ARO Report 82-2

PROCEEDINGS OF THE TWENTY-SEVENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Sponsored by

The Army Mathematics Steering Committee

on Behalf of

PROCEEDINGS OF THE TWENTY-SEVENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Sponsored by
The Army Mathematics Steering Committee
on Behalf of

THE CHIEF OF RESEARCH, DEVELOPMENT AND ACQUISITION

U. S. Army Research Office

Report No. 82-2

June 1982

Conference on the Design of Experiments in Army Research Development and Testing.

PROCEEDINGS OF THE (TWENTY-SEVENTH) CONFERENCE ON THE DESIGN OF EXPERIMENTS

Sponsored by the Army Mathematics Steering Committee

HOST

U. S. Army Research Office
Research Triangle Park, North Carolina

HELD AT

The McKimmon Center for Continuing Education
North Carolina State University
Raleigh, North Carolina
21-23 October 1981

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

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

FOREWORD

The Twenty-Seventh Conference on the Design of Experiments in Army Research, Development and Testing was held in the Jane S. McKimmon Center on the campus of North Carolina State University, Raleigh, on 21-23 October 1981. The U. S. Army Research Office served as host for this meeting. The Army Mathematics Steering Committee (AMSC) continues to be the sponsor for this series of meetings. Members of this committee would like to thank Dr. Robert L. Launer for serving as Chairperson on Local Arrangements, and Mrs. Sherry Duke, who handled many of the administrative details. These individuals did an outstanding job of carrying out the many tasks associated with conducting a conference of this size.

Each year the Program Committee is asked to select invited speakers who can discuss in an informative and stimulating manner statistical areas of current interest. At least one of the speakers, who has expertise in areas of current interest to the Army, is asked to present new developments in these fields. The selection criteria were certainly met by the gentlemen giving the talks in the General Sessions. The names of the invited speakers and their topics are noted below.

Speaker and Affiliation	Title of Address
Professor Norman L. Johnson University of North Carolina	RECENT TRENDS IN DISCRETE DISTRIBUTIONS
Professor Nozer D. Singpurwalla George Washington University	ROBUSTNESS OF SEQUENTIAL EXPONENTIAL LIFE TESTING PROCEDURES FOR RESTRICTED CLASSES OF DISTRIBUTIONS
Professor Douglas A. Wolfe Ohio State University	COMPARING SEVERAL GROUPS IN A TWO-WAY LAYOUT SETTING
Professor David C. Hoaglin Harvard University	APPLICATION OF EXPLORATORY DATA ANALYSIS TECHNIQUES IN MORE COMPLEX MODELS
Professor Walter L. Smith University of North Carolina	ASYMPTOTIC BEHAVIOR OF CUMULATIVE PROCESSES

In addition to the two invited addresses on the first day of this meeting, there were five solicited talks on product assurance. These were delivered by Army scientists that are specialists in this area. Another event associated with this meeting was a tutorial seminar on "Quality Control". It was held just preceding the conference on 19-20 October and was given by Professors P. M. Ghare and D. R. Jensen of the Virginia Polytechnic Institute and State University. This course gave the standard procedures for monitoring a process by variables or by attributes in the case of a single quality characteristic.

The winner of the first Wilks Award for Contributions to Statistical Methodologies in Army Research, Development and Testing* was presented to Professor Robert E. Bechhofer of Cornell University at a luncheon on the first day of the conference. He richly deserves this honor for his many scientific contributions to ranking and selection procedures as well as other statistical areas. He has given freely of his time to help Army scientists develop statistical skills. Recently he solved a very important problem in ballistic testing related to kinetic energy penetrators.

Members of the AMSC would like to take this opportunity to express their thanks to Mr. Philip G. Rust of Thomasville, Georgia for endowing both of the Wilks Awards. His generous gifts in memory of his friend, Sam Wilks, will contribute to the welfare of the military services as well as foster statistical science in general.

*For more information on this award see "Open letter on the New Wilks Award". It is printed following this foreword.

OPEN LETTER ON THE NEW WILKS AWARD

TO: Dr. Jagdish Chandra, Chairman of the Army Mathematics Steering Committee (AMSC)

FROM: Dr. Robert L. Launer

This letter is being written as a result of recent developments concerning the establishment of a new Wilks award. The reason for bringing this matter to your attention will become clear.

The Samuel S. Wilks Memorial Medal and Award was initiated in 1964 with a $5000 gift from Mr. Philip B. Rust of Thomasville, Georgia. This award, presented each year at the Army Design of Experiments Conference, has been coordinated and administered by the American Statistical Association, because of the difficulties involved at that time in administering the funds internally. Mr. Rust intended that this award be given to a statistician for contributions to Army technology. Unfortunately, he did not compose a specific citation which would be used as a guide in choosing the annual winner of the award. A complete record of the ceremonial remarks made at the time of the initial award are given in the proceedings of the Tenth Design of Experiments Conference.

Questions have been raised about some of the recent winners of the Wilks award. These were communicated to Dr. Ralph Bradley of Florida State University and President of the American Statistical Association. As the result of the ensuing protracted conversations, Mr. Rust made another gift for the establishment of a second Wilks award, and to augment the endowment for the original Wilks award.

The Army Gifts Fund Office (AGFO), under the Office of Secretary of the Army (AR1-100), acts as the custodian of gifts to the Army. In the case of monetary gifts the AGFO channels the funds through the commander of an Army installation having close ties to the donors expressed purpose of the award. Since the Wilks award is for academic or scientific excellence, the Adjutant General of the Army has established that the Army Research Office (ARO) can serve as such an installation. The Commander of ARO has agreed to have the funds channeled through his Finance and Accounting Office and to have the other aspects of this award handled by the AMSC, an intra-Army committee.

In order that the new Wilks award begin on a firm basis, with the first award given in 1981 if possible, and to avoid unforeseen difficulties and criticisms, it is proposed that the new Army Wilks award be administered in accordance with the following points.

1) The award should be called "The Wilks Award for Contributions to Statistical Methodologies in Army Research, Development and Testing".

2) The award winner should be chosen by a majority vote of an ad-hoc committee of five statisticians. The committee should consist of at least two statisticians from the Army community and at least two statisticians from academia, each serving for two-year terms. The initial committee should be appointed by the Chairman of the AMSC. Thereafter, the subcommittee on Statistics and Probability shall annually nominate one new member. The winner of the award may replace one member of the selection committee for the following year. The Chairman of the AMSC may act as a non-voting chairman of the selection committee, or he may appoint a representative to act in his stead (voting or non-voting).

3) The Chairman of the selection committee should convene the selection committee early enough that the award can be made at the annual Design of Experiments Conference.

4) Nominations of candidates for the award may be forwarded to any member of the selection committee. Solicitations may also be made through announcement letters for the annual DOE conference, or any other appropriate method.

5) The chairman of the selection committee should serve as the coordinator of the award money. This should be in the form of a check made payable to the winner of the award.

TABLE OF CONTENTS*

*This table of contents contains only the papers that are published in this technical manual. For a list of all papers presented at the Twenty-Sixth Conference on the Design of Experiments, see the Program of the meeting.

A G E N D A

for the

TWENTY-SEVENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN

ARMY RESEARCH, DEVELOPMENT AND TESTING

21-23 October 1981

Host: U. S. Army Research Office

Location: Jane S. McKimmon Center
North Carolina State University
Raleigh, North Carolina 27650

***** Wednesday, 21 October *****

0815-0915	REGISTRATION -- Lobby, McKimmon Center
0915-0930	CALLING OF THE CONFERENCE TO ORDER WELCOMING REMARKS COL Duff G. Manges, Commander, U. S. Army Research Office
0930-1200	<u>GENERAL SESSION I</u> <u>Chairman</u> - Francis G. Dressel, U. S. Army Research Office
0930-1030	<u>KEYNOTE ADDRESS</u> RECENT TRENDS IN DISCRETE DISTRIBUTIONS Norman L. Johnson, University of North Carolina-Chapel Hill
1030-1100	BREAK
1100-1200	ROBUSTNESS OF SEQUENTIAL EXPONENTIAL LIFE TESTING PROCEDURES FOR RESTRICTED CLASSES OF DISTRIBUTIONS Nozer D. Singpurwalla, George Washington University
1200-1330	LUNCH AND WILKS AWARD PRESENTATION

1330-1500 SPECIAL PRODUCT ASSURANCE SESSION

Chairman - Ray Bell, U. S. Army Materiel Systems Analysis Activity

ON THE AMSAA RELIABILITY GROWTH MODEL FOR ONE-SHOT SYSTEMS

Larry H. Crow and William P. Clay, U. S. Army Materiel Systems Analysis Activity

RELIABILITY GROWTH EXPERIENCE FOR U. S. ARMY MISSILE SYSTEMS

Robert Mioduski, U. S. Army Materiel Systems Analysis Activity

1330-1500 TECHNICAL SESSION I - "STOCHASTIC PROCESSES"

Chairman - Grady Miller, U. S. Army Materiel Systems Analysis Activity

COMPUTING THE STEADY STATE OF A TWO-DIMENSIONAL POSITIVE RANDOM WALK

Mary Anne Maher, US Army White Sands Missile Range

EVALUATION OF PARAMETERS IN ARMA ANALYSIS OF TIME SERIES BY A LEAST-CHI SQUARE METHOD

Richard L. Moore and F. Luzzi, U. S. Army Armament Research and Development Command

AN APPLICATION OF RENEWAL THEORY TO SOFTWARE RELIABILITY

Leonard A. Stefanski, U. S. Army Research Office and University of North Carolina-Chapel Hill

1530-1700 SPECIAL PRODUCT ASSURANCE SESSION (cont'd)

CONFIDENCE BOUNDS ON THE MISSION RELIABILITY OF SYSTEMS WITH NON-CONSTANT FAILURE RATE

George M. Hanna et al., U. S. Army Materiel Systems Analysis Activity

METHODOLOGY FOR ESTIMATING MISSION AVAILABILITY AND RELIABILITY FOR A MULTIMODAL SYSTEM

Henry P. Betz, U. S. Army Materiel Systems Analysis Activity

BLACK HAWK RELIABILITY PREDICTIONS FOR PRODUCTION AIRCRAFT

Clarke Fox, U. S. Army Materiel Systems Analysis Activity

1530-1700 TECHNICAL SESSION II - "STATISTICAL SAMPLING"

Chairman - Donald S. Burdick, Duke University

SOLUTION TO A CLASS OF TWO-PHASE SAMPLING PROBLEMS

Patrick D. Allen and Stephen M. Rasey, U. S. Army Concepts Analysis Activity

EMPIRICAL AND THEORETICAL COMPARISON OF SEVERAL VARIANCE ESTIMATORS IN SURVEY SAMPLING

Chien-Fu Wu, Mathematics Research Center

DESIGN OF A MULTIPLE SAMPLE WESTENBERG TYPE TEST FOR SMALL SAMPLE SIZES

Jim Knaub, U. S. Army Logistics Center

***** Thursday, 22 October *****

0830-1000 TECHNICAL SESSION III - "ERROR ANALYSIS AND SIMULATION"

Chairman - Breton Graham, U. S. Army Concepts Analysis Agency

ERROR PROPAGATION IN PHYSICAL MODELS

J. Richard Moore and Jerry Thomas, Ballistic Research Laboratory

ERROR MODELS FOR RADIOSONDE HEIGHT INDEXING AND WIND VELOCITY DETERMINATIONS

Walter B. Miller, U. S. Army White Sands Missile Range

A DATA BASED RANDOM NUMBER GENERATOR FOR MULTIVARIATE DISTRIBUTIONS

James R. Thompson, Rice University and Malcolm Taylor, Ballistic Research Laboratory

0830-1000 CLINICAL SESSION A

Chairman - Walter Foster, Armed Forces Institute of Pathology

Panelists - William E. Baker, Ballistics Research Laboratory
David C. Hoaglin, Harvard University
Donald B. Rubin, Mathematics Research Center

TERRAIN CLASSIFICATION FOR INTERVISIBILITY PURPOSES

Warren Olson and Donald McCoy, U. S. Army TRADOC Systems Analysis Activity

IRREGULARITIES IN THE ERROR ANALYSIS OF A PIECE-WISE CONTINUOUS FUNCTION

Paul H. Thrasher, U. S. Army White Sands Missile Range

1030-1200 TECHNICAL SESSION IV - "RELIABILITY AND FAILURE ANALYSIS"

Chairman - Edward W. Ross, U. S. Army Natick R&D Laboratories

INFERENCE ON A FUTURE RELIABILITY PARAMETER WITH THE WEIBULL PROCESS MODEL

Grady Miller, U. S. Army Materiel Systems Analysis Activity

A COMPARISON OF METHODS FOR LOWER CONFIDENCE LIMITS IN THE RELIABILITY OF SERIES SYSTEMS

Bernard Harris and Andrew Soms, University of Wisconsin-Madison

EXAMINATION OF SIZE EFFECTS IN THE FAILURE PREDICTION OF CERAMIC MATERIAL

Donald M. Neal and Edward M. Lenoe, U. S. Army Materials and Mechanics Research Center

1030-1200 CLINICAL SESSION B

Chairman - Carl B. Bates, U. S. Army Concepts Analysis Agency

Panelists - Robert Bechhofer, Cornell University
Douglas A. Wolfe, Ohio State University

SPECIFYING A DETECTABLE 3-FACTOR INTERACTION WITH A NON-CENTRAL F

Walter D. Foster and Jack L. Wray, Armed Forces Institute of Pathology

SHOULD CRITERIA FOR FIELD TESTS BE FORMULATED AS STATISTICAL HYPOTHESES

Carl T. Russell, U. S. Army Cold Regions Test Center

1200-1330 LUNCH

1330-1500 TECHNICAL SESSION V - "STATISTICAL METHODOLOGY"

Chairman - Larry H. Crow, U. S. Army Materiel Systems Analysis Activity

STATISTICAL TESTING OF LARGE COMPLEX COMPUTER SIMULATION MODELS

Carl B. Bates, U. S. Army Concepts Analysis Agency

SELLING A COMPLICATED EXPERIMENTAL DESIGN TO THE FIELD TEST OPERATOR

Carl T. Russell, U. S. Army Cold Regions Test Center

1330-1500 CLINICAL SESSION C

Chairman - Jerry Thomas, Ballistic Research Laboratory

Panelists - J. Richard Moore, Ballistic Research Laboratory
Nozer D. Singpurwalla, George Washington University
Chien-Fu Wu, Mathematics Research Center

LONG-TERM STORAGE OF ARMY RATIONS

Edward W. Ross, Jr., U. S. Army Natick R&D Laboratories

ASSURING QUALITY THROUGH BALLISTIC TESTING

Michael McMiller, U. S. Army Aviation Research and Development Center

1530-1700 GENERAL SESSION II

Chairman - Daniel L. Solomon, North Carolina State University

COMPARING SEVERAL GROUPS IN A TWO-WAY LAYOUT SETTING

Douglas A. Wolfe, Ohio State University

APPLICATION OF EXPLORATORY DATA ANALYSIS TECHNIQUES IN MORE COMPLEX MODELS

David C. Hoaglin, Harvard University

***** Friday, 23 October *****

0830-1000 CLINICAL SESSION D

Chairman - Richard L. Moore, U. S. Army Armament Research and Development Command

Panelists - Bernard Harris, Mathematics Research Center
Edward A. Saibel, U. S. Army Research Office
Walter L. Smith, University of North Carolina-Chapel Hill

A MATHEMATICAL BASIS FOR TRACKING MANEUVERING AIRCRAFT WITH DOPPLER RADAR

Donald W. Rankin, U. S. Army White Sands Missile Range

FORCES DUE TO ICE-STRUCTURE INTERACTION

Devinder Singh Sodhi, U. S. Army Cold Regions Research and Engineering Laboratory

1030-1200 GENERAL SESSION III

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

OPEN MEETING OF THE AMSC SUBCOMMITTEE ON PROBABILITY AND STATISTICS

ASYMPTOTIC BEHAVIOR OF CUMULATIVE PROCESSES IN DISCRETE TIME

Walter L. Smith, University of North Carolina-Chapel Hill

1200 ADJOURN

* * * * * * *

PROGRAM COMMITTEE

Carl Bates	Robert Launer
Larry Crow	Douglas Tang
Walter Foster	Malcolm Taylor
Frank Grubbs	Jerry Thomas
Bernard Harris	Langhorne Withers

RECENT DEVELOPMENTS IN DISCRETE DISTRIBUTIONS

N.L. Johnson
University of North Carolina
Chapel Hill, North Carolina

1. Introduction. In accordance with the title of this talk, I plan to give a (very) concise description of work, (very) nearly all on discrete distributions, published in the period 1969-81. This seems a rather strange topic for a "keynote" speech, so I will provide some explanation.

My major interest is to illustrate some personal attitudes towards research results. Many fields could serve as sources of illustrative material. I have chosen discrete distributions because, for about the last two years, Dr. Samuel Kotz and I have been collecting material on this topic for an article for the International Statistical Review (ISR). So it is a field wherein I feel at home, and some of the illustrative material may be of interest to some of you, even if my comments are not.

The bibliography for the article contains nearly 700 entries, so you will not be surprised to learn that the Editor of the ISR has asked us to reduce the length of the article to about 20% of the original (though he has generously agreed to publish the whole of the bibliography). So there is (for me) a second attraction - the opportunity to draw attention to a few matters which will be mentioned inadequately - if at all - in the published article.

In order to make it possible to cover a considerable amount of material, I will make use of a number of summaries in note-form, of which you should have copies. Much of my talk will consist of comments on these.

The collection and classification of the material was done jointly by Dr. Kotz and myself, but all expressions of opinion are my own.

2. Prejudices. Before coming to details, a few words about prejudices. It is, unfortunately, true that many of us - most, I suppose - sometimes react negatively to the mere title, or general content of a report or paper. (Let me admit to an unreasoning, and unreasonable hostility to papers on characterizations and on Least Squares - plain or modified). We may be influenced by fashion - a need to appear in accord with currently dominant prejudices - or we may, in our own experience, have found the topics to be unrewarding and so, by some sort of extrapolation decide they are of little importance generally.

Here are two examples of ways in which prejudice can arise - the first reflecting types of attitude which are still not unknown; the second relevant to our present topic.

1) (R. and J. Peto, replying to discussion of their paper "Asymptotically Efficient Rank Invariant Test Procedures" - J.R. Statist. Soc. 135A (1972), p. 205).

"Every analysis of survival data should either be efficient against Lehmann alternatives or have a very clear reason for not being so; the Lehmann family is the "normal" distribution of survival theory."

(My italics. No evidence for the last statement is presented.)

A more balanced (and honest) approach is provided by N. Mantel in "Evaluation of Survival Data..." in Cancer Chemotherapy Reports, 50, (1966) p. 167.

"It is unlikely that in any real instance in which the two force-of-mortality functions $Z_1(t)$ and $Z_2(t)$ differ that any simple relationship exists between them. In principle, however, it is possible to determine the power...for alternatives - $Z_1(t) = kZ_2(t)$, $k \neq 1$." (These are Lehmann alternatives.)

2) From a review by E.J. Williams of RCSW, Vols. 2/3 (see References - III) in Australian J. Statist., (1973), p. 183:

"After reading some of the contributions, this reviewer would question whether, at this stage of development of the statistical art:

(1) finding a distribution that fits well to data is a contribution to statistical science, and

(2) the study of a particular form of statistical distribution is a contribution to statistical theory."

I am not claiming that these authors were unjustified in their specific comments, but I do think that their modes of expression can lead to development of prejudice.

There will be plenty of prejudice in this talk. As a counterbalance, it is well to keep in mind that:

(a) The fact that one does not understand something is evidence neither of its value or its lack thereof, though it _may_ reflect lack of skill in exposition.

(b) It is lazy to judge a paper by its title alone, or a piece of work by its field.

(c) In judging "applicability," it is important to consider as many _aspects_ of application as possible and not just those based on personal experience, however lengthy or distinguished.

On the other hand,

(d) The appearance of a large number of papers on the same topic in a brief period of time may not signify that the topic is of lasting importance. It may just mean that it appears to offer the prospect of quick results from small-scale investigations. (This is often the genesis of a "fashion"), and finally,

(e) It is a salutary exercise to read copies of journals from 10 or more years ago and to try to assess the reasons why the articles contained therein were accepted for publication.

3. Families (Systems, Classes, etc.) of Distributions. The construction of new families of distributions still fascinates a number of workers, myself among them. Age has tempered some earlier enthusiasm, and I am now more aware that usefulness, more than novelty, is the essential property of value. "Usefulness" can be interpreted quite broadly but a new family should not be too similar to an existing one unless it is also simpler in some important respects.

I believe there is real value in suitable systems - particularly because they can assist appreciation of relationships among distributions. The requirement of suitability, however, is important. The mere variety and number of distributions included in a system are not, in themselves, measures of its importance - still less, of its practical value. (For example, the class defined by $\sum_x P_x = 1$ is very broad but contributes nothing to understanding.) What matters is the inclusion of as wide a variety as possible within as specific a formulation as possible.

(We now refer to Table 1a.) An outstanding example is the class of power series distributions (PSD). In fact, it is only the θ^x part that stops them being uselessly general, but this is enough. A competitive - and complementary - system, the factorial series distributions (FSD), has been introduced in our period.

Generalized hypergeometric series distributions have explicitly structured $b(x)$ functions. The class is very broad. Dacey (1972) lists some 50 members of the class, utilizing values of h and k not exceeding 3. In practical use it is rare to have either h or k even as great as 3. For one thing, there are $(h+k+1)$ parameters and Occam's Razor is well-established in distribution construction. ("Fashion" and/or "Prejudice"?) For another, if we consider the quantities

$$U_{x+1} = (x+1)P_{x+1}/P_x = \theta \prod_{i=1}^{h} (c_i+x)/\{\prod_{i=1}^{k} (b_i+x)\}$$

we see that (i) if h and k are not known, it is likely to be difficult to estimate them,

(ii) even if h and k are known, values of U_x for quite an extensive range of values of x will be needed to estimate θ, a_i's and b_i's.

Even supposing $\theta = 1$ (generalized hypergeometric distributions) it is very doubtful whether much is gained by increasing h and/or k above 2.

The use of U_x (and similar functions) to systematize search for an appropriate distribution, has much to commend it. Ord (1972) has provided rules for using this approach when $h = 1$ and $k = 0$ (so that U_x is a linear function of x). (See Table 1b.)

It is easy to see

(i) how (in principle) this approach can be extended to higher values of h and k, and

(ii) the considerable technical difficulties likely to arise.

Despite (ii), I think that this kind of approach, with suitable (approximate) allowance for sampling variation, can be of real value. It can also provide a way of assessing when two (or more) distributions - however attractively named - are likely to be indistinguishable in particular kinds of applications.

The modified power series distributions (MPSD) are extensions of the PSD in the sense that when $h(\theta)$ is invertible they are just PSD's. By taking $h(\theta) = \theta/g(\theta)$ the interesting <u>Lagrange distributions</u> are obtained. These use the Lagrange expansion

$$f(s) = f(0) + \sum_{j=1}^{\infty} \frac{(s/G(s))^j}{j!} D^{j-1}\{[G(t)]^j f'(t)\}\Big|_{t=0}$$

for the p.g.f. $f(s)$. $G(s)$ is also taken to be a pgf with $g'(0) < 1$. Then

$$P_0 = f(0); \quad P_x = \frac{1}{x!} D^{x-1}\{[G(t)]^x f'(t)\}\Big|_{t=0}$$

A table from Consul and Shenton (1972) shows a number of special cases.

Assessment of the "value" of this class of distributions needs some careful balancing. There are useful general formulae for moments, in terms of cumulants corresponding to $f(\cdot)$ and $G(\cdot)$. It is possible to approximate binomial and negative binomial distributions quite well with the Lagrange double Poisson (e.g. Jain (1974)).

On the other hand, the formal development does not as yet provide (for me) a very helpful background to comprehend the nature of the distributions. Some of these do arise "naturally" in queueing theory (Shenton and Consul (1973), Kumar (1981)) and in ballot theory (Narayana (1979)).

A feature of these distributions is that it is often not immediately obvious (to me) that $\sum P_x = 1$. For the "double binomial," for example, this relationship can be derived from Abel's identity. There are related identities (e.g. Riordan (1979)) from which other distributions can be concocted in a formal way. (See Table 1c). Whether they are all Lagrange distributions I do not know. In fact, I do not know a general method for deciding whether a given distribution is a Lagrange Distribution. (The "practical value" of such knowledge is at present uncertain - prejudice may suggest it would be only of intellectual value - but it would interest me.)

4. Modifications. Now we look at Table 2. There has been considerable interest in "monkeying about" with distributions. A very simple form is just increasing (or decreasing) one probability (usually P_0) and decreasing (or increasing) all the others proportionately - giving rise to "inflated" (or "deflated") distributions. Truncated distributions are limiting cases.

If the adjustments are proportional to the original P_x values we have weighted distributions with

$$P_x{}^* = w_x P_x / (\sum_y w_y P_y) \ .$$

Particular cases, such as w_x a linear function of x, so that $P_x{}^*/P_x = \alpha + \beta x$ are of interest, because one can fit a standard $\{P_x\}$ to data $\{f_x\}$ and then study the ratios $\{f_x/P_x\}$ as a function of x, which may lead to a suitable and simple modification.

"Mixing" (or "compounding") is a well-established form of modification. The table shows the notation. Note that $F_1 \underset{\tilde{\theta}}{\wedge} F_2$ is usually called a F_2-F_1 distribution (though sometimes it is called F_1-F_2). By choosing different pairs of F_1, F_2's quite a wide range of distributions can be obtained. Some relatively recent examples are described in the references. While interesting discoveries can be made from speculative F_1, F_2 pairings, it is usually more attractive when the mixing arises from a natural model. The subclass of compound Poisson (i.e. with F_1 Poisson) has been especially popular. There is a respectable reason for this (i.e. apart from mathematical simplicity). Poisson corresponds to "independence in time and/or space"; compound

Poisson can be detected by comparing data with a fitted Poisson distribution. (See e.g. Shaked (1980).)

We note, in passing, that "Random Sum" or "generalization" ($F_1 \vee F_2$) can be expressed as a mixture of convolutions

$$F_1^{*N} \wedge F_2 \quad (\text{"}F_2\text{-generalization of }F_1\text{"})$$

of F_1.

New distributions can sometimes be found as <u>limits</u> of "old" ones. Recently Sibuya (1979) obtained <u>digamma</u> and <u>trigamma distributions</u> as limits of a zero-truncated inverse Polya-Eggenberger distribution. These are just special sorts of hypergeometric distributions; of special interest because of their relation to logseries distributions.

More sweeping modifications can lead to very broad systems. The <u>Poisson modification of Bernoulli trials</u> (allowing success probability to change from trial to trial) which leads to Poisson binomial (to be distinguished from Poisson-binomial and binomial-Poisson) distribution can be modified by allowing for dependence between trials (see Table 2). The references given develop, <u>inter alia</u>, (joint) distribution(s) of total number(s) of successes. I find these studies of more use for the general nature of the results than for their specific forms.

5. <u>Damage Models and Characterizations</u>. I have already admitted to a negative attitude to work on characterizations. This is based, I

believe, on a feeling that characterization depends on a distribution being followed <u>exactly</u>; the nature and amount of departure from characterization corresponding to given departure from the assumed distribution seems to be of little account. It would be an encouraging sign if characterization research were to be accompanied by some indication of robustness.

A rather prominent example of the baleful influence of characterization is in the study of "damage models." In the simplest form, these models include three random variables X, Y and Z with X = Y + Z; X represents "undamaged" value, Y represents observed ("damaged") value, and Z represents the "damage." (X might be the number of nonconforming items in a sample; Y the number detected by an inspection process; then the "damage," Z, is the number of nonconforming items not detected in the inspection.)

In 1968, Rao and Rubin showed that <u>if</u> Y and Z are independent <u>and</u> the conditional distribution of Y, given X is binomial, then Y and Z each have Poisson distributions. Since then, there have been many variations on this theme (see Table 3), with the common feature that Y and Z are mutually independent. Is this assumption likely to be realistic? I have not seen any investigation of this.

A related topic, which <u>does</u> seem to have some virtue, was studied by Samaniego (1976). He defined "convoluted Poisson" variables

($Y = X_1 + X_2$ where X_1 and X_2 are independent and X_1 is Poisson) and finds characterizations of the distribution of Y. It is suggested that Y might represent an "overcounted" Poisson variable.

6. Approximations. Before the advent of powerful computing aids, approximations were useful because they made calculations feasible which would otherwise have been impossible. One might expect that with the present profusion of computing power there would be decreasing interest in approximations. One would be wrong. Just considering approximation of the tail probabilities of binomial by those of unit normal distributions we have the quite imposing list shown in Table 4, all published in 1968-80. Similar tables can be constructed for approximations of Poisson and hypergeometric distributions.

There are varied reasons for these phenomena. Of course, many of the computer programs, themselves, use approximations. Some approximations (such as that of a normal by a lognormal or logistic distribution) can be used to simplify more extensive theoretical analysis. Sometimes, also, it is useful to have a quick way of calculating an approximate value. Their major value (in my opinion) is their ability to present easily comprehended pictures of whole sets of results.

In addition to these, more or less valid, reasons, we are, unfortunately, left with the impression that some approximations have only their elegance, and some, even, only their novelty, to recommend them. As a general rule the approximation should be rather less complicated than the quantity being approximated. This does not seem to be true for some items in Table 4.

Sometimes it is not entirely clear in which direction the approximation is most useful. For example the Lagrange double Poisson with parameters $\lambda_1 = N\{(1+P)^{1/2}-(1+P)^{-1/2}\}$, $\lambda_2 = 1-(1+P)^{-1/2}$ and the negative binomial with parameters N,P are very similar to each other. Jain (1974) implies that the latter is a useful approximation to the former, yet many persons would consider the negative binomial "simpler" than the double Poisson.

Recently an increase in research on accuracy of established approximations has been a welcome, though often tedious and rarely elegant, feature of statistical literature. It is to be hoped that some way will be found of presenting the results of such research in both more digestible and more permanent forms (e.g. monographs), combining attractive production with careful non-partisan effectiveness.

7. <u>Concluding Remarks</u>. You may still feel that I have chosen an unfortunate topic as a vehicle for these remarks on evaluation of research. I do not think this is so. Taking the risk of undue repetitiveness, I believe similar assessments would be reached in very many, if not all fields, especially if one discounts the effect of current, transitory fashions with their new-found but often ill-founded enthusiams.

I think that we also should allow for the tendency (perhaps unconscious) to welcome the idea that much research (in other fields than one's own, and even in one's own field, by other workers) is of little permanent value, so that there is no need to spend time and effort in understanding it. It is undoubtedly true that the present organization of research effort in the world is such that there is much waste - both in redundancy and in publication piecemeal of special results of little intrinsic interest or value except "novelty." But this does not mean that we should underestimate the value of all that is published. We still tend to start with an overoptimistic idea of how much to expect from a piece of research, quickly followed by disillusionment when reality does not match up to our preconceptions. We should try to attain a realistic view of what to expect of "good" research. I suggest this means that evaluation has to be delayed for a few years (3 or more, perhaps) to see more clearly where the research results stand in relation to the general development of a subject.

In the specific context of the present talk we can say, I think, that

(1) Work on new families of discrete distributions has increased

(i) our power to construct useful mathematical frameworks

and (ii) our ability to appreciate relationships between different frameworks.

(2) Modification of distributions has been systematized, and its possibilities are becoming more clearly realized.

(3) The power and usefulness of approximations have been increased and is now more generally appreciated, and methods of assessing the accuracy of approximations are becoming better understood.

(4) Although much of the development of new multivariate discrete distributions has been rather formal, there is a slow growth in appreciation of the kinds of such distributions now available to the analyst.

REFERENCES

ABBREVIATIONS

AS: *Applications of Statistics* (Ed: P.R. Krishnaiah) North Holland, Amsterdam

NATO, Trieste: *International Summer School on Statistical Distributions in Scientific Work* (Eds: C. Taillie, G.P. Patil and B. Baldessari)

RCSW: *Random Counts in Scientific Work* (Ed: G. P. Patil) Pennsylvania State University Press, University Park, Pennsylvania

SDSW: *Statistical Distributions in Scientific Work* (Eds: G. P. Patil, S. Kotz and J. K. Ord) Reidel, Dordrecht and Boston

---

TABLE 1a

(I) Patil, G. P. (1962) *Ann. Inst. Statist. Math., Tokyo*, 14, 179-182.

(I)' ____________ (1965) *Sankhyā*, 28A, 225-237.

(II) Berg, S. (1974) *Scand. J. Statist.*, 1, 145-152; (1975) *Scand. Actu. J.*, 233-237.

________ (1979) *Dept. Statistics, Univ. Lund*, Report #11.

(II)' ________ (1977) *Scand. J. Statist.*, 4, 25-30.

(III) Patil, G. P. (1963) *Ann. Math. Statist.*, 34, 1050-1056.

(IV) Consul, P. C. and Shenton, L. R. (1972) *SIAM J. Appl. Math.*, 23, 239-248; (1974) *Comm. Statist.*, 2, 41-57.

Consul, P. C. (1981) *Comm. Statist.*, A10, 2039-2046.

Jain, G.C. (1975a) *Biom. Z.*, 17, 85-97; (1975b) *Biom. Z.*, 17, 501-506.

Shenton, L. R. and Consul, P. C. (1973) *Sankhyā*, 35A, 229-236.

Kumar, A. (1981) *Comm. Statist.*, A10, 1429-1436.

Narayana, T. V. (1979) *Lattice Path Combinatorics and Statistical Applications*, University of Toronto Press.

(V) Kemp, A. W. (1968) *Sankhyā*, 30A, 401-410.

Dacey, M. F. (1972) *Sankhyā*, 34B, 243-250.

(VI) Steyn, H. S. (1951) *Proc. Kon. Akad. Weten, Amsterdam*, 54A, 23-30; (1955) Ibid., 58A, 588-595.

Janardan, K. G. and Patil, G. P. (1972) *Sankhyā*, 27B, 158-164.

(VII) Sibuya, M. (1979) *Ann. Inst. Statist. Math., Tokyo*, 31A, 373-390.

________ (1980) *Ann. Inst. Statist. Math., Tokyo*, 32A, 25-36.

TABLE 1b

(VIII) Ord, J. K. (1972) *Families of Frequency Distributions*, Wiley, New York.

(IX) Altham, P. M. E. (1978) *Appl. Statist.*, 27, 162-167.

(X) Steyn, H. S. (1975) *S. Afr. Statist., J.*, 9, 163-172.

Doss, D. C. and Graham. R. C. (1975) *Calcutta Statist. Assoc. Bull.*, 24, 93-99.

Kawamura, K (1979) *Kodai Math. J.*, 2, 337-345.

Widdra, W. (1972) *Metrika*, 19, 68-71.

TABLE 1c

(XI) Consul, P. C. (1974) *Sankhyā*, 36B, 391-399.

Consul, P. C. and Mittal, S. P. (1975) *Biom. Z.*, 17, 67-75.

(XI)' ______________________________(1977) *Biom. Z.*, 19, 161-173.

Riordan, J. (1979) *Combinatorial Indentities*, Krieger; Huntington, New York

---

TABLE 2

(XII) Grzegorska, L. (1973a) *Ann. Univ. Maria Curie-S., Lublin*, 27A, 19-29; (1973b) Ibid., 31-42.

(XIII) Patil, G. P. and Rao, C. R. (1977) *AS*, 383-405; (1978) *Biometrics*, 34, 179-189.

(XIV) Shaked, M. (1980) *J. R. Statist. Soc., Ser. B*, 42, 192-198.

(XV) Rai, G. (1971) *Trab. Estad.*, 22, 207-211.

(XVI) Bulmer, M. G. (1974) *Biometrika*, 30, 101-110.

(XVII) Shaban, S. A. (1981) *Comm. Statist.*, A10, 1389-1399.

(XVIII) Pavlik, M. (1970) *Aplik. Mat.*, 15, 97-105. (Czech; English summary)

(XIX) Sichel, H. S. (1974) *J. R. Statist. Soc., Ser. A.*, 137, 25-34.

(XIX)' Ong, S. H. and Lee, P. A. (1979) *Biom. J.*, 21, 611-627.

(XX) Subrahmaniam, Kathleen (1978) *Comm. Statist.*, A7, 673-683.

(XXI) Morrison, D. G. (1978) *Amer. Statist.*, 32, 23-26.

Morrison, D. G. and Brockway, G. (1979) *Psychometrika*, 44, 427-442.

(XXII) (Goodman, L. A. (1961) *Ann. Math. Statist.*, 32, 148-170.)

Berg, S. (1978, 1979) *Dept. Statistic, Univ. Lund*, Reports.

(XXIII) Kemp, C. D. and Kemp, A. W. (1965) *Biometrika*, 52, 381-394.

(XXIV) Gupta, R. P. and Jain, G. C. (1974) *SIAM J. Appl. Math.*, 27, 359-363.

(XXIV)' Jain, G. C. and Plunkett, I. G. (1977) *Biom. J.*, 19, 347-354.

TABLE 3

(XXV) Rao, C. R. and Rubin, H. (1964) *Sankhyā*, 26A, 295-298.

