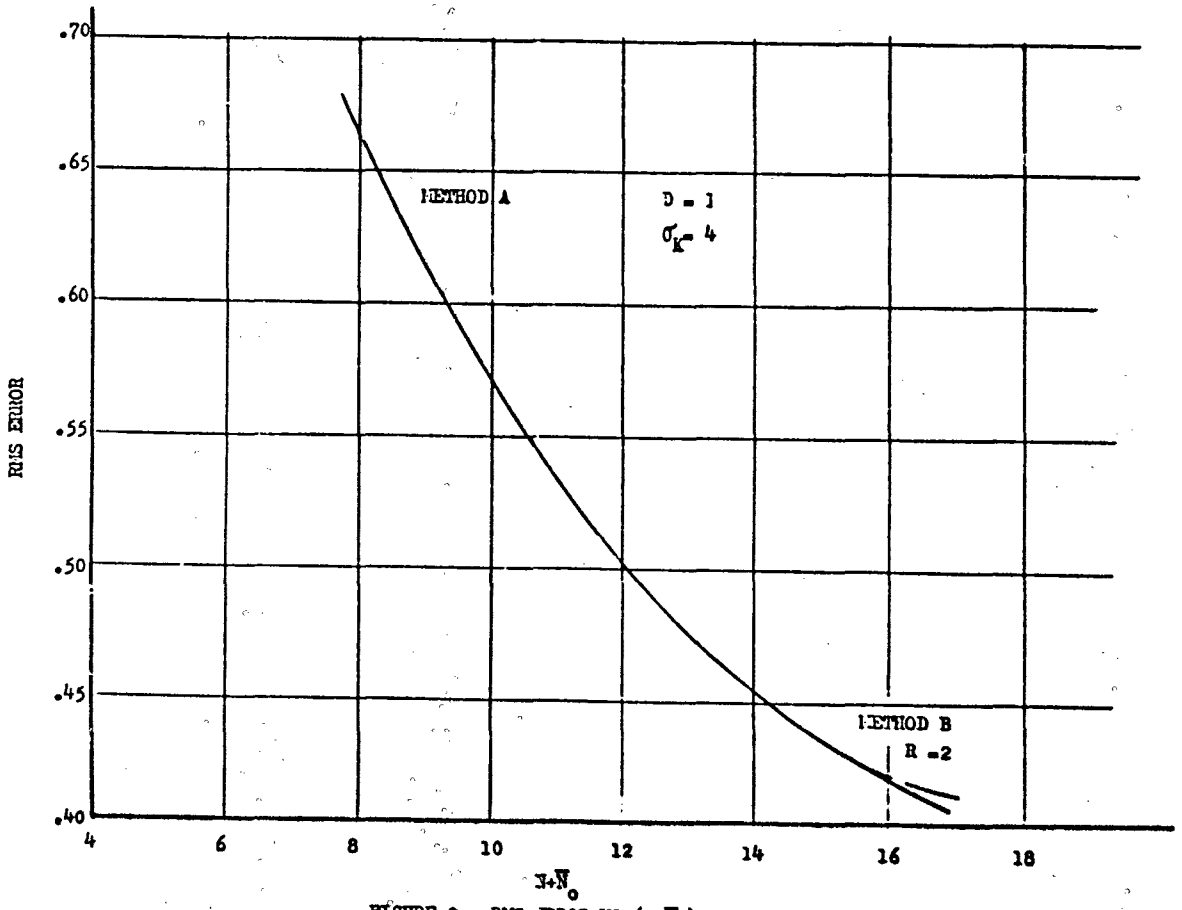


FIGURE 8 - RMS ERROR VS $(N+\bar{N}_o)$

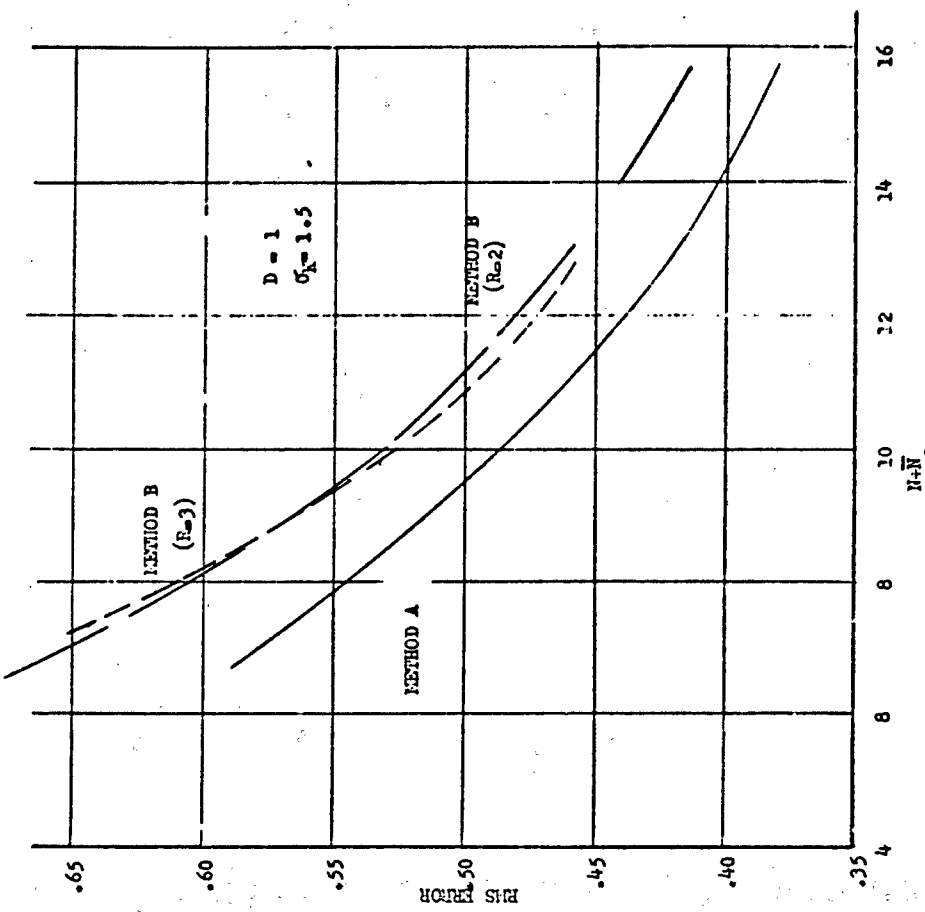


FIGURE 9 - RMS ERROR VS $(M+\bar{N}_o)$

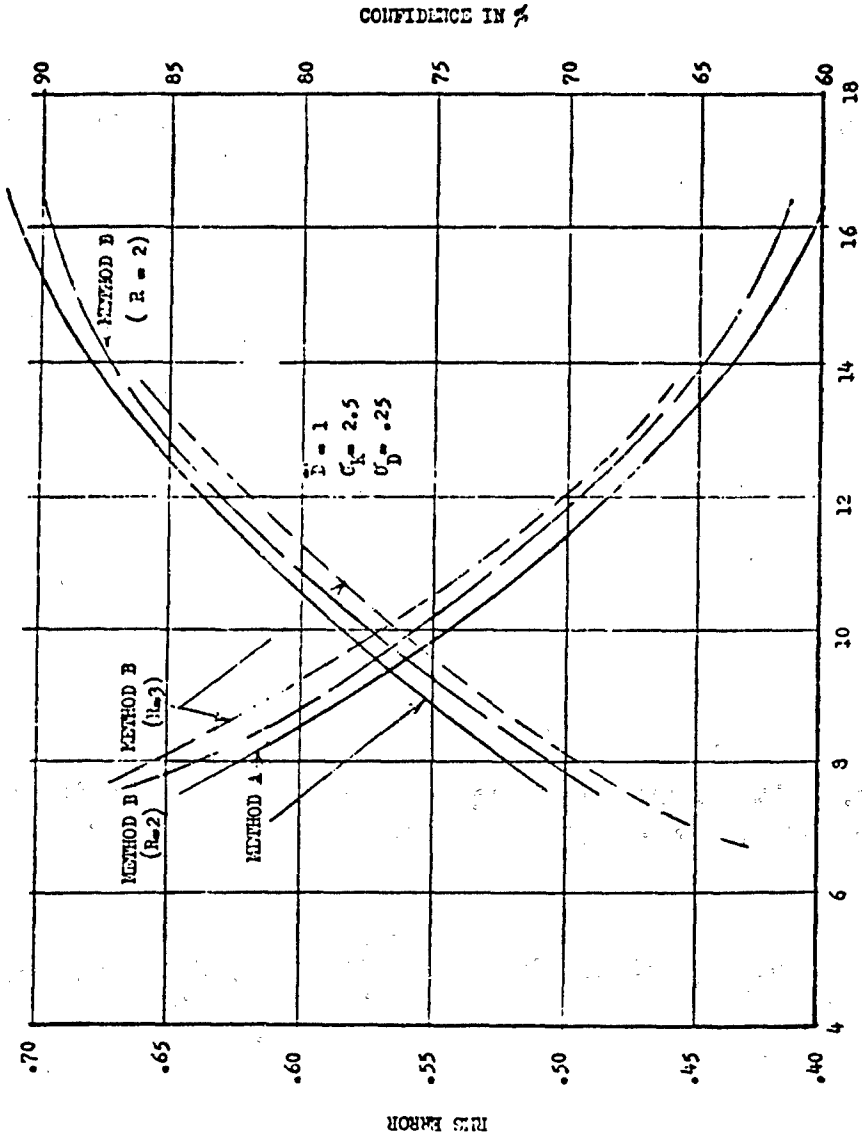


FIGURE 10 - RMS ERROR AND CONFIDENCE VS $(N+\bar{N}_o)$

EVALUATING AND SCHEDULING PROTOTYPE REQUIREMENTS FOR SUITABILITY TESTING

Major Richard B. Cole & Major William J. Owen
U. S. Army Infantry Board
Fort Benning, Georgia

ABSTRACT. This paper addresses the problem of developing a schedule for suitability testing of the prototype of a complex item. The sequential approach discussed involves ordering the requirements against which the prototype is to be evaluated and then using this ordered set of requirements as a basis for sequencing the subtests included in the suitability test. Emphasis is placed on developing the ordered set of requirements. A model based on the criteria recommended by Fishburn and by Moore and Baker is developed for mapping the requirements from a randomly arranged set to an unconstrained test sequence. A linear model is developed which is determined to be an acceptable normative model. Based on the results of this model a method is proposed for partitioning requirements into subtests and for sequencing subtests into a constrained test schedule.

PURPOSE. The purpose of this presentation is to develop a method useful in scheduling a suitability test for the prototype of a complex item. A suitability test is a test designed to evaluate the prototype in order to determine if the item represented by the prototype is suitable for production. The overall test of the prototype will usually involve evaluating the prototype against many, often related, requirements. Generally, a separate test is required for the evaluation of one or more related requirements. Consequently, the suitability test will actually consist of a series of individual tests, or subtests. The specific problem to be addressed is to determine a method of scheduling the subtests to maximize the rate in which information, relative to the potential suitability of the item, is generated during the suitability test.

BACKGROUND. In the development of a complex item of equipment, it is common for the equipment to undergo a research and development (R&D) cycle of several years in length and to incur R&D costs of several million dollars. One of the last phases of the R&D cycle is the development and test of a prototype of the item. The actual test of the prototype can be quite expensive and time-consuming and can directly affect the final cost and final availability date of the end item. Consequently, if this phase of the cycle could most efficiently serve its purpose, then an important portion of the cost and developmental time of the end item could be minimized.

The purpose of the suitability test of a prototype is to provide information upon which a decision of item disposition can be made. The decision usually will be to determine whether the item represented by the prototype should be accepted and placed into production, accepted contingent upon certain modifications, retained for further development, or rejected from further consideration. This decision may have to be made prior to the completion of the suitability test; thus it is essential to maximize the flow of information.

The overall suitability test will consist of a series of intermediate tests each designed to evaluate the prototype against one or more specifications or operational requirements. Each of the subtests derives specific information about the prototype. The information accumulated from all subtests then serves as a basis for the decision relating to the final disposition of the item.

The time required to make the decision on equipment disposition directly relates to the time required to accumulate sufficient information upon which the decision can be based. Consequently, it is desirable that the subtests be scheduled so as to maximize the rate of information generated. This is obviously a particularly important criterion in the scheduling of prototypes of items required for an immediate need. On the other hand, care must be exercised so as to prevent a premature decision on item disposition. Obviously, an incorrect decision could result in accepting an expensive but unsatisfactory piece of equipment, or it could result in delaying the production of a suitable item.

The problem of developing a test schedule which will maximize the rate of information generated is compounded and made more important by the fact that there is frequently no predetermined stopping rule upon which the decision on item disposition can be made. For example, it may be undesirable to decide before the test that if a certain per cent of the operational requirements are not met, then the testing will stop and the item will be rejected. This type of stopping rule may be unsatisfactory since the performance of the prototype against other requirements may be so outstanding as to overshadow its failures, or the degree of failure may be more important than the failure itself.

CONCEPT. There are multiple factors relating to a suitability test which influence the desired sequencing of its subtests. These factors must, of course, relate to the amount of potential information which could be gained from executing the subtest. An illustrative factor pertaining to the amount of information is the importance of the requirements tested. For example, the information gained from evaluating the prototype against an essential requirement would contribute more information upon which to base the decision of item disposition than would evaluating against a relatively minor requirement. However, there may be several factors which warrant consideration. In the tests considered in this research, five factors were identified as influencing the desired relative placement in the testing sequence, and these factors were found to be of varying degrees of relative importance.

In addition to the factors, the degree to which each factor would apply to each requirement must be considered. The possible degrees of applicability of a factor to the requirements in the suitability test are defined as the categories of the factor. In this research, methods were developed for identifying, weighting as to relative importance, and categorizing each factor applicable to the suitability test of a prototype. The factors, factor weights, and factor categories are considered to be suitability test dependent.

Once the factors, factor weights, and factor categories have been determined for a prototype, each requirement is described in terms of the degree to which the factors apply to the requirement and to the subtest needed for the evaluation.

Obviously there are trade-offs to be made between the desire to place essential requirements early in the test sequence and the desire to maintain the prototype in testable condition. These trade-offs become unmanageable when several factors of several categories each must be considered.

A model is developed for mapping the requirements from a randomly ordered collection of requirements to an ordered set of requirements. This model maps the requirement against which the prototype should first be evaluated to the first place in the ordered set.

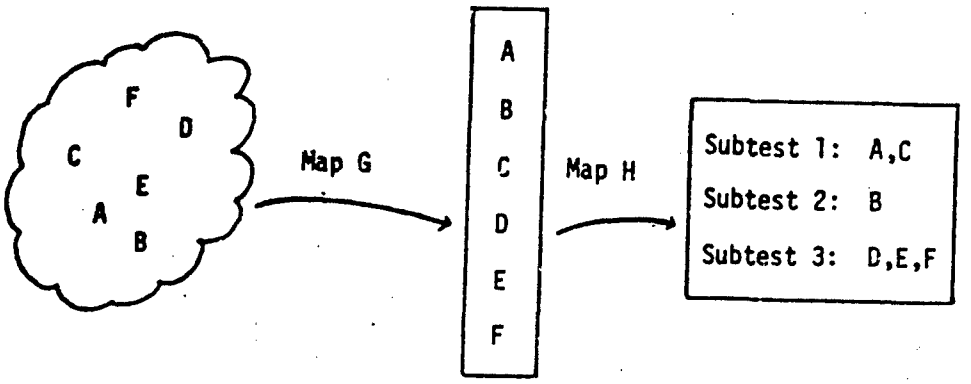
When the requirements are ordered, they are in the proper sequence for unconstrained testing of requirements. However, in developing the actual test schedule, there may be constraints which require that several requirements be grouped into one subtest, or which prevent the bests being sequenced as desired, or which affect the test schedule in other ways. Consequently, a second model is then needed to map the requirements from their positions in the ordered set to their final position in the test schedule. The concept upon which this research was based is shown in Figure 1.

Three vital tasks must be accomplished to execute this schematic concept. First, for each prototype to be tested, the appropriate factors, categories and weighting values must be determined. Secondly, a model must be developed which will map the set of random requirements according to the parameters determined. The third task to be accomplished is to develop a scheduling algorithm to transform the set of ordered requirements into a constrained sequence of ordered subtests.

The first two tasks were accomplished by experimenting with actual suitability tests that were being conducted in the US Army R&D community. A brief summary of this portion of the research will be given. The third task of developing a scheduling algorithm has been partially accomplished and is included for future consideration. However, it should be noted that this algorithm has not been used on an actual suitability test as of this date.

DETERMINATION OF MODEL PARAMETERS. The primary concern is to develop a model to maximize the flow of information upon which to base the disposition decision. It is desired to place the most important requirements first in the testing sequence. The desired placement is a function of the requirement importance and the effect that the testing of the prototype against the requirement may have on the overall rate of information flow. If a model can be constructed which develops a measure that represents the "Requirement Importance vs Effect on Overall Information Flow" tradeoff for each requirement, then these measures can be used in specifying and sequencing subtests. This measure of criticality will be indicated as C_i.

The next task is to identify the factors which are relevant to measuring C_i. These factors will be applicable to any suitability test, but will be in varying degrees for each prototype. This phase of the research was con-



SCHEMATIC OF CONCEPT

Figure 1

ducted by extensive interviews and group discussions with experienced test officers at a test installation. As a result, five major factors were identified for consideration and for inclusion into the model. These factors are shown in Figure 2.

These five factors should not be considered as being an exhaustive list applicable to all suitability tests. There should be a flexible method for selecting the factors appropriate to each test analyzed. Consequently, it is proposed that the first step in analyzing a suitability test is to present the above five factors to the test supervisory personnel and ask them to consider the applicability of each factor. In addition to considering applicability of these five, it is also essential that they be given the opportunity to add other factors if needed.

Once the factors applicable to testing a prototype have been identified, the importance of each factor, j, relative to the other factors must be determined. These are simply factor weights and will be indicated by W_j. In their discussion of scoring models, Moore and Baker (1) stress the importance of assigning weights to factors in order to insure that the model reflects the priorities of the decision makers. Similarly, in the model being developed, it is essential that weights be determined to reflect the relative importance of the factors.

Numerous means are available to determine relative importance. These include simple rank ordering, correlated simple rankings, ratings, and successive ratings. The method of successive ratings was selected for this research because:

a. It is a simple and fast method;
b. It will allow the decision-maker to determine the weights considered appropriate by each judge as well as the overall group weights;
c. It forces each judge to develop ratings which he feels to be consistent and;
d. The method is intuitively appealing.

For this portion of the research, test supervisory personnel ranked the factors applicable to the test of a prototype by a simplified version of the Delphi Technique. This simplified ranking scheme converged rapidly to a ranking acceptable to each judge. Once the factors had been ranked, the method of successive ratings was used to assign weights to each factor.

This portion of the research used two major suitability tests as experimental vehicles. The rankings and weights assigned in one test differed from those assigned in the other. Whether these differences are due to differences in the prototypes for the two tests, or due to differences between the groups is an unanswered question. However, it appears that the prototype tested is the most important factor since the members of both groups agreed that they could rank and weight the factors for any particular test.

After the factors which were considered to be important for inclusion into the model have been selected and weighted, the next task is to categorize each of them. Categorization is merely the partitioning of each factor into levels.

PROBABILITY OF FAILURE	The estimated probability that the prototype will not meet the specific requirement. (Relates to the importance of the requirement.)
CONFIDENCE LEVEL	The estimated accuracy of the estimated probability of failure. (Relates to the importance of the requirement.)
IMPACT	The importance of the requirement to the potential suitability of the item. (Relates to the importance of the requirement.)
DESTRUCTIVENESS	The potential destructiveness of the subtest required for testing the prototype against the requirement. (Relates to the effect on information flow.)
CONSEQUENCE	The effect that the results of the subtest evaluating the requirement would have on the test schedule if the requirement is not met. (Relates to the effect on information flow.)

FACTORS, j

Figure 2

The categorization of the factors in this research was accomplished by two test project groups and was generally based on the guidance found in the US Army Test and Evaluation Command Regulation 70-34 on Risk Analysis (2). The categories shown in Figures 3, 4, and 5 were agreed upon by both test project groups. It must be emphasized that these categories are merely suggested and may in fact be prototype dependent.

In categorizing the estimated probability that a requirement would not be met, the test groups indicated a preference to use the point estimate of the probability. This same technique was also used in categorizing the confidence level with which the estimate of probability of failure is made.

Now with each factor partitioned into categories, it is necessary to weight the categories of each factor. This portion of the research again used a modified Delphi Technique to order the categories of each factor. Then each member of each test panel weighted the categories of each factor on a scale from one to ten with ten being applied to the most important category and one being applied to the least important category. The other categories were scaled between one and ten. The category score was assigned as the average score for each category.

The net category weight for each factor was then computed as the product of factor weight and category score. An example of this is shown in Figure 6.

It was found during this portion that the categories of Impact and Probability of Failure were constant in both tests. This may be a random occurrence or it may be true for all suitability testing. The categories for the other factors were not so clear and this indicates prototype dependence.

The result thus far has been the determination of the parameters which may be included in the model. The next task is to determine the parameters applicable to each requirement. The results of this portion of the research indicate that each member of the test group should categorize each requirement. Then a composite of these is given to the test project officer for a final determination as to the category of each requirement. This step in the procedure is seen as a simplified version of the Delphi Technique.

A summary of the parameter development is shown in Figure 7.

CONSTRUCTING AND TESTING THE MODEL. The problem now is to develop a model using these parameters to transform the random requirements into a set of ordered requirements. It is hypothesized that such a model would be of the form

$$C_i = G\{[N_1^k(i)], [N_2^k(i)], \ldots, [N_j^k(i)], \ldots, [N_n^k(i)]\}$$

where C_i is the measure of criticality (i.e., a number which reflects the "Requirement Importance vs Effect on Overall Information Flow" trade-off of requirement i) and $N_j^k(i)$ is the importance of category k of factor j relative to the other categories of j for requirement i.

DESTRUCTIVE	Testing against the requirement is potentially destructive to the test item.
DAMAGING	Testing against the requirement is potentially damaging to the test item or to components not under test.
SENSITIVE	The requirement relates to a component which is delicate and which could be easily damaged during the course of unrelated tests.
STABLE	The requirement does not require potentially destructive testing and does not relate to a delicate component.

CATEGORIES OF THE FACTOR DESTRUCTIVENESS

Figure 3

CRITICAL	Failure to meet the requirement is sufficient for declaring the item unsuitable.
IMPORTANT	Failure to meet the requirement is not sufficient for declaring the item to be unsuitable but the requirement will be given major consideration in making the final determination of suitability.
DESIRED	The requirement will be given some consideration in making the final determination of suitability.
MINOR	The requirement will be given little or no consideration in making the final determination of suitability.

CATEGORIES OF THE FACTOR IMPACT

Figure 4

If the requirement is not met, the consequence to the test plan may be:

STOP TESTING	The test will be stopped for an undetermined length of time or will be terminated.
SUSPEND TESTING	The test will result in a slippage of more than 5 days in the test schedule.
TEST DELAY	There will be a test schedule slippage of from 1 to 5 days.
DEGRADE TEST	Testing may continue in a degraded mode while the deficiency is being corrected. There will be no test schedule slippage nor significant effect in the determination of suitability of the item under test.
OVERTIME REQUIRED	Retesting or additional work will be required but there should be no test schedule slippage.
RESCHEDULING	Testing will continue but rescheduling of subsequent requirements will be required. However, neither rescheduling nor retesting should result in test schedule slippage.
REPEAT TEST	Testing will continue, but the failed requirement will require re-evaluation during other planned tests.
WAIVE	The requirement will probably be waived due to being overly stringent or beyond the current state of the art. Failing the requirement will have no effect on the test schedule.
NONESSENTIAL	The requirement will not affect the determination of suitability and failing the requirement will have no effect on the test schedule.

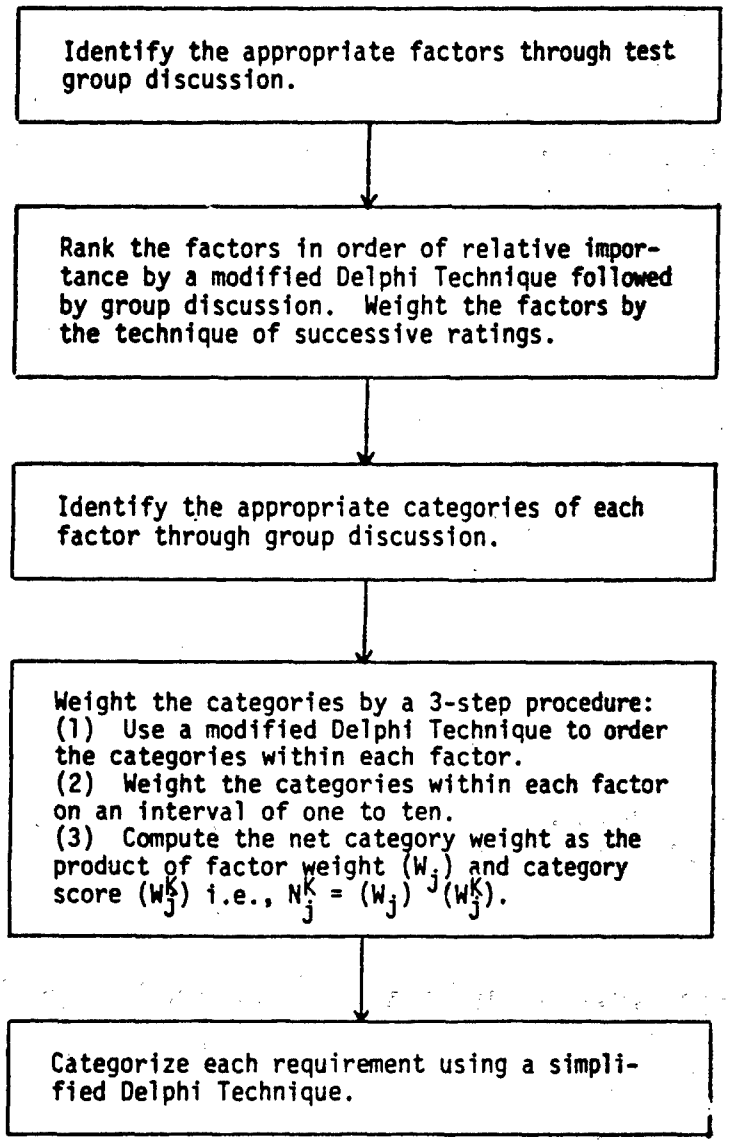
CATEGORIES OF THE FACTOR CONSEQUENCE

Figure 5

FACTOR(j)	FACTOR WEIGHT W_j (1,10)	CATEGORIES D_j^k	CATEGORY SCORE W_j^k (1,10)	NET CATEGORY WEIGHT N_j^k
Destructiveness	3.0	Destructive	1.0	3.0
		Damaging	5.8	17.4
		Sensitive	10.0	30.0
		Stable	6.1	18.3

EXAMPLE OF FACTOR DESTRUCTIVENESS

Figure 6



DETERMINATION OF MODEL PARAMETERS

Figure 7

In determining G, only linear and simple multiplicative functions are evaluated. Disjunctive and conjunctive functions described by Einhorn (3, 4) and logarithmic functions were considered but not included in this research. The primary reason for not investigating these forms is that the potential benefits from the more complicated model would be offset by its computational difficulties. It must be stressed that this research was oriented towards the suitability test planner who cannot be expected to have an operations research or other strong mathematical background.

Six models were evaluated during this research. Models 1 and 2 are formulated as:

$$C_i = \sum_{j=1}^{n} N_j^k (i) \qquad (1)$$

$$C_i = \prod_{j=i}^{n} N_j^k (i) \qquad (2)$$

Models 3 and 4 are respectively formulated as:

$$C_i = \sum_{j=1}^{m} N_j^k (i) \qquad (3)$$

$$C_i = \prod_{j=1}^{m} N_j^k (i) \qquad (4)$$

where the factor confidence level is not included in the set j = (1,2,...,m).

Models 5 and 6 are respectively formulated as:

$$C_i = \sum_{j=1}^{3} N_j^k (i) \qquad (5)$$

$$C_i = \prod_{j=1}^{3} N_j^k (i) \qquad (6)$$

where j denotes the three factors of Impact, Probability of Failure and Consequence as specified in TECOM Regulation 70-34.

This phase of the research involved designing an experiment in which significant indications of the relative desirability of additive and multiplicative models could be determined. Models 1 and 2 are included since it is hypothesized that one of them is the desired model. Since Model 6 was proposed by the US Army Test and Evaluation Command for identifying "high risk" requirements, it was included. Models 3 and 4 were considered since the factor Confidence Level was not deemed appropriate in one of the test projects considered in this research. Models 3 and 4 are essentially compromises between Models 1 and 2 and Models 5 and 6.

Each model was used to compute the measures of criticality for each requirement. For ease of reading, the term "score" is used as being synonomous with the term "measure of criticality."

Based on the scores computed by the six models, six sequences of requirements were generated. The requirement placed first in each sequence was the one receiving the highest score by the corresponding model. The other requirements are then sequenced in order of decreasing scores. As expected, the sequences were not identical.

Procedures were developed for identifying the most desirable sequence. These procedures were developed for determining a ranking of the sequences and consequently a ranking of the models. The purpose was to identify the better function (linear or simple multiplicative) and to identify whether the model should include the five factors identified earlier or only the three factors identified by TECOM. It is assumed that since the sequences are determined by the models, the sequence identified as being the most desirable must be the output of the best model. Three procedures were used in attempting to identify the best model.

The first procedure involved the test officer's attempting to rank the sequences generated by each model. This procedure was found unsatisfactory since it required him to consider too many variables. For example, one test had 59 requirements, 5 factors and 7 categories per factor. This is more than 2000 decision variables in comparing just two sequences.

The second procedure involved the test personnel discriminating between sequences indirectly. The technique used was to have the test officers compare requirements which had received appreciably different rankings in linear and multiplicative models. The procedure appeared feasible but no significant results were obtained. There was no detectable fault in the procedure used so it was concluded either the results indicate none of the models is a particularly good predictive model or that the judges were not consistent in their evaluations.

The third procedure involved simulating the actual results which would have been experienced if tests had been conducted according to each model. This simulating procedure addressed the normative side of model building in that the results of the rankings rather than the rankings themselves are considered. This approach was found to be successful in that a ranking of sequences (and consequently a ranking of models) is generated with a significant level of concordance among evaluations.

In this procedure, it was hypothesized that if the judges could identify the simulated tests which they considered to be better scheduled and if these tests could be ranked in order of desirability, then an ordering of the relative desirability of the models would result. For this procedure a seventh sequence based upon random placement of requirements was generated and was identified as Model 7.

Based upon the simulations for one of the prototype tests (Test A), the test sequences were ranked by test personnel judges. The results are shown in Figure 8.

The judges on this test agreed that each of the models produced sequences superior to Model 7. It was also concluded that the additive models (i.e., 1, 3, and 5) were respectively superior to the multiplicative models (i.e., 2, 4, and 6). It was further concluded that the five factor models (i.e., 1 and 2) were superior to the three factor models (i.e., 5 and 6). These conclusions were reinforced by the results of simulations on another prototype test (Test B).

JUDGE	MODEL 1	2	3	4	5	6	7	W*
A	3	2	4	1	5	6	7	
B	1	5	2	4	3	6	7	
C	2	6	1	4	3	5	7	
Correlated	1	5	2	3	4	6	7	.686

W* is the concordance coefficient described by Kendall (5). The value shown indicates agreement between judges significant at the .05 level.

RANKING OF MODELS BASED ON TEST A

Figure 8

Based on these conclusions, Models 1, 2, 5, and 6 were simulated for Test B and presented to a new panel of judges (i.e., new test supervisory personnel). These judges ranked the models as shown in Figure 9.

The judges concurred that the sequence generated by Model 1 produced a test schedule superior to that generated by the others. This sequence was so obviously superior that further opinions were not obtained.

Test A was ready to begin, so no rescheduling was allowed for that prototype based on model results. Test B had sufficient planning time to make use of the model results. The test officer for Test B was given the requirement sequence of 58 requirements for his test as generated by Model 1. He was asked to use this ranking in any way he saw fit in scheduling the test.

In scheduling the test the test officer first identified the constraints active for Test B. As it turned out, only technological constraints were required and these dictated that the test consist of four subtests of multiple requirements. Three of these subtests were required to be conducted sequentially and the fourth subtest consisted of requirements which required evaluation throughout the entire testing period. The requirements which were required to be placed in each subtest were identified and grouped within their appropriate subtests. The ranking of requirements generated by Model No. 1 were then used to order the requirements within each subtest to form the final test sequence. The ordering of requirements in each subtest was rank order consistent with the ordering of the requirements in Sequence No.1. Finally, the time and personnel requirements for the subtests were identified and a tentative test schedule which required four personnel and 2 weeks was established. The test officer found the ranking of requirements generated by Model No. 1 to be of appreciable assistance when establishing the order in which the requirements would be addressed within each subtest. He also considered the resulting test schedule to be "optimum", or as nearly "optimum" as he could determine.

Prior to the conduct of this research, a tentative test schedule for Test B had been developed. According to the previously developed schedule, a planning figure of 16 weeks was established for the time required to complete the suitability test. Of course, this planning figure is a pessimistic estimate. A most likely estimate of the time required had not been determined.

Through the application of the methods and model described herein, a test schedule was developed with a most likely estimate of the time required being established at 2 weeks. The test officer did not wish to establish a new planning figure, or pessimistic estimate, until he had re-evaluated all possible contingencies. However, he was confident that the new planning figure would be no more than 4 weeks. No claims are made that through the use of the procedures and model developed in this research a test schedule will be developed which will require less than one-fourth of the time which would otherwise be required. However, it appears that the procedures can result in either a substantial savings in test time or a more accurate estimate of the test time required.

JUDGE	1	2	5	6	W*
A	1	3	2	4	
B	1	3	2	4	
C	1	4	2	3	
Correlated	1	3	2	4	.91

RANKINGS OF MODELS BASED ON TEST B

Figure 9

Finally, the authors and the test officer discussed the resulting test schedule. Without benefit of the sequence generated by the model or of the categorizations of the requirements established earlier, the test officer was asked to justify his test schedule based only upon the verbal descriptions of the requirements and his knowledge of the overall test. He was able to justify convincingly the relative placement of each requirement and found the schedule to be, somewhat to his own surprise, "optimum" from his point of view. Of course, this determination of optimality is based upon subjective judgment and the validity of the conclusions is only as valid as the judgment of the officer making them. However, this officer is an experienced test officer and could reasonably be considered to be an expert in his field. Until a constrained optimization model is developed which will replace expert judgment in qualitative analysis, the opinions of the experts in the field will have to be used in the determination of optimality.

Based on the apparent success of this method, it was concluded that Model 1 was applicable in developing an ordered set of requirements for use in scheduling suitability tests for a prototype item. Model 1 is a simple linear model and includes all factors considered important by the decision makers involved.

To recap the research thus far, we have determined the parameters deemed important for inclusion in the model and we have selected an acceptable normative model to map a set of randomly placed requirements into a set of ordered requirements. This set of ordered requirements represents an ordered unconstrained test sequence of requirements. If we could evaluate each requirement sequentially we would maximize the rate of information flow by the measure of criticality. However, this unconstrained testing is not practical since it does not consider time and personnel constraints on the testing sequence. This brings us to the third task outlined in the research concept.

SCHEDULING PROCEDURES. The problem under consideration now is to determine a procedure for developing an actual test schedule which will result in the optimum rate of information being generated during the test. The possibilities of evaluating each requirement simultaneously and evaluating the requirements individually and sequentially are considered infeasible and are not addressed. This portion of the research assumes that there are technological, precedence or proximity constraints which make these type tests impractical. The following discussion of scheduling procedures raquires the adoption of the assumptions shown in Figure 10.

Three approaches were investigated in developing test schedules. The first two approaches considered situations in which the requirements could be easily partitioned into logical and practical subtests. The last approach investigated a procedure to quantitatively assign the requirements to subtests and then order the subtests into a constrained test sequence.

The first approach considers unconstrained test scheduling where one or more requirements have already been assigned to each subtest. The value of each subtest is determined from the model previously developed and the assumptions shown in Figure 10. The time required for each subtest must be estimated by the test planner. These time estimates are presently being accomplished so there is no new requirement for the evaluating organization.

1. The measure of criticality, C_i, is the same as the value of the information which will be gained from evaluating the prototype against the requirements.

2. The value of the requirement is the same as the value of the information to be gained from evaluating the prototype against the requirement.

3. The value of a subtest is the sum of the values of the included requirements.

4. There is a linear relationship between the value of the information obtained from a subtest and the length of time which will be spent on the subtest.

5. The time required to complete a subtest is the same as the time required to test the prototype against the most time consuming requirement included in the subtest.

6. The personnel required to conduct a subtest is the sum of the personnel required to individually evaluate each of the requirements.

SCHEDULING ASSUMPTIONS

Figure 10

Procedure A is a two-step algorithm recommended for determining a test schedule under these conditions. This procedure is shown in Figure 11. This procedure is a direct application of Theorem 3-10 stated and proved by Conway, Maxwell and Miller in their book, The Theory of Scheduling (6).

In situations where the subtests have been established and where precedence constraints are active, a second procedure is recommended. Procedure B is recommended under these circumstances. This procedure is an application of the constrained least cost testing sequence described by Mankekar and Mitten (7). See Figure 12.

Basically this procedure involves isolating those subtests for which the precedence constraints are active and then systematically satisfying the constraints. After this is done, Procedure A is applied in a manner which does not violate any of the constraints previously satisfied.

Set 1 consists of those subtests for which precedence constraints are active and Set 2 consists of those subtests for which there are no precedence constraints. The matrix R is an m x m matrix.

$R = \{r_{ij}\}$ where m is the number of subtests in Set 1 and

$r_{ij} = 1$ if subtest i

must precede subtest j; otherwise

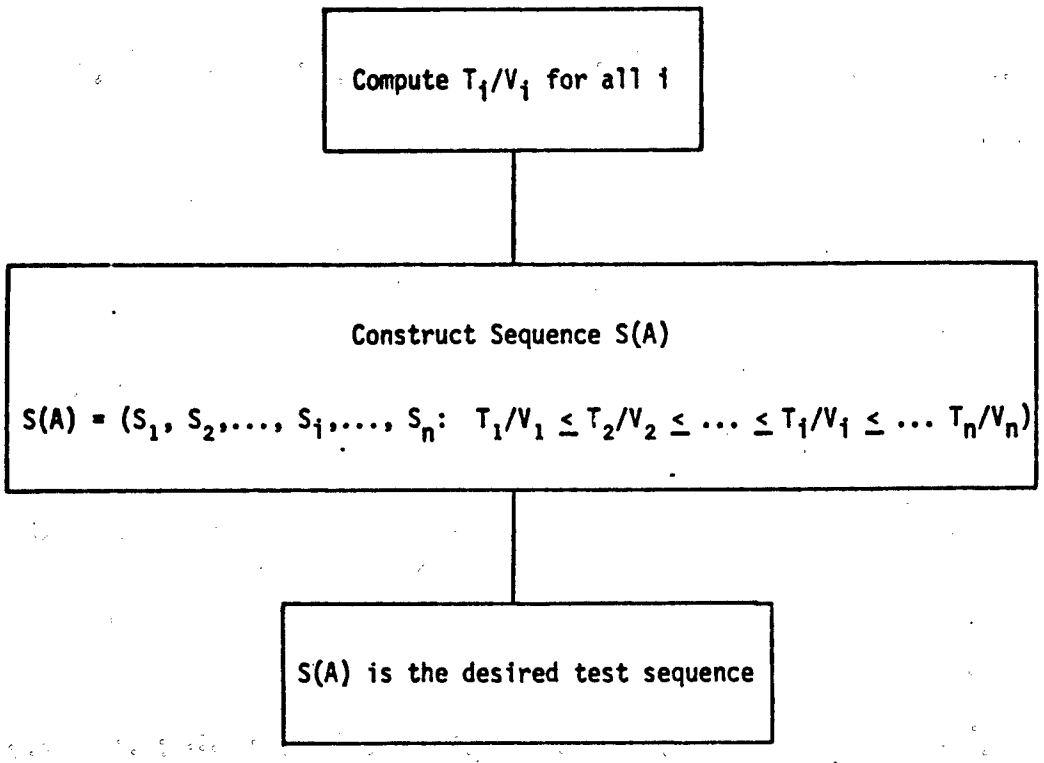
$r_{ij} = 0$. ($r_{ii} \equiv 0$).

The matrix R reflects all precedence constraints on Set 1. The matrix R' is identical to R and is merely used as a working matrix. With these definitions in hand, the procedure will lead to an optimal least time test sequence under the assumptions noted. The development, proof of finiteness and proof of optimality were developed by Mankekar and Mitten (7). A computational algorithm for using this precedure is shown in Appendix 1.

In these two approaches at scheduling, it was assumed that the subtests were predetermined. The next approach attempts to quantitatively assign the requirements to appropriate subtests.

It is assumed that the only constraint is personnel where only N personnel are available for commitment to a subtest. It is further assumed that V_i, P_i and T_i are known where these variables are the value, personnel required, and the time required for requirement i, respectively. V_i can be determined from the model previously developed and P_i and T_i can be estimated as they are presently being done. The problem now becomes one of designing the best set of subtests which can be sequenced by Procedure A.

This problem is analogous to the n/m job shop problem where the n jobs (test requirements) are assigned to m machines (subtests). Procedure C, shown in Figure 13, is presented here only for consideration as a solution to the problem. It draws heavily from the work done by Conway, Maxwell and Miller. It has received very little testing and has not been applied to an actual suitability test. However, it is simple, intuitively appealing, and it can be carried out by hand or coded for computer use.



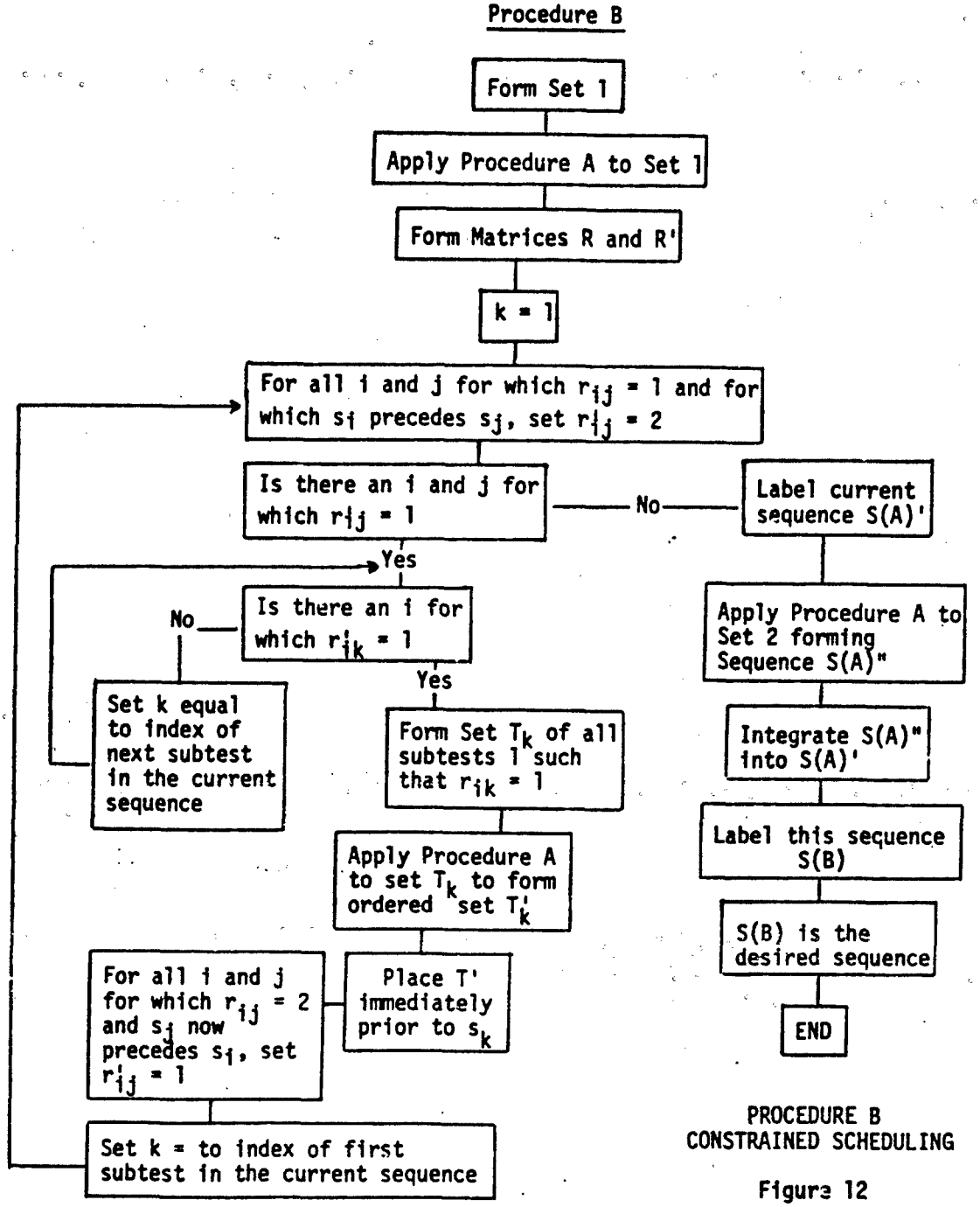
T_i = time required to conduct subtest i for each i ε {1,n}

V_i = value of subtest i for each i ε {1,n}

n = number of subtests

PROCEDURE A
UNCONSTRAINED SCHEDULING

Figure 11



PROCEDURE B
CONSTRAINED SCHEDULING

Figure 12

Procedure C

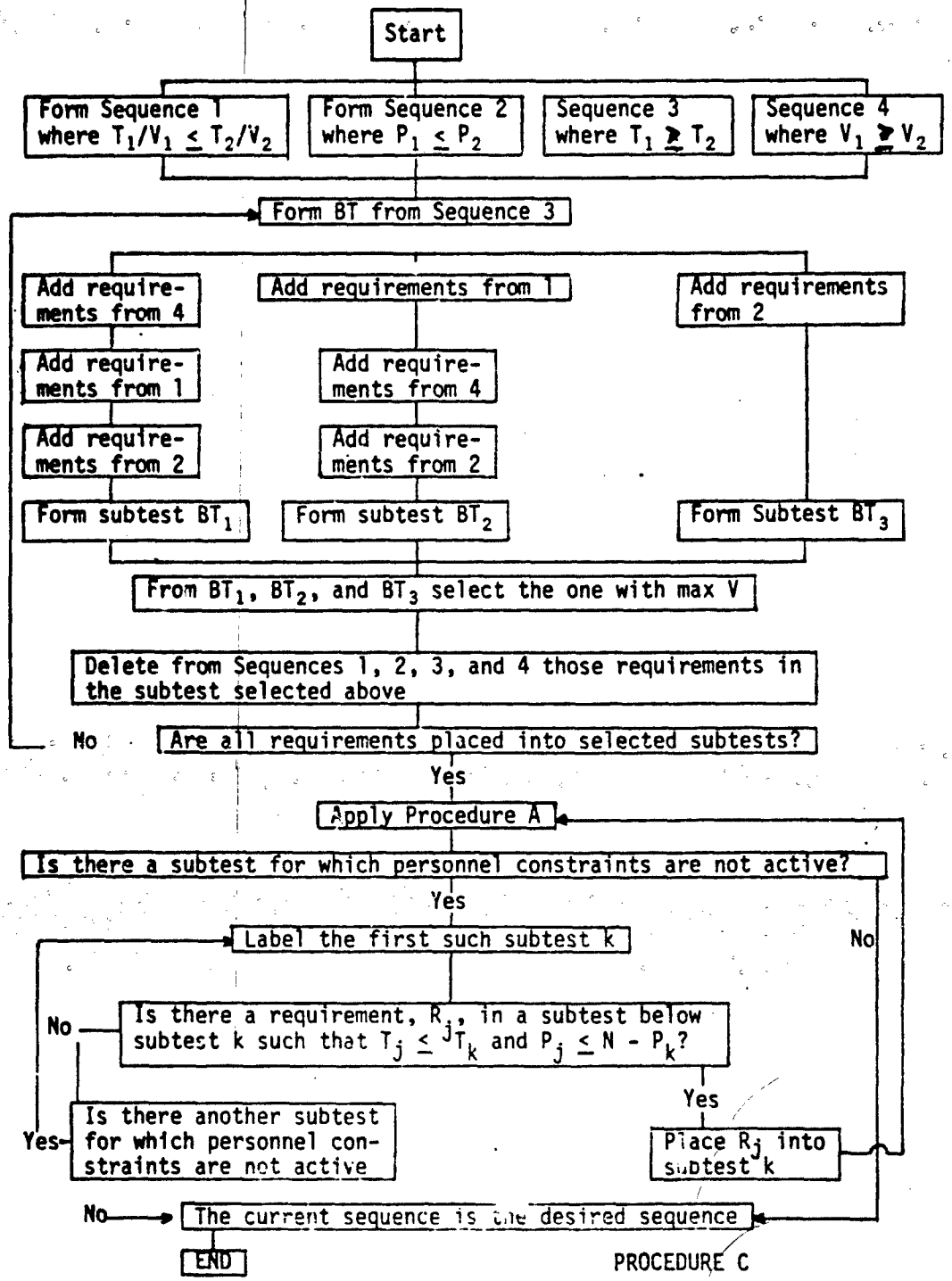
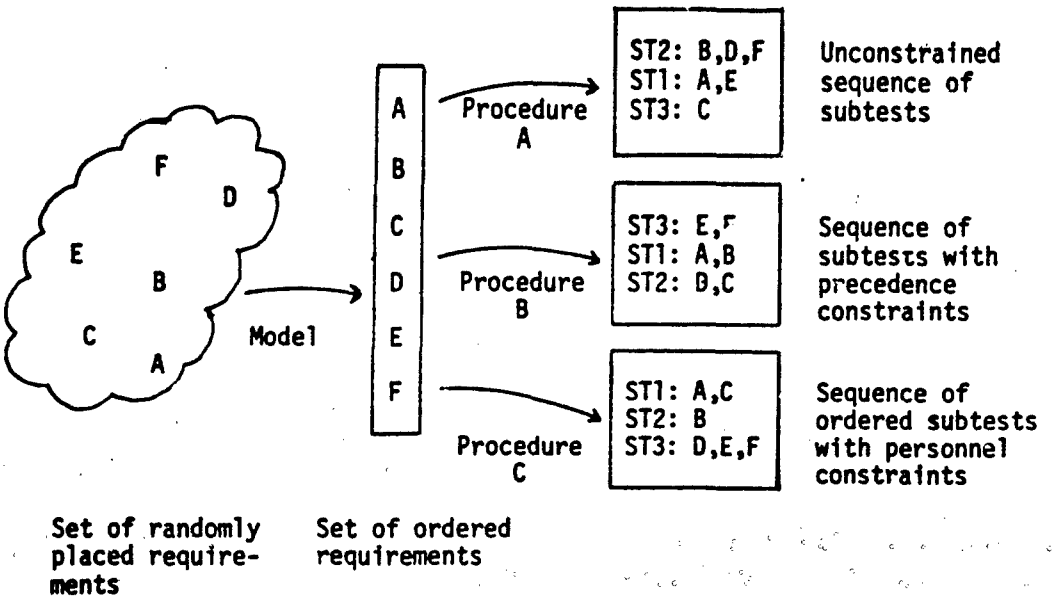


Figure 13



CONCEPT SUMMARY

Figure 14

Basically, Procedure C involves forming four sequences of requirements and alternately drawing from the sequences to form trial subtests. The subtest which contains the most value is then selected as the best of the trial subtests and the procedure is repeated until each requirement is assigned to a subtest. A computational algorithm for using Procedure C in included in Appendix II.