Aczel, J. (1972) *J. Appl. Prob.*, 9, 852-856; (1975) *Abh. Math. Sem. Univ. Hamburg*, 44, 91-100.

Govindarajulu, Z. and Leslie, R. T. (1970) *RCSW*, 1, 77-96.

Krishnaji, N. (1974) *Sankhyā*, **36A**, 204-213.

Patil, G. P. and Ratnaparkhi, M. V. (1977) *Sankhyā*, **39B**, 65-75.

Shanbhag, D. N. (1974) *J. Appl. Prob.*, **11**, 211-215; Ibid. (1977), **14**, 640-646.

Srivastava, R. C. and Singh, J. (1974) *SDSW*, **3**, 271-277; (1980) *NATO, Trieste.*

Steele, J. M. (1973) *Amer. Statist.*, **27**, 232.

Talwalker, S. (1970) *Sankhyā*, **32A**, 265-270; (1980) *J. Appl. Prob.*, **17**, 563-569.

(XXVI) Shanbhag, D. N. and Basawa, I. N. (1974) *Trab. Estad.*, **25**, 109-111.

Panaretos, J. (1980) *NATO, Trieste.*

(XXVII) Consul, P. C. (1974) *SDSW*, **3**, 279-290.

Janardan, K. G. (1974) *SDSW*, **3**, 359-364.

Samaniego, F. J. (1976) *J. Amer. Statist. Assoc.*, **71**, 475-479.

TABLE 4

(XXVIII) Samiuddin, M. and Mallick, S. A. (1970) *J. Natur. Sci. Math.*, **10**, 59-63.

(XXIX) Cressie, N. (1978) *Ann. Inst. Statist. Math., Tokyo*, **30A**, 435-442.

(XXX) Peizer, D. B. and Pratt, J. W. (1968) *J. Amer. Statist. Assoc.*, **63**, 1416-1456.

(XXXI) Borges, R. (1970) *Z. Wahr. verw. Geb.*, **14**, 189-199.

(XXXII) Molenaar, W. (1970) "Approximations..." *Math. Centrum Tracts No. 31*; Math Centrum, Amsterdam.

(XXXIII) Fujino, Y. (1980) *Biometrika*, **67**, 677-681.

(XXXIV) Ghosh, B. K. (1980) *Comm. Statist.*, A9, 427-438.

* Littlewood, J. E. (1969) *Adv. Appl. Prob.*, **1**, 43-72.

(Jain, G. C. (1974) *Math. Opforsch. Statist.*, **5**, 207-211.)

TABLE 1a SYSTEMS OF DISCRETE DISTRIBUTIONS - I

(I) (II) POWER SERIES (PSD(θ;g(θ)))

$$P_x = \frac{\theta^x}{x!} \cdot \frac{b(x)}{g(\theta)} \quad (x=0,1,\ldots)$$

$$b(x) = D^x g(\theta)\big|_{\theta=0}$$

$$PSD^{*n}(\theta;g(\theta)) \sim PSD(\theta;\{g(\theta)\}^n)$$

FACTORIAL SERIES (FSD(N;h(N)))

$$P_x = \frac{N^{(x)}}{x!} \cdot \frac{c(x)}{h(N)} \quad (x=0,1,\ldots,N)$$

$$c(x) = \Delta^x h(N)\big|_{N=0} = \Delta^x h(0)$$

*

(I)' (II)' Multivariate

$$P_{\underset{\sim}{x}} = \{\prod_{j=1}^{m} \frac{\theta_j^{x_j}}{x_j!}\} b(\underset{\sim}{x})/g(\underset{\sim}{\theta})$$

$$b(\underset{\sim}{x}) = (\prod_{j=1}^{m} D_{\theta_j}^{x_j})\, g(\underset{\sim}{\theta})\big|_{\underset{\sim}{\theta}=\underset{\sim}{0}}$$

$$P_{\underset{\sim}{x}} = \{\prod_{j=1}^{m} \frac{N_j^{(x_j)}}{x_j!}\} c(\underset{\sim}{x})/h(\underset{\sim}{N})$$

$$c(\underset{\sim}{x}) = (\prod_{j=1}^{m} \Delta_{N_j}^{x_j})\, h(\underset{\sim}{0})$$

(III) MODIFIED POWER SERIES (MPSD (h(θ); g(θ))

$$P_x = \frac{\{h(\theta)\}^x}{x!} \cdot \frac{b(x)}{g(\theta)} \quad (x=0,1,\ldots)$$

(IV) LAGRANGE (L(G($\cdot$), f($\cdot$)) (from MPSD with h(θ) = θ/G(θ))

Probability generating function (pgf) is $f(s)=f(0) + \sum_{j=1}^{\infty} \frac{\{s/G(s)\}^j}{j!} [D^{j-1}\{G(t)\}^j f'(t)\}|_{t=0}]$.

Usual to take G($\cdot$) also a pgf with $|G'(0)| < 1$.

$$P_0 = f(0); \quad P_x = D^{x-1}\{[G(t)]^x f'(t)\}\big|_{t=0} \qquad (x=1,2,\ldots)$$

(V) GENERALIZED HYPERGEOMETRIC SERIES ($_hG_k$Hypg ($\underset{\sim}{a};\underset{\sim}{b};\theta$))

$$P_x = [{}_hF_k(\underset{\sim}{a};\underset{\sim}{b};\theta)]^{-1}\{[\prod_{i=1}^{h} a_i^{[x]}]/[\prod_{j=1}^{k} b_j^{[x]}]\}\cdot(\theta^x/x!) \qquad (x=0,1,\ldots)$$

(If θ = 1 => GENERALIZED HYPERGEOMETRIC)

(VI) Multivariate

$$P_{\underset{\sim}{x}} \propto [a^{[x]}/c^{[x]}][\prod_{j=1}^{m} \{b_j^{[x_j]}\theta^{x_j}/x_j!\}] \qquad (x = \sum_{j=1}^{m} x_j;\ x_j=0,1,2,\ldots)$$

LIMITS OF INVERSE PÓLYA-EGGENBERGER (TR(0)): $P_x = \frac{\gamma^{[\alpha]}\gamma^{[\beta]}}{\gamma^{[\alpha+\beta]}} \cdot \frac{\alpha^{[x]}\beta^{[x]}}{x!\,(\alpha+\beta+\gamma)^{[x]}} \left(1-\frac{\gamma^{[\alpha]}\gamma^{[\beta]}}{\gamma^{[\alpha+\beta]}}\right)^{-1}$

$\beta \to 0$ leads to DIGAMMA; $P_x = \{\psi(\alpha+\gamma) - \psi(\gamma)\}^{-1}\, \alpha^{[x]}/\{(\alpha+\gamma)^{[x]}x\}$ $(x=1,2,\ldots)$

$\beta \to 0$, $\alpha \to 0$ leads to TRIGAMMA; $P_x = \{\psi'(\gamma)\}^{-1}(x-1)!/(\gamma^{(x)}x)$ $(x=1,2,\ldots)$

Multivariate

(VII)' INV. PÓLYA-EGGENBERGER (TR($\underset{\sim}{0}$)): $P_{\underset{\sim}{x}} = \frac{\gamma^{[\alpha]}\gamma^{[\beta]}}{\gamma^{[\alpha+\beta]}} \cdot \frac{\beta^{[x]}}{(\alpha+\beta+\gamma)^{[x]}} \prod_{j=1}^{m} \left(\frac{\alpha_j^{[x_j]}}{x_j!}\right) \Big/ \left(1-\frac{\gamma^{[\alpha]}\gamma^{[\beta]}}{\gamma^{[\alpha+\beta]}}\right)$

$(\alpha = \sum_{j=1}^{m}\alpha_j;\ x = \sum_{j=1}^{m} x_j;\ x_j \geq 0,\ \underset{\sim}{x} = \underset{\sim}{0}$ excluded)

$\beta \to 0$ leads to MULTIVARIATE DIGAMMA: $P_x = \{\psi(\alpha+\gamma) - \psi(\gamma)\}^{-1} \frac{(x-1)!}{(\alpha+\gamma)^{[x]}} \prod_{j=1}^{m} \left(\frac{\alpha_j^{[x_j]}}{x_j!}\right)$

$\beta \to 0$, $\alpha_j \to \infty$, $\gamma \to \infty$ with $\alpha_j/(\alpha+\gamma) = \theta_j$ and $\sum_{j=1}^{m}\theta_j = \theta$ leads to

MULTIVARIATE LOGSERIES: $P_x = \frac{(x-1)!}{\log(1-\theta)} \prod_{j=1}^{m} \left(\theta_j^{x_j}/x_j\right)$

[$\beta \to 0$, $\alpha_j \to 0$ with $\alpha_j/\alpha = t_j$, γ fixed gives a degenerate distribution]

TABLE 1b SYSTEMS OF DISCRETE DISTRIBUTIONS - II

(Probability ratios and choice of distribution)

(VIII) $$P_x - P_{x-1} = \frac{(a-x)P_{x-1}}{b_0+b_1x+b_2x^{(2)}} \Rightarrow U_x = \frac{xP_x}{P_{x-1}} = x + \frac{a-x}{b_1+b_2(x-1)}$$

If plot of U_x against x is *linear* then $b_2 = 0$ and distribution is indicated by properties of the linear relation $U_x = x + (a-x)/b_1$

INTERCEPT ("U_0")	SLOPE	=>	DISTRIBUTION
$\phi > 0$	0		Poisson (ϕ)
$(n+1)p/q > 0$	$-p/q < 0$		Binomial (n,p)
$(N-1)P/Q > 0$	$P/Q > 0$		Negative Binomial (N,P)
$-\phi < 0$	$\phi > 0$		Logseries (ϕ)
0	1		Discrete rectangular

(General deviations from 'Bernoulli sampling')

[POISSON BINOMIAL]

(IX) $$P_{\underset{\sim}{i}} = P_{i_1 i_2 \ldots i_n} = \Pr\left[\bigcap_{j=1}^{n} (X_j = x_j)\right] \qquad (i_j = 0,1)$$

Additive system: $$P_{\underset{\sim}{i}} / \left\{\prod_{j=1}^{m} P_{i_j}\right\} = \sum_{a<b}\sum\{p_{i_a i_b}/(p_{i_a}p_{i_b})\} - \binom{n}{2} + 1$$

Multiplicative system: $$P_{\underset{\sim}{i}} = \prod_{a<b} \phi_{i_a i_b} \qquad (\phi_{01} = \phi_{10})$$

(X) Multivariate (pgf of multinomial is $(\sum_{j=1}^{m} p_j s_j)^n$)

MULTINOMIAL BINOMIAL: pgf is $\left(\sum_{a_1=0}^{1}\sum_{a_2=0}^{1}\ldots\sum_{a_m=0}^{1} p_{\underset{\sim}{a}} \prod_{j=1}^{m} s_j^{a_j}\right)^n$ $[\sum\ldots\sum p_{\underset{\sim}{a}} = 1]$

MULTINOMIAL MULTINOMIAL: pgf is as above, but some (at least) of the a_h's have upper limits greater than 1.

TABLE 1c SYSTEMS OF DISTRIBUTIONS - III

(XI) "ABEL" DISTRIBUTIONS

$(x = 0,1,\ldots,n)$

QUASI-BINOMIAL I: $P_x = \alpha(\alpha+\beta+n\tau)^{-n}\binom{n}{x}(\alpha+\tau x)^{x-1}\{\beta+\tau(n-x)\}^{n-x}$

QUASI-BINOMIAL II: $P_x=\alpha\beta(\alpha+\beta)^{-1}(\alpha+\beta+n\tau)^{-(n-1)}\binom{n}{x}(\alpha+\tau x)^{x-1}\{\beta+\tau(n-x)\}^{n-x-1}$

QUASI-HYPERGEOMETRIC: $P_x= \alpha\beta(\alpha+\beta+n\tau)(\alpha+\beta)^{-1}\{\beta+\tau(n-x)\}^{-1}$
$$\binom{\alpha+\tau x}{n}\binom{\beta+\tau(n-x)}{n-x}\Big/\binom{\alpha+\beta+\tau n}{n}$$

QUASI-POLYA: $P_x=[(\alpha+\beta+n\tau)^{[n]}]^{-1}\alpha\binom{n}{x}\{\alpha+\tau(x-1)\}^{[n-1]}\{\beta+\tau(n-x)\}^{[n-x]}$.

Multivariate

QUASI-MULTINOMIAL I: $P_{\underset{\sim}{x}}= (\sum_{j=1}^{m}\alpha_j+\beta)^{-n}\binom{n}{x_1,\ldots,x_m,n-\sum_{j=1}^{m}x_j}\{\prod_{j=1}^{m}\alpha_j(\alpha_j+x_j)^{x_j-1}\}$

$$(\beta-\sum_{j=1}^{m}x_j)^{n-\sum_1^m x_j} \quad (0\le x_j;\ \sum_{j=1}^{m}x_j\le n)$$

TABLE 2   MODIFICATION METHODS

(XII) a) INFLATION AND DEFLATION

$$P^*_{x_0} = P_{x_0} + \alpha; \; P^*_x = P_x(1-P_{x_0}-\alpha)/(1-P_{x_0}) \quad (x \neq x_0; -P_{x_0} \leq \alpha \leq 1-P_{x_0})$$

$\alpha > 0 \Rightarrow$ *inflated* distribution; $\alpha < 0 \Rightarrow$ *deflated* distribution

$\alpha = -P_{x_0} \Rightarrow$ *truncated* distribution (especially with $x_0 = 0$)

$x_0 = 0, \alpha > 0 \Rightarrow$ 'added zeroes' distribution

$$\mu'^*_r = \left(1-\frac{\alpha}{1-P_{x_0}}\right)\mu'_r + \frac{\alpha}{1-P_{x_0}} x_0^r$$

(XIII) b) WEIGHTING

$$P^*_x = w_x P_x / \sum_j w_j P_j \qquad (w_x \geq 0)$$

(XIV) c) MIXING (Some examples of COMPOUND POISSON $\mathrm{Poi}(\theta) \underset{\theta}{\wedge} G$ - "G-POISSON")

[NOTE: POISSON-BINOMIAL is $\mathrm{Bin}(n,p) \underset{n/k}{\wedge} \mathrm{Poi}(\theta)$]

(XV) POWER FUNCTION-POISSON $\mathrm{Poi}(\theta) \underset{\theta}{\wedge}$ Power function $(h,>)$

(XVI) LOGNORMAL-POISSON $\mathrm{Poi}(\theta) \underset{\log\theta}{\wedge} N(\mu,\sigma^2)$

(XVII) INVERSE GAUSSIAN-POISSON $\mathrm{Poi}(\theta) \underset{\theta}{\wedge} \mathrm{Gauss}^{-1}(m,\sigma^2)$

(XVIII) ? $\mathrm{Poi}(\theta) \underset{1-e^{-\theta}}{\wedge} \mathrm{Beta}(\alpha,\beta)$

(XIX) (XIX)' BESSEL A-POISSON, BESSEL B-POISSON

(XX) NEGATIVE BINOMIAL-POISSON $\mathrm{Poi}(\theta) \underset{\theta/\phi}{\wedge} \mathrm{Neg.Bin}(N,P)$

BINOMIAL-POISSON $\mathrm{Poi}(\theta) \underset{\theta/\phi}{\wedge} \mathrm{Bin}(n,p)$

BETA-BINOMIAL $\mathrm{Bin}(n,p) \underset{p}{\wedge} \mathrm{Beta}(\alpha,\beta)$

(XXI) Generalized by taking $p = g(c)$ with $c \frown \mathrm{Beta}(\alpha,\beta)$

(XXII) $\mathrm{FSD}(N;\{h(N)\}^n) \underset{n}{\wedge} \mathrm{PSD}(\theta;g(\theta))$ ('snowball sampling')

RANDOM SUM is a special sort of mixing

$$F^{*N} \underset{N}{\wedge} G, \text{ (also called 'G-generalized F')}$$

(XXIII) (Poisson (θ) has pgf $\exp\{\theta(s-1)\}$)

HERMITE has pgf $\exp\{\theta_1(s-1)+\theta_2(s^2-1)\}$.

(XXIV) GENERALIZED HERMITE has pgf $\exp\{\theta_1(s-1)+\theta_m(s^m-1)\}$ $(m>2)$.

(XXIV)' GENERALIZED G_1,G_2-HERMITE has pgf $\exp\{\theta_1(G_1(s)-1)+\theta_m(G_2(s^m)-1)\}$.

TABLE 3 DAMAGE MODEL CHARACTERIZATIONS

If $\underset{\sim}{Y}$ and $\underset{\sim}{Z}$ are mutually independent and	distribution of $\underset{\sim}{Y}$, given $(\underset{\sim}{Y}+\underset{\sim}{Z})$ is:	then distributions of $\underset{\sim}{Y}$ and $\underset{\sim}{Z}$ are:
(XXV)	Binomial	Poisson
	Hypergeometric	Negative Binomial
	Pólya	Hypergeometric
(XXVI)	Multinomial	Independent Poisson
	Multivariate Hypergeometric	Multivariate Negative Binomial
(XXVII)	Quasi-binomial I	Lagrange double Poisson
	Quasi-hypergeometric	Quasi-binomial I
	Quasi-Pólya	Quasi-hypergeometric

TABLE 4 APPROXIMATE SOLUTIONS OF $\Phi(u_p) = \sum_{j=0}^{x} \binom{n}{j} p^j q^{n-j}$ (=P)

$(q = 1 - p;\ \delta_x = (x+\frac{1}{2}-np)/\sqrt{(npq)};\ \Phi(u) = (\sqrt{2\pi})^{-1}\int_{-\infty}^{u} \exp(-\frac{1}{2}t^2)dt)$

(XXVIII) $$\frac{(n-x-\frac{1}{2})(x+\frac{1}{2})}{n}\{\log(\frac{x+\frac{1}{2}}{np}) - \log(\frac{n-x-\frac{1}{2}}{nq})\}$$

(XXIX) $$\delta_x + \frac{1}{6}(q-p)(1-\delta_x^2)(npq)^{-\frac{1}{2}} - \frac{1}{72}\{2(1-pq)\delta_x - (5-14pq)\delta_x^3\}(npq)^{-1}$$

(XXX) $$\frac{x+\frac{2}{3}-(n+\frac{1}{3})p + \frac{1}{50}(\frac{q}{x+1} - \frac{p}{n-x} + \frac{\frac{1}{2}-p}{n+1})}{\sqrt{\{(n+\frac{1}{6})pq\}}}$$

$$\times \frac{1}{\delta_x}[2\{(x+\tfrac{1}{2})\log(\frac{x+\frac{1}{2}}{np}) + (n-x-\tfrac{1}{2})\log(\frac{n-x-\frac{1}{2}}{nq})\}]^{\frac{1}{2}}$$

(XXXI) $$\int_{p}^{(x+\frac{1}{6})/(n+\frac{1}{3})} \{t/(1-t)\}^{\frac{1}{3}} dt$$

(XXXII) $$\begin{cases} \sqrt{\{(4x+3)q\}} - \sqrt{\{(4n-4x-1)p\}} & (0.05 \le P \le 0.95) \\ 2[\sqrt{\{(x+1)q\}} - \sqrt{\{(n-x)p\}}] & (P < 0.05,\ P > 0.95) \end{cases}$$

(XXXIII) $$(x^*-n^*p)/\sqrt{(n^*pq)} \text{ with } x^* = x + \frac{1}{3}(u_p^2+2);\ n^* = n + \frac{1}{3}(2u_p^2+2)$$

(Iteration needed)

(XXXIV) $$(1+2c\delta_x+c^2)^{\frac{1}{2}}\,\mathrm{sgn}(q-p) - c \quad \text{for } p \ne \tfrac{1}{2} \quad (c=3(q-p)^{-1}(npq)^{\frac{1}{2}})$$

$$\frac{2}{\sqrt{3}}\sqrt{(12n+1)}\cos[\frac{1}{3}\pi + \frac{1}{3}\cos^{-1}\{36\sqrt{3}\delta_x\, n(12n+1)^{-\frac{3}{2}}\}] \quad \text{for } p = \tfrac{1}{2}$$

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Reliability and Risk Analysis

Abstract
of
Serial GWU/IRRA/TR-81/4
21 December 1981

THE U.S. ARMY (BRL'S) KINETIC ENERGY PENETRATOR PROBLEM: ESTIMATING THE PROBABILITY OF RESPONSE FOR A GIVEN STIMULUS

Thomas A. Mazzuchi
Nozer D. Singpurwalla

The crew compartment of an army vehicle is protected by an armor plate. It is desired to test the strength of this armor plate in order to assess its appropriateness for use in the vehicle.

A specimen of the plate is taken and projectiles are fired at different points on the plate at different striking velocities. If a projectile penetrates the armor it is said to have defeated the armor. Our goal is to determine the relationship between the striking velocity and the probability of penetration. Due to the expensive nature of all items involved, this goal must be achieved with a minimum amount of testing. A Bayesian approach for solving this problem is presented here and illustrated using some real data.

Research Supported by

U.S. Army Research Office
Grant DAAG-29-80-C-0067
and
Naval Surface Weapons Center under
Contract N00014-77-C-0263, with the
Office of Naval Research

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Reliability and Risk Analysis

THE U.S. ARMY (BRL'S) KINETIC ENERGY PENETRATOR PROBLEM: ESTIMATING THE PROBABILITY OF RESPONSE FOR A GIVEN STIMULUS

Thomas A. Mazzuchi
Nozer D. Singpurwalla

1. STATEMENT OF THE PROBLEM

The following statement of the problem is based on our several discussions with Dr. Robert L. Launer of the Army Research Office, Research Triangle Park, North Carolina, and Dr. J. Richard Moore of the Ballistic Research Laboratory (BRL), Aberdeen Proving Ground, Maryland.

The crew compartment of an army vehicle is protected by a certain kind of material which we will refer to as an "armor plate." It is desired to test the strength of this armor plate so that we may be able to assess its appropriateness for use on the vehicle.

In order to do this, a 10' × 10' specimen of the armor plate is taken, and a projectile is fired from a gun which is aimed at different points on the plate. In Figure 1.1 below, we indicate a possible firing pattern according to which the gun is aimed.

Typically, the distance between the muzzle of the gun and the target is about 200 meters, and the velocity of the projectile, measured

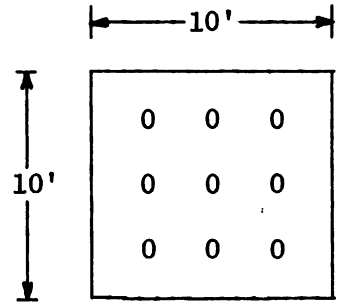


Figure 1.1--Illustration of a firing pattern of a gun.

between two conveniently located points between the gun and the target is about 5000 feet per second.

The projectile is known as the "penetrator," and the outcome of each firing is described by a binary variable which takes the value 1 if the penetrator defeats the target, and the value 0 if the penetrator fails to defeat the target. The penetrator induces a stress on the armor; the stress is a function of two quantities, the "striking velocity" and the "angle of fire." The striking velocity, also known as the "stimulus," is the velocity with which the penetrator strikes the armor, whereas the angle of fire θ (indicated in Figure 1.2 below) is the amount by which the armor plate is tilted.

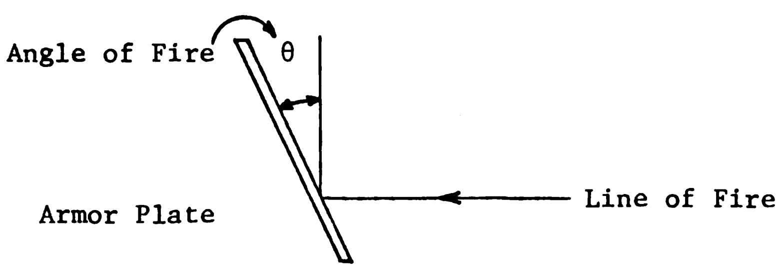


Figure 1.2--Illustration of the angle of fire.

Both the armor specimen and the penetrator are very expensive and thus the testing has to be kept to a bare minimum. One strategy that has been adopted is to fix the angle of fire, say at θ^*, and then to fire the penetrator at different striking velocities. After each firing, a record is made of whether the penetrator defeated the target or not. It is assumed that the striking velocity can be measured without any error.

2. GOALS, OBJECTIVES, AND SOME COMMENTS ON CURRENT APPROACHES

Given that our goal is to be able to assess the appropriateness of the armor plate for use on a vehicle, our objective should be to estimate the relationship between the striking velocity (the stimulus) and the probability of penetration (a response of 1). This is illustrated in Figure 2.1, wherein it is assumed that the probability of penetration is a nondecreasing function of the stimulus.

The situation described above is identical to the one encountered in "bioassay experiments," and "low dose radiation experiments," in which the relationship mentioned before is known as the <u>quantal response curve</u>. The dose level of a drug is the stimulus, and interest generally centers around $V_{.5}$, the stimulus at which the probability of response is $.5$. Since it is possible to subject more than one animal to a particular dose level, the number of tests at each value of the stimulus can be more than one. Furthermore, tests are often conducted at several dose levels, and thus the large sample theory which typically justifies inference from bioassay experiments is adequately substantiated.

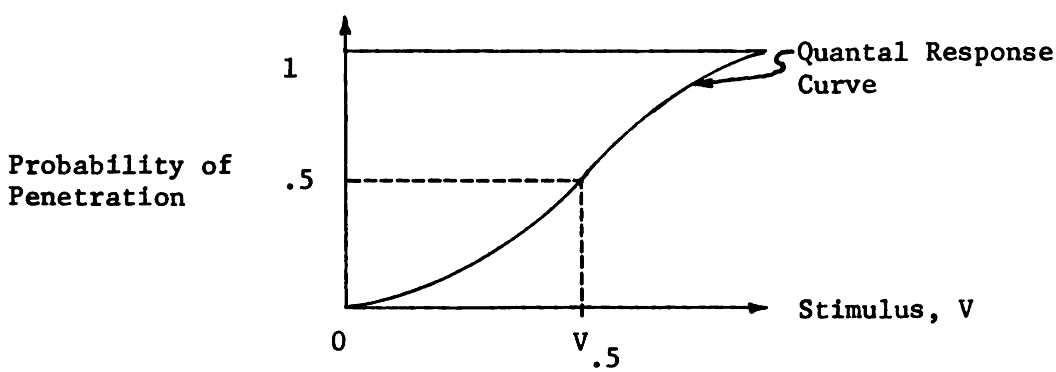


Figure 2.1--Probability of penetration vs. stimulus.

Despite these conspicuous differences between bioassay experimentation and the problem described here, the methodology and techniques of the former have been directly adopted for use in the latter. In so doing, a serious compromise has been made--the estimation of $V_{.5}$, rather than the entire quantal response curve, has been made the dominant issue of the kinetic energy penetration problem. Specifically, the BRL's commonly used "Langley Method" [Rothman, Alexander, and Zimmerman (1965, pp. 55-58)] and the "Up and Down Method" [op. cit., pp. 101-103] focus exclusive attention on the estimation of $V_{.5}$.

The typical approach used in bioassay for estimating $V_{.5}$ is to assume that the probability of response p is an arbitrary nondecreasing function of the stimulus V, specified via the relationship

$$p = F((v-\mu)/\sigma) ,$$

where F is a distribution function determined by a symmetrical density function with location parameter μ and scale parameter σ. Often F is taken to be the normal distribution function

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds ,$$

or the logistic distribution function $F(x) = (1 - e^{-x})^{-1}$.

The data from a bioassay experiment consists of n_i, the number of subjects receiving stimulus V_i, $i=1,\ldots,K$, and X_{ij}, $j=1,\ldots,n_i$, where

$$X_{ij} = 1 , \text{ if the j subject responds under stimulus } V_i, \text{ and}$$
$$= 0 , \text{ otherwise.}$$

Given the data (n_i,X_{ij}), $i=1,\ldots,K$, $j=1,\ldots,n_i$, the parameters μ and σ are estimated using the method of maximum likelihood, under the assumption that the test results can be judged independent. Once μ and σ are estimated, the estimation of $V_{.5}$ follows from the fact that F, the tolerance distribution, has been specified. Nonparametric and robust estimators of $V_{.5}$, such as the Spearman-Karber estimator, the L-estimator, the M-estimator, and the Tukey Biweight estimator, have also been obtained, all under the assumption that the density function giving F is symmetric. These estimators have been discussed by Miller and Halpern (1979). Furthermore, it has been empirically shown that for the estimation of $V_{.5}$ it does not matter what specific form is chosen for F; many of the commonly used nonparametric estimators yield identical estimates of $V_{.5}$, as long as symmetry is assumed.

A drawback of the assumption of symmetry is that the estimate of the probability of response when the stimulus is zero is nonzero. Whereas this may not be too disturbing in bioassay with its emphasis on $V_{.5}$, in the problem considered here and the low dose radiation experimentation, such an estimate would be clearly unacceptable. A zero value of the stimulus should correspond to a zero value for the probability of response.

In view of the above difficulty, the paucity of data at each level of the stimulus, and our inability to specify a functional form of F which has some practical merit, we are motivated to advocate a Bayesian approach for the solution of this problem. Our approach is described in Section 3.

3. AN OUTLINE OF A BAYESIAN APPROACH

A Bayesian approach to the bioassay problem was first proposed by Kraft and Van Eeden in 1964, and was more fully developed by Ramsey in 1972. We consider here the theme proposed by Ramsey; extensions of this theme are considered by Shaked and Singpurwalla (1982).

Let $0 \equiv V_0 < V_1 < \ldots < V_M < V_{M+1} \leqq \infty$, be M distinct levels of the stimulus at which the target (armor plate) is tested; M is chosen in advance. The outcome of a test at V_i is described by a binary (0,1) variable X_i , where $X_i = 1$ if the penetrator with a striking velocity V_i defeats the target. Let $p_i = P\{X_i=1\}$, $i=1,\ldots,M$, and without loss of generality, we assume that

$$0 \equiv p_0 < p_1 < p_2 < \ldots < p_M < p_{M+1} \equiv 1 \; ; \tag{3.1}$$

it is always possible to choose V_1 and V_M which satisfy the above inequality.

Given $\underset{\sim}{X} = (X_1,\ldots,X_M)$, one goal is to estimate the unknown p_i's , $i=1,\ldots,M$, subject to the inequalities (3.1). Another goal is to estimate p_j , for some $j \neq i$, such that if $V_i < V_j < V_{i+1}$, the estimates satisfy $p_i < p_j < p_{i+1}$, $i=1,\ldots,M$; this pertains to estimating the probability of response at a stimulus where no target was tested. Yet a third goal would be to estimate the largest stimulus, say V_α , for which $p_\alpha \leqq \alpha$, where $0 < \alpha < 1$ is specified.

Ramsey's approach for achieving the above goals is to assign a Dirichlet as a prior distribution for the successive differences $p_1, p_2-p_1, \ldots, p_M-p_{M-1}$, and then to use the <u>modal</u> value of the resulting joint posterior distribution as a Bayes point estimate of $(p_1,\ldots,p_M)$. The modal value is computed with the inequalities (3.1) being satisfied. The modal value of the posterior distribution, if unique, is also known as the <u>generalized maximum likelihood estimator</u> [see DeGroot (1970, p. 236)], and is used as a Bayes estimator when we do not wish to specify a particular loss function. Having estimated the p_i's , the estimation of p_j and V_α is undertaken via an interpolation procedure.

Specifically, if $\alpha_i > 0$, $i=1,\ldots,M$, and $\beta > 0$ are constants such that $\sum_{i=1}^{M+1} \alpha_i = 1$, then the prior density function π is of the form

$$\pi \propto \left\{ \prod_{i=1}^{M+1} (p_i - p_{i-1})^{\alpha_i} \right\}^{\beta} . \tag{3.2}$$

It is important to note that when averaging according to π integration must be done with respect to $\Pi_{i=1}^{M} dp_i \;/\; \Pi_{i=1}^{M+1} (p_i - p_{i-1})$.

Since M has been prechosen, the stopping rule is clearly delineated, and so the likelihood for the response probabilities at the observed stresses is

$$\prod_{i=1}^{M} p_i^{X_i} (1 - p_i)^{1-X_i} . \tag{3.3}$$

The joint density function of the posterior distribution of $p_1,\ldots,p_M$ is proportional to the product of the prior density function (3.2) and the likelihood function (3.3). Thus

$$f(p_1,\ldots,p_M \mid X_1,\ldots,X_M)$$

$$\propto \prod_{i=1}^{M+1} p_i^{X_i}(1-p_i)^{1-X_i}\left[\frac{\Gamma(\beta)}{\prod_{i=1}^{M+1}\Gamma(\beta\alpha_i)}\right]\left\{\prod_{i=1}^{M+1}(p_i-p_{i-1})^{\alpha_i}\right\}^{\beta} . \quad (3.4)$$

Ramsey has not been able to obtain the posterior marginal distributions of p_i , $i=1,\ldots,M$, nor has he commented on any aspects of these distributions. He uses a nonlinear programming algorithm to obtain $(\hat{p}_1,\ldots,\hat{p}_M)$, the modal value of (3.4), subject to the constraint that $\hat{p}_1 \leq \hat{p}_2 \leq \cdots \leq \hat{p}_M$; this is his Bayes estimator of $(p_1,\ldots,p_M)$. In contrast to this Mazzuchi (1982) has been able to obtain all the moments of the marginal posterior distribution of the p_i , $i=1,\ldots,M$. This work of Mazzuchi's represents an extension of Ramsey's results, and is one that takes us a step closer to a fully Bayesian analysis. The moments can be used to approximate the marginal posterior distributions of the p_i's using the techniques given in Elderton and Johnson (1969). The approximated posterior distributions give us a measure of uncertainty associated with our using the first moment of the marginal posterior distribution of p_i , $i=1,\ldots,M$, as our Bayes estimate of p_i . The first moment of the marginal posterior distribution is used as a Bayes estimator when we are willing to assume the square error as a loss function. The formulae for the moments and their use for approximating the marginal posterior distributions are given in Appendix A.

The computational effort required to compute the moments mentioned above increases with M . Thus there is a trade-off between the convenience of using an optimization algorithm to obtain the modal value of (3.4), versus the laborious computational effort involved in obtaining several moments of each of the M posterior marginal distributions. The optimization algorithm cited above is based on the "Sequential Unconstrained Minimization Technique" (SUMT) of Fiacco and McCormick (1968). A computer code which adopts SUMT for the problem considered here is described by Mazzuchi and Soyer (1982). This code can also be used for the computation of the moments of the marginal posterior distributions of the p_i , $i=1,\ldots,M$.

3.1 Specification of the Prior Parameters

In order to implement the Bayesian procedure, we need to specify the prior parameters α_i , $i=1,\ldots,M$, and β , given in (3.2). In order to do this, we observe (see Ramsey) that $u_i = p_i - p_{i-1}$, $i=1,\ldots,M$, has a beta distribution on the unit interval (denoted as $u_i \sim \text{Beta}(\beta\alpha_i, \beta(1-\alpha_i); 0,1))$,

$$f(u_i;\ \beta\alpha_i,\ \beta(1-\alpha_i)) = \frac{\Gamma(\beta)}{\Gamma(\beta\alpha_i)\ \Gamma(\beta(1-\alpha_i))}\, u_i^{\beta\alpha_i}\ (1-u_i)^{\beta(1-\alpha_i)} \quad , \quad 0 \leqq u_i \leqq 1 \ ,$$

with

$$E(u_i) = \alpha_i \ , \text{ and} \tag{3.5}$$

$$\text{Var}(u_i) = \frac{\alpha_i(1-\alpha_i)}{(\beta+1)} \ . \tag{3.6}$$

If P_i^* denotes our best prior guess about P_i , consistent with the fact that the P_i^*'s increase in i , then the α_i's can be obtained via (3.5) as

$$\alpha_1 = P_1^*$$

$$\alpha_i = P_i^* - P_{i-1}^* , \quad i=2,\ldots,M ,$$

and

$$\alpha_{M+1} = 1 - P_M^* .$$

In order to choose the parameter β , we need to have some idea about the uncertainty associated with our choice of p_1^* . This in practice can be done in one of the following two ways:

(i) Suppose that in addition to p_1^* , our best guess about the variance of p_1 is $\mathrm{Var}(p_1)$. Then, substituting $\alpha_1 = p_1^*$ in (3.6), we have

$$\mathrm{Var}(u_1) = \mathrm{Var}(p_1) = \frac{\alpha_1(1 - \alpha_1)}{(\beta + 1)} ,$$

so that

$$\beta = \begin{cases} \dfrac{p_1^*(1 - p_1^*)}{\mathrm{Var}(p_1)} - 1 , & \text{if } \beta > 0 , \\ 0 , & \text{otherwise.} \end{cases}$$

Note that $\beta = 0$ corresponds to the case of isotonic regression.