SUMMARY. The three procedures used during this research can be put into perspective by relating to the original concept of the research. See Figure 14. The model developed mapped a set of randomly placed requirements into a set of ordered requirements. Procedure A maps this set into an unconstrained sequence of subtests when the requirements have already been assigned to subtests. Procedure B maps the set of ordered requirements into a constrained sequence of subtests when the requirements have already been established and there are active precedence constraints. Procedure C maps the set of ordered requirements into subtests and then develops the order or sequence for these subtests when there are active personnel constraints.

None of the scheduling procedures used during this research are completely satisfying. Each procedure is only a partial answer. What is required is a procedure which will map a set of ordered test requirements into an ordered sequence of subtests when there are active precedence, proximity, personnel, time and economic constraints. This is an area for future research.

However, it is felt that the portions of the research dealing with parameter identification and model development are a worthwhile basis for further research into the problem of test scheduling when there are multiple active constraints.

REFERENCES

1. John R. Moore and Norman R. Baker. "An Analytical Approach to Scoring Model Design - Application to Research and Development Project Selection". IEEE Transactions on Engineering Management Volume EM-16, No. 3, August 1969.

2. Headquarters, US Army Test and Evaluation Command, Risk Analysis of Suitability Testing (USATECOMR 70-34), Aberdeen Proving Ground, Maryland, 1971.

3. Hillel J. Einhorn, "The Use of Nonlinear, Noncompensatory Models in Decision Making", Psychological Bulletin, Volume 73, No. 3, 1970.

4. ———. "Use of Nonlinear, Noncompensatory Models as a Function of Task and Amount of Information". Organizational Behavior and Human Performance, Volume 6, 1971.

5. M. C. Kendall, Rank Correlation Methods. Charles Griffin and Company Limited. 4th Edition, London, 1970.

6. Richard W. Conway, William L. Maxwell and Louis W. Miller, Theory of Scheduling. Addison-Wesley Publishing Company, Reading, Massachusetts, 1967.

7. P.S. Mankekar and L.G. Mitten, "The Constrained Least Cost Test Sequencing Problem", The Journal of Industrial Engineering, Volume XVI, No. 2 March-April 1965.

APPENDIX A

COMPUTATIONAL ALGORITHM FOR PROCEDURE B

STEP 1: Form two sets of subtests. Let Set 1 consist of those subtests for which precedence constraints are active. Let Set 2 consist of those subtests for which no precedence constraints are active. Steps 2 through 15 refer to Set 1 only.

STEP 2: Order the subtests in Set 1 by Procedure A. Index the subtests according to their relative position in S(A) with the first subtest in the sequence being denoted s_1.

STEP 3: Form an m x m matrix $R = \{r_{ij}\}$ where

$r_{ij} = 1$ if subtest i must precede subtest j,

otherwise $r_{ij} = 0$; and

$r_{ii} \equiv 0$; and

$r_{ij} = 1$ implies $r_{ji} = 0$; and

if $r_{ij} = 1$ and $r_{jk} = 1$, then $r_{ik} = 1$.

STEP 4: Form a matrix $R' = \{r'_{ij}\}$ identical to matrix R.

STEP 5: Set the index k = 1.

STEP 6: Consider each pair of subtests i and j. If $r'_{ij} = 1$ and subtest i precedes subtest j in the current sequence, set $r'_{ij} = 2$.

STEP 7: If $r'_{ij} \neq$ for all i and j go to Step 15.

STEP 8: Scan R' to determine if $r'_{ik} = 1$ for any i. If there exists an i such that $r'_{ik} = 1$, go to Step 9. If $r'_{ik} \neq 1$ for all i, set k equal to the index of the next subtest in the current sequence and repeat this step.

STEP 9: Form set T_k of all subtests i for which $r_{ik} = 1$.

STEP 10: Apply Procedure A to set T_k to form the ordered set T_k'.

STEP 11: Place the ordered set T_k' immediately in front of subtest S_k.

STEP 12: Consider each pair of subtests i and j for which $r_{ij}' = 2$. If subtest j now precedes subtest i, set $r_{ij}' = 1$.

STEP 13: Set k equal to the index of the first subtest in the current sequence of subtests in Set 1.

STEP 14: Go to Step 6.

STEP 15: Label the current sequence S(A)'.

STEP 16: Apply Procedure A to Set 2. Label the resulting sequence S(A)". S(A)" = $\{S_j''\}$.

STEP 17: Form sequence S(B) from sequences S(A)' and S(A)" by iteratively integrating the S_j'' into S(A)' such that $T_i'/V_i' \leq T_j''/V_j'' \leq T_{i+1}'/V_{i+1}'$. S(B) is the desired sequence for testing.

APPENDIX B

COMPUTATIONAL ALGORITHM FOR PROCEDURE C

STEP 1: Apply Procedure A to the requirements to form Sequence 1. For those requirements in which $T_i/V_i = T_j/V_j$ place the more time consuming requirement first in the sequence.

STEP 2: Construct Sequence 2 of requirements with the requirements being ranked in order of increasing value of personnel required. Resolve ties by placing the more time consuming requirement first.

STEP 3: Construct Sequence 3 of requirements with the requirements being ranked in order of decreasing value of T_i. Resolve ties by placing the more valuable requirement first.

STEP 4: Construct Sequence 4 of requirements with the requirements being ranked in order of decreasing value of V_i. Resolve ties by placing the more time consuming requirement first.

STEP 5: Construct a base-subtest by including in the subtest the first consecutive requirements from Sequence 3 until the inclusion of the next requirement in the sequence would violate the personnel constraint. Call this base subtest BT.

STEP 6: Construct three tentative subtests as follows:

a. Add the first consecutive requirements from Sequence 4 to BT until the inclusion of the next requirement in the sequence would violate the personnel constraint. Next add the first consecutive requirements from Sequence 1 to the current subtest until the inclusion of the next requirement would violate the personnel constraint. Finally, to this subtest add the first consecutive requirements from Sequence 2 until the inclusion of the next requirement would violate personnel constraint. Label this subtest BT_1.

b. Construct subtest BT_2 in a manner similar to constructing BT_1. However, in forming BT_1 requirements were added to BT from Sequences 4, 1, and 2 in that order. In forming BT_2 add requirements to BT from Sequences 1, 4, and 2 in that order.

c. Construct subtest BT_3 by adding the first consecutive requirements from Sequence 2 until the inclusion of the next requirement in the sequence would violate the personnel constraint.

STEP 7: From subtests BT_1, BT_2, and BT_3 select the subtest with the greatest value. Note that the time required for each subtest is the same as the time required for each of the other subtests since each subtest is based upon BT. Consequently, this step involves selecting the subtest with the minimum value of T/V.

STEP 8: Delete from Sequences 1, 2, 3, and 4 those requirements included in the subtest selected in Step 7.

STEP 9: If each of the requirements has been included in selected subtests, go to Step 10. Otherwise return to Step 5.

STEP 10: Apply procedure to the subtests generated to form Test Sequence S(A).

STEP 11: Scan S(A) until the first subtest is found in which the personnel constraints are not active. Call this subtest k with test time required being T_k and personnel required being P_k. If no such subtests are located, then go to Step 14.

STEP 12: Continue to scan S(A) until the first requirement R_j is found such that $T_j \leq T_k$ and $P_j \leq N - P_k$. Place R_j into subtest k. If no such requirement is located, then return to Step 11. Scan immediately below subtest k.

STEP 13: Return to Step 10.

STEP 14: Stop. The current sequence is the desired sequence which should be the basis of the testing schedule.

STOPPING RULES FOR SEQUENCING WITH PARTICULAR REFERENCE TO MISSILE RANGE SCHEDULING*

Paul H. Randolph
Department of Mathematical Sciences
New Mexico State University
Las Cruces, New Mexico

ABSTRACT. Monte Carlo methods have been proposed for finding solutions to scheduling problems. One deficiency of these methods has been the absence of appropriate rules for stopping the sampling processes. This paper presents stopping rules that not only have been found effective for a variety of sequencing problems, but also provide a measure of the quality of the sequence chosen. Reference to missile range scheduling is made.

MISSILE RANGE SCHEDULING. At a missile range a set of missions are requested each day. One way to schedule these missions is to take a permutation of the missions and schedule the missions as early in the day as possible in the order of the permutation, but with no conflict in the resources required for each mission. Because of the nature of the standard work day, it may not be possible to schedule some missions when using the given permutation. Different permutations will give schedules with different sets of missions that are scheduled and not scheduled.

With each mission there is associated a payoff, so that a schedule payoff is the sum of the payoffs of the scheduled missions. If a permutation is selected by a random procedure, then the corresponding schedule payoff can be considered a random variable. By taking a sequence of random permutations, a sequence of random variables of schedule payoffs is obtained. If "enough" of these random payoffs are obtained, the random schedule generation can be terminated and the schedule corresponding to the best of the observed schedule payoffs can be used for the set of missions requested for the day. The problem, of course, is to determine how much is "enough"; or, in other words, when to stop the random generation of schedule or sequences.

STOPPING RULES FOR INTEGER PAYOFFS. Let $X_1, X_2, \ldots$ denote the random variable of the payoffs associated with generating successive sequences by a Monte Carlo sampling process. For the present, assume that each sequence payoff is an integer and that the objective of the sequencing problem is to find a sequence for which the payoff is maximized. Furthermore, without loss of generality, assume that all payoffs are positive and bounded above by the known integer ℓ. Also, let y_n denote the maximum of the payoffs, $x_1, \ldots, x_n$, obtained from the first n sequences; that is $y_n = \max(x_1, \ldots, x_n)$. Thus, it is assumed that sampling will be with recall.

---

*Research for this paper was partially supported under AROD Contract No. DAHC04-C-0011 at the Instrumentation Directorate, White Sands Missile Range, New Mexico.

The probability function for each sequence payoff is the multinomial characterized by $P(X=k) = p(k)$, $k = 1,\ldots,\ell$. When the values of $p(k)$ are known, then Chow and Robbins [1], [5] have shown that the optimal stopping rule is obtained by calculating the expected increase in gross payoff associated with generating another sequence.

$$T(y_n) = \sum_{k=y_n}^{\ell} (k-y_n)p(k),$$

and comparing this with the relative cost, c, of generating a single sequence on the computer; that is, if $T(y_n) > c$, continue for another observation; if $T(y_n) \leq c$, stop. The function $T(y_n)$ is sometimes called the stopping rule function.

Note that this stopping rule states that at each stage of the process, the experimenter computes the expected gain from taking exactly one further observation and then terminating the process. If the expected net gain from taking this observation is not positive, then the process is terminated. Otherwise, the next observation is taken, and a similar computation is again performed. This procedure is called the myopic procedure because at each stage the experimenter does not look beyond the possible outcomes of his very next observation when making his decision, and for sampling with a known distribution this myopic procedure is optimal.

Unfortunately, for sequencing problems the values of $P(X=k) = p(k)$, $k = 1,\ldots,\ell$, are almost never available. However, even though these probabilities are not known, it is possible to obtain estimates of these probabilities through a Bayesian analysis, and then substitute these Bayesian estimates for the $p(k)$ into the above myopic stopping rule function to obtain what might be called a Bayesian stopping rule for multinomial observations [4].

To obtain these estimates define the ℓ-dimensional vector $\Theta=(\Theta_1,\ldots,\Theta_\ell)$ of probabilities such that for the n-th observation, X_n, the conditional probability function is given by $P(X_n = k|\Theta) = \Theta_k$, $k = 1,\ldots, \ell$, where Θ is an element of the simplex

$$S = \{\Theta \in E^{\ell}: \sum_{k=1}^{\ell} \Theta_k = 1, \Theta_k \geq 0, k = 1,\ldots,\ell\}.$$

Since the conjugate prior density [3] for the multinomial is the Dirichlet, the initial prior density of Θ can be written as

$$f_0(\Theta) = \Gamma(m) \prod_{k=1}^{\ell} \left\{ \Theta_k^{m_k-1} \Big/ \Gamma(m_k) \right\}$$

This is the Bayesian prior density of Θ for observation X_{n+1}.

Furthermore, since the joint density function for X_{n+1} and θ_k is $\theta_k f_n(\theta)$, then the marginal distribution

$$p_n(k) = (m_k+n_k)/(m+n), \quad k = 1,\ldots,\ell,$$

is the probability that X_{n+1} will take on the value k.

This value of $P(X_{n+1}=k) = p_n(k)$ can be substituted for p(k) in the stopping rule function, $T(y_n)$, to obtain what might be called the "Bayesian stopping rule function", which will be denoted $T_B(y_n)$, and is given by

$$T_B(y_n) = \sum_{k=y_n}^{\ell} (k-y_n)p_n(k) = (m+n)^{-1} \sum_{k=y_n}^{\ell} (k-y_n)m_k.$$

Comparing the value of this function with the value of c will determine a stopping point; that is, if $T_B(y_n) \leq c$, the sampling of sequence payoffs should be stopped. Since y_n is a monotonically non-decreasing function of n, then $T_B(y_n)$ is a decreasing function of n, which approaches zero as n increases. Thus, sampling always will eventually stop.

This Bayesian stopping rule function depends on the specification of a set of parameters associated with the Dirichlet prior density function. If these parameters, $m_1,\ldots,m_\ell$, are examined, it will be noted that they can be written in terms of the initial probabilities as $m_k = mp_0(k)$, $k = 1,\ldots,\ell$. Since the $p_0(k)$ are essentially normalized values of the m_k, it may be preferable to specify these initial probabilities, $p_0(k)$, $k = 1,\ldots,\ell$, (of which only $\ell-1$ are independent) and the parameter m, rather than to estimate the m_k directly. This can be done by an arbitrary selection of probability values, by specifying a discrete probability function, or even by integrating a continuous function over a unit interval containing k.

The parameter m has some interesting characteristics. A lower bound for m is zero, and this can be a greatest lower bound only when $p_0(k) = 1/\ell$, $k = 1,\ldots,\ell$. As $m \to 0$, then $T_B(y_n) \to 0$, and the Monte Carlo process stops with the first observation, implying no confidence in the initial probabilities. On the other hand, as $m \to \infty$, then

$$T_{n+1}(y_n) = m(m+n)^{-1} \sum_{k=y_n}^{\ell} (k-y_n)p_0(k) \to \sum_{k=y_n}^{\ell} (k-y_n)p_0(k) = T(y_n)$$

which is the expected improvement for a <u>known</u> multinomial distribution, indicating a complete confidence in the initial probabilities. Thus, the parameter m can be interpreted as a coefficient of confidence in the initial probabilities. In fact, it can be considered as being analogous to the sample size that would be needed to obtain through a random sample the same quality of estimate of $p_0(k)$ as those given by the specified prior probabilities.

STOPPING RULES FOR CONTINUOUS PAYOFFS. In missile range scheduling the schedule payoff can be taken to be the sum of the payoffs of scheduled missions. Each mission payoff is assumed to be a quadratic function of the mission priority, weighted by a "project readiness" factor. Project readiness is the probability that the contractor will not cancel the mission after it has been scheduled. If the priority for mission i is given by r_i and the probability of noncancellation is q_i, then the schedule payoff is

$$X = \Sigma q_i r_i^2$$

where the summation is over all missions that can be scheduled when using a given permutation. Since q_i is a number between 0 and 1, it is evident that the payoffs will not be integers. Furthermore, the number of different possible values of the payoffs is large and also the values of these payoffs are unknown. Thus, for missile range problems the payoff can essentially be considered a continuous random variable.

To determine the prior distribution of θ for discrete payoffs, it was proposed that values of m_k be obtained through the specification of the initial probabilities $p_0(k)$. One way of estimating these initial probabilities is by integrating a continuous function over a unit interval that contains the point k. This suggests that a limiting procedure could result in a stopping rule for continuous payoffs.

Suppose that the sequence payoffs can assume arbitrary values in the interval $[0,\ell]$, and let $A_1,\ldots,A_\nu$ be any partition of this interval, where A_k is defined as $A_k = (x'_{k-1},x'_k)$, $k = 2,3,\ldots,\nu$, $A_1 = [0,x'_1]$. Suppose H(x) is a distribution function of $[0,\ell]$ such that

$$p_0(A_k) = \int_{A_k} dH(x) = H(x'_k) - H(x'_{k-1})$$

reflects the experimenter's prior intuition for the initial probabilities for each partition A_k, $k = 1,\ldots,\ell$, regardless of the method of partitioning. If x''_k is any point in A_k, then the integral defined by

$$m(m+n)^{-1}\int_{y_n}^{\ell}(x-y_n)dH(x) = \lim_{\nu\to\infty} m(m+n)^{-1}\sum_{k=1}^{\nu}(x''_k-y_n)(H(x'_k) - H(x'_{k-1}))I(x''_k \geq y),$$

is the expected gross improvement in payoff for an additional observation, and is denoted by $T_B(y_n)$. I(x) is the usual indicator or characteristic function.

As an example, assume the normal distribution reflects the experimentor's beliefs for a particular set of initial probabilities. If Φ denotes the standarized normal distribution function (zero mean and unit variance) and ϕ its corresponding density function, then the stopping rule function becomes

$$T_{n+1}(y_n) = m(m+n)^{-1}(\sigma(\phi((y_n-\mu)/\sigma)-\phi((\ell-\mu)/\sigma)) + (\mu-y_n)(\phi((\ell-\mu)/\sigma)-\Phi((y_n-\mu)/\sigma))),$$

where μ and σ are the mean and variance, respectively, for the prior distribution. It is suggested that a set of initial observations be used to estimate the parameters μ and σ.

EXAMPLE. In Figure 1 is given the results when 40 missions were requested, and where the total number of different resources for these missions was 116. If all 40 missions had been scheduled, the schedule payoff would have been 189.5.

By scheduling via a Monte Carlo procedure, a total of 19 schedules were generated. The corresponding schedule payoffs, x_n, are indicated in the second column, with the maximum payoffs, y_n, given in the third column. In the next column are the expected payoffs, $T_B(y_n)$. In this problem the value of c was 0.001. Three missions, all of low payoff value, were not scheduled.

It should be noted that the expected payoff from continued sampling is 0.00090. That is, if the computer were permitted to continue generating schedules, the amount of expected improvement of schedule payoff would be only this much over the maximum of 186.88 that was obtained by stopping with the nineteenth observation. This, of course, is a measure of the quality of the schedule finally chosen.

CONCLUSIONS AND LIMITATIONS. Myopic stopping rules have been applied to missile range scheduling at White Sands Missile Range with very satisfactory results. It is possible to generate a schedule in one-twentieth of a second and thus in a few minutes hundreds of schedules can be observed, and the Bayesian stopping rules are very effective in determining the stopping point in the sampling procedure.

However, there exists one problem. It has been shown that the myopic procedure is optimal for random variables with known distributions. When the probability values are not known, then the myopic rule is not appropriate. The fundamental distinction is that when the distribution is completely specified, the observations are independent; that is, knowledge of the values of some of the observations provides the experimenter with no additional information about the values of the other observations. On the other hand, if the distribution involves the value of one or more parameters that have prior distributions, the observations are dependent under their joint marginal distribution. Hence, knowledge of the values of some of the observations will, by providing information about the value of the parameters, also provide information about the values of the other observations. This difference between independent and dependent observations distinguishes these two types of problems. The observations in a random sample from a distribution involving unknown values of parameters will no longer be independent.

In general it is felt that myopic rules applied to the random variable with unknown probabilities, using a Bayesian analysis, will provide stopping rules that are "near-optimal". Preliminary analysis for the multinomial indicates that such a rule may be conservative, that is, requiring more observations than necessary before stopping, but this is not certain. So, until more accurate results are available, the myopic rules will be used as good approximations that will yield "near optimal" sequences.

LARGEST POSSIBLE TOTAL EXPECTED SCHEDULE PAYOFF IS 189.500

N	SCHEDULE PAYOFF	LARGEST PAYOFF SO FAR	EXPECTED IMPROVEMENT
1	184.249973	184.249973	1.001000
2	166.039972	184.249973	1.001000
3	170.569973	184.249973	1.001000
4	170.699972	184.249973	1.001000
5	178.319973	184.249973	1.001000
6	181.079973	184.249973	1.001000
7	160.679972	184.249973	.013012
8	186.879972	186.879972	.003502
9	160.919971	186.879972	.003118
10	176.789972	186.879972	.002458
11	178.149973	186.879972	.002012
12	177.989973	186.879972	.001645
13	179.209972	186.879972	.001358
14	178.949972	186.879972	.001151
15	166.519972	186.879972	.001168
16	179.889973	186.879972	.001069
17	182.769972	186.879972	.001101
18	181.309973	186.879972	.001056
19	173.019972	186.879972	.000900

FIGURE 1

Computer Printout of
Monte Carlo Scheduling

REFERENCES

1. Chow, Y. S. and Robbins, H., "On Optimal Stopping Rules," Ziet. Wahrscheinlichkeitstheorie, 2:33-49 (1963).

2. DeGroot, M., "Some Problems of Optimal Stopping," Jour. of Royal Stat. Soc., 30:108-122 (1968).

3. Raiffa, H. and Schlaifer, R., Applied Statistical Decision Theory, Harvard Business School, 1961.

4. Randolph, P. H., "Optimal Stopping Rules for Multinomial Observations", Metrika, 14:48-61 (1968).

5. Robbins, H., "Optimal Stopping," Am. Math Monthly, 77:333-343 (1970).

APPROXIMATE CONFIDENCE LIMITS FOR P(Y < X)

J. R. Moore and M. S. Taylor

U. S. Army Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

ABSTRACT

A procedure is given for constructing approximate confidence limits for P(Y < X), where X and Y are independent random variables; the distribution of Y being known and normal and the distribution of X being unknown and positively skewed. A problem of determining the probability that the sidewall of a combustible cartridge case will not be burned through prior to firing an artillery round in an automatic firing cycle, given that it is ignited by smoldering residue after chambering, is used to illustrate the technique. A listing of a computer subroutine for the procedure is also given.

INTRODUCTION

An artillery round with a combustible cartridge case is fired from a weapon using a control system which loads the round, aims the weapon and fires the weapon, all automatically. At least two rounds of ammunition are fired in this fashion and some smoldering residue from the preceding round may remain in the chamber of the weapon when a round is loaded. Let R be the conditional probability that the sidewall of the cartridge case of the chambered round is not burned through prior to firing, given that it is instantaneously ignited by smoldering residue. The problem is to find both a point estimate and a 95% lower confidence limit for R, using information concerning the gun cycle time and data on cartridge case burn-through time obtained from laboratory tests.

There was sufficient gun cycle time data available to justify the assumption that the elapsed time between chambering and firing a round is normally distributed with (true) mean $\mu_2 = 2.9$ seconds and standard deviation $\sigma_2 = 0.13$ seconds.

An experiment was conducted to estimate the statistical distribution of cartridge case sidewall burn-through times. One hundred and fifty samples of sidewall material were taken from several cartridge cases and tested. Each sample of sidewall material was ignited and the elapsed time between ignition and burn-through was measured by three observers using stop watches. No data were obtained for two samples and there were some missing data for some of the item under consideration, it was necessary to sample cartridge cases from only one lot and assume that this sample was randomly selected from the conceptual population consisting of all such cartridge cases which will be manufactured. It was suggested that the validity of this assumption should be verified by further testing when a sample which is more representative of the production item can be selected.

PRELIMINARY STATISTICAL ANALYSIS

The burn-through data were analyzed assuming a one way classification, components of variance (random effects), analysis of variance model with unequal number of observations per cell. The component of variance attributable to observers was much smaller than the among samples component of variance (0.08 sec.^{2} vs. 2.23 sec.^{2}) and was not statistically significant at the $\alpha = 0.001$ level of significance. It was concluded that the precision of measurement resulting from the use of observers with stop watches was adequate.

The data were used to test the hypothesis of normality of the distribution of burn-through times, a requirement of the Church-Harris procedure used for estimating R. A chi-square goodness of fit test rejected this hypothesis at the 0.05 level of significance but accepted it at the 0.01 level. Since the chi-square test, which is relatively insensitive to departures from normality in the region of the tails of a distribution was inconclusive, the statistic $b_1 = n[\Sigma(x_i - \bar{x})^3]^2[\Sigma(x_i - \bar{x})^2]^{-3}$ was calculated and used to test the hypothesis that the distribution is not skewed. This hypothesis was rejected at the 0.02 level of significance (a two tail test with 0.01 probability in each tail was used) so it was inferred that the distribution is positively skewed. Next, Craig's procedure [2] was used to determine which member of the Pearson system of frequency curves best describes the data. It was found that the Pearson Type III curve (a gamma density function) fits the data best. From Carver's table of the standardized Type III function [3], it was verified that the lower tail of the Pearson Type III curve contains less area in the interval $-\infty < x \leq \mu + \alpha$ than a normal curve with the same mean and variance. This indicated that a normality assumption would leave to conservative point and interval estimates of R, i. e., if the estimates are biased, the bias will be such that R is underestimated.

For the purpose of estimating R, it was assumed that cartridge case sidewall burn-through time is normally distributed with estimates of the mean and standard deviation of the distribution being $\bar{X} = 9.71$ seconds and, $S = 1.49$ seconds, respectively.

APPLICATION OF A MODIFIED CHURCH-HARRIS PROCEDURE

The procedure used for estimating R is based on the work of Church and Harris [1]. Let the random variable X be the cartridge case sidewall burn-through time in seconds and the random variable Y be the gun cycle time in seconds. Assume that X and Y are statistically independent and both normally distributed. Introducing the notation

$$E(X) = \mu_1$$

$$VAR(X) = \sigma_1^2$$

$$E(Y) = \mu_2$$

$$VAR(Y) = \sigma_2^2$$

and defining the random variable $W = Y - X$, it follows that W is normally distributed with $E(W) = \mu_2 - \mu_1$ and $VAR(W) = \sigma_1^2 + \sigma_2^2$. Then

$R = P\{Y < X\} = P\{Y - X < 0\} = P\{W < 0\}$.

We next make the transformation $Z = \dfrac{W - (\mu_2 - \mu_1)}{\sqrt{\sigma_1^2 + \sigma_2^2}}$ so that Z is distributed normally with $E(Z) = 0$ and $VAR(Z) = 1$. Then

$$R = P\{W < 0\} = P\left\{Z\sqrt{\sigma_1^2 + \sigma_2^2} + \mu_2 - \mu_1 < 0\right\} = P\left\{Z < \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right\}$$

$$= \Phi\left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) \qquad (1)$$

where $\phi(\cdot)$ is the standard normal cumulative distribution function. Substituting the known values μ_2 and σ_2^2 and the estimated values $\hat{\mu}_1 = \bar{X}$ and $\hat{\sigma}_1^2 = S^2$ into (1) yields the point estimate

$$\hat{R} = \phi\left(\frac{\bar{X} - \mu_2}{\sqrt{S^2 + \sigma_2^2}}\right)$$

Now, having a point estimate of R, we proceed to approximate the probability distribution of the random variable $\hat{R}$ and use this to construct an approximate $100(1 - \gamma)\%$ lower confidence limit for R. In doing so, we make use of the fact that $\bar{X}$ is normally distributed with $E(\bar{X}) = \mu_1$ and $VAR(\bar{X}) = \sigma_1^2/n$; S^2 is asymptotically normally distributed with $E(S^2) = \sigma_1^2$ and $VAR(S^2) = 2\sigma_1^4/(n-1)$; and that $\bar{X}$ and S^2 are statistically independent.

Let $T = S^2 - \sigma_1^2$ so that $E(T) = 0$ and $VAR(T) = \dfrac{2\sigma_1^4}{n-1}$

and define

$$V = \frac{\bar{X} - \mu_2}{\sqrt{S^2 + \sigma_2^2}} = \frac{\bar{X} - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 + T}}$$

Expanding V in a Taylor's series about the point $[E(\bar{X}), E(T)] = (\mu_1, 0)$ we obtain

$$V = \frac{\bar{X} - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} - \frac{1}{2}\frac{(\mu_1 - \mu_2)T}{(\sqrt{\sigma_1^2 + \sigma_2^2})^3} + O\left(\frac{1}{n}\right).$$

with probability one.

Because of the indenendence of $\bar{X}$ and T, the distribution of V is asyasymptotically normal with

$$E(V) = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$

and

$$\mathrm{VAR}(V) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\left[\frac{1}{n} + \frac{1}{2}\,\frac{\sigma_1^2(\mu_1 - \mu_2)^2}{(n-1)(\sigma_1^2 + \sigma_2^2)^2}\right].$$

Using S^2 to estimate σ_1^2 and $\bar{X}$ to estimate μ_1 we obtain as an estimate of the standard deviation of V

$$\hat{\sigma}_V = \frac{S}{\sqrt{S^2 + \sigma_2^2}}\left[\frac{1}{n} + \frac{1}{2}\,\frac{S^2(\bar{X} - \mu_2)^2}{(n-1)(S^2 + \sigma_2^2)^2}\right]^{1/2}$$

$$\text{Since } R = \phi\left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) = \phi[E(V)]$$

and

$$P\left\{\frac{V - E(V)}{\sigma_V} < \phi^{-1}(1 - \gamma)\right\} = 1 - \gamma$$

it follows that

$$P\left\{R > \phi[V - \phi^{-1}(1 - \gamma)\hat{\sigma}_V]\right\} \approx 1 - \gamma \qquad (2)$$

We use (2) to obtain the $100(1 - \gamma)$ % lower confidence limit for R as

$$\mathrm{LCL}_{1-\gamma} = \phi[V - \phi^{-1}(1 - \gamma)\hat{\sigma}_V].$$

COMPUTATIONAL RESULTS

For the combustible cartridge case problem we want a 95% lower confidence limit for R. Since $\phi^{-1}(.95) = 1.645$, this limit is given by

$$LCL_{.95} = \phi(V - 1.645\hat{\sigma}_V).$$

A computer program (Appendix A) was prepared to calculate V, $\hat{R}$, $\hat{\sigma}_V$ and a 95% lower confidence limit for R. In this particular problem a 95% lower confidence limit of $1 - 10^{-4}$ was considered satisfactory to assure the margin of safety required. The 95% lower confidence limit was determined to be $1 - 0.2 \times 10^{-4}$.

Next, the question of how much the mean gun cycle time can be increased and still leave a 95% lower confidence limit of $1 - 10^{-4}$ for R was considered. To answer this question it was assumed that the coefficient of variation of the random variable Y, σ_2/μ_2, remained constant as μ_2 increased, from 2.9 seconds to 3.9 seconds, in steps of 0.1 seconds. V, $\hat{R}$, $\hat{\sigma}_V$ and the 95% lower confidence limit for R were calculated at each step. It was found that the mean gun cycle time can be increased as much as 0.5 seconds without the 95% lower confidence limit for R falling below $1 - 10^{-4}$.

SUMMARY AND CONCLUSIONS

The application of the Church-Harris technique to the combustible cartridge case problem can be summarized as follows: The analysis depended on two critical assumptions; a) the sample of cartridge case sidewall material used to obtain burn-through time measurements was a random sample from the conceptual population consisting of all cartridge cases of the same type which will be manufactured in the future and b) the distributions of X and Y are normal. Assumption a) seemed questionable and it was suggested that further testing be done to verify it. Assumption b) was not satisfactorily established by the data but the analysis indicated that, if the assumption is not valid the inferences drawn from the study will be on the safe side, i. e.,

R will be underestimated. The conditional probability that the sidewall of a cartridge case of a chambered round will not be burned through prior to firing, given that it is instantaneously ignited by smoldering residue remaining from a round previously fired, was estimated to be 0.9999971 with a 95% lower confidence limit of 0.9999779. It was also determined that the mean gun cycle time can be increased as much as 0.5 seconds (a 17% increase) without the 95% lower confidence limit exceeding 0.9999, provided that the coefficient of variation remains constant.

REFERENCES

[1] Church, J. D. and Harris, B. (1970). "The Estimation of Reliability from Stress-Strength Relationships". Technometrics 12, 49-54.

[2] Craig, C. C. (1936). "A New Exposition and Chart for the Pearson System of Frequency Curves". The Annals of Mathematical Statistics, 7, 16-28.

[3] Carver, H. C. (1940). "Statistical Tables". Edwards Brothers, Inc., Ann Arbor, Michigan.

APPENDIX A

The following is a listing of a subroutine CHAR (CHurch-HARris) and a representative driving program written for this study. Notice that CHAR requires as input X(a vector of data), μ_2, σ_2, γ, N and outputs $\bar{X}$,S,$\hat{R}$,$\hat{\sigma}_V$, and LCL. Surboutine CHAR calls subroutines FND and FINVND to evaluate the normal distribution function and the inverse of the normal distribution function respectively.

The output of the sample program is formatted as appears below. The dimension statements (DIMENSION X(500)) appearing in the driving program may be modified to accommodate the data available.

```
      REAL LCL,MU2
      DIMENSION X(500)
C
C        READ INPUT PARAMETERS
C
      READ(5,101)MU2,SIGMA2,GAMMA,N
  101 FORMAT(3F10.3,I10)
C
C        READ DATA
C
      READ(5,102)(X(I),I=1,N)
  102 FORMAT(10X,3F10.5)
      CALL CHAR(X,MU2,SIGMA2,GAMMA,N,XBAR,S,RHAT,V,SIGVHAT,LCL)
      RHAT=1.0-RHAT
      LCL=1.0-LCL
C
C        WRITE COLUMN HEADINGS
C
      WRITE(6,103)
  103 FORMAT(6X,3HMU2,7X,6HSIGMA2,6X,1HN,6X,4HXBAR,10X,1HS,8X,6H1-RHAT,
     19X,1HV,8X,7HSIGVHAT,6X,5HGAMMA,7X,5H1-LCL)
      WRITE(6,104)
  104 FORMAT(1HO)
C
C        WRITE OUTPUT
C
      WRITE(6,105)MU2,SIGMA2,N,XBAR,S,RHAT,V,SIGVHAT,GAMMA,LCL
  105 FORMAT(2E12.3,I6,7E12.3)
      STOP
      END
      SUBROUTINE CHAR(X,MU2,SIGMA2,GAMMA,N,XBAR,S,RHAT,V,SIGVHAT,LCL)   CHAR    1
      REAL LCL,MU2                                                      CHAR    2
      DIMENSION X(1)                                                    CHAR    3
      XN=N                                                              CHAR    4
      XBAR=0                                                            CHAR    5
      SX2=0                                                             CHAR    6
      DO 11 I=1,N                                                       CHAR    7
      XBAR=XBAR+X(I)                                                    CHAR    8
      SX2=SX2+X(I)*X(I)                                                 CHAR    9
   11 CONTINUE                                                          CHAR   10
      XBAR=XBAR/XN                                                      CHAR   11
      S2=(SX2-XN*XBAR**2)/(XN-1.0)                                      CHAR   12
      S=SQRT(S2)                                                        CHAR   13
      VAR=S2+SIGMA2**2                                                  CHAR   14
      V=(XBAR-MU2)/SQRT(VAR)                                            CHAR   15
      RHAT=FND(V)                                                       CHAR   16
      SIGVHAT=(S/SQRT(VAR))*SQRT(1.0/XN+0.5*S2*(XBAR-MU2)**2/((XN-1.0)  CHAR   17
     1*VAR**2))                                                         CHAR   18
      LCL=FND(V-FINVND(1.0-GAMMA)*SIGVHAT)                              CHAR   19
      RETURN                                                            CHAR   20
      END                                                               CHAR   21
```

MU2	SIGMA2	N	XBAR	S	1-RHAT	V	SIGVHAT	GAMMA	1-LCL
0.290E 01	0.130E 00	444	0.920E 01	0.261E 01	0.795E-02	0.241E 01	0.937E-01	0.500E-01	0.120E-01

STATISTICAL EVALUATION OF FLIGHT TEST PERFORMANCE OF THE HELICOPTER LIFT MARGIN SYSTEM (HLMS)

Erwin Biser and Ronald Kurowsky
Avionics Laboratory, USAECOM
Fort Monmouth, New Jersey

ABSTRACT. An item urgently needed by utility aircraft is a measurement of lift margin. Lift Margin is defined as the maximum available lift minus the effective gross weight of the aircraft. It is the intent of this program to define a method whereby this value is automatically presented to the pilot at all times.

Such a system is presently being designed and built under the auspices of the Joint-Army-Navy-Aircraft-Instrument-Research (JANAIR) program. This system will also have the added capability of forcasting Lift Margin (L.M.) to and at a given destination if altitude and ambient air temperature are known. The factors that are most likely to affect the performance of HLMS are torque, air temperature, altitude, fuel weight, and aircraft weight.

The flight test evaluation of HLMS will be performed at US Army Systems Test Activity (ASTA), Edwards AFB, California; and data collected on a pulse-code-FM modulated (PCM) system will be reduced to digital format for evaluation. A reference air density system, corrected for humidity and calibrated by the National Bureau of Standards will be used to define errors caused by calculating air density directly from pressure and temperature.

A statistical analysis of the performance of the entire HLMS and its subsystems (components/sensors) is being undertaken by means of error models to determine and validate the effectiveness of HLMS. The objective is to obtain regression equations of lift margin as a function of torque, temperature, pressure, etc.

Lift Margin (L.M.) = Maximum Available Lift (MAL) - Effective Gross Weight (EGW) LM=0=the point at which the aircraft cannot hover at a higher altitude under its present loading conditions.

INTRODUCTION. The present work was undertaken because it was realized that the safety and utility of helicopters would be substantially improved if the pilot knew at all times the lift capability of the vehicle. There have been a number of accidents wherein the major cause was attributed to the inability of the pilot to accurately assess the lifting ability of the vehicle and the helicopter "stalled". This phenomenon, usually referred to as "settling-with power", typically occurs when the helicopter attempts to take off or land at high vehicle gross weights and high density altitudes. If a helicopter should land in the early morning when the air is cool and relatively dense, it might be impossible for the helicopter to take off with the same load in the afternoon when the air is warm.

Theoretically, it is possible to account for temperature, pressure, humidity and varying loads by the use of slide rules and charts supplied in aircraft handbooks or manuals but such is simply not convenient or fast enough to be practical in most operating situations. In addition, there is the tendency to feel that once the computation has been made, the results will hold for a substantial length of time. Changes in ambient conditions over a period of just a few hours may completely alter the capability of the helicopter.

For these reasons a Helicopter Lift Margin System is extremely desirable. Such a system will indicate to the pilot the present potential lift in excess of the weight of the helicopter. Therefore, before slowing down and landing the pilot can check the helicopter's lift margin and determine if it is safe to hover and complete a vertical landing, or if a rolling landing (STOL type) should be made, or if in fact it is totally unsafe to land in a restricted space.

The major objective of the Helicopter Lift Margin System is to demonstrate the feasibility of continuously computing helicopter lift margin with a desired accuracy under varying operational conditions. Other objectives such as determining the lift variations due to air density measurements with and without humidity inputs and the empirical use of torque to measure fuel consumed will be studied.

DEFINITIONS OF PRINCIPAL PARAMETERS:

Helicopter Lift Margin = Maximum Available Lift - Effective Gross Weight
Maximum Available Lift (MAL)- The maximum left that can be generated by the rotor under ambient conditions of air temperature, air density, altitude, air speed, ground effect and engine characteristics.
Effective Gross Weight (EGW)- The apparent weight of the helicopter as "seen" by the rotor under hover conditions considering air density, ground effect and engine characteristics.

UNDERLYING CONCEPT. A concept by which lift margin may be obtained is illustrated in Figures 1 through 3.

As shown in Figure 1, lift margin is obtained by generating a current proportional to "Potential Lift". A second current that is proportional to weight is subtracted from potential lift. The remainder represents lift margin.

Figure 2 illustrates that available lift is obtained from the computation of potential horsepower multiplied by the ratio of lift to horsepower. Potential horsepower is that horsepower that can be obtained from the engine under the existing ambient conditions. Similarly, the ratio of lift to horsepower is that ratio which holds for the existing ambient conditions.

Figure 3 shows the general concept for obtaining helicopter weight. The horsepower actually being used is measured and converted to weight by

multyplying with the lift-to-horsepower ratio appropriate for the existing ambient conditions. This ratio is identical with the ratio used for the potential lift computation.

The weight of the helicopter is not computed continuously but is computed under flight conditions suitable for the measurement. At other times the weight servo stores the weight so that it is available continuously. The weight may be measured while the helicopter is hovering, with or without wind, with its wheels within a few feet of the ground, i.e., within a fraction of the rotor diameter of the ground or in-ground effect (IGE). The other suitable flight condition is hovering out-of-ground effect (OGE) at zero-indicated air speed; i.e., the helicopter is hovering in the air mass, not necessarily with respect to the ground.

The rest of this article was reproduced photographically from the authors' manuscript.

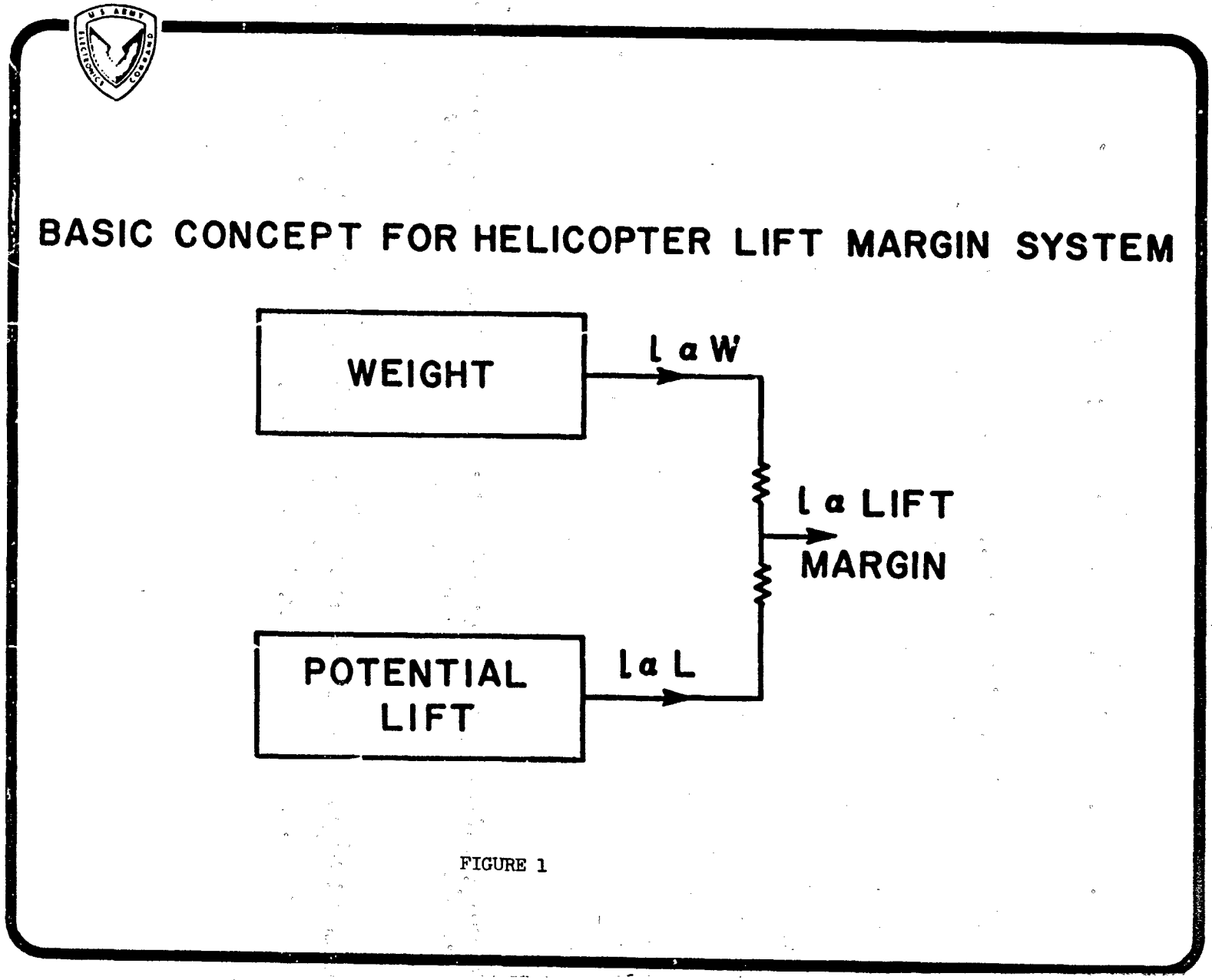

BASIC CONCEPT FOR HELICOPTER LIFT MARGIN SYSTEM
WEIGHT
I a W
I a LIFT
MARGIN
POTENTIAL
LIFT
I a L

FIGURE 1

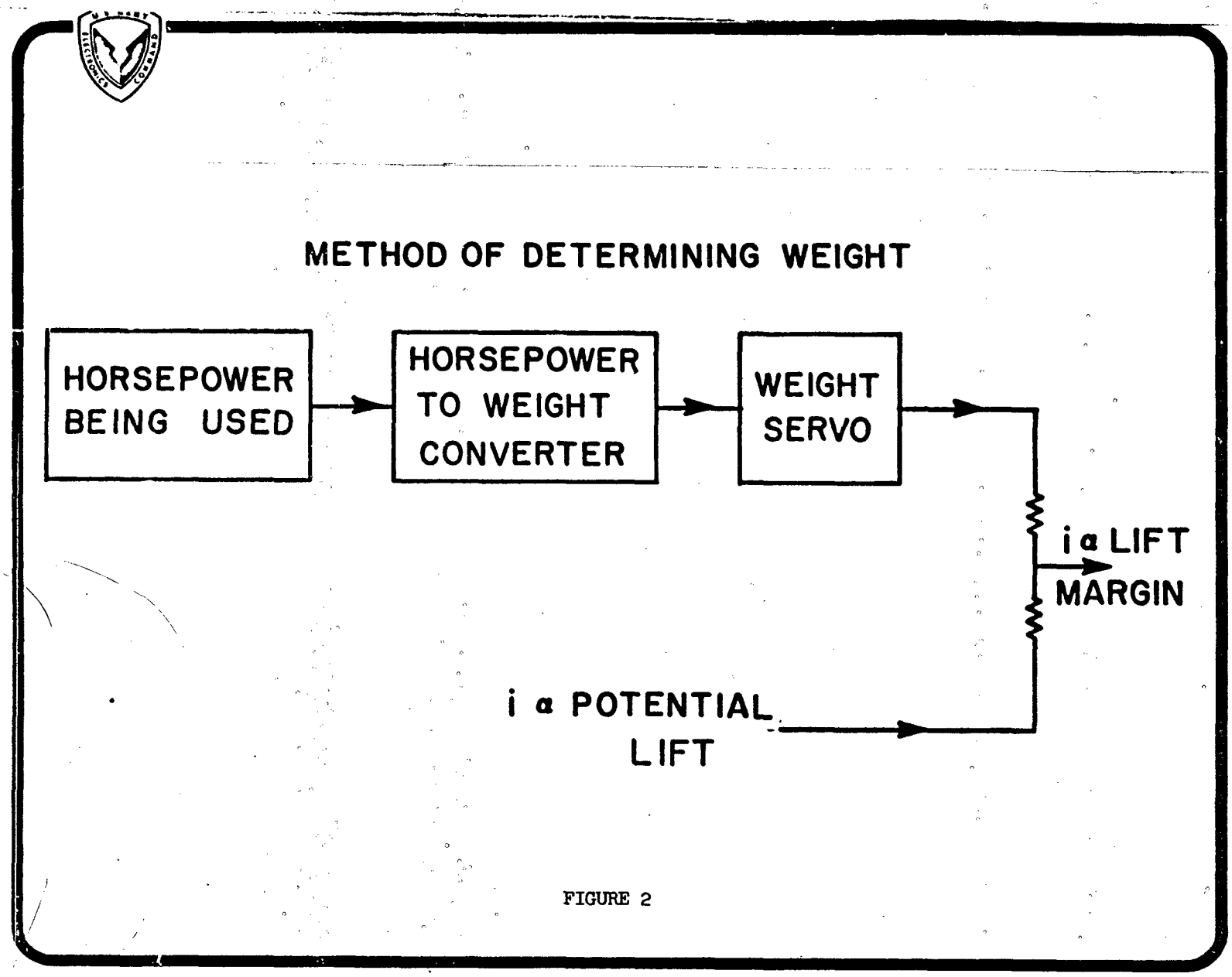

METHOD OF DETERMINING WEIGHT
HORSEPOWER BEING USED
HORSEPOWER TO WEIGHT CONVERTER
WEIGHT SERVO
i α LIFT MARGIN
i α POTENTIAL LIFT

FIGURE 2

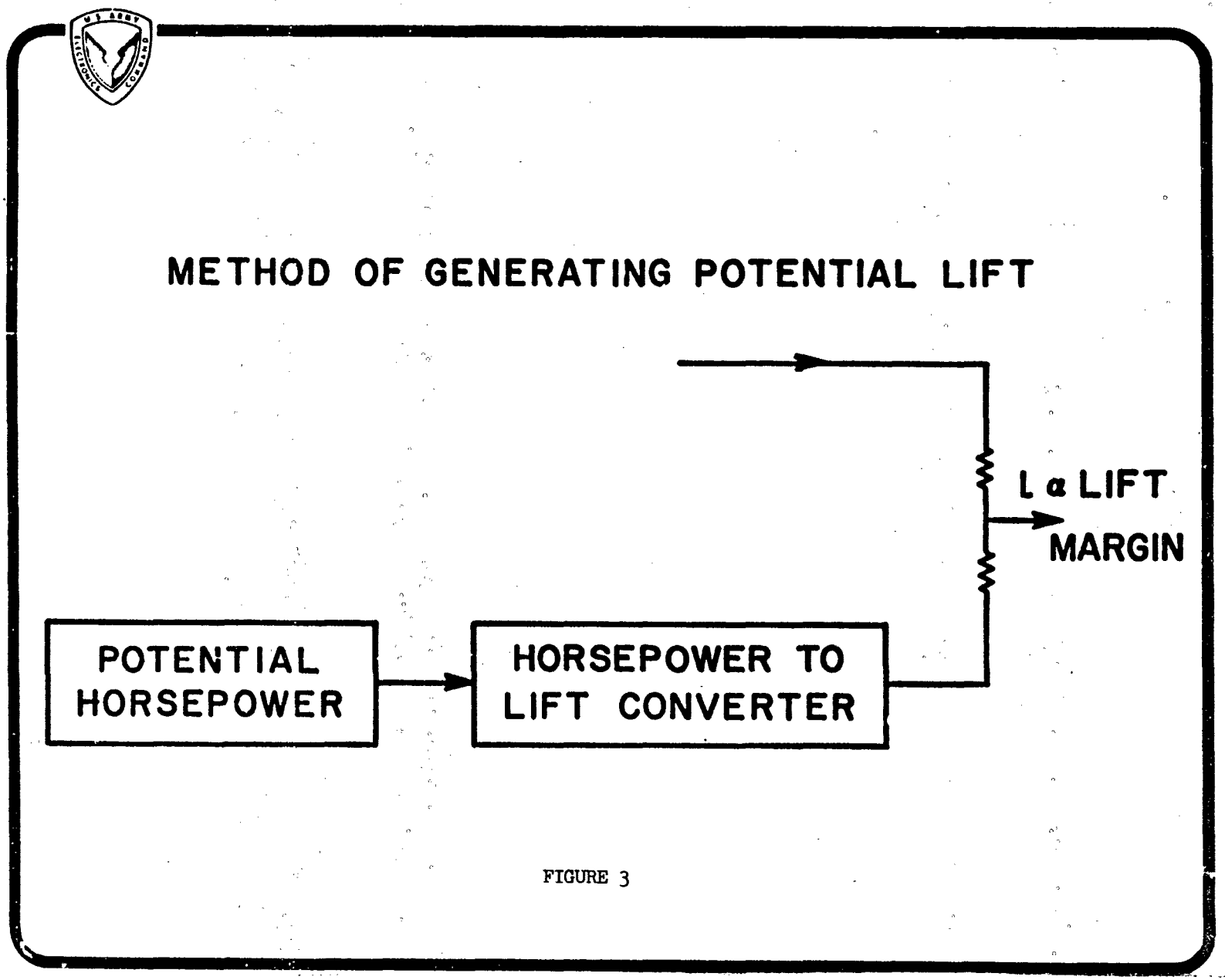

METHOD OF GENERATING POTENTIAL LIFT
POTENTIAL HORSEPOWER
HORSEPOWER TO LIFT CONVERTER
L α LIFT
MARGIN

FIGURE 3

II. SYSTEM DESCRIPTION

FLIGHT TEST CONSIDERATION:

Question: Can the feasibility of the helicopter lift margin system be demonstrated with a meaningful flight test program which avoids the complexity, cost, and time required to implement the complete system?