(ii) Often in practice [cf. McDonald (1979)], associated with the best guess value p_1^* , a user is able to specify two numbers $a_1^* > 0$ and $b_1^* < 1$, such that for some γ_1 (specified by the user), $0 < \gamma_1 < 1$,

$$P(a_1^* < p_1 < b_1^*) = 1 - \gamma_1 .$$

Since $p_1 \sim \mathrm{Beta}(\beta\alpha_i, \beta(1-\alpha_i); 0,1)$, given p_1^* , we set $\alpha_1 = p_1^*$, and find that value of β such that

$$\int_{a_1^*}^{b_1^*} \frac{\Gamma(\beta)}{\Gamma(\beta\alpha_1)\,\Gamma(\beta(1-\alpha_1))}\, p_1^{\beta\alpha_1 - 1}\,(1-p_1)^{\beta(1-\alpha_1)-1}\, dp_1 = 1-\gamma_1 \tag{3.7}$$

Suppose, further, that for any one or more of the indices i , $i=2,\ldots,M$, a user is also able to specify two numbers $a_i^* > p_{i-1}^*$, and $b_i^* < 1$, such that for some γ_i (specified by the user), $0 < \gamma_i < 1$,

$$P(a_i^* < (p_i^* \mid p_{i-1}^*,\ldots,p_1^*) < b_i^*) = 1 - \gamma_i \ .$$

Then, using the fact (see Ramsey) that

$$\begin{aligned}(p_i \mid p_{i-1}^*) &\sim \mathrm{Beta}(\beta\alpha_i,\ \beta(1 - \alpha_1 - \ldots - \alpha_i);\ p_{i-1}^*,\ 1)\\ &= f(p_i \mid p_{i-1}^*;\ \beta,\alpha_i)\ \text{, say,}\end{aligned}$$

we can find the <u>smallest</u> value of β , β^* , which satisfies (3.7) and (3.8), where

$$\int_{a_i^*}^{b_i^*} f(p_i \mid p_{i-1}^*;\ \beta,\alpha_i)\,dp_i = 1 - \gamma_i \ , \tag{3.8}$$

with $\alpha_i = p_i^* - p_{i-1}^*$, $i=2,\ldots,M$.

A computer code which determines the smallest value of β described above is available; the details of this program are given by Mazzuchi and Soyer (1982). Our reason for choosing the smallest value of β stems from the fact that large values of β give a very strong prior, with the result that even a large amount of failure data will not change our prior distribution.

3.2 Interpolation Procedure and the Estimation of Quantiles

Let the M-dimensional point

$$(p_1^+,\ldots,p_M^+) = \begin{cases} (\hat{p}_1,\ldots,\hat{p}_M) \text{ , if the mode of the joint posterior is} \\ (\tilde{p}_1,\ldots,\tilde{p}_M) \text{ , if the first moments of the marginal posterior are} \end{cases}$$

used as the Bayes estimator of $(p_1,\ldots,p_M)$.

Suppose that we wish to estimate p_j , for some $j \neq i$, $i=1,\ldots,M$, where $V_i < V_j < V_{i+1}$. Let p_j^* be our best prior guess of p_j , the probability of response at a nonexperimental impulse V_j . Then, following Ramsey, we pick p_j^+ in such a manner that

$$\frac{p_{i+1}^* - p_j^*}{p_{i+1}^+ - p_j^+} = \frac{p_j^* - p_i^*}{p_j^+ - p_i^+} \text{ .} \tag{3.9}$$

For the estimation of V_α , the αth quantile $(0 < \alpha < 1)$, we first see if there is an observation stimulus, say V_i , for which $p_i^+ = \alpha$. If so, then V_i is our Bayes estimate of V_α . If not, we determine the pair of observational impulses, say V_i and V_{i+1} , for which $p_i^+ < \alpha < p_{i+1}^+$. Since the probability of response curve is assumed to be increasing, the straight line segment joining the points $0, p_1^+, \ldots, p_i^+, p_{i+1}^+, \ldots, p_M^+, 1$, will be an increasing function of i . We shall find that value of the impulse, say V_α^+ , $V_i < V^+ < V_{i+1}$, for which $p_\alpha^+ = \alpha$.

4. APPLICATION TO SOME BRL DATA

In Appendix B we present eight sets of data labelled 1, 2, 3, 4, 6, 7, 8, and 9, pertaining to 60 kinetic energy penetration tests. These data were given to us by Dr. Moore of BRL and have been carefully sanitized to maintain confidentiality. Data sets labelled 5 and 10, also given to us by Dr. Moore, have been eliminated from consideration because the striking velocity for these data is much too different from those of the other sets. All the 10 sets of data were obtained sequentially over time, in the sense that data set 1 was the first one to be obtained, followed by data set 2 (obtained after some lapse of time), and so on, until we reach data set 9, which is the last considered here. To the best of our knowledge, all eight data sets are assumed to have been collected under identical conditions. That is, there is no indication that, except for differences in striking velocity, the material and the methods of testing used for data set 1 are different from those used in data set 2, and so on. This, plus the sequential nature of the data, enables us to use the posterior obtained from one data set as the prior for the next set, and so on, until we obtain the posterior using data set 9, which then gives our final estimate of the response curve.

Data set 1 consists of 13 observations taken at striking velocities ranging from 128.60 (in some unspecified units) to 166.16. The result of each test is indicated by a binary variable X_i . The best prior guess values p_i^* , necessary to choose the prior parameters α_i , were not specified by BRL. However, what appears to be reasonable is to assume that the probability of response at a striking velocity of 100 is close to zero, and that at a striking velocity of 200 it is almost 1.

Thus we make an arbitrary choice for p_i^* , say p_{i0}^* , by letting $p_{i0}^* = 1 - \exp[-.07(V_i - 100)]$. Data on striking velocities outside the range of 100 to 200 were excluded. Despite this arbitrary choice of p_{i0}^* , we shall see how even a scant amount of data significantly changes the posterior response curve, provided that the smoothing parameter β is not too large. Three values of β were also chosen arbitrarily; these are 1, 10, and 25. Recall that small values of β tend to emphasize the data, whereas large values of β tend to emphasize the prior distribution. In Appendix B we show our analysis for the case of $\beta = 10$.

Since, in reality, the data are generated sequentially over time, our first step would be to revise the best prior guess values p_{i0}^* , $i=1,\ldots,61$, based on data set 1 alone. The posterior (modal) values corresponding to the striking velocities of data set 1, p_{i1}^+ , will be the revised values of p_{i0}^* , for $i=1,\ldots,13$; these are given in column 5 of the table in Appendix B. The revised values of p_{i0}^* , for $i=4,\ldots,61$, are obtained via the interpolation formula (3.9), using p_{i1}^+ , $i=1,\ldots,13$, and p_{i0}^* , $i=14,\ldots,61$. Let the revised values of p_{i0}^* , $i=14,\ldots,61$, be denoted by p_{i1}^* ; these too are shown in column 5 of the table in Appendix B.

Upon receiving data set 2, we revise the values p_{i1}^* , $i=14,\ldots,19$, by the posterior modal values corresponding to the six striking velocities of data set 2. We denote these revised values by p_{i2}^+ , $i=14,\ldots,19$; these are given in column 5 of the table in Appendix B. The revised values of p_{i1}^* , $i=20,\ldots,61$, are obtained by interpolation, using

p^{+}_{i1} , i=1,...,13 , p^{+}_{i2} , i=14,...,19 , and p^{*}_{i1} , i=20,...,61 ; we denote these revised values by p^{*}_{i2} , i=20,...,61 , and show them in column 6.

We continue the above scheme of systematically revising the $p_{i.}$'s , either via the posterior modal values or by interpolation, until we incorporate the effect of all eight sets of data. Data set 9, the last one considered here, consists of eight observations taken at starting velocities ranging from $V_{54} = 144.83$ to $V_{61} = 198.94$. The posterior modal values corresponding to the striking velocities of data set 9, p^{+}_{i8} , i=54,...,61 , are given in column 12; the interpolated values p^{*}_{i7} required to obtain the p^{+}_{i8}'s are given in column 11. Since the p^{*}_{i7}'s incorporate the results of the previous seven sets of data, we claim that the final posterior modal values p^{+}_{i8} , i=54,...,61 , are based on the results of all the testing. Had we ignored the sequential nature of the data and computed the posterior modal values by using Bayes Theorem on the best prior guess values p^{*}_{i0} , i=1,...,61 , then the posterior modal values corresponding to V_{14} through V_{61} would be different from the $p^{+}_{i.}$ values, i=14,...,61 , given in the table. This difference is due to the interpolation scheme that is used to constantly revise the best prior guess values, when we consider the data sets sequentially.

A plot of p^{+}_{i8} versus V_i , i=54,...,61 , represents our final estimate of the quantal response curve. Estimates of the probabilities of response at striking velocities different from V_i , i=54,...,61 , can be obtained using the interpolation formula (3.9). When we use the interpolation formula to obtain an estimate of p_j , for some

$j=1,\ldots,53$, we need to specify a value p_j^* , the best prior guess value of p_j . Suppose that the index j appears in data set k , for some $k < 9$; then for p_j^* we will use p_{jk}^+ . In so doing, we will have incorporated the effect of the last data set, data set 9, in our obtaining the estimate of p_j , and thus achieve a certain amount of smoothness. Note that the effect of the data sets between k and 9 is already present in our estimates p_{i8}^+ , $i=54,\ldots,61$, and these are used in our interpolation scheme. For example, suppose that we wish to estimate the probability of response at a striking velocity of 158.52. This striking velocity occurs in data set 2, and lies between the striking velocities 148.97 and 159.15 of data set 9. The index j corresponding to the value 158.82 is 17. To use (3.9), we identify p_{i+1}^* and p_{i+1}^+ as being .70499 and .53014, respectively, p_i^* and p_i^+ as .62881 and .42386 (see data set 9), and p_j^* as .64436 (see data set 2), and compute p_j^+ as our estimate of p_j .

In Figures 4.1, 4.2, and 4.3, we show plots of our Bayes estimate of the probability of response at the eight striking velocities of data set 9, for $\beta = 1$, 10, and 25 , respectively. Also shown are the 90% probability of coverage intervals for each estimate. These intervals are obtained using the moments of the posterior distributions of p_i , $i=54,\ldots,61$, and then using the techniques of Elderton and Johnson (1969) to approximate the posterior distributions--see Appendix A. On each of these figures we also show a graph of our best guess values p_{i0}^* , $i=1,\ldots,61$; these enable us to see how the data have changed our prior estimates. We observe that the 90% probability of coverage intervals tend to be small in the middle of the range of the striking velocities.

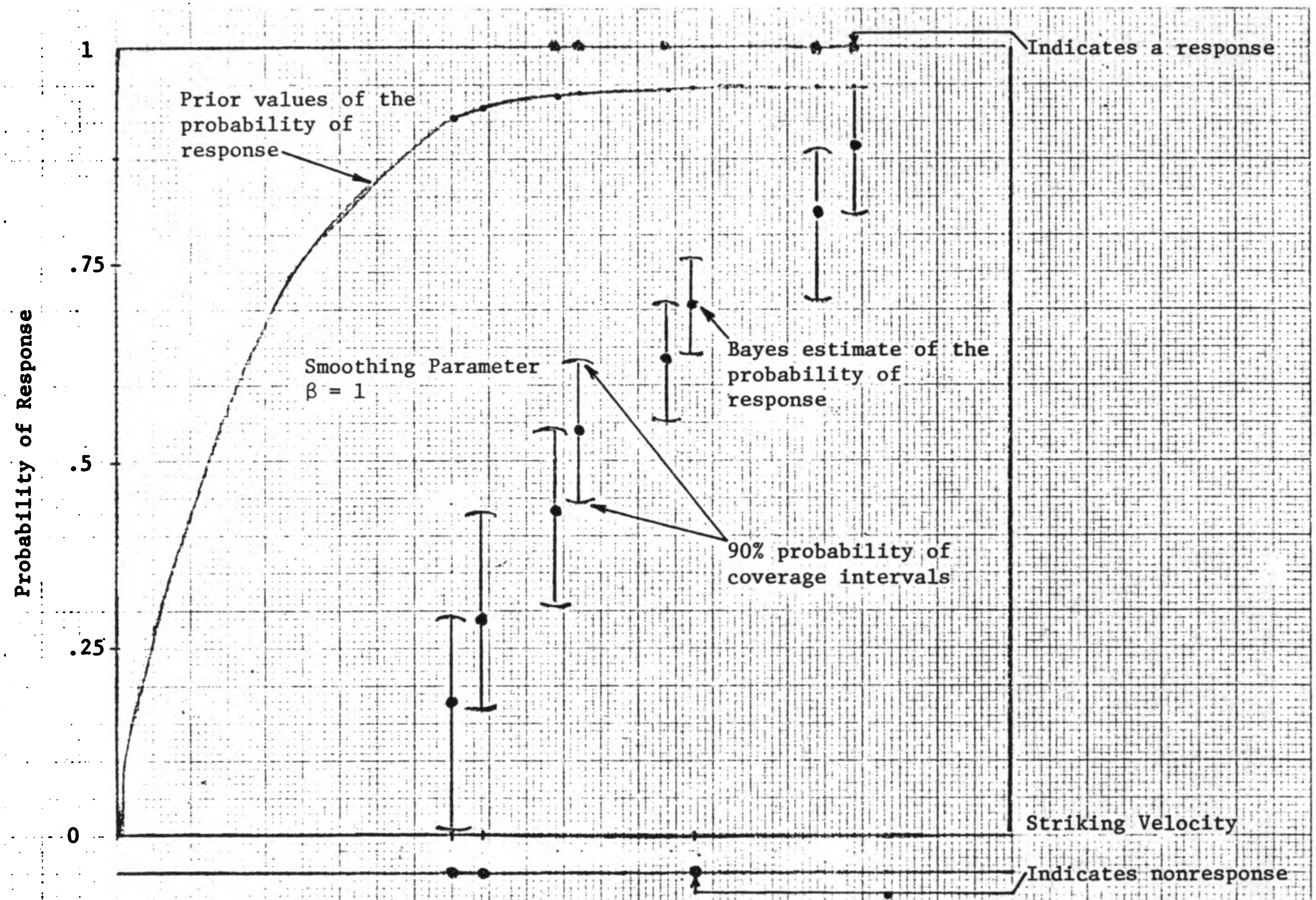


Figure 4.1--Prior and posterior values of the probability of response for data set 9, with β = 1.

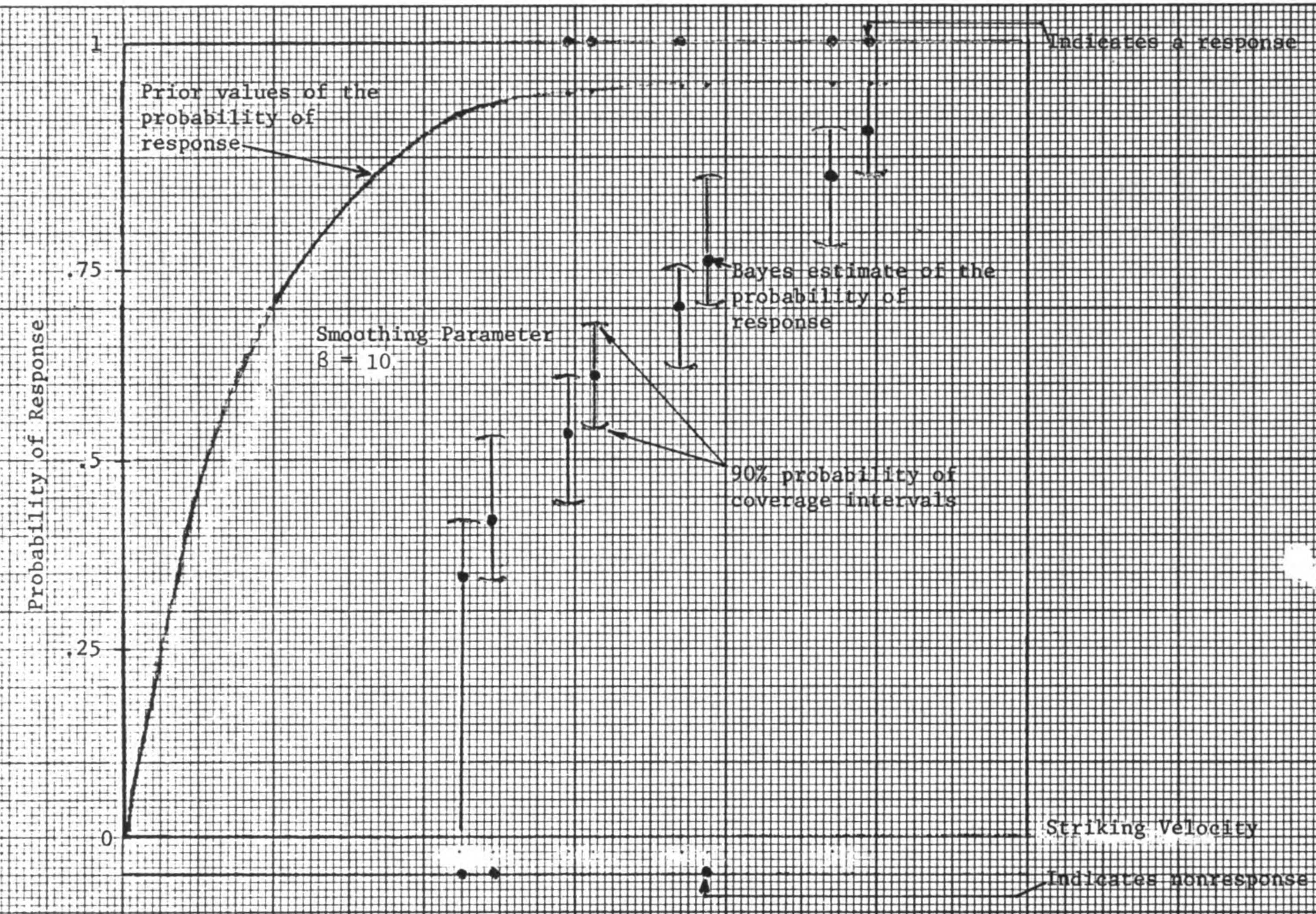


Figure 4.2—Prior and posterior values of the probability of response for data set 9, with β = 10.

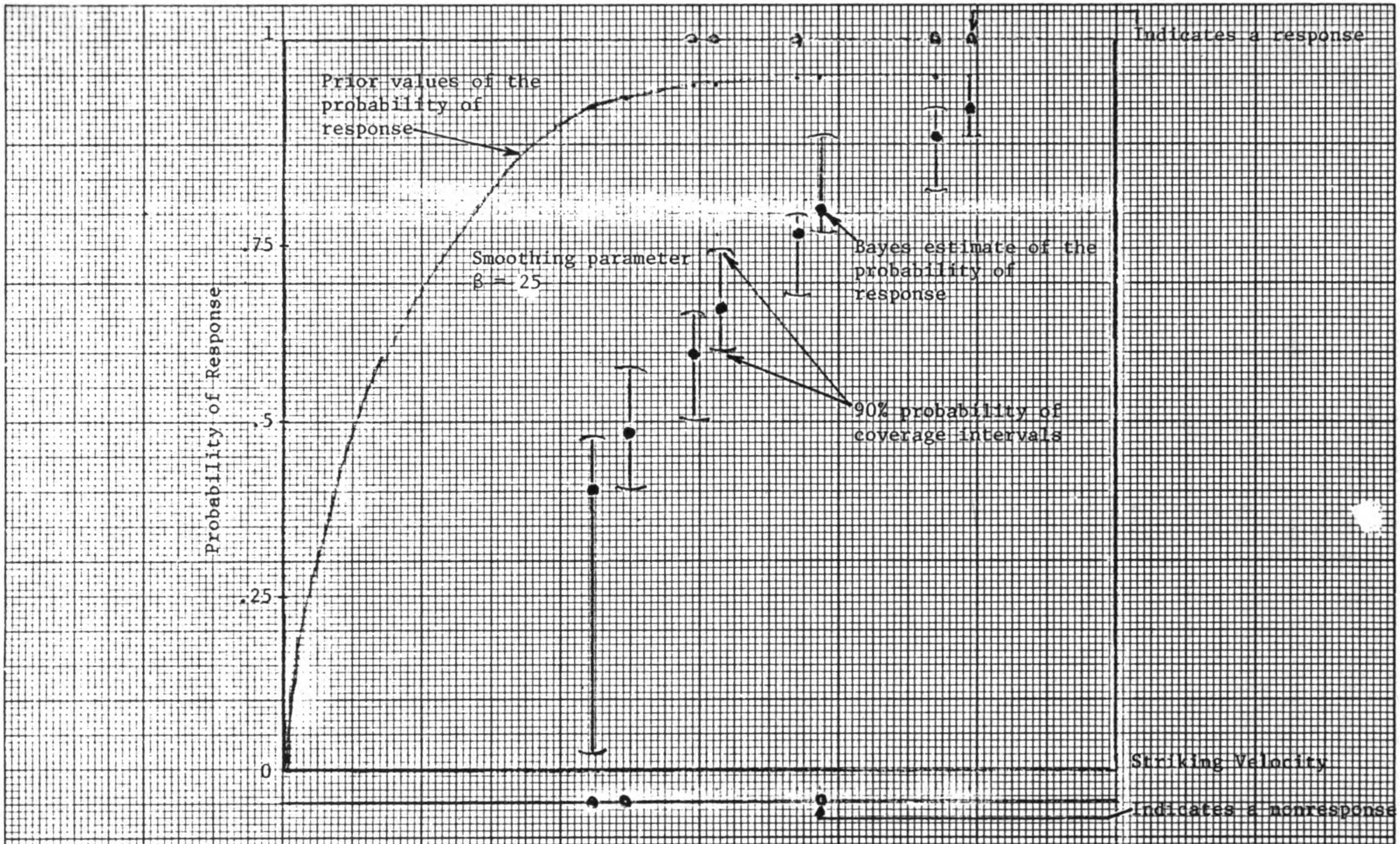


Figure 4.3--Prior and posterior values of the probability of response for data set 9, with $\beta = 25$.

In Figure 4.4, we superimpose the plots of Figures 4.1, 4.2, and 4.3, in order to give a perspective of the effect of β in our computations. It appears that our Bayes estimates for the three cases of $\beta = 1, 10,$ and 25 tend to converge toward each other; this is to be expected, since we have 61 observations with which we revise our prior probabilities.

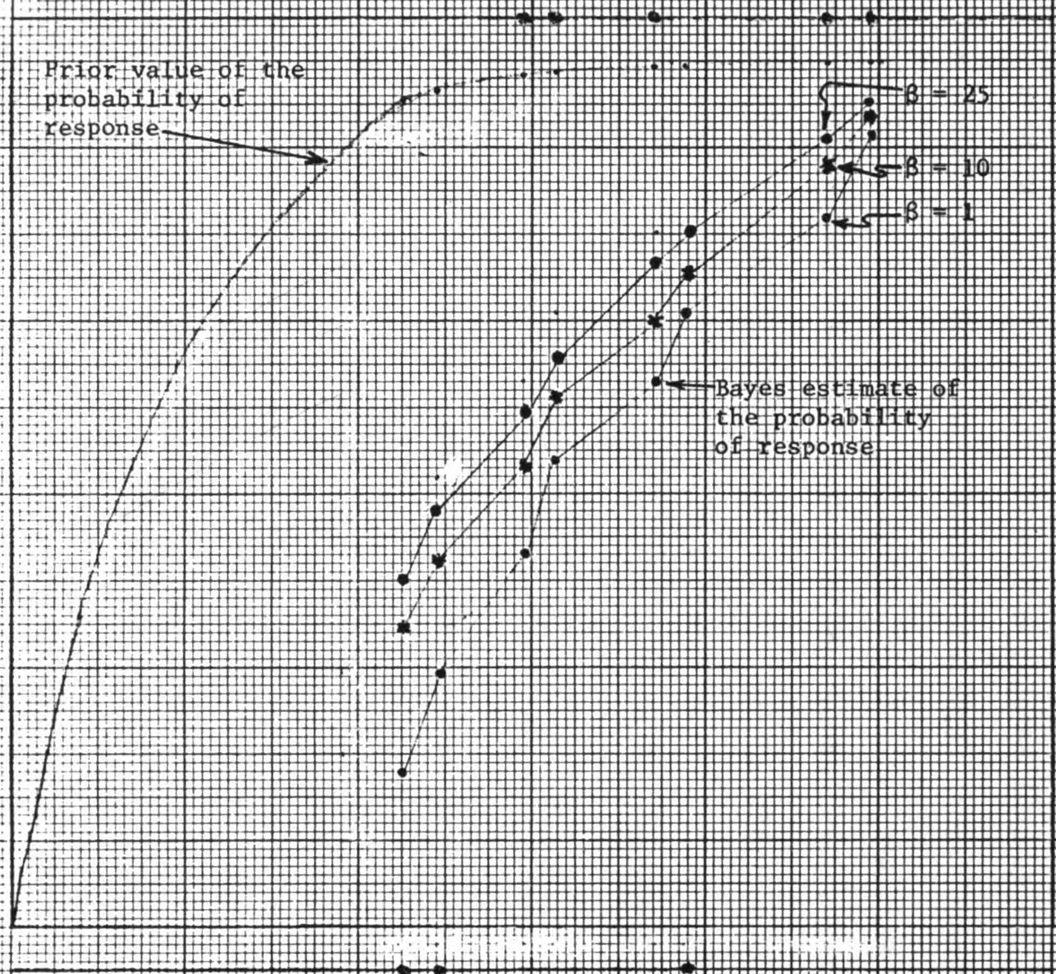


Figure 4.4--Prior and posterior values of the probability of response for β = 1, 10, and 25.

ACKNOWLEDGMENTS

The use of the "Sequential Unconstrained Minimization Technique" (SUMT) was central to our undertaking and accomplishing the work reported here; without it, we would not have been able to implement our Bayesian ideas. We gratefully acknowledge the help of Professors A. V. Fiacco and G. P. McCormick in connection with our use of SUMT. We also acknowledge the several helpful conversations with Drs. J. Richard Moore and Robert Launer regarding several aspects of this problem, and are particularly grateful to the latter for encouraging us to write this up. Mr. Refik Soyer's assistance with developing the computer programs cited here is also recognized. Mr. W. McDonald of the Naval Surface Weapons Center, made several comments which led us to develop the material of Section 3.1.

APPENDIX A

Moments of the Marginal Posterior Distributions

The moments of the posterior distribution of p_i , $i=1,\ldots,M$, have been obtained by Mazzuchi (1982); a formula for obtaining these is given below. A computer code which facilitates the computation of the moments is described by Mazzuchi and Soyer (1982).

Let $\bar{X}_i = 1 - X_i$, $i=1,\ldots,M$, $B(a,b) = \Gamma(a)\Gamma(b) / \Gamma(a+b)$, and

$$K = \sum_{r_1=0}^{\bar{X}_1} \cdots \sum_{r_M=0}^{\bar{X}_M} (-1)^{\sum_{i=1}^{M} r_i} \prod_{i=1}^{M} B\left(\sum_{j=1}^{i} X_j + \beta\alpha_j + r_j ,\ \beta\alpha_{i+1} \right) .$$

Then, for $\ell=1,2,\ldots$,

$$E(p_s^{\ell}) = \frac{1}{K} \sum_{r_1=0}^{\bar{X}_1} \cdots \sum_{r_M=0}^{\bar{X}_M} (-1)^{\sum_{i=1}^{M} r_i} \prod_{i=1}^{M} B\left(\sum_{j=1}^{i} X_j^* + \beta\alpha_j + r_j ,\ \beta\alpha_{i+1} \right) ,$$

where

$$X_j^* = \begin{cases} X_j + \ell , & j = s \\ X_j , & \text{otherwise.} \end{cases}$$

These moments can be used to approximate the posterior distribution of p_i , $f(p_i)$, $i=1,\ldots,M$. In order to do this, we consider a system of frequency curves described by Elderton and Johnson (1969) which are based on the transforms of a standard normal variate Z . The system of curves which is appropriate to our problem is that referred to as the "bounded system of curves," denoted by Elderton and Johnson (1969, p. 123) as S_B , and described by

$$Z = \gamma + \delta \operatorname{Ln}[(p_i - \varepsilon) / (\varepsilon + \lambda - p_i)] , \quad \varepsilon < p_i < \varepsilon+\lambda ,$$

where γ , δ , λ , and ε are parameters whose values are determined by the first four moments of $f(p_i)$ about its mean.

Hill, Hill, and Holder (1976) give a computer code which determines γ , δ , λ , and ε from the first four moments of $f(p_i)$ about its mean. Since it was assumed that $p_{i-1} < p_i < p_{i+1}$, we estimate λ and ε from the Bayesian estimates of the p_i ; γ and δ are obtained from the computer code. Having obtained these parameters, the distribution $f(p_i)$ is obtained from Elderton and Johnson (1969, p. 130) as

$$f(p_i) = \frac{N}{\lambda\sqrt{2\pi}} \left[\left(\frac{p_i-\varepsilon}{\lambda}\right)\left(1 - \frac{p_i-\varepsilon}{\lambda}\right)\right]^{-1} \exp\left[-\tfrac{1}{2}\left(\gamma + \delta\ell n\left(\frac{p_i-\varepsilon}{\varepsilon+\lambda-p_i}\right)\right)^2\right],$$

$$\varepsilon < p_i < \varepsilon + \lambda ,$$

where N in our case is 1.

In order to obtain the approximate $(1-\gamma)\%$ probability of coverage intervals for each p_i , which contain its Bayes estimate $\overset{+}{p}_i$ (mode or mean), we use the fact that since

$$z = \gamma + \delta\ell n[(p_i-\varepsilon)/(\varepsilon+\lambda-p_i)] , \quad \varepsilon < p_i < \varepsilon + \lambda ,$$

$$p_i = \lambda \exp[(\tfrac{\gamma-z}{\delta}) + 1]^{-1} + \varepsilon .$$

Thus, to find two numbers, a and b , such that

$$P\{\overset{+}{p}_i-a \leqq p_i \leqq \overset{+}{p}_i+b\} = 1 - \delta ,$$

we use

$$P\left\{-\delta\ell n\left(\frac{\lambda}{\overset{+}{p}_i-a-\varepsilon} - 1\right) + \gamma \leqq z \leqq -\delta\ell n\left(\frac{\lambda}{\overset{+}{p}_i+b-\varepsilon} - 1\right) + \gamma\right\} = 1 - \delta ,$$

and solve for a and b by setting

$$-\delta \ell n \left(\frac{\lambda}{p_i^+ - a - \varepsilon} - 1 \right) + \gamma = z_{1-(\delta/2)}$$

and

$$-\delta \ell n \left(\frac{\lambda}{p_i^+ + b - \varepsilon} - 1 \right) + \gamma = z_{\delta/2} ,$$

where $z_{\delta/2}$ is the $(1-(\delta/2))$th percentile of a standard normal distribution. Taking $c = \max(a,b)$, we form our interval

$$\Pr\{p_i^+ - c < p_i < p_i^+ + c\} \geq 1 - \delta .$$

These intervals may not be symmetric about the mean or modal estimate. This case arises when the boundaries of the probability of coverage interval exceed the boundary of the variable. In such cases the variable boundary is used as the boundary of the probability of coverage interval. The probability of any symmetric interval about the mean or modal estimate may be obtained by proceeding in the reverse or the above and evaluating the interval for the standard normal variate.

APPENDIX B

In the table below we give values of the striking velocity V_i, the response X_i, and the best prior guess values p^*_{i0}, $i=1,\ldots,61$, for the eight sets of data described in Section 4. We also show, for $\beta = 10$, the revised values of p^*_{i0}, p^+_{ij}, or p^*_{ij} based on data set j, $j=1,2,3,4,6,7,8,9$.

TABLE B.1

Data Set	Striking Velocity v_i	Response x_i	Best Guess Values p_{i0}	Revised Values p^+_{i1}, i=1,...,13 and p^*_{i1}, i=14,...,61	Revised Values p^+_{i2}, i=14,...,19 and p^*_{i2}, i=20,...,61	Revised Values p^+_{i3}, i=20,...,26 and p^*_{i3}, i=27,...,61	Revised Values p^+_{i4}, i=27,...,33 and p^*_{i4}, i=34,...,61	Revised Values p^+_{i6}, i=34,...,39 and p^*_{i6}, i=40,...,61	Revised Values p^+_{i7}, i=40,...,45 and p^*_{i7}, i=46,...,61	Revised Values p^+_{i8}, i=46,...,53 and p^*_{i8}, i=54,...,61	Revised Values p^+_{i9}, i=54,...,61	Observation No.
1	128.60	0	.86141	.29717								1
	132.74	0	.89589	.34071								2
	143.56	0	.95071	.39334								3
	146.74	0	.96043	.43289								4
	147.70	1	.96297	.47221								5
	149.29	0	.96682	.50928								6
	153.11	0	.97452	.55059								7
	154.70	0	.97717	.59361								8
	155.34	0	.97816	.64076								9
	159.15	0	.98321	.69702								10
	161.70	1	.98593	.76390								11
	162.79	1	.98695	.82463								12
	166.16	0	.98966	.88210								13
2	139.42	0	.93438	.37766	.25028							14
	145.15	0	.95584	.41421	.32728							15
	145.79	0	.95775	.42199	.39349							16
	158.52	1	.98245	.68855	.64436							17
	163.61	1	.98767	.83990	.79980							18
	169.98	1	.99206	.90947	.89721							19
3	110.14	0	.50375	.17378	.11517	.13815						20
	128.28	1	.85831	.29610	.19623	.26379						21
	130.51	0	.87855	.31881	.21128	.32700						22
	133.05	1	.89810	.34283	.22720	.39629						23
	146.24	1	.95904	.42723	.39842	.53722						24
	148.01	0	.96376	.47982	.44792	.60804						25
	161.06	1	.98529	.74816	.70558	.80077						26
4	118.09	0	.71350	.24614	.16312	.21247	.16467					27
	132.10	0	.89112	.33478	.22186	.37305	.31127					28
	137.83	0	.92676	.37035	.24544	.41130	.39592					29
	141.97	1	.94498	.38784	.27173	.43294	.47888					30
	148.01	1	.96376	.47982	.44792	.60804	.64292					31
	155.65	1	.97892	.64589	.60421	.72495	.76144					32
	183.03	1	.99678	.96328	.95831	.97179	.93809					33
6	120.32	0	.75442	.26026	.17248	.22698	.17792	.15360				34
	130.83	0	.88120	.32216	.21350	.33666	.27805	.27191				35
	138.15	0	.92836	.37188	.24645	.41213	.39910	.41606				36
	155.65	1	.97862	.64589	.60421	.72495	.76144	.74847				37
	155.65	1	.97862	.64589	.60421	.72495	.76144	.74847				38
	156.29	1	.97955	.65265	.61396	.73224	.76666	.81215				39
7	158.52	0	.98247	.68878	.64458	.75514	.78805	.82534	.48173			40
	159.42	0	.98353	.70489	.66114	.76753	.79191	.83248	.54375			41
	169.34	1	.99170	.90536	.89146	.92655	.90571	.92409	.67075			42
	173.43	0	.99376	.92885	.91921	.94533	.91915	.93491	.73758			43
	185.26	1	.99724	.96853	.96427	.97582	.94693	.95728	.83340			44

Table B.1--*Continued*

Data Set	Striking Velocity V_i	Response X_i	Best Guess Values P^*_{i0}	Revised Values P^+_{i1}, i=1,...,13 and P^*_{i1}, i=14,...,61	Revised Values P^+_{i2}, i=14,...,19 and P^*_{i2}, i=20,...,61	Revised Values P^+_{i3}, i=20,...,26 and P^*_{i3}, i=27,...,61	Revised Values P^+_{i4}, i=27,...,33 and P^*_{i4}, i=34,...,61	Revised Values P^+_{i6}, i=34,...,39 and P^*_{i6}, i=40,...,61	Revised Values P^+_{i7}, i=40,...,45 and P^*_{i7}, i=46,...,61	Revised Values P^+_{i8}, i=46,...,53 and P^*_{i8}, i=54,...,61	Revised Values P^+_{i9}, i=54,...,61	Observation No.
	187.17	1	.99758	.97241	.96867	.97880	.95347	.96254	.90878			45
8	102.18	0	.13984	.04824	.03197	.03835	.02972	.02566	.01498	.05849		46
	116.50	0	.68023	.23467	.15552	.20069	.15554	.13428	.07838	.14365		47
	119.37	0	.73775	.25451	.16867	.22107	.17252	.14894	.08693	.20861		48
	124.14	0	.81139	.27991	.18550	.24716	.19634	.17536	.10235	.27999		49
	134.01	1	.90464	.34911	.23136	.39971	.37027	.38021	.22192	.42858		50
	134.33	1	.90673	.35112	.23269	.40081	.37271	.38307	.22359	.48811		51
	141.01	1	.94121	.38422	.26410	.42666	.45480	.46547	.27168	.56532		52
	175.07	1	.99441	.93626	.92763	.95103	.92323	.93820	.75167	.84252		53
9	144.83	0	.95485	.41019	.31881	.47169	.51518	.52119	.30421	.58411	.34483	54
	148.97	0	.96608	.50216	.46894	.62376	.65886	.65380	.38161	.62881	.42386	55
	159.15	1	.98321	.69702	.65306	.76149	.78759	.82900	.51352	.70499	.53014	56
	162.34	1	.98654	.77890	.73715	.82213	.83099	.86394	.58736	.74763	.60749	57
	174.12	1	.99403	.93193	.92271	.94770	.92085	.93628	.74345	.83777	.70205	58
	177.30	0	.99521	.94538	.93798	.95803	.92824	.94223	.76893	.85347	.75580	59
	194.17	1	.99851	.98301	.98071	.98695	.97136	.97694	.94385	.96439	.87678	60
	198.94	1	.99893	.98780	.98615	.99063	.97944	.98345	.95970	.97444	.93611	61

REFERENCES

DE GROOT, M. H. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.

ELDERTON, W. P. and N. L. JOHNSON (1969). *Systems of Frequency Curves.* Cambridge University Press, London.

FIACCO, A. V. and G. P. McCORMICK (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Technique.* Wiley, New York.

HILL, L. T., R. HILL, and R. L. HOLDER (1976). Fitting Johnson curves by moments. *Appl. Statist.* *25*, 180-190.

KRAFT, C. H. and C. VAN EEDEN (1964). Bayesian bioassay. *Ann. Math. Statist.* *7*, 163-186.

MAZZUCHI, T. A. (1982). Bayesian nonparametric estimation of the failure rate. D.Sc. dissertation, George Washington University.

MAZZUCHI, T. A. and R. SOYER (1982). Computer programs for 'A Bayesian approach to quantile and response probability estimation using binary response data'--A user's guide. Technical Paper Serial GWU/IRRA/TR-82/5, George Washington University.

McDONALD, W. (1979). A statistical treatment of the explosion damage problem, with applications to the case of damage to submarine pressure hulls by very low yield nuclear weapons. Technical Report, Naval Surface Weapons Center, White Oak, Maryland.

MILLER, R. G. and J. W. HALPERN (1979). Robust estimators for quantile bioassay. Technical Report No. 42, Division of Biostatistics, Stanford University.

RAMSEY, L. L. (1972). A Bayesian approach to bioassay. *Biometrics 28*, 841-858.

ROTHMAN, D., M. J. ALEXANDER, and J. M. ZIMMERMAN (1965). The design and analysis of sensitivity experiments, Volume 1. Technical Report NAS8-11061, National Aeronautics and Space Administration.

SHAKED, M. and N. D. SINGPURWALLA (1982). A Bayesian approach to quantile and response probability estimation using binary response data. Technical Report in progress.

SUPERMARTINGALES AND CRITERIA FOR
RECURRENCE AND TRANSIENCE OF MARKOV CHAINS

Mary Anne Maher
Army Materiel Test and Evaluation Directorate
US Army White Sands Missile Range
White Sands Missile Range, NM 88002

Abstract

For an irreducible Markov chain whose time parameter is discrete and whose state space is a countable discrete set, criteria for recurrence and transience are obtained by constructing supermartingales. These constructions are extensions of Foster's criteria for recurrence and transience in terms of inequalities; and they are similar to the construction of Lyapunov functions in dynamical systems. Examples to which the criteria are applied include: pairs of queues with priorities, pairs of queues in parallel, two-dimensional positive random walks, and competition processes.

1. Introduction

Either the states of an irreducible Markov chain are recurrent or they are transient. To determine which occurs is the first step in analysing the chain. For recurrent chains, the next step is to determine the frequency of visits of the sample paths to each state. For transient chains, the next step is to determine the asymptotic behavior of the sample paths. This paper deals with criteria to distinguish between recurrence and transience which are obtained by the construction of non-negative supermartingales. This method is similar to the construction of Lyapunov functions to analyze the stability of dynamical systems.

Section Two contains the statements of the criteria and their proofs. The results presented there are extensions of the criteria obtained by Foster (1951, 1952, 1953) for recurrence and transience in terms of inequalities. Recently Mertens, Samuel-Cahn, and Zamir (1977, 1978) have also used supermartingales to obtain necessary and sufficient conditions for recurrence and transience of Markov chains.

Section Three deals with specific examples: pairs of queues with priorities, pairs of queues in parallel, two-dimensional positive random walks, and competition processes. Criteria for positive recurrence and transience have been obtained previously for these examples: for priority queues, by Kesten and Runnenburg (1957); for parallel queues, by Kingman (1961b); for two-dimensional positive random walks, by Kingman (1961a) and Malyšev (1972); and for competition processes, by Iglehart (1964) and Reuter (1961). However, the proofs presented here are different from those in the papers cited above.

2. Criteria for Recurrence and Transience

Let $\{Z_n: n \geq 0\}$ denote a stationary Markov chain whose time parameter is discrete and whose state space is a countable discrete set, S. Throughout this paper, the underlying probability space (Ω,F) is fixed with Ω the set of sequences of elements of S and with F the sigma-field generated by the finite-dimensional cylinder sets. Then the random variables $\{Z_n: n \geq 0\}$ are simply the coordinate functions on the product space Ω. The chain is governed by a transition function $P(z,z')$ defined for elements z, z' of S. Such a transition function determines a family of probability measures $\{P_z: z \in S\}$ on the measure space (Ω,F). Namely, to a cylinder set of the form $\bigcap_{i=0}^{k} \{Z_i = z_i\}$, the measure P_z assigns the value $\delta(z_0,z)\Pi_{i=0}^{k-1} P(z_i, z_{i+1})$. (Here, the value $\delta(z_0,z)$ is equal to one if $z=z_0$ and to zero otherwise). Finally, it is assumed that the chain is irreducible. Then, given two points z and z' of S, there are a positive integer k and points $z_1, z_2,\ldots,z_k$ of S such that the product $P(z,z_1)\{\Pi_{i=1}^{k-1}P(z_i, z_{i+1})\}P(z_k,z')$ is positive.

Either the states of an irreducible Markov chain are recurrent or they are transient. The distinction is that the typical sample path of a recurrent chain visits every state infinitely often while the typical path of a transient chain does not visit any state infinitely often. Formally, the chain is recurrent if there exist a point z' of S and a finite subset A' of S such that the probability $P_{z'}\{Z_n \in A'$ infinitely often$\}$ is positive. Moreover, if this occurs then for any point z and any finite subset A of S, the probability $P_z\{Z_n \in A$ i.o.$\}$ is equal to one. The chain is transient if there exist a point z' of S and a finite subset A' of S for which the probability $P_{z'}\{Z_n \in A'$ i.o.$\}$ is equal to zero.

For a subset A of S, the stopping time T_A is the non-negative random variable defined by:

$T_A = \min \{n \geq 1: Z_n \varepsilon A\}$ if the set is not empty;

$= + \infty$ otherwise.

For recurrent chains, the probability $P_z\{T_A < \infty\}$ is equal to one for any point z of S and any subset A of S. A recurrent chain is positive recurrent if there exist a point z′ of S and a finite subset A′ of S - {z′} for which $E_{z'}(T_{A'})$ is finite. (Here, E_z refers to the integral with respect to the measure P_z.) Positive recurrent chains are exactly those which admit an invariant probability measure. See Chung (1967) or Freedman (1971) for a more complete discussion of these results.

To show that an irreducible chain is transient it is enough to show that its sample paths cannot visit certain states infinitely often. To show that an irreducible chain is recurrent it is enough to show that paths whose initial point is outside a fixed finite set are certain to hit that set. These are the basic ideas of the results which follow. Their proofs make use of the notion of a supermartingale and of the convergence theorem for non-negative supermartingales. The necessary material is summarized in Appendix A.

<u>Theorem 2.1</u>. Let $\{Z_n: n \geq 0\}$ denote a stationary Markov chain taking values in countable discrete set S. Suppose that there exists a non-negative function ϕ, defined on S, for which

(2.2) $E(\phi(Z_{n+1})|Z_n = z) \leq \phi(z)$ for all z S. If ϕ is not constant and if all states of the chain communicate, then the chain is transient.

Proof. Let z and z′ be two points of S for which $\phi(z) \neq \phi(z')$. The sequence $\{\phi(Z_n): n \geq 0\}$ is a non-negative supermartingale; it converges almost everywhere with respect to the measure P_z. Since all states communicate, paths of a recurrent chain would visit infinitely often each of the points z and z´. But this contradicts the convergence of the sequence $\{\phi(Z_n): n\geq 0\}$.

__Theorem 2.3.__ Let $\{Z_n: n\geq 0\}$ denote a stationary Markov chain whose state space is the countable discrete set S. Suppose that there exists a sequence $\{\phi_n: n\geq 0\}$ of non-negative functions on S for which

(2.4) $E(\phi(Z_{n+1})|Z_n = z) \leq \phi_n(z)$ for all $z\epsilon S$. If all states of the chain communicate, and if there is a point $z\epsilon S$ for which

(2.5) $\lim_{n\to\infty} \phi_n(z) = +\infty$, then the chain is transient.

__Proof.__ Let z be a state of S for which the property (2.5) holds. The non-negative supermartingale $\{\phi_n(Z_n): n\geq 0\}$ converges almost everywhere to a finite limit with respect to the measure P_z. Sample paths of a recurrent chain would return infinitely often to the initial state z; but then (2.5) contradicts the result of the convergence theorem.

Mertens, Samuel-Cahn, and Zamir (1977) have obtained independently the next result with a similar proof.

__Theorem 2.6.__ Let $\{Z_n: n\geq 0\}$ denote a stationary, irreducible Markov chain taking values in a countable, discrete set S. If there exists a finite subset A of S and a non-negative function ϕ, defined on S, for which

(2.7) $E(\phi(Z_{n+1})|Z_n = z) \leq \phi(z)$ for $z \notin A$, and

(2.8) $\{z: \phi(z)<M\}$ is a finite set for all $M>0$,

then the chain is recurrent.

__Proof.__ Let $m = \inf\{\phi(z): z\epsilon S\}$. From (2.8), it follows that there are states z_1 and z_2 of S for which $\phi(z_1) = m$, $\phi(z_2)>\phi(z_1)$, and also $P\{Z_{n+1} = z_2|Z_n = z_1\}>0$. Since the inequality (2.7) cannot hold for z_1, the exceptional set A is not empty.

Let $z \varepsilon S-A$. To show that the chain is recurrent, it is sufficient to show that its sample paths hit A almost surely with respect to the probability measure P_z. The non-negative supermartingale $\{\phi(Z(n \wedge T_A))\}$ converges almost everywhere. Because the chain is irreducible, it follows from (2.8) that $P_z\{\lim\sup_{n\to\infty} \phi(Z_n) < \infty\} = 0$. Thus, convergence of the sequence $\{\phi(Z(n \wedge T_A))\}$ implies that paths from z hit A almost surely.

The remaining results of this section refer specifically to irreducible chains $\{Z_n: n \geq o\}$ on the integral lattice Z_+^d in R^d of points whose coordinates are non-negative integers. For any vector $W = (W_1, W_2 \ldots, W_d) \varepsilon \; R^d$, the "exponential" W^{Z_n} denotes the product

$$W^{Z_n} = \prod_{i=1}^{d} W_i^{Z_n^{(i)}} \text{ with } Z_n = (Z_n^{(1)}, Z_n^{(2)}, \ldots, Z_n^{(d)}).$$

Proposition 2.9. If there exists a vector $W = (W_1, W_2, \ldots, W_d)$ whose components are strictly positive and which satisfies

(2.10) $E(W^{Z_1 - Z_0} | Z_0 = z) < 1$ for all $z \varepsilon Z_+^d$, then the chain is transient.

Proof. The assertion follows from Theorem 2.1 with the function $\phi(Z) = W^Z$.

Proposition 2.11. Suppose that there exists a vector $W \varepsilon R^d$ whose components satisfy the inequality $W_i > 1$ for $1 \leq i \leq d$ and a finite subset A of Z_+^d such that, for all $n \geq 0$,

(2.12) $E(W^{Z_{n+1} - Z_n} | Z_n = z) < 1$ for $z \varepsilon A$, then the chain is recurrent.

Proof. Note that the origin belongs to the exceptional set A and apply Theorem 2.6 with the function $\phi(Z) = W^Z$.

Proposition 2.13. Suppose that there exists a vector $W_i \varepsilon R^d$ whose components satisfy $W_i > 1$ for $1 \leq i \leq d$ and a finite subset A of Z_+^d such that $K < 1$ where

(2.14) $K = \sup\{E(W^{Z_{n+1} - Z_n} | Z_n = z): n \geq 0, z \in Z_+^d - A\}$. Then the chain is positive recurrent.

<u>Proof</u>. Note that the exceptional set contains the origin. Let z be a point of $Z_+^d - A$. For $n \geq 0$, it follows from (2.14) that

$$E_z(W^{Z_{n+1}} \cdot 1_{(T_A > n+1)}) \leq E_z(E(W^{Z_{n+1}} \cdot 1_{(T_A \geq n+1)}) | Z_n))$$

$$= E_z(W^{Z_n} \cdot 1_{(T_A > n)}(E(W^{Z_{n+1} - Z_n}) | Z_n))$$

$$\leq K\, E_z(W^{Z_n} \cdot 1_{(T_A > n)}).$$

So for $n \geq 0$, $E(W^{Z_n} \cdot 1_{(T_A > n)}) \leq K^n W^z$. Finally, since $W_i > 1$ for $1 \leq i \leq d$:

$$P_z(T_A > n) \leq E_z(1_{(T_A > n)} W^{Z_n}) \leq K^n W^z.$$

Therefore, $P_z(T_A > n)$ converges to zero geometrically. Hence $E_z(T_A^k)$ is finite for each positive integer k. This shows that the chain is positive recurrent.

Note that the construction of "exponential" supermartingales of the type described in (2.13) provides an estimate of the distribution of the hitting time T_A in terms of a geometric distribution. When the explicit forms of the iterates of the transition function are not simple to obtain, such estimates are useful for approximating the moments of hitting times. This technique has been used by Kemperman (1961). A related technique for continuous-parameter chains has been proposed by Aldous (1981).

3. Examples

Example A: Pairs of Queues with Priorities. Priority customers and non-priority customers queue up at a counter for service. Non-priority customers are served only when there are no priority customers in the system. The arrival rates are λ_1, for priority customers, and λ_2, for non-priority customers. Their service rates are μ_1 and μ_2 respectively. Inter-arrival times and service times are exponentially distributed.

Let $Z_n = (X_n, Y_n)$ represent the number of priority customers and the number of non-priority customers in the system, observed just after the n-th change of state. The transition function for this discrete parameter chain is, if $x>0$:

$$P\{Z_1 = (x+1,y) | Z_0 = (x,y)\} = \lambda_1/\Gamma,$$

$$P\{Z_1 = (x-1,y) | Z_0 = (x,y)\} = \mu_1/\Gamma,$$

$$P\{Z_1 = (x,y+1) | Z_0 = (x,y)\} = \lambda_2/\Gamma.$$

The normalizing constant Γ is equal to $\lambda_1+\lambda_2+\mu_1$. If $x=0$ and $y>0$:

$$P\{Z_1 = (0,y+1) | Z_0 = (0,y)\} = \lambda_2/\Gamma'',$$

$$P\{Z_1 = (0,y-1) | Z_0 = (0,y)\} = \mu_2/\Gamma'',$$

$$P\{Z_1 = (1,y) | Z_0 = (0,y)\} = \lambda_1/\Gamma''.$$

The normalizing constant Γ'' is $\lambda_1+\lambda_2+\mu_2$. For jumps from $(0,0)$:

$$P\{Z_1 = (1,0) | Z_0 = (0,0)\} = \lambda_1/(\lambda_1+\lambda_2),$$

$$P\{Z_1 = (0,1) | Z_0 = (0,0)\} = \lambda_2/(\lambda_1+\lambda_2).$$

Criteria for recurrence and transience can be described simply in terms of the parameters λ_1, λ_2, μ_1, and μ_2.

Theorem 3.1. The queue with priorities is transient if:

(3.2) $\lambda_1 > \mu_1$,

or if both

(3.3) $\lambda_1 < \mu_1$ and $\lambda_1/\mu_1 + \lambda_2/\mu_2 > 1$,

or if

(3.4) $\lambda_1 = \mu_1$.

It is positive recurrent if

(3.5) $\lambda_1/\mu_1 + \lambda_2/\mu_2 < 1$.

It is recurrent if

(3.6) $\lambda_1 < \mu_1$ and $\lambda_1/\mu_1 + \lambda_2/\mu_2 = 1$.

The following notation will be used in the proof. Let $M = E(Z_{n+1}|Z_n = (x,y))$ for $x>0$; so $M_1 = (\lambda_1-\mu_1)/\Gamma$ and $M_2 = \lambda_2/\Gamma$. Let $M'' = E(Z_{n+1}|Z_n = (0,y))$ for $y>0$; then $M_1'' = \lambda_1/\Gamma''$ and $M_2'' = (\lambda_2-\mu_2)/\Gamma''$. Finally, let $M^o = E(Z_{n+1}|Z_n = (0,0))$; so $M_1^o = \lambda_1/(\lambda_1+\lambda_2)$ and $M_2^o = \lambda_2/(\lambda_1+\lambda_2)$. These vectors represent the mean displacement due to a single step of the chain. Note that the components M_1^o, M_2^o, M_2, and M_1'' are positive.

<u>Proof of Theorem 3.1</u>. Given a vector $W = (W_1,W_2)$, set $V = (W_1-1, W_2-1)$. If $|V_i| < 1$, then $W_i^{-1} = 1 - V_i + W_i^{-1}V_i^2$. Thus

$$
(3.7)\qquad E(W^{Z_{n+1} - Z_n}|Z_n = (x,y)) \begin{aligned}[t] &= 1 + (M,V) + (\mu_1/\Gamma)V_1^2W_1^{-1} && \text{if } x>0; \\ &= 1 + (M'',V) + (\mu_2/\Gamma)V_2^2W_2^{-1} && \text{if } x=0,\ y>0; \\ &= 1 + (M^o,V) && \text{if } x=y=0. \end{aligned}
$$

If it is possible to choose the vector V so that each of the inner products (M,V) (M'',V), and (M^o,V) are strictly negative, then by choosing $||V||$ sufficiently small, it is possible to have the entire right side of equation (3.7) strictly negative.

Now, if (3.2) holds, apply (2.9) with $W = (1-V_1,1)$ for $V_1\epsilon(0,1)$ sufficiently small to conclude that the chain is transient.

If either (3.3) or (3.4) holds, it is possible to choose V_1 and V_2 from the interval (-1,0) so that $(M,V) < 0$ and $(M'',V) < 0$. For such a vector V, automatically $(M^o,V) < 0$, so again it follows from (2.9) that the chain is transient.

If (3.5) holds, it is possible to choose V_1 and V_2 from the interval (0,1) so that $(M,V) < 0$ and $(M'',V) < 0$. With $||V||$ small enough and the exceptional set $A = \{(0,0)\}$, it follows from (2.13)that the chain is positive recurrent.

If (3.6) holds, then the vectors M and M'' point in opposite directions. Let $U = (M_2,-M_1)$. This vector has positive components so the random variables $\phi(Z_n) = (Z_n,U)$ are non-negative. With $A = \{(0,0)\}$ as the exceptional set, it follows from (2.6) that the chain is recurrent.

Example B: Pairs of Queues in Parallel. Consider a counter where two servers wait on arriving customers. The servers work independently but at equal rates. An arriving customer joins the shorter of the two lines; if both lines have the same length, he is equally likely to join either. It is assumed that the inter-arrival times have the exponential distribution with rate λ and that the service times have the exponential distribution with rate μ.

Let $Z_n = (X_n,Y_n)$ denote the number of customers to be handled by the servers, just after the n-th change of state. The transition function of this discrete-time chain is shown in Figure 3.1. There it is assumed that the time scale has been fixed so that $\lambda + 2\mu = 1$.

Theorem 3.8. The discrete parameter chain associated with the pair of queues in parallel is transient if $\lambda > 2\mu$. It is positive recurrent if $\lambda < 2\mu$; and it is recurrent if $\lambda = 2\mu$.

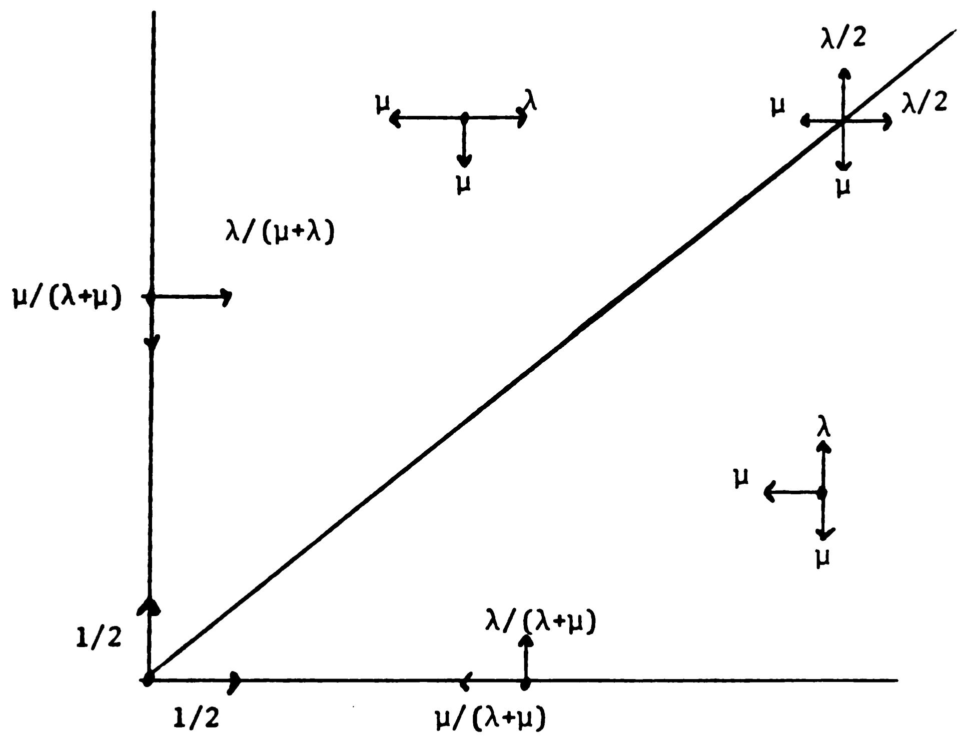


Figure 3.1

Transition Probabilities for the Parallel Queues

Let $M_n = E(Z_{n+1} - Z_n)|Z_n)$ and note that

$$(3.9)\quad \begin{aligned} M_n &= (\lambda/2 - \mu, \lambda/2 - \mu) && \text{if } X_n = Y_n \text{ and } X_n > 0; \\ &= (-\mu, \lambda-\mu) && \text{if } X_n > Y_n > 0; \\ &= (\lambda-\mu, -\mu) && \text{if } Y_n > X_n > 0; \\ &= (1/(\lambda+\mu))(-\mu, \lambda) && \text{if } Y_n = 0 \text{ and } X_n > 0; \\ &= (1/(\lambda+\mu))(\lambda, -\mu) && \text{if } X_n = 0 \text{ and } Y_n > 0; \\ &= (1/2, 1/2) && \text{if } X_n = Y_n = 0. \end{aligned}$$

Proof of Theorem 3.8. If $\lambda > 2\mu$, then the inner product of any vector in the direction of (-1,-1) with each of the vectors displayed in (3.9) is negative. Let $V_1 \varepsilon (0,1)$ and set $V = (-V_1, -V_1)$ and $W = (1-V_1, 1-V_1)$. For $n \geq 0$,

$$E(W^{Z_{n+1} - Z_n}|Z_n) = 1 + (M_n, V) + V_1^2\Theta_n,$$

where $\{\Theta_n : n \geq 0\}$ is a sequence of uniformly bounded random variables. Thus, if V_1 is sufficiently small, then almost surely:

$$E(W^{Z_{n+1} - Z_n}|Z_n) < 1 \quad \text{for } n \geq 0.$$

That the chain is transient follows from (2.1).

To conclude that the chain is positive recurrent, apply Foster's criterion with the function $\phi(z) = x + y$ and the exceptional set $A = \{(0,0)\}$, if $\lambda < \mu$. If $\mu < \lambda < 2\mu$, the function $\phi(z) = x^2 + xy + y^2$ will do.

In the critical case, consider the sequence of random variables (X_n^*, Y_n^*)

defined by $X_n^* = \min(X_n, Y_n)$ and $Y_n^* = \max(X_n, Y_n)$. (See Figure 3.2 for the transition diagram of the chain (X_n^*, Y_n^*).) It follows from (2.6) with the exceptional set $\{(0,0)\}$ and the function $\phi(z) = 2x + y$ that the chain $\{(X_n^*, Y_n^*): n \geq 0\}$ is recurrent. Since $(X_n^*, Y_n^*) = (0,0)$ if and only if $(X_n, Y_n) = (0,0)$, it follows that the original pair of queues is also recurrent.

Example C: Two-Dimensional Positive Random Walks. Recall that Z_+^2 is the lattice in R^2 of points whose coordinates are non-negative integers. This lattice can be decomposed into four types of states: interior points, for which both coordinates are positive; boundary points whose first coordinate is positive, boundary points whose second coordinate is positive; and the origin. The transition function for a two-dimensional positive random walk is homogenenous with respect to these four types of states, and single steps of the chain lead only to neighboring points in Z_+^2. Let

$$p_{ij} = P\{Z_{n+1} = (x+i, y+j) | Z_n = (x,y)\} \quad \text{for } x > 0,\ y > 0;\ |i| \leq 1,\ |j| \leq 1;$$
$$p'_{ij} = P\{Z_{n+1} = (x+i, j) | Z_n = (x,0)\} \quad \text{for } x > 0,\ |i| \leq 1,\ 0 \leq j \leq 1;$$
$$p''_{ij} = P\{Z_{n+1} = (i, y+j) | Z_n = (0,y)\} \quad \text{for } y > 0,\ |j| \leq 1,\ 0 \leq i \leq 1;$$
$$p^0_{ij} = P\{Z_{n+1} = (i, j) | Z_n = (0, 0)\} \quad \text{for } 0 \leq i,j \leq 1.$$

Criteria for recurrence and transience can be stated in terms of the drift vectors due to single steps of the chain. Let $M = E(Z_{n+1} - Z_n | Z_n = (x,y))$ for $x > 0$ and $y > 0$; and let $M' = E(Z_{n+1} - Z_n | Z_n = (x,0))$ for $x > 0$, $M'' = E(Z_{n+1} - Z_n | Z_n = (0,y))$ for $y > 0$. Then:

Theorem 3.10. When $M_1 < 0$ and $M_2 < 0$, the two-dimensional positive random walk is positive recurrent if both of the determinants

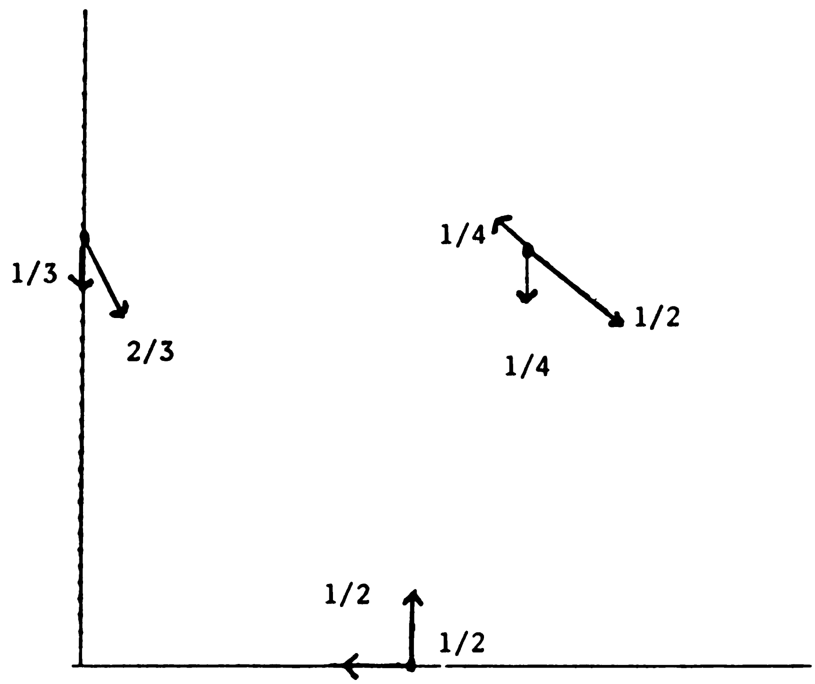


Figure 3.2

Transition Probabilities for $\{(X_n^*, Y_n^*)\}$

$$\begin{vmatrix} M_1 & M_1' \\ M_2 & M_2' \end{vmatrix} \qquad\qquad \begin{vmatrix} M_2 & M_2'' \\ M_1 & M_1'' \end{vmatrix}$$

are strictly negative. It is recurrent if at least one of the determinants is strictly negative.

Proof. It is assumed that sufficiently many of the terms of the distributions p, p', p'', and p^0 are positive that the chain is irreducible.

The criteria for transience follow from Theorem 3 of Kesten (1976).

Consider the function $\phi(z) = ax^2 + 2bxy + cy^2$. Note that if $b^2 \leq ac$, and if $a \geq 0$ and $c \geq 0$, then ϕ takes only non-negative values. For $n \geq 0$:

$$E\{\phi(Z_{n+1})|Z_n\} = \phi(Z_n) + ((2aX_n + 2bY_n, 2bX_n + 2cY_n), M_n) + E\{a(X_{n+1}-X_n)^2 + 2b(X_{n+1} - X_n)(Y_{n+1}-Y_n) + c(Y_{n+1}-Y_n)^2|Z_n$$

Since a single jump from Z_n leads only to neighboring lattice points, it follows that:

$$E\{\phi(Z_{n+1})|Z_n\} \leq \phi(Z_n) + a + 2|b| + c + ((2aX_n + 2bY_n, 2bX_n + 2cY_n), M_n).$$

Moreover:

$$\begin{aligned} ((2aX_n + 2bY_n, 2bX_n + 2cY_n), M_n) &= 2X_n((a,b),(M_1',M_2')) \quad \text{if } Y_n=0,\ X_n>0; \\ &= 2Y_n((b,c),(M_1'',M_2'')) \quad \text{if } Y_n>0,\ X_n=0; \\ &= 2X_n((a,b),(M_1,M_2)) + 2Y_n((b,c),(M_1,M_2)) \\ &\qquad \text{if } X_n \cdot Y_n > 0. \end{aligned}$$

Thus, if the numbers a,b, and c can be chosen so that $a>0$, $c>0$, and $b^2< ac$; and both

$$(3.11)\quad ((a,b),\ (M_1,M_2)) < 0 \quad\text{and}\quad ((a,b),\ (M_1',M_2')) < 0;$$

as well as

$$(3.12)\quad ((b,c),\ (M_1,M_2)) < 0 \quad\text{and}\quad ((b,c),\ (M_1'',M_2'')) < 0;$$

then there is a finite subset A of Z_+^2 such that

$$\sup_{z\notin A} E\{\phi(Z_{n+1})|Z_n = z\} < \phi(z) - 1.$$

In this case, that the chain is positive recurrent follows from Foster's theorem. Given that both determinants are strictly negative, there is a vector (a,b) with $a > 0$ and $b<0$ for which (3.11) holds and a vector (b,c) with b the number already chosen and $c>0$ such that (3.12) holds. Since $((a,b),\ (M_1,M_2)) < 0$ and $((b,c),\ (M_1,M_2)) < 0$, it follows that

$$-a|M_1| + |b||M_2| <0 \quad\text{and}\quad |b||M_1|- c|M_2| < 0;$$

Thus $|b|/a < |M_1/M_2|$ and $|b|/c < |M_2/M_1|$ Finally, $b^2 < ac$; and this completes the proof.

Example D: Birth and Death Processes. Let $\{N(t): t\geq 0\}$ denote a Markov chain with continuous time parameter whose state space is Z_+^2. For the two examples con-

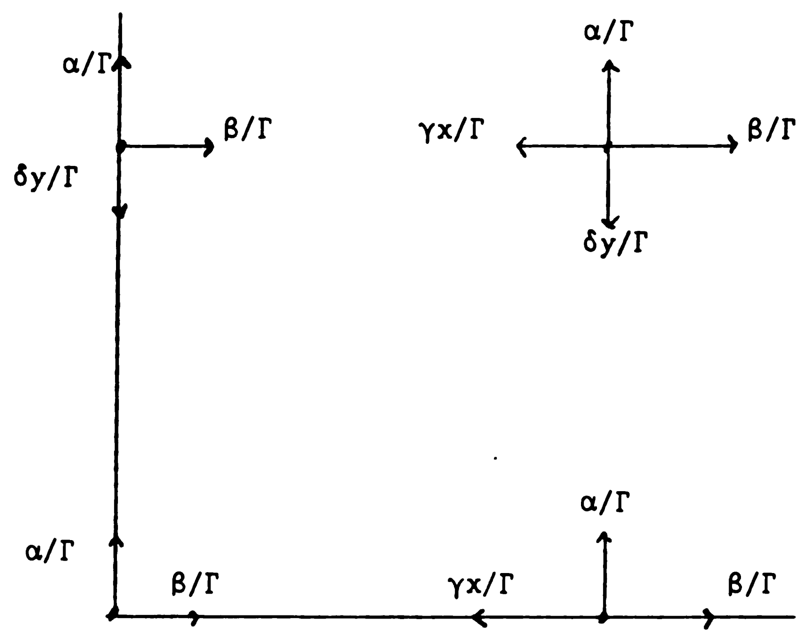


Figure 3.3

Transition Probabilities for

Reuter's Birth-and-Death Processes

sidered here, with probability one, the sample paths of the chain have only finitely many jumps in a finite time interval. If $N(t) = z$, then a single jump leads only to one of the neighboring points in Z_+^2 of the form $z \pm e_i$. (Here e_i denotes the unit vector for which the i-th coordinate is equal to one and all of the other coordinates are equal to zero.) Kesten (1976) has given criteria for the recurrence and transience of two-dimensional birth-and-death processes with linear transition rates. Milch (1968) used generating functions of several variables to study special examples of birth-and-death processes with linear transition rates. Other examples of multi-dimensional birth-and-death processes have appeared in the literature with the name "competition processes." Such processes have been studied by Iglehart (1964) and by Reuter (1961).

In the following examples, the results of Section Two lead to simple proofs of criteria for recurrence and transience.

Suppose that the transition rates are given by:

$$P\{N(t+h) = (x+1,y) \;|N(t) = (x,y)\} = \alpha h + o(h);$$
$$P\{N(t+h) = (x,y+1) \;|N(t) = (x,y)\} = \beta h + o(h);$$
$$P\{N(t+h) = (x-1,y) \;|N(t) = (x,y)\} = \gamma x h + o(h);$$
$$P\{N(t+h) = (x,y-1) \;|N(t) = (x,y)\} = \delta y h + o(h).$$

This example was considered by Reuter (1961). If the pair (x,y) of non-negative integers represents the sizes of twc populations, then these transition probabilities correspond to a process in which the populations grow independently.

Suppose that the numbers α,β,γ, and δ are strictly positive. Then all states of the chain communicate. Let $\{Z_n: n\geq 0\}$ be the embedded Markov chain of the successive states of the continuous time process. Its transition function is displayed in Figure 3.3. Note that the normalizing constant $\Gamma(x,y)$ is defined by $\Gamma(x,y) = \alpha+\beta+\gamma x+\delta y$.

Proposition 3.13. This embedded Markov chain is positive recurrent.

Proof. Let $V = (V_1,V_2)$ be a vector of R^2 for which $V_1 > 0$ and $V_2 > 0$. Set $W = (1+V_1,1+V_2)$ and note that:

$$E(W^{Z_{n+1} - Z_n} \mid Z_n = (x,y)) = 1 + (\alpha-\gamma x)V_1/\Gamma + (\beta-\delta y)V_2/\Gamma + K_{x,y}V_1^2 + K'_{x,y}V_2^2$$

where $K_{x,y} = W_1^{-1}\gamma x/\Gamma \leq 1$ and $K'_{x,y} = W_2^{-1}\delta y/\Gamma \leq 1$.

Thus

$$E(W^{Z_{n+1} - Z_n} \mid Z_n = (x,y)) \leq 1 + (\alpha-\gamma x)V_1/\Gamma + (\beta-\delta y)V_2/\Gamma + ||v||^2.$$

If $||v||$ is sufficiently small, there is a finite subset A of Z_+^2, necessarily containing $(0,0)$, for which

$$\sup_{z \notin A} E(W^{Z_{n+1} - Z_n} \mid Z_n = z) < 1.$$

It follows from (2.13) that the chain is positive recurrent.

For the final example, assume that the transition probabilities for the continuous parameter chain $\{N(t): t \geq 0\}$ are given by:

$$P\{N(t+h) = (x+1,y) \mid N(t) = (x,y)\} = ((1-\alpha)x + \beta y)h + o(h);$$

$$P\{N(t+h) = (x,y+1) \mid N(t) = (x,y)\} = (\alpha x + (1-\beta)y)h + o(h);$$

$$P\{N(t+h) = (x-1,y) \mid N(t) = (x,y)\} = xh + o(h);$$

$$P\{N(t+h) = (x,y-1) \mid N(t) = (x,y)\} = yh + o(h).$$

This example was proposed by Milch(1968). Here, the numbers α and β satisfy $0 < \alpha < 1$ and $0 < \beta < 1$. Note that $(0,0)$ is an absorbing state. At each other point (x,y), the sum of the birth rates is equal to the sum of the death rates.