Answer: Yes, the basic system feasibility can be demonstrated by flight testing the simplified version of the helicopter lift margin system as shown in Figure 4.

Figure 4 illustrates a method of representing the system required to determine MAL and ECW. It is noted that the amount of fuel consumed is obtained by means of an empirical equation based on Torque (Q). The effect of IGE/OGE increases lift by approximately 15% in-ground effect.

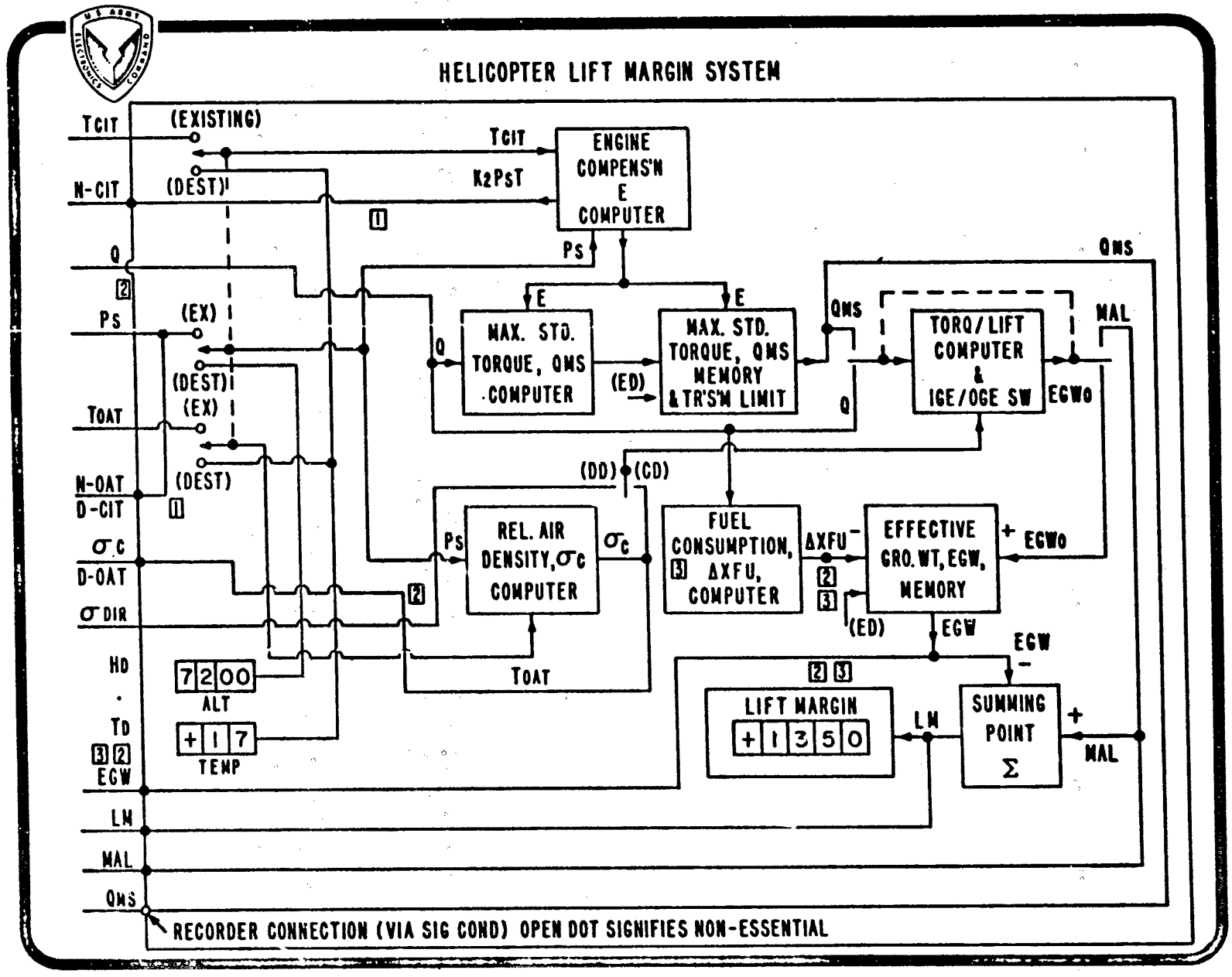
Another item is destination lift margin. An attempt to obtain a measure of this parameter consists of mechanically inserting altitude and temperature into the system with the readout presenting the expected lift margin.

Implementation: To simulate the helicopter/engine characteristics in a computer.

To apply inputs to the computer representing engine torque, rotor speed, air temperature, altitude, air speed, fuel weight, load changes and ground effect.

To compute and display continuously and automatically helicopter lift margin and/or effective gross weight.

To prove feasibility, three basic concepts of the helicopter lift margin system must be verified.


HELICOPTER LIFT MARGIN SYSTEM
TCIT
(EXISTING)
N-CIT
(DEST)
Q
PS
(EX)
(DEST)
(EX)
TOAT
N-OAT
D-CIT
(DEST)
σc
D-OAT
σ DIR
HD
TD
EGW
LM
MAL
QMS
7200
ALT
+17
TEMP
TCIT
K2PST
ENGINE COMPENS'N E COMPUTER
PS
E
E
MAX. STD. TORQUE, QMS COMPUTER
(ED)
MAX. STD. TORQUE, QMS MEMORY & TR'S'M LIMIT
QMS
Q
TORQ/LIFT COMPUTER & IGE/OGE SW
MAL
EGWo
(DD)
(CD)
PS
REL. AIR DENSITY, σc COMPUTER
σc
TOAT
FUEL CONSUMPTION, ΔXFU, COMPUTER
ΔXFU
EFFECTIVE GRO. WT, EGW, MEMORY
+ EGWo
(ED)
EGW
EGW
LIFT MARGIN
+1350
LM
SUMMING POINT Σ
+
MAL
RECORDER CONNECTION (VIA SIG COND) OPEN DOT SIGNIFIES NON-ESSENTIAL

1. The helicopter/engine dynamic characteristics can be simulated in a computer such that maximum available lift can be dynamically computed.

2. The effective gross weight of a helicopter can be accurately measured during a hover maneuver.

3. The difference between maximum available lift and effective gross weight is helicopter lift margin.

Flight test of a hover lift computer will verify the above basic concepts.

Lift is derived from the basic equations of helicopter performance (Gesson & Meyers, "Aerodynamics of the Helicopter").

$L = C_T \pi R^2 \rho (\omega R)^2$ Lift

$Q = C_Q \pi R^2 \rho (\omega R)^2 R$ Torque

$P_H = C_p \pi R^2 \rho (\omega R)^3$ Horsepower

where:

R = rotor radius

ρ = air density

ω = rotor angular velocity

$C_T = 1.414\, C_p^{2/3} M$

M = Figure of merit of rotor system

By operating the aircraft engine at a constant speed (maximum rpm) and using charts available in UH-1B helicopter manual we obtain lift as

$\text{Lift} = m_o (HP)^{2/3} \rho^{1/3} + b_o \rho$

$M_o = 97.4/(HP)^{2/3}$

b_o = -997 (a constant)

By proper manipulation of this equation and charts from the operator's handbook

$$MAL = K_6K_7NQ_{ms}\frac{P}{P_o}(K_1 + K_2T) + K_8\frac{P}{P_o}(K_3 + K_4T)^{-1}$$

$$EGW = K_6K_7NQ_w + K_8\frac{P}{P_o}(K_3 + K_4T)^{-1}$$

$$Qms = Q_T\left[\frac{P}{P_o}(K_1 + K_2T)\right]^{-1}$$

It is these equations that will be used to compute Lift Margin in the HLMS.

III. OBJECTIVE OF EXPERIMENT

Helicopter Lift Margin is defined as the difference between Maximum Available Lift and Effective Gross Weight. The accuracy of system is dependent upon the accuracy of Maximum Available Lift and Effective Gross Weight. The objective of the experiment is therefore to measure the errors in Maximum Available Lift and Effective Gross Weight.

INDEPENDENT VARIABLES OF THE SYSTEM:

1. Ambient temperature (T)
2. Compressor Inlet Temperature (CIT)
3. Absolute Pressure (P)
4. Relative Humidity (RH)
5. Torque (Q)

DEPENDENT VARIABLES:

1. Effective Gross Weight (EGW)
2. Maximum Available Lift (MAL)
3. Fuel Flow
4. Air Density (ρ)

EGW = f(Fuel Flow, Hover Torque, Air Density)

MAL = g(Fuel Flow, Maximum Available Torque, Air Density)

IV. OPERATIONS AND PROCEDURES

Below is described a typical operation of the HLMS, including the initial elements of acquiring the aircraft effective gross weight and aircraft engine potential.

1.1 TOPPING. To store the engine potential, the pilot will, after the aircraft is airborne and the HLMS is turned on, operate the engine at maximum take-off power, putting the craft into a sufficient climb angle to limit the airspeed to a suitable value. The (ENTER DATA) switch is operated along with the out-of-ground effect (OGE) and (EXISTING) switches, the (TOP) pushbutton depressed to enter the engine maximum available torque, Q_M. When the (TOP) pushbutton becomes illuminated, Q_M has been stored in the Q_{MS} memory, correction of Q_M to the maximum standard-condition torque, Q_{MS}, being performed automatically. At this point the (TOP) pushbutton may be released and engine power reduced, and the (ENTER DATA) switch restored to the (DISPLAY DATA) condition. If the (TOP) pushbutton is operated after the (DISPLAY DATA) switch is activated, Q_{MS} will be displayed, proper attention being given (automatically) to the transmission limit.

1.2 WEIGHING. To store the aircraft effective gross weight (EGW), the aircraft is brought to a hover in level flight out-of-ground effect, the (ENTER DATA), (OGE), and (EXISTING) switches all being activated. The hover pushbutton is

depressed until it becomes illuminated, indicating that the aircraft weight has been stored in the EGW memory. This weight is automatically updated for fuel consumption, as may be demonstrated by operating the (DISPLAY DATA) and (HOVER WEIGH) (weight) switches upon which the steadily-decreasing gross weight will be displayed.

1.3 Lift Margin. Once the topping (Para 1.1) and weighing (para 1.2) operations have been performed, the HLMS will read out lift margin continuously in pounds of lift capability for the ambient conditions surrounding the aircraft. This data will be qualified appropriately by the operation of three switches whose functions are described below.

1.3.1 (COMP DENS/DIR DENS) PUSHBUTTON. This switch selects either a computed or externally-supplied value of ambient air density for calculation of rotor lift. The selection is specific to this model of HLMS only, and is used to compare the two methods of deriving air density as applied to the computation of rotor lift.

1.3.2 (OGE/IGE) PUSHBUTTON. This switch is used to increase the computed maximum available lift (MAL) by about 15 percent to approximate the effect of the ground (2-foot skid height only) in lift margin and weighing operations. For instance, an OGE lift margin can be computed from an in-ground-effect (IGE) weight determination. The manually-controlled Hover Lift Computer was not so arranged.

In order to develop error models over the entire operational range of the Lift Margin System, altitude should vary from sea level to 10,000 feet; temperature from 0°C to +25°C and relative humidity from 10% to 100%. To accomplish these variations, four test locations are to be used:

1.	Oxnard, Calif.	Sea level
2.	Edwards, AFB, Calif.	2300 feet
3.	Bishop, Calif.	5300 feet
4.	Coyote Flats, Calif.	9400 feet

Eleven flights per location will be flown. The first flight to check MAL and engine performance, flights 2-10 will consist of three flights per day; one in early morning, one in late morning, and one in early evening for 3 days. The eleventh flight will again check engine degradation over this period.

By use of these locations it is expected that a suitable variation in all these parameters may be achieved.

Another hypothesis to be tested is that relative humidity will have little or no effect on Lift Margin. By use of the modified air density equations given by the National Bureau of Standards, error models will be developed to show the effect of air density measurements made with and without the relative humidity input on the performance of the Lift Margin System.

V. ERROR MODELS AND REGRESSION EQUATIONS

It is assumed that the data, the output of the flight tests, will permit of the construction of regression equations. These equations will establish (probalistic) functional relations between Lift Margin, air density and Torque; and pressure (a subsystem relationship).

It is also assumed that in view of the anticipated complexities of the system and the difficulties of obtaining independent estimates of the effects of the control variables, a predictive linear model will yield the salient characteristics of the behavior of the Lift Margin System response.

The selected independent variables will be tried and tested in the regression equations. Confidence intervals about the coefficients will be determined. This type of model will provide insight as to the response of the lift margin system; and suggest guidelines for more meaningful experimental design in this area. The model will test the stability of the parameters over the sample space, i.e. the operative range of the controlled (and uncontrolled) variables.

1. Types of Linear Model

a. $$E(LM) = \alpha_o + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_1^2 + \alpha_5 X_2^2 + \alpha_6 X_1 X_3 + \alpha_7 X_1 X_2 + \alpha_8 X_2 X_3$$

This is a linear model, linear in the parameters α_i; $i = 0,1,\ldots,8$. The highest power of an independent variable is called the order of the model. It is to be noted that the linearity of a model refers to the linearity (or

non-linearity) of the coefficients.

The test data will enable us to obtain estimates of the true coefficients $a_o, a_1, \ldots, a_8$

X_1 = Torque

X_2 = Air density

X_3 = Pressure

Estimation theory will be applied to obtain estimates of these coefficients: b_o, b_1, b_2, etc. with tolerable confidence limits. The ultimate purpose it to obtain an "optimal" regression equation.

ERROR MODELS

The error models are of significant impact in that they will establish criteria for measuring the performance of the Helicopter Lift Margin system. The error models will also enable us to compare the performance of the system with respect to the standard reference provided by a strain gauge. Distributions of errors for various levels of torque, pressure (and altitude) will be obtained. Estimates of the means and the respective standard errors (of the means) will be obtained.

The error model comparing the effect of air density as computed by the NBS Air Density Equation (on the Helicopter Lift Margin System) with that computed by the algorithms of the Transonics Computer is of special importance. It will be used to validate the equations developed by the National Bureau of Standards (NBS). One of the error models will compare the air density computed by the NBS equations with that computed by ASTA.

SYMBOLOGY

LM	-	Lift Margin
MAL	-	Maximum Available Lift
EGW	-	Effective Gross Weight
T	-	Temperature
CIT	-	Compressor Inlit Temperature
P	-	Pressure
ρ	-	Air Density
Q	-	Torque
Qms	-	Maximum Standard Torque
Ps	-	Static Pressure (absolute)
OGE	-	Out of Ground Effect
IGE	-	In Ground Effect
K_1	-	1.0834
K_2	-	-0.00556
K_3	-	0.94690
N	-	Engine rpm
K_4	-	.0003472
K_6	-	.000352
K_7	-	6.493, K_8 = 225 (Set 1)
K_7	-	7.639, K_8 = 1325 (Set 2)
Q_w	-	Torque at weighing (psi)
P_o	-	Standard pressure 29.921 (in Hg.)

REFERENCES

1. Gessow and Meyers, "Aerodynamics of the Helicopter". Frederick Ungar Pub Co., N.Y. 1952.

2. Edgerton, B., "Final Report for Helicopter Lift Margin System" Vol II, Dec 1970.

3. UH-1B Operator's Manual TM55-1520-219-10, 16 January 1969.

4. Jones, F., "Air Density and Helicopter Lift" Draft Report, NBS Project 2130439.

5. Biser, Dworzak, and Santanelli, "Design of Experiment for Evaluation Flights of Helicopter Lift Margin System Feasibility Model", Draft report, July 1970.

BOX AWARDED THE 1972 SAMUEL S. WILKS MEMORIAL MEDAL

Introductory Remarks Made by Frank E. Grubbs, Conference Chairman

Again, it is so nice for all of us to be together at another Army Design of Experiments Banquet, and it is a pleasure to see all of you once again this year for the Eighteenth Conference. This is not the first time we have had the Army Design of Experiments Conference at Aberdeen Proving Ground. In fact, our Sixth Conference was held here twelve years ago, and Sam Wilks was with us then. As I recall, George Box, Churchill Eisenhart, Stu Hunter, Boyd Harshbarger and Bill Cochran were present for that meeting, but John Tukey was unable to attend, although we are glad to have him back for this year's conference.

Of all things, the program for the Sixth Army Design of Experiments Conference at Aberdeen Proving Ground had a 10 X 10 Graeco Latin Square on its cover! Thus, our first conference at Aberdeen Proving Ground occurred just a year or so after the so-called "Euler-Spoilers" came along. Way back, the great mathematician, Euler, conjectured that it was not possible in general to construct Graeco Latin Squares of <u>even</u> sizes (2n + 2) for the Greek and Latin letters. (I might say, parenthetically, that even though Tukey, Box and others present, are thinking about that statement it is of sufficient accuracy for 90% to 95% of us present anyway!) In any event, with high speed electronic computation capability available, an attempt was made in the late 1950's to construct 10 X 10 Graeco Latin Squares on a computer. A program was set up to generate 10 X 10 Latin Squares, attempting to pair them up to satisfy the Graeco Latin Square condition. For several hundred hours of running time, the unfortunate computer tried to "marry" a Latin Square to a Greek Square and failed to do so. The "Euler-Spoilers" (R. C. Bose, S. S. Shrikande, and E. T. Parker of the University of North Carolina) on hearing about this computer failure proved with the help of advanced group theory that Euler's conjecture was wrong, and if the computer had been left to run the way it was set up, it might have had a 50:50 chance of constructing a 10 X 10 Graeco Latin Square in too many years! Thus, we were just in time for the correct construction of a 10 X 10 Graeco Latin Square and one appears on this cover of our Sixth Army Design of Experiments Conference. Anyone who cares to may check it out! I'll pass it around.

We now turn to the Samuel S. Wilks Memorial Medal.

The Samuel S. Wilks Memorial Medal Award, initiated jointly in 1964 by the U. S. Army and the American Statistical Association, is administered by the American Statistical Association, a non-profit, educational and scientific society founded in 1839. The Wilks Award is given each year

to a statistician and is based primarily on his contribution to the advancement of scientific or technical knowledge in Army statistics, ingenious application of such knowledge, or successful activity in the fostering of cooperative scientific matters which coincidentally benefit the Army, the Department of Defense, the U. S. Government, and our country in general.

The Awards consists of a medal, with a profile of Professor Wilks and the name of the recipient on the reverse, and a citation and honorarium related to the magnitude of the Award funds. The annual Army Design of Experiments Conferences, at which the Award is given each year, are sponsored by the Army Mathematics Steering Committee on behalf of the Office of the Chief of Research and Development, Department of the Army.

The funds for the S. S. Wilks Memorial Medal Award were donated by Philip G. Rust, retired industrialist, Thomasville, Georgia.

Previous recipients of the Samuel S. Wilks Memorial Medal include John W. Tukey of Princeton University (1965), Major General Leslie E. Simon retired (1966), William G. Cochran of Harvard University (1967), Jerzy Neyman of the University of California, Berkeley (1968), Jack Youden (1969) retired from the National Bureau of Standards and deceased, George W. Snedecor (1970) retired from Iowa State University, and Harold Dodge (1971) retired from Bell Telephone Laboratories.

With the approval of ASA President William H. Shaw, the 1972 Wilks Medal Committee consisted of:

Professor Robert E. Bechhofer	- Cornell University
Dr. Fred Frishman	- Army Research Office, Washington,D.C.
Professor J. Stuart Hunter	- Princeton University
Professor Oscar Kempthorne	- Iowa State University
Dr. Badrig Kurkjian	- US Army Materiel Command, Washington, D. C.
Professor Fred Leone	- The University of Iowa
Dr. William R. Pabst, Jr.	- Washington, D. C.
Major General Leslie E. Simon	- Retired
Dr. Frank E. Grubbs, Chairman	- US Army Ballistic Research Labs Aberdeen Proving Ground, Maryland

As many of you conferees are aware, our process of selecting the Wilks Memorial Medalist each year turns out to be a statistically significant event, having to screen 25-30 nominees, fighting out the basic purposes of the Wilks Medal (including what statistics have found wide application to government work and are highly relevant to the Army), and committee members occasionally exchanging insults as the situation demands! Again, however, we certainly got the right man.

The 1972 Samuel S. Wilks Memorial Medalist is an internationally recognized authority on statistics and has contributed greatly to the design and

analysis of scientific experiments. He was born at a place that sounds a bit "redundant", Gravesend, Kent, England, in 1919. He began his career as a Statistician during World War II, and completed his formal education in statistics with a Bachelor of Science Degree (1947) and a PhD (1952) from the University of London. Now as I go along, the Wilks Medalist for 1972 should be shown to be above or go beyond the 95% statistical level of significance, so to speak! So, we will begin accumulating the points. Now this man, during World War II, and as a consequence of his cleaver use of experimental design in the analysis of enemy toxic materials, was awarded the British Empire Medal in 1946. For that we will give him 15 percentage points. After obtaining his PhD in Statistics from the University of London, he left to work for Imperial Chemical Industries Ltd., where he had the opportunity to come into contact with real world experimentation. For that we will give him 10 percentage points and he is up to 25. On leave of absence from Imperial Chemical Industries, he spent the year 52-53 at the Institute of Statistics, Raleigh, North Carolina, and of all things on a research grant supported by the Army Research Office - Durham (ARO-D). Now anyone who would take on some of the statistical problems of the Army deserves special recognition, so we will give him 20 percentage points for succeeding at that. My word, we are already up to 45 percentage points. While working for ARO-D, a famous expository paper on the Exploration and Exploitation of Response Surfaces appeared, along with ideas of "robustness" in the analysis of variance, and also those important rotatable designs. On his return to Imperial Chemical Industries, our 1972 Wilks Medalist delved into statistical methods for the elucidation of basic mechanisms and advanced the concept of Evolutionary Operations. In 1957, he became the Director of the Statistical Techniques Research Group at Princeton University, sponsored also by the Army Research Office, and during the years at "Gauss House" at Princeton he came to know Sam Wilks quite well and established a vigorous statistical center there. Papers on design for non-linear models, simplex sum and three-level designs, the generation of random normal deviates, and papers on adaptive optimization and rebustness to non-normality of regression were completed. For all of this work, most of which was very useful to the Army and others as well, he must be given a significant number of points, anyway, say 20. It seems we just hit 65 percentage points and still counting. In 1960, our 1972 Wilks Medalist left Princeton to become a member of the then Army Mathematics Research Center (now just the Mathematics Research Center since that place was bombed out!) at the University of Wisconsin. At Wisconsin, he immediately took the lead in establishing the Department of Statistics. His research interests while at Wisconsin have steadily broadened and in addition to further contributions to fractional factorials and sequential designs for non-linear models, he became concerned with problems of non-linear estimation, the dynamic control of industrial processes and he "joined the opposition" in contributing the Bayesian methods (!), and finally parametric time series analysis. In fact, last year our 1972 Wilks Medalist was elevated to the R. A. Fisher Chair of Statistics at Wisconsin. Furthermore, we are reminded that back in

1964 he was awarded the Guy Medal of the Royal Statistical Society, London. Now for all this we must give him 30 points and oops we hit the 95% level. But wait a minute! In 1968, our 1972 Wilks Medalist got the Shewhart Medal of the American Society for Quality Control for his contributions there. Now that is a fine honor indeed, but I am reminded that a couple of Army Statisticians got that too, so we will deduct five percentage points bringing him back to only 90 which is not quite enough. Oh my, I have nevertheless forgotten something important. A former summer student employee of mine 20 years ago here at Aberdeen Proving Ground had enough inherent capability in statistics to work with our Wilks Medalist during the year (1952-53) he was turning out all that fine statistical work at the Institute of Statistics, Raleigh, N. C., for ARO-Durham. Furthermore, that former summer student employee of mine, with such excellent training has become a famous statistician in his own right. Now with all of this good work going on and hopefully as a result of some well chosen Army contacts we must add more points, but one can't throw in too many more points, so we will settle for a final 8 points, and award George Box the 1972 Samuel S. Wilks Memorial Medal at 98 percentage points! Congratulations, George Box, and I'll now call on Churchill Eisenhart, Past President of ASA, to present the 1972 Wilks Medal.

GEORGE E. P. BOX RECEIVES THE 1972 SAMUEL S. WILKS MEMORIAL MEDAL

The Presentation of the Award Made by Churchill Eisenhart, Past President of the ASA

The following citation was read:

> "To George E. P. Box, in recognition of his many significant contributions to experimental design, robustness, Evolutionary Operations, Bayesian methods and time series analysis, and for his leadership in relating theoretical results to practical problems."

ACCEPTANCE REMARKS OF GEORGE E. P. BOX ON RECEIVING THE SAMUEL S. WILKS MEMORIAL MEDAL FOR 1972

General Koster and fellow Statisticians: -

You do me especial honor in presenting me with an award commemorating Sam Wilks. Wilks as you all may know was a wise and greatly loved man, who also was a distinguished statistician. He made fundamental contributions to statistical theory, but he was also a man who believed in statistics as a key to solving practical problems. This was evidenced especially by his work beginning in World War II for the National Defense Research Committee, his setting up and his direction of the Princeton Statistical Group, his work on Quality Control, his originating of the unique yearly Princeton Conference and his initiating of the Army

Conference on the Design of Experiments, the Eighteenth of which we presently celebrate.

It has sometimes seemed to me a great pity that his very proper attitude towards the complimentary functions of theory and practice has not been more widely understood. Practice alone uses a too little cook book to produce dishes which are often stale, tasteless and inappropriate to the occasion. Theory alone is directionless and can wander into labyrinths of pointless abstraction. But practice inspiring new theory and theory tested with new practice can produce wonders.

I have sometimes heard members of the statistical profession avow in voices that admired their own liberalism that yes! There should be some schools of applied statistics as well as some concerned with mathematical statistics. The implication was that there really could be no harm in this so long as they remained far enough apart!

It is rather like trying to produce a fine race of children by encouraging the development of a healthy group of men and the separate but equal development of a corresponding group of women, and at the same time taking precaution to ensure that they never see each other.

Sam Wilks was very conscious of these communication problems and it was by instituting such conferences as the present one and by many parallel efforts, some of which I have already mentioned that he contributed to their solution.

One of the difficulties that gets in the way of fruitful interaction between scientific experimenters and statisticians is intellectual arrogance. The fault can lie on either or both sides, but I blush to confess that most often it is the statistician who is in error.

There are various levels of knowledge and ignorance which have been recognised by philosophers. Among these are knowing that you know, not knowing that you know, knowing that you don't know, and not knowing that you don't know. The tragedy of the last category is that once you are in it you remain in it.

Perhaps one story of near disaster will serve as illustration. Twenty some odd years ago I remember being involved in the design of an experiment to compare two treatments A and B from a batch chemical process. I made up my mind early on (too early on as it turned out) that what they needed was a standard paired design in which batches consecutive in time would constitute a pair, and the order (AB) or (BA) would be randomly assigned within each pair. It was quite clear that the chemists thought this was a bad idea. Without further thought I attributed their objection to laziness. I more than hinted that they couldn't be bothered to do the job right! It was only after much argument that I allowed myself to hear

what was the real nature of their objections. The process was one where what is called "carry over" occurs. In these circumstances, the most efficient design is not a paired arrangement but one in which several B's follow several A's which is what they had proposed!

The technical explanation is of course that the carry over phenomenon produces a time series in which successive observations are negatively correlated. The usual assumption that propinquity produces high positive correlation which leads to pairing breaks down. Once I understood this, I was not slow in explaining it to them. In discussion, I pointed out its implications in terms of spectral analysis and the insights it provided on randomization theory. They were very polite about it; they said they didn't know anything about all that but they seemed glad that I had at last seen sense.

I accept this medal aware of my luck in having had patient scientific colleagues and acknowledge the debt I owe them for taking the time to educate a statistician. I know I would have learned more if I had listened better.

AUTOMATED RADAR DATA PROCESSING AT WHITE SANDS MISSILE RANGE FEATURING ADAPTIVE FILTERING WITH BIAS ESTIMATION

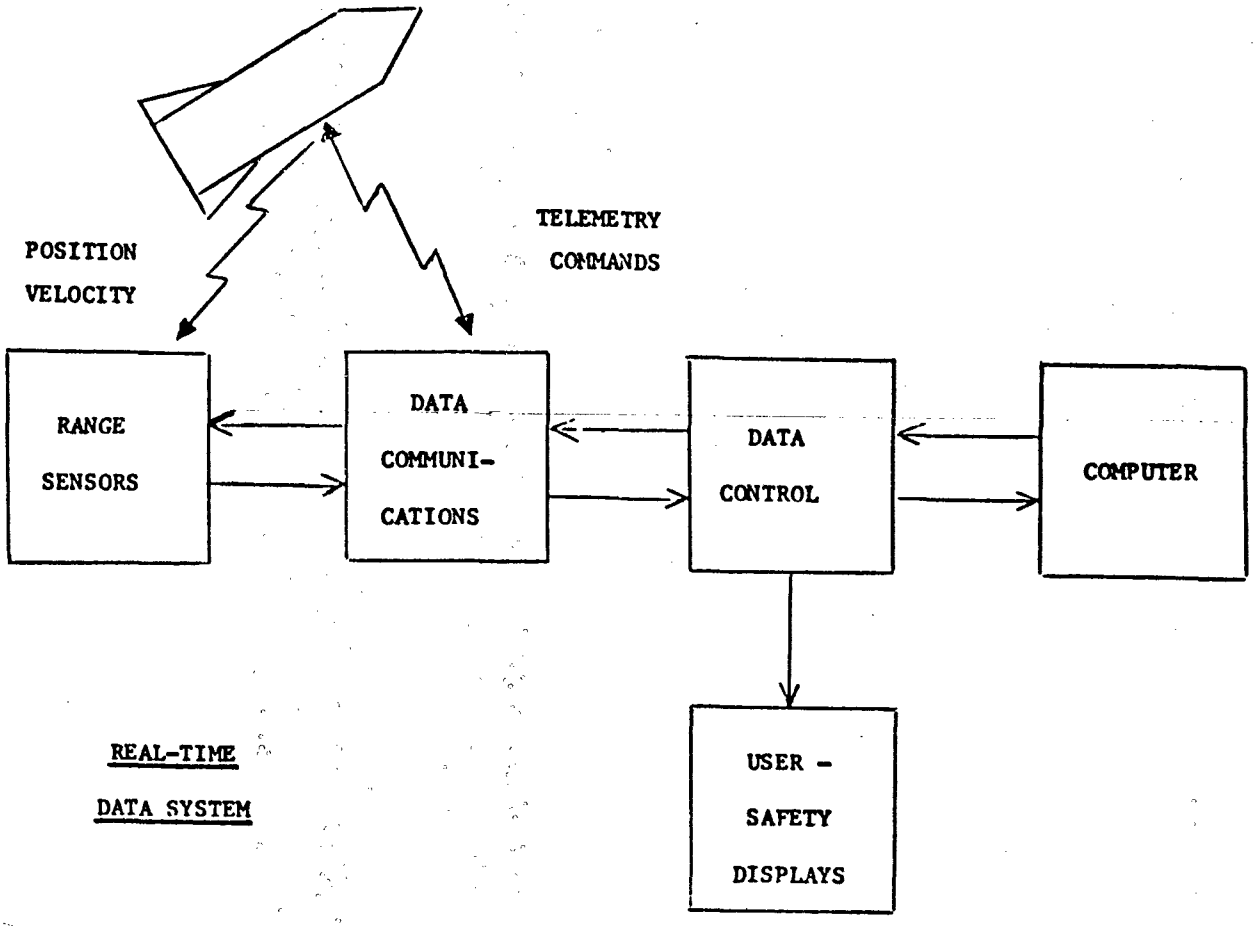
W.A. McCool
Analysis and Computation Division
National Range Operations Directorate
White Sands Missile Range, New Mexico

ABSTRACT. This paper describes the automated system for radar data processing now being implemented at White Sands Missile Range. Such a system will significantly reduce the average delivery time of the data reports containing the processed data and, in priority situations, will provide such data reports essentially on-line, i.e. within minutes after mission completion. Furthermore, the number of instrumentation radars generating data to be processed has been steadily increasing as a result of Range modernization and special radar projects (AMRAD, RAMPART, and others). The automation has become feasible with the availability of: (1) large-scale third generation computing facilities (three UNIVAC 1108 systems); (2) the capability to record all radar data at the central computing site; and (3) software containing new data editing techniques and adaptive digital filters which are based on Kalman filter concepts and, at the same time, are computationally efficient. Feasibility of the automated system has already been demonstrated and its implementation has now advanced to the point of developing operating procedures.

INTRODUCTION. As its major mission, the White Sands Missile Range (WSMR) provides instrumentation, air-space, and supporting facilities for testing missile weapon systems and conducting scientific experiments, e.g. the Army's PERSHING missile system, the Air Force's ATHENA missile system for re-entry research, and the Air Force's Project 621B for proving out a new satellite navigation system. This mission requires a total of 18 precise instrumentation radars for tracking missiles, satellites, aircraft, bombs, balloons, parachutes, and any other flying objects and generating corresponding trajectory (metric) data. During the past year WSMR radars, sampling observed target positions 20 times every flight second, generated more than 250,000,000 space-point measurements. The expeditious processing of this huge volume of radar data is critical to acceptable project support. At the present time, the average delivery time of the standard post-flight reports containing the processed metric data is about ten days. "Quick-look" radar data reports, with a limited number of processed metric parameters, are usually delivered in less than one working day after mission completion. The major objective of the automated radar processing system is to reduce the average report delivery time by an order of magnitude and, in priority situations, to an on-line situation.

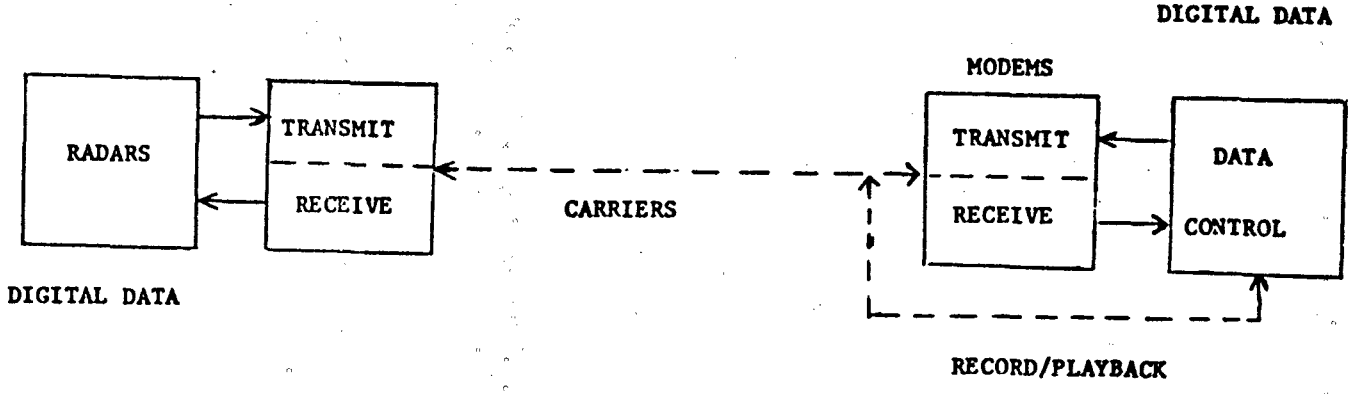
WSMR DATA PROCESSING FACILITIES. The Department of Army has provided WSMR with large-scale third-generation multi-processing, digital computing facilities consisting of two systems (denoted as the A and B systems) each containing two UNIVAC 1108 main-frames and another system (C) containing a single 1108 main-frame. The A system is devoted primarily to real-time mission support, the B system to remote terminal support and normal batch processing of unclassified data, and the C system to batch processing of classified data. The automated processing of radar data will be handled as "background" workload in the A multi-processing system. There is a possibility that, at some future time, part or all of the automated processing will be transferred to the C system. At the present time the WSMR UNIVAC computers are: (1) processing about 16,000 jobs per month producing more than 100,000,000 lines of listing; (2) supporting about 25 real-time missions per month; (3) producing between 800 and 1200 data reports of all types per month; and (4) supporting about 30 remote terminals (a total of 66 terminals is planned).

Figure 1 depicts the basic functions comprising the "Real-Time Data System" (RTDS) at WSMR which was instituted on a very small scale in 1961 and through the Range modernization program has become a highly sophisticated system for satisfying current and future real-time mission support requirements. In addition to the UNIVAC computer A system with its real-time interfaces, the RTDS has a very flexible and reliable data communication subsystem which transmits data from 39 sensors, including radars, to the real-time computer via a Data Control subsystem and from the computer to 44 sensors requiring acquisition (pointing) data. The Data Control subsystem includes a central recording facility which is capable of simultaneously recording/playing-back the data from all the transmitting sensors as well as from the computer when it is generating acquisition data to all the receiving sensors. All WSMR instrumentation radars, when so scheduled, transmit their data in real-time to the central recording facility. The play-back capability not only provides a permanent mission data record but also is an essential component of the automated radar data processing system because, in general, the processing will not be accomplished during missions. Figure 2 indicates how data are transmitted between a sensor and the central data control site.



REAL-TIME
DATA SYSTEM

FIGURE 1



RADAR DATA TRANSMISSION

FIGURE 2

The modulation-demodulation process is employed to allow sensor data to be efficiently and reliably two-way transmitted over relatively inexpensive first-class telephone lines without distance limitation. The transmitting-receiving devices which are called MODEMS, one at each end of the line, utilize tone-modulated "carriers". Recording these carriers with an instrumentation recorder, similar to a home type recorder, is an extremely simple and reliable technique (called analog recording) which has been standard practice at WSMR since 1964. On playback, each carrier is demodulated by a MODEM to reconstruct the original digital data, generated by the sensor, to be received by the real-time computer.

The standard vendor-supplied executive program for the UNIVAC 1108 computer is called EXEC-8 which has been augmented at WSMR for efficient real-time data processing. This augmented EXEC-8 along with specific real-time applications programs can support six missions concurrently in a wide variety of mission types, being limited primarily by the number of available input/output data channels from and to Range sensors. The average number of concurrent missions requiring real-time support at any given time during a typical Range day for the next several years will be no more than two. This means that plenty of computing capacity will be available for back-ground processing required by the automated system.

Real-time data processing utilizing a large-scale digital computer began at WSMR in 1962. Because of the computing time constraints inherent in this type of data processing, there has been a continuing evolution since that time in the development of more efficient data handling, editing, selection, and filtering techniques as well as a wide variety of mathematical procedures, algebraic equations, computation of Kepplerian orbital trajectories for instantaneous impact prediction, coordinate transformations, and command generation.

It should be clear from the foregoing discourse that WSMR now has the necessary facilities to automate its radar data processing, i.e. the third generation digital 1108 computer system with real-time interfaces, the Real-Time Data System with a versatile central recording capability, complete operational software for supporting real-time multi-processing, and a vast experience and "know-how" in real-time data processing techniques

and procedures. This picture would not be complete without clearly emphasizing an operational constraint around which this automated system must be designed: conduct of every Range mission has unqualified priority over all post-flight automated system requirements. More specifically, there are two areas in which this constraint will occasionally arise; suspension of background computing during real-time mission support and data communication problems which would preclude real-time recording but not mission support. The impact of the first situation would merely be a short, insignificant delay; the impact of the second situation would incur the delays involved in physically transporting radar data tapes recorded on-site to the Data Control area in the Range Control Center.

DESIGN OBJECTIVES OF THE AUTOMATED SYSTEM. The design of the automated system for processing radar data has four major objectives.

1. An average delivery time of one working day after mission completion for all standard data reports

Achievement of this objective will effectively combine the post-flight and quick-look types of reports now being provided to the Range users in ten and one days, respectively. Priority delivery may be provided with "on-line" processing, the reports being computer listings generated almost immediately after missions. Reports requiring precise plots will be delivered in two days until an on-line high-speed plotter is acquired. A standard data report limits the user to prescribed options which have been established by frequency of use. Delivery of non-standard reports, i.e. those having unique requirements, must be considered on a case-by-case basis. This objective is motivated by the fact that, in many cases, the value of missile test data to the user decreases rapidly with every passing day after a mission. It should be clear that this objective could not be achieved without central recording and real-time data processing capabilities.

2. Use of Kalman[1,5] filter concepts and real-time editing techniques to optimize the quality of the processed radar data and the computing efficiency of the processing.

This objective has already been accomplished with the proven editing and filtering techniques described later on in this paper. Historically,

radar data at WSMR has been processed using typical classical procedures (e.g. moving-arc least-squares smoothing) which are computationally inefficient and less effective compared to the performance of some of the more modern recursives methods, such as the QD digital filter.[3,4,6] The latter, for example, is employed in all the programs which have been generating quick-look radar reports.

3. Providing multi-station trajectory and bias estimates and all required types of error estimates on a point-by-point basis.

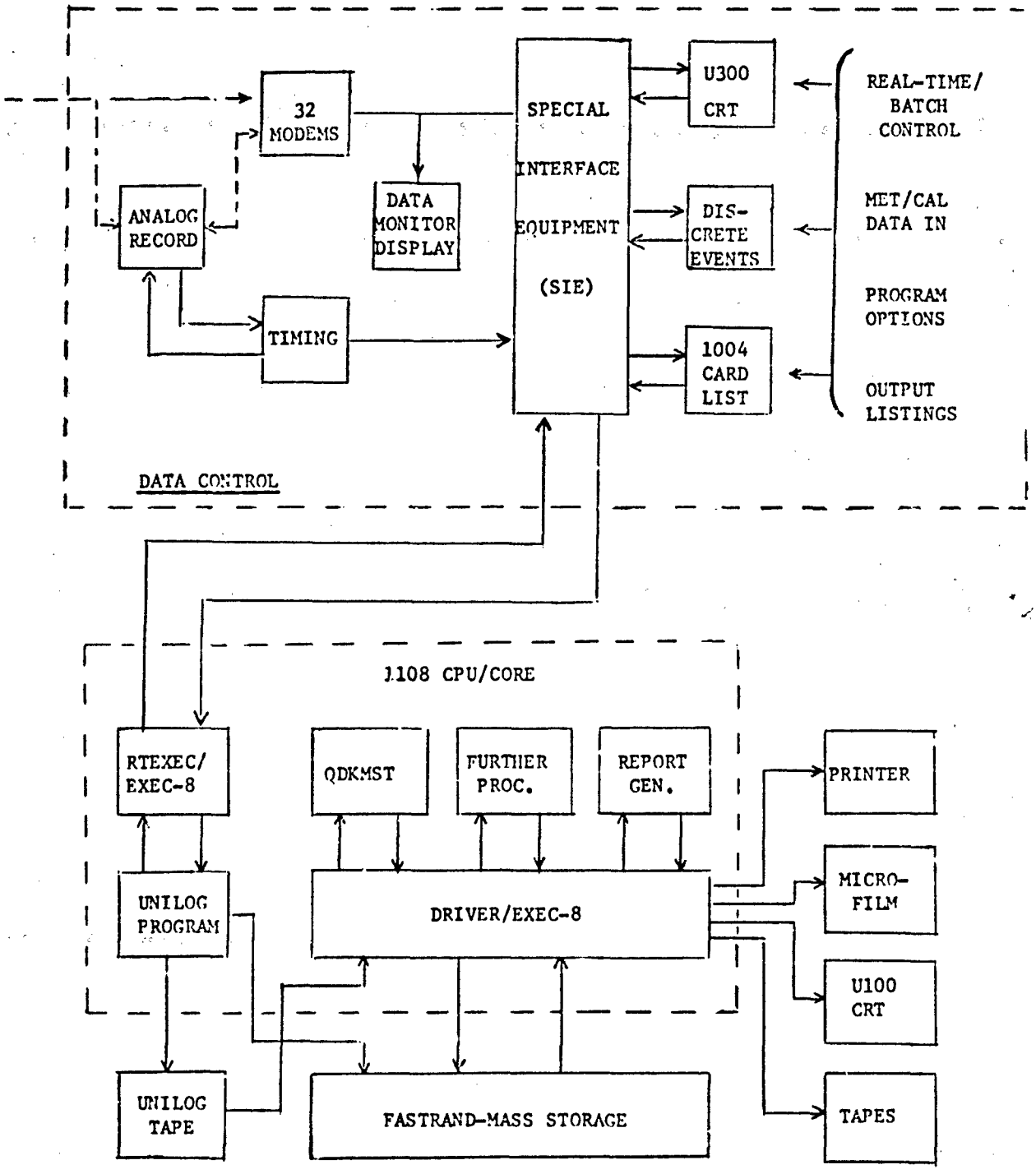
These estimates are included in the automated processing computer program with very little expense in computing time. In the past, multi-station trajectory estimates (sometimes called an N-station solution) have been provided with an average delivery time of 20 days; bias estimates, based on radar data only, have not been provided at all; and the variety of error estimates has been limited by computing time to a few available types. Quick-look data reports, for example, have never provided any error estimates.

4. Simplification and standardization of data report formats.

Current data formats have been established by user specification and historical development. In the automated system, the formats of the data listing, the explanatory information, and plots are being standardized with a variety of available options.

FUNCTIONAL DESCRIPTION OF THE AUTOMATED RADAR DATA PROCESSING SYSTEM. The essential functions of the automated system are shown in the data flow diagram, Figure 3. The system naturally divides into two areas, Data Control and the UNIVAC 1108-A computer. Data Control accepts the radar data from the Range data communication network and other pertinent data; the computer generates the radar data reports.

The radar data entering Data Control is contained in the modulation of the modem carriers, as indicated by the broken lines in Figure 3. The data in this analog form are recorded along with timing data and simultaneously transmitted to the modem receivers for demodulation into digital form. Transmission to the modem receivers is often delayed with a playback of the recorded data. At this point the digital radar data are monitored by Data Control personnel to evaluate data quality and equipment performances. One of the major functions of the Special Interface equipment is to simultaneously accept, reformat, and store the digital data from as many as 32 radars (or other sensors) and transmit the data to the computer. Prior to accepting radar data during a mission or a play-back in Data Control, radar calibration data (i.e. corrections for tilt, beam misalignment, etc.),



AUTOMATED RADAR DATA PROCESSING SYSTEM

FIGURE 3

radar meteorological data for refraction corrections, radar identification data, and program and report parameter options must be fed into the SIE by a system controller via a CRT terminal (U-100) keyboard and/or punch-cards (1004 card reader-punch-printer).

When the radar data and the pre-processing information are accepted by the computer from the SIE, they are recorded on a log tape (UNILOG) and/or magnetic drum mass storage (FASTRAND) under control of RTEXEC, the real-time monitor. In other words, the data logging is a real-time process. After logging is completed, a batch job with high priority is automatically initiated to process the logged data and generate a data report. As indicated in Figure 3, the processing consists of three phases under the control of a Driver program (in turn controlled by EXEC-8, of course); QDKMST (which is described in the later chapter) concurrently edits and filters the data from several radars and generates derivative parameters and error estimates; "further processing" software consists of a variety of optional sub-programs to generate ancillary parameters as required by the Range users; report generation consists of conversion of parameters to specified engineering units, setting up data report formats, paging, etc.; and interfacing with one or more of the selected output devices, line printer, microfilm printer, cathode ray tube (CRT) display, or magnetic tape transports.

From the functional description, it should be clear that generating a data report for a single Range mission involving several radars is a straightforward process. The question may be raised, however, as to the effectiveness of the automated system under a heavy workload. In response to this question, it should be emphasized that the automated system may be viewed as a "next generation" version of the quick-look system which is now generating about half of all WSMR radar data reports.

As noted before, the average delivery time for quick-look radar data reports is one working day after mission completion and this has been accomplished with an IBM 7094II/7044 Direct Couple System (DCS) whose job stream execution is sequential, i.e. one job at a time. In contrast, the automated system has a dual 1108 multi-processor (A system) which executes six jobs and supervises all input-output activity concurrently. Even though System A is dedicated to real-time mission support, it has more than ample capacity to satisfy the automated system's computing requirements. Thus, the major bottleneck will be the manual time and effort to set up each job in accordance with user requirements and options and collecting and inserting pre-processing data.

REVIEW OF THE QDK[7] FILTER/SMOOTHER THEORY. There has been a long-standing, urgent need for a process to filter radar data which would be significantly superior to those currently in use with respect to computing speed and filtering effectiveness. It was recognized that such a process would have to adaptively effect an optimum trade-off between noise suppression (attenuation) and best fit (minimum distortion of the true signal) in accordance with the characteristics of the noise content and the kinematics of the input data. It was also recognized that the increased computing speed constraint would almost surely dictate the use of a recursive procedure. The QDK filter[7] was developed to satisfy this need. The basic QDK structure is identical to that of the discrete Dalman filter but its detailed formulation is greatly simplified using the QD filter theory (hence, the QDK acronym) for rapid computation without seriously degrading the Kalman optimality. In order to indicate how these simplifications are applied we begin with the computational step-by-step formulas for the discrete Kalman filter.

$$\bar{x} = \Phi\hat{\bar{x}} + \Gamma u \qquad \text{- state prediction} \qquad (1)$$

$$\bar{P} = \Phi\hat{\bar{P}}\Phi^t + \Psi Q \Psi^t \qquad \text{- state covariance prediction} \qquad (2)$$

$$W = \bar{P}H^t(H\bar{P}H^t + R)^{-1} \qquad \text{- weighting matrix} \qquad (3)$$

$$\hat{x} = \bar{x} + W(z - H\bar{x}) \qquad \text{- state estimation} \qquad (4)$$

$$\hat{P} = (I - WH)\bar{P} \qquad \text{- state covariance estimation} \qquad (5)$$

where:

$\bar{x}$ - predicted state vector

$\hat{x}$ - current estimated state vector

$\hat{\bar{x}}$ - preceding estimated state vector

z - input data vector

Φ - transition matrix

Γ - control coupling matrix

Q - uncertainty matrix of the plant-process model

Ψ - model uncertainty coupling matrix

H - input/state Jacobian

$\bar{P}$ - predicted covariance matrix of the plant-process model

$\hat{P}$ - current estimated model covariance matrix

$\hat{\bar{P}}$ - preceding estimated model covariance matrix

The data from instrumentation radars at WSMR are generated at 20 samples per second in spherical coordinates, i.e. range (R), azimuth (A), and elevation (E) in the local radar reference systems. Each system is established by the local tangent plane with zero azimuth being true or grid north. It is common practice to transform these data, prior to filtering, into an arbitrarily located and oriented Cartesian coordinate system with components z_a, z_b, and z_c. For Kalman filtering the z_a, z_b, z_c data from a single radar, the dimensions of the vectors and matrices in the computational formulas (1) through (5) are identified as follows:

$\bar{x}, \hat{x}, \hat{\bar{x}}$	- 6x1	Φ	- 6x6
u	- 6x1	Γ	- 6x6
$\bar{P}, \hat{P}, \hat{\bar{P}}$	- 6x6	Q	- 6x6
Ψ	- 6x6	H	- 3x6
R	- 3x3	I	- 6x6 (unit)
W	- 6x3	z	- 3x1

The elements of each of the state vectors $\bar{x}$, $\hat{x}$, and $\hat{\bar{x}}$ are three position and three velocity components while the elements of the z vector are the three filter "inputs". The u vector and the matrices are correspondingly dimensioned. It is clear from these dimensional data that the straightforward computations, including the matrix inversion, involved in using the Kalman formulas (1) through (5) for each computing step would be prohibitive, particularly when the plant process parameters in Φ, Γ, and Ψ, as well as in H are time-varying.

The first simplifications in our application of Kalman filtering to radar data segment the basic formulas into three independent filters, one corresponding to each Cartesian component. To achieve these simplifications, the following reasonable assumptions and approximations are made:

(1) The plant process is the radar target whose mathematical model represents a point mass with acceleration components being independent and slowly varying with time. Thus, each acceleration component is assumed to be constant during any integration interval. This assumption effects independence of the component models so that all of the model matrices and vectors of the basic Kalman formulas can be suitably partitioned and separated.

(2) The R matrix, which describes the noise characteristics of the input data components z_a, z_b, and z_c has non-zero cross-correlation elements due to the non-linearity of the R,A,E to z_a, z_b, z_c coordinate transformation. In practice, these elements are so small that the variances of the input data components can be assumed to be independent, i.e. R is a diagonal matrix.

(3) The foregoing assumptions and approximations allow partitioning of all the vectors and matrices so that each component of input data can be independently Kalman filtered. In each of these three filters it is convenient to augment the Φ matrix, which is the same for each filter, increasing its dimensions from 2x2 to 3x3 in order to eliminate the u vector and the Γ matrix.

With the above simplifications the dimensions of the component Kalman filters are:

$\bar{x},\hat{x},\hat{\bar{x}}$	- 3x1	Φ	- 3x3
$\bar{P},\hat{P},\hat{\bar{P}}$	- 3x3	Q	- 3x3
R	- 1x1	H	- 1x3
W	- 3x1	z	- 1x1

which are defined as follows:

$$\bar{\mathbf{x}} \equiv \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \dot{\bar{x}} \\ \ddot{\bar{x}} \end{bmatrix}, \quad (6) \qquad \bar{P} \equiv \begin{bmatrix} \bar{P}_{11} & \bar{P}_{12} & \bar{P}_{13} \\ \bar{P}_{21} & \bar{P}_{22} & \bar{P}_{23} \\ \bar{P}_{31} & \bar{P}_{31} & \bar{P}_{33} \end{bmatrix} \quad (7)$$

$\hat{\mathbf{x}}$, $\bar{\hat{\mathbf{x}}}$, $\hat{P}$, $\bar{\hat{P}}$ are analogous to (6) and (7), respectively. If Taylor series integration is used to generate the augmented Φ matrix from the model differential equation, i.e.

$$\ddot{\mathbf{x}} = \ddot{\hat{\mathbf{x}}}(t) = \text{constant}, \quad (8)$$

then

$$\Phi = \begin{bmatrix} 1 & h & h^2/2 \\ 0 & 1 & h \\ 0 & 0 & 1 \end{bmatrix} \quad (9)$$

where

$h = \Delta t$ = sampling inverval, .05 sec.

Also,

$$R = r \quad (10)$$

$$z = z \quad (11)$$

$$H = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \quad (12$$

$$W = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \quad (13)$$

When (7), (10), and (12) are substituted into (3), it reduces to the simple formula,

$$W = \frac{1}{p_{11} + r} \begin{bmatrix} p_{11} \\ p_{21} \\ p_{31} \end{bmatrix}, \tag{14}$$

in which the matrix inversion in (3) becomes a scalar division operation.

A further dramatic reduction in computations is achieved by applying the OD filter theory to formulas (2) and (5) of the component Kalman filters. This is the first step in the evolution of the ODK formulas from the basic Kalman formulas (1) through (5). Starting with (1), in which Γ and u have been eliminated in the component filter, we have the simplified state prediction formula:

$$\bar{x} = \phi\hat{x}, \tag{15}$$

in which ϕ is given by (9) and $\bar{x}$ and $\hat{x}$ by (6). Thus, (15) and (4) are the second-order OD filter formulas if, in (4) and (13),

$$w_1 = \frac{60M^2}{10M^3 + 33M^2 + 23M - 6}, \tag{16}$$

$$w_2 = 2w_1/h, \tag{17}$$

$$w_3 = 2w_1/h^2 = w_2/h, \tag{18}$$

in which the development of w_1 in (16) is based on a second-degree curve fitting an arbitrary span of the last M input data values in a constrained least-squares sense. The intercept and slope of the fitted curve are constrained at the time point corresponding to the last input data value implicitly dropped from the span. Using the OD theory, approximate functional relationships among the elements of the $\bar{P}$ matrix have been developed and applied to formula (2) to obtain the scalar recursive formulas:

$$\bar{P}_{11} = \beta^4 \hat{\bar{P}}_{11} + q, \tag{19}$$

$$\bar{P}_{21} = 2\alpha(\beta^3 \hat{\bar{P}}_{11} + q), \tag{20}$$

$$\bar{P}_{31} = 2\alpha^2(\beta^2 \hat{\bar{P}}_{11} + q), \tag{21}$$

where:

$$\alpha = \frac{1}{Mh}, \tag{22}$$

$$\beta = 1 + \frac{\dot{\hat{\bar{P}}}}{M}, \tag{23}$$

and q is the model uncertainty scalar corresponding to Q in (2). (19), (20), and (21) are required in (14) to compute the elements of the QDK weighting vector. The value of M, the equivalent OD filter span, in (22) and (23) is retained from the preceding computing step in which M is estimated with a polynomial inversion of (16) for the corresponding computed value of w_1, i.e.

$$M = a_1 + a_2 w_1 + a_3 w_1^2 + a_4 w_1^3. \tag{24}$$

The value of $\hat{\bar{P}}_{11}$ in (19), (20), and (21) is retained from the preceding computing step as $\hat{P}_{11}$ which is estimated with the scalar formula

$$\hat{P} = (1 - w_1)\,\bar{P}_{11} \tag{25}$$

derived by substituting (7), (12), and (13) into (5).

From the foregoing exposition we can now summarize the QDK computational formulas. The first three are the prediction formulas obtained by using (6) and (9) in (15). The estimation formulas (32), (33), and (34) are obtained by using (6), (12), and (13) in (4).

$$\bar{x} = \hat{x} + h\hat{\dot{x}} + \frac{h^2}{2}\hat{\ddot{x}} \tag{26}$$

$$\bar{\dot{x}} = \hat{\dot{x}} + h\hat{\ddot{x}} \tag{27}$$

$$\bar{\ddot{x}} = \hat{\ddot{x}} \tag{28}$$

$$\alpha = 1/(Mh), \tag{22}$$

$$\beta = 1 + 1/M, \tag{23}$$

$$p_{11} = \beta^4 p_{11} + q, \tag{19}$$

$$p_{21} = 2\alpha(\beta^3 p_{11} + q) \tag{20}$$

$$p_{31} = 2\alpha^2(\beta^2 p_{11} + q), \tag{21}$$

$$w_1 = p_{11}/(\bar{p}_{11} + r), \tag{29}$$

$$w_2 = \bar{p}_{21}/(\bar{p}_{11} + r), \tag{30}$$

$$w_3 = \bar{p}_{31}/(\bar{p}_{11} + r), \tag{31}$$

$$\hat{x} = \bar{x} + w_1(z - \bar{x}), \tag{32}$$

$$\hat{\dot{x}} = \bar{\dot{x}} + w_2(z - \bar{x}), \tag{33}$$

$$\hat{\ddot{x}} = \bar{\ddot{x}} + w_3(z - \bar{x}), \tag{34}$$

$$M = a_1 + a_2 w_1 + a_3 w_1^2 + a_4 w_1^3, \tag{24}$$

$$\hat{p}_{11} = (1 - w_1)\bar{p}. \tag{25}$$

The QDK formulation for independently filtering each data component is simply extended to the multi-station formulation. To do this, the inverse variance weighted average for each of the Cartesian components in a common reference system is used as the input to the single component QDK filter. These weighted averages z_m for N radars are computed with

$$z_m = \left(\sum_{i=1}^{N} z_i/r_i\right) \Big/ \sum_{i=1}^{N} 1/r_i \tag{35}$$

in which the z_i are the corresponding component bias-free inputs with their respective variances r_i. In practice, the z_i are not bias-free. If b_i are the bias estimates then (35) is modified as

$$z_m = \left(\sum_{i=1}^{N} (z_i - b_i)/r_i\right) \Big/ \sum_{i=1}^{N} 1/r_i. \tag{36}$$

Thus, the QDK multi-station formulation must include bias estimation procedures. In the development of the bias estimation formulas, which are also based on Kalman and QD theories, the biases are assumed to be slowly varying and independent of each other as well as the multi-station estimation. The bias computation formulas are:

$$\bar{P}_i = \hat{\bar{P}}_i + kq_i, \; i = 1,2,\text{----}N \tag{37}$$

$$w_i = \bar{P}_i/(\bar{P}_i + r_i), \tag{38}$$

$$\hat{b}_i = \hat{\bar{b}}_i + w_i\,(z_i - z_b - \hat{\bar{b}}_i), \tag{39}$$

$$\hat{P}_i = (1 - w_i)\,\bar{P}_i. \tag{40}$$

where:

$\bar{P}_i$, $\hat{P}_i$, $\hat{\bar{P}}_i$ - bias model variance scalars,

q_i - model uncertainty scalars,

k - q_i weighting factor,

w_i - bias weighting coefficients

z_b - arbitrary bias reference.

z_b may be optical data components known to be essentially bias free, the component inputs of one particular radar, or the unweighted average of the z_i

$$z_b = \frac{1}{N} \sum_{i=1}^{N} z_i. \tag{41}$$

THE QDKMST PROGRAM FOR THE AUTOMATED RADAR DATA PROCESSING SYSTEM. QDKMST employs the QDK filter described in the preceding chapter and is modularized into ten subroutines one of which contains three QDK filter/smoothers for each set of radar data. QDKMST includes pre-execution, initializing, and data subprograms.

In addition to concurrently generated trajectory and derivative estimates for each set of radar data, QDKMST also optionally provides corresponding multi-station estimates along with relative trajectory bias estimates at the R, A, E level for each set of radar data. An alternate option consists of estimates selected, on a point-by-point basis, from those of the "minimum total variance" radar whose data have been bias corrected.

QDKMST employs a total of 354 zero, first, and second order QD filters[3,4,5] in addition to the 33 QDK filter/smoothers to generate the estimates for ten sets of radar data and a multi-station solution.

Execution time per set of observed data per radar for QDKMST on any of the UNIVAC 1108 computer systems is less than 4.0 milliseconds. For ten sets of radar data the multi-station estimates per sample time require approximately 15 milliseconds. If the radar data are edited at the R, A, E level at 20 samples per second and then averaged down to five samples per second prior to transforming the data to the common Cartesian coordinate system, then the data from ten radars can be reduced, including the multi-station or minimum variance estimates, in less than 40 milliseconds, allowing 160 milliseconds for data handling, executive overhead, other batch processing, and a reasonable safety margin within the 200 millisecond on-line execution period.

QDKMST contains the same radar error model for calibration corrections as that described in the WSMR Data Reduction Handbook[2]. The same refraction index computation is also used. The refraction correction, however, with spherical earth geometry is the same as that used in the real-time programs which support ATHENA and PERSHING.

QDKMST itself is the calling routine for each of its subroutines whose functions are outlined as follows. The FORTRAN source program includes definitions of all the parameters and variables of the COMMON statements.

INTQDK

a. Initializes QDKMST variables.

EDIT

a. Accepts range, azimuth, and elevation observations from each tracking radar;

b. Initializes QD editing filters when several consecutive 4th differences of the R, A, E data are less than some limit;

c. After initialization, the QD editing filters test for spikes, and when spikes are detected, substitute predicted filter values;

d. Reinitializes the editing process after total data dropouts;

e. Estimates variances of edited R, A, E data;

f. Generates acceptable data flags for each set of radar data; and

g. Optionally averages edited data from 20 samples per second down to five samples per second.

RCXFRM

a. Corrects edited range and elevation data for refraction and applies calibration corrections; and

b. Transforms edited R, A, E data for each radar to an arbitrarily selected common Cartesian reference system.

MSBS (Optional)

a. Estimates R, A, E biases for each radar relative to one of three references, optionally selected;

b. Generates a variance weighted average of available sets of bias corrected data in the common Cartesian coordinate system, and

c. Identifies the minimum total variance radar.

QDK

a. Generates smoothed trajectory and derivative estimates for each set of radar data and the multi-station set of data (optional);

b. Estimates variances of QDK Cartesian coordinate input data and generates point-by-point scalar quantities corresponding to the Kalman R, Q, and P matrices; and

c. As an alternate to the multi-station estimates, optionally selects trajectory and derivative estimates from the minimum total variance radar on a point-by-point basis.

STAT

a. Generates error estimates (variances) of the QDK smoothed Cartesian component positions, velocities, and accelerations.

b. Estimates the means of the Cartesian component residuals as a measure of smoothing distortion.

c. QDK noise attenuation factors.

d. Generates the means of the R, A, E residuals and bias estimates.

e. All estimates are generated on a point-by-point basis.

BLOCK DATA

Contains all of the input parameters and constants which are compiled into QDKMST.

PREMIS

a. Computes a variety of parameters and constants not contained in BLOCK DATA;

b. Computes QD filter coefficients;

c. Computes parameters for refraction corrections.

INTCAL (called by PREMIS)

a. Computes parameters for calibration corrections and applies tilt corrections to rotation matrices.

REFERENCES

1. R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", Journal of Basic Engineering, 1960.

2. "Handbook of Data Reduction Methods", Technical Report, August 1964, Data Reduction Division, White Sands Missile Range, New Mexico.

3. W. A. McCool, "QD - A New Efficient Digital Filter", Internal Memorandum No. 60, August 1967, Analysis and Computation Directorate, White Sands Missile Range, New Mexico.

4. W. A. McCool, "A Matrix-Vector Formulation of the QD Digital Filter", Technical Report No. 3, December 1968, Analysis and Computation Directorate, White Sands Missile Range, New Mexico.

5. W. A. McCool, "A Derivation of the Discrete Kalman Filter Formulation Based on the QD Filter and Classical Least-Squares Estimation Theories", Technical Report No. 9, April 1969, Analysis and Computation Directorate, White Sands Missile Range, New Mexico.

6. W. A. McCool, "Zero and First Order QD Digital Filters", Technical Report No. 15, March 1970, Analysis and Computation Directorate White Sands Missile Range, New Mexico.

7. W. A. McCool, "QDK - A New Adaptive Digital Filter", Technical Report No. 24, June 1971, Analysis and Computation Division, White Sands Missile Range, New Mexico.

ALGORITHM FOR EDITING BIVARIATE DATA FILES WITH RANDOM SPACING IN THE INDEPENDENT VARIABLE

1LT L.D. Clements
Data Reduction Section
Yuma Proving Ground
Yuma, Arizona

Methods for smoothing equally spaced, bivariate data have been under development for the past one hundred years. A wide variety of techniques have been adopted for this purpose (see Whittaker and Robinson, The Calculus of Observations, Dover, 1967). A special case within the larger problem of data smoothing is the need to remove points which are grossly in error with respect to the surrounding data. Again, techniques are available for use with evenly spaced data (see Handbook of Data Reduction Methods, White Sands Missile Range, 1964). In the literature available to the author, however, there appears to be no method available for directly editing gross values out of a bivariate data file with uneven spacing in the independent variable. The intent of this paper is to introduce such a technique.

DEVELOPMENT OF TECHNIQUE. Consider first the equally spaced string of numbers given in Table 1 and plotted in Figure 1. It is obvious from the figure that point number 11 is in error, as are points 21, 22, and 23. The question then becomes, does a single "bad" point or series of "bad" points generate a discernible pattern in the derivatives (actually differences) which may be taken? The table and figure show that indeed there is a definite pattern of anomalous difference values generated by erroneous points.

The approach used was then to say, if in the sequence of derivatives a given pattern of anomalous values is present, then the point(s) generating this pattern must be incorrect. Symbolically, let us define for a datum pair (x_i,t_i), the first difference pair $(v_{i,i+1},t_{i,i+1})$, where

$$v_{i,i+1} = \frac{x_{i+1}-x_i}{t_{i+1}-t_i} \qquad (1)$$

and

$$t_{i,i+1} = \frac{t_{i+1}+t_i}{2} \qquad (2)$$

and the second difference

$$a_{i,i+1,i+2} = \frac{v_{i+1,i+2} - v_{i,i+1}}{t_{i+1,i+2} - t_{i,i+1}} \qquad (3)$$

Now if the point x_i is bad, then $v_{i-1,i}$, $v_{i,i+1}$, $a_{i-2,i-1}$, $a_{i-1,i,i+1}$, and $a_{i,i+1,i+2}$ are abnormal with respect to surrounding values. Similarly, if points x_i, x_{i+1}, ..., x_{i+n} are offset in the same direction such as points 21, 22, 23 in Figure 1, $v_{i-1,i}$ and $v_{i+n,i+n+1}$ and $a_{i-2,i-1,i}$, $a_{i-1,i,i+1}$, $a_{i+n-1,i+n,i+n+1}$, and $a_{i+n,i+n+1,i+n+2}$ are abnormal. Although to carry out this sequence of calculations by hand is ridiculous, a digital computer may rapidly use this pattern identification technique to locate bad points in an unevenly spaced bivariate data stream.

FEATURES OF THE COMPUTER ALGORITHM. The key feature in determining a suitable algorithm for editing using the pattern recognition scheme outlined above is answering the question: When is a point grossly in error? Erroneous points on a graph stand out only if they are far outside of the normal spread of the data. In the algorithm described above, editing is dependent upon identifying anomalous values of $\Delta x/\Delta t$ and $\Delta^2 x/\Delta t^2$. The test used is if $\Delta x/\Delta t$ or $\Delta^2 x/\Delta t^2$ is greater than 2.5-3.5 times the previously established values, then the derivative magnitude is excessive. The choice of this constant multiplier is necessarily arbitrary, but dependent upon the level of noise present in the bivariate data stream.

If an isolated point is identified as being bad the program user can either drop the point out of the file, or it may be replaced. When the replacement option is exercised, a quadratic equation is fit to eight points surrounding the point in error and a corrected value calculated. However, among the eight points, if more than two have been identified as bad points themselves, the fit is not made and the point in question is dropped from the file. A sequence of erroneous points, once identified is dropped from the file.

TESTING OF THE ALGORITHM. Noisy data were generated from two analytical equations in the following manner. A starting value of the independent variable was decided upon and a standard increment size. A uniform distribution random number generator then was used to calculate a fraction of the increment size to be used and either

$$t_{i+1} = t_i + f_i \cdot \Delta t \qquad (4)$$

or

$$t_{i+1} = t_i + \Delta t + f_1 \cdot \Delta t \qquad (5)$$

was used to calculate a new value of the independent variable. A second random number was generated and the ratio

$$A = f_1/f_2; \quad \text{if } f_1<.5, A = -A \tag{6}$$

was used to generate the noisy dependent variable from the relation

$$x_i = F(t_i) + A \cdot \left.\frac{dF}{dt}\right|_{t=t_i} \tag{7}$$

where $F(t_i)$ is the bivariate functional form being used. The two analytical equations used were

$$x_i = 100. - 10.2t_i - 0.5t_i^2 \tag{8}$$

and

$$x_i = 0.1t_i \sin(t_i/\pi) \tag{9}$$

Equation (8) was used as a crude simulation of a terminal trajectory record, while equation (9) was of interest to study the effects of low order oscillations on the algorithm.

The noisy input data from equations (4) and (8) are shown in Figure 2 and the edited data in Figure 3. Note there is a difference of an order of magnitude in the x_i scale which accounts for the apparent increased spread of the data. Similar results from equations (5) and (8) are shown in Figures 4 and 5. As is evident from Figures 6 and 7, generated by equations (5) and (9), oscillations in the data file do degrade performance somewhat. Overall results are encouraging, however, because, as is evident, the algorithm does identify and remove most of the gross points. Data files, once edited may then be passed on to more refined smoothing routines to eliminate inherent noise.

CONCLUSIONS. Grossly erroneous points in a varyingly spaced, bivariate data file may be identified and either corrected or removed using the algorithm described in this paper. Data screened in this manner are then suitable for further smoothing and/or processing.
Note: Copies of the algorithm described here in FORTRAN IV are available upon request from

Commander
Yuma Proving Ground
STEYP-MTS (C., Data Reduction Section)
Yuma, Arizona 85364

Table 1: Patterns of Anomalous Differences in a Bivariate Data Stream

Data No.	t	x	$\frac{\Delta x}{\Delta t}$	$\frac{\Delta^2 x}{\Delta t^2}$
9	1.0	- 9.5	- 6.5	497.5
10	1.2	-10.8	-106.0	-987.5
11	1.4	-32.0	91.5	-420.0
12	1.6	-13.7	- 7.5	- 2.5
.	.	.	.	.
.	.	.	.	.
18	2.8	-22.7	- 7.5	- 7.5
19	3.0	-24.2	- 9.0	245.
20	3.2	-26.0	40.	-250.
21	3.4	-18.0	-10.	- 50.
22	3.6	-20.0	-20.	-227.5
23	3.8	-24.0	-65.5	260.
24	4.0	-37.1	-13.5	- 7.5
.	.	.	.	.
.	.	.	.	.

Figure 1: Generated Patterns in First and Second Differences from Erroneous Data

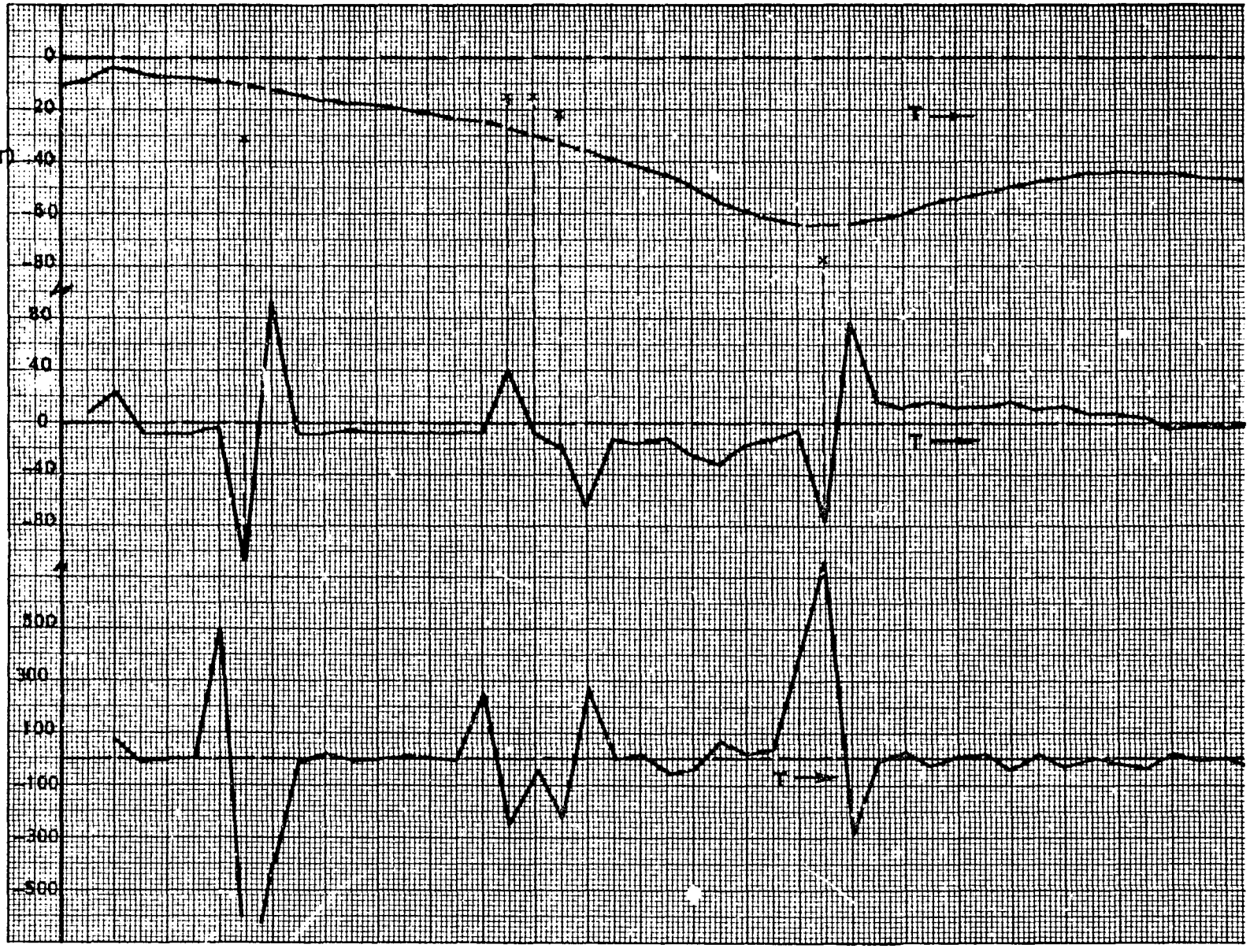
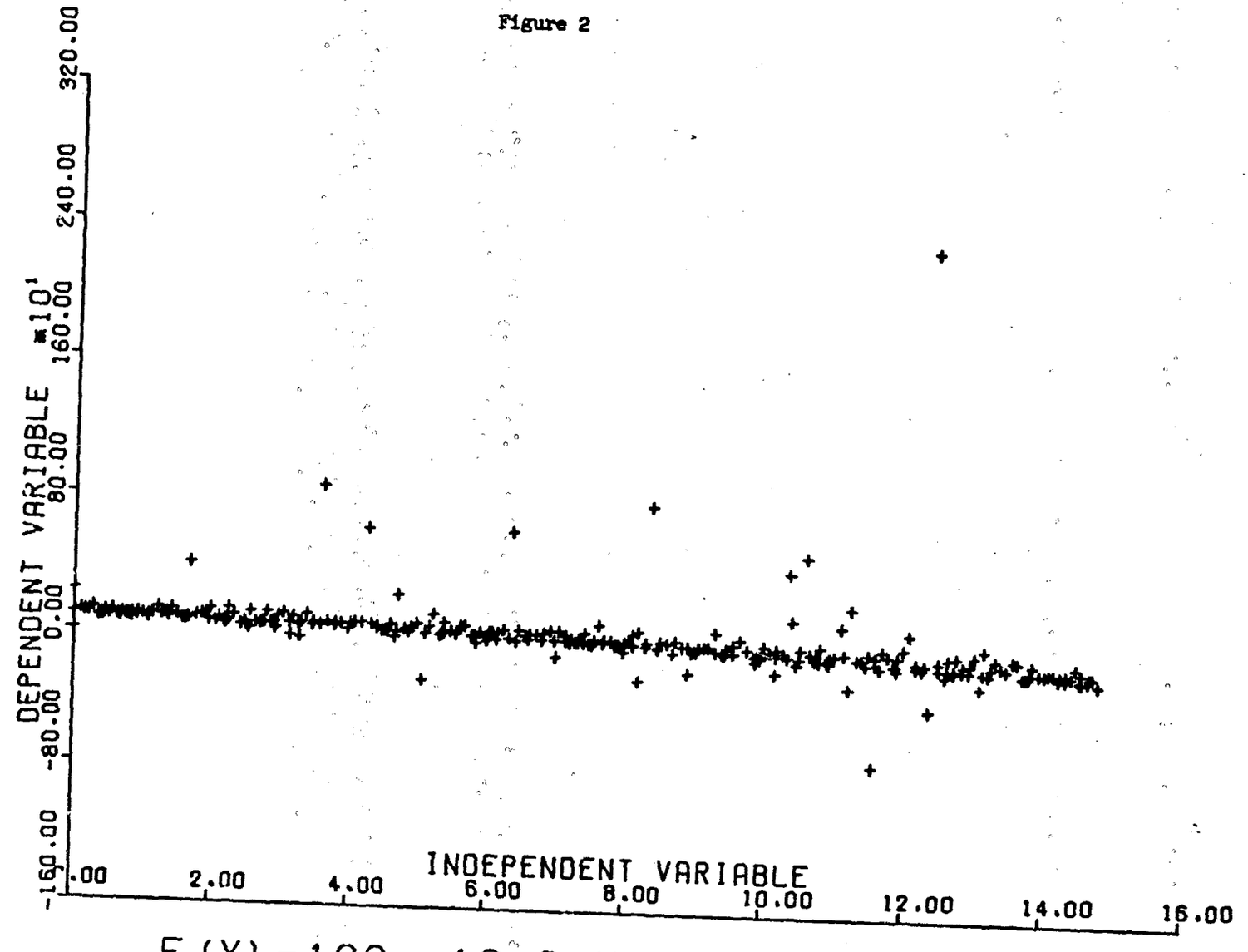


Figure 2



F(X) = 100. - 10.2*X - 0.5*X*X, EQN. 4 UNEDIT

Figure 3

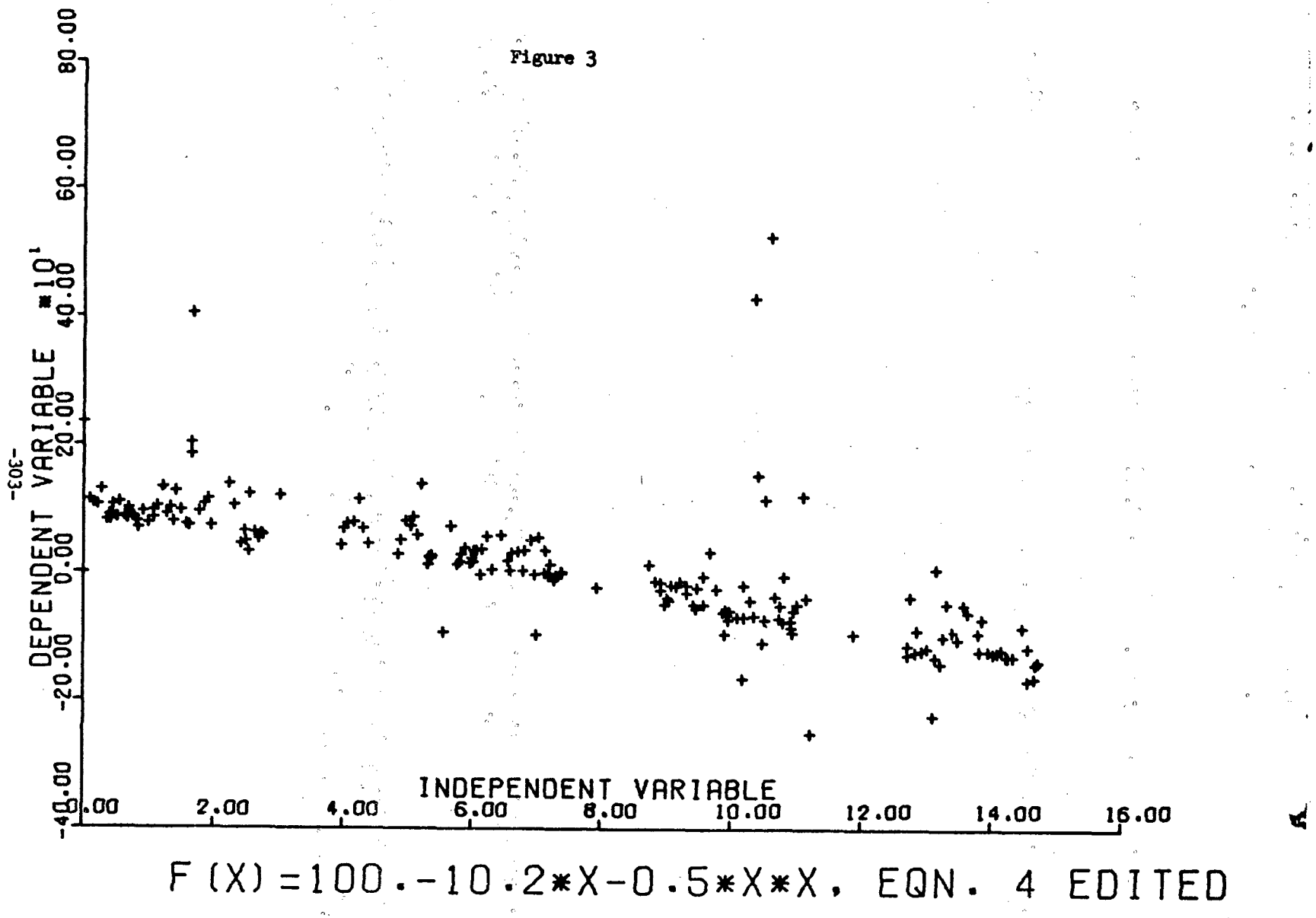

DEPENDENT VARIABLE
-303-
*10¹
80.00
60.00
40.00
20.00
0.00
-20.00
-40.00
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
INDEPENDENT VARIABLE
F(X)=100.-10.2*X-0.5*X*X, EQN. 4 EDITED

Figure 4

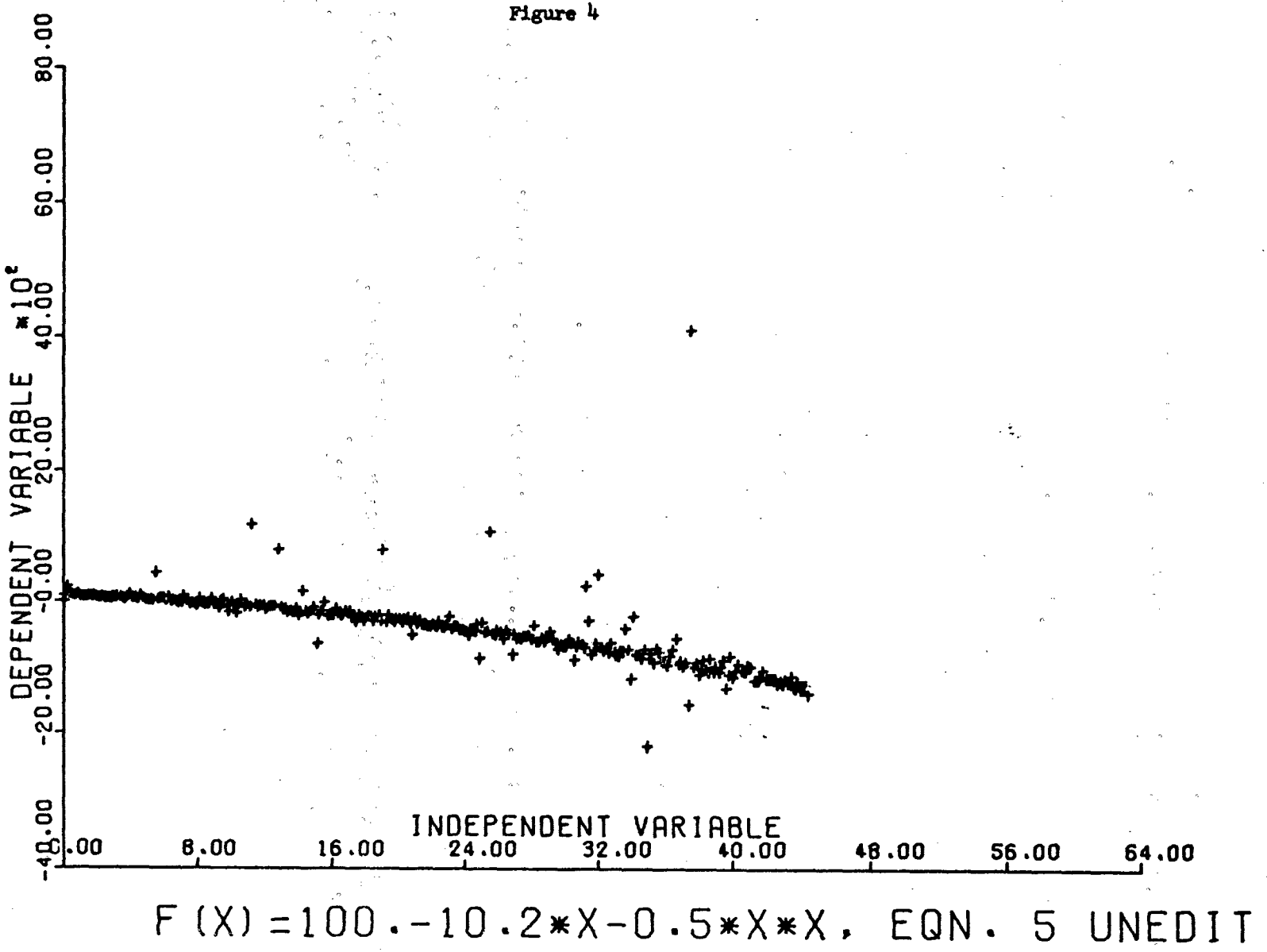
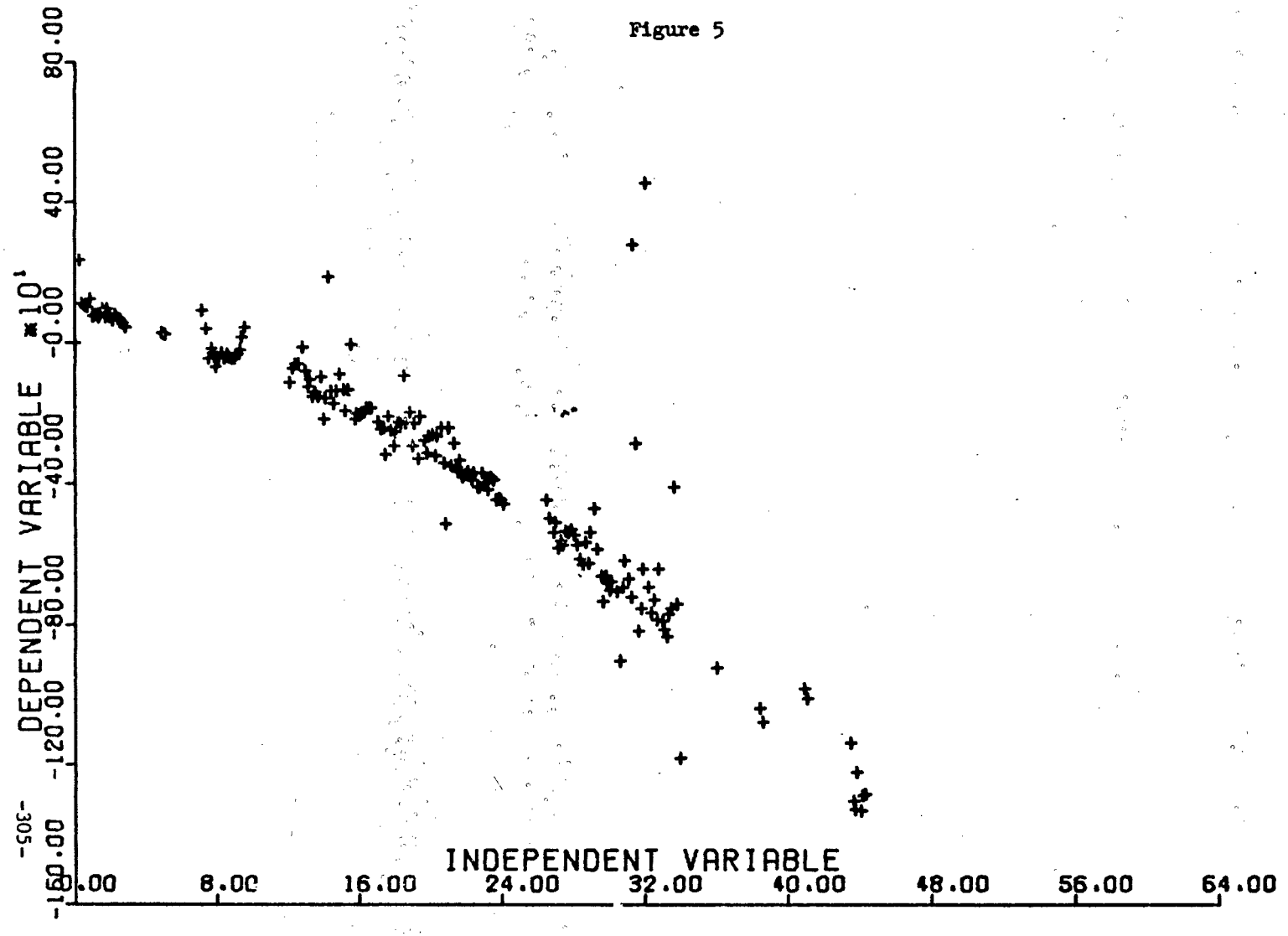
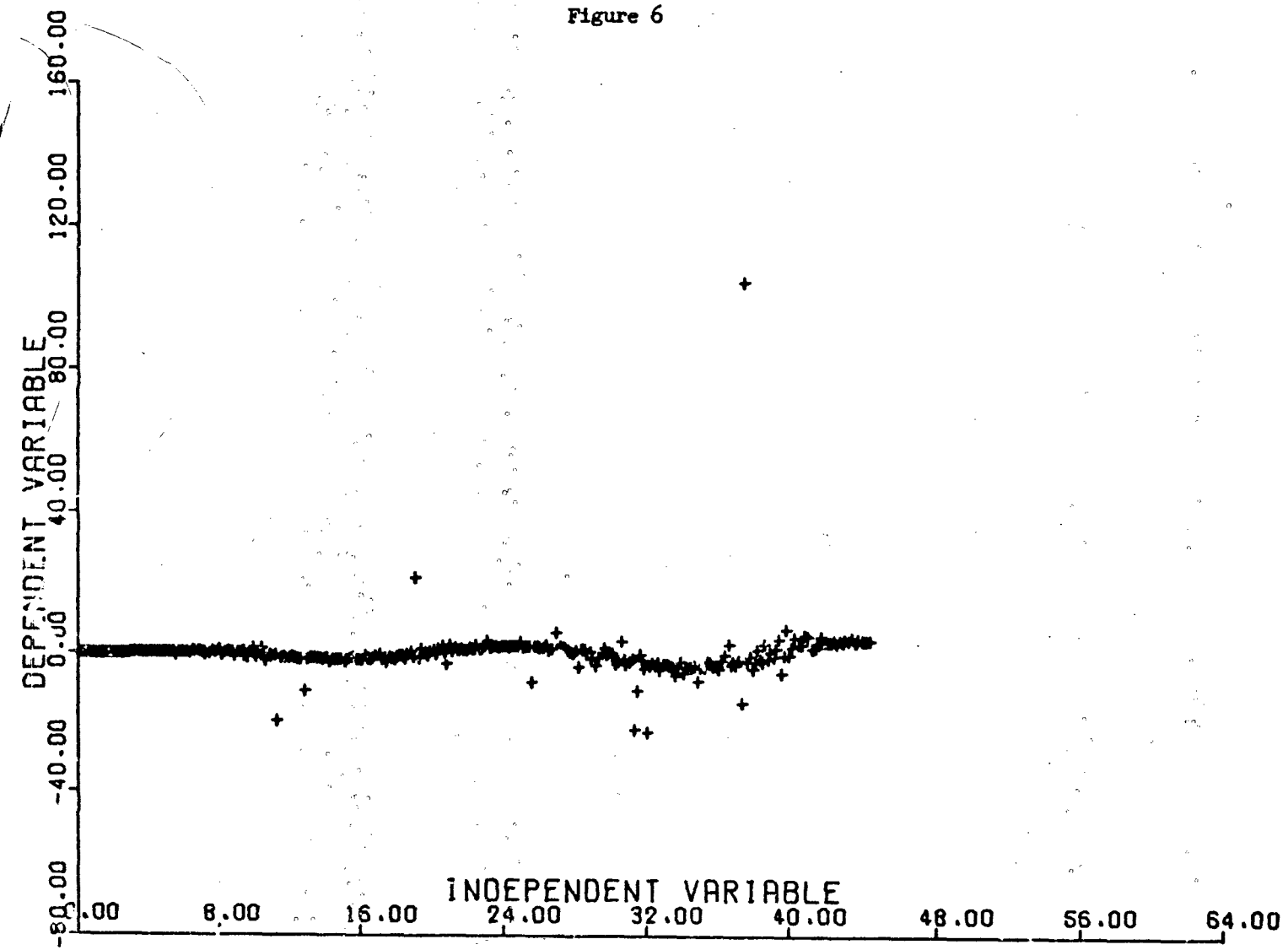

80.00
60.00
40.00
20.00
-0.00
-20.00
-40.00
DEPENDENT VARIABLE *10**2
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
64.00
INDEPENDENT VARIABLE
F(X)=100.-10.2*X-0.5*X*X, EQN. 5 UNEDIT

Figure 5



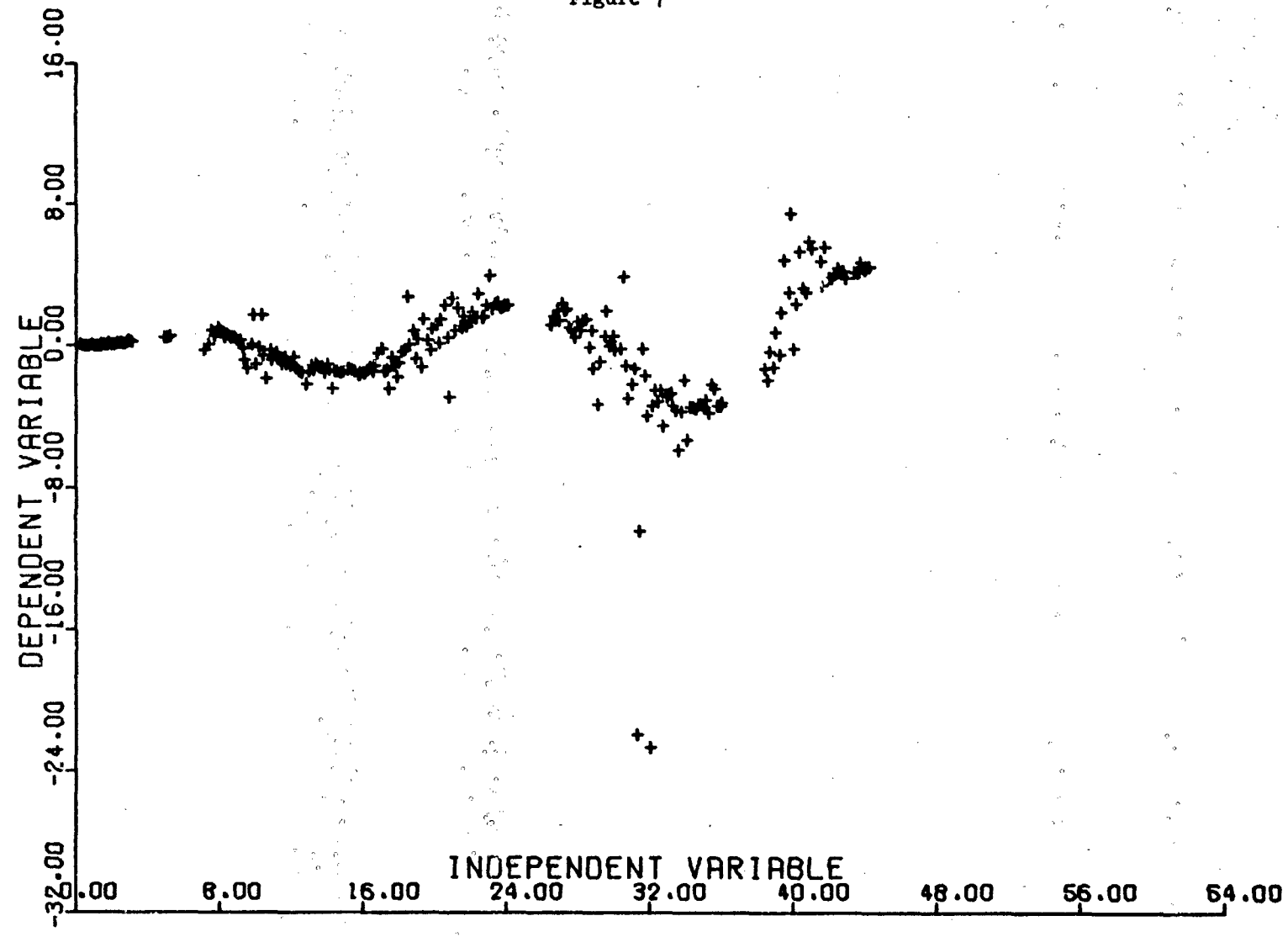
F(X)=100.-10.2*X-0.5*X*X, EQN. 5 EDITED

Figure 6



F(X)=0.1*X*SIN(X/3.14159), UNEDITED

Figure 7



F(X)=0.1*X*SIN(X/3.14159), EDITED

STATISTICAL ANALYSIS OF H. F. OBLIQUE AND VERTICAL INCIDENCE IONOSPHERIC DATA APPLICABLE TO FIELD ARMY DISTANCES

Richard J. D'Accardi
Chris P. Tsokos*

U. S. Army Electronics Command
Fort Monmouth, New Jersey

ABSTRACT. The object of this paper is to present a statistical approach for the analysis and interpretation of short-path oblique incidence and vertical incidence ionospheric soundings over typical field army distances.