Proposition 3.14. The embedded Markov chain for these transition rates is absorbed at $(0,0)$ with probability one.

Proof. Let $Z_n = (X_n, Y_n)$ denote the state after the n-th jump. For the function $\phi(z) = x + y$, observe that $E(\phi(Z_{n+1})|Z_n) = \phi(Z_n)$ if $Z_n \neq (0,0)$. Let T_0 denote the time of the first visit to $(0,0)$ of paths of the embedded Markov chain $\{Z_n: n \geq 0\}$. It follows that the stopped sequence $\{\phi(Z_{n \wedge T_0}): n \geq 0\}$ is a non-negative martingale. Thus, it converges almost everywhere. Such convergence is possible only if the stopping time T_0 is finite with probability one. This completes the proof.

Appendix A. Non-negative Supermartingales

Let (Ω, F, P) denote a probability space on which there is defined an increasing sequence $\{F_n: n \geq 0\}$ of sigma-fields contained in F. A random variable is an extended real-valued function defined on the sample space Ω which is measurable with respect to the sigma-field F. Conditional expectations refer to the probability measure P.

Definition A.1. A sequence $\{W_n: n \geq 0\}$ of non-negative random variables is a non-negative supermartingale if, for $n \geq 0$, W_n is F_n-measurable and $E(W_{n+1}|F_n) \leq W_n$.

Theorem A.2. If $\{W_n: n \geq 0\}$ is a non-negative supermartingale, then the sequence W_n converges almost everywhere. The limit W satisfies the inequality $E(W) \leq E(W_0)$.

Note that the inequality $E(W) \leq E(W_0)$ implies that the limit is finite almost everywhere when the expected value of W_0 is finite.

Definition A.3. A random variable T, taking values from the set $\{n: n \geq 0\} \cup \{+\infty\}$, is a stopping time if, for each non-negative integer n, the set $\{T=n\}$ belongs to the sigma-field F_n.

Definition A.4. Let T be a stopping time and let $\{W_n: n \geq 0\}$ be a sequence of random variables with the property that W_n is F_n-measurable. The random variable W_T is defined on the set $\{T<\infty\}$ by $W_T = W_n$ when $T = n$.

In the next proposition, the notation $n \wedge T$ refers to the stopping time obtained by truncating T at the integer n.

Proposition A.5. Let $\{W_n: n \geq 0\}$ be a non-negative supermartingale and let T be a stopping time. Then the sequence $\{W_{n\wedge T}: n \geq 0\}$ is also a non-negative supermartingale.

Two references for this material are the books by Doob (1953) and Neveu (1975).

References

Aldous, D. J. (1981). Markov chains with almost exponential hitting times. Preprint.

Chung, K. L. (1967). Markov Chains with Stationary Transition Probabilities, second edition. Springer, New York.

Doob, J. L. (1953). Stochastic Processes. Wiley, New York.

Foster, F. G. (L951). Markoff chains with an enumerable number of states and a class of cascade processes. Proc. Cambridge Phil. Soc. 47, 77-85.

Foster, F. G. (1952). On Markov chains with an enumerable infinity of states. Proc. Cambridge Phil. Soc. 48, 587-591.

Foster, F. G. (1953). On the stochastic matrices associated with certain queuing processes. Ann. Math. Statist. 24, 355-360.

Freedman, D. (1971). Markov Chains. Holden-Day, San Francisco.

Iglehart, D. (1964). Multivariate Competition Processes. Z. Wahrscheinlichkeitstheorie 2, 314-331.

Kemperman, J. H. B. (1961). The Passage Problem for a Stationary Markov Chain. Univ. of Chicago Press, Chicago.

Kesten, H. (1976). Recurrence criteria for multi-dimensional Markov chains and multi-dimensional linear birth and death processes. Adv. Appl. Probability 8, 58-87.

Kesten, H. and Runnenburg, J. Th. (1957). Priority in waiting line problems, I and II. Indag. Math. 19, 312-336.

Kingman, J. F. C. (1961a). The ergodic behavior of random walks. Biometrika 48, 391-396.

Kingman, J. F. C. (1961b). Two similar queues in parallel. Ann. Math. Statis. 32, 1314-1323.

Malyšev, V. A. (1972). Classification of two dimensional positive random walks and almost linear supermartingales. Sov. Math. Dokl. 13, 136-139.

Mertens, J. F., Samuel-Cahn, E., and Zamir, S. (1977). Necessary and sufficient conditions for recurrence and transience of Markov chains, in terms of inequalities. Abstract 77t-141. Bull. Inst. Math. Statis. 6, 287.

Mertens, J. F. Samuel-Cahn, E., and Zamir, S. (1978). Necessary and sufficient conditions for recurrence and transience of Markov chains, in terms of equalities. J. Appl. Probab. 15, 848-851.

Milch, P.R. (1968). A multi-dimensional linear growth birth and death process. Ann. Math. Statist. 39, 727-754.

Neveu, J. (1975). Discrete-Parameter Martingales. North-Holland, Amsterdam.

Reuter, G.E.H. (1961). Competition Processes. Proc. Fourth Berkeley Symp. 2, J. Neyman, ed. U. California Press, Berkeley.

EVALUATION OF PARAMETERS IN ARMA ANALYSIS OF TIME SERIES BY A LEAST CHI-SQUARE METHOD

RICHARD L. MOORE AND FRANCIS J. LUZZI
SYSTEMS ANALYSIS DIVISION
REQUIREMENTS AND ANALYSIS OFFICE
US ARMY ARMAMENT RESEARCH AND DEVELOPMENT COMMAND
DOVER, NJ 07801

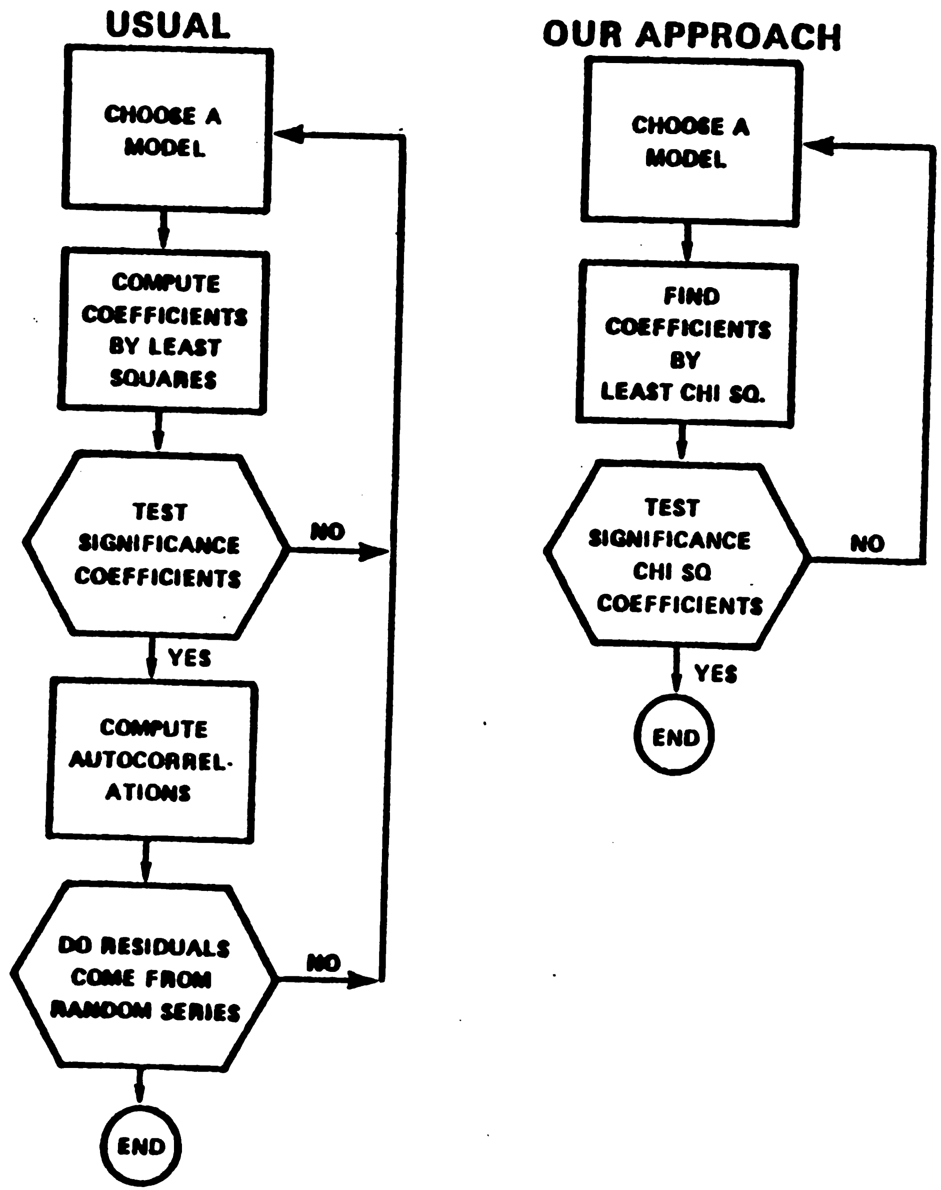
ABSTRACT. We have programmed a "least chi-square" procedure for multiple independent variables to analyze time series - especially those of the auto-correlated moving average type. We compare this procedure with the usual procedure theoritically and experimentally. The new procedure combines in one criterion the previous method which uses two independent criteria in a serial manner - first one and then the other. (These two criteria are the sum of squares of the residuals and the Box-Pierce test for their randomness.) Our analysis of three well-known series illustrates the advantage of our method. The results indicate that it obtains a more probable set of parameters than the older method. We conclude that the method has promise in simplifying the fitting of time series. By having a single measure of goodness of fit, the latitude for variation in the choice of fit due to individual judgment appears to be reduced.

I. INTRODUCTION. In the usual methods to find the best-fitting model for a time-series[1], the procedure follows the schematic shown on the left side of Figure 1. The investigator chose a mathematical model as an initial hypothesis. He uses a computer program which finds the unknown coefficients in the model using the principle of least squares to estimate the best coefficients. If the least squares test of significance indicates a reasonably good fit, he then computes the autocorrelation coefficients, forms them into a weighted sum of squares of autocorrelation coefficients (a criterion called Box-Pierce number) and uses the resulting figure of merit to test whether the residuals are consistent with sampling from a series each of whose terms is selected randomly (or independently) from an error population which has a normal distribution. If the residuals don't pass this test - (or any of the other tests which may be used for the same purpose) then he revises the mathematical model he has used. The new model is selected to reduce those values of the autocorrelations which are largest. In this way he uses a two step procedure: First, minimize the squares of the residuals and then minimize the Box-Pierce number. (For an easy to understand review of this process see Roberts(2)).

On the other hand, one of us (3,4) has shown that the two steps can be combined by adopting a combined criteria: The sum of squares of the residuals normalized by dividing by the measured or estimated variance of measurement, (this has a Chi-Square distribution.), plus the Box-Pierce criteria (which also has a Chi-Square distribution, if the residuals are independent). In this paper, we report a computer program developed to obtain the best fitting coefficients of a given model using the above least chi-square criteria. Our approach to fitting time series is to use this program to help determine, if not the best fitting models, at least a good fitting model.

FIGURE 1

ANALYSIS OF TIME SERIES



In this paper, we will provide a summary of the mathematics used to program our central CDC-6500 computer. We will give examples of data fitting on three well-known series - two from Box-Jenkins (1) and one from Roberts (2), and compare the results with that reached by the usual approach.

II. MATHEMATICAL FORMULATION. We let the mathematical model be such that the theoretical values are represented by the n dimensional vector y^* and the independent variables: x, u, v, w,....

$$y^* = \theta_0 + \theta_1 x + \theta_2 u + \theta_3 v + \quad . . .$$

$$\text{Or} \quad y^* = \theta x$$

We let y be the vector of observed values of y^*, and x_i, u_i, v_i be values of the appropriate series at the "ith" observation. We will refer the reader to previous work (4) for a detailed derivation. In that we differentiate the criteria CHI SQ TOT (χ_T^2) with respect to each of the unknown coefficients and set them to zero to find the following equation in matrix form for the estimated values of θ, denoted as θ*.

$$\frac{\partial \chi_T^2/2}{\partial \theta^*} = \sigma_e^{-2} \{(P' \Gamma P\theta^*) - P' \Gamma y\} = 0$$

where

$$P' = \begin{vmatrix} 1 & 1 & 1 & 1 & \cdots \\ x_1 & x_2 & x_3 & x_4 & \\ u_1 & u_2 & u_3 & u_4 & \\ & \cdots & & & \end{vmatrix}$$

$$\text{and} \quad \alpha_j = \frac{+\; 2r_j/\upsilon_j}{(d)'(d)/\sigma_e^2 - 2\sum_{i=1}^{s} (r_i)^2/\upsilon_i}$$

$$\Gamma = I + \sum_{i=1}^{s} \alpha_i [V_i^{-1} + V_{-i}^{-1}]$$

I is the n dimensional unit matrix. V_i^{-1} is the matrix resulting when we shift the columns of I by "i" columns to the right and insert zeros in the columns which remain. r_i is the "i" th autocorrelation. σ^2 is the variance of the measurement error. d is the vector of the residuals. υ_i is the variance of $(r_i)^2$. On multiplying out $P'\Gamma$ and $[P'\Gamma P]$ we find the usual least squares expression as given in Figure 2A and 2B, but modified by the addition of the terms which contain α in their expression.

All terms except the value of the α s can be evaluated at once from the data input. We assume initially each r_i is zero; calculate them from the

FIGURE 2A

LEFT SIDE OF REGRESSION EQUATION

$$\sum_{1}^{n} y_i + \sum_{p=1}^{s} \alpha_p \sum_{i=1}^{n-p} (y_i + y_{i+p}) = \theta_o \left(n + \sum_{p=1}^{s} 2(n-p)\alpha_p\right) +$$

$$\sum_{i=1}^{n} x_i y_i + \sum_{p=1}^{s} \alpha_p \sum_{i=1}^{n-p} (x_{i+p} y_i + x_i y_{i+p}) = \theta_o \left(\begin{array}{l}\text{SAME AS COEFFICIENT}\\ \text{OF } \theta_1 \text{ IN ROW 1}\end{array}\right) +$$

$$\sum u_i y_i + \sum_{1}^{s} \alpha_p \sum_{i=1}^{n-p} (u_{i+p} y_i + u_i y_{i+p}) = \theta_o \left(\begin{array}{l}\text{SAME AS COEFFICIENT}\\ \text{OF } \theta_2 \text{ IN ROW 1}\end{array}\right) +$$

FIGURE 2B

RIGHT SIDE OF REGRESSION EQUATION

$$\theta_1\left(\sum_{i=1}^{n} x_i + \sum_{p=1}^{s} \alpha_p \sum_{i=1}^{n-p} (x_i + x_{i+p})\right) + \theta_2\left(\sum_{i=1}^{n} U_i + \sum_{p=1}^{s} \alpha_p \sum_{i=1}^{n-p} (U_i + U_{i+p})\right) + \cdots$$

$$\theta_1\left(\sum_{i}^{n} x_i^2 + \sum_{p=1}^{s} \alpha_p \sum_{i=1}^{n-p} 2 x_i x_{i+p}\right) + \theta_2\left(\sum_{i=1}^{n} U_i x_i + \sum_{i}^{s} \alpha_p \sum_{i=1}^{n-p} (U_i x_{i+p} + U_{i+p} x_i)\right) + \cdots$$

$$\theta_1\left(\begin{array}{l}\text{SAME AS COEFFICIENT}\\ \text{OF } \theta_2 \text{ IN ROW 2}\end{array}\right) + \theta_2\left(\sum_{i}^{n} U_i^2 + \sum_{p=1}^{s} \alpha_p \sum_{i=1}^{n-p} 2 U_i U_{i+p}\right) + \cdots$$

autocorrelations of the residuals, and iterate to find the new values of the vector θ^*. We usually chose an incremental decrease of 1% or less in the value of CHI-SQ TOT to stop iteration; however this criteria is an input parameter on each run. The significance of the value of CHI SQ TOT has been calculated from the formula

$$\text{Significance} = \sqrt{2\chi_T^2} - \sqrt{2(n + s - p) - 1}$$

where p is the number of parameters, and s is the number of autocorrelations used. This is well-known to have a normal distribution with a mean of zero and a standard deviation of 1. It will be used to compare the results of the computations using the same initial data but possibly different numbers of data points, autocorrelations and parameters. If the "significance" of a given model has a larger positive value then another, it is less significant.

III. EXAMPLES OF TIME SERIES. In order to provide the reader with some readily available data to which we have applied our technique, we used two series supplied by Box and Jenkins (1). Figure 3 shows a plot taken from Figure 4.1 of reference (1). We have analyzed both Series C and Series D. In addition we have analyzed the GNP data of Roberts. We will begin with the Series C.

III-1. TIME SERIES C. In Figure 3 this series appears quite smooth - from this we might expect that a rather long term lag might be present in addition to short term fluctuals (and indeed we found such a long term lag). We had no prior knowledge of the error of measurement for this series. We can, nevertheless, estimate the lowest possible value of this error. We do this from the observation that only a simple decimal point accuracy is carried in the data analysis. The digitizing error - assuming a uniformly distributed population - is estimated as .029. Later in the process we examined the sensitivity of the results to this assumption, by multiplying the error by five thus increasing it to .145.

Turning now to our analysis, the autocorrelations in the four cases selected to illustrate our results are listed in Table 1. The first step in fitting this series was to analyze the data in terms of the first and second lagged series. The results as indicated for Case 10 in Table 2 (Series C Results) indicate a rather low level of significance (73.8) (see section II for the definition of "significance") with the largest part of CHI SQ total coming from the sume of squares of the residuals. The largest terms in the autocorrelations are of rank 9, 10, 11, and 12. (See Table 1.) The next step, we computed case 14, with the estimated variance increased by a factor of 5. This reduced the value of CHI SQ 1 to 180.1; the value of CHI SQ 2 increased basically due to the increase in the number of autocorrelations from 12 to 20. The autocorrelations of rank 9, 10, 11 and 12 remained the largest.

Case 1 in Table 2 shows the effect of changing the model to using a differenced independent variable and the first, eighth, and ninth lagged series as predictors (using again the small value (.029) of the "error of measurement"). The value of CHI SQ 2 is substantially reduced; and the sum

FIGURE 3

SOME TYPICAL TIME SERIES ARISING IN FORECASTING AND CONTROL PROBLEMS

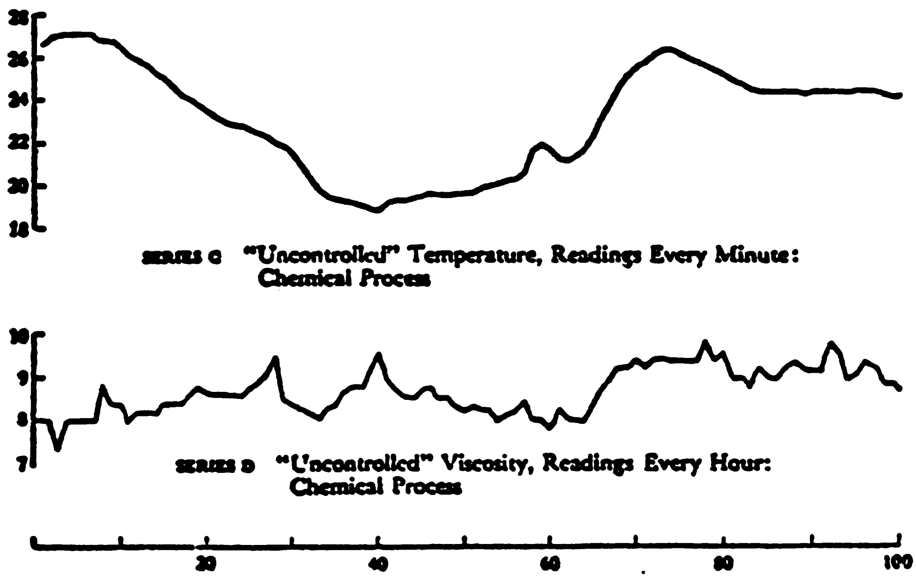


TABLE 1 AUTOCORRELATIONS FOR VARIOUS MODELS FOR SERIES C

Case	1	4	10	14
Rank				
1	.0338	.0049	-.0193	.0093
2	.0140	.0086	-.0018	.0194
3	-.099	-.0597	-.0648	-.0472
4	-.0277	-.0194	-.0203	-.0077
5	.0654	.0641	.0607	.0684
6	-.0134	.0165	.0197	.0261
7	.0477	.0637	-.0688	.0715
8	.0120	-.0267	-.0296	-.0256
9	-.0157	-.0876	-.0844	-.0800
10	.0908	.1327	.1431	.1397
11	-.1504	-.1126	-.1143	-.1101
12	.0576	.0922	.1061	.103
13				-.0726
14				.0480
15				.0114
16				-.0549
17				.1892
18				-.0884
19				-.0014
20				-.0131

TABLE 2 SERIES C RESULTS

		1	4	10	14
CASE		1	4	10	14
Independent Variables Lagged	1			X	X
	2			X	X
Differenced	1	X	X		
	8	X			
	9	X			
Auto-correlations		12	12	12	20
Error		.029	.029	.029	.145
CHI SQ 1		$4.3.10^3$	$4.7.10^3$	$4.5.10^3$	180.1
CHI SQ 2		10.25	12.3	13.8	23.8
CHI SQ TOT		$4.3.10^3$	$4.7.10^3$	$4.5.10^3$	205.1
Significance		71.9	75.5	73.8	-1.54
Coefficient	1	-.004	-.008	.2745	.231
	2	.817	.816	1.807	1.78
	3	-.107	-	-.819	-.792
	4	.113	-	-	-

of squares is slightly reduced. The 9th, 10th, and 12th autocorrelations (Table 1) are slightly reduced, but the 11th autocorrelation is increased. The "significance" is improved; since the value decreased from 73.8 to 71.9. If this same case is run without the 8th and 9th lag, the values for case 4, indicate that the CHI SQ TOT increases substantially, even above the Case 10. This indicates the significance of the coefficients 3 and 4 for Case 1, since that is the only difference between the two cases. In Case 4, the autocorrelations of rank 9, 10, 11, and 12 are all comparable with Case 10.

Comparison with Box-Jenkins. Case 4 can be compared with one of the results of Box-Jenkins. In their terminology this series was given as a single difference in the time series z_t, (∇z_t), as a function of the once-lagged difference ∇z_{t-1}, and the residuals a_t. They found

$$\nabla z_t = a_t + .82 \nabla z_{t-1}$$

In our least squares analysis we found that an additional constant value of .00769 could be added to the right of this term, and that the coefficient of ∇z_{t-1} was .8153. Using LCS these values were changed only slightly. The result which we have obtained previously that Case 1 is more significant than Case 4 applies equally well to the conclusion that Case 1 is more significant than the Box-Jenkins result. In Box-Jenkins terms our result is:

$$\nabla z_t = a_t - .004 + .817 \nabla z_{t-1} - .107 \nabla z_{t-8}$$

$$+ .113 \nabla z_{t-9}$$

We stopped the analysis of this series at this point, although we might have well considered the addition of a lag of 10, and possibly 11 to the lagged variables used in Case 1, because of the rather large values of the autocorrelations of rank 10 and 11 in Table 1.

<u>III-2. SERIES D</u>. As in the case of Series C, we initially used two lagged series to determine the regression of the series. We see from Table 3 that all the autocorrelations from 1 to 20 are less than .1, however we find that if we use only <u>one</u> lagged parameter our fit is better since the significance is better. (In case 2, we used fewer terms than in Case 1, so that a direct comparison of the values of CHI SQ TOT would be misleading.) When we added two differenced parameters to the calculation, as in Case 5, we found an increase in the value of CHI SQ 2, as indicated by the increased value of the 2nd to 4th autocorrelation coefficients. Other changes have been made here, the number of autocorrelations has been reduced to six from 20, and the assumed measurement error has been reduced to .029 from .030. It is surprising that the use of

TABLE 3 AUTOCORRELATIONS FOR VARIOUS MODELS FOR SERIES D

Rank \ Case	1	2	5	7	8
1	-.0050	.0255	.0072	-.0145	-.0080
2	-.0034	-.0036	-.0695	-.0667	-.0676
3	-.0161	-.0219	-.0693	-.1653	-.0667
4	-.0328	-.0326	-.1035	-.1009	-.1018
5	-.0258	-.0011	-.04257	-.0396	-.0407
6	.0345	.0154	-.0150	-.0135	-.0141
7	.0062	-.0081			
8	.0052	.0103			
9	-.0103	.0114			
10	-.0476	-.0579			
11	-.0014	-.0143			
12	.0531	.0528			
13	.0303	.0089			
14	-.0622	-.0634			
15	-.0233	-.0152			
16	-.0769	-.0720			
17	-.0118	-.0098			
18	.0617	.0554			
19	-.0004	.0056			
20	.0723	.0866			

TABLE 4 SERIES D RESULTS

CASE	1	2	5	6	8
Independent Variables					
Lagged 1	X	X	X	X	X
Lagged 2	X				
Differenced 1			X	X	X
2			X		
Auto-correlations	20	20	6	6	6
Error	.030	.030	.029	.290	.290
CHI SQ 1	$3.1.10^4$	$2.9.10^4$	$3.3.10^4$	$3.3.10^2$	$3.3.10^2$
CHI SQ 2	8.34	8.21	6.409	5.957	6.08
CHI SQ TOT	$3.1.10^4$	$2.9.10^4$	$3.3.10^4$	$3.3.10^2$	$3.4.10^2$
Significance	222	217	232	1.547	1.503
Coefficient 1	1.305	1.354	.00056	.0028	0.0
2	.873	.852	-.044	-.028	-.028
3	-.016	-	-	-	-

lagged difference independent variables does not improve the fit. On increasing the expected measurement error by a factor of ten when using one lagged and one differenced variable, as in Cases 6 and 8, we find a slight reduction in CHI SQ 2 since the program puts more emphasis on it in such a case. On setting the mean arbitrarily to zero (i.e. coefficient 1 is zero, as in Case 8) we find we can make the usual argument that the number of adjustable parameters is less than before. We interpret the reduction of significance in Cases 1, 5, and 6 (as compared to 2, and 8) as due to the fact the additional terms add "noise" to the theoretical value. Similarly, we attribute the relative increase of CHI SQ 2 from Case 2 to Case 8, to the increased noise introduced by using the first difference. Case 2 is an approximation to the result of Box-Jenkins who found:

$$z_t = .87z_{t-1} + 1.17 + a_t$$

A subset of the basic data was used so that the difference between the coefficients: .87 vs .85 and 1.17 vs 1.35; is not surprising.

III-3. GROSS NATIONAL PRODUCT. The data we used in our studies were given by Roberts in Column 1 of his Table 11-1. We chose this set to analyze because Roberts has given a detailed analysis of these data which is easy for anyone to follow and with which we can make a simple but meaningful comparison. Our first step in the discussion is to present and compare least-square and LCS computations. We have repeated Roberts LS analysis and used the results both as validation of our computer program and as a basis of comparison between Roberts' results and ours.

To show the difference between the least square, and the least chi-square results, we have prepared Tables 5 and 6. Consider the residuals in Table 5 for four cases of interest: 1 through 4. In Case 1, the autocorrelations of orders of 2, 4, 6, and 11 are large (greater than .1). In Case 2, rank 6 is reduced but 5 and 7 are increased. On going to Case 3, ranks 2, 4, 5, 7 and 11 are still large and the Box-Pierce coefficient is larger than Case 2. For Case 4, ranks 2, 3, and 8 are large, but the Box-Pierce coefficient, (CHI SQ 2) has become less than half of Case 1.

In Table 6, Cases 1-4 are given to compare least squares (LS) and least chi square (LCS). In all cases of LCS the value of the sum of squares of the residuals is slightly increased thus increasing CHI SQ 1. On the other hand, in every case CHI SQ 2 is substantially decreased, and results in a decreased value of CHI SQ TOT. The regression coefficients for the lagged variables change in all cases and occasionally change by a substantial amount, as would expect on going from the LS to the LCS analysis. The column labelled "L" in Table 6 indicates the amount of lag which has generated the independent series which corresponds to the coefficient given in the column next to it.

Based on results of the first case the variable lagged by 7 (corresponding to coefficient 4) was replaced by the series which lagged by 8. The difference between Case 2 and Case 3 is that in the former, the error was assumed as .06 while the latter it was assumed as .04. Thus Case 2 puts more weight on the reduction of CHI SQ 2 than does Case 3, and the results indicate the same. The regression coefficients are little different in Case 3 from those in Case 2.

TABLE 5 AUTOCORRELATIONS FOR VARIOUS MODELS FOR GNP SERIES

Rank \ Case	1	2	3	4	11	12	14*
1	-.0843	.0685	.07126	.0117	-.0472	.0781	-.0294
2	.2175	.2348	.2371	.1786	-.0022	-.0289	.0146
3	.0871	.0136	.0159	-.1633	-.0764	-.0734	-.1457
4	-.3087	-.2261	-.2377	-.0696	-.1168	-.1209	-.1100
5	.0799	.1553	.1618	.0754	-.0024	-.0771	-.0183
6	-.1935	-.0198	-.0148	.0026	.0413	.0482	.0387
7	-.0282	.1379	.1494	.0728	-.0905	-.1357	-.0537
8	.0389	.0827	.0767	-.1322	.1864	-.1929	-.1717
9	.0031	.0490	.0564	.0311	-.0112	-.0007	.0061
10	.0243	.0223	.0252	.0448	-.0704	-.0794	-.0618
11	-.1179	-.1345	-.1317	-.0733	-.0752	-.0673	-.0782
12	.0790	-.0200	-.0308	-.0237	.0144	-.0045	-.0063

* The autocorrelations of rank 13-20 are: -.0726, .0480, .0114, -.0549, .1892, -.0884, -.0014, and -.0131

TABLE 6 GNP SERIES RESULTS

Case	1			2			3			4		
	L	LS	LCS	L	LS	LCS	L	LS	LCS	L	LS	LCS
Coefficient												
1	-	-.0392	-.0395	-	-.0547	-.0569	-	-.0547	-.0557	-	.0153	-.0102
2	1	1.117	1.064	1	.9335	.9331	1	.9335	.9337	1	.9592	.9528
3	2	-.3886	-.3298	2	-.1353	-.1349	2	-.1353	-.1356	9	-.3877	-.3756
4	7	.0095	.0314	8	.2539	.2414	8	.2539	.2482	8	.6170	.5999
5	4	.9438	.9197	4	.7172	.7085	4	.7172	.7150	4	.3611	.3721
6	5	-1.1226	-1.078	5	-.9330	-.9219	5	-.9330	-.9287	5	-.5843	-.5854
7	6	.4543	.4066	6	.1835	.1938	6	.1835	.1883	6	.0454	.0464
8												
CHI SQ 1		257	259.		227.2	227.9		511.2	511.41		191.4	191.5
CHI SQ 2		20.7	16.9		15.9	13.57		15.0	14.5		7.8	7.4
CHI SQ TOT		278	275.9		242.2	241.5		526.2	525.9		199.2	198.9
Significance												6.42

In Case 4, the only change from Case 2 was the use of the 9th lagged series in place of the second lagged series. This change was made because the coefficient of the twice lagged series was small; we replaced the twice lagged series with the 9 times lagged series to attempt to reduce the large autocorrelation at ranks 4 and larger, which none of the previous models had been able to do. This change made a big improvement (from Case 2 to Case 4).

Since the coefficient of the sixth term is now small, when we continue, we will eliminate it and test to see if using the twice-lagged series might reduce the second, third and ninth autocorrelations.

In order to compare with Roberts' analysis of GNP, we took the 4th difference of the first difference and obtained a dependent variable which is "twice" differentiated". (The autocorrelations for Case 13 are not shown in Table 5 since they were similar to those of Case 14.) Of the four cases analyzed (refer to Tables 5 and 7) the autocorrelations and the Box-Pierce number (CHI SQ 2) of Case 11, are smaller than those of the others. In 12, the second lagged variable was omitted, and the significance test applied. The results indicate that Case 11 is slightly more significant than Case 13, and more significant than Case 12.

Case 13 was used as a base case to find the effect of varying the "error" or estimated precision of measurement of the transformed GNP. The number .06 corresponds to an estimate of a 1% error in the measurement of the GNP. Correspondingly Case 14 corresponds to an approximate error of 1½% in the GNP. The estimated measurement error was introduced in a progression of computations going from .03, .06, .12, .24, to .48. The "deviation" from the expected value was plotted on Figure 4, as a function of the assumed error. An iterative procedure was used to interpolate on the curve around the "0" deviation value; Case 14 was the result. We infer from this curve that approximately .093 is the "best" estimate of the standard deviation of the measurement of the GNP.

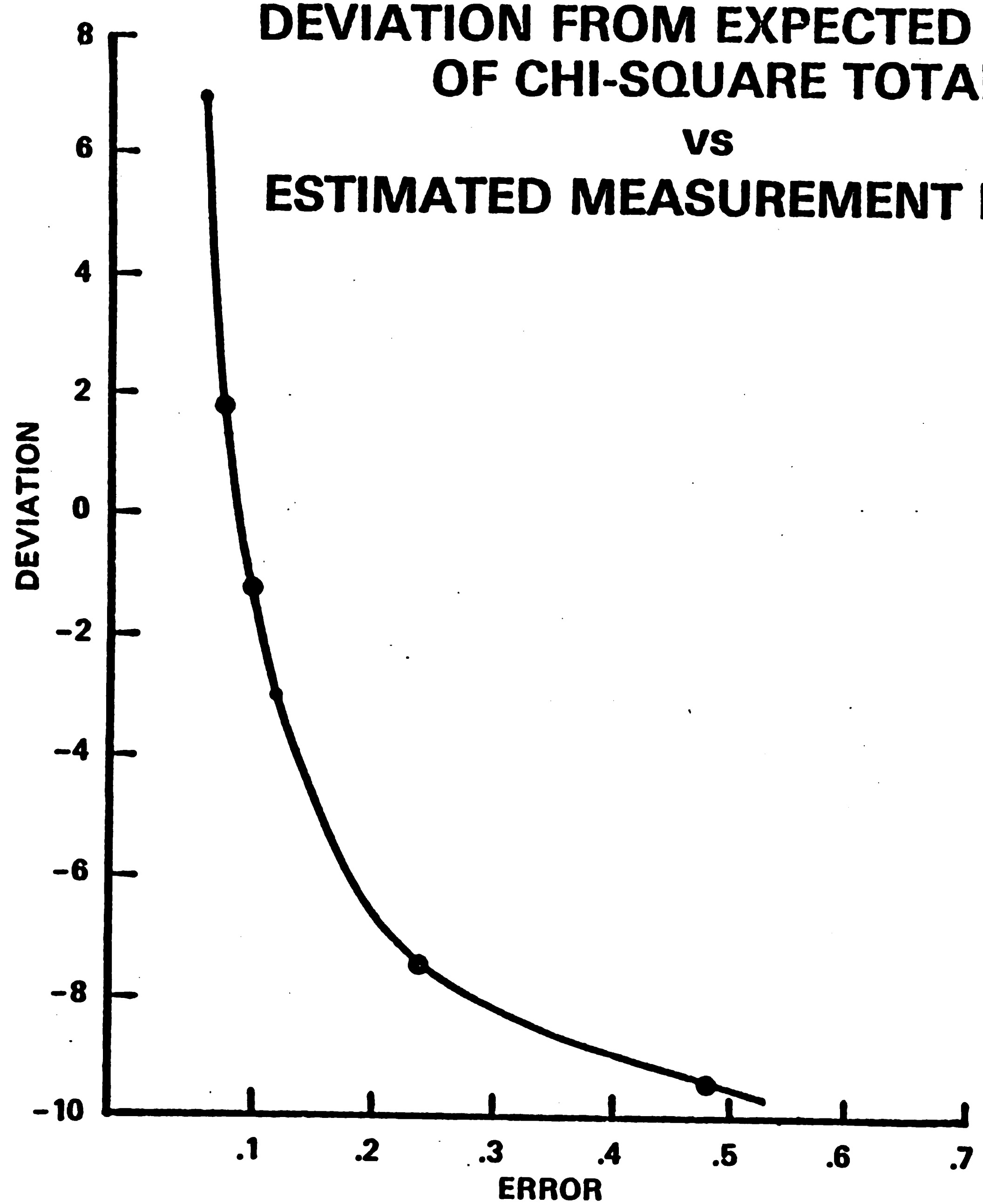
When we compare the results of the analysis of the twice differentialed series of Case 11 with that of the single differentiated series of Case 4 we observe that Case 4 has a smaller value of the significance parameter 6.42, as compared with 6.89. Thus if the assumptions behind these two cases are valid, we would conclude that the use of the single differenced procedure is better than the twice differenced procedure. However, we prefer to conclude that there is no real difference between the two. This conclusion is subject to a further caveat that the comparison depends crucially on the value of the measurement error assumed; as we have seen from our previous example, its expected value is about .093. Further study of this matter is clearly indicated. The least squares analysis (Case 13) may be used to compare the Roberts' analysis of GNP with ours. Roberts found coefficients of (0, .1261, .2395, -.58); in our first iteration we found (.0033, .1242, .2362 and -.5712). The CHI SQ TOT was 13.68, and the significance is 7.02. Comparing this result with our final iteration of Case 13, we see that the significance is decreased by .10. Comparing with our best fit "Case 4", we found an increase of "significance" from 6.42 to 7.02. Thus we see our procedure, using the six lagged series gives a better fit since its "significance" is smaller.