Univariate spectral analysis is performed on the non-stationary stochastic realization of the oblique and vertical incidence data taken over the 60 Km path between Fort Monmouth, New Jersey, and Fort Dix, New Jersey. Estimates of the power spectrum are obtained using three "lag windows," namely, those of Bartlett, Tukey, and Parzen, respectively. A specific truncation length has been obtained for each of the above windows so that, regardless of which one is utilized, the same approximate estimate of the power spectrum will be obtained. In addition, bivariate spectral analysis of the vertical and oblique incidence data is given.

---

* Department of Mathematics, University of South Florida, Tampa, Florida

1. INTRODUCTION

The deployment of a Field Army necessitates many means of communication, especially those not limited by line-of-sight, extended distances, and intervening terrain obstacles. High frequency (HF) communications systems are not so stringently limited, and provide excellent back-up for higher frequency and high density systems. It, too, has limitations, foremost of which is the ionosphere, a medium which is time variant, random, and to say the least, highly unstable. Although much attention has been given to developing and fielding superior equipment, far less attention has been given to improving the use of the propagation media. It is in this regard that the Communications/ADP Laboratory, of the U. S. Army Electronics Command, has sponsored three experiments aimed at developing media-system parameters to provide tactical (HF) communicators with propagation predictions, in near real-time, for a typical Field Army area of influence. Specifically, both vertical incidence and short-path oblique incidence ionosonde data were taken over 60 Km, 200 Km, and 500 Km paths and analysed.

Work by D'Accardi-Tsokos-Kulinyi [1971] was the first in dealing with short-path ionospheric data as a stochastic realization as opposed to analysis and forecasting on the basis of specific "blocks" of time of day, Ames-Egan [1958], each being considered as independent and homogeneous. Their results introduced a new statistical concept to the estimation of short-path oblique incidence (OI) ionospheric data, and provided statistical models to forecast either oblique or vertical incidence soundings over specific paths. With respect to their first objective, regression techniques were used to relate vertical incidence (VI) soundings to OI soundings as the first part of the forecasting problem. This was a practical alternative to the widely accepted secant φ law, which, due to the lack of mid-path data, to the assumption of a stratified ionosphere, and to the difficulty in scaling virtual height at the critical frequencies, yielded poor results when applied over the longer path experiments. With regard to their second objective, the actual forecasting, they have shown that both the vertical and oblique soundings are non-stationary stochastic realizations; that is, they form a discrete time series that is not in statistical equilibrium. Their data was characterized by autoregressive, moving averages, and combination processes. This approach has pointed out that more information can be obtained from the data with respect to the development of system parameters, Krause et al [1970]. As a continuation of this effort, the aim of this paper is to utilize the information gained by D'Accardi, Kulinyi, and Tsokos in the analysis of the power spectrum. That is, to describe in detail how the variance of the stochastic realization (non-stationary ionospheric data) is distributed with frequency.

In section 3, we shall give some basic concepts of time series analysis, defining stationary and non-stationary processes. The procedure for forecasting is also given, including the difference equation for forecasting the average oblique incidence soundings for the 60 Km experiment. The model is used in the spectral analysis of the oblique incidence series.

Section 4 is devoted to basic concepts and a systematic procedure for the spectral analysis of OI soundings. The theoretical spectrum, $\Gamma_{yy}(f)$, is shown for the general discrete autoregressive (ar) process, and it is shown specifically for a third order ar process. Estimates of the spectral density

function using the lag windows of Bartlett, Tukey, and Parzen are given including the criteria for choosing the best one for estimating $\Gamma_{yy}(f)$. For various truncation lengths, L, the bandwidth, 95% confidence intervals, and degrees of freedom are given for each window. The bivariate behavior of the oblique and vertical incidence ionospheric soundings for the 60 Km experiment is discussed in section 5. More specifically, we obtain estimates of the smoothed coquadrature, phase, and cross-amplitude spectra using all three lag windows. Estimates of the coherency spectrum are also given.

2. DESIGN OF THE EXPERIMENT

The experimental distances of 60 Km, 200 Km, and 500 Km were chosen to fall within the idealized 300 x 300 kilometer tactical Field Army area of responsibility. The diagonal distance of such an area is approximately 440 Km, and represents the longest distance of an internal communications path. With Fort Monmouth, N. J., as a base station, mobile ionosonde terminals were set up at Fort Dix, N. J. (60 Km), Aberdeen Proving Ground, Md. (200 Km), and at Camp Drum, N. Y. (500 Km), as shown in Figure 1. The analysis presented herein, is based upon the 60 Km path data, but similar results were achieved over the other paths and will be presented at a later time.

Each ionosonde terminal of the 60 Km path, operating in the 2-16 MHz range, made scheduled soundings every 10 minutes for eighteen days over a twenty-one day period. While the Fort Monmouth terminal was transmitting and receiving its own signal, the Fort Dix terminal would simultaneously receive the same transmission; likewise for the Fort Dix to Fort Monmouth station (see Figure 2). Both ionosndes* were synchronized so that the "remote" sounder scans would be precise with the fixed station. The number of days that the experiment was performed has so significance with respect to the results obtained, but was a matter of funding.

The frequency range of the sounders spanned three "octaves," 2-4 MHz, 4-8 MHz, and 8-16 MHz, each of which contained 400 discrete channels. Transmissions consisted of successively "stepping" through the channels of each octave with 100 μs pulses. The resulting data is a recording of these pulses, as they return from the ionosphere, parametric in frequency and time delay. The time delay is a measure of virtual height of reflection from the ionospheric layers. Figure 3 shows an ionogram record of frequency vs. time delay. These records were taken on 35 mm. film at Fort Monmouth and on light-sensitive oscillograph paper at the remote terminal. After collection and development, the ionograms were scaled for the extraordinary critical frequencies, f_xF_2 (see Figure 3), and the resulting data was compiled for computer analysis. Some data (ionograms) were unreadable due to man-made noise, and solar and geomagnetic activity. For those records that were unreadable (though signal was detected), simulated data was prepared. The occurence of obscured data was negligible over the experiments.

*These instruments were Granger Associates Model 3905-5 systems, matched with wide-response delta antennas.

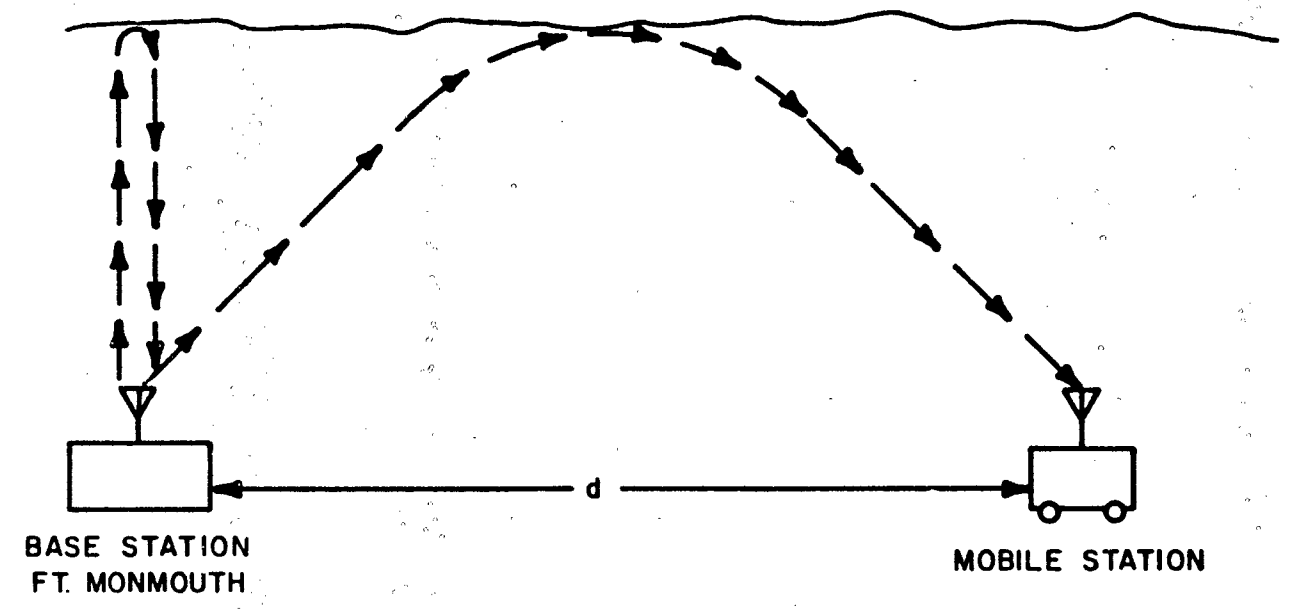


Figure 1. FIELD TESTS FOR A NEAR-REAL TIME IONOSPHERIC FORCASTING SCHEME

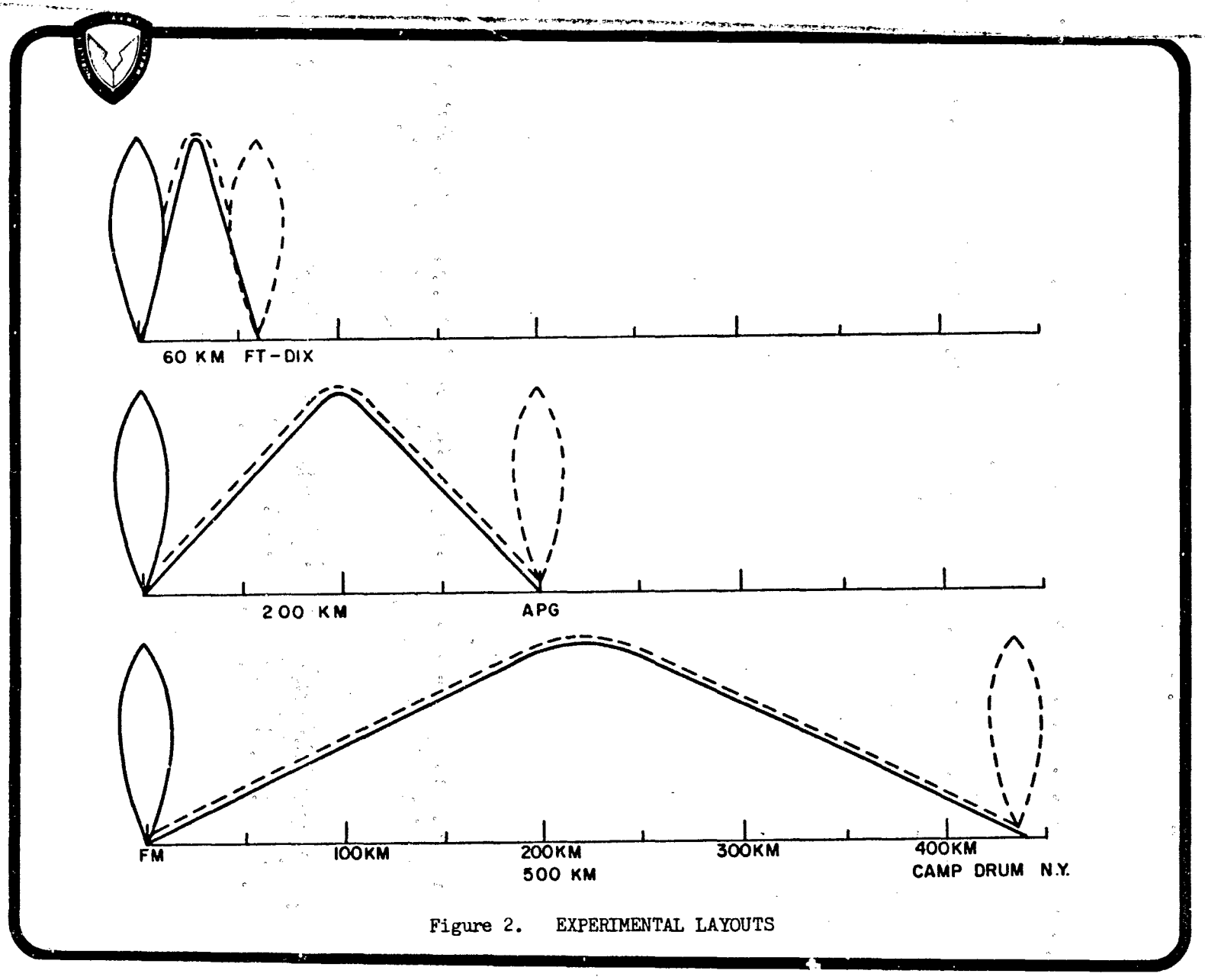


Figure 2. EXPERIMENTAL LAYOUTS

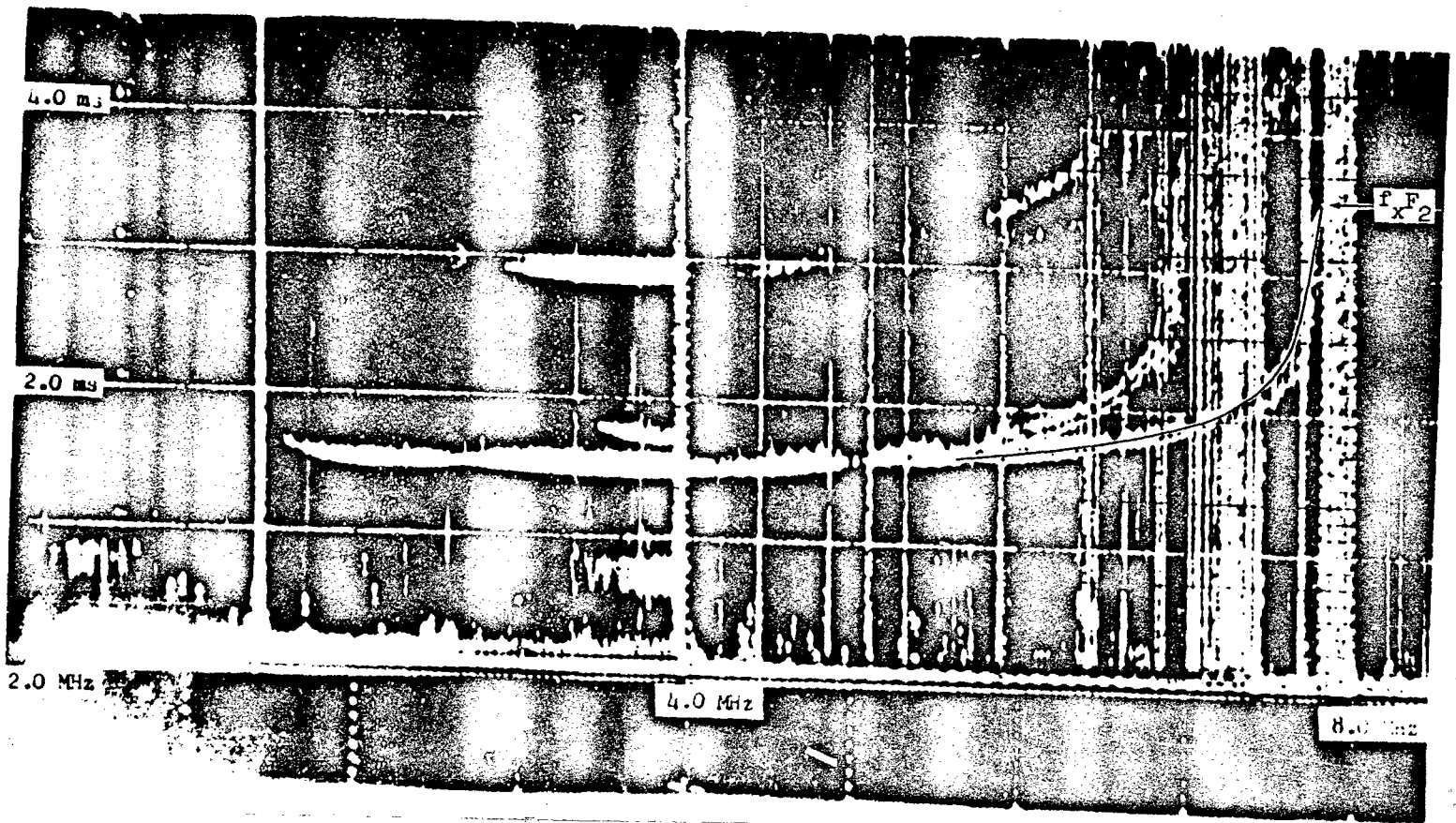


Figure 3. Vertical Incidence Ionogram. Ionosphere Near Normal Two Hours After Spread F Condition. Taken at Ft. Monmouth, N. J. - 0842 Hours EST, 28 January 1969.

3. BASIC PROCEDURE FOR FORMULATING THE FORECASTING MODEL

3.1 Basic Concepts in Time Series Analysis:

Any phenomenon such as the oblique incidence soundings, which changes with time, and any collection of data that measures the aspects of such a phenomenon can be considered as a time series. A time series can either be a deterministic function or a non-deterministic function of an independent variable, usually time; but, in most physical situations, it will be a non-deterministic function. A non-deterministic function exhibits random or fluctuating properties and, hence, it is not possible to forecast its future values exactly. Thus, a non-deterministic time series can only be described by statistical laws or models. We begin by assuming that one can describe a time series at a given time, t, by a random variable and its associated probability distribution function. In this manner, we may describe at all instances, the behavior of a time series by an ordered set of random variables and the associated probability distributions. Such an ordered set of random variables is called a stochastic process. Thus, an observed time series, y_t, can be considered as one realization of an infinite ensemble of functions that might have been generated by a stochastic process. Such a process is said to be strictly stationary, is a joint probability distribution of any set of observations, and is not affected by shifting all times of the observations ahead or backward by any integer amount, k. A stationary stochastic process can be described in terms of its mean, μ, which is estimated by:

$$\bar{y} = \frac{1}{n} \sum_{t=1}^{n} y_t \ , \qquad (3.1.1)$$

its variance, σ^2, which is estimated by:

$$s_y^2 = \frac{1}{n} \sum_{t=1}^{n} (y_t - \bar{y})^2 \ , \qquad (3.1.2)$$

its sample autocovariance function, which measures the extent to which two random variables are linearly independent, estimated by:

$$c_{yy}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y}) \ , \qquad (3.1.3)$$

where $k = 0, 1, 2, \ldots, n-1$,

and its sample autocorrelation function, which acts like a correlation coefficient and is estimated by:

$$r_{yy}(k) = \frac{c_{yy}(k)}{c_{yy}(0)} \ , \quad k = 0, 1, \ldots, n-1 \qquad (3.1.4)$$

3.2 Stationary vs. Non-Stationary Time Series:

A stationary time series is a series that is in statistical equilibrium in the sense that its properties do not change significantly with respect to time. A non-stationary time series is such that its properties change with time. The information with which the present study is concerned, namely, ionospheric soundings, is non-stationary in nature. In general,

non-stationary phenomena can be divided into three basic classes:

(a) Those time series that exhibit stationary properties over a long period of time.

(b) Those that are approximately stationary over short periods of time.

(c) Those time series that exhibit non-stationary properties; that is, their visual properties change continuously with respect to time.

In the present state-of-the-art, there exist techniques to analyze stationary time series information; however, the techniques available for the analysis and interpretation of non-stationary time series information, such as ionospheric data, are inadequate and do not lend themselves to meaningful interpretations of physical situations. It is possible, however, to adjust non-stationary time series so as to be able to apply the existing techniques of stationary time series analysis. This adjustment takes the form of applying a proper filter to the observed non-stationary time series to filter out the non-stationary components.

The search for a mathematical function to transform the non-stationary time series into a stationary series is in some respects a trial-and-error procedure. One of the most popular and most efficient methods of accomplishing this purpose is the application of a difference equation (see Jenkins and Watts [1968]; Box and Jenkins [1970], among others). A first order difference equation is defined by:

$$x_t = y_t - y_{t-1} \,, \tag{3.2.1}$$

where:

y_t = observed non-stationary series,

and:

x_t = the first difference series.

Similarly, a second order difference equation is defined by:

$$w_t = y_t - 2y_{t-1} + y_{t-2} \,, \tag{3.2.2}$$

and so on. In practice, a first or second order difference equation is usually sufficient to transform most non-stationary time series.

To identify whether or not the observed time series exhibits stationary or non-stationary properties, we make use of the following three basic concepts:

(a) Visual interpretation of the series.

(b) A plot of the sample autocorrelation function of the observed series.

(c) Application of various trend tests to the observed series.

The graphical representation of the observed series can be of practical help. However, for a more rigorous classification of the series, we must rely on the latter two concepts. For the observed series and its first and second differences, one computes the sample autocorrelation function using equation 3.1.4, and conducts trend tests, such as Kendall's tau, Kendall and Stewart [1966]. The sample autocorrelation function of a stationary phenomenon has the basic property that it dampens out fairly rapidly; that is, it approaches zero. Also, a stationary series will be such that it contains no trend. Following this procedure, one can obtain sufficient information to determine if the observed series exhibits stationary or non-stationary components; and if it exhibits non-stationary components, whether or not a first or second order difference equation would filter them out.

Having reduced the given information to a stationary time series, our aim is to fit a parametric model to this series, either an autoregressive, a moving averages, or a combination of the two. These stationary stochastic models assume that the process (series) remains in equilibrium about a constant mean level and they are of great value in modeling stationary time series. The general autoregressive process is given by:

$$Y_t - \mu = \alpha_1 (Y_{t-1} - \mu) + \ldots + \alpha_m (Y_{t-m} - \mu) + Z_t \ , \qquad (3.2.3)$$

where μ is the mean of Y_t, Z_t is a purely random process (Jenkins and Watts [1968], and m is the order of the process. The general moving average process is given by:

$$Y_t - \mu = Z_t - \beta_1 Z_{t-1} - \cdots - \beta_q Z_{t-q} \ , \qquad (3.2.4)$$

where μ and Z_t are as defined above, and q is the order of the process. The general mixed autoregressive-moving average process is given by:

$$Y_t - \mu = \alpha_1 (Y_{t-1} - \mu) + \ldots + \alpha_m (Y_{t-m} - \mu) + Z_t - \beta_1 Z_{t-1} - \cdots - \beta_q Z_{t-q} \ ,$$

where q is independent of m. (3.2.5)

In a recent paper, D'Accardi, Kulinyi, and Tsokos [1971], working with ionospheric information, have outlined a procedural approach for fitting the above models. They discuss in detail the criteria for selecting the process, its order (which gives the best fit to the observed series), the procedure to estimate its parameters, diagnostic check of goodness of fit, and how the model can be employed in forecasting ionospheric soundings.

3.3 Forecasting Model for the Average of the Oblique Incidence Soundings for the 60 Km Path:

Following the above procedure, we have fitted a third order autoregressive model to the average oblique incidence soundings of the 60 Km path series. The model which best characterized this series is given by the following difference equation:

$$\bar{y}_t = -0.5987\, \bar{y}_{t-1} - 0.2034\, \bar{y}_{t-2} + 0.1608\, \bar{y}_{t-3} \qquad (3.3.1)$$

In formulating this model, we utilized a second order difference filter, given by equation (3.2.2), for t = 3,...,85, and the estimates for the above

model were found to be:

$$\hat{a}_1 = -0.59867$$
$$\hat{a}_2 = -0.20345$$
$$\hat{a}_3 = -0.16082$$

Figure 4 gives a graphical display of the average of the original oblique incidence soundings of the 60 Km experiment, and the predicted values of the series. Note that we began predicting the average oblique soundings after having observed the first four observations, and utilizing this information, we continued to forecast until a difference of 0.5 units occurred between actual and predicted information. Even though our last observed average oblique soundings was recorded at time slot 85, we continued predicting up to time slot 99.

The above model, equation (3.3.1), will be used in a later section where we will be concerned with the spectral analysis of the oblique ionospheric series.

4. SOME BASIC CONCEPTS OF THE SPECTRAL ANALYSIS OF OBLIQUE INCIDENCE SOUNDINGS

In this section, we shall present a spectral analysis of the oblique incidence soundings of the 60 Km experiment. We shall be using some of the basic concepts of spectral analysis given by Jenkins and Watts [1968], and Box and Jenkins [1971].

4.1 The Theoretical Spectrum:

The fou[illegible]sform of the autocovariance function is called the power spectrum or [illegible]ectrum of the time series, and its plot shows how the variance of the stochastic realization is distributed with respect to frequency (time). The theoretical spectrum, denoted by $\Gamma_{yy}(f)$, can be written by the following equation:

$$\Gamma_{yy}(f) = \Delta\sigma_z^2 |H(f)|^2, -\frac{1}{2\Delta} \le f \le \frac{1}{2\Delta} , \qquad (4.1.1)$$

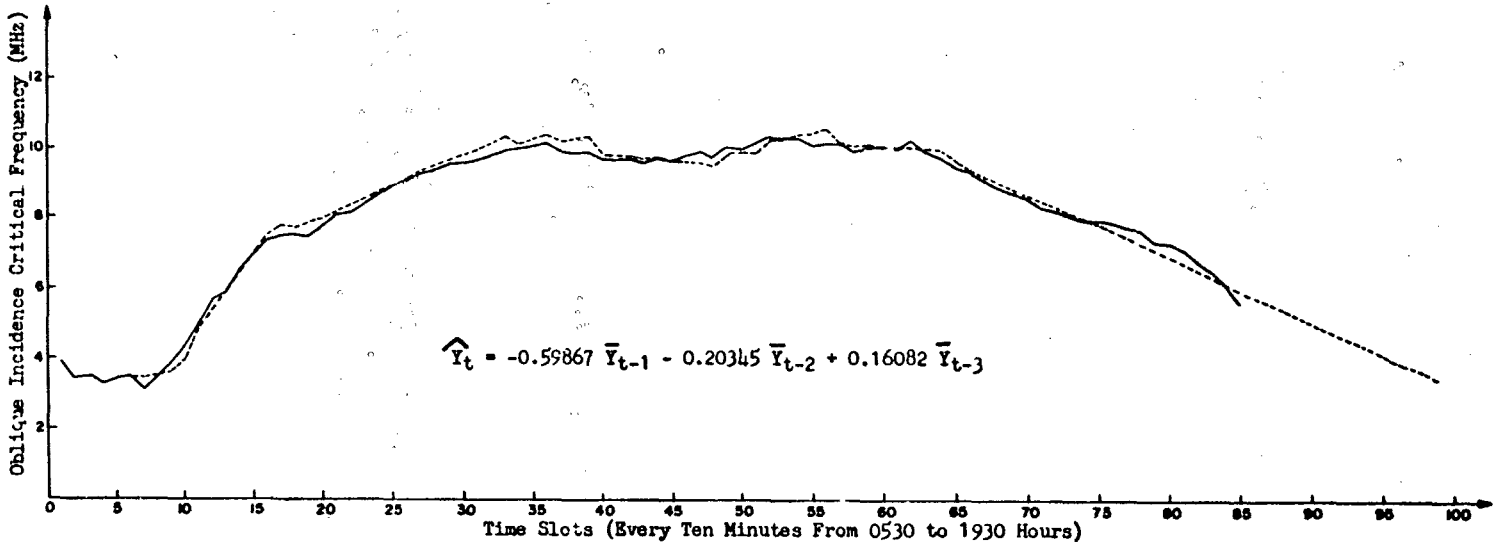
where:

$H(f)$ = frequency response function,
σ_z^2 = the variance of a purely random process,
Δ = incrementation interval between observations.

For the general discrete autoregressive process, equation (3.1.1), the theoretical spectrum is given by:

$$\Gamma_{yy}(f) = \Delta\sigma_z^2 \left| \frac{1}{1-\alpha_1 e^{-j2\pi f\Delta} - \ldots - \alpha_m e^{-j2\pi f\Delta^m}} \right| \qquad (4.1.2)$$

For the oblique incidence soundings, we have formulated a third order (m=3) autoregressive process. Thus, equation (4.1.1) can be written as follows:



Legend: ——— Oberserved Series
- - - Predicted Swries

Figure 4. PIECEWISE FIT OF THE AUTOREGRESSIVE FORECASTING MODEL, Y_t, TO ORIGINAL SHORT-PATH OBLIQUE INCIDENCE DATA FOR THE 60 KM PATH

$$|H(f)|^2 = \frac{1}{1+\alpha_1^2+\alpha_2^2+\alpha_3^2+(2\alpha_2\alpha_3-2\alpha_1+2\alpha_1\alpha_2)\cos 2\pi f+(2\alpha_1\alpha_3-2\alpha_2)\cos 4\pi f - 2\alpha_3\cos 6\pi f|} \qquad (4.1.3)$$

Substituting $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\alpha}_3$ for α_1, α_2, and α_3 respectively, and obtaining a σ_z^2 estimate (assuming $\Delta=1$), we obtained an estimate of the theoretical spectral density function,* $\Gamma_{yy}(f)/\sigma_z^2$ for the filtered process Y_t = 1,...,85, as seen from equation (3.3.1). From the process, we obtain:

$$\begin{aligned}\sigma_z^2 &= \mathrm{Var}(Z_t) = \mathrm{Var}(Y_t-\alpha_1 Y_{t-1}-\alpha_2 Y_{t-2}-\alpha_3 Y_{t-3})\\ &= \sigma_y^2+\alpha_1^2\sigma_y^2+\alpha_2^2\sigma_y^2+\alpha_3^2\sigma_y^2-2\alpha_1\mathrm{Cov}(Y_t,Y_{t-1})-2\alpha_2\mathrm{Cov}(Y_t,Y_{t-2})\\ &\quad -2\alpha_3\mathrm{Cov}(Y_t,Y_{t-3})+2\alpha_1\alpha_2\mathrm{Cov}(Y_t,Y_{t-1})+2\alpha_2\alpha_3\mathrm{Cov}(Y_t,Y_{t-1})\\ &\quad +2\alpha_1\alpha_3\mathrm{Cov}(Y_t,Y_{t-2})\\ &= \sigma_y^2\,[1+\alpha_1^2+\alpha_2^2+\alpha_3^2-2\alpha_1\rho_{yy}(1)-2\alpha_2\rho_{yy}(2)-2\alpha_3\rho_{yy}(3)+2\alpha_1\alpha_2\rho_{yy}(1)\\ &\quad +2\alpha_2\alpha_3\rho_{yy}(1)+2\alpha_1\alpha_3\rho_{yy}(2)] \qquad (4.1.4)\end{aligned}$$

Hence, the estimate of σ_z^2, substituting the α_i, i = 1,2,3, and $r_{yy}(k)$, the estimate of $\rho_{yy}(k)$ is given by:

$$\tilde{\sigma}_z^2 = \sigma_y^2(2.69387)$$

Therefore, the estimate of the theoretical spectral density of the autoregressive process fitted to the filtered data is:

$$\frac{\Gamma_{yy}(f)}{\sigma_y^2} = \frac{2.6938}{1.425 + 1.375\cos 2\pi f + 0.214\cos 4\pi f - 0.322\cos 6\pi f},$$

where: $-\frac{1}{2} \le f \le \frac{1}{2}$

(4.1.5)

A smoothed estimate of the theoretical spectral density function can be obtained using the following equation:

$$\bar{R}_{yy}(\ell) = 2\left\{1 + 2\sum_{k=1}^{L-1} r_{yy}(k)w(k)\cos\frac{\pi\ell k}{F}\right\}, \qquad (4.1.6)$$

ℓ = 0,1,...,F, where F = 2L, L being the truncation length, and w(k) is the lag window. In the above equation, the lag window plays a major role in obtaining a good estimate of the spectral density function. In practice,

*Frequently in practice, we have to compare time series with different scales of measurement. In order to do this, it is necessary to normalize the spectrum, that is, simply divide the theoretical spectrum by the variance of the process.

there are three basic windows that are commonly used, namely those of Bartlett, Tukey, and Parzen. We shall briefly define these windows, and for more specific details and the properties concerning them, see Jenkins and Watts [1968]. Bartlett's lag window is given by the following expression:

$$w_B(u) = \begin{cases} 1 - \frac{|u|}{M}, & u \leq M \\ 0 & \text{otherwise} \end{cases} \tag{4.1.7}$$

Tukey's lag window is given by:

$$w_T(u) = \begin{cases} \frac{1}{2}(1+\text{Cos}\,\frac{\pi u}{M}), & |u| \leq M \\ 0 & \text{otherwise} \end{cases} \tag{4.1.8}$$

Parzen's lag window is given by:

$$\begin{aligned} w_p(u) &= \begin{cases} 1-6\left(\frac{|u|}{M}\right)^2 + 6\left(\frac{|u|}{M}\right)^3, & |u| \leq \frac{M}{2} \end{cases} \\ &= \begin{cases} 2\left(1-\frac{|u|}{M}\right)^3, & \frac{M}{2} \leq |u| \leq M \end{cases} \\ &= \begin{cases} 0 & \text{otherwise} \end{cases} \end{aligned} \tag{4.1.9}$$

A rectangular window is another alternative not mentioned above, which is defined by:

$$w_R(u) = \begin{cases} 1, & |u| \leq m \\ 0 & \text{otherwise} \end{cases} \tag{4.1.10}$$

where M is a truncation point.

Scientists who have been involved in choosing the proper shape of a lag window, w(u), have taken into consideration the fact that the spectral window, W(f), that is, the fourier transform of the lag window, should be concentrated near the zero frequency. Blackman and Tukey [1959], looking at the problem from the communications engineering point of view, almost identified it with that of choosing the intensity distribution along an antenna, so that the variation will fall in a narrow beam. The principal maximum and the subsidiary extreme of W(f) are called, respectively, main and side lobes. A window should be an even function so that it can equally treat positive and negative values of the spectral density function on both sides of a given point of the time series. It should integrate to unity, that is,

$$\int_{-\infty}^{\infty} W(f)df = 1$$

and should achieve a maximum value at the frequency f = 0. That is,

$$|W(f)| \leq W(0), \text{ for all } f.$$

It should be concentrated as much as possible about f = 0 in order that the behavior of the spectral density function be reflected as much as possible in that neighborhood.

There has been no agreed valid criterion for comparing the degree of concentration of any window. One criterion could be the ratio of the size of the second largest peak to the size of the largest peak. However, again this would be powerful only in the case where the second largest peak would occur at the same point. This fact explains why one has to consider all the different windows, not the most popular, in one's search for the most appropriate case.

For the main lobe of $W(f)$ to be concentrated, the graph of $w(u)$ should be flat due to the way the two concepts are related. Also, for the side lobes to be small $w(u)$ should be smooth and should not change rapidly as in the case of the rectangular window. Therefore, one should compromise. The authors' analyses have been done along these lines, and this is the reason why we have, as a result, numerous windows among which we should choose.

Taking Bartlett's spectral window, $w_B(f)$, as an example, we find that when it is graphed against frequency, it is symmetric about the origin and has zeros at $f = \pm \frac{1}{M}, \pm \frac{2}{M}, \pm \frac{3}{M}, \ldots$.

We shall call base width the distance between the first zeros on either side of the origin. The base width for Bartlett's window is equal to $\frac{2}{M}$. It is inversely proportional to M and the variance. By increasing the base width, the bias, $B(f)$, increases as well. Thus, we are forced to compromise between bias and variance in choosing a particular window.

The rectangular window is more concentrated about the center frequency than any other of the windows under consideration. Nevertheless, although it has the smallest bandwidth, which implies small bias, it also has the largest side lobes. This makes it very impractical. The first side lobe is about 1/5 of the height of the main lobe which shows an unrealistic characterization of the estimate of the power spectrum.

Thus, in view of the above remarks, for the density function of the oblique incidence soundings, we shall utilize the Bartlett, Tukey, and Parzen lag windows in search of the best estimate of the spectral density function.

4.2 Estimate of the Spectral Density Function Using Bartlett's Lag Window:

The values of the estimate of the spectral density function using Bartlett's lag window, equation (4.1.7), were calculated and plotted versus frequency for L = 8, 12, 16, 20, 24, 28, and 32 units. As a basis of comparison, we plotted our estimate on the same set of axes for L = 8, 12, 16, 24, and 32. For these values of L, we calculated the bandwidth, the confidence intervals, and the degrees of freedom which are shown in the following table:

TABLE I: TRUNCATION POINT, BANDWIDTH, DEGREES OF FREEDOM, AND CONFIDENCE INTERVALS FOR BARTLETT'S LAG WINDOW

L	Bandwidth	d.f.	95% C.I.	$\bar{R}_{yy}(f)$
8	.188	31	.63	1.56
12	.125	20	.58	2.10
16	.094	15	.54	2.35
24	.063	10	.49	3.00
32	.047	7	.42	4.10

The formula used for the bandwidth of the estimate of the spectral density function is given by:

$$b = \frac{b_1}{L\Delta} = \frac{1.5}{L}$$

and the equation for the degrees of freedom is given by:

$$\nu = 2\left(\frac{T}{L}\right)b = 2Tb = 2(83)b = 166b.$$

Note that since we have chosen $\Delta = 1$, we have $L = M$.

Figure 5 gives a comparison of the theoretical spectral density function of the autoregressive process and its smoothed estimate for the various truncation points, along with the 95% confidence intervals.

It is a known fact that increasing the bandwidth of the estimate of the spectral density means increasing the amount of bias and decreasing the variance; thus, a compromise has to be reached as to the best value of L. In making such a decision, we should take into consideration the confidence interval, the degrees of freedom, and the visual appearance of the plot of our estimate. For L = 8, the plot is very smooth and has a shape very similar to the theoretical spectrum with the bandwidth being wide enough to conceal any peaks that may be present. By increasing L to 12, we obtain an indication of another peak that appears at f = 3/16 cycles per second, in addition to the major peak in our theoretical spectral density. The plot is still quite smooth and the bandwidth is wide enough to give a great deal of faith in our estimate. Increasing L to 16, the bandwidth seems to be in a very shaky range. However, the curve has changed very little from the one for which L = 12. It displays the peak in the theoretical spectral density and also the extra peak at f = 3/16 cycles per second. Since larger values of L produce many small erratic peaks, we chose L = 16 to estimate the spectral density using Bartlett's lag window.

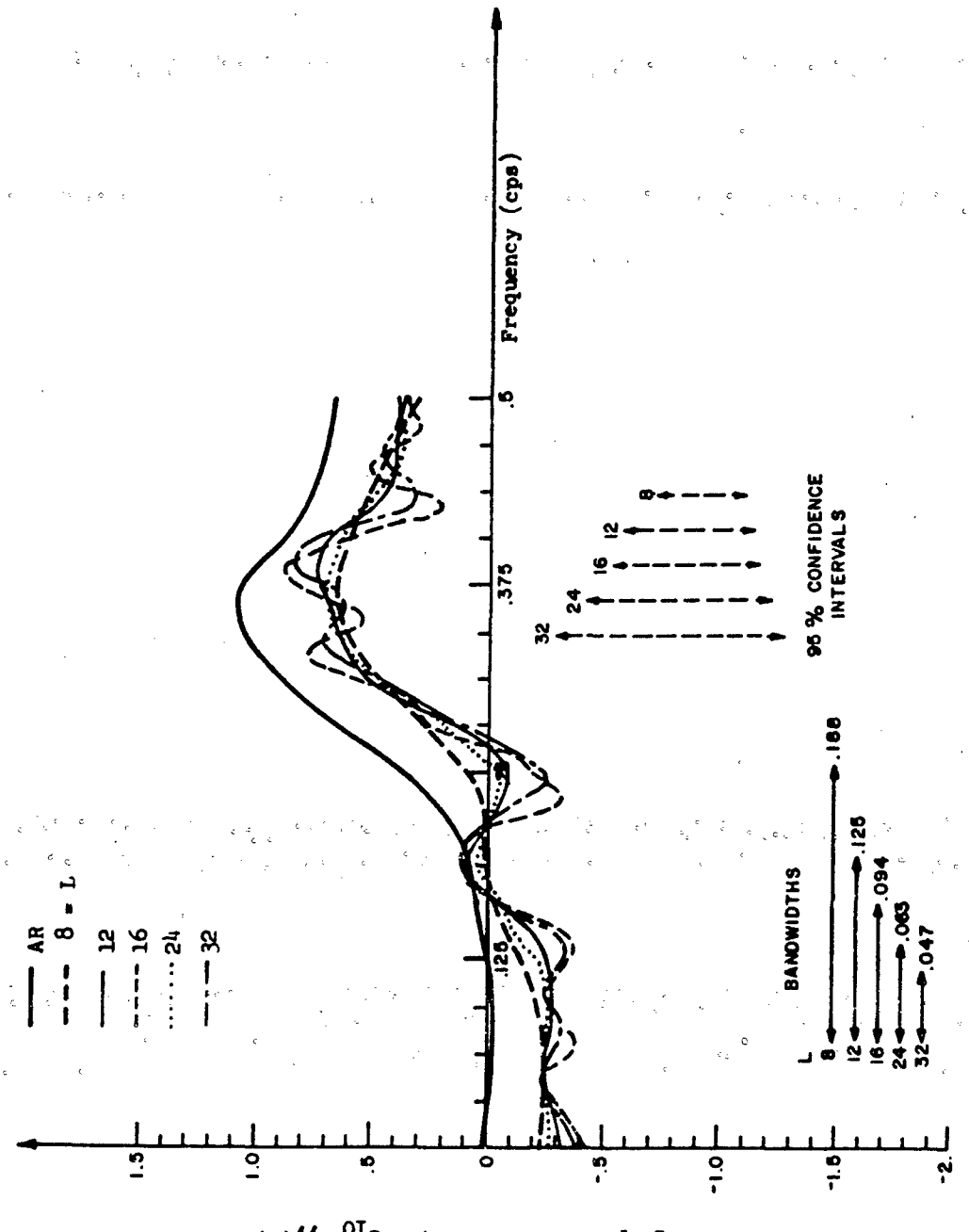


Figure 5. ESTIMATE OF THE SPECTRAL DENSITY USING THE BARTLETT LAG WINDOW

4.3 Estimate of the Spectral Density Function Using Tukey's Lag Window:

Using Tukey's lag window given by equation (4.1.8), the smooth spectral density estimate $\bar{R}_{yy}(f)$ was calculated for L = 8, 12, 16, 20, 24, 28, and 32 units. In Figure 6, we display the spectral density estimates along with an estimate of the theoretical spectrum of the third order autoregressive process. In addition, we display the 95% confidence intervals, and the bandwidths which correspond to the various truncation lengths. It is clear that for L = 8, the sample spectrum has the same general shape as the theoretical spectrum, and the curve is very smooth. Increasing the truncation value to 12, the plot is still fairly smooth, but a peak appears at about f = .186 cycles per second. At L = 16, the peak is slightly more pronounced, and as L is increased to 20 and above, more peaks appear at higher frequencies. This indicates that the variance is increasing and, thus, the sample spectrum is becoming more erratic for $L \geq 20$. On this basis, we decided that for L = 12, 14, or 16, we would try to obtain better estimates than those calculated for other values of L. We computed the spectral density estimates for L = 14, 18 units, respectively.

Table II displays, for the various truncation points, the bandwidth, degrees of freedom, and confidence intervals using Tukey's lag window.

TABLE II: TRUNCATION POINT, BANDWIDTH, DEGREES OF FREEDOM, AND CONFIDENCE INTERVALS FOR TUKEY'S LAG WINDOW

L	b	d.f.	95% C.I.	$\Gamma_{yy}(f)$
8	.166	27	.61	1.58
12	.112	18	.57	2.25
14	.095	16	.54	2.35
16	.083	13	.51	2.50
20	.067	11	.49	2.85
32	.049	6	.41	4.30

Table II is quite helpful in deciding that for L = 14 units, we will have the best estimate of the spectrum using Tukey's lag window. The degrees of freedom, $\nu = 15$, are sufficient for fairly small 95% confidence intervals, and this gave a bandwidth of .095 so that peaks in the time spectrum of bandwidths larger than .095 will be detected. Decreasing the bandwidth to .083, that is, L = 16, causes a loss of two degrees of freedom and a slight increase in the confidence interval width. For L = 12, the bandwidth is considerably larger (.112), and there is not much change in the confidence interval even though there are eighteen degees of freedom. Therefore, for a truncation length of 14 units, we obtain the best estimate for the spectrum using Tukey's lag window. Figure 7 shows the spectral density estimates of the filtered data using the Tukey lag window for truncation lengths L = 8, 12, 14, 16 and 32, along with the 95% confidence intervals and the various

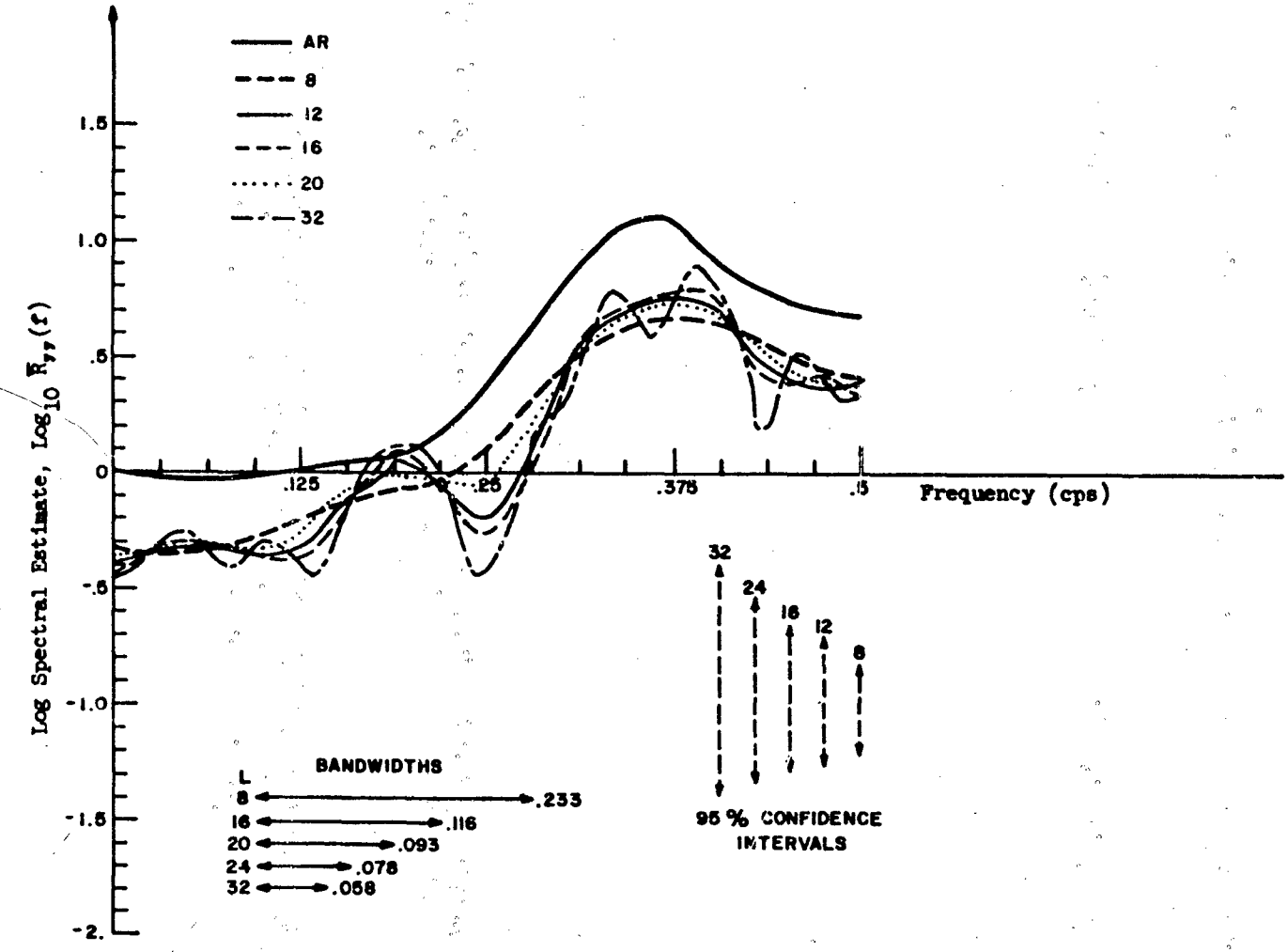


Figure 6. ESTIMATES OF THE SPECTRAL DENSITY USING THE TUKEY LAG WINDOW

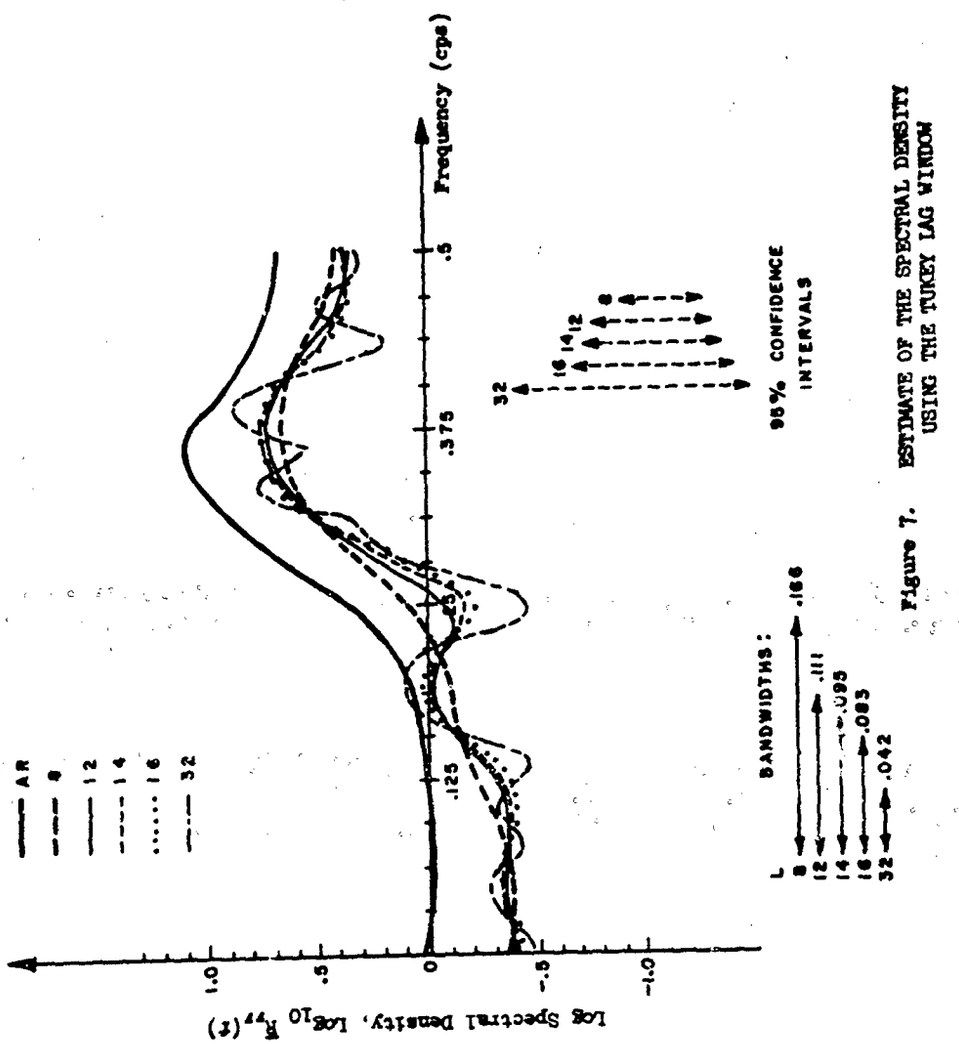


Figure 7. ESTIMATE OF THE SPECTRAL DENSITY USING THE TUKEY LAG WINDOW

bandwidths associated with these truncation points. As we mentioned previously, the plot of the estimate of the spectral density is given in the logarithmic scale to show more detail in the spectrum over a wider amplitude range.

4.4 Estimates of the Spectral Density Function Using Parzen's Lag Window:

Using Parzen's lag window, given by equation (4.1.9), we obtained estimates of the spectral density function for various truncation points. As before, we shall let $\Delta = 1$, so that $L = M$, the truncation points of the smoothed spectral estimator. We varied L from 8 to 32 in intervals of four units.

Figure 8 shows the spectral density estimates of the filtered data for the various truncation points along with the theoretical spectral density of the third order autoregressive process. In addition, a 95% confidence interval and the corresponding bandwidths are displayed. The bandwidth using Parzen's lag window is given by:

$$b = \frac{1.86}{L\Delta} = \frac{1.86}{L}$$

The degrees of freedom for the confidence intervals were found using the following relationship:

$$\nu = 2\left(\frac{T}{M}\right) b_1$$

where $b_1 = 1.86$ for the Parzen window and T = total number of observations, which in our case, is 85 oblique incidence soundings. Table III gives, for the various truncation points, the corresponding bandwidths, degrees of freedom, and a 95% confidence interval for the theoretical spectrum, $\Gamma_{yy}(f)$, for the Parzen lag window.

TABLE III: BANDWIDTH, DEGREES OF FREEDOM, AND 95% CONFIDENCE INTERVALS FOR SELECTED VALUES OF L FOR PARZEN'S WINDOW

L	Bandwidth	d.f.	95% C. I.	$\Gamma_{yy}(f)$
8	.233	38	.65	1.54
16	.116	19	.58	2.20
20	.093	15	.54	2.35
24	.078	12	.50	2.75
32	.058	9	.48	3.30

In selecting a proper value for L for our spectral density, we want to be able to detect peaks in the spectrum, have a reasonable confidence interval, and a bandwidth which affords us a reasonable bias. For an L value

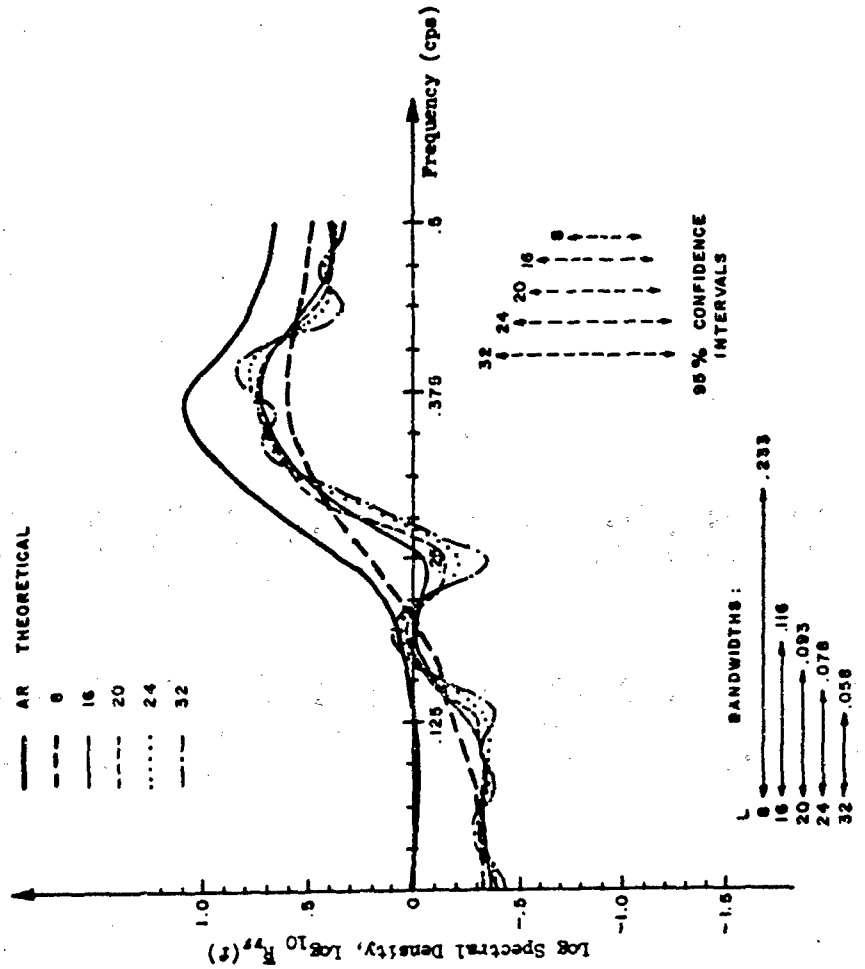


Figure 8. ESTIMATE OF THE SPECTRAL DENSITY USING THE PARZEN LAG WINDOW

Of 8 units, the spectral density was too smooth, and we were unable to detect peaks less than 0.233 wide. Increasing the L values from 16 to 20 units, gives a fairly reasonable display of the spectral density, that is, two major peaks occur which are quite similar to those of the theoretical density. For a truncation point of 24 units, very small peaks begin to appear which indicate that the variance may be influencing the density. This was also seen at L = 28 and 32, where the peaks became very erratic, and very noticeable. Thus, our choice was narrowed very quickly to choosing L = 16 or 20 units. The confidence intervals for L = 16 and L = 20 units are almost identical. The bandwidth for L = 20, however, has been reduced by about 20% from that of L = 16. Therefore, the spectral density corresponding to L = 20 units was selected as the most reasonable truncation point. The spectral density estimate clearly shows that most of the power is concentrated at high frequencies. A major peak is located at f = .375 cycles per second with a smaller peak located at .18 cycles per second. The bandwidth for L = 20 units is .093, which means that we can detect peaks with a width of this value or greater. The above remarks are graphically verified in Figure 8 where the theoretical spectral density for the third order autoregressive model is compared with the spectral estimate for L = 8, 16, 20, 24, and 32 units. In addition, the 95% confidence interval and the corresponding bandwidths for the truncation points are given.

5. BIVARIATE SPECTRAL ANALYSIS OF THE OI SOUNDINGS:

In this section, we shall be concerned with analyzing the bivariate behavior of the oblique and vertical incidence ionospheric soundings for the 60 Km experiment. More specifically, we shall obtain estimates of the smooth coquadrature, phase, and cross-amplitude spectra using the three lag windows we discussed in Section 4. In addition, we shall obtain estimates of the coherency spectrum.

With respect to the aims of the present study, we will only give the equations (estimates) which characterize the above concepts and we will not discuss the theoretical implications. For complete details of these concepts, see Jenkins and Watts [1968] and Box and Jenkins [1971].

The sample cross-correlation function is defined by:

$$r_{yx}(k) = \frac{c_{yx}(k)}{\sqrt{c_{yy}(o)\, c_{xx}(o)}} \qquad (5.1.1)$$

where:

$$c_{yx}(k) = \frac{1}{N} \sum_{t=1}^{N-K} (Y_{1t} - \bar{Y})(X_{2t+k} - \bar{X}) \;,\; 0 \le k \le L-1 \qquad (5.1.2)$$

As in the univariate case, the sample cross spectrum is obtained by taking the fourier transform of the sample cross covariance function. The sample cospectral estimate is given by:

$$\bar{L}_{yx}(i) = \left\{ \ell_{yx}(o) + 2 \sum_{k=1}^{L-1} \ell_{yx}(k) w(k) \cos \frac{\pi i k}{F} \right\} \;,\; 0 \le i \le F \qquad (5.1.3)$$

where:

$$l_{yx}(k) = \tfrac{1}{2}\{c_{yx}(k) + c_{yx}(-k)\},\ 0 \le k \le L-1 \qquad (5.1.4)$$

The equation used to calculate the quadrature spectral estimate is given by:

$$\bar{Q}_{yx}(i) = 4\sum_{k=1}^{L-1} q_{yx}(k)w(k)\sin\frac{\pi ik}{F},\ 1 \le i \le F-1 \qquad (5.1.5)$$

where:

$$q_{yx}(k) = \tfrac{1}{2}\{c_{yx}(k) - c_{yx}(-k)\},\ 0 \le k \le L-1 \qquad (5.1.6)$$

Note that $Q_{yx}(0) = \bar{Q}_{yx}(F) = 0$. The smoothed cross-amplitude spectral estimate was calculated using the following equation:

$$A_{yx}(i) = \sqrt{\bar{F}^2_{yx}(i) + \bar{Q}^2_{yx}(i)},\ 0 \le i \le F \qquad (5.1.7)$$

where the smoothed phase spectral estimate is given by:

$$\bar{F}_{yx}(i) = \arctan - \frac{\bar{Q}_{yx}(i)}{\bar{L}_{yx}(i)},\ 0 \le i \le F \qquad (5.1.8)$$

and $\bar{Q}_{yx}(i)$ and $\bar{L}_{yx}(i)$ are as previously defined. The smoothed squared coherency spectral estimate is given by:

$$\bar{K}^2_{yx}(i) = \frac{\bar{A}^2_{yx}(i)}{\bar{C}_{yy}(i)\bar{C}_{xx}(i)},\ 0 \le i \le F \qquad (5.1.9)$$

where: $\bar{A}_{yx}(i)$ is the smoothed cross-amplitude spectral estimate and

$\bar{C}_{yy}(i)$ is the smoothed spectral estimate given by:

$$\bar{C}_{yy}(i) = 2\left\{c_{yy}(0) + 2\sum_{k=1}^{L-1} c_{yy}(k)w(k)\cos\frac{\pi ki}{F}\right\},\ 0 \le i \le F \qquad (5.1.10)$$

and

$$\bar{C}_{xx}(i) = 2\left\{c_{xx}(0) + 2\sum_{k=1}^{L-1} c_{xx}(k)w(k)\cos\frac{\pi ki}{F}\right\},\ 0 \le i \le F \qquad (5.1.11)$$

Having calculated and plotted the cross amplitude spectrum, we can detect whether or not frequency components in the vertical incidence soundings are associated with large or small amplitudes at the same frequency in the oblique incidence series. The estimate of the phase spectrum of the two stochastic realizations helps us in determining whether or not frequency components in the vertical incidence series are in phase or out of phase (lag or lead) with components, at the same frequency, in the oblique incidence series.

An estimate of the cross-amplitude spectrum and the phase spectrum would suffice to provide a complete description of the behavior of the two series. The square coherency spectrum is the plot of the $\overline{K}^2_{yx}(f)$ vs. frequency. The cross amplitude spectrum, $\overline{A}_{yx}(f)$, is a measure of the covariance which exists between the oblique and vertical incidence soundings at frequency, f. In general, the coherency spectrum plays the role of a correlation coefficient with respect to frequency. Its usefulness lies in the fact that dimensions do not enter the picture when the correlation is measured with respect to frequency. Unlike the square coherency spectrum, the cross amplitude spectrum depends upon the dimensions of the oblique and vertical incidence soundings. This is the reason why the square coherency spectrum is sometimes preferred over the cross amplitude spectrum, and together with the phase spectrum, will give a complete picture of the cross correlation behavior of the oblique and vertical incidence soundings.

We shall, in what follows, obtain estimates for the coquadrature, phase, and cross amplitude spectral estimates using Bartlett's lag window. These smoothed estimates were obtained using the truncation points L = M = 8, 12, 16, 20, and 24 units for the cross spectral estimate and L = M = 8, 12, 16, 24, and 32 units for the smoothed coquadrature spectral estimate. These truncation points correspond to decreasing the bandwidth to $b = b_1/L = 1.5/L$.

Figure 9 shows the smoothed cospectral estimate. Similarly, Figure 10 shows, on the same axes, the various smoothed quadrature spectral estimates. It is clear that for $L \geq 20$ units, the estimates in both cases, i.e., vertical and oblique, become very erratic. As we mentioned previously, compromising between bias and variance, it appears that for L = 16 units, we have the best estimate using Bartlett's lag window with b = .094 and $\nu = 15$ degrees of freedom. The smoothed phase spectral estimate and the smoothed cross spectral estimate, plotted for L = 16, each on separate sets of axes to enhance the details of the series, are shown in Figures 11 and 12, respectively.

The smoothed coquadrature, phase, and cross amplitude spectral estimates were similarly obtained using Tukey's lag window for truncation points L = 8, 12, 14, 16, and 32. Figure 13 displays the smoothed cospectral estimates. The smoothed quadrature spectral estimates are plotted in Figure 14 for the same truncation points. For both of these cases, the estimates become more erratic as L is increased beyond 20 units. Taking the bandwidth into consideration, we choose the estimate for which L = 14 units as the best compromise between bias and variance. Thus, the bandwidth resulted in $b = 1.33/L = .095$ for L = 14 and $\nu = 15$ degrees of freedom for the Tukey lag window. Decreasing b to .083, the degrees of freedom are decreased considerably, therefore, having chosen L = 14 units will give the best estimate of the co- and quadrature spectra for the Tukey lag window. The smoothed phase and smoothed cross amplitude spectra were then plotted for L = 15 units to enhance the details. Figures 15 and 16 display the smoothed cross amplitude spectral estimate and the smoothed phase spectral estimate respectively, using the Tukey lag window for L = 14.

A similar analysis was performed to obtain smoothed estimates for the co- and quadrature spectra using Parzen's lag window for L = 8, 16, 20, 24, and 32 units. Figures 17 and 18 display the above smoothed estimates.

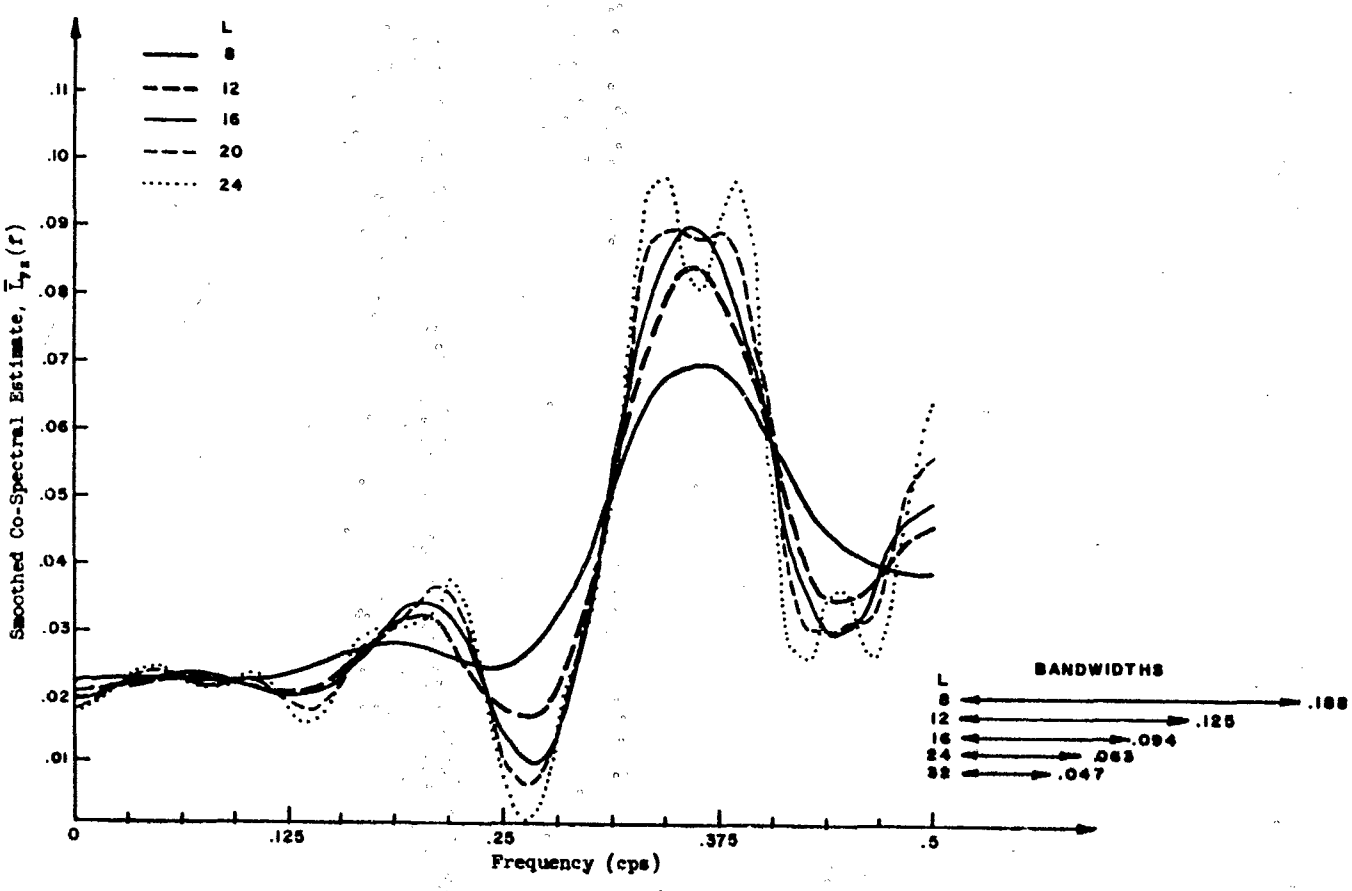


Figure 9. SMOOTHED CO-SPECTRAL ESTIMATES USING THE BARTLETT LAG WINDOW

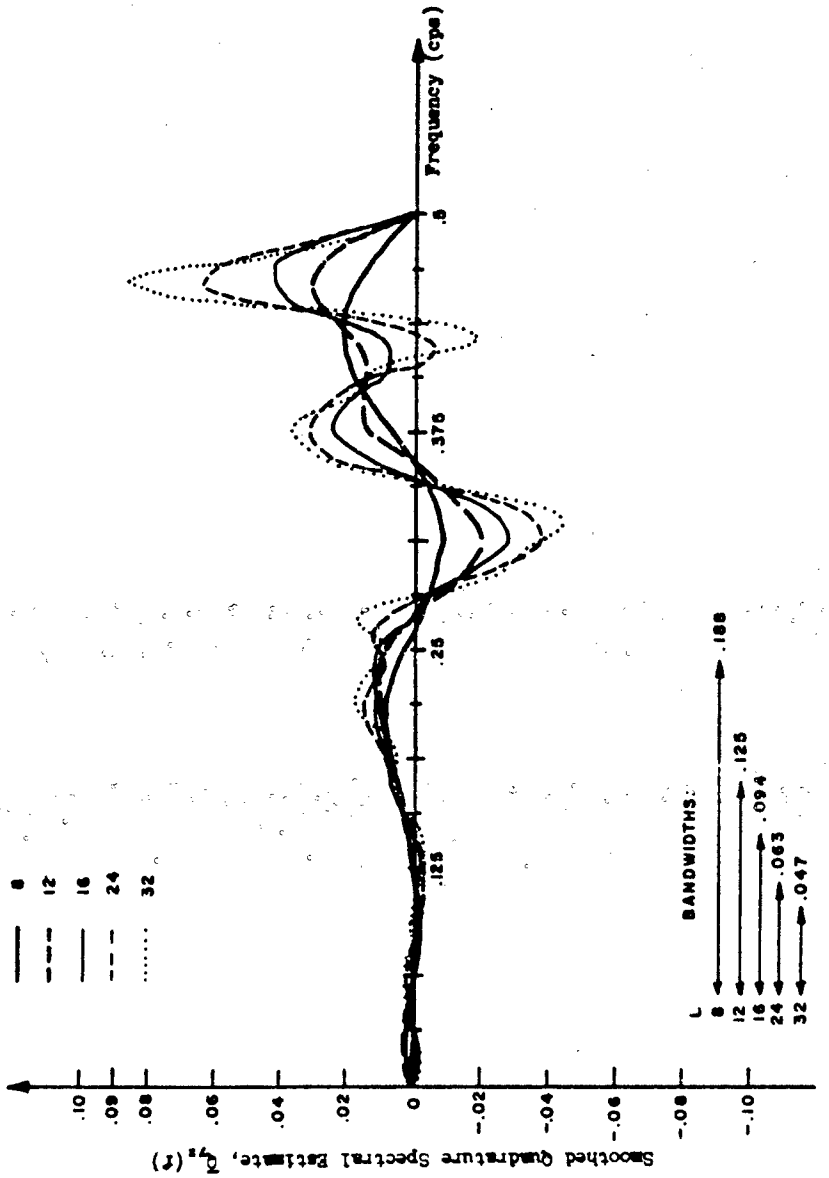


Figure 10. SMOOTHED QUADRATURE SPECTRAL ESTIMATES USING THE BARTLETT LAG WINDOW

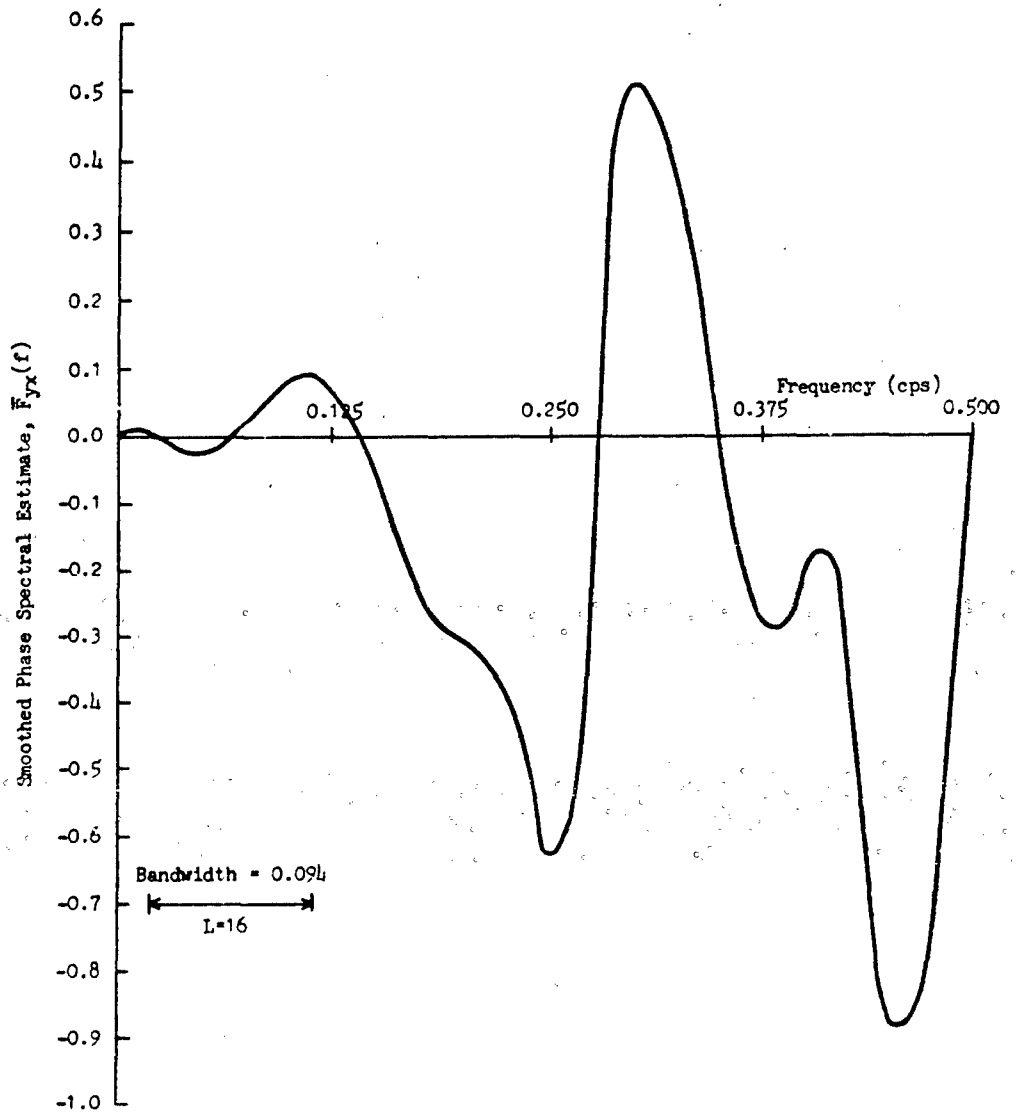


Figure 11. SMOOTHED PHASE SPECTRAL ESTIMATE USING THE BARTLETT LAG WINDOW

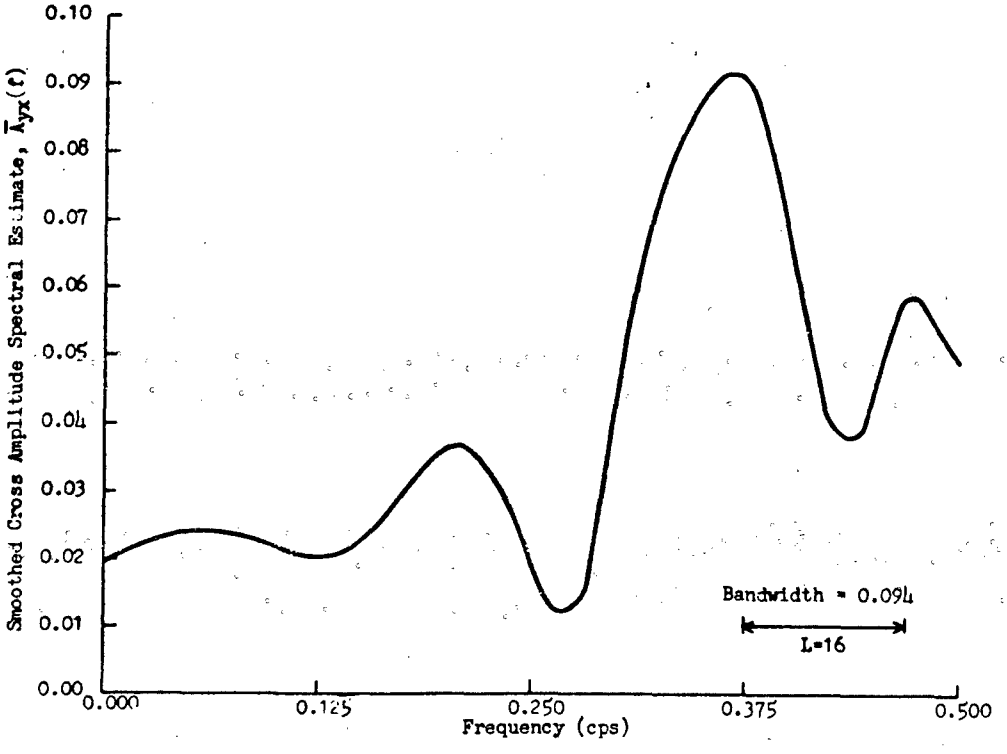


Figure 12. SMOOTHED CROSS AMPLITUDE SPECTRAL ESTIMATE USING THE BARTLETT LAG WINDOW

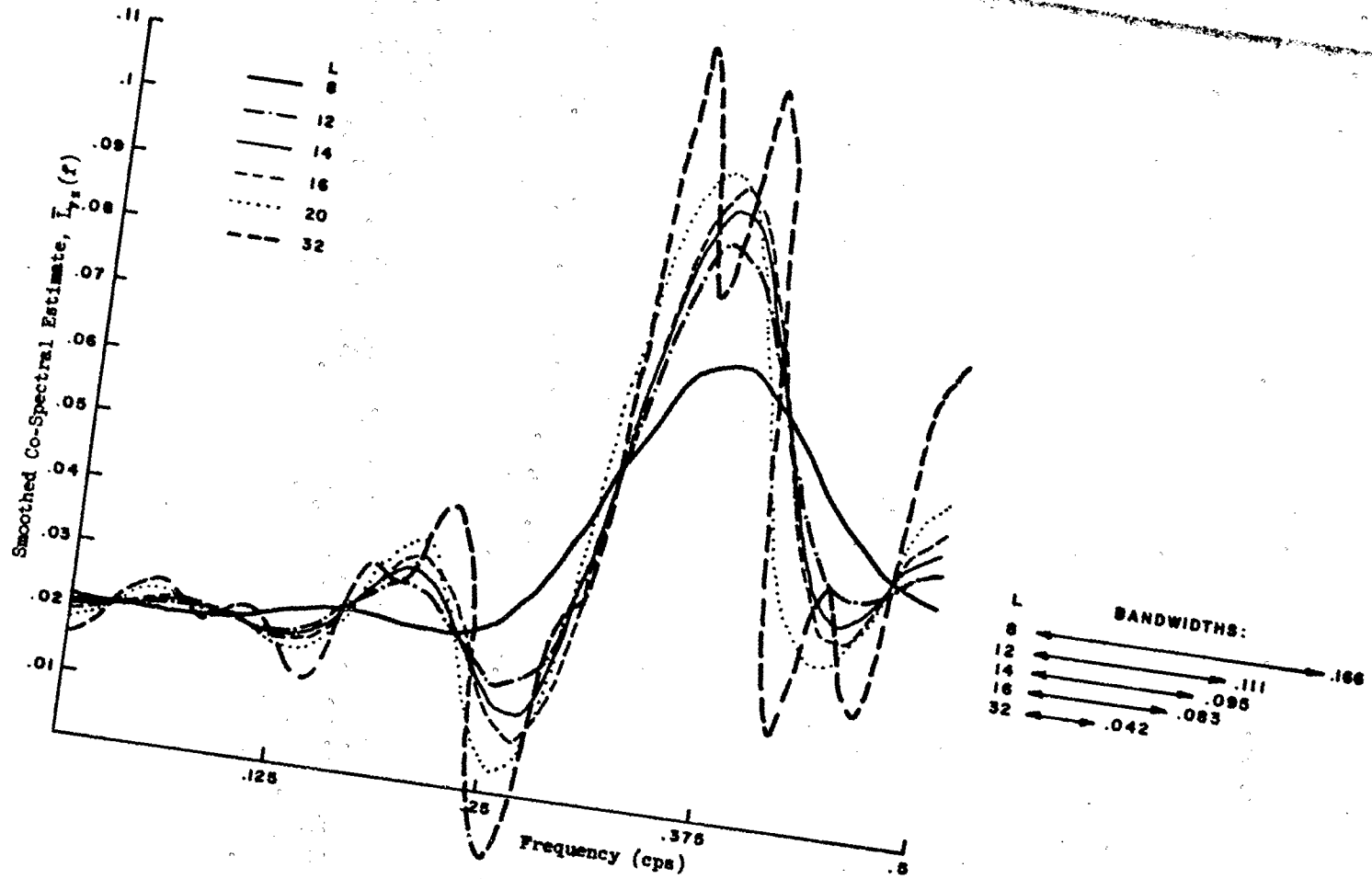


Figure 13. SMOOTHED CO-SPECTRAL ESTIMATES USING THE TUKEY LAG WINDOW

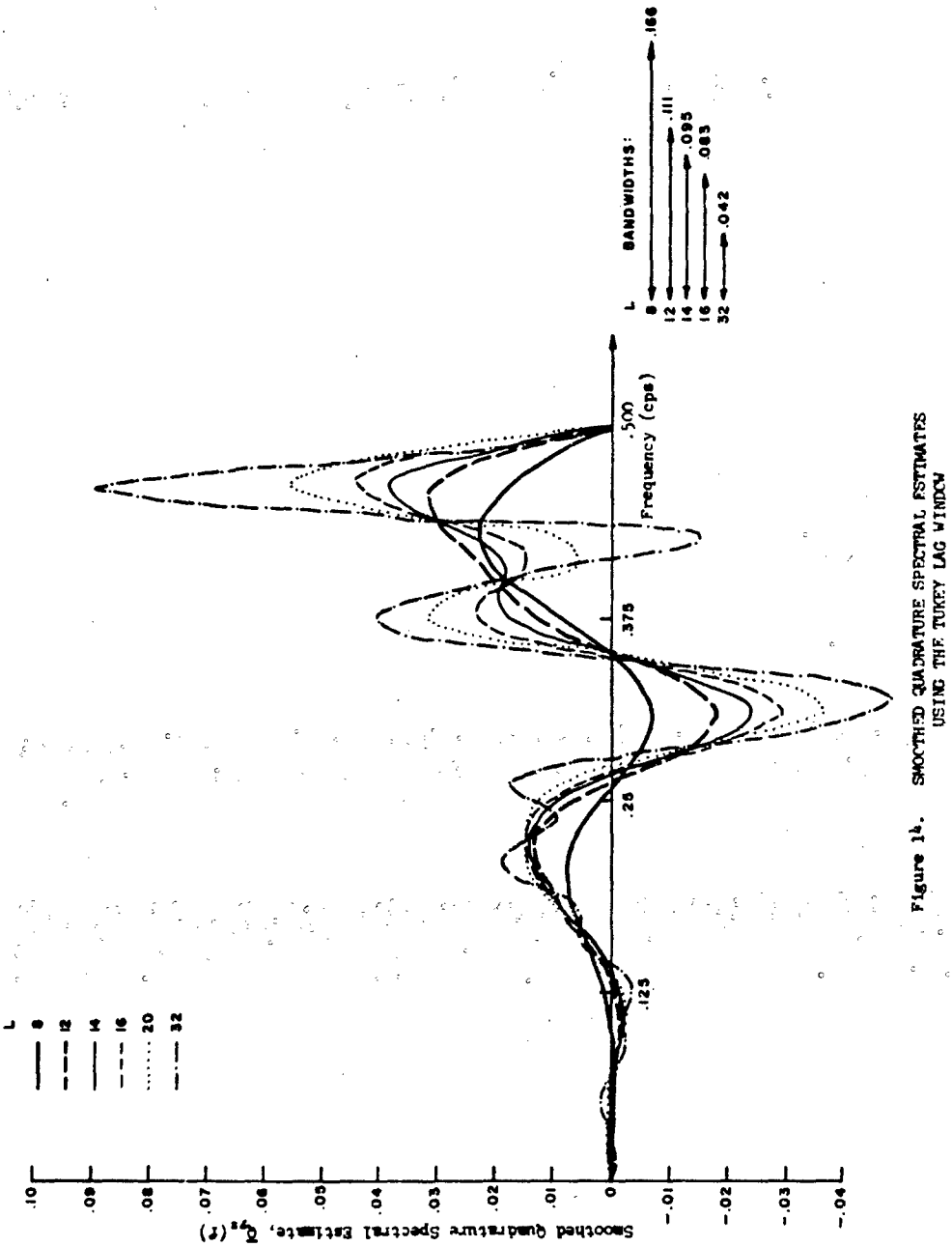


Figure 14. SMOOTHED QUADRATURE SPECTRAL ESTIMATES USING THE TUKEY LAG WINDOW

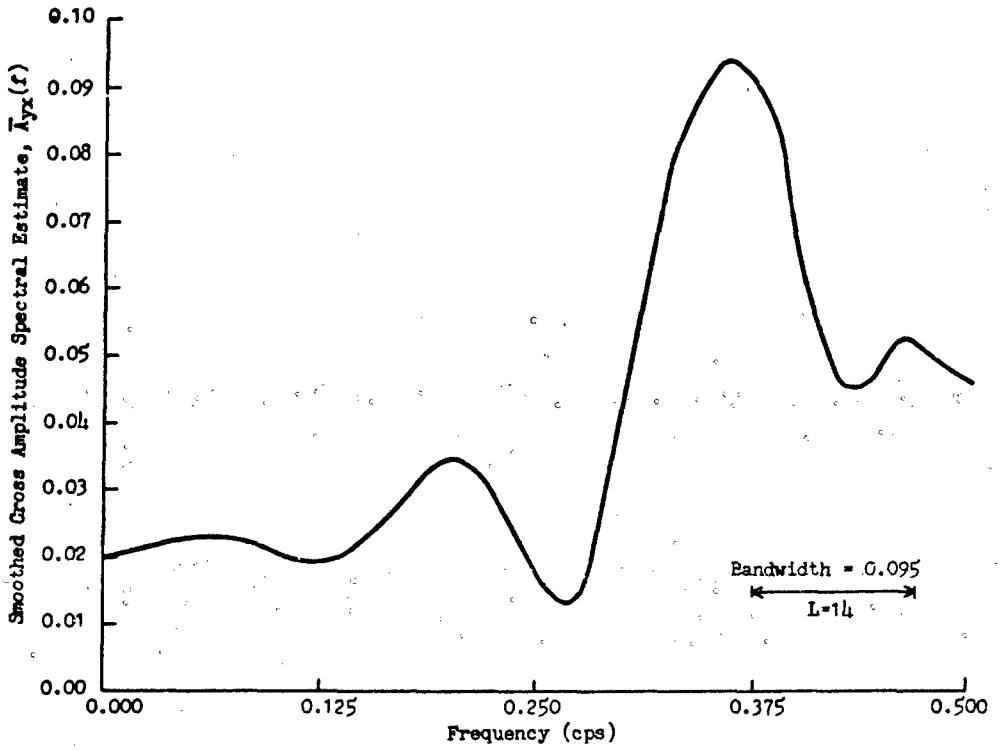


Figure 15. SMOOTHED CROSS AMPLITUDE SPECTRAL ESTIMATE USING THE TUKEY LAG WINDOW

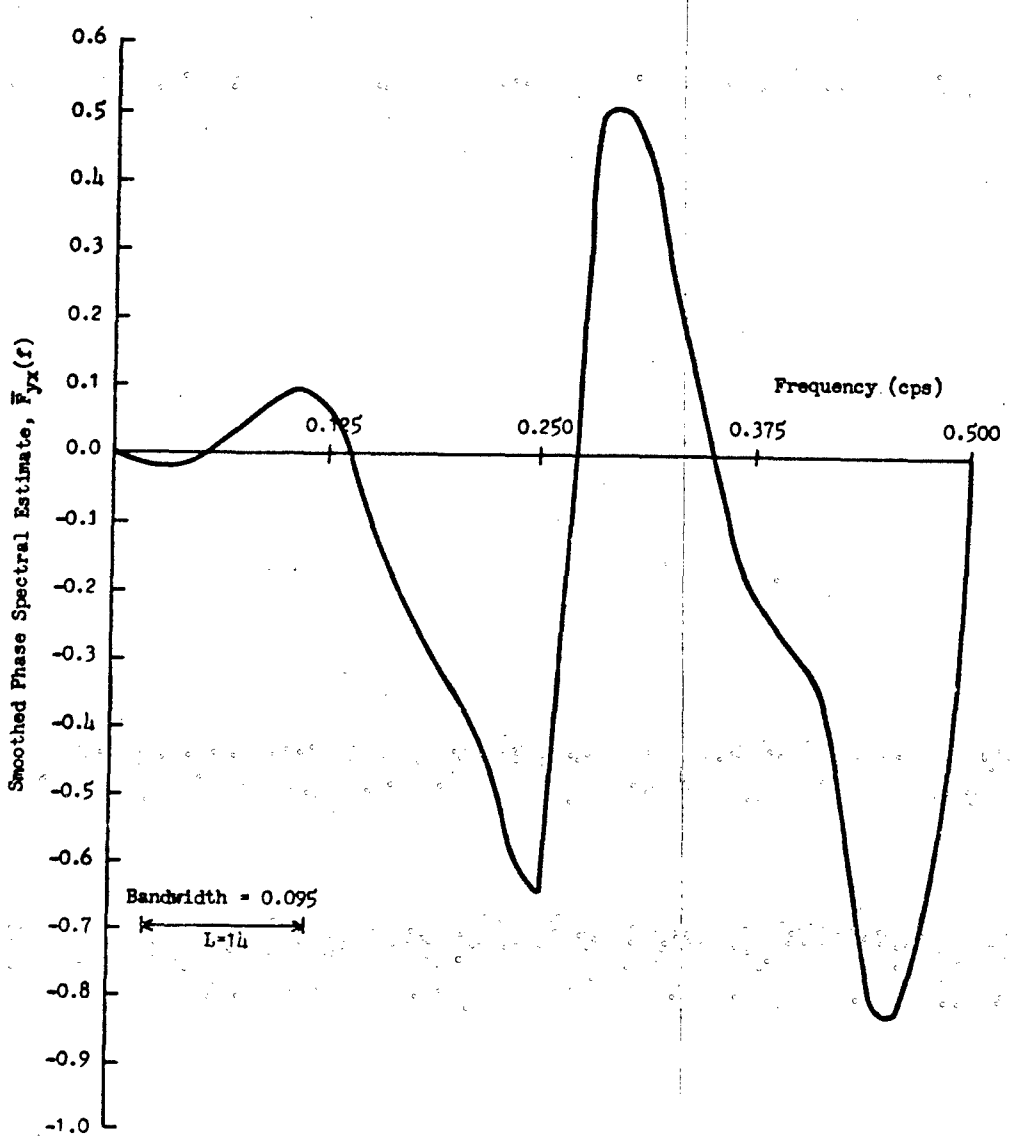


Figure 16. SMOOTHED PHASE SPECTRAL ESTIMATE USING THE TUKEY LAG WINDOW

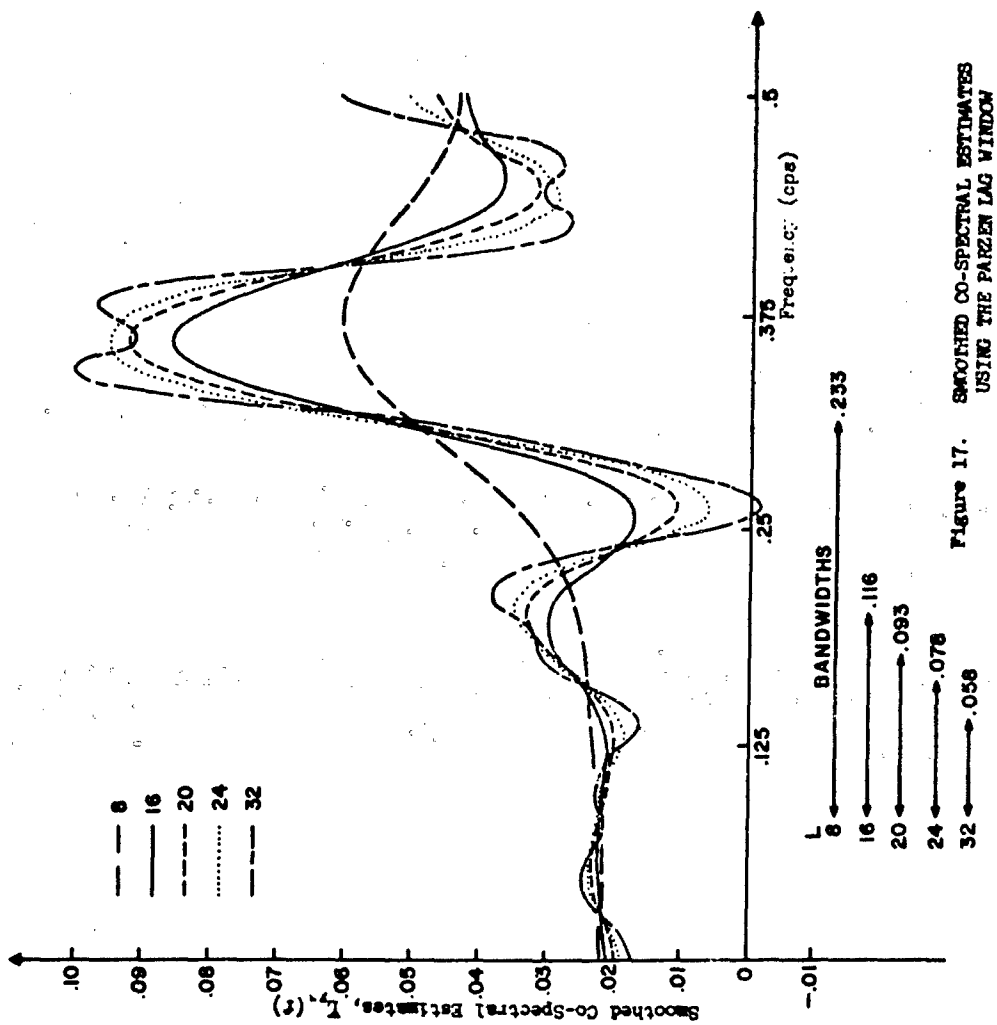


Figure 17. SMOOTHED CO-SPECTRAL ESTIMATES USING THE PARZEN LAG WINDOW

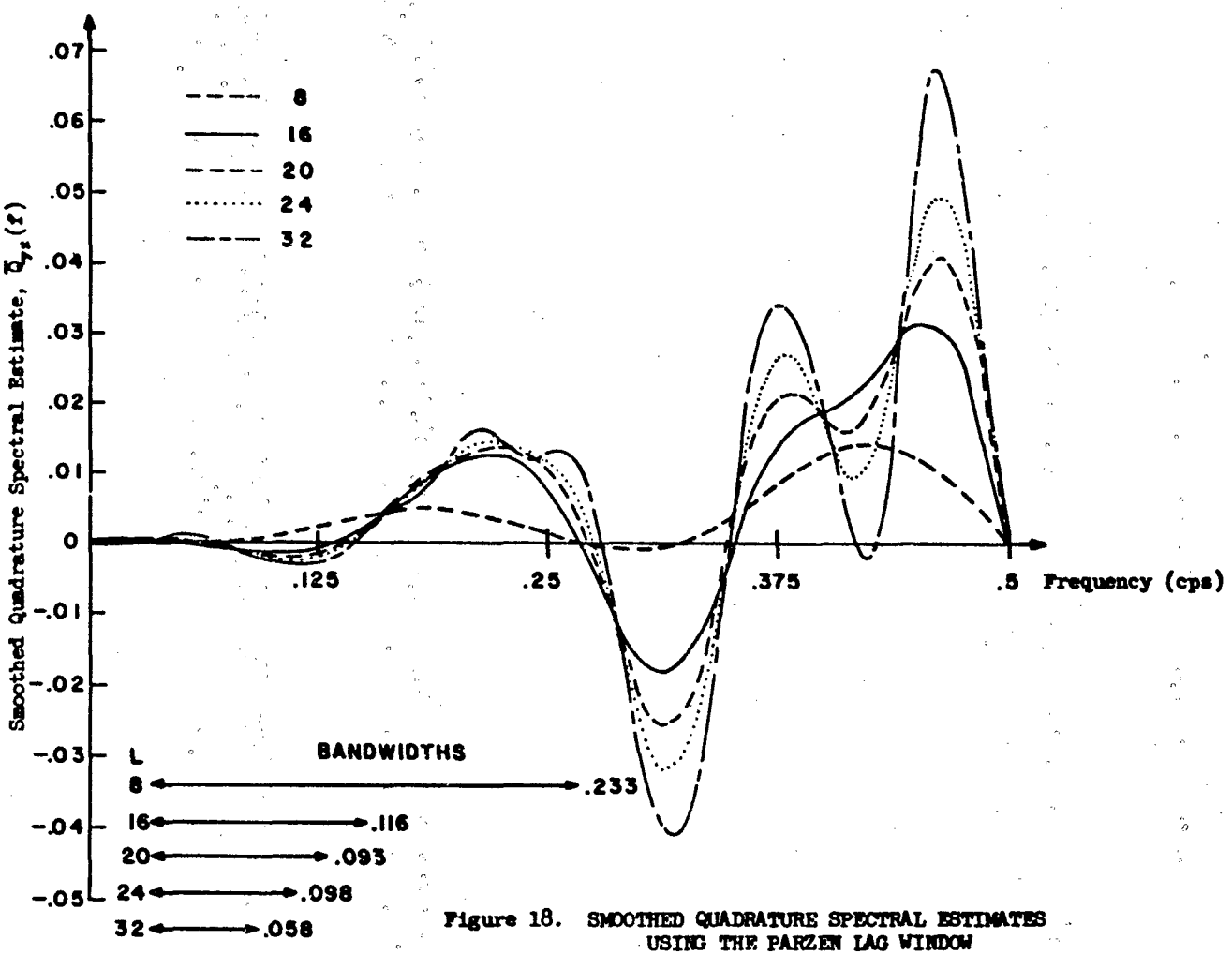


Figure 18. SMOOTHED QUADRATURE SPECTRAL ESTIMATES USING THE PARZEN LAG WINDOW

The bandwidths for the Parzen lag window are given by $b = 1.86/L$ and the degrees of freedom can be obtained from $\nu = 166\ b$. For values of $L \geq 24$, the estimates become quite erratic and the bandwidth and degrees of freedom are decreased. However, the decrease in bandwidth from .093 to .078 for L = 20 and 24 units, respectively, is not worth the decrease in variance. Hence, we choose L = 20 as our best estimates of the co- and quadrature spectra. This gives a bandwidth of b = .093. Figures 19 and 20 show the smoothed phase and cross amplitude spectral estimates, respectively, for L = 20, using the Parzen lag window, along with their corresponding bandwidth.