TABLE 7 EFFECT OF LAGGED VARIABLES ON TWICE DIFFERENCED DATA

CASE		11	12	13	14
INDEPENDENT VARIABLES (LAG)	1	X		X	X
	2	X	X	X	X
	3	X	X		
	4	X	X	X	X
ERROR		.06	.06	.06	.093
CHI SQ 1		205.7	211.4	207.6	86.5
CHI SQ 2		5.65	7.67	6.02	5.9
CHI SQ TOT		211.40	218.98	213.58	92.4
SIGNIFICANCE		6.89	7.18	6.92	-0.15
COEFFICIENT	1	.00355	.00333	.0034	.0028
	2	.1343	-	.1214	.1180
	3	.252	.297	.238	.2405
	4	-.073	-.047	-	-
	5	-.535	-.569	-.559	-.5465

FIGURE 4

DEVIATION FROM EXPECTED VALUE OF CHI-SQUARE TOTAL vs ESTIMATED MEASUREMENT ERROR



IV. CONCLUSION. Comparison of the least chi-square (LCS) analysis of three well-known time series with the usual least squares analysis indicates that in every case a more significant set of parameters is obtained by the LCS analysis. We think that this procedure should be studied in more detail to further validate this conclusion.

REFERENCES

1. Box, G. E. P., and Jenkins, G. M. Time Series Analysis (Rev Ed) Holden-Day, San Francisco, CA, 1976.

2. Roberts, H. V. Conversational Statistics The Sciéntific Press, Palo-Alto, CA 94304.

3. Moore, R. L. Non-Linear Lease Chi-Square Algorithm, An Improvement on Non-Linear Least Squares, pp. 37-58, Proc. 2nd Meeting of the Coordinating Group on Modern Control Theory. Part I, 10-11 December 1980, AAMSA, APG, MD 21005.

4. Moore, R. L. 23rd Proc. Army Conference on Design of Experiments, pp. 53-68, ARO Report 78-2.

AN APPLICATION OF RENEWAL THEORY TO SOFTWARE RELIABILITY

Leonard A. Stefanski
U.S. Army Research Office

ABSTRACT

In this paper we introduce a probabilistic model for describing software failures which generalizes several models appearing in the literature. Associated with this model is a *Superimposed Delayed Transient Renewal Process*, (SDTRP). The structure of such processes is exploited to obtain several quantitative measures of software performance. In addition we are able to consolidate the literature on software reliability by pointing out that several models are special cases of SDTRP's. Finally, some results relevant to inference and goodness-of-fit tests for a restricted class of models are presented.

0. INTRODUCTION

In the past decade several studies of the stochastic behavior of computer software failures have been undertaken with the objective of developing analytical models to be used to obtain quantitative measures of software performance. For the most part the development and analysis of these models has proceeded along highly individualistic lines. This has led to some duplication of effort and, more importantly, has precluded a coherent study of software reliability modeling. This paper attempts to remedy this situation by codifying a small but important segment of the literature.

In Section 1 we describe a scenario for the generation of software failures which leads us to consider superimposed delayed transient renewal processes as models of software reliability. Several measures of software performance are then obtained.

In Section 2 we show how this model generalizes the works of Jelinski and Moranda (1972), Littlewood (1981), and Goel and Okumoto (1978). We also point out an interesting relationship between the above models and that of Goel and Okumoto (1979).

Finally, Section 3 contains results relevant to inference and goodness-of-fit tests for the class of order statistic models. These are statistical questions which have not been adequately addressed in the past. For Jelinski and Moranda's model the use of maximum likelihood to estimate parameters often yields useless results. The problem is shown to be equivalent to estimating population size when observations are obtained by truncated sampling. Thus the work of Blumenthal and Marcus (1975) can be brought to bear on this subject. In addition a procedure for comparing the fit of these models to existing data sets is outlined. This is illustrated with data from a software development project.

1. MODEL DEVELOPMENT

A software program may contain errors yet still be capable of successful execution for certain types of inputs. A failure occurs when an input is processed which causes one of these errors to manifest itself. Since debugging takes place we expect the number of errors to decrease and program reliability to increase. With this in mind we offer the following interpretation of the software failure process.

Let N represent the initial number of errors in the software and for ease of reference call these $E_1, E_2, \ldots, E_N$. Furthermore, assume that the detection of E_i is independent of that of E_j for $i \neq j$. Let us focus attention on just one of the E_j's, say E and let M(t) count the number of software failures experienced in [0,t] *due to* E. X_1 is the time until the first of these failures and has distribution $F(x) = \Pr(X_1 \le x)$. If errors, when detected, are corrected with probability one, then M(t) is a very simple point process, assuming only the values 0 and 1,

$$M(t) = \begin{cases} 0 & 0 \le t < X_1 \\ 1 & t \ge X_1 \end{cases}$$

M(t) is a Bernoulli random variable with $\Pr[M(t) = 1] = F(t)$. However, if debugging is not successful then we will experience a second failure after some time X_2. A reasonable assumption is that X_1 and X_2 are independent and identically distributed. Similarly, if the first k attempts at correcting E are unsuccessful, we would reason that the inter-failure times $X_1, X_2, \ldots, X_{k+1}$ are i.i.d. with common distribution F. If, however, errors are corrected with probability p, $0 < p < 1$, there is positive probability that the number of software failures caused by E terminates at some finite value $n = 1,2,\ldots$. In particular, if $Y_1, Y_2, \ldots$ are the inter-failure times of the M(t) process *under imperfect debugging* then

$$Y_1 = X_1$$

$$Y_2 = \begin{cases} X_2 & \text{with probability } q \\ +\infty & \text{with probability } p \end{cases}$$

$$\vdots$$

$$Y_j = \begin{cases} X_j & \text{with probability } q \\ \infty & \text{with probability } p, \end{cases}$$

where $q = 1-p$. Thus Y_1 is a proper random variable having distribution F and for $j \geq 2$ Y_j is a *defective* random variable with

$$\Pr[Y_j \leq t] = qF(t) \quad 0 \leq t < \infty$$

$$\Pr[Y_j = +\infty] = P \quad .$$

The M(t) process is a delayed transient renewal process.

Let $S_j = Y_1 + \ldots + Y_j$ be the time of the j-th failure due to E. Then for any finite $t \geq 0$ we have

$$\begin{aligned} \Pr(S_j \leq t) &= \Pr(Y_1 + \ldots + Y_j \leq t) \\ &= \Pr(Y_1 + \ldots + Y_j \leq t, \text{ and } Y_i = X_i \quad i=1,\ldots,j) \\ &= \Pr(X_1 + \ldots + X_j \leq t) \prod_{i=2}^{j} P(X_i = Y_i) \\ &= q^{j-1} F_{(j)}(t) \end{aligned} \tag{1.1}$$

where $F_{(j)}(t)$ is the j-fold convolution of F with itself. In deriving (1.1) it has been tacitly assumed that the debugging attempts are independent of one another and of the occurrence of failures. Since $[M(t) \geq k] \equiv [S_k \leq t]$ we find that

$$\Pr(M(t) \geq k) = q^{k-1} F_{(k)}(t)$$

and therefore the renewal function $H_q(t)$ is given by

$$H_q(t) = E[M(t)] = \sum_{k=1}^{\infty} \Pr[M(t) \geq k] = \sum_{k=1}^{\infty} q^{k-1} F_{(k)}(t). \tag{1.2}$$

One then obtains the following relationship

$$H_q(t) = F(t) + q(F*H_q)(t) \tag{1.3}$$

where * denotes Lebesgue-Stieltjes convolution. Letting

$$h_q^{\circ}(s) = \int_0^{\infty} e^{-st} dH_q(t)$$

and

$$f^{\circ}(s) = \int_0^{\infty} e^{-st} dF(t) \qquad \text{we have}$$

the relationship

$$h_q^{\circ}(s) = \frac{f^{\circ}(s)}{1-qf^{\circ}(s)} \tag{1.4}$$

which can be used to determine $H_q(t)$ given any distribution F. The probability generating function of M(t) is defined as $\phi_{M(t)}(s) = E\{\exp[M(t)\log s]\}$ and it can be established that

$$\phi_{M(t)}(s) = 1 + (s-1)\, H_{qs}(t), \qquad 0 < s < 1.$$

Now let $J(y) = H_y(t)$ for $0 < y < 1$ then we have

$$\phi_{M(t)}(s) = 1 + (s-1)\, J(qs) \qquad \text{and}$$

$$\frac{d^k}{ds^k}[\phi_{M(t)}(s)] = q^k(s-1)\, J^{(k)}(qs) + kq^{k-1}\, J^{(k-1)}(qs).$$

Thus the k-th factorial moment of M(t) is given by

$$E\{M(t)[M(t)-1]\cdot\ldots\cdot[M(t)-k+1]\} = \frac{d^k}{ds^k}\phi_{M(t)}(s)\Big|_{s=1} = kq^{k-1}\, J^{(k-1)}(q),$$

and the variance of M(t) is

$$\mathrm{Var}[M(t)] = 2q\, J'(q) + J(q) - [J(q)]^2$$

Finally note that $M(\infty)$ is a negative binomial random variable with

$\Pr\{M(\infty) = k\} = pq^{k-1} \qquad k=1,2,\ldots$.

A random variable of considerable interest is W, the time until error E is corrected. Clearly, $W = X_1+\ldots+X_k$ with probability pq^{k-1} and hence

$$P[W\le t] = \sum_{n=1}^{\infty} pq^{n-1}F_{(n)}(t) = pH_q(t).$$

This is a proper distribution function. We can find the mean and variance of W by noting that

$$E(W) = \sum_{n=1}^{\infty} pq^{n-1} \int_0^{\infty} tdF_{(n)}(t) = \sum_{n=1}^{\infty} pq^{n-1} n\mu_1 = \frac{\mu_1}{p}$$

and

$$\begin{aligned}
\mathrm{Var}(W) &= E(W-\frac{\mu_1}{p})^2 \\
&= \sum_{n=1}^{\infty} pq^{n-1} \int_0^{\infty} [t-\frac{\mu_1}{p}]^2 \, dF_{(n)}(t) \\
&= \sum_{n-1}^{\infty} pq^{n-1} [n\sigma^2+\mu_1^2(n-\frac{1}{p})^2] \\
&= \frac{1}{p}\{\sigma^2+\mu_1^2(\frac{1}{p}-1)\}
\end{aligned}$$

where μ_1 and σ^2 are the mean and variance of X_1.

A natural assumption is that each of the N errors gives rise to a SDTRP, not necessarily stochastically identical. In particular we assume that error E_i, $i=1,\ldots,N$ has associated with it the following quantities:

a) F^i = Distribution of the time until first failure due to E_i.

b) P_i = Probability of correcting E_i on any trial, $q_i = 1-P_i$.

c) $M_i(t)$ = Number of failures in [0,t] due to E_i. $\Pr\{M_i(t)\ge k\} = (q_i)^{k-1}F^i_{(k)}(t)$

d) $H^i_{q_i}(t) = E[M_i(t)] = \sum_{k=1}^{\infty} (q_i)^{k-1}F^i_{(k)}(t)$

e) $J_i(y) = H^i_y(t)$.

f) $\phi_i(s) = \phi_{M_i(t)}(s)$ = Probability generating function of $M_i(t)$.

g) W_i = Time until E_i is corrected. $\Pr\{W_i\le t\} = P_iH^i_{q_i}(t)$.

Now let $M^*(t)$ count the number of failures *of all types* experienced in $[0,t]$. Then $M^*(t) = M_1(t)+\ldots+M_N(t)$ and hence $H(t) = E[M^*(t)] = H^1_{q_1}(t)+\ldots+H^N_{q_N}(t)$. Also, since we have assumed the detection of E_i to be independent of that of E_j, $i \neq j$, we also have $\mathrm{Var}[M^*(t)] = \sum_{i=1}^{N} \mathrm{Var}[M_i(t)] = \sum_{i=1}^{N}\{2q_i J_i'(q_i) + J_i(q_i) - J_i^2(q_i)\}$ and $E\{\exp(M^*(t)\log s)\} = \phi_{M(t)}(s) = \prod_{i=1}^{N} \phi_i(s)$.

1.1 Measures of Software Performance

We are now in a position to derive the distributions of 1) T, the time to a completely debugged system, 2) T_n, the time to a specified number (n) of remaining errors, and 3) X(t) the number of errors in the program at time t. We will make use of the ordered random variables $W_{(1)} < W_{(2)} < \ldots < W_{(N)}$ where $W_{(i)}$ is the i-th order statistic from $(W_1, W_2, \ldots, W_N)$.

The software is completely debugged if and only if each of the N errors has been corrected. Thus $\{T \le t\}$ iff $\{W_i \le t\}$ $i=1,\ldots N$. And we find

$$
\begin{aligned}
\Pr\{T \le t\} &= \Pr\{\max\{W_1,\ldots,W_N\} \le t\} \\
&= \Pr\{W_{(N)} \le t\} \\
&= \prod_{i=1}^{N} \{p_i H^i_{q_i}(t)\},\ t \ge 0.
\end{aligned}
$$

Economic and/or time considerations may make us willing to tolerate an upper bound n on the number of errors remaining in the program. Thus the distribution of T_n is of interest. The event $\{T_n \le t\}$ occurs iff at least N-n errors have been corrected by time t. Thus we have

$$\Pr\{T_n \le t\} = \Pr\{W_{(N-n)} \le t\}.$$

Proceeding similarly one finds that $\{X(t) \ge n\}$ if and only if $\{W_{(N-n+1)} > t\}$.

Thus

$$\Pr\{X(t) \geq n\} = \Pr\{W_{(N-n+1)} > t\}$$

and

$$\Pr\{X(t) = n\} = \Pr\{W_{(N-n)} \leq t\} - \Pr\{W_{(N-n+1)} \leq t\},$$

also

$$E[X(t)] = \sum_{n=1}^{N} \Pr\{W_{(N-n+1)} > t\} .$$

Remarks: For arbitrary choices of $P_1,\ldots,P_N$ and $F^1,\ldots,F^N$ these formulae are quite cumbersome. However, it is often assumed that $P_1 = P_2 = \ldots = P_N$ and $F^1 = F^2 = \ldots = F^N$ in which case $W_1,\ldots,W_N$ are identically as well as independently distributed. Expressions for the joint distribution of order statistics are well known for the i.i.d. case. In fact if the common distribution function is $F(x) = 1 - \exp(-\lambda x)$ then the W_i's have distribution $\Pr\{W_i \leq t\} = 1 - \exp(-\lambda pt)$ and the above expressions simplify considerably.

2. MODEL CONSOLIDATION

Consider the simplification obtained by taking $P_1 = P_2 = \ldots = P_N = P$ and $F^1 = F^2 = \ldots = F^N = F$. If we further restrict $F(x) = 1 - e^{-\lambda x}$ we have the model analyzed by Goel and Okumoto (1978). Their model in turn is a generalization of that of Jelinski and Moranda (1972) to include the possibility of imperfect debugging. We will show that the model proposed by Littlewood (1981) is of the same "type" as the J-M model. To allow for imperfect debugging under Littlewood's framework we need only take $P_1 = P_2 = \ldots = P_N = P$ and $F_1 = F_2 = \ldots = F_N = F$ where $F(x) = 1 - (\frac{\beta}{\beta+x})^{\alpha}$. In addition to simplifying the analysis in studying these models (the primary obstacle is finding $H_q(t)$) the development in terms of SDTRP's provides an alternative interpretation of the failure process.

One of the earliest and certainly most referenced models of software failures is that of Jelinski and Moranda (1972). Our framework includes this under the restrictions $P_1 = P_2 = \ldots = P_N = 1$ and $F^1 = F^2 = \ldots = F^N = F = 1 - e^{-\lambda x}$. Recall that for the case of perfect debugging, ($P = 1$) each error gives rise to a Bernoulli point process, $M_i(t) = 0$ if E_i has not been corrected by time t, $M_i(t) = 1$ otherwise. Thus if the M_i's(t) are i.i.d. then $M^*(t) = M_1(t)+\ldots+M_N(t)$ has a binomial distribution;

$$\Pr\{M^*(t) = k\} = \binom{N}{k}[F(t)]^k[1 - F(t)]^{N-k}, \; k=1,\ldots,N.$$

$M^*(t)$ is called an Order Statistic process since the time of occurrence of the j-th event is distributed as the j-th order statistic from a population of size N with common distribution function F. Jelinski and Moranda originally specified the joint distribution of $Y_1,\ldots,Y_n$, $n \le N$ where Y_i is the time between discovery of the (i-1)-st and i-th error;

$$f(y_1\ldots y_n) = \prod_{i=1}^{n} (N-i+1)\lambda e^{-\lambda(N-i+1)Y_i}. \tag{2.1}$$

An easy computation establishes that the random variables $X_{(i)} = Y_1+\ldots+Y_i$ $i=1,\ldots,n$ are indeed the order statistics from an exponential population. This observation, until now overlooked, has important implications when it is desired to estimate the unknown parameters N and λ. This problem is addressed in Section 3.

We stated earlier that the model of Littlewood (1981) is the same "type" as that of Jelinski and Moranda. In fact, Littlewood has characterized an order statistic process generated by a Pareto population. That is, $X_{(j)}$, the time of occurrence of the j-th failure, is distributed as the j-th order statistic from a population of size N with common distribution $F(x) = 1 - \left(\frac{\beta}{\beta+x}\right)^\alpha$.

Order statistic models can be motivated via the following argument. Assume there are N errors, $E_1,\ldots,E_N$. Let X_i be the debugging time needed to reveal E_i. Then the time of discovery of the *first* error is just $X_{(1)}$ the first order statistic from $X_1,\ldots,X_N$. Similarly for $X_{(j)}$ $j=1,\ldots,N$.

In a later paper Goel and Okumoto (1979) suggest modeling software error detections via a non-homogeneous Poisson process. They choose a mean function of the form $\mu(t) = a(1-e^{-bt})$. This model is closely related to the order statistic processes as the following theorem demonstrates.

<u>Theorem 1</u>: If for a Poisson process it is given that exactly N events have occurred in $[0,t_0]$, then the event times are distributed as order statistics from the distribution $F(x) = \frac{\mu(x)}{\mu(t_0)}$, $0 \le x \le t_0$.

<u>Proof</u>: See Thompson (1981) for a proof of this theorem and related results.

To apply this result in the present context take $t_0 = \infty$, (i.e., $t_0 \to \infty$). Then conditioned on the event that exhaustive debugging uncovers a total of N errors, the times of error discoveries, $X_{(1)} < X_{(2)} <\ldots< X_{(N)}$, are order statistics from the distribution $F(x) = \frac{\mu(x)}{\mu(\infty)} = 1 - e^{-bx}$. The resulting failure time structure is identical to that of the Jelinski-Moranda model. It is also possible to assign a Poisson prior on the number of errors, N, in an order statistic model and obtain a Poisson process with mean function $\mu(t) = aF(t)$, where a is the prior mean. This was pointed out by Langberg and Singpurwalla (1981). A summary of the relationships between the models discussed thus far appears in Figure 1.

We end this section by pointing out that a subclass of the order statistic processes is well adapted to modeling software failures (or more generally problems in which reliability growth is occurring). Littlewood has suggested that any software reliability model should possess certain features among which are the stochastic ordering and decreasing failure rate (DFR) property of the random variables $Y_1, Y_2, \ldots$ where Y_i is the time between the (i-1)-st and i-th failure. Stochastic ordering will be designated by $\underset{ST}{\le}$, $\underset{ST}{\ge}$. In an order statistic model $Y_1, Y_2, \ldots$ are the ordered spacings from the distribution F. If F is DFR, the Y_i's do possess the desired properties as the following result shows.

Theorem 2: Let $Y_1,\ldots,Y_n$ be the interarrival times of events in an order statistic process. If the distribution F is DFR then

a) $Y_1 \underset{ST}{\leq} Y_2 \underset{ST}{\leq} \cdots \underset{ST}{\leq} Y_N$

b) $Y_1,\ldots Y_N$ are associated

c) Y_i is DFR $\quad i=1,\ldots,N$.

Proof: Part (a) follows directly from Theorem 6.1 in Barlow *et al.* (1972). Let h(t) be the hazard rate of the distribution F(t), that is $F(t) = 1 - \exp\{-\int_0^t h(x)dx\}$, then $h(t) \downarrow t$ by assumption. An easy calculation shows the conditional hazard rate of $Y_k|(Y_1,Y_2,\ldots,Y_{k-1})$ to be $(N-k+1)\ h(t+\sum_1^{k-1} Y_i)$. Since this decreases in Y_i, $i=1,\ldots,k-1$ it follows that $Y_1,\ldots,Y_N$ are *conditionally increasing in sequence* which in turn implies the association of $Y_1,\ldots,Y_N$. Y_i is DFR since its distribution function is a mixture of DFR distributions, Barlow and Proschan (1975).

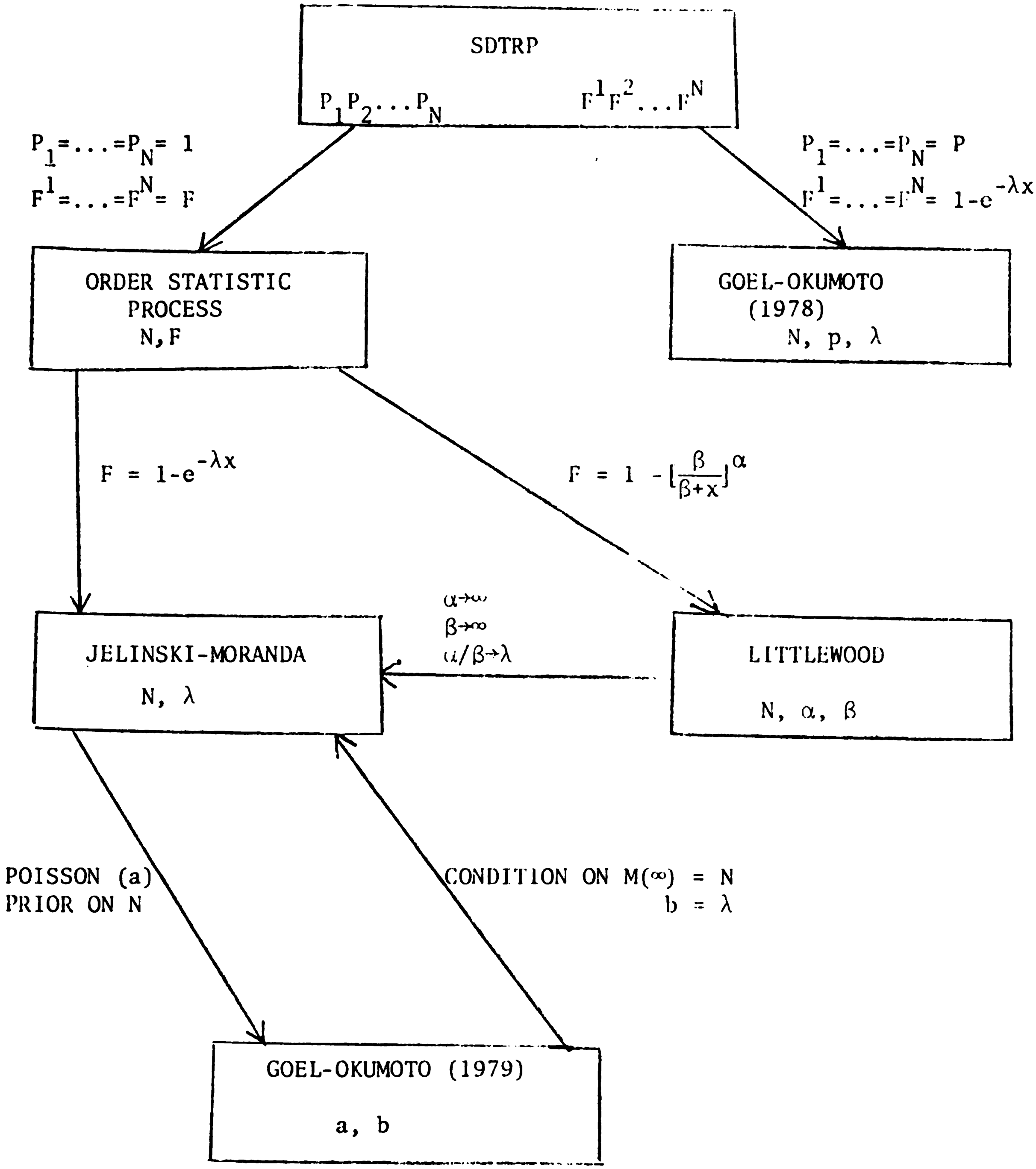

SDTRP
$P_1 P_2 \ldots P_N$
$F^1 F^2 \ldots F^N$
$P_1 = \ldots = P_N = 1$
$F^1 = \ldots = F^N = F$
$P_1 = \ldots = P_N = P$
$F^1 = \ldots = F^N = 1-e^{-\lambda x}$
ORDER STATISTIC PROCESS
N,F
GOEL-OKUMOTO (1978)
N, p, λ
$F = 1-e^{-\lambda x}$
$F = 1 - [\frac{\beta}{\beta+x}]^{\alpha}$
$\alpha \to \infty$
$\beta \to \infty$
$\alpha/\beta \to \lambda$
JELINSKI-MORANDA
N, λ
LITTLEWOOD
N, α, β
POISSON (a)
PRIOR ON N
CONDITION ON $M(\infty) = N$
$b = \lambda$
GOEL-OKUMOTO (1979)
a, b

FIGURE 1

3. INFERENCE AND GOODNESS-OF-FIT

In the application of software reliability models it has been the practice to use information contained in the first n failure times $X_1,\ldots,X_n$, to estimate unknown parameters. For a majority of cases the quality of estimates so obtained has been far from acceptable. In the present section we study this problem for the class of order statistic models, and in particular the model of Jelinski and Moranda. The parameter of paramount interest is N, the initial number of errors. For an order statistic model $X_1 < X_2 <\ldots<X_n$ are the first n ordered random variables from a population of size N with distribution function $F(x) = F(x|\underset{\sim}{\theta})$, where $\underset{\sim}{\theta}$ is a vector of parameters, often unknown. Thus, estimating error content is equivalent to estimating population size when observations are obtained by truncated sampling. In what follows we take $F(x) = 1 - \exp(-\lambda x)$ (Jelinski-Moranda model) although many of our comments pertain to the general case as well.

There are several data sets from completed software projects which have been used to examine the validity of the Jelinski-Moranda model. Since the software has been thoroughly debugged N is known as well as the failure times $X_1,\ldots,X_N$. Maximum likelihood estimates of N are calculated from the first n < N failure times and compared to the true value. Typically it has been found that either $\hat{N}$ grossly overestimates N (often $\hat{N} = \infty$) or on the other extreme $\hat{N}$ significantly underestimates N (i.e., $\hat{N}$ = n or n+1). We will show that the former behavior is to be expected. However, the latter is not and suggests the inappropriateness of the model.

For now, consider the case where each of the times X_i, i=1,...,n are observable. If testing is stopped after a fixed time t_0, and in that time n failures are recorded, the likelihood function based on $X_1,\ldots,X_n$ is

$$L(N,\lambda) = \left\{ \prod_1^n (N-i+1)\lambda\exp(-\lambda X_i)\right\} \exp[-\lambda t_0(N-n)] \quad . \tag{3.1}$$

If estimates are desired after exactly n_0 errors have been detected, the likelihood function is obtained from (3.1) by replacing n with n_0 and t_0 with X_{n_0}. Maximum likelihood estimates for both sampling schemes have been studied by Blumenthal and Marcus (1975). The MLE is finite if an only if

$$\sum_{i=1}^{n} X_i < (n+1)t_0 \tag{3.2}$$

for the case of truncated sampling, and

$$\sum_{i=1}^{n_0} X_i < (n_0+1)X_{n_0} \tag{3.3}$$

when sampling stops after n_0 failures. More importantly, their results indicate that for moderately small values of $\delta = 1 - \exp(-\lambda t_0)$ a fairly large positive bias persists even when $\hat{N}$ is finite. Blumenthal and Marcus also consider the conditional MLE and a class of Bayes modal estimates. On the basis of second order asymptotic properties, one of the Bayes estimates is preferred by the authors.

Let us examine more closely the MLE of N when sampling stops after n_0 observations. Set $Y_1 = X_1$ and $Y_i = X_i - X_{i-1}$, $i=2,\ldots,n_0$. Then the inequality in (3.3) is satisfied iff the slope of the regression line of Y_i on i is positive. Furthermore if we let m equal this slope, then, conditional on $\sum_{i=1}^{n_0} Y_i = X_{n_0}$, $\hat{N}$ can be shown to be a decreasing function of m. Thus if the Y_i's exhibit marked reliability growth (m large) $\hat{N}$ tends to be small. Under the assumption of exponential failure times one would expect positive but not large values of m; however, sampling from a (strictly) DFR population would account for this.

The data in Table 1 is from J. Musa's "Software Reliability Data", available from DACS, Rome Air Development Center, NY. It contains the ordered times between failures $Y_1,\ldots,Y_{38}$, N = 38. Table 2 shows the value of $\hat{N}$ computed from sample sizes $n_0 = 5,10,\ldots,35,38$. As can be seen, $\hat{N}$ significantly underestimates N. This phenomenon has been observed in other data sets as well (Forman and Singpurwalla, 1977). Our analysis would suggest that the assumption of exponentiallity (Jelinski-Moranda model) is not warranted for this data. To verify this conjecture a test of the hypothesis H_0: F is exponential against H_1: F is strictly DFR was performed using the *cumulative total time on test* statistic of Barlow *et al.* (1972). V_k is defined by

$$V_k = \frac{\sum_{j=1}^{k} (k-j)(k-j+1)Y_j}{\sum_{j=1}^{k} (k-j+1)Y_j} .$$

For the data in Table 1 $V_{38} = 13.43$. Using a normal approximation to the distribution of V_k one finds that

$$\Pr\{V_{38} \le 13.43\} \cong .002 .$$

Since small values of V_k favor DFR populations H_0 is rejected. For this data the Jelinski-Moranda model is inappropriate.

This type of posterior analysis should be useful in determining the relative merit of various order statistic models. In asking whether the model of Jelinski and Moranda provides a better fit than does Littlewood's we are in fact questioning whether the variability in $X_1,\ldots,X_N$ is better explained by an exponential distribution than a Pareto distribution. Procedures for answering such questions are well-known.

We make one final comment pertaining to the analysis of grouped data. Quite often the only information available is of the form r_i errors detected in the interval $(X_{i-1},X_i]$ $i=1,\ldots,k$. In the past it was thought necessary to assume the r_i failure times uniformly distributed over the interval in order to obtain estimates from the Jelinski-Moranda model. With the knowledge that we are observing an order statistic process it becomes evident that the number of failures in disjoint intervals follows a multinomial distribution. Thus we have the likelihood,

$$L(N,\lambda) = \frac{N!}{(N-r)!} \left[\prod_{i=1}^{k} \frac{[\exp(-\lambda X_{i-1}) - \exp(-\lambda X_i)]^{r_i}}{r_i!} \right] \exp[-\lambda(N-r)X_k] \qquad (3.4)$$

where r_i is the number of failures in $(x_{i-1}, x_i]$ $i=1,\ldots,k$, $0 = X_0<\ldots<X_k$, and $r = r_1 +\ldots+r_k$. Sanathanan (1972) considered estimating the size of multinomial populations when cell probabilities are differentiable functions of an unknown parameter. The interested reader should consult this paper for details.

4. SUMMARY AND CONCLUSIONS

The problem of modeling software failures is a challenging problem which to date has no satisfactory solution. In an attempt to organize the literature we have shown that several well-known models are special cases of superimposed delayed transient renewal processes. In particular the models of Jelinski and Moranda (1972) and Littlewood (1981) were shown to be members of the class of order statistic processes, which in turn is related, via conditioning arguments, to a subclass of the Poisson processes, namely those with bounded mean functions. More importantly, the estimation problems encountered in using these models were put in their proper perspective. The relevance of the Blumenthal and Marcus (1975) results had previously escaped the attention of workers in this field.

There is currently a considerable amount of interest in determining which model works best. We have pointed out a simple procedure for verifying model assumptions against existing data sets and comparing the fit of certain models. Unfortunately, having the correct model does not guarantee one of obtaining reliable maximum likelihood estimates of error content. Since this is the parameter of paramount interest, it is necessary to examine alternative forms of estimation. Langberg and Singpurwalla (1981) have addressed this problem with a Bayesian approach which also allows them to unify some of the literature on software reliability models.

TABLE 1

Times between failures $(Y_1,...,Y_{38})$

115	3	83	178	194	136	1077	15	15	92	50	71	606
1189	40	788	222	72	615	589	15	390	1863	1337	4508	834
3400	6	4561	3186	10571	563	2770	652	5593	11696	6724	2546	

TABLE 2

MLE of N = 38

Sample size (n_0)	5	10	15	20	25	30	35	38
$\hat{N}$	6	19	18	25	25	30	35	38

REFERENCES

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972) Statistical Inference Under Order Restrictions, New York: John Wiley & Sons, 265-266.

Barlow, R. E. and Proschan, F. (1975) Statistical Theory of Reliability and Left Testing, New York: Holt, Rinehart and Winston.

Blumenthal, S. and Marcus, R. (1975) Estimating population size with exponential failure, *Journal of the American Statistical Association, 70,* 913-922.

Forman, E. H. and Singpurwalla, N. D. (1977) An empirical stopping rule for debugging and testing computer software, *Journal of the American Statistical Association, 72,* 750-757.

Goel, A. L. and Okumoto, K. (1978) An imperfect debugging model for reliability and other quantitative measures of software systems, Technical Report 78-1, Department of Industrial Engineering and Operations Research, Syracuse University, Syracuse, NY 13210

Goel, A. L. and Okumoto, K. (1979) Time-dependent error-detection rate model for software reliability and other performance measures, *IEEE Transactions on Reliability, R-28,* 206-211.

Jelinski, Z. and Moranda, P. B. (1972) Software reliability research, in Statisical Computer Performance Evaluation, ed. W. Freiberger, New York: Academic Press, 485-502.

Langberg, N. and Singpurwalla, N. D. (1981) A unification of some software reliability models via the Bayesian approach, Technical Memorandum, Serial TM-66571, The George Washington University, School of Engineering and Applied Sciences.

Littlewood, B. (1981) Stochastic realiability growth: a model for fault removal in computer programs and hardware designs, *IEEE Transactions on Reliability,* to appear.

Sanathanan, L. P. (1972) Estimating the size of a multinomial population, *Annals of Mathematical Statistics, 43,* 142-152.

Thompson, W. A., Jr. (1981) On the foundations of reliability, *Technometrics, 23,* 1-13.

METHODOLOGY FOR ESTIMATING MISSION AVAILABILITY AND RELIABILITY FOR A MULTIMODAL SYSTEM

Henry P. Betz
U.S. Army Materiel Systems Analysis Agency
Aberdeen Proving Ground, Maryland

1. INTRODUCTION

Developmental Testing constraints sometimes require that a system be tested according to a profile that is different from the mission profile for which the system's reliability requirements were specified. For example, a surface to air missile system whose capabilities include movement, surveillance, and target engagement might, because of accelerated testing requirements, be tested extensively in the target engagement mode (in order to assure that all engagement performance requirements are met) and only minimally in the movement and surveillance modes (Figure 1). However, in the tactical mission profile, surveillance functions might encompass the majority of the mission (Figure 2). It would be incorrect to compare the system availability and mean time between failure (MTBF) demonstrated in the test scenario to the requirements specified for the tactical scenario. This is due to the fact that the engagement mode of operation is more complex and therefore, many more failures associated with it would be expected. Although the rate of failure detections experienced in the engagement mode of the test scenario would remain the same as in the tactical one, the amount of time spent in the engagement mode of the tactical scenario is much less than the test scenario which means that a smaller number of engagement mode failures should be expected on a per mission basis. In this situation, it can be seen that evaluating the system MTBF based on the test scenario would understate the MTBF value. In order to determine if the system meets its reliability specifications, the reliability of the system in the tactical mission must be evaluated from data collected in a test scenario which is entirely different.

This report will develop a methodology that can be used to evaluate a system which is operated in a series of n modes with the i^{th} mode being defined as having a certain number of subsystems operating in it and mode $i + 1$ consists of mode i subsystems plus additional subsystems operating. That is, subsystems operating in mode i are nested in mode $i + 1$ (Figure 3).

In addition, the corrective maintenance time and logistics delay time that will be seen in the field are not always known at the time of development testing, either because maintenance procedures are not fully specified at that time or for expediency's sake contractor personnel perform maintenance normally done by the soldier. This report allows for the insertion of maintainability parameters derived from other sources i.e. maintainability demonstrations, logistics simulations, etc.

ACKNOWLEDGEMENT

The author wishes to recognize Dr. Alan W. Benton and Mr. Robert E. Mioduski for their contributions to this report.

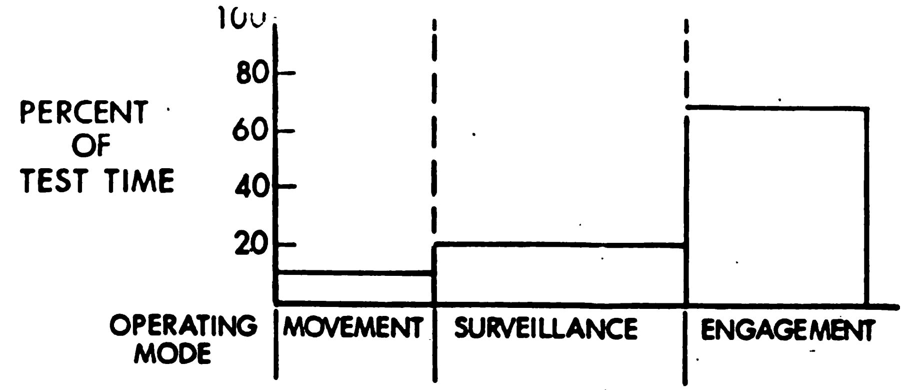


Figure 1. Possible Test Scenario.

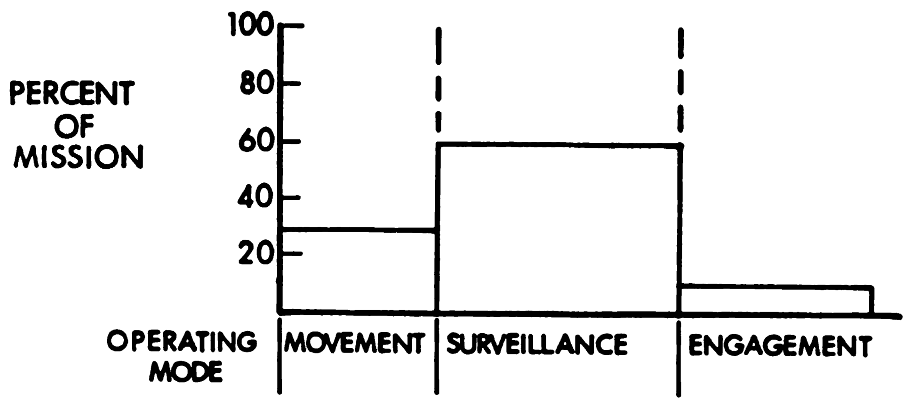


Figure 2. Tactical Mission.

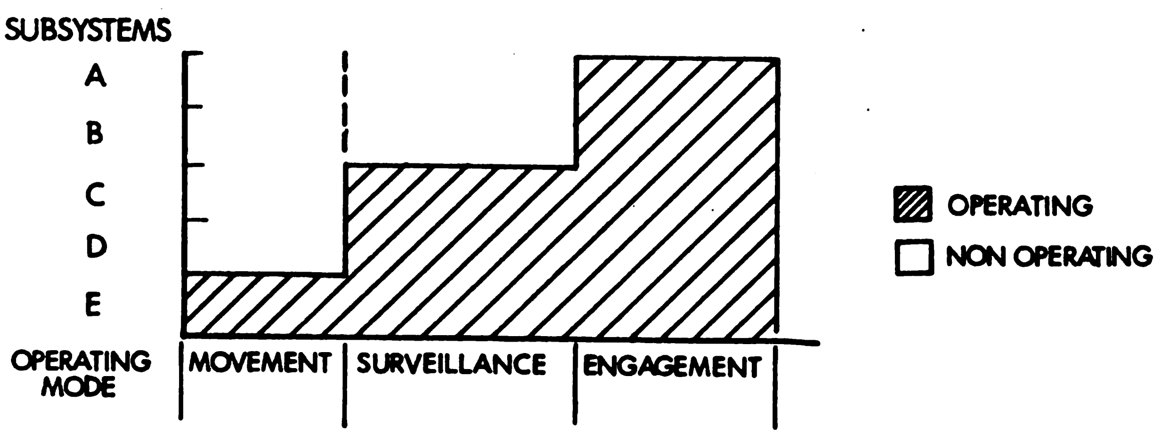


Figure 3. System Operation.

2. DERIVATION OF EQUATIONS

The first step in deriving MTBF and operational availability estimates is to begin with the basic definition of operational availability from which we will derive the mission profile MTBF. It is assumed that estimates of failure rates or MTBFs for each of the operating modes are available. The basic definition of operational availability [1] is given by

$$A_o = \frac{\text{UPTIME}}{\text{UPTIME} + \text{DOWNTIME}} \tag{1}$$

Let us define

k_i = time spent in mode i as specified by desired scenario

$MTBF_i$ = mean time between failure detections of operating mode i (derived from test scenario)

MDT_i = mean downtime including logistics downtime in operating mode i (either specified in tactical scenario or determined by other means - maintainability demonstration, simulation, etc.) If these individual values cannot be determined, use MDT for all values of MDT_i.

MDT = Overall system mean downtime calculated as

$$\frac{\sum\left(\frac{k_i}{MTBF_i + MDT_i}\right) MDT_i}{\sum\left(\frac{k_i}{MTBF_i + MDT_i}\right)} \quad \text{if } MDT_i\text{'s are available.}$$

Otherwise, use an overall MDT from test, simulation, etc.

Now, total time is uptime plus downtime, or uptime equals total time minus downtime. From the definitions it may be noted that total mission time, T, is given by

$$T = \sum_i k_i$$

In determining downtime, it may be noted that

$$\frac{k_i}{MTBF_i + MDT_i}$$

is the expected number of failures in operating mode i and multiplying this by the expected downtime for mode i, MDT_i, gives the expected downtime in mode i. Therefore,

$$\text{UPTIME} = \sum_i \left[k_i - \left(\frac{k_i}{\text{MTBF}_i + \text{MDT}_i} \right) \text{MDT}_i \right] \tag{2}$$

and thus, it follows that

$$A_o = \frac{\sum_i \left[k_i - \left(\frac{k_i}{\text{MTBF}_i + \text{MDT}_i} \right) \text{MDT}_i \right]}{T} \tag{3}$$

It may be shown that A_o is a weighted average of the mode availabilities. On rearranging (3), we have

$$A_o = \frac{\sum_i \frac{k_i \left(\text{MTBF}_i + \text{MDT}_i \right) - k_i \ \text{MDT}_i}{\text{MTBF}_i + \text{MDT}_i}}{T}$$

$$= \frac{1}{T} \sum_i \frac{k_i \text{MTBF}_i}{\text{MTBF}_i + \text{MDT}_i}$$

$$= \sum_i \left(\frac{k_i}{T} \right) \left(\frac{\text{MTBF}_i}{\text{MTBF}_i + \text{MDT}_i} \right)$$

$$= \sum_i W_i \ A_{oi}$$

Now, we may also view A_o as

$$A_o = \frac{\text{MTBF}_{sys}}{\text{MTBF}_{sys} + \text{MDT}} \tag{4}$$

where MTBF_{sys} is the system MTBF. Equating (3) and (4) we have

$$\frac{\text{MTBF}_{sys}}{\text{MTBF}_{sys} + \text{MDT}} = \frac{\sum_i \left[k_i - \frac{k_i}{\text{MTBF}_i + \text{MDT}_i} \ \text{MDT}_i \right]}{T}$$

Rearranging and solving for system MTBF yields

$$MTBF_{sys} = \frac{T}{\sum_i \frac{k_i}{MTBF_i + MDT_i}} - MDT \qquad (5)$$

3. EXAMPLE OF APPLICATION

Consider a surface to air missile system which is characterized by three modes of operation - travel, surveillance and target engagement. A system MTBF requirement of 100 hours and an operational availability requirement of 0.90 have been set. The typical 24 hour scenario for which the requirements were set is as follows:

	Time (hours)
Travel	1
Surveillance	21
Engagement	2

During the test program the following failure detection rates were observed:

	$MTBF_i$
Travel	1000
Surveillance	500
Engagement	50

Overall mean time to repair was determined to be 6 hours and mean logistics delay time was found by a logistics simulation to be 14 hours. No other information is available.

The question is then, has the system demonstrated requirements? Using equation (5) we have

$$MTBF_{sys} = \frac{T}{\sum \frac{k_i}{MTBF_i + MDT_i}} - MDT$$

Therefore $MTBF_{sys} = \dfrac{24}{\dfrac{1}{1000 + 20} + \dfrac{21}{500 + 20} + \dfrac{2}{50 + 20}} - 20$

$MTBF_{sys} = 323$ hrs.

Using Equation 3, we have:

$$A_o = \frac{\sum_i \left[k_i - \left(\dfrac{k_i}{MTBF_i + MDT_i} \right) MDT_i \right]}{\text{TOTAL TIME IN MISSION}}$$

$$A_o = \frac{\left[1 - \left(\dfrac{1}{1000 + 20}\right) 20\right] + \left[21 - \left(\dfrac{21}{500 + 20}\right) 20\right] + \left[2 - \left(\dfrac{2}{50 + 20}\right) 20\right]}{24}$$

$A_o = .94$

Therefore, the system requirements have been demonstrated.

REFERENCE

1. AMCP 706-132, ENGINEERING DESIGN HANDBOOK, Maintenance Engineering Techniques, pp 4-12.

RELIABILITY PREDICTIONS FOR BLACK HAWK PRODUCTION AIRCRAFT

Clarke J. Fox
U.S. Army Materiel Systems Analysis Activity
Aberdeen Proving Ground, Maryland

ABSTRACT

In the final stage of the BLACK HAWK helicopter prototype development, a large number of delayed fixes were proposed for the production aircraft. The BLACK HAWK Project Manager requested that AMSAA evaluate the impact of these delayed fixes on the reliability of the production aircraft. A methodology was developed to predict the reliability of the production aircraft based on estimates of the effectiveness of the delayed fixes and estimates of the rate of occurrence of new failure modes.

1. INTRODUCTION

In December 1976, Sikorsky Aircraft was awarded a production contract for the BLACK HAWK utility helicopter (UH-60A). However, further development work was needed on the prototype aircraft prior to production deliveries. This period of advanced prototype development was called the Maturity Phase. During this period, a significant degradation in the aircraft system reliability was observed. This degradation was attributed primarily to a more severe contractor test environment and wear-out of some aircraft components. An additional aggravating factor was the deferral of all reliability improvements to the production phase. Consequently, no fixes or engineering modifications would be tested on the prototype aircraft. These circumstances prevented the BLACK HAWK Project Manager from determining the progress of the aircraft toward meeting the established system reliability goal of 4.0 hours MTBF. In order to determine the likelihood of meeting this goal, the Project Manager requested that the Army Materiel Systems Analysis Activity (AMSAA) conduct an evaluation of the effectiveness of the delayed fixes and predict the system MTBF of the production aircraft. This evaluation would allow the Project Manager to make timely managerial and technical decisions in order to correct problem areas prior to production.

2. METHODOLOGY

2.1 Basis for Prediction. The prediction of the reliability of the production BLACK HAWK was based on analysis of existing Maturity Phase failure modes in light of proposed engineering fixes, and the merging of this analysis with an estimate of the rate of occurrence of new failure modes. Existing Maturity Phase failure modes were also evaluated to identify any modes which exhibited wear-out characteristics or which were discovered during a modification or special test procedure peculiar to the contractor test environment. These type of failure modes were considered to be "unlikely occurrences" on new production aircraft.

2.2 <u>Analysis of Existing Failure Modes</u>. The analysis of existing failure modes was based on failures occurring during 457 flight hours of Maturity Phase testing. A total of 273 failures were charged to Sikorsky during this 457 flight hour period. An Army team of engineering personnel was organized by AMSAA to estimate the effectiveness of the contractor fixes for these failures. The effectiveness of a modification on any given failure mode was evaluated by means of the average effectiveness factor (k factor) assigned to that mode by the Army team. For example, a k factor of .60 assigned to a fix for a particular failure mode over a certain time interval would indicate that after the fix is incorporated, 60% of the number of failures of that particular mode over the same time interval would not be expected to occur; that is, only 40% of the number of failures would be expected to occur over the same time interval if the fix were incorporated.

Thus, if N represents the number of failures of a particular mode occurring over a certain time period, and E(N) represents the expected number of failures over the same time period after the fix with effectiveness factor k is incorporated, then

$$E(N) = N(1-k).$$

Fixes were proposed by Sikorsky for 243 of the 273 failures during the Maturity Phase. Table 1 provides a breakdown of the 243 failures by major subsystems and the expected number of failures in each subsystem after applying to each failure in the subsystem its respective k factor as determined by the Army team. After applying the k factors to the 243 failures for which fixes were proposed, the expected number of failures based on 457 flight hours was reduced from 243 to 91.85.

TABLE 1. FAILURES FOR WHICH FIXES WERE PROPOSED

SUBSYSTEM	NUMBER OF FAILURES	EXPECTED NUMBER OF FAILURES*	EXPECTED FAILAURE RATE (BASED ON 457 HOURS)
Rotor	42	11.61	0.025
Transmission	6	2.08	0.005
Propulsion (GFE)	13	5.55	0.012
Propulsion (CFE)	25	8.75	0.019
Electrical	27	12.11	0.026
Avionics	33	20.64	0.045
Airframe	32	11.65	0.025
Electronic Controls	15	2.56	0.006
Hydraulic/ Flight Controls	50	16.90	0.037
TOTAL	243	91.85	0.200

*Obtained by applying to each failure the average k factor assigned by the Government Evaluation Team

It was anticipated that prior to production deliveries, fixes would be installed for the remaining 30 failures out of the total of 273 failures. The effectiveness of these fixes was estimated by categorizing the 30 failures by subsystem, and applying the average subsystem k factor derived from the total 243 failures reviewed to the respective number of failures in each subsystem. Table 2 provides the breakdown of the 30 failures by subsystem and the expected number of failures after application of the k factors discussed above. The expected number of failures was thus reduced from 30 to 11.6.

The 273 failures which were charged to Sikorsky during the 457 flight hour period were evaluated to determine those modes which exhibited wear-out characteristics or which were discovered during a modification, special inspection, or experimental procedure peculiar to the contractor testing. These modes were not considered likely to occur on production aircraft. The Army team identified 46 unlikely occurrences among the 243 failures with fixes. An AMSAA independent analysis identified an additional 16 unlikely occurrences among the 243 failures with fixes and the 30 failures without fixes. The AMSAA analysis thus considered 62 failures as unlikely occurrences among the total 273 failures. Fifty-eight of the 62 unlikely occurrences were discovered among the 243 failures with fixes, yielding a total of 185 failures with fixes and purged of unlikely occurrences. Four of the 62 unlikely occurrences were discovered among the 30 failures without fixes, leaving 26 failures without fixes and purged of unlikely occurrences.

Table 3 provides a breadown by subsystem of the 185 failures with fixes and purged of unlikely occurrences, and the number of these expected to occur after applying the k factors of the Army team. The expected number of failures after the k factors are applied was reduced from 185 to 70.2.

TABLE 2. FAILURES WITHOUT FIXES

SUBSYSTEM	NUMBER OF FAILURES	SUBSYSTEM K FACTOR	EXPECTED NUMBER OF FAILURES*	EXPECTED FAILURE RATE (BASED ON 457 HOURS)
Rotor	2	.723	.55	0.001
Transmission	3	.653	1.04	0.002
Propulsion (GFE)	0	.573	0.00	0.000
Propulsion (CFE)	3	.650	1.05	0.002
Electrical	4	.551	1.80	0.004
Avionics	4	.374	2.50	0.005
Airframe	2	.635	1.27	0.003
Electronic Controls	4	.829	0.68	0.001
Hydraulics/ Flight Controls	8	.662	2.70	0.006
TOTAL	30		11.59	0.024

*Obtained by applying the subsystem average k factor to the failures in that subsystem

TABLE 3. FAILURES FOR WHICH FIXES WERE PROVIDED (PURGED OF UNLIKELY OCCURRENCES)

SUBSYSTEM	NUMBER OF FAILURES	EXPECTED NUMBER OF FAILURES*	EXPECTED FAILURE RATE (BASED ON 457 HOURS)
Rotor	21	5.80	0.013
Transmission	3	0.97	0.002
Propulsion (GFE)	8	3.92	0.009
Propulsion (CFE)	20	6.32	0.014
Electrical	24	11.77	0.026
Avionics	29	18.36	0.040
Airframe	23	6.85	0.015
Electronic Controls	15	2.56	0.006
Hydraulics/ Flight Controls	42	13.63	0.030
TOTAL	185	70.18	0.155

*Obtained by applying to each failure the average k factor assigned by the Government Evaluation Team (Appendix A)

Table 4 provides the breakdown of the 26 failures without fixes and purged of unlikely occurrences and the expected number of failures after application of the k factors discussed above. The expected number of failures was thus reduced from 26 to 10.3.

Table 5 combines the results of Table 3 and Table 4 to give the total expected number of failures (purged of unlikely occurrences) after proposed corrective actions (Table 3) and after anticipated corrective actions (Table 4) are incorporated. A total of 80.5 expected failures (purged of unlikely occurrences) was thus obtained, yielding a failure rate of 0.178 based on 457 flight hours.

2.3 Analysis of New Failure Modes. It was expected that the production aircraft would experience new failure modes which had not been seen on prototype aircraft. An estimate of the rate of occurrence of new failure modes on production aircraft was obtained by considering the rate of occurrence of new failure modes through all Basic Engineering Development (BED) Phase and Maturity Phase flight testing.

Figure 1 presents a log-log plot of the cumulative rate of occurrence of new failure modes versus the cumulative test time during the BED Phase and Maturity Phase flight testing. The linear fit of the data on the log-log plot indicates that the occurrence of new failure modes follows a non-homogeneous Poisson process with intensity function given by $r(t) = \lambda B t^{B-1}$ (Reference 1). The function $r(t)$ represents the instantaneous rate of occurrence of failure modes. This function is shown in Figure 2 with estimates of λ and B obtained from the AMSAA Reliability Growth Model (Reference 1). The expected rate of occurrence of new failure modes on the production aircraft was estimated as follows:

if $E(N(t))$ represents the expected number of new failure modes in time t,

TABLE 4. FAILURES WITHOUT FIXES (PURGED OF UNLIKELY OCCURRENCES)

SUBSYSTEM	NUMBER OF FAILURES	SUBSYSTEM K FACTOR	EXPECTED NUMBER OF FAILURES*	EXPECTED FAILURE RATE (BASED ON 457 HOURS)
Rotor	0	.723	0.00	0.000
Transmission	3	.653	1.04	0.002
Propulsion (GFE)	0	.573	0.00	0.00
Propulsion (CFE)	1	.650	0.35	0.001
Electrical	4	.551	1.80	0.004
Avionics	4	.374	2.50	0.005
Airframe	2	.635	1.27	0.003
Electronic Controls	4	.829	0.68	0.001
Hydraulic/ Flight Controls	8	.662	2.70	0.006
TOTAL	26		10.34	0.022

*Obtained by applying the subsystem average k factor to the failures in that subsystem

TABLE 5. TOTAL FAILURES (PURGED OF UNLIKELY OCCURRENCES)

SUBSYSTEM	NUMBER OF FAILURES	EXPECTED NUMBER OF FAILURES*	EXPECTED FAILURE RATE (BASED ON 457 HOURS)
Rotor	21	5.80	0.013
Transmission	6	2.01	0.004
Propulsion (GFE)	8	3.92	0.009
Propulsion (CFE)	21	6.67	0.015
Electrical	28	13.57	0.030
Avionics	33	20.86	0.046
Airframe	25	8.12	0.018
Electronic Controls	19	3.24	0.007
Hydraulics/ Flight Controls	50	16.33	0.036
TOTAL	211	80.52	0.178

*Obtained by adding the number of failures in each subsystem from Table 4 and Table 5

FIGURE 1

CUMULATIVE RATE OF OCCURRENCE OF NEW SYSTEM FAILURE MODES VS. CUMULATIVE FLIGHT HOURS

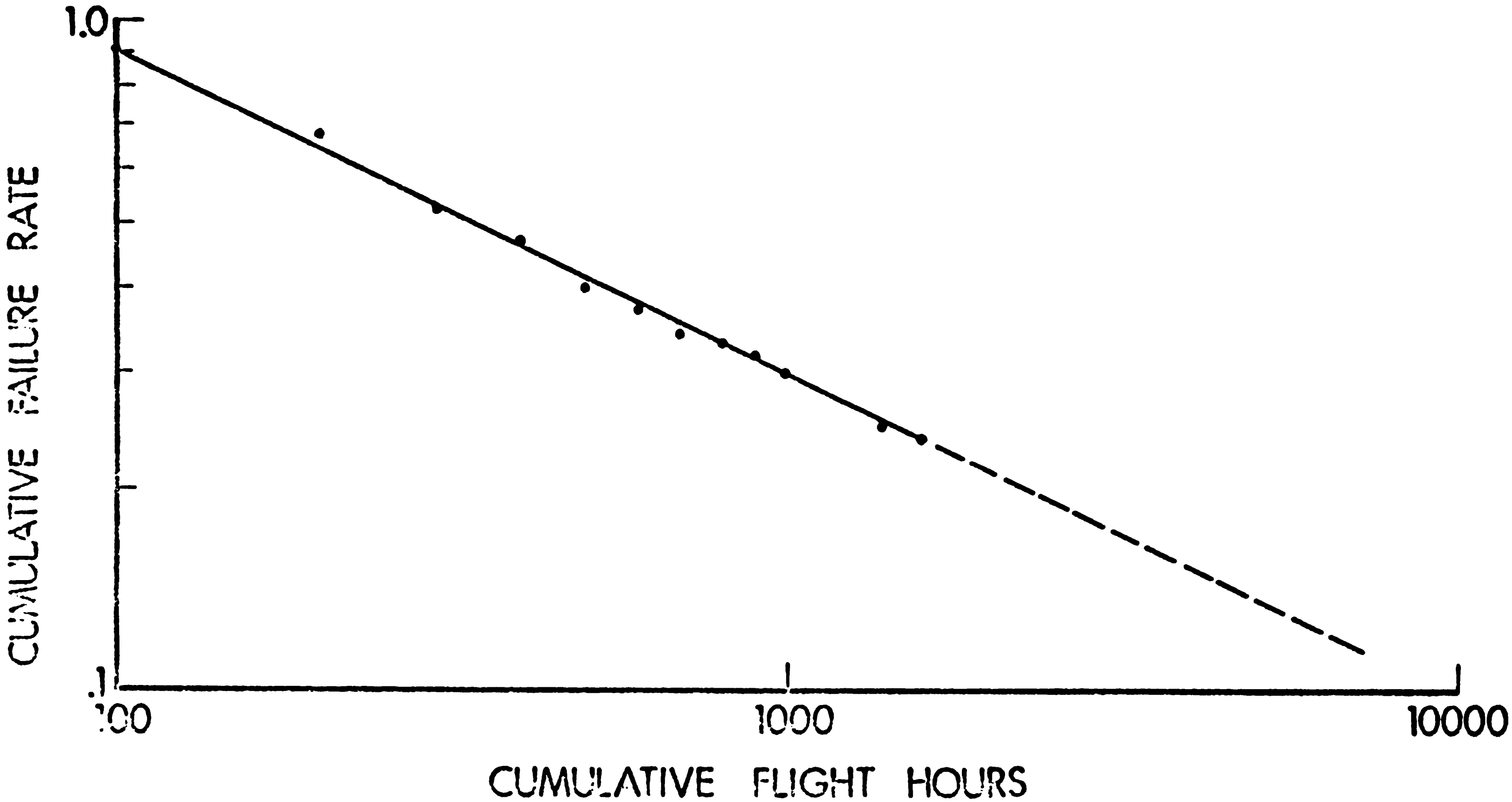
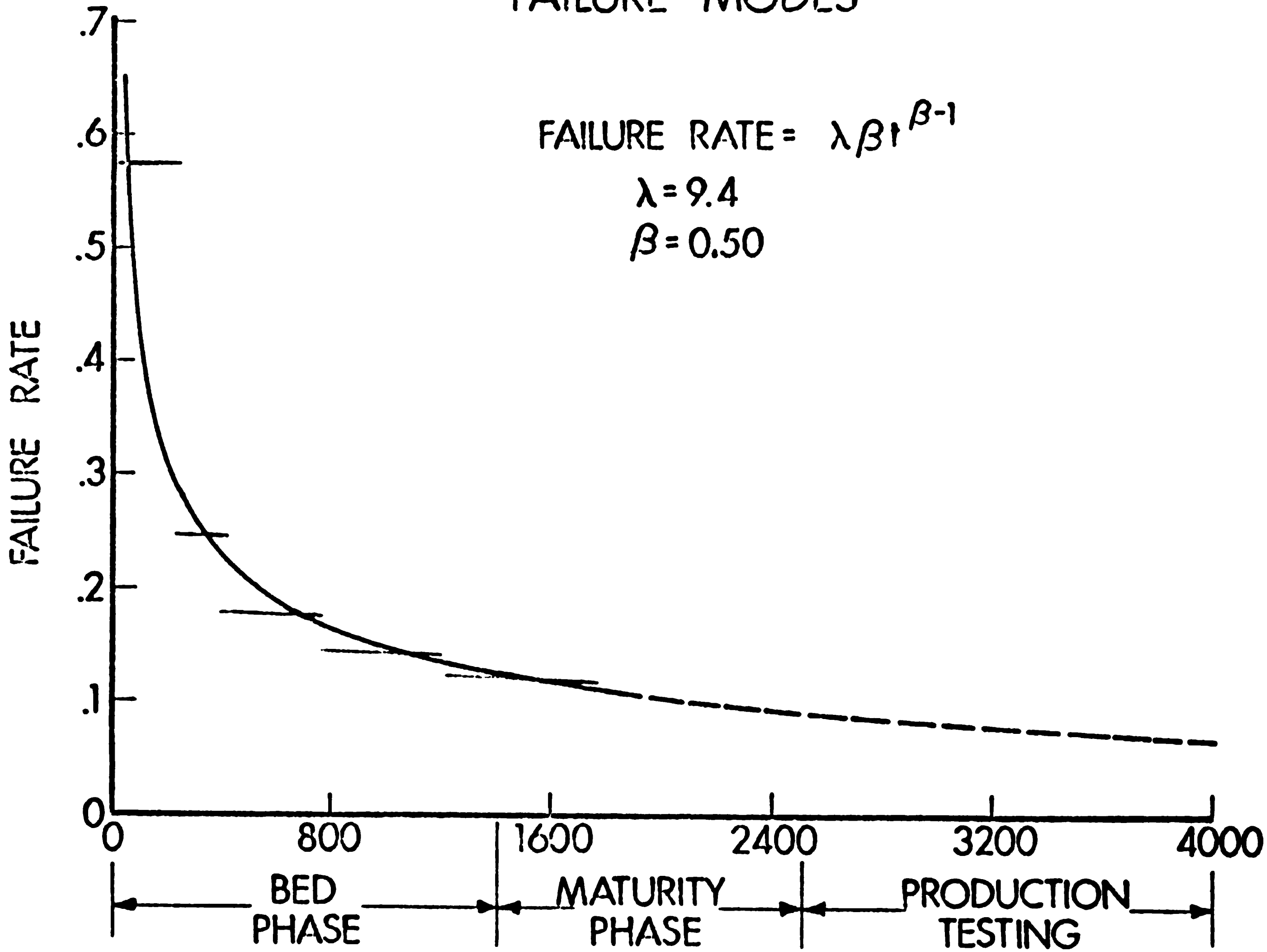


FIGURE 2

INSTANTANEOUS RATE OF OCCURRENCE OF NEW SYSTEM FAILURE MODES



then $E(N(t)) = \int_0^t \lambda B t^{B-1} dt = \lambda t^B$.

As depicted in Figure 2, production testing on the BLACK HAWK began after approximately 2500 flight hours of testing during the Basic Engineering Development (BED) Phase and the Maturity Phase. This testing on first production year aircraft was expected to accumulate 1500 flight hours which would extend the total flight time on prototype and production aircraft to 4000 flight hours. The number of new failure modes occurring on production aircraft would then be those occurring between 2500 and 4000 flight hours on Figure 2. The expected number of new failure modes between 2500 and 4000 flight hours was thus estimated by $E(4000) - E(2500) = \lambda(4000)^B - \lambda(2500)^B = 124$.

This value was obtained with the estimates of λ and B shown in Figure 2. The expected rate of occurrence of new failure modes was then estimated to be 124/1500 = 0.08.

2.4 Prediction of System MTBF. The prediction of system MTBF was obtained by adding the rate of occurrence of new failure modes to the failure rate of Table 5 which represents the rate for existing failure modes purged of unlikely occurrences and adjusted for contractor fixes. The calculation is given below:

0.08 (unseen failure mode rate) + 0.178 (Table 5) = 0.258 failures per hour.

The failure rate of 0.258 corresponds to a system MTBF of 3.9 hours.

3. CONCLUSIONS

The system MTBF prediction of 3.9 hours indicated to the Project Manager that no drastic measures in terms of program cost or testing would be required to meet the MTBF goal of 4.0 hours. At the same time, however, it was apparent that the contractor could not afford to reduce his efforts in improving reliability.

The production BLACK HAWK aircraft was subsequently delivered to the Army and demonstrated an MTBF of 3.7 hours as reported by the US Army Aviation Board during initial production testing. This value compares favorably with the prediction of 3.9 hours MTBF developed by the methodology in this paper. However, the MTBF of 3.7 reported by the Aviation Board did not include a large number of deferred maintenance actions. These deferred actions were largely quality control defects peculiar to initial production deliveries. If these defects are included, the MTBF of the early production aircraft is reduced from 3.7 hours to 2.9 hours. In future applications of the methodology in this paper, an adjustment will be required to account for large numbers of production line discrepancies.

REFERENCE

1. Crow, L. H. (1974). Reliability Analysis for Complex Repairable Systems, Reliability and Biometry, Edited by F. Proschan and R. J. Serfling, SIAM, Philadelphia, PA, pp 379-410.

POTENTIAL MILITARY APPLICATIONS OF TWO-PHASE SAMPLING

Patrick D. Allen
US Army Concepts Analysis Agency
8120 Woodmont Avenue
Bethesda, MD 20814
and
Stephen M. Rasey
Getty Oil Company
3810 Wilshire Blvd.
Los Angeles, CA 90010

ABSTRACT. Two-phase sampling is a relatively unknown technique which, in military applications, could save money in several areas of measurement. By first classifying the population through the use of an inexpensive method, one reduces the number of expensive measurements which would be required if traditional one-phase sampling were used. Two-phase sampling achieves this by better utilizing information gained by classification than does traditional stratified sampling.

I. Introduction. The purpose of this paper is to inform a larger portion of the military analytic community about the benefits and techniques of two-phase sampling. In these times of reduced budgets and increased emphasis on efficiency, any technique which can reduce the total cost of an experiment or measurement should be considered as an alternative approach to the problem.

The first part of the paper describes several military application areas in which two-phase sampling appears to be feasible and beneficial.

The second part familiarizes the reader with a brief description of the basic theory behind two-phase sampling. The third part gives the general problem formulation and solution, along with a computational example.

<u>II. Possible Areas of Application</u>. Many military studies require descriptive statistics on large populations for which comprehensive and exhaustive measurement is either infeasible or too costly. Sampling is used to extrapolate information about a set by measuring one or more attributes of a subset selected at random from this set. One can either choose the number of samples given a desired level of confidence, or one can obtain a confidence level given a fixed number of samples or budget limitation. In some cases, however, the costs associated with taking the number of samples required for a given confidence level are prohibitive or unnecessarily expensive. Two-phase sampling may allow the experimenter to obtain the desired level of accuracy within specified budget constraints.

There are several potential areas of application, both military and nonmilitary, in which two-phase sampling may be better than single-phase sampling. One situation involves the destructive/nondestructive methods of testing the components of a system. Destructive testing is not desirable when the component is expensive. If non-destructive testing is applied instead, the individual tests are less costly, but, since a much larger number of them might be required, total costs may be no less expensive. Successive use of both nondestructive (inexpensive) and destructive (expensive) measurements may provide the desired accuracy with less total cost than only a single-phase sample.

Another potential area of military applications is in inventory management. The Army maintains large numbers of items on stock which are inventoried annually. Usually, this is done on a 10 percent per month inventory cycle so that an annual inventory is completed by the end of the year; however, complete physical count inventories are expensive and time-consuming. Two-phase sampling may help in this area as well. The stock may be partitioned into high cost and low cost items. The high cost items will be completely inventoried by a physical count and the low cost items will be sampled. (One does not wish to have to count every nail, screw, or washer, although it has been attempted! Some inventory operations base a count on the weighing of many small homogeneous items.) Information about the state of the complete stock may be obtained by this method.

Another way of improving inventory control is to first segregate the items to be counted according to demand records. The first phase of sampling is the sorting of the potential inventory into categories comprising the most popular demand items. The second is the physical counting of those high demand items. The two-phase sampling technique will show how to do this at least cost.

A third area of military application is in personnel studies. The Army maintains large data bases describing various subsets of the military population relevant to military performance. (For example, the proportion of new recruits who went AWOL after receiving a monetary bonus was found to be quite high at one time.) Measuring a certain attribute of the population by personal interviews, or even questionnaires, can be both time-consuming and expensive. However, use of a first phase to subdivide the population before taking the second, more expensive set of measurements can give the same amount of accuracy at less cost.

Two-phase sampling is not the same thing as stratified random sampling. Most questionnaires are based upon stratified random sampling, which simply samples a random group from the stratified categories defined by the experimenter. However, if the data defining the strata are old or not currently accurate, the results of the estimation will be biased.[1] This is because two-phase sampling estimates the strata weights in the first phase, which provides more current estimates than in stratified random sampling.

III. A Brief Description of the Basic Theory. The classic presentation of two-phase sampling was given by Cochran in his book, Sampling Techniques. The total variance of the unbiased estimator for the stratified mean ($\bar{y}_{st}$) of the variable y in two-phase sampling is given by the equation:

$$V(\bar{y}_{st}) = S^2\left(\frac{1}{n'} - \frac{1}{N}\right) + \sum_{h=1}^{L} \frac{W_h S_h^2}{n'}\left(\frac{1}{v_h} - 1\right) \qquad [1]$$

where: S^2 is the variance of the total population,

S_h^2 is the variance of each stratum ($h=1,\ldots,L$),

W_h is the weight given to each stratum of the population where $W_h = N_h/N$,

N is the total population,

N_h is the subpopulation of stratum h,

n' is the size of the first sample,

n'_h is the number of stratum h in the first sample,

n_h is the number of stratum h in the second sample,

v_h is the proportion of the second sample taken from the population given by the first sample in that stratum, or

$$n_h = v_h n_h'. \tag{2}$$

The assumptions required for this to be true are:

(1) The first sample must be random.

(2) The second sample is a random subsample of the first sampling.

(3) The first sample is large enough so that the estimated weights (W_h) are all nonzero.

(4) Every proportion found in the optimal solution is less than the total number chosen in that stratum in the first sample, i.e. n_h' is greater than n_h.

The purpose of the first sampling is to determine the strata weights. The purpose of the second sample is to estimate the strata means so that the population mean may be estimated in an unbiased way. Stratified random sampling may give biased estimators, whereas two-phase sampling will not.[2]

Equation [1] may be rewritten as:

$$\frac{1}{n'}(S^2 - \sum W_h S_h^2) + \frac{1}{n}\sum \frac{W_h S_h^2}{v_h} = Var(\bar{y}_{st}) + \frac{S^2}{N} \tag{3}$$

Notice that if all of the variances of each stratum are identical, then the first term becomes zero and we are back to a single-phase sampling problem. Two-phase sampling seems to work the best when a small percentage of the population has high variance while the rest of the population has low variance. This also shows how two-phase sampling has a slight edge over stratified sampling. Notice that the variance terms on the right-hand side of equation [3] are larger than in the single sample case, which includes only $\mathrm{Var}(\bar{y}_{st})$. Also, in each term on the left-hand side, the numerator is smaller than in the single sample case. Of course, the sum of the terms on the left-hand side must be less than the right-hand side when the confidence interval is fixed. In the case where the optimal number of samples is small and the number of strata is large, two-phase sampling may do worse than single-phase sampling; however, such conditions are rare.

IV. General Problem Formulation and Solution. In general, the experimenter will wish to determine how many samples should be taken in each phase and in each stratum. Even though the problem formulation is nonlinear, there is a relatively simple method of solution. The general formulation is:

$$V_j = V(\bar{y}_{st}) = \sum_{i=1}^{L} a_{ij}x_i \qquad [4]$$

where

$$a_{ij} = W_i S_{ij}^2 \qquad [5]$$

and

$$x_i = (1/n_i - 1/N_i). \qquad [6]$$

The objective is to minimize K, where c_i is cost of measurement n_i, or

$$\mathrm{Min}\ K = \sum_{i=1}^{L} c_i n_i \qquad [7]$$

subject to:

$$\sum_{i=1}^{L} a_{ij}x_i \leq v_j \qquad [8]$$

and

$$0 \leq x_i \leq (1 - 1/N_i); \qquad [9]$$

where $i=1,\ldots,p$, where p is the number of characteristics addressed in the problem.