To compare the estimates obtained for the Bartlett, Tukey, and Parzen lag windows, the estimates corresponding to the best value of L (chosen for each window) were plotted on the same axes, see Figure 21. The estimates for the co- and quadrature spectra coincided almost exactly. Each estimate has 15 degrees of freedom for the autospectrum analysis. The Parzen lag window has a slightly smaller bandwidth than the others. It was difficult to choose the best window, but since Parzen's lag window for L = 20 units gave a bandwidth of .093, we chose it as the best smoothed estimate of the co- and quadrature spectra. The smoothed estimates for the phase and cross amplitude spectra are also best represented by this lag window for L = 20 units. The smoothed sample cospectral estimate estimates the covariance due to the in-phase components. There is a peak at about 0.2 cycles per second which corresponds to the peaks in the autospectra due to the fact that the variance is a special case of the covariance. At frequencies less than 0.125 cycles per second, the covariance between the vertical and oblique incidence realizations is reasonably small and constant over the frequency range 0 to .125 cycles per second. The variance at most frequencies in the autospectra is fairly large. However, the covariance distribution of the in-phase components of the filtered ionospheric series is small, and therefore, the series in-phase components are not very dependent. The larger value of the sample cospectrum is near .375 cycles per second corresponding to variance values of autospectra of about 10 at the same frequency for the Parzen lag window. L = 20 units, and hence, the correlation is small as will be verified by the square coherency spectral estimate.

The smoothed quadrature spectral estimate estimates the covariance of the out-of-phase components of the two filtered time series. This also shows that there is small covariance between the out-of-phase components of the two filtered series and, hence, that they are not very correlated. The largest value is 0.041 for the chosen lag window (Parzen, L = 20) and the smallest value is -.025. There is little or no covariance exhibited at all in the range 0 to 0.25 cps., but the out-of-phase components begin to vary in a sinusoidal manner at high frequencies.

The smoothed phase spectral estimate estimates the phase angle in radians by which one filtered time series leads or lags the other. At frequencies 0 to .0625 cps., the phases are approximately the same (phase spectral estimate is near 0). At frequencies between .0625 cps. and 0.13 cps. the in-phase components of the two time series lag the out-of-phase components very slightly. From 0.13 cps. to approximately 0.27 cps., the out-of-phase components lag the in-phase components. From 0.27 cps. to 0.35 cps., the in-phase is lagging, and from 0.35 cps. to 0.5 cps., the out-of-phase components lag the in-phase components of the two time series. Since the phases alternate leading, there is no reason to assume or conclude that one time series leads or lags the other at all frequencies.

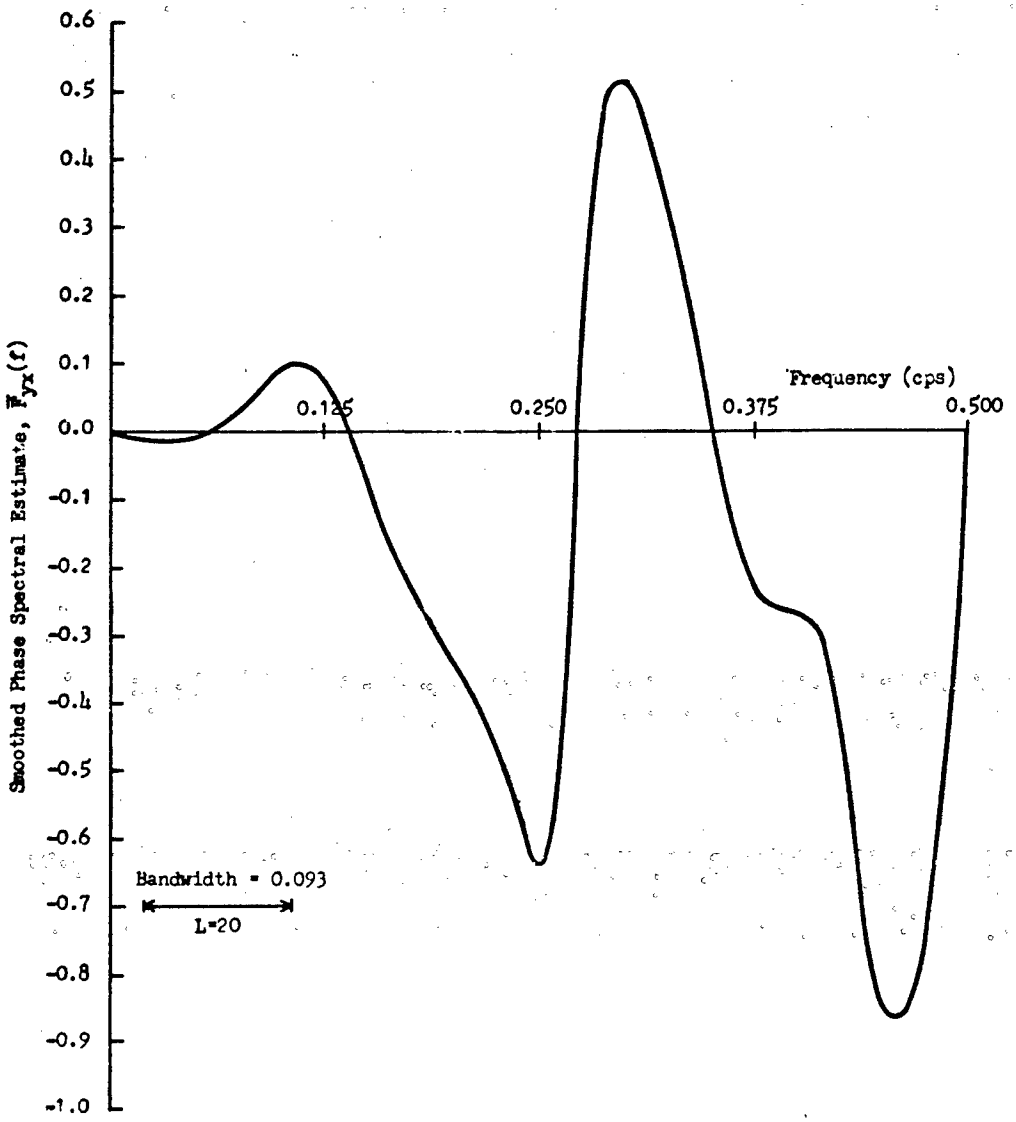


Figure 19. SMOOTHED PHASE SPECTRAL ESTIMATE USING THE PARZEN LAG WINDOW

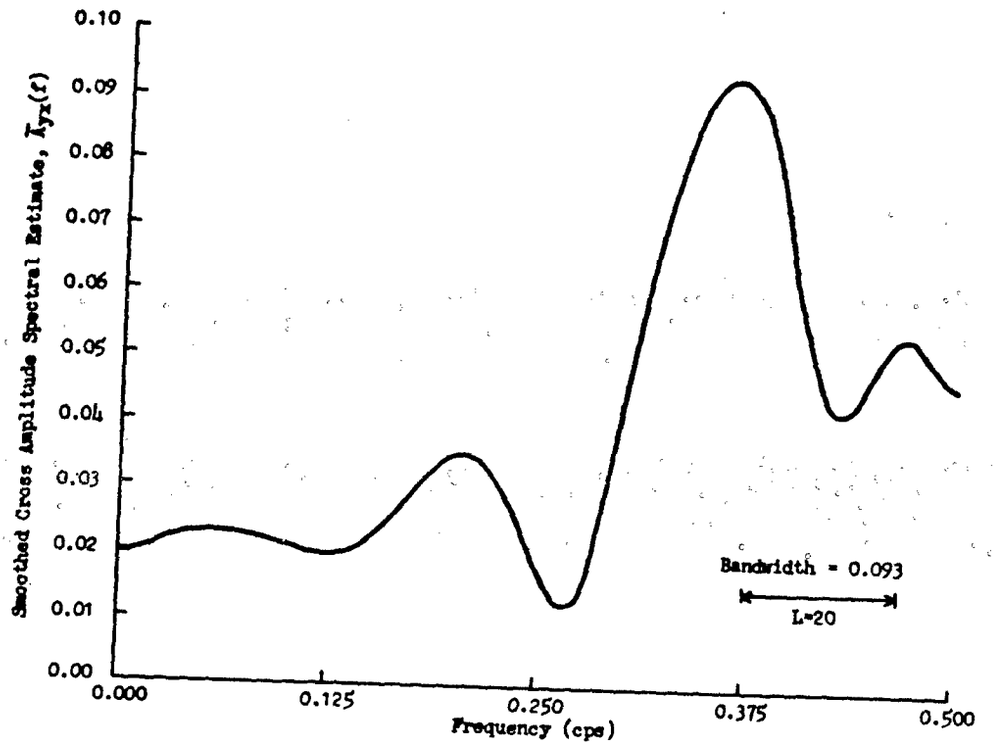


Figure 20. SMOOTHED CROSS AMPLITUDE SPECTRAL ESTIMATE USING THE PARZEN LAG WINDOW

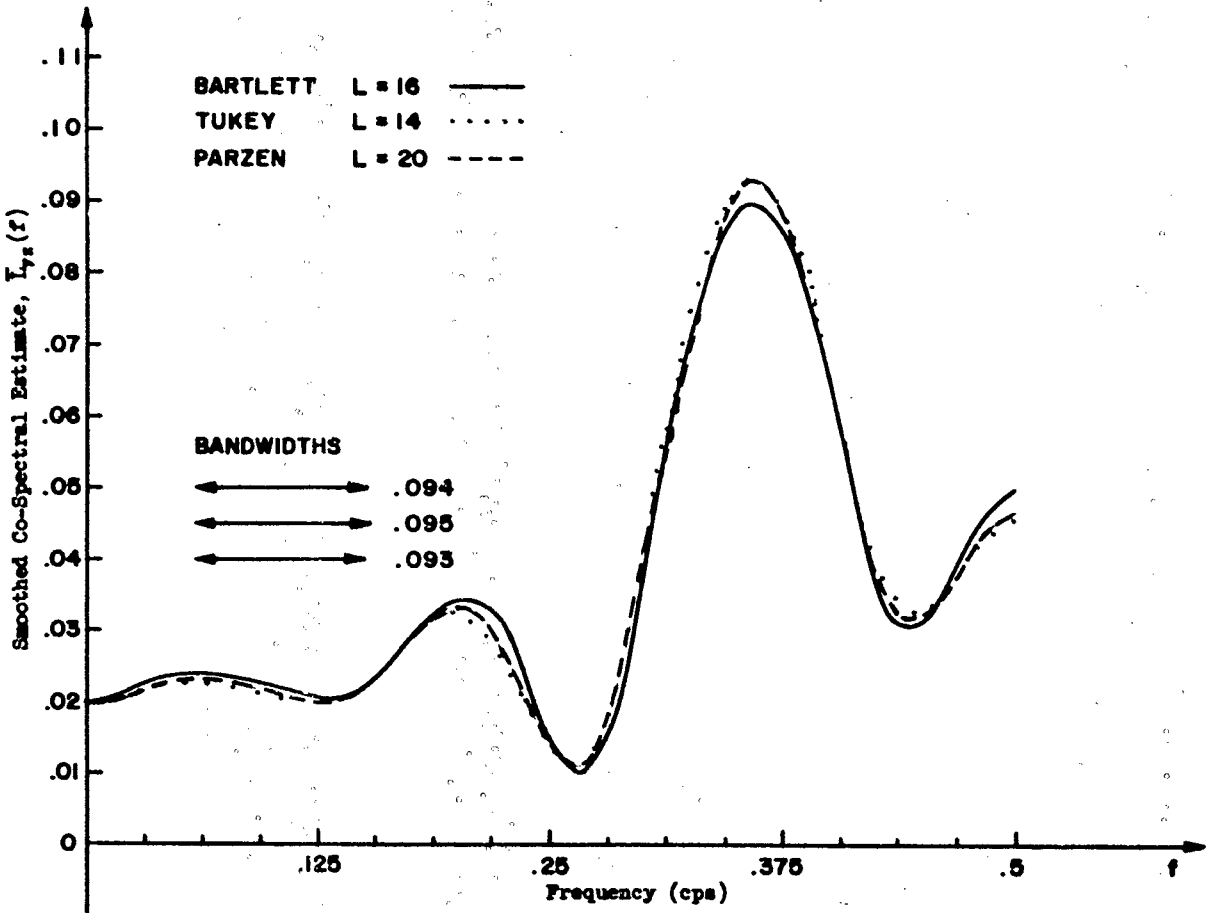


Figure 21. COMPARISON OF THE SMOOTHED CO-SPECTRAL ESTIMATES OF THE BARTLETT, TUKEY, AND PARZEN LAG WINDOWS

The smoothed cross amplitude spectral estimate shows whether or not the amplitude of the components at a particular frequency in one time series is associated with a large or small amplitude of the same order at the same frequency in the other time series. The spectral density of the autospectra shows that the variance is about 10 in both filtered $\bar{x}_t$ and $\bar{y}_t$ series so that, at frequencies from 0.3 cps to 0.4 cps, the amplitude of the components of one time series is associated with corresponding large or small amplitudes at the same frequency in the other. Again, this seems to indicate that covariance between the component amplitudes is near zero at other frequencies. In Figure 21, the best smoothed cospectral estimate is displayed and, in Figure 22, the best smoothed quadrature spectral estimate is shown.

6. SUMMARY AND CONCLUSIONS

A plot, see Figure 21, is given for the selected best estimates of the spectral densities for each of the three lag windows, namely, those of Bartlett, Tukey and Parzen. Although the truncation is different for each lag window, the bandwidth, degrees of freedom, and confidence intervals are almost identical. Thus, it is quite difficult to choose which lag window gives the best smoothed estimate of the spectral density function. However, calculating the approximate bias for each of the above lag windows, we found that the bias for Parzen's lag window is somewhat smaller than that for the Tukey and Bartlett lag windows. That is:

$$B_p(f) = \frac{.152}{400} \Gamma''_{yy}(f)$$

Furthermore, the variance ratio, that is, the proportional reduction in variance as the result of using the smoothed estimator as compared to the sample spectrum estimate, is approximately equal to 0.128. On the basis of these two criteria, we choose the best estimate of the spectral density using Parzen's lag window. In addition, the bandwidth of this lag window is slightly smaller than that of the Tukey and Bartlett lag windows. Therefore, the best estimate of the spectral density of the average oblique incidence soundings was obtained using Parzen's lag window for L = 20 units. This value of L resulted in a 95% confidence interval width of 2.25 with 15 degrees of freedom, and a bandwidth of b = .093. The bandwidth is less than 1/5 of the total frequency range over which the spectral density function is estimated. Since we are detecting peaks with widths of .093 or more, the two peaks appearing in the estimated spectral density at frequencies f = 3/16 cps and f = 3/8 cps are valid peaks, and they should be taken into consideration in interpreting the behavior of the average oblique incidence soundings. The process generating the resultant soundings exhibits large variance around these two frequencies for the filtered data. Such information should be taken into account in the design of the system. Frequencies below f = .125 cps on the spectral estimates gives the lowest power, that is, the least variance.

The Parzen lag window for L = 20 units and b = .093, was used to obtain smoothed estimates of the co- and quadrature spectra. The smoothed estimates of the phase and cross amplitude spectra were also obtained using the same lag window and L = 20 units.

The smoothed sample spectral estimate estimates the covariance due to the in-phase components. There is a peak at about .20 cps and one at .375 cps which correspond to the peaks in the autospectra. At frequencies

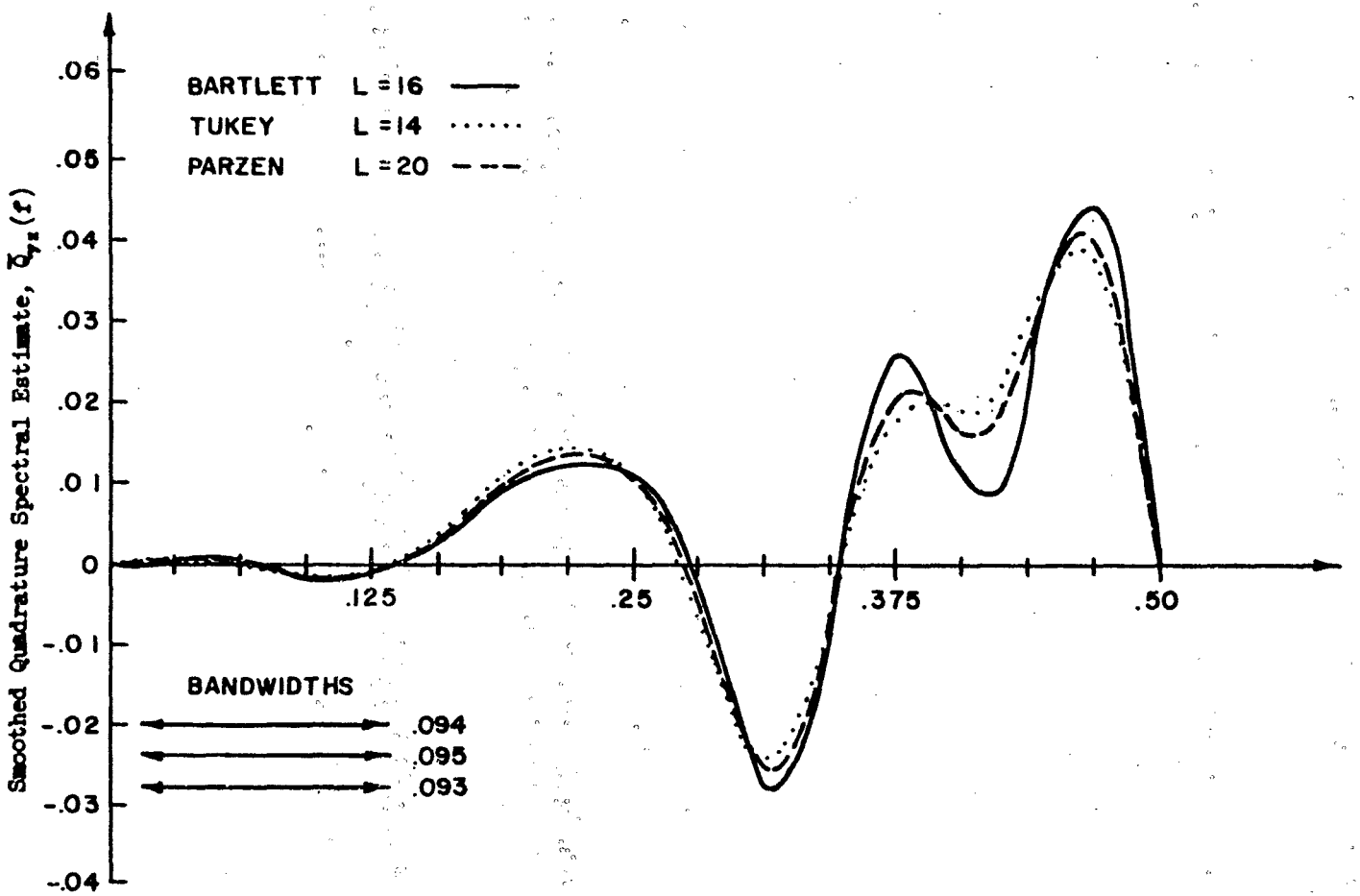


Figure 22. COMPARISON OF THE SMOOTHED QUADRATURE SPECTRAL ESTIMATE OF THE BARTLETT, TUKEY, AND PARZEN LAG WINDOWS

less than .125 cps, the covariance is reasonably small and approximately constant over the frequency range of 0 to .125 cps. The variance at most frequencies in the autospectra is fairly large. However, the covariance distribution of the in-phase components of the two filtered series is small and has, due to the soundings series, in-phase components that are not very dependent. The larger value of the sample spectra is near .375 cps, corresponding to variance values of the autospectra of about 10, at the same frequency, using the Parzen lag window for L = 20 units. Hence, the correlation between the average oblique and vertical incidence soundings is small as was verified by the squared coherency spectral estimate.

The smoothed quadrature spectral estimate estimates the covariance of the out-of-phase components of the filtered oblique and vertical incidence soundings. It showed that the covariance between the out-of-phase components of the two filtered series is small, and hence, that they are not very correlated. The largest value is .041 for the chosen lag window, and the smallest value is -.025. There is little or no covariance exhibited in the range from 0 to .25 cps, but the out-of-phase components begin to vary in a sinusoidal manner at higher frequencies.

The smoothed phase spectral estimate estimates the phase angle in radians by which one filtered time series leads or lags another. At frequencies 0 to .0625 cps, the phases are approximately the same; that is, the phase spectral estimate is near zero. At frequencies between .0625 cps and .13 cps, the in-phase components of the two time series lag the out-of-phase components very slightly. From .13 cps to approximately .27 cps, the out-of-phase components lag the in-phase components. From .27 cps to .35 cps, the in-phase components are lagging, and from .35 cps to .50 cps, the out-of-phase components lag the in-phase components of the two time series (the average oblique and vertical incidence soundings). Since the phase is alternately leading, there is no reason to assume or conclude that one time series leads or lags the other at all frequencies.

The smoothed cross amplitude spectral estimate shows whether or not the amplitude of the components at a particular frequency in one time series is associated with a large or small amplitude of the same order, at the same frequency, in the other time series. The spectral density of the autospectra shows that the variance is about 10 in both the filtered average oblique incidence and filtered average vertical incidence soundings, so that, at frequencies from .30 cps to .40 cps, the amplitude of the components of one time series is associated with corresponding large or small amplitudes (at the same frequency) of the other. Again, this indicates that the covariance between the component amplitudes is near zero at other frequencies. In Figures 21 and 22, we displayed the best smoothed estimates of the cospectral and the quadrature spectral estimates.

In order to obtain a better representation of the important peaks and a confidence interval, the square coherency was calculated and plotted (see Figure 23) on the truncation scale, Jenkins and Watts [1968], given by:

$$y(f) = \arctan \left| \bar{K}_{yz}(f) \right| .$$

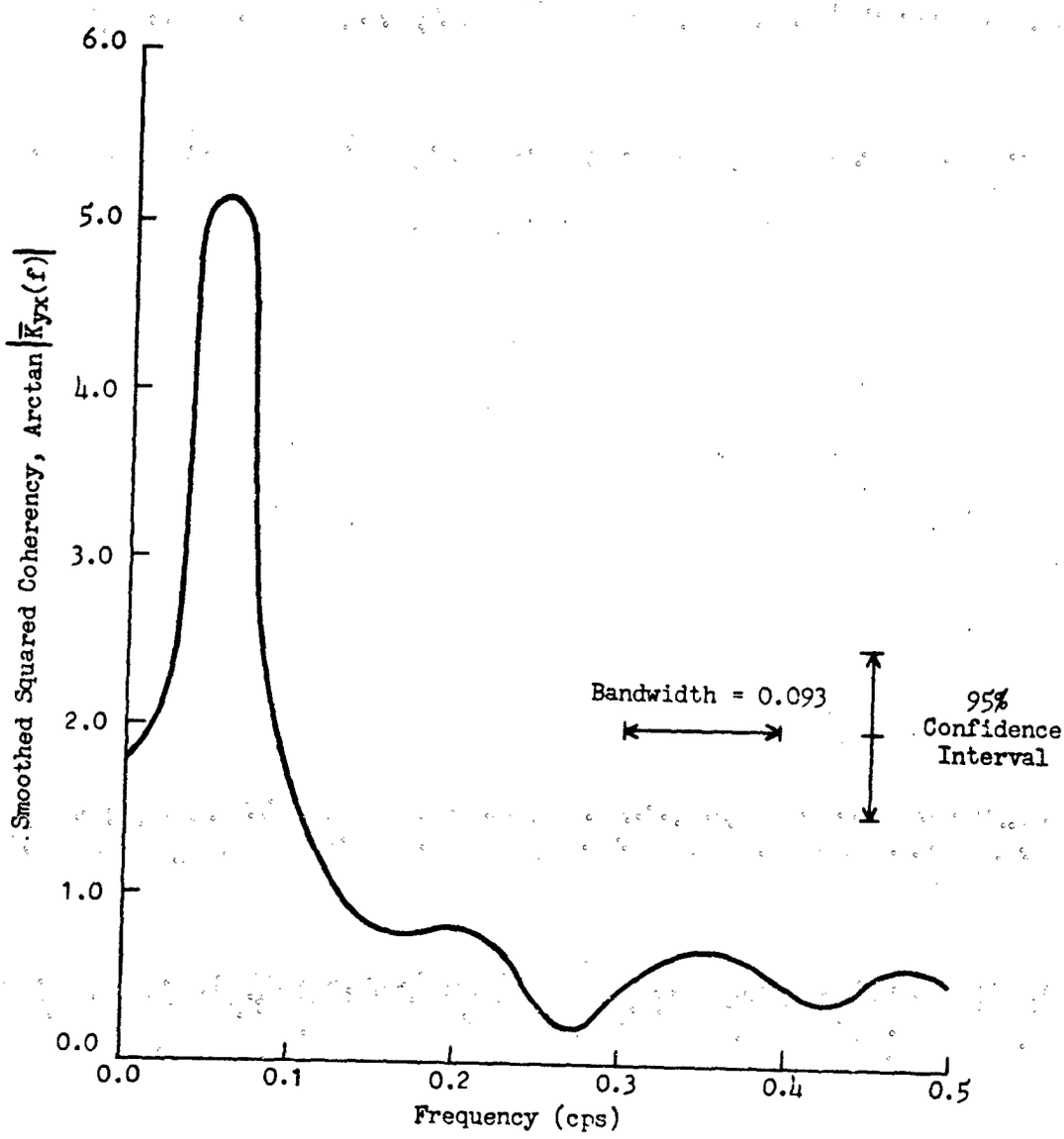


Figure 23. SMOOTHED SQUARED COHERENCY ON ARCTAN SCALE FOR THE PARZEN LAG WINDOW

A 95% confidence interval was obtained using the following expression:

$$\bar{\gamma}_{yx}(f) = \pm 1.96 \sqrt{L/2b_1 N}$$

$$= \pm 1.96 \sqrt{20/2(1.86)83} = \pm .499 ,$$

and is shown on the graph of the smoothed squared coherency spectrum. This squared coherency spectral estimate gives the correlation between the average oblique incidence soundings and the average vertical incidence soundings for the 60 Km experiment. At low frequencies, we have almost perfect correlation between the two filtered series, but this dampens out near zero at about .25 cps, and again at .50 cps. Furthermore, it never becomes greater than .33 cps. This frequency range shows virtually no correlation. Between these two frequencies, .25 cps and 50 cps, the squared coherency is near zero which indicates that the noise level is high in the filtered series for components of this frequency. This is consistent with the results obtained by the autospectra analysis, that is, the distribution of power or variance is larger at high frequencies (between .25 cps and .50 cps). At low frequencies, the squared coherency is high, which indicates low noise or variance in the autospectra for the corresponding frequencies and again, this is the same result obtained in the autospectra analysis.

REFERENCES

D'Accardi, R. J., C. P. Tsokos, and R. A. Kulinyi [1971]. "Statistical Models for HF Ionospheric Forecasting for Field Army Distances." Vol. Proceedings of the 17th Confereuce on Design of Experiments and Testing in Army R&D.

Ames, J. W. and R. D. Egan [1967]. "Digital Recording and Short Term Prediction of Oblique Ionospheric Propagation." IEEE Transcript, Vol. AP-15.

Krause, G. E. et al [1970]. "Field Test of a Near Real Time Ionospheric Forecasting Scheme (60 Km)." ECOM Technical Report #3345.

Jenkins, G. M., and D. G. Watts [1968]. Spectral Analysis and Its Applications. Holden-Day, San Francisco.

Box, G. E. P. and G. M. Jenkins [1970]. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.

Kendall, M. G. and A. Stuart [1966]. The Advanced Theory of Statistics. Vol. 3. Griffin, London.

Blackman, R. B. and J. W. Tukey [1959]. Measurement of Power Spectra from the Point of View of Communication Engineering. Dover, New York.

U. S. Army Materiel Systems Analysis Agency
Aberdeen Proving Ground, Maryland

MAXIMUM LIKELIHOOD ESTIMATION PROCEDURES IN RELIABILITY GROWTH

Larry H. Crow

Introduction

A development program is generally recognized as being a necessity for most systems since they usually exhibit initial design and engineering deficiencies. Attempts are made during the development program to find and remove these deficiencies to a point where certain levels of performance with respect to reliability and other requirements are met.

The development of a system usually evolves as a repeated process of system examination and testing, determination of system failure modes, and design and engineering changes as attempts to eliminate these modes. Because of the scarcity of data, it is often a difficult task for one to obtain directly good estimates of the progress of the development program and to project future progress. In this regard, program managers generally need specialized techniques and methodology which will allow them to evaluate the progress of the development program from a limited amount of test data. The area of reliability growth modeling is a management tool directed toward this need of the program managers.

It is usually assumed that the system reliability will increase during the development program and, thus, mathematical models describing this phenomenon have come to be called "reliability growth" models. Most of the reliability growth models considered in the literature assume that a mathematical formula (or curve), as a function of time, represents the reliability of the system during the development program. It is commonly assumed, also, that these curves are nondecreasing. That is, once the system's reliability has reached a certain level, it will not drop below this level during the remainder of the development program. It is important to note that this is equivalent to assuming that any design or engineering changes made during the

development program do not decrease the system's reliability.

The central purpose of most reliability growth models includes one or both of the following objectives:

- Inference on the present system reliability;
- Projection on the system reliability at some future development time.

This paper will consider a commonly used reliability growth model proposed by Duane [1]. For this model maximum likelihood estimates of the unknown parameters will be given along with appropriate confidence interval and hypotheses testing procedures.

Background

For a nonrepairable system let $F(x)$ be the cumulative distribution function of the time to failure. Let $f(x)$ be the corresponding probability density function. Then the system failure rate is

$$r(x) = \begin{cases} f(x)/1-F(x), & F(x) < 1 \\ \infty, & F(x) = 1. \end{cases}$$

Note that

$$r(x)dx$$

is the probability that a system of age x will fail in the interval $(x,x+dx)$. This probability is conditional on no failure during $[0,x]$.

Examples

1. Exponential

$$F(x) = 1-e^{-\lambda x}, \quad x > 0, \ \lambda > 0.$$

$$r(x) = \lambda, \quad x > 0.$$

2. Weibull

$$F(x) = 1-e^{-\lambda x^{\beta}}, \quad x > 0, \ \lambda > 0, \ \beta > 0.$$

$$r(x) = \lambda\beta x^{\beta-1}, \quad x > 0.$$

Observe that for $\beta=1$, $r(x)$ is constant. If $\beta>1$ ($\beta<1$) then $r(x)$ is increasing (decreasing) which implies that the system is wearing out (improving) with age.

The above definition of failure rate is appropriate when one is interested primarily in time to first failure. However, during a development program the system is repaired or modified after each failure and tested further.

The failure rate of a (complex) repairable system may be defined by

$$r(x)dx = \text{(unconditional) probability that a system of age } x \text{ will fail in } (x,x+dx).$$

This probability is independent of the failure history of the system during $[0,x]$. Again if $\beta>1$ ($\beta<1$) then $r(x)$ is increasing (decreasing) which implies that the system is wearing out (improving) with age.

Examples

1. Constant failure rate

$$r(x) = \lambda,\ \lambda > 0,\ x > 0.$$

2. Weibull failure rate

$$r(x) = \lambda\beta x^{\beta-1},\ \lambda > 0,\ \beta > 0,\ x > 0.$$

The Duane Reliability Growth Model [1] is usually written as

$$r(x) = (1-\alpha)\lambda x^{-\alpha}$$

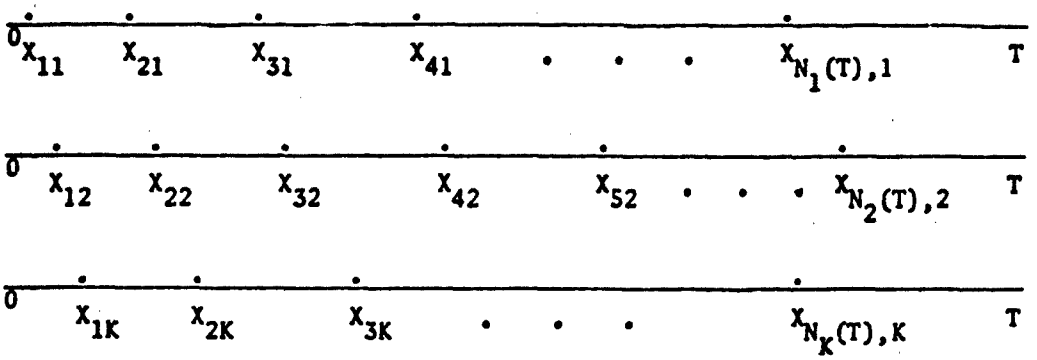
$x > 0$, $\lambda > 0$, $0 \leq \alpha \leq 1$, where $r(x)$ is the failure rate of a repairable system. Replacing $-\alpha$ by $\beta-1$, we see that the Duane Model and the Weibull repairable system failure rate model are the same. For a system with a constant failure rate for a fixed configuration, this model is equivalent to assuming that the mean time between failure (MTBF) of the system at time x is

$$M(x) = [r(x)]^{-1} = \frac{x^{1-\beta}}{\lambda\beta}.$$

That is, the MTBF is proportioned to $x^{1-\beta}$.

Maximum Likelihood Estimates of λ and β

Suppose K systems have each experienced T units of operation since the development program began. Let $N_r(T)$ be the random number of failures observed for the r-th system, r=1,...,K. Let X_{ir} be the age of the r-th system (regarding the age at the beginning of development as 0) at the i-th failure, $i=1,\ldots,N_r(T), r=1,\ldots,K$.



The maximum likelihood estimate (MLE) of β is

$$\hat{\beta} = \frac{\sum_{r=1}^{K} N_r(T)}{\sum_{r=1}^{K} \sum_{i=1}^{N_r(T)} \log \frac{T}{X_{ir}}}.$$

The MLE of λ is

$$\hat{\lambda} = \frac{\sum_{r=1}^{K} N_r(T)}{K\, T^{\hat{\beta}}}.$$

(All logs are with respect to base e.)

Example

Suppose K=3 systems were tested for time T=200. This experiment was simulated on a computer when λ=0.6 and β=0.5. These results are given in

Table 1 where X_{ir} is the age of the r-th system at the i-th failure. From this simulation the MLE of β is

$$\hat{\beta} = 0.615$$

and the MLE of λ is

$$\hat{\lambda} = 0.461.$$

The Duane Model states that if development of the system is stopped at T=200 hours of testing, then the times between failures of the system thereafter will follow the exponential distribution

$$F(x) = 1-e^{-x/M(T)}$$

$x > 0$, where

$$M(T) = [r(T)]^{-1} = \frac{(200)^{1-\beta}}{\lambda\beta}.$$

Based on 200 hours of testing the MLE of M(T) is

$$\hat{M}(200) = \frac{(200)^{1-\hat{\beta}}}{\hat{\lambda}\hat{\beta}} = 27.12.$$

If development is stopped at, say, T=300 hours of testing the model states that future times between failure will, also, follow the exponential distribution but with mean

$$M(300) = \frac{(300)^{1-\beta}}{\lambda\beta}.$$

Based on 200 hours of testing the projection of the MTBF at 300 hours of testing is

$$\hat{M}(300) = \frac{(300)^{1-\hat{\beta}}}{\hat{\lambda}\hat{\beta}} = 31.70.$$

TABLE 1

Simulated Data for K=3 Systems Operated For Time T=200 when $\lambda = 0.6$ and $\beta = 0.5$

Sys. 1	Sys. 2	Sys. 3
X_{i1}	X_{i2}	X_{i3}
4.3	0.1	8.4
4.4	5.6	32.5
10.2	18.6	44.7
23.5	19.5	48.4
23.8	24.2	50.6
26.4	26.7	73.6
74.0	45.1	98.7
77.1	45.8	112.2
92.1	75.7	129.8
197.2	79.7	136.0
	98.6	195.8
	120.1	
	161.8	
	180.6	
	190.8	
$N_1(T) = 10$	$N_2(T) = 15$	$N_3(T) = 11$
$\sum_{i=1}^{10} \log \frac{T}{X_{i1}}$	$\sum_{i=1}^{15} \log \frac{T}{X_{i2}}$	$\sum_{i=1}^{11} \log \frac{T}{X_{i3}}$
$= 19.661$	$= 26.434$	$= 12.398$

$$N(T) = N_1(T)+N_2(T)+N_3(T) = 36$$

$$\sum_{r=1}^{3} \sum_{i=1}^{N_r(T)} \log \frac{T}{X_{ir}} = 58.493$$

$$\hat{\beta} = \frac{N(T)}{\sum_{r=1}^{3} \sum_{i=1}^{N_r(T)} \log \frac{T}{X_{ir}}} = 0.615$$

$$\hat{\lambda} = \frac{N(T)}{K\, T^{\hat{\beta}}} = 0.461$$

Hypotheses Tests on β

Let N(T) be the total number of failures for the K systems. That is,

$$N(T) = N_1(T)+N_2(T)+\ldots+N_K(T).$$

Conditioned on N(T)=n (n a fixed integer), the random variable

$$2\beta \sum_{r=1}^{K} \sum_{i=1}^{N_r(T)} \log\left(\frac{T}{X_{ir}}\right)$$

has the Chi-Square distribution with 2n degrees of freedom.

This result may be used in the usual fashion to test hypotheses on the true value of β.

When n is moderate in size then one may use the fact that

$$\frac{\beta \sum_{r=1}^{K} \sum_{i=1}^{N_r(T)} \log\left(\frac{T}{X_{ir}}\right) - n}{\sqrt{n}}$$

is approximately normally distributed with mean 0 and variance 1 to test hypotheses on the true value of β.

Confidence Bounds on β

To construct exact confidence bounds on β one may again use the result that conditioned on N(T)=n

$$2\beta \sum_{r=1}^{K} \sum_{i=1}^{N_r(T)} \log\left(\frac{T}{X_{ir}}\right)$$

has the Chi-Square distribution with 2n degrees of freedom. Exact (1-α)x100 percent lower and upper confidence bounds are

$$LCB = \hat{\beta}\left(\frac{\chi^2(\frac{\alpha}{2}, 2n)}{2n}\right)$$

$$UCB = \hat{\beta}\left(\frac{\chi^2(1-\frac{\alpha}{2}, 2n)}{2n}\right)$$

respectively, where $\hat{\beta}$ is the MLE of β, and $\chi^2(\frac{\alpha}{2}, 2n)$ $[\chi^2(1-\frac{\alpha}{2}, 2n)]$ is the $\frac{\alpha}{2}$-th $[1-\frac{\alpha}{2}$ -th] percentile for the Chi-Square distribution with 2n degrees of freedom. When n is moderate, however, the normal approximation may be used. This approximation yields (1-α)x100 percent lower and upper confidence bounds

$$LCB = \hat{\beta}\left(1-\frac{P_{\alpha/2}}{\sqrt{n}}\right)$$

$$UCB = \hat{\beta}\left(1+\frac{P_{\alpha/2}}{\sqrt{n}}\right)$$

respectively, where $\hat{\beta}$ is the MLE of β, and $P_{\alpha/2}$ is the $\frac{\alpha}{2}$-th percentile for the normal distribution with mean 0 and variance 1.

Example

Consider again the simulated results presented in Table 1 when K=3 and T=200.

The MLE of β was computed to be

$$\hat{\beta} = 0.615.$$

Conditioned on N=36, 90 percent approximate confidence bounds on β are

$$\text{LCB} = \hat{\beta}(1 - \frac{1.645}{6}) = 0.446$$

$$\text{UCB} = \hat{\beta}(1 + \frac{1.645}{6}) = 0.784.$$

Hypotheses Tests and Confidence Bounds on λ (β known).

Suppose β is equal to some known value β_o, say, and that K systems have operated for time T during development. Then the random number of failures, N(T), for the K systems during [0,T] has the Poisson distribution with mean

$$\theta = K\lambda T^{\beta_o}.$$

This result may be used to test hypotheses or construct confidence bounds on λ when β is known.

Example

Assume K=3 systems were operated for time T=200 and N(T)=36 failures were observed. Suppose, also, that β is known to equal 0.5. Two-sided 95 percent confidence bounds on θ are (25.1, 49.8). Consequently, two-sided 95 percent

confidence bounds on λ are (λ_1, λ_2), where

$$\lambda_1 = \frac{25.1}{K\,T^{\beta}} = .348$$

$$\lambda_2 = \frac{49.8}{K\,T^{\beta}} = .691.$$

Confidence Bounds on Current Failure Rate and MTBF.

Let N(T)=n and let

$$\theta_1 = \frac{\chi^2(\frac{\gamma}{2}, 2n)}{2}$$

and

$$\theta_2 = \frac{\chi^2(1-\frac{\gamma}{2}, 2n+2)}{2}$$

be $(1-\gamma)$x100 percent lower and upper confidence bounds, respectively, on

$$\theta = K\lambda T^{\beta}.$$

Also, let β_1 and β_2 be $(1-\alpha)$x100 percent lower and upper (conditioned on N=n) confidence bounds on β. Then $(1-\alpha)(1-\gamma)$x100 percent lower and upper (conservative)* confidence bounds on the failure rate, r(T), at time T are

$$r_1 = \frac{\beta_1 \theta_1}{KT}$$

$$r_2 = \frac{\beta_2 \theta_2}{KT}$$

*That is, our assurance is at least, instead of exactly equal to, a specified value that the parameter of interest will lie within the stated bounds.

respectively. Consequently, $(1-\alpha)(1-\gamma)x100$ percent lower and upper (conservative) confidence bounds on the MTBF at time T, M(T), are

$$M_1 = \frac{1}{r_2} = \frac{KT}{\beta_2 \theta_2}$$

$$M_2 = \frac{1}{r_1} = \frac{KT}{\beta_1 \theta_1}.$$

Example

Consider again the simulated results presented in Table 1 when K=3 and T=200. Approximate 95 percent upper confidence bound on β is

$$\beta_2 = 0.784.$$

Also, based on N=36 failures, 97.5 percent upper confidence bound on

$$\theta = K\lambda T^{\beta}$$

is

$$\theta_2 = 49.8.$$

Hence, $(95.0)x(97.5) = 92.625$ percent (conservative) upper confidence bound on the failure rate

$$r(T) = \lambda\beta T^{\beta-1}$$

at time T is

$$r_2 = .065.$$

Consequently, 92.625 percent (conservative) lower confidence bound on the MTBF

$$M(T) = \frac{1}{r(T)}$$

at time T is

$$M_1 = \frac{1}{r_2} = 15.385.$$

REFERENCE

[1] Duane, J. T., Learning Curve Approach to Reliability Monitoring, IEEE Transactions on Aerospace, Vol. 2, No. 2, 1964.

MODIFIED PROPAGATION OF ERRORS WITH APPLICATIONS TO MAINTAINABILITY AND AVAILABILITY

Paul C. Cox
Quality Assurance Office
White Sands Missile Range, New Mexico

ABSTRACT. A modification of the conventional method of "Propagation of Errors" is proposed. This modified method promises to have numerous applications, is frequently more easily applied than conventional propagation of errors, and for a few functions of random variables which have been studied, provides improved approximations of confidence limits over conventional propagation of errors as well as over other well known methods. Modified Propagation of Errors (MPE) is described, applications to "mean time to repair" and "availability" are illustrated, and the extent of error caused by using MPE is discussed. Finally, to illustrate another application, MPE is used to approximate confidence limits for system reliability from confidence limits for component reliability.

1. INTRODUCTION .

a. The method of "Propagation of Errors," by which the variance of a function of variables is determined from the variances of the individual variable is well known. After obtaining the variance, it is then possible to at least approximate errors, confidence limits, and levels of significance for the function of variables. This discussion will be centered around determining confidence limits for a function of variables.

b. In the event of a linear function of independent, random, normal variables, the function is also normal; and there is no error in the variance obtained by propagation of errors, assuming there is no error in the individual variances. It follows that the concept of propagation of errors is very useful when evaluating a linear function of independent sample means.

c. Propagation of errors is frequently used when the function is not linear and/or the variables are not normal. Since the application of propagation of errors is usually followed by an assumption of normality for the function, the procedure can be expected to result in an error in the confidence limits approximated by this method. It is the purpose of MPE to broaden the application of propagation of errors and to reduce the error when certain variables are not normal. The procedure proposed to modify the propagation of errors has the following characteristics:

(1) Simple in concept and easy to apply.

(2) Can be applied to a large variety of functions of independent variables--there are many functions in which MPE can easily be applied but for some of these functions it may be extremely difficult to apply conventional propagation of errors.

(3) In almost all areas studied, MPE provided results which were as good or better than those provided by conventional propagation of errors. In numerous cases, MPE provided almost errorless estimates while large errors were noted when using conventional propagation of errors.

d. While MPE is applicable to a wide variety of functions of independent variables (the main requirement being that confidence limits can be obtained for each variable), the extent of error must be determined on a case-by-case basis. A number of applications have been studied. For example, the old problem of obtaining confidence limits for a system if the confidence limits for the components are known is discussed in Section 10. However, this paper is primarily concerned with the application of MPE to approximating confidence limits for mean time to repair (MTTR) and availability (A). Fortunately certain recent reports, reference e-h, have provided tables of exact confidence limits, thus providing a means for determining the error when approximating confidence limits by MPE.

2. THE METHOD.

a. MPE is applicable to virtually any function of any type of random, independent variables, as long as confidence limits can be obtained for all of the random variables within the function. The method of MPE will be described and compared with conventional propagation of errors. The two methods will be illustrated using a linear function y of three random, normal variables (x_1, x_2, and x_3), and two sided 90% confidence limits for the mean μ_y.

b. The method of conventional propagation of errors.