There are two ways to solve this formulation with a linear objective function and nonlinear constraints. One uses Lagrangian multipliers and the other uses geometric programing. A third solution uses a transformation which makes the objective function nonlinear and the constraints linear, as shown by Kokan and Khan (1967). All methods give the same solution,

$$x_{ij} = a_{ij} \sum_{i=1}^{L} \sqrt{c_i a_{ij}} \Big/ \left(\sqrt{c_i a_{ij}} (K_j) \right) \qquad [10]$$

where

$$K_j = v_j + \sum_{i=1}^{L} a_{ij}/N_i. \qquad [11]$$

The procedure for finding the solution to a given problem is to first find a constraint which is not dominated by any other constraint, and then calculate the solution given by equation [10]. If this solution satisfies all constraints, the optimal solution has been found. The following example should help explain the procedure.

A problem of estimating plant biomass in Range Science has been defined as a two-phase sampling problem. There are three strata of plants to be measured at minimum cost subject to constraints on total variance. In the constraints given below, the numbers for the a_{ij} have been combined into one coefficient, and the right-hand side has been divided through. The resulting formulation is:

$$\text{Minimize } .5n_1 + 5n_2 \qquad [12]$$

such that

$$200/n_1 + 12/n_2 \le 1$$

$$175/n_1 + 40/n_2 \le 1$$

$$150/n_1 + 30/n_2 \le 1$$

$$170/n_1 + 20/n_2 \le 1$$

$$175/n_1 + 25/n_2 \le 1$$

and

$$n_1 \ge n_2 > 0,$$

where n_i is the number of samples of phase i to be drawn. We can see from the first constraint that a lower bound on n_1 and n_2 is required to be $n_2 \ge 12$ and $n_1 \ge 200$ for the first constraint to be feasible. Similarly, for constraint two, $n_1 \ge 175$ and $n_2 \ge 40$. For the rest of the constraints we have $n_1 \ge 150$ and $n_2 \ge 30$, $n_1 \ge 170$ and $n_2 \ge 20$, and $n_1 \ge 175$ and $n_2 \ge 25$. But notice that the last three constraints are totally dominated by the second constraint. Therefore, the last three constraints can be ignored since they do not influence the problem in the presence of constraint two. This brief anaysis gives us an immediate lower bound on the minimum cost at 300.

When we try our solution from equation [10], $n_1 = 354.92$ and $n_2 = 27.49$ if we chose the first constraint as being active. However, this violates the second constraint. If we choose the second constraint as being active, then the optimal solution is obtained: $n_1 = 439.58$, $n_2 = 66.46$, and minimum cost is 552.08. Notice that since the solutions are noninteger, one must round up.

V. Conclusions. The two-phase sampling technique appears to be applicable to a large number of military problems. When applicable, it has potential for cost-saving relative to traditional single-phase sampling. Although not as easily applied as stratified random sampling, two-phase sampling usually will do better especially when the data are not up-to-date. The general solution shown in this paper can be applied with the use of a hand calculator. Two-phase sampling is not applicable when no data or a priori assumptions about the variances of the strata are known or when they are expensive to obtain.

Footnotes

1) Cochran, Sampling Techniques, pp. 117-119.

2) Ibid.

Bibliography

Cochran, William G., Sampling Techniques, New York, John Wiley and Sons, 1977.

Kokan, A. R., and S. Khan, "Optimum Allocation in Multivariate Surveys: An Analytical Solution," Journal of the Royal Statistical Society, Series B, 1967.

Pechanec, J. F., and G. D. Pickford, "A Weight Estimate Method for Determination of Range or Pasture Production," Journal of the American Society of Agronomy, Vol. 29, 1937.

Stuart, Alan, Basic Ideas of Scientific Sampling, London, Griffin and Company, LTD., 1976.

DESIGN OF A MULTIPLE SAMPLE WESTENBERG TYPE TEST FOR SMALL SAMPLE SIZES

James R. Knaub, Jr.
US Army Logistics Center
Fort Lee, Virginia

ABSTRACT. This paper discusses an experimental design which is based on a multiple sample extension of Westenberg's Interquartile Range Test (random effects) and the Westenberg-Mood Median Test (fixed effects). Both type I and type II error probabilities have been computerized for small sample sizes using an exact test which is a multiple sample extension of a two-tailed Fisher Exact Test. Unlike other distribution free analysis of variance tests such as Kruskal-Wallis or Friedman, random effects may be investigated, and practical type II error analyses are available. When sample sizes are large, the chi-square distribution will provide a reasonable approximation; however, when sample sizes are small this test is needed.

Current plans are to use this test in coordination with other statistical methods to analyze data from a study being conducted concerning manpower requirements in the various types of units found in the US Army. This will be discussed briefly in this paper also.

INTRODUCTION

This paper is based upon a part of Annex B to reference 2, which is a study plan for gathering data on the amount of time expected to be spent by soldiers, particularly in a wartime environment, on duties other than their primary Military Occupational Specialty (MOS) duties. An earlier study, reference 4, indicated that, for example, median times to be spent on Kitchen Police (KP) activities were expected to be the same when comparing questionnaire results from a number of US Army posts. (The Kruskal-Wallis test was used as the sample distributions were decidedly non-normal.) However, a cursory study of the data indicated strongly that the variances were quite different. One explanation would be that various types of Army units stationed at these posts could have influenced the data in such a way. In the current study, reference 2, units are broken out by category (combat, combat support, combat service support) and logical region (division, corps and echelons above corps), and in some cases by unit type (e.g., transportation, chemical, etc.). In reference 2, Annex B suggests a number of analytical methods for examining the results of this study as designed in that study plan. For comparing units in three logical regions (LRs) and three categories (CATs) for both fixed and random effects when the assumption of normality may not be reasonable, the methods of this paper have been developed. The Statistical Package for the Social Sciences (SPSS) will be used for much of the analysis, but the computer programs for the hypothesis tests to be given here are given as an appendix to this paper since they are, of course, not found in the SPSS.

NEW STATISTICAL HYPOTHESIS TEST
FOR DISTRIBUTION-FREE ANOVA

This test is basically a multiple sample extension of the Westenberg Interquartile Range Test and the Westenberg-Mood Median Test. As such, the underlying statistics are from Fisher's Exact Test for proportions. At the foundation of Fisher's Exact Test is the following mathematical expression:

$$\frac{[(A+B)!\,r_1^{B}(1-r_1)^{A}(C+D)!\,r_2^{D}(1-r_2)^{C}]/[A!B!C!D!]}{[(A+B+C+D)!\,r_0^{B+D}(1-r_0)^{A+C}]/[(B+D)!(A+C)!]}$$

This is the probability of having A "failures" in one sample when the probability of "success" is r_1, and C "failures" in the other sample with a probability of "success" r_2, normalized to having a total of A+C "failures" in the combined sample with overall probability of "success" r_0. ("Success" and "failure" must be defined in each application. These words are used more accurately when Fisher's Exact Test is applied in the comparison of system reliabilities, however, when used to judge, for example, whether two (or more) samples appear to come from populations with the same interquartile range, "success" could denote an observed value within the interquartile range of the combined sample, or vice versa.)

When determining the probability of a type I error, since the null hypothesis is that of equal probabilities, r_1, r_2 and r_0 are set equal to each other, and the above expression reduces to:

$$\frac{(A+B)!(C+D)!(B+D)!(A+C)!}{A!B!C!D!(A+B+C+D)!}$$

This is the probability of obtaining the event observed given that the null hypothesis is true. It is found in reference 6 and in other forms in reference 1 and other sources. In order to determine the probability of a type I error, this case and all more extreme cases must be analyzed and their probabilities added. [This is a one-tailed test, however, the multiple sample version (i.e., more than two samples), is an extended two-tailed test.]

The probability of a type II error was derived for the Fisher Exact Test and tabulated in a study at White Sands Missile Range, to become part of the material held at the Defense Documentation Center, under the title "Missile Round Sample Size Considerations for Test Planning and Reporting," which is reference 3. Under the alternative hypothesis of unequal r_1, r_2 and r_0, the probability of the event observed becomes:

$$\frac{(A+B)!\,r_1^{B}(1-r_1)^{A}(C+D)!\,r_2^{D}(1-r_2)^{C}}{A!B!C!D!p}$$

where $p \equiv P$[total of A+C failures in the combined sample]

$$= \sum_{x=0}^{A+C} \binom{A+B}{x} (1 - r_1)^x r_1^{A+B-x} \binom{C+D}{A+C-x} r_2^{D-A+x} (1 - r_2)^{A+C-x}$$

(When $B \geq C$ and $D \geq A$, otherwise, not all terms will be present)

To obtain the probability of a type II error, the case observed and all less extreme cases must be analyzed and their probabilities added.

The general expression for a multiple sample version of this test is

$$\frac{\{\prod_{k=1}^{n}[(A_k + B_k)! r_k^{B_k}(1 - r_k)^{A_k}]\}\{[\sum_{k=1}^{n} A_k]! [\sum_{k=1}^{n} B_k]!\}}{\{\prod_{k=1}^{n} A_k! B_k! r_0^{B_k}(1 - r_0)^{A_k}\}\{[\sum_{k=1}^{n}(A_k + B_k)]!\}}$$

Where,

$A_k \equiv$The number of "failures" in sample k,

$B_k \equiv$The number of "successes" in sample k,

$r_k \equiv$The probability of "successes" in the kth sample,

and

$r_0 \equiv$The overall probability of "success."

Also, $\sum_k A_k$ is a constant.

In order to determine which cases are "more extreme" for the purposes of calculation, consider that the idealized result under the null hypothesis constitutes a point in n-space, where n is the number of samples being compared, and so does the idealized result under the alternative hypothesis, as well as the actual result obtained. The square of the geometric distance between the result obtained and that ideally obtained for the null hypothesis is denoted DA, and similarly for the alternative hypothesis one has DB in the computer program which was developed for this test. The test can be adjusted for unequal sample sizes, however, writing a general program for this is difficult, so the current form of the program only considers equal sample sizes.

Of all the ANOVA methods available, this method can be the most informative and accurate because it makes no distributional assumption (unlike the F-test); it cannot be easily fooled by random effects (unlike the Kruskal-Wallis Test which is basically a fixed effects test); and it lends itself to interpretable alternative hypotheses for meaningful power analyses (essentially unlike any other ANOVA).

The alternative hypothesis that will be used in the case of comparing four samples could, for example, be that two of the proportions will be 0.4:0.6 and two will be 0.6:0.4. In the case of three samples, 0.4:0.6, 0.5:0.5, and 0.6:0.4 could be used. This is a more stringent alternative than in the four sample cases; however, the odd number of samples makes this unavoidable, and these alternatives are relatively easy to interpret and communicate to the decision maker. For future reference, in this example of alternative hypotheses, one has W = 0.40.

APPLICATION

It was found in reference 4 that data of the nature to be gathered for reference 2 may be decidedly non-normally distributed. Other distributional forms could be experimented with, and/or transformations used. However, one will not be certain of the effects of such manipulations. In reference 4, the Kruskal-Wallis One-Way Analysis of Variance (ANOVA) was used. This is a distribution free test which concentrates on location. (It is basically an extension of the Wilcoxon Rank Sum Test to more than two samples.) From this test, in an example given on KP, no significant difference was found between data sources as far as location was concerned. However, it was obvious that large differences in dispersion did exist. The Kruskal-Wallis Test could not discern this, nor can the power of the test be clearly described. For these reasons, a k-sample extension of Westenberg's Interquartile Range Test and the Westenberg-Mood Median Test (see reference 1), where k is greater than or equal to 3, has been derived based upon work performed at White Sands Missile Range (see reference 3). Using this new hypothesis test, both random and fixed effects can be investigated (dispersion and location), and the probability of a type II error (in a simplified sense, the complement of power, see reference 3), will be provided for an understandable alternative hypothesis. (It should be emphasized that being distribution-free, bimodality, etc., will not be a problem.) The computerization of this new hypothesis test runs into practical limitations for most cases when k is greater than 5. However, the Chi-square test for proportions will adequately approximate this test in most practical situations. If all sample sizes are 20 or better, Chi-square can be used to compare all nine samples, or even all 18 when considering both CONUS and Europe.

Evaluation is proposed as follows: the LRXCAT cells will be compared using this new test, and also using normal-theory ANOVA (which can be implemented by use of the SPSS). If the results are substantially identical, then Duncan's multiple range test can be used (see reference 7) to discern which LRXCAT's have mean values which are indistinguishable at this level of testing. However, if the results of the new vs normal-theory ANOVA's are not compatible, further use of the new test is dictated.

(Note that the LRXCAT cells are being treated independently; i.e., neither LR nor CAT effect is being studied separately. Thus, this is a one-way analysis of variance.)

Let A represent LR1XCATI, B represent LR1XCATII, C be LR1XCATIII, D be LR2XCATI, etc., so that A through I can be filled into the following table as shown:

CAT:	I	II	III
LR:			
1	A	B	C
2	D	E	F
3	G	H	I

(Note that if some of these cells do not exist, this analysis will not suffer.)

Make the following four sets of comparisons: 1) A, B, D, E; 2) B, C, E, F; 3) D, E, G, H; 4) E, F, H, I. If α =5% is used, for each comparison, then the actual probability of finding at least one "significant" difference is between 5% and 18.6%, where the latter case would apply if all of the comparisons were independent. (See reference 5.) Therefore, in reality, 5% is less than α , which is less than 18.6%. However, given that there is one comparison (in CONUS) out of the four enumerated above which shows a "significant" difference, purely by chance, then the probability that the same comparison under OCONUS will do likewise is $\binom{4}{1}(0.95)^3(0.05)(0.25)$ + $\binom{4}{2}(0.95)^2(0.05)^2(0.5) + \binom{4}{3}(0.95)(0.05)^3(0.75) + \binom{4}{4}(0.05)^4(1) = 0.05$. (This assumes that CONUS and OCONUS are equivalent. Something which is addressed elsewhere in reference 2.) Therefore, the fact that comparisons will be done in OCONUS on the same LRXCAT groupings as in CONUS can be used to analytical advantage if CONUS and OCONUS are identical. Also, applying the same reasoning as above, if interactive effects are ignored and if samples can be combined first by LR, and later by CAT, then a 10% significance level for each comparison translates into a 10% probability that should one of these two comparisons show a "significant" difference purely by chance, then the same comparison under OCONUS will do likewise. Interactive effects can be handled as on pages 139-140 of reference 1. However, the more comparisons that are to be made, the less certainty there is in the stand-alone analysis of either CONUS or European generated data.

If, under this plan, B, C, E and F appear to have essentially the same variances, but none of the others do, then their variances, (i.e., for B, C, E and F) could be "averaged" and considered as equal for purposes of feeding a model, and the others kept as distinctive from one another. If D, E, G and H appear to have essentially the same locations and so do B, C, E and F, then E will be averaged with the group in which the comparison was most significant. Further, in this example since B, C, E and F had indistinguishable locations and dispersions (at some acceptable level of power), future data may be collected on them synonymously. (If it is decided that the risk in doing this should first be reduced further, then increase the sample sizes in the comparison.) In this way, the number of distinct tables of organization and equipment (TOEs) influenced by distinct sets of nonavailable time factor values, can be determined.

Note that although the extended Westenberg Interquartile Range Test is basically a random effects test and the extended Westenberg-Mood Median Test is a fixed effects test, the random effects test can be confounded although this could only occur in cases where a definite difference in location should be found, and still the chance of confounding the random effects test would be small. All things considered, this methodology has fewer disadvantages for this application than any alternative available.

The fact that CONUS and OCONUS units will both be studied can be very helpful here. If a certain set of LRXCAT groupings seem similar in the CONUS study, and that same set appears similar in the OCONUS study, then this would support the conclusion that they be labeled that way. If, however, one set of LRXCAT groupings appear similar in CONUS and not in OCONUS, for example, then either the CONUS/OCONUS distinction was important, or the supposed similarities and dissimilarities may have been by chance. (As an alternative approach, if two sets of units appear identical in CONUS, one may pinpoint those two in the OCONUS study and compare them specifically with a two-sample test.)

The following will also be considered: If two observations per unit XCATXLR grouping can be taken, where "unit" represents a unit type such as transportation, the new test or a more straight forward binomial comparison can yield some over-all dispersion and/or location information for a relatively large number of unitXCATXLRs. The computer programs marked WB1 and WC are designed for this purpose (see Appendix A). Program WC can be used to analyze several hundred samples of two observations each.

APPENDIX A

COMPUTER PROGRAMS

A-1. Program marked WA is for a two-tailed Fisher and for three and four samples.

A-2. Program WA1 is for type I errors only (it runs faster and can be used if the "consumer risk" is unimportant).

A-3. Program WB1 is for type I errors with samples of size 2 when there are up to 14 samples--computer time for larger number of samples would be prohibitive.

A-4. Program WC is also for samples of size 2 but can be used for up to several hundred samples--note that the hypotheses are different from the other three programs.

In these programs, NS is the number of samples being considered, IS is the size of each sample, DA is the "distance" of the observed result from the null hypothesis and DB is its "distance" from the alternative hypothesis where 100 x W% of the populations from which half of the samples are drawn is found outside (or inside) of the interquartile range of the combined sample, or above (or below) the median of the combined sample and vice versa for the other half. For example, if NS = 4, IS = 20, and W = 0.25, then the null hypothesis, H_o, and the alternative hypothesis, H_1, can be represented as

H_o:	10	10	10	10
	10	10	10	10
H_1:	5	5	15	15
	15	15	5	5

If the observed values are

7	8	12	13
13	12	8	7

then DA = $(10-7)^2 + (10-8)^2 + (10-12)^2 + (10-13)^2 = 26$ and DB = $(5-7)^2 + (5-8)^2 + (15-12)^2 + (15-13)^2 = 26$. Note that DB is calculated using the order that makes it the smallest possible.

The programs are written such that the denominator of the basic mathematical expression shown near the beginning of this paper will have $r_o = R$. If R is replaced by a number slightly greater than 0.5 then the distribution for the alternative hypothesis can be "normalized" so that the area under the representative curve is unity. Some values for R are given in appendix C.

In the fourth program, WC, the number of zeroes, NZ, along with RA and RB are explained in the first "FORMAT" statement. (Note that NZ is the same number as DA in program WB1.) The "FORMAT" statement is written in terms of the interquartile range of the combined sample, but could easily be in terms of the median of the combined sample.

PA and PB have common meanings across these programs. PA is the probability that if the null hypothesis were true, the result obtained, or a less likely one, would occur. PB is the probability that if the alternative hypothesis were true, the result obtained, or a less likely one under these conditions, would occur.

```
@FTN,S K.WA,KEK
FTN 10R1    01/05/82-11:57(24,)
                      COMMON F(500),G(500),H(500),DA,DB,I1,I2,I3,I4,A,B,IS,W,R
                      DOUBLE PRECISION A,B,DORP,ANS,AIS,AICAF
                      WRITE (6,1100)
                 1100 FORMAT(5X,'THE FORMAT IS  I2,1X,I3,1X,F6.2,1X,F4.2,1X,F6.2,'
                    1        '1X,F8.6 '
                    1   /5X,' FOR  NS,IS,DA,W,DB,R')
                 5000 READ(5,1,END=2010)NS,IS,DA,W,DB,R
                    1 FORMAT(I2,1X,I3,1X,F6.2,1X,F4.2,1X,F6.2,1X,F8.6)
                      A=IS
                      L=IS+1
                      DO 18 I=1,L
 1                    AA=I-1
 1                    F(I) = (0.5*A-AA)*(0.5*A-AA)
 1                    G(I) = (W*A-AA)*(W*A-AA)
 1                 18 H(I) = ((1.0-W)*A-AA)*((1.0-W)*A-AA)
                      A=0
                      B=0
                      C=0
                      J=IS*NS/2
                      GO TO (1000,2,3,4)NS
                 1000 WRITE (6,1001)
                 1001 FORMAT (5X,'BAD VALUE OF NS')
                      GO TO 5000
                    3 DO 30 I1=0,IS
 1                    DO 30 I2=0,IS
 2                    I3 = J - I1 - I2
 2                    IF((I3.LT.0).OR.(I3.GT.IS)) GO TO 31
 2                    CALL  SUB3
 2                 31 CONTINUE
 2                 30 CONTINUE
                   70 DORP=1.0
                      ICAFI = (NS*IS/2)-1
                      DO 71 ICAF = 0,ICAFI
 1                    ANS=NS
 1                    AIS=IS
 1                    AICAF=ICAF
 1                 71 DORP = DORP * (((ANS*AIS/2.0)-AICAF)/((ANS*AIS)-AICAF))
                      PA = A*DORP
                      PB = B*DORP
                      GO TO 2000
                    4 DO 130 I1=0,IS
 1                    DO 130 I2=0,IS
 2                    DO 130 I3=0,IS
 3                    I4 = J - I1 - I2 - I3
 3                    IF((I4.LT.0).OR.(I4.GT.IS)) GO TO 131
 3                    CALL  SUB4
 3                131 CONTINUE
 3                130 CONTINUE
                      GO TO 70
                    2 DO 230 I1 = 0,IS
 1                    I2 = J - I1
 1                    IF((I2.LT.0).OR.(I2.GT.IS)) GO TO 231
 1                    CALL SUB2
 1                231 CONTINUE
 1                230 CONTINUE
                      GO TO 70
                 2000 WRITE(6,2001) PA,PB
                 2001 FORMAT(1X,F6.4,5X,F9.7)
                      GO TO 5000
```

```
 2010 STOP
      END
```

```
      SUBROUTINE  SUB3
      COMMON F(500),G(500),H(500),DA,DB,I1,I2,I3,I4,A,B,IS,W,R
      DOUBLE PRECISION A,B,PROD,X,Y
      III = I1+1
      II2 = I2+1
      II3 =I3+1
      DAA = F(II1)+F(II2)+F(II3)
      DBB1 = G(II1)+F(II2)+H(II3)
      DBB2 = G(II1)+F(II3)+H(II2)
      DBB3 = G(II2)+F(II1)+H(II3)
      DBB4 = G(II2)+F(II3)+H(II1)
      DBB5 = G(II3)+F(II1)+H(II2)
      DBB6 = G(II3)+F(II2)+H(II1)
      DBB = MIN(DBB1,DBB2,DBB3,DBB4,DBB5,DBB6)
      PROD = 1.0
      IF(I1.EQ.IS) GO TO 52
      III = IS -1 -I1
      DO 51 IFAC = 0,III
      X = IS-IFAC
      Y = X - I1
   51 PROD = PROD*X/Y
   52 IF(I2.EQ.IS) GO TO 54
      III2 = IS-1-I2
      DO 53 IFAC = 0,III2
      X = IS-IFAC
      Y = X - I2
   53 PROD = PROD*X/Y
   54 IF (I3.EQ.IS) GO TO 56
      III3 = IS-1-I3
      DO 55 IFAC = 0,III3
      X = IS-IFAC
      Y = X-I3
   55 PROD = PROD*X/Y
   56 IF(DAA.LT.DA) GO TO 61
      A=A+PROD
   61 IF(DBB.GE.DB) GO TO 62
      RETURN
   62 B = B+PROD*(((((1.0/R)*(1.0-W)))**(I1*2))*((((1.0/R)*W))
     1      **(I3*2)))
      RETURN
      END
```

```
      SUBROUTINE  SUB4
      COMMON F(500),G(500),H(500),DA,DB,I1,I2,I3,I4,A,B,IS,W,R
```

```
      DOUBLE PRECISION A,B,PROD,X,Y
      II1=I1+1
      II2=I2+1
      II3=I3+1
      II4=I4+1
      DAA = F(II1)+F(II2)+F(II3)+F(II4)
      DBB1 = G(II1)+G(II2)+H(II3)+H(II4)
      DBB2 = G(II1)+G(II3)+H(II2)+H(II4)
      DBB3 = G(II1)+G(II4)+H(II2)+H(II3)
      DBB4 = G(II2)+G(II3)+H(II1)+H(II4)
      DBB5 = G(II2)+G(II4)+H(II1)+H(II3)
      DBB6 = G(II3)+G(II4)+H(II1)+H(II2)
      DBB = MIN(DBB1,DBB2,DBB3,DBB4,DBB5,DBB6)
      PROD = 1.0
      IF(I1.EQ.IS) GO TO 152
      III = IS-1-I1
      DO 151 IFAC =0,III
      X=IS-IFAC
      Y=X-I1
  151 PROD = PROD *X /Y
  152 IF(I2.EQ.IS) GO TO 154
      III2 = IS-1-I2
      DO 153 IFAC =0,III2
      X = IS-IFAC
      Y = X-I2
  153 PROD = PROD *X /Y
  154 IF(I3.EQ.IS) GO TO 156
      III3 = IS-1-I3
      DO 155 IFAC =0,III3
      X = IS-IFAC
      Y = X-I3
  155 PROD = PROD *X /Y
  156 IF (I4.EQ.IS) GO TO 158
      III4 = IS-1-I4
      DO 157 IFAC =0,III4
      X = IS-IFAC
      Y = X-I4
  157 PROD = PROD *X /Y
  158 IF(DAA.LT.DA) GO TO 161
      A=A+PROD
  161 IF(DBB.GE.DB) GO TO 162
      RETURN
  162 B = B+PROD*((((1.0/R)*(1.0-W))**((I1+I2)*2))*(((1.0/R)*W)**
     1   ((I3+I4)*2)))
      RETURN
      END
```

```
      SUBROUTINE SUB2
      COMMON F(500),G(500),H(500),DA,DB,I1,I2,I3,I4,A,B,IS,W,R
      DOUBLE PRECISION A,B,PROD,X,Y
      II1=I1+1
      II2=I2+1
      DAA = F(II1) + F(II2)
      DBB1 = G(II1) + H(II2)
```

```
          DBB2 = G(II2) + H(III)
          DBB = MIN(DBB1,DBB2)
          PROD = 1.0
          IF(I1.EQ.IS) GO TO 252
          III = IS - 1 - I1
          DO 251 IFAC = 0,III
1         X = IS - IFAC
1         Y = X - I1
1     251 PROD = PROD*X/Y
      252 IF(I2.EQ.IS) GO TO 254
          III2 = IS-1-I2
          DO 253 IFAC = 0,III2
1         X = IS-IFAC
1         Y = X-I2
1     253 PROD = PROD*X/Y
      254 IF(DAA.LT.DA) GO TO 257
          A = A+PROD
      257 IF(DBB.GE.DB) GO TO 258
          RETURN
      258 B = B+PROD*((((1.0/R)*(1.0-W)))**(I1*2))*((((1.0/R)*W))
         1      **(I2*2)))
          RETURN
          END

END FTN 916 IBANK 252 DBANK 1513 COMMON
```

```
@FTN,S K.WA1,KEK
FTN 10R1   01/05/82-11:57(61,)
                    COMMON F(200),DA,I1,I2,I3,I4,A,IS
                    DOUBLE PRECISION A,DORP,ANS,AIS,AICAF
                    WRITE (6,1100)
              1100  FORMAT(5X,'THE FORMAT IS  I2,1X,I3,1X,F6.2 '
                 1   /5X,' FOR  NS,IS,DA')
              5000  READ(5,1,END=2010)NS,IS,DA
                 1  FORMAT(I2,1X,I3,1X,F6.2)
                    A=IS
                    L=IS+1
                    DO 18 I=1,L
 1                  AA=I-1
 1              18  F(I) = (0.5*A-AA)*(0.5*A-AA)
                    A=0
                    J=IS*NS/2
                    GO TO (1000,1000,3,4)NS
              1000  WRITE (6,1001)
              1001  FORMAT (5X,'BAD VALUE OF NS')
                    GO TO 5000
                 3  DO 30 I1=0,IS
 1                  DO 30 I2=0,IS
 2                  I3 = J - I1 - I2
 2                  IF((I3.LT.0).OR.(I3.GT.IS)) GO TO 31
 2                  CALL  SUB3
 2              31  CONTINUE
 2              30  CONTINUE
                70  DORP=1.0
                    ICAFI = (NS*IS/2)-1
                    DO 71 ICAF = 0,ICAFI
 1                  ANS=NS
 1                  AIS=IS
 1                  AICAF=ICAF
 1              71  DORP = DORP * (((ANS*AIS/2.0)-AICAF)/((ANS*AIS)-AICAF))
                    PA = A*DORP
                    GO TO 2000
                 4  DO 130 I1=0,IS
 1                  DO 130 I2=0,IS
 2                  DO 130 I3=0,IS
 3                  I4 = J - I1 - I2 - I3
 3                  IF((I4.LT.0).OR.(I4.GT.IS)) GO TO 131
 3                  CALL  SUB4
 3             131  CONTINUE
 3             130  CONTINUE
                    GO TO 70
              2000  WRITE(6,2001) PA
              2001  FORMAT(1X,F6.4)
                    GO TO 5000
              2010  STOP
                    END

                    SUBROUTINE  SUB3
                    COMMON F(200),DA,I1,I2,I3,I4,A,IS
                    DOUBLE PRECISION A,PROD,X,Y
```

```
          III = I1+1
          II2 = I2+1
          II3 =I3+1
          DAA = F(III1)+F(II2)+F(II3)
          IF(DAA.GE.DA) GO TO 61
          RETURN
       61 PROD = 1.0
          IF(I1.EQ.IS) GO TO 52
          III = IS -1 -I1
          DO 51 IFAC = 0,III
1         X = IS-IFAC
1         Y = X - I1
1      51 PROD = PROD*X/Y
       52 IF(I2.EQ.IS) GO TO 54
          III2 = IS-1-I2
          DO 53 IFAC = 0,III2
1         X = IS-IFAC
1         Y = X - I2
1      53 PROD = PROD*X/Y
       54 IF (I3.EQ.IS) GO TO 56
          III3 = IS-1-I3
          DO 55 IFAC = 0,III3
1         X = IS-IFAC
1         Y = X-I3
1      55 PROD = PROD*X/Y
       56 A=A+PROD
          RETURN
          END
```

```
          SUBROUTINE  SUB4
          COMMON F(200),DA,I1,I2,I3,I4,A,IS
          DOUBLE PRECISION A,PROD,X,Y
          II1=I1+1
          II2=I2+1
          II3=I3+1
          II4=I4+1
          DAA = F(II1)+F(II2)+F(II3)+F(II4)
          IF(DAA.GE.DA) GO TO 161
          RETURN
      161 PROD = 1.0
          IF(I1.EQ.IS) GO TO 152
          III = IS-1-I1
          DO 151 IFAC =0,III
1         X=IS-IFAC
1         Y=X-I1
1     151 PROD = PROD *X /Y
      152 IF(I2.EQ.IS) GO TO 154
          III2 = IS-1-I2
          DO 153 IFAC =0,III2
1         X = IS-IFAC
1         Y = X-I2
1     153 PROD = PROD *X /Y
      154 IF(I3.EQ.IS) GO TO 156
          III3 = IS-1-I3
```

```
              DO 155 IFAC =0,III3
1             X = IS-IFAC
1             Y = X-I3
1         155 PROD = PROD *X /Y
          156 IF (I4.EQ.IS) GO TO 158
              III4 = IS-1-I4
              DO 157 IFAC =0,III4
1             X = IS-IFAC
1             Y = X-I4
1         157 PROD = PROD *X /Y
          158 A=A+PROD
              RETURN
              END

END FTN 553 IBANK 156 DBANK 208 COMMON
```

```
@FTN,S K.WB1,KEK
FTN 10R1   01/05/82-11:57(45,)
                    COMMON F(200),DA,I1,I2,I3,I4,I5,I6,I7,I8,I9,I10,I11,I12,I13,
                 1        I14,A,IS
                    DOUBLE PRECISION A,DORP,ANS,AIS,AICAF
              C  READ AS MANY CARDS AS DESIRED, ONE AT A TIME
              C  END BY USING '@'
                    WRITE (6,1100)
               1100 FORMAT (5X,' THE FORMAT IS     I2,1X,F6.2 '
                 1     /5X, ' FOR    NS,DA ')
               5000 READ(5,1,END=2010)NS,DA
                  1 FORMAT(I2,1X,F6.2,1X,F4.2,1X,F6.2)
                    IS = 2
                    A=IS
                    L=IS+1
                    DO 118 I=1,L
 1                  AA=I-1
 1              118 F(I) = (0.5*A-AA)*(0.5*A-AA)
                    A=0
                    J=IS*NS/2
                    GO TO (1000,1000,1000,4,1000,6,1000,8,1000,10,1000,12,
                 1       1000,14)NS
               1000 WRITE (6,1001)
               1001 FORMAT (5X,'BAD VALUE OF NS')
                    GO TO 5000
                 70 DORP=1.0
                    ICAFI = (NS*IS/2)-1
                    DO 71 ICAF = 0,ICAFI
 1                  ANS=NS
 1                  AIS=IS
 1                  AICAF=ICAF
 1               71 DORP = DORP * (((ANS*AIS/2.0)-AICAF)/((ANS*AIS)-AICAF))
                    PA = A*DORP
                    GO TO 2000
                  4 DO 130 I1=0,IS
 1                  DO 130 I2=0,IS
 2                  DO 130 I3=0,IS
 3                  I4 = J - I1 - I2 - I3
 3                  IF((I4.LT.0).OR.(I4.GT.IS)) GO TO 131
 3                  CALL  SUB4
 3              131 CONTINUE
 3              130 CONTINUE
                    GO TO 70
                  6 DO 330 I1=0,IS
 1                  DO 330 I2=0,IS
 2                  DO 330 I3=0,IS
 3                  DO 330 I4=0,IS
 4                  DO 330 I5=0,IS
 5                  I6 = J - I1 - I2 - I3 - I4 - I5
 5                  IF((I6.LT.0).OR.(I6.GT.IS)) GO TO 331
 5                  CALL  SUB6
 5              331 CONTINUE
 5              330 CONTINUE
                    GO TO 70
                  8 DO 530 I1=0,IS
 1                  DO 530 I2=0,IS
 2                  DO 530 I3=0,IS
 3                  DO 530 I4=0,IS
 4                  DO 530 I5=0,IS
 5                  DO 530 I6=0,IS
 6                  DO 530 I7=0,IS
```

```
 7           I8 = J-I1-I2-I3-I4-I5-I6-I7
 7           IF((I8.LT.0).OR.(I8.GT.IS))  GO TO 531
 7           CALL   SUB8
 7       531 CONTINUE
 7       530 CONTINUE
             GO TO 70
          10 DO 730 I1=0,IS
 1           DO 730 I2=0,IS
 2           DO 730 I3=0,IS
 3           DO 730 I4 = 0,IS
 4           DO 730 I5=0,IS
 5           DO 730 I6=0,IS
 6           DO 730 I7=0,IS
 7           DO 730 I8=0,IS
 8           DO 730 I9=0,IS
 9           I10 = J-I1-I2-I3-I4-I5-I6-I7-I8-I9
 9           IF((I10.LT.0).OR.(I10.GT.IS)) GO TO 731
 9           CALL   SUB10
 9       731 CONTINUE
 9       730 CONTINUE
             GO TO 70
          12 DO 830 I1=0,2
 1           DO 830 I2=0,2
 2           DO 830 I3=0,2
 3           DO 830 I4 = 0,2
 4           DO 830 I5=0,2
 5           DO 830 I6=0,2
 6           DO 830 I7=0,2
 7           DO 830 I8=0,2
 8           DO 830 I9=0,2
 9           DO 830 I10=0,2
10           DO 830 I11=0,2
11           I12 = J-I1-I2-I3-I4-I5-I6-I7-I8-I9-I10-I11
11           IF((I12.LT.0).OR.(I12.GT.2)) GO TO 831
11           CALL   SUB12
11       831 CONTINUE
11       830 CONTINUE
             GO TO 70
          14 DO 930 I1=0,2
 1           DO 930 I2=0,2
 2           DO 930 I3=0,2
 3           DO 930 I4 = 0,2
 4           DO 930 I5=0,2
 5           DO 930 I6=0,2
 6           DO 930 I7=0,2
 7           DO 930 I8=0,2
 8           DO 930 I9=0,2
 9           DO 930 I10=0,2
10           DO 930 I11=0,2
11           DO 930 I12=0,2
12           DO 930 I13=0,2
13           I14 = J-I1-I2-I3-I4-I5-I6-I7-I8-I9-I10-I11-I12-I13
13           IF((I14.LT.0).OR.(I14.GT.2)) GO TO 931
13           CALL   SUB14
13       931 CONTINUE
13       930 CONTINUE
             GO TO 70
        2000 WRITE(6,2001) PA
        2001 FORMAT(1X,F8.4)
             GO TO 5000
        2010 STOP
```

```
      END
```

```
      SUBROUTINE  SUB4
      COMMON F(200),DA,I1,I2,I3,I4,I5,I6,I7,I8,I9,I10,I11,I12,I13,
     1     I14,A,IS
      DOUBLE PRECISION A,X,Y,PROD
      II1=I1+1
      II2=I2+1
      II3=I3+1
      II4=I4+1
      DAA = F(II1)+F(II2)+F(II3)+F(II4)
      IF(DAA.GE.DA) GO TO 100
      RETURN
  100 PROD = 1.0
      IF(I1.EQ.IS) GO TO 152
      III = IS-1-I1
      DO 151 IFAC =0,III
1     X=IS-IFAC