Let $y = f(x_1, x_2, x_3, \ldots, x_k)$

and $\hat{y} = f(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k)$

where $\hat{y}$ is some type of average

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \ldots + \left(\frac{\partial f}{\partial x_k}\right)^2 \sigma_{x_k}^2$$

if y is a linear function:

$$y = a_1x_1 + a_2x_2 + a_kx_k$$

$$\mu_y = a_1\mu_{x_1} + a_2\mu_{x_2} + \cdots + a_k\mu_{x_k} \quad \text{and}$$

$$\bar{y} = a_1\bar{x}_1 + a_2\bar{x}_2 + \cdots + a_k\bar{x}_k \quad \text{also}$$

$$\sigma_y^2 = a_1^2\sigma_{x_1}^2 + a_2^2\sigma_{x_2}^2 + \cdots + a_k^2\sigma_{x_k}^2$$

$$s_y^2 = a_1^2 s_{x_1}^2 + a_2^2 s_{x_2}^2 + \cdots + a_k^2 s_{x_k}^2$$

$$s_{\bar{y}}^2 = \frac{a_1^2}{n_1} s_{x_1}^2 + \frac{a_2^2}{n_2} s_{x_2}^2 + \cdots + \frac{a_k^2}{n_k} s_{x_k}^2$$

and the values of $\bar{y}$, the appropriate variance, and the assumption of normality are used to determine (or approximate) the desired confidence limits.

c. The method of modified propagation of errors (MPE).

$$d_h^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 h_1^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 h_2^2 + \cdots + \left(\frac{\partial f}{\partial x_k}\right)^2 h_k^2$$

$$d_\ell^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \ell_1^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \ell_2^2 + \cdots + \left(\frac{\partial f}{\partial x_k}\right)^2 \ell_k^2$$

where

$$h_i = (\mu cl)_i - \hat{x}_i \text{ [1]}$$

$$\ell_i = \hat{x}_i - (lcl)_i,$$

and where $\hat{x}_i$ is the desired average (mean, ratio, median, etc.); $(\mu cl)_i$ and $(lcl)_i$ are the upper and lower confidence limits respectively associated with the random variable x_i.

[1] MPE replaces σ with an interval in the propagation of errors formula. Another example of the use of this concept may be found on P. 91 in reference k.

Then:

Upper confidence limits = $\hat{y} + d_h$

Lower confidence limits = $\hat{y} - d_\ell$.

d. An example of obtaining confidence limits for the mean when y is a linear function of $x_1, x_2, \ldots, x_k$.

Given: $y = 3x_1 + 4x_2 + 5x_3$. The x's are random normal variables and mutually independent.

$\bar{x}_1 = 10 \quad \sigma_{x_1} = 4 \quad n_1 = 9$

$\bar{x}_2 = 12 \quad \sigma_{x_2} = 5 \quad n_2 = 16$

$\bar{x}_3 = 16 \quad \sigma_{x_3} = 6 \quad n_3 = 25$

To find 90%, 2 sided confidence limits for μ_y.

(1) Using conventional propagation of errors.

$\bar{y} = 3 \cdot 10 + 4 \cdot 12 + 5 \cdot 16 = 158$

$\sigma^2_{\bar{y}} = \frac{9}{9} \cdot 16 + \frac{16}{16} \cdot 25 + \frac{25}{25} \cdot 36 = 77$

$\sigma_{\bar{y}} = 8.7750$

90%, 2 sided confidence limits:

$158 - 1.645 \cdot 8.775 \leq \mu_y \leq 158 + 1.645 \cdot 8.775$

$143.565 \leq \mu_y \leq 172.435$

(2) Using MPE

$h_1 = \ell_1 = 1.645(4/3) = 2.1933$

$h_2 = \ell_2 = 1.645(5/4) = 2.0563$

$$h_3 = \ell_3 = 1.645(6/5) = 1.9740$$

$$d_h = d_\ell = \sqrt{(3 \cdot 2.1933)^2 + (4 \cdot 2.0563)^2 + (5 \cdot 1.9749)^2}$$

$$= \sqrt{208.3626} = 14.435$$

90% 2 sided confidence limits:

$$158 \pm 14.435$$

$$143.565 \leq \mu_y \leq 172.435.$$

e. In the above example, the confidence limits are exact and it really made no difference whether the conventional or MPE method was used. This is because y is a linear function of the x's and because each x is normally distributed. The advantages of MPE become evident when the variables are not normal, the function is not linear, and the confidence limits are only approximated.

3. APPLICATION OF MPE TO OBTAIN CONFIDENCE LIMITS FOR AVAILABILITY.

a. Assume that time to failure is distributed as the expon ntial, then MTBF is distributed as the χ^2. Assume that down time is distrib·ed as the log normal.

b. There appears to be no known solution to the problem of obtaining confidence limits for availability, using the above assumptions. The following exceptions are known:

(1) References c and d provide tables and procedures for confidence limits for availability under the assumption that σ_z^2 is known. The tables provided by reference c are brief and may require involved interpolation.

(2) Reference i provides a solution if time to failure is distributed as the exponential.

c. From 2e, Appendix A.

$$B = \exp[\ln(1/\mu_y) + (1/2)\sigma_z^2 + \mu_z]$$

$$\hat{B} = \exp[\ln(1/y) + (1/2)s_z^2 + \bar{z}]$$

$$\ln(1/\bar{y}) + (1/2)s_z^2/2 + \bar{z} = -2.3026 + 0.3750 + 0.9163 = -1.0113$$

The first step is to obtain estimates and 90% 2 sided C.L. for

$$\ln(1/\mu_y) + (1/2)\sigma_z^2/2 + \mu_2$$

d. $\ln(1/\bar{y}) = -2.3026$

From 3b, Appendix A, 90% C. L.:

$0.05217 \leq 1/\mu_y \leq 0.16055$; taking natural logs:

$-2.9533 \leq \ln(1/\mu_y) \leq -1.8292$

$h_1 = -1.8292 - (-2.3026) = .4734$

$\ell_1 = -2.3026 - (-2.9533) = .6507$

e. $s_z^2/2 = .3750$ (1f, Appendix A).

Using the χ^2 with 8 d/f, 90% -2 sided C.L.

$0.1935 \leq \sigma_z^2/2 \leq 1.0977$

$h_2 = 1.0977 - .3750 = .7277$

$\ell_2 = 0.3750 - .1935 = .1815$

f. $\bar{z} = 0.9163$, $s_z = 0.8660$, z is normally distributed (1e, f, App A).

$t_{.05}$ (8 d/f) = 1.8595

$h_3 = \ell_3 = (t \cdot s)/\sqrt{n} = 0.5368$

g. $d_h = \sqrt{(.4734)^2 + (.7277)^2 + (.5368)^2} = 1.0207$

$d_\ell = \sqrt{(.6507)^2 + (.1815)^2 + (.5368)^2} = 0.8628$

$\hat{B} = \exp(-1.0113) = 0.3637$

Upper C.L.: $\exp(-1.0113 + 1.0207) = 1.0094$

Lower C.L.: $\exp(-1.0113 - .8628) = 0.1535$

90% C.L. for B: $0.1535 \leq B \leq 1.0094$

h. Estimate of Availability:

$$\frac{1}{1 + \hat{B}} = \frac{1}{1 + .3637} = .7332$$

90% -2 sided C.L. for availability.

$$\frac{1}{1+1.0094} = 0.4977 \leq A \leq \frac{1}{1+0.1535} = .8669.$$

4. CONFIDENCE LIMITS FOR AVAILABILITY WHEN σ_z^2 IS KNOWN.

a. As stated in Section 3b(1), a solution may be obtained, using tables provided by reference c, if σ_z^2 is known and if values for m and n are within the scope of these tables. A small section of this table is included in Appendix B, and the problem, illustrated in Section 3, will be reworked by MPE using $\sigma_z^2 = s_z^2 = 0.75$ and the results compared with the exact values obtained by using the tables in reference c.

b. Solution by MPE.

(1) $\hat{B} = \exp(-2.3026 + 0.3750 + 0.9163) = \exp(-1.0113) = 0.3637$ as in Section 3g.

(2) $h_1 = 0.4734$; $\ell_1 = 0.6507$ as in Section 3d.

(3) $h_2 = \ell_2 = 0$, since σ_z^2 does not vary.

(4) $h_3 = \ell_3 = (t \cdot \sigma)/\sqrt{n} = (1.645 \cdot 0.8660)/3 = 0.4749$

(Note that the normal t is used here instead of the student t as in Section 3f).

(5) $d_h = \sqrt{(.4734)^2 + (.4749)^2} = .6704$

$d_\ell = \sqrt{(.6597)^2 + (.4749)^2} = .8055$

(6) 90% -2 sided C.L. for B:

Upper C.L.: $\exp(-1.0113 + .6704) = .7111$

Lower C.L.: $\exp(-1.0113 - .8055) = .1626$

c. Solution using tables, reference c.

(1) C.L. for $B = \dfrac{\exp(\sigma_z^2/2) \cdot b \cdot \bar{x}_g}{2m\bar{y}}$,

where the appropriate values for b are obtained from the tables of reference c. (Note Appendix B for an extract from this table).

(2) 90% -2 sided C.L. for B:

From Appendix B, $b_{.05} = 7.995$; $b_{.95} = 34.852$

Upper C.L. for $B = \dfrac{(1.455)(34.853)(2.50)}{2 \cdot 9 \cdot 10} = 0.7043$

Lower C.L. for $B = \dfrac{(1.455)(7.995)(2.50)}{2 \cdot 9 \cdot 10} = 0.1616$

d. 90% -2 sided C.L. for availability:

By MPE: $.5844 \leq A \leq .8602$

By Ref c: $.5868 \leq A \leq .8610$

5. CONFIDENCE LIMITS FOR MEAN TIME TO REPAIR (MTTR).

a. Frequently, the assumption is made that time to repair is distributed as the exponential. If this be the case, confidence limits

for MTTR can be obtained in the same way as for MTBF (Note 3b of Appendix A). This is discussed in detail in reference i.

b. The usual assumption is that time to repair is distributed as the log normal. Using this assumption, the four reports by Charles E. Land, references e-h, provide the necessary procedures and tables for obtaining the confidence limits. The tables listed under reference h are necessary for computing the confidence limits and are extremely comprehensive. They are presently unpublished,[1] and therefore are generally unavailable. A brief extract is included in Appendix B.

c. If time to repair (x) is distributed as the log normal, and $z = \ln x$, then an estimate of MTTR = $\exp(\bar{z} + s_z^2/2) = \exp(0.9163+0.3750) = \exp(1.2913) = 3.6375$.

d. Referring to the extract from Charles Land's tables in Appendix 2, for $s_z = 0.866$, using linear interpolation, -0.5725 is obtained for .05 and 1.0431 for .95. These values are multiplied by s_z, giving -0.4958 and 0.9033.

e. Lower C.L. = $\exp(1.2913 - 0.4958) = 2.2155$
Upper C.L. = $\exp(1.2913 + 0.9033) = 8.9770$

f. The solution by MPE:

From Section 3e, $h_2 = .7277$ and $\ell_2 = .1815$

From Section 3f, $h_3 = \ell_3 = 0.5368$

$d_\ell = \sqrt{(.1815)^2 + (.5368)^2} = 0.5667$

$d_h = \sqrt{(.7277)^2 + (.5368)^2} = 0.9043$

Lower C.L. = $\exp(1.2913 - 0.5667) = 2.0637$
Upper C.L. = $\exp(1.2913 + 0.9043) = 8.9850$

g. Comparing these results indicates a conservative error for MPE in each case, and a negligible error for the more important case of upper confidence limits.

6. CONFIDENCE LIMITS FOR AVAILABILITY IF μ_y IS KNOWN.

a. Assume that MTBF = 10 hrs, as before, but this value was obtained from long history instead of a small sample.
Then $\mu_y = 10$ and B = (0.10)(MTTR).

[1]Dr. Land is presently negotiating with certain statistical journals, and it is expected that these tables will be published before the end of 1973, probably considerably reduced in size.

b. $\hat{B} = (0.10)(3.6375) = 0.36375$, as in Para 3f.

c. Multiplying confidence limits for MTTR by 0.10, gives 90% -2 sided C.L. for B:

by Land's Tables, $0.2216 \leq B \leq 0.8977$

and by MPE, $0.2064 \leq B \leq 0.8985$

d. 90% confidence limits for availability by Land's Tables:

$0.5270 \leq A \leq 0.8186$

by MPE, $0.5267 \leq A \leq 0.8289$.

e. The error resulting from using MPE is again negligible in the important (lower C.L.) case.

7. COMBINING LAND'S TABLES AND MPE TO OBTAIN CONFIDENCE LIMITS FOR AVAILABILITY.

a. If the tables by Charles Land, reference h, are available, it appears that the error in obtaining confidence limits for availability would be reduced to a minimum if Land's Tables were combined with MPE to obtain the confidence limits.

b. The problem of Section 3 will be reworked by combining these two methods.

c. From Section 3d, $h_1 = 0.4734$ and $\ell_1 = 0.6507$.

d. The application of Land's Tables in the solution of this problem, can be obtained from Section 5d.

$$h_{2,3} = 0.9033 \text{ and } \ell_{2,3} = 0.4958$$

e. $d_h = \sqrt{(.4734)^2 + (.9033)^2} = 1.0197$

$d_\ell = \sqrt{(.6507)^2 + (.4958)^2} = 0.8180$

f. From Para 3f, $\hat{B} = \exp(-1.0113) = 0.3637$

Then, 90% -2 sided C.L. for B:

$\exp(-1.0113-.8180) \leq B \leq \exp(-1.0113 + 1.0197)$

$0.1605 \leq B \leq 1.0084$.

g. 90%, 2 sided C.L. for availability:

$$\frac{1}{1+1.0084} = 0.4979 \leq A \leq \frac{1}{1+0.1605} = 0.8617$$

8. COMPARISON OF CONFIDENCE LIMITS BY VARIOUS METHODS.

a. The following table summarizes the 90% - 2 sided confidence limits for availability, obtained in the previous paragraphs.

CONDITIONS	METHOD	SEC.	90% - 2 sided C.L. for B		90% - 2 sided C.L. for Availability	
			LOWER	UPPER	LOWER	UPPER
None	MPE	3g,h	0.1535	1.0094	0.4977	0.8669
"	MPE&LAND	7f,g	0.1605	1.0084	0.4979	0.8617
σ_z^2	MPE	4b,d	0.1626	0.7111	0.5844	0.8602
"	Ref c	4c,d	0.1616	0.7043	0.5868	0.8610
μ_y Known	MPE	6c,d	0.2064	0.8985	0.5267	0.8289
"	Ref h	6c,d	0.2216	0.8977	0.5270	0.8186

Table 1. Comparison of 90% - 2 sided confidence limits, obtained by various methods.

b. A study of the above table suggests the following:

(1) The use of MPE usually provides conservative approximations. That is to say, MPE approximations appear to be a little larger for upper C.L. and a little smaller for lower C.L. than the true value.

(2) If the root sum square of the differences between the MPE approximations and the actual confidence limits is obtained, when σ_z^2 and μ_z^2 are known, the following is determined:

RSS, lower: $\sqrt{(.0024)^2 + (.0003)^2} = .0024$

RSS, upper: $\sqrt{(.0008)^2 + (.0103)^2} = .0103,$

which suggests that the errors in the MPE approximations for the problem of Section 3 may be less than .01 for the upper C.L. and less than .0024 for the lower C.L.

9. ERROR IN THE USE OF MPE.

a. It is clear that the usefulness of MPE depends upon its accuracy as well as its ease of application. In Section 8, certain errors related to the specific example of this report were stated. In this Section, MPE values will be compared with exact values for a wide spectrum of parameters, using data from references c, f, and h to provide the exact comparison.

b. First, a comparison will be made for confidence limits for availability for the special case in which σ_z^2 is known. For this study, availability will be approximated by MPE, using the methods of Section 4. These approximate values will then be compared with exact values obtained by using the tables of reference c. Table 2 shows the error in using MPE for select values of m = n, n/σ_z^2, and levels of probability and for the special case of $\bar{y} = 4\bar{x}_g$.

		PROBABILITY					
m = n	n/σ_z^2	.95	.90	.75	.25	.10	.05
5	5	.005	.007	.008	.004	.002	.002
	12	.003	.004	.004	.002	.000	-.002
	40	.001	.001	.002	.000	.000	-.002
9	5	.004	.005	.007	.005	.003	.001
	12	.002	.003	.004	.002	.002	.001
	40	.001	.001	.001	.001	.001	.000
13	5	.003	.004	.005	.005	.003	.002
	12	.002	.003	.003	.002	.001	.001
	40	.001	.001	.001	.001	.000	.000

Table 2. [A (exact) - A(MPE)], σ_z^2 known and $\bar{y} = 4\bar{x}_g$.

(1) From Table 2, it appears that under the conditions of this paragraph, and if m and n both exceed 5, the error should not exceed .008 if availability is approximated by MPE. As n/σ_z^2 increases, the error rapidly approaches zero. As m or n increases, the error approaches zero, but rather slowly. Errors appear to be smaller when probabilities are close to 0 or 1; and for lower confidence limits, the error is conservative.

(2) To investigate the error if $\bar{y} \neq 4\bar{x}_g$, the worst case situation in Table 2 was selected; i.e., $m = n = n/\sigma_z^2 = 5$ and prob. = 0.75. Here it was found that for some examples of $\bar{y} < 4x_g$, the error became as large as .009. As the ratio of y to $4\bar{x}_g$ increases, the error slowly decreases. For example, when the ratio reaches 20, the error is .03.

(3) From the comparisons of this Section, it appears that the error resulting from the use of MPE when determining confidence limits for availability is less than .01, in the special case in which σ_z^2 is known. MPE has the following additional advantages.

(a) Can be used when the tables of reference c are unavailable.

(b) Can be used when parameters that are not included in reference c. For example, these tables provide entrees only for $5 \leq m \leq 13$; have no values for 2 sided -95% confidence limits; and provide a limited selection of values for n/σ_z^2.

(c) MPE may be applied when σ_z^2 can only be approximated from a sample.

c. The second study of this Section is to provide a comparison of the data contained in the tables of reference h, and the corresponding values if obtained by MPE. Furthermore, values obtained by the method of minimum variance unbiased estimators (MVUE) will be compared with corresponding exact and MPE values. The technique of MVUE was developed by Dr. Land at the suggestion of Prof. D. R. Cox. Dr. Land has compared his exact values with several well known approximations, and generally MVUE provided better approximations than any of the other procedures. Furthermore, MVUE is [1] essentially equivalent to the procedure of conventional propagation of errors.

(1) Tables 3 and 4 are adaptations of two tables prepared by Dr. Charles Land and included in reference f. In this reference, Land compares exact values of confidence limits for $\ln\mu_x = \mu_z + \frac{1}{2}\sigma_z^2$ with the MVUE approximation

[1] MVUE uses $s_w^2 = s_z^2/n + s_z^4/2(n+1)$, conventional propagation of error uses $s_w^2 = s_z^2/n + (n-1)s_z^4/2n^2$, where $w = \bar{z} + (\frac{1}{2})s_z^2$.

as well as three other approximations. Tables 3 and 4 copy Land's data for exact and MVUE values and includes the values obtained by MPE. Table 3 is for $\bar{z} = 1.219$ and $s_z^2 = 0.208$. Table 4 is for $\bar{z} = 7.650$ and $s_z^2 = 4.632$.

LEVEL	n = 11			n = 101			n = 1001		
	EXACT	MPE	MVUE	EXACT	MPE	MVUE	EXACT	MPE	MVUE
.005	.950	.883	.952	1.205	1.200	1.200	1.285	1.284	1.284
.010	.990	.938	.988	1.216	1.212	1.212	1.288	1.288	1.288
.025	1.045	1.012	1.041	1.232	1.230	1.230	1.294	1.293	1.293
.050	1.091	1.069	1.086	1.247	1.245	1.245	1.298	1.298	1.298
.100	1.142	1.130	1.139	1.263	1.262	1.262	1.304	1.304	1.304
.250	1.227	1.225	1.226	1.291	1.291	1.291	1.313	1.313	1.313
.500	1.324		1.323	1.323		1.323	1.323		1.323
.750	1.431	1.432	1.420	1.356	1.356	1.355	1.333	1.333	1.333
.900	1.546	1.541	1.507	1.387	1.386	1.384	1.343	1.343	1.342
.950	1.629	1.619	1.560	1.405	1.404	1.401	1.348	1.348	1.348
.975	1.712	1.698	1.605	1.422	1.420	1.416	1.353	1.353	1.353
.990	1.829	1.809	1.658	1.442	1.439	1.434	1.359	1.359	1.358
.995	1.925	1.900	1.727	1.456	1.453	1.446	1.363	1.362	1.362

Table 3. One sided confidence limits for $\ln\mu_x = \mu_z + \frac{1}{2}\sigma_z^2$, comparing MPE with exact values from Land's Tables and MVUE.

$\bar{z} = 1.219$ and $s_z^2 = .208$.

LEVEL	n = 11			n = 101			n = 1001		
	EXACT	MPE	MVUE	EXACT	MPE	MVUE	EXACT	MPE	MVUE
.005	7.774	7.480	7.012	9.119	9.096	8.965	9.666	9.664	9.647
.010	8.-62	7.576	7.298	9.191	9.173	9.062	9.693	9.692	9.678
.025	8.297	8.096	7.718	9.300	9.288	9.204	9.735	9.733	9.725
.050	8.514	8.398	8.080	9.397	9.390	9.327	9.771	9.770	9.763
.100	8.787	8.723	8.496	9.515	9.511	9.468	9.813	9.812	9.808
.250	9.319	9.312	8.193	9.724	9.724	9.704	9.885	9.885	9.883
.500	10.083		9.966	9.977		9.966	9.967		9.966
.750	11.151	11.176	10.740	10.257	10.258	10.228	10.052	10.052	10.050
.900	12.555	12.568	11.436	10.[illegible]35	10.534	10.464	10.131	10.131	10.125
.950	13.712	13.717	11.853	10.716	10.712	10.606	10.179	10.179	10.170
.975	14.999	14.995	12.214	10.882	10.876	10.728	10.222	10.220	10.209
.990	16.950	16.938	12.634	11.086	11.077	10.870	10.272	10.271	10.254
.995	18.658	18.634	12.920	11.234	11.223	10.967	10.307	10.305	10.285

Table 4. One sided confidence limits for $\ln\mu_x = \mu_z + (\frac{1}{2})\sigma_z^2$, comparing the MPE approximation with exact values from Land's tables[2] and MVUE.

$\bar{z} = 7.650$ and $s_z^2 = 4.632$.

(a) Table 3 provides comparative data for an example for which $s_z^2 = 0.208$ is relatively small, and the following is evident:

1. For n = 100, either MPE or MVUE should provide a satisfactory approximation.

2. For n = 11 and the probability level less than 0.25,[2] MVUE is somewhat better than MPE.

3. For n = 11 and the level greater than 0.75, MPE provides a much better fit than MVUE.

4. As will be shown soon, MVUE is generally under 0.25 and frequently very poor at levels above 0.75. MPE is usually at its best at levels above 0.75. For most applications, the higher levels provide the more useful levels of confidence.

(b) Table 4 uses much larger values for both $\bar{z}$ and s_z^2 than table 3. For this example, MPE provides better results than MVUE in almost all instances. Specifically, for n = 11 and for levels $\geq$ .750, MPE provides excellent approximations while the use of MVUE results in large errors.

(2) Table 5 contains a sample of data for n = 3 taken from reference h (Tables by Charles Land),[3] and compares these data with corresponding values obtained by the methods of MVUE and MPE. Table 6 provides these comparisons in graphical form for levels 0.90 and 0.005. From these tables, it appears that even for the very small sample of 3, MPE provides a close approximation if the level $\geq$ 0.25. MVUE provides acceptable approximations in some areas, frequently better than MPE. However, MVUE generally does not provide satisfactory approximations for any value of s for levels > 50%.

(3) Table 7 provides a graphical comparison at 4 levels for exact, MPE, and MVUE values for n = 11. These graphs suggest that for a sample of this size, MPE is clearly superior to MVUE for almost all levels, and MPE should be a reasonably satisfactory approximation for any level of probability greater $\geq$ 0.05.

---

[2]Levels under 0.50 correspond to lower confidence limits for maintainability and upper confidence limits for availability (note Sections 5 and 6). Levels above 0.50 correspond to the reverse.

[3]Data in reference h are multiplied by s_z and added to $\bar{z} + (\frac{1}{2})s_z^2$ to obtain confidence limits for $\mu_z + (\frac{1}{2})\sigma_z^2$, assuming z is normal. Appendix B contains an extract from these tables for n = 9.

S	METHOD	.005	.010	.025	.050	.100	.250	.750	.900	.950	.975	.990	.995
	EXACT	4.104	3.136	2.113	1.506	1.012	.457	.491	1.192	1.945	3.088	5.889	10.364
00	MPE	5.730	4.021	2.484	1.686	1.089	.471	.487	1.168	1.921	3.134	7.589	11.482
	MVUE	1.490	1.346	1.134	.951	.741	.390	.390	.741	.951	1.134	1.346	1.490
	EXACT	2.271	1.919	1.478	1.162	.858	.438	.655	2.114	4.593	9.608	24.618	49.623
0	MPE	5.733	4.026	2.491	1.694	1.098	.477	.779	2.383	4.903	9.872	25.410	50.079
	MVUE	1.555	1.408	1.183	.993	.774	.407	.407	.774	.993	1.183	1.405	1.555
	EXACT	1.753	1.546	1.269	1.053	.825	.453	1.150	4.211	9.231	19.240	49.297	99.208
00	MPE	5.744	4.040	2.511	1.718	1.125	.492	1.326	4.377	9.362	19.269	49.833	99.649
	MVUE	1.744	1.575	1.327	1.114	.868	.457	.457	.868	1.114	1.327	1.575	1.744
	EXACT	1.622	1.489	1.296	1.130	.926	.503	2.429	8.474	18.488	38.494	98.495	
00	MPE	5.787	4.096	2.589	1.813	1.227	.548	2.523	8.548	18.494	38.296	99.167	
	MVUE	2.351	2.124	1.789	1.501	1.170	.616	.616	1.170	1.502	1.789	2.124	
	EXACT	2.381	2.264	2.086	1.865	1.568	.789	6.172	21.221				
00	MPE	6.078	4.472	3.081	2.370	1.785	.841	6.214	21.225	46.075	95.572		
	MVUE	4.790	4.326	3.645	3.059	2.383	1.254	1.254	2.383	3.059	3.645		
	EXACT	3.277	3.140	2.898	2.632	2.224	1.107	9.273	32.204				
50	MPE	6.488	4.978	3.694	3.014	2.384	1.147	9.305	31.814	69.086	143.33		
	MVUE	6.990	6.313	5.319	4.464	3.478	1.830	1.830	3.478	4.464	5.319		
	EXACT	4.233	4.068	3.768	3.431	2.905	1.440	12.371					
00	MPE	7.021	5.612	4.411	3.733	3.031	1.471	12.400	42.407	92.103	191.095		
	MVUE	9.227	8.334	7.021	5.893	4.591	2.416	2.416	4.591	5.893	7.021		
		NEGATIVE						POSITIVE					

ble 5. Comparison of MPE and MVUE with exact values extracted from reference h (n = 3).

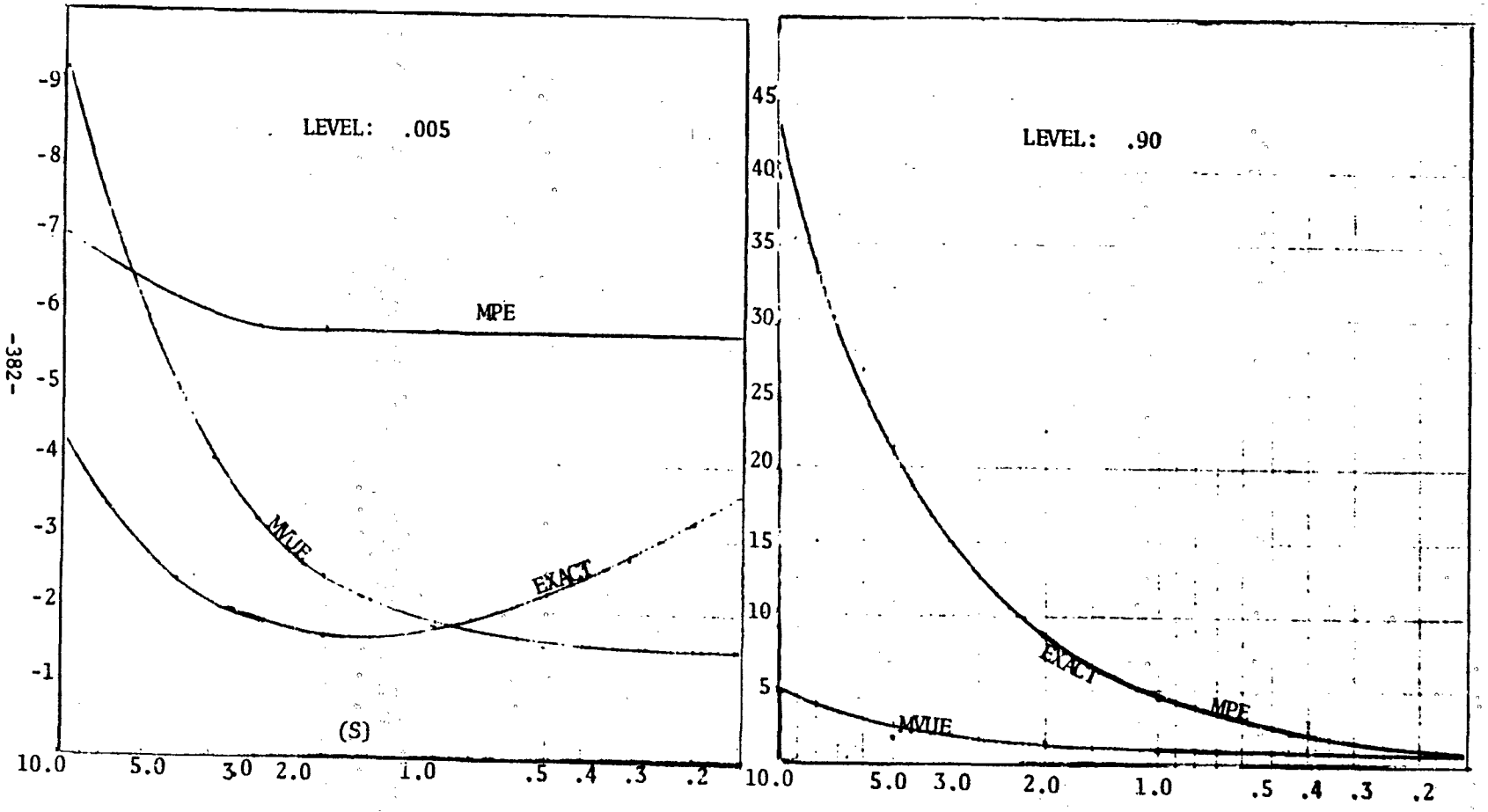


TABLE 6

Comparison of EXACT, MPE, and MVUE, n = 3 and for levels .005 and .90 (Exact values from Reference h).

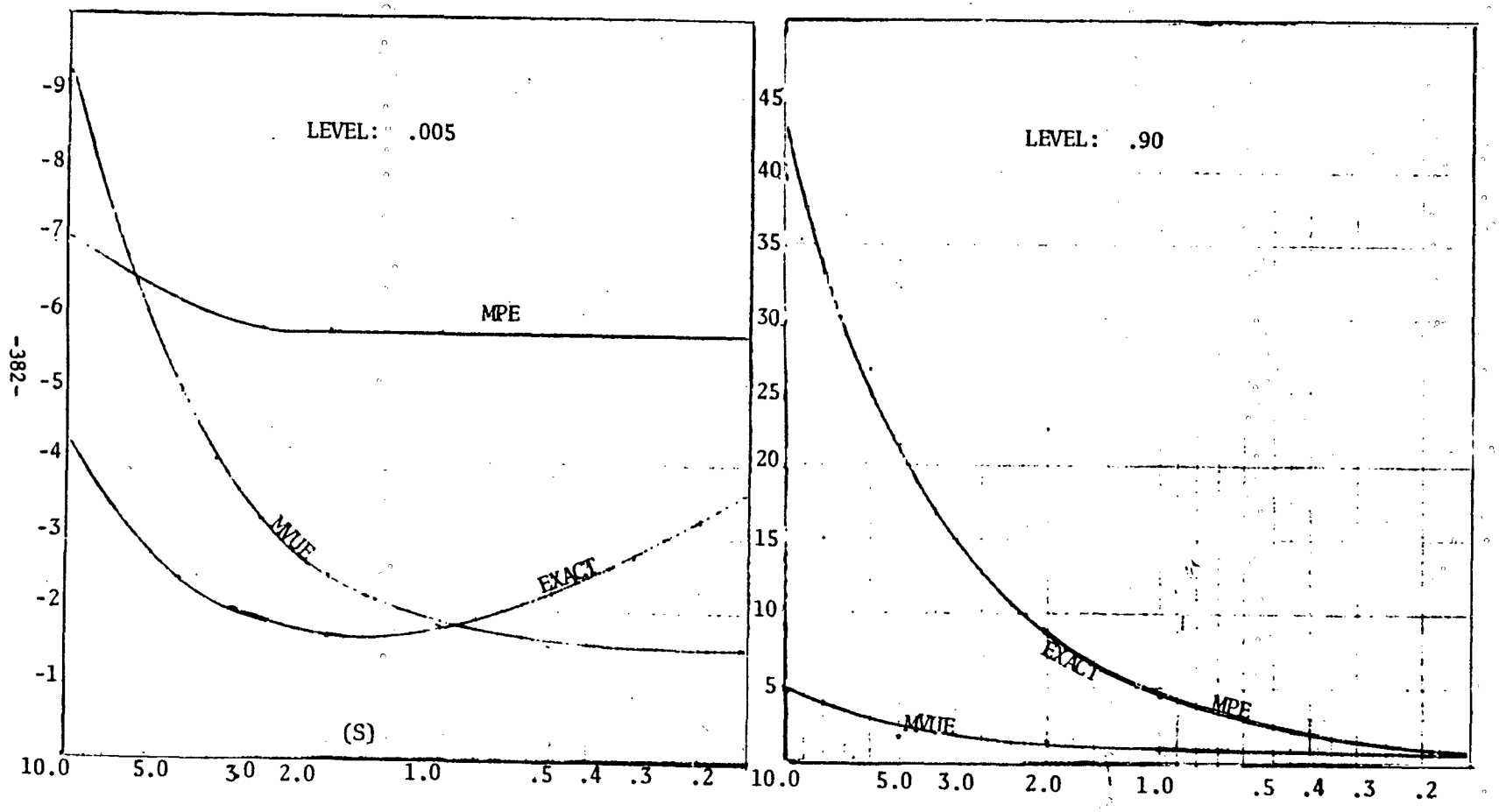


TABLE 6

Comparison of EXACT, MPE, and MVUE, n = 3 and for levels .005 and .90 (Exact values from Reference h).

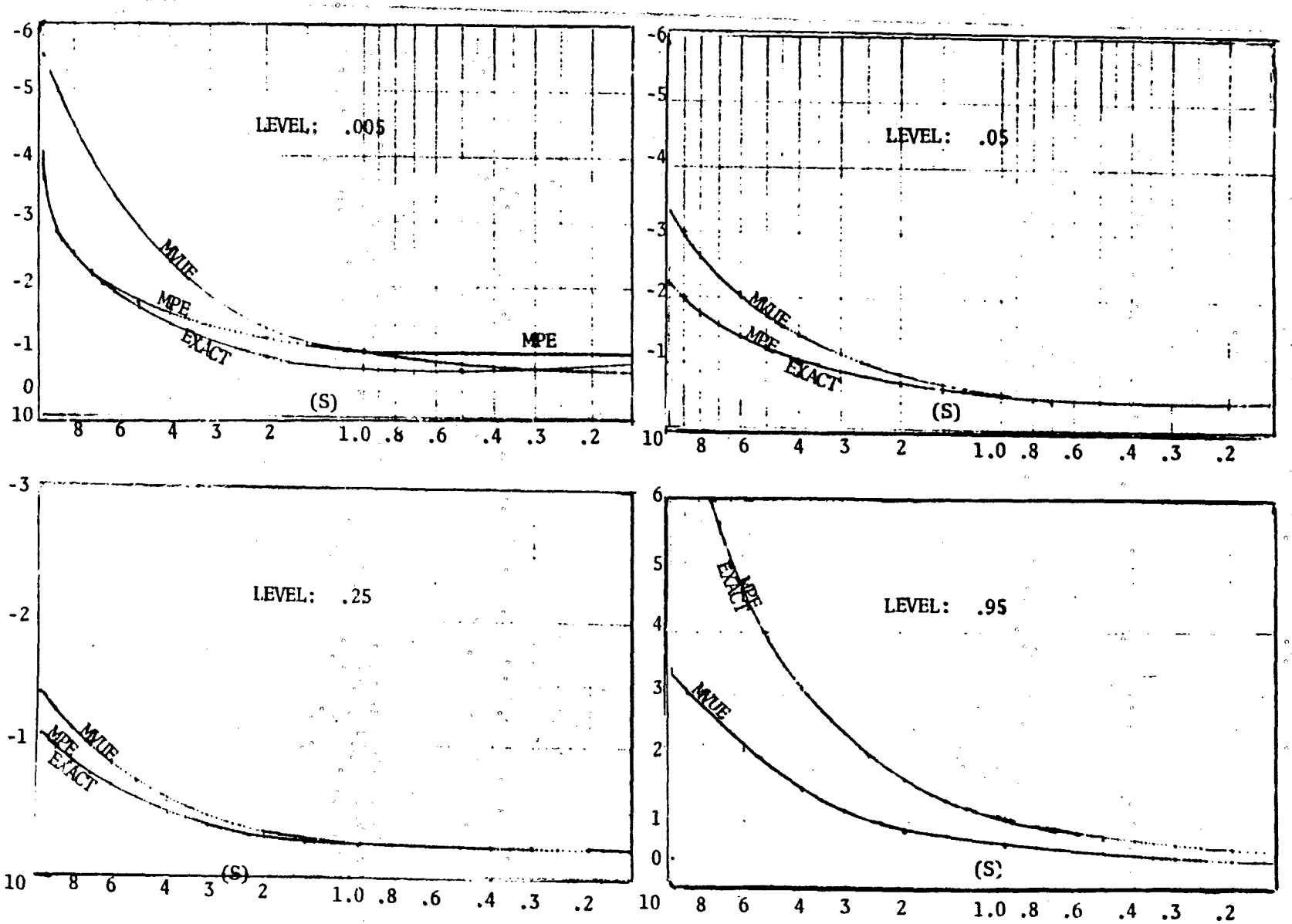


TABLE 7

Comparison of EXACT, MPE, and MVUE, n = 11 and for four probability levels (Exact values from Reference h).

(4) Using the criteria that the relative error for MPE shall not exceed 0.1, the following table indicates generally safe areas for the application of MPE.

SAMPLE SIZE	LEVELS OF PROB. GREATER THAN OR EQUAL
3	.250
6	.100
11	.050
25	.010
50	.005
Above 50	No restriction

d. From the discussion of this Section, the following conclusions are drawn about the errors resulting from the use of MPE.

(1) In the event that σ_z^2 is known and MPE is used to approximate the methods of reference c (note Section 4 and 9b of this report), the error can be expected to be less than .01; for lower confidence limits, the MPE error is conservative, thus the true lower confidence limit can be expected to be slightly larger than the approximation. For upper confidence limits, the error is not conservative but is usually very small.

(2) If μ_y is known and MPE is used to approximate confidence limits for availability, as discussed in Section 6, both upper and lower approximations will almost always be conservative, and in the very few instances in which not conservative, the error will be small. If the regions in which the relative error may exceed 0.1 are avoided [note para 9c(4)], the method of MPE should provide errors no larger than .02 in the approximate confidence limits for availability.

(3) If the RSS of the maximum errors, discussed in the two previous paragraphs, is obtained, it would suggest that the method of MPE, when applied to the total problem of approximating confidence limits for availability (as described in Section 3), should provide an error which is conservative and will not exceed 0.025.

10. CONFIDENCE LIMITS FOR SYSTEM RELIABILITY FROM COMPONENT RELIABILITY DATA.

a. MPE will be used to obtain approximate confidence limits for a system which consists of three components in series and rel ·bility estimates for each component has been obtained in three separate t·sts. It is not claimed that MPE is the best method for obtaining these approximate confidence limits, but it is offered to illustrate this method when the function is a product of independent variables.

b. Assume the results of the three component tests are these provided by the following table.

	COMP. ONE (C_1)	COMP. TWO (C_2)	COMP THREE (C_3)
No. Tested (n_i)	50	40	30
Successes (s_i)	42	38	27
Est. of Comp. Rel(r_i)	0.840	0.950	0.900
90% - 2sided C. L.	$.730 \leq \rho_1 \leq .920$	$.851 \leq \rho_2 \leq .991$	$.761 \leq \rho_3 \leq .973$
ℓ_i	.840 - .730=.110	.950-.851=.099	.900-.761=.139
h_i	.920-.840=.080	.991-.950=.041	.973-.900=.073
Est. of Systems Reliability (R) = (.840)(.950)(.900) = .718			

c. Since $R = r_1.r_2.r_3$, the formulas of para 2b,c become:

$$s_R^2 = r_2^2.r_3^2.s_{r_1}^2 + r_1^2.r_3^2.s_{r_2}^2 + r_1^2.r_2^2.s_{r_3}^2$$

$$d_h^2 = r_2^2.r_3^2.h_1^2 + r_1^2.r_3^2.h_2^2 + r_1^2.r_2^2.h_3^2$$

$$d_\ell^2 = r_2^2.r_3^2.l_1^2 + r_1^2.r_3^2.l_2^2 + r_1^2.r_2^2.l_3^2$$

d. One question arises, when dealing with a nonlinear function, what should be substituted for the r_i, since they are unknown? The estimate may be used or the confidence limit itself might be substituted. Some studies have indicated that substitution of confidence limits usually gives better results. Both procedures will be illustrated.

e. Using the estimate of r_i.

$$d_h^2 = (.95 \cdot .90 \cdot .08)^2 + (.84 \cdot .90 \cdot .041)^2 + (.84 \cdot .95 \cdot .073)^2 = .00904; \; d_h = .095$$

$$d_\ell^2 = (.95 \cdot .90 \cdot .11)^2 + (.84 \cdot .90 \cdot .041)^2 + (.84 \cdot .95 \cdot .139)^2 = .02675; \; d_\ell = .164$$

90% - 2 sided C.L. for the system:

$$.718 - .164 = .554 \leq R \leq .718 + .095 = .813$$

f. Using confidence limits for r_i.

$$d_h^2 = (.991 \cdot .973 \cdot .080)^2 + (.920 \cdot .973 \cdot .041)^2 + (.920 \cdot .991 \cdot .073)^2 = .01172; \; d_h = .1083$$

$$d_\ell^2 = (.851 \cdot .761 \cdot .11)^2 + (.730 \cdot .851 \cdot .139)^2 + (.730 \cdot .851 \cdot .139)^2 = .01556; \; d_\ell = .1247$$

90% - 2 sided C.L. for the system:

$$.718 - .125 = .593 \leq R \leq .718 + .108 = .826$$

g. An alternate approach, eliminating the dilemma of a proper value to be substituted for r_1, r_2, and r_3, is to use logarithms and thus transform the given function into a linear function.

$$\log R = \log r_1 + \log r_2 + \log r_3.$$

$$\text{est. of } \log R = -.0757 - .0223 - .0458 = -.1438.$$

log of lower C.L.	log of upper C.L.
log .730 = -.1367	log .920 = -.0362
log .851 = -.0701	log .991 = -.0039
log .761 = -.1186	log .973 = -.0119

ℓ_1 = -.0757+.1367=.0610; h_1=-.0562+.0757=.0395

ℓ_2 = -.0223+.0701=.0478; h_2=-.0039+.0223=.0184

ℓ_3 = -.0458+.1186=.0728; h_3=-.0119+.0458=.0339

$d_\ell^2 = (.0610)^2+(.0478)^2+(.0728)^2=.01130$ $d_\ell=.1063$

$d_h^2 = (.0395)^2+(.0184)^2+(.0339)^2=.00305$ $d_h=.0552$

lower C.L. = antilog (-.1438-.1063) = .562
upper C.L. = anitlog (-.1438+.0552) = .815

11. REFERENCES

a. Scarbrough, James B., Numerical Mathematical Analyses,6th Edition, 1966, Johns Hopkins Press, Baltimore, Md.

b. Ku, Harry H., Notes on the Use of Propagation of Error Formulas, 1965, National Bureau of Standards Report No. 9011.

c. Gray, H. L. and Lewis, T. O., A Confidence Interval for the Availability Ratio, Technometrics, Vol. 9, No. 3, August 1967, Pages 465-471.

d. Gray, H. L. and Schucany, W.R., Lower Confidence Limits for Availability Assuming Lognormally Distributed Repair Times, IEEE Transactions on Reliability, Vol. R-18, No. 4, November 1969, Pages 157-162.

e. Land, C. E. (1971), Confidence Intervals for Linear Functions of the Normal Mean and Variance. Ann. Math-Stat. 42, 1187-1205.

f. Land, C.E. (1972) An Evaluation of Approximate Confidence Interval Estimation Methods for Lognormal Means, Technometrics, 14, 145-158.

g. Land, C.E. (1971) Tables of Confidence Limits for Linear Functions of the Normal Mean and Variance, TR#25, Oregon State University, Nov 1971.

h. Land,C.E. (1971) Tables of Confidence Limits for Linear Functions of the Normal Mean and Variance. Unpublished Tables.

i. Nelson, W. B., Statistical Methods for Reliability, Maintainability, and Availability Data, General Electric Research and Development Center, Schenectady, N.Y., Report No. 68-C-020, February 1968.

j. Nelson, W. B., A Statistical Prediction Interval for Availability, General Electric Research and Development Center, Schenectady, N.Y., Report No. 70-C-277, August 1970.

k. Bowker, A. H. and Lieberman, G. J., Engineering Statistics, 2nd Ed., 1972, Prentice-Hall, Englewood Cliffs, New Jersey.

APPENDIX A

SYMBOLS AND DEFINITIONS AND DATA FOR THE EXAMPLE

1. General Definitions and Data.

   a. m = no. of failures = 9

   b. n = no. of repairs = 9

   c. t = total operating time = 90 hours

   d. x = repair time

   e. z = ln repair time--assume that z is normally distributed

   f. $\bar{z} = 0.9163$; $s_z^2 = 0.75$; $s_z = 0.8660$

   g. $\bar{x}_g = \exp(\bar{z})$ = geometric mean of the sample of x's = 2.50 hours

   h. $\exp(s_z^2/2) = \exp(0.375) \approx 1.455$

   i. estimate of mean time to repair (MTTR) = $\exp(\bar{z}+s_z^2/2) = 3.6375 \approx \bar{x}$

   j. y is time to failure. Assume that y is distributed as the exponential

   k. $\bar{y} = t/m$ = sample mean time between failure (MTBF) = 10 hours

   l. μ_x = population mean time to repair = $\exp(\mu_z + \sigma_z^2/2)$

   m. μ_y = population mean time between failure

   n. C.L.--confidence limits

2. Availability Formulas.

   a. $A = \mu_x/(\mu_y + \mu_x)$

   b. Est. of $A = \hat{A} = \bar{y}/(\bar{y} + \bar{x})$

   c. $A = 1/(1 + B)$, where $B = \mu_x/\mu_y$

d. Est. of $B = \hat{B} = \overline{x}/\overline{y} = (1/\overline{y})\ \exp(s_z^2/2 + \overline{z})$

e. $B = (1/\mu_y)\ \exp(\sigma_z^2/2 + \mu_z) = \exp[\ln(1/\mu_y) + \sigma_z^2/2 + \mu_z]$

f. $\hat{B} = \exp[\ln(1/\overline{y}) + s_z^2/2 + \overline{z}]$

3. Reliability (MTBF) formulas.

a. Consider: $1/\overline{y} = m/t = 9/90 = 0.10$; $\ln(1/\overline{y}) = -2.3026$

b. Confidence limits for $1/\mu_y$ (m is fixed, t is variable)

$$\frac{\chi^2_{2m;\ \alpha/2}}{2T} \leq 1/\mu_y \leq \frac{\chi^2_{2m;\ (1\ -\ \alpha/2)}}{2T}$$

for 90% - 2 sided C.L. $9.39/180 \leq 1/\mu_y \leq 28.9/180$

$$0.05217 \leq 1/\mu_y \leq 0.16055$$

c. If t is fixed and m is variable, then confidence limits for MTBF become:

$$\frac{\chi^2_{2m;\ \alpha/2}}{2T} \leq 1/\mu_y \leq \frac{\chi^2_{(2m\ +2);\ (1\ -\ \alpha/2)}}{2T}$$

APPENDIX B

1. Extract from table listed in reference 3. ($n/\sigma_z^2 = 9 \div 0.75 = 12$):

g/m →	8	9	10
.050	6.865	7.995	9.171
.100	8.227	9.541	10.855
.250	11.072	12.674	14.279
.500	15.211	17.201	19.191
.750	20.653	23.116	25.577
.900	26.960	29.941	32.910
.950	31.516	34.852	38.173

2. Extract from the tables of reference 8 (Charles Land). The extract is for n-1 = 8 d/f.

s ↓	.025	.05	.10	.90	.95	.975
.50	-.6783	-.5657	-.4404	.5763	.8050	1.0475
.60	-.6736	-.5646	-.4419	.6115	.8620	1.1329
.70	-.6721	-.5658	-.4451	.6508	.9254	1.2264
.80	-.6733	-.5691	-.4497	.6939	.9945	1.3272
.90	-.6769	-.5743	-.4556	.7402	1.0682	1.4340
1.00	-.6825	-.5811	-.4627	.7896	1.1452	1.5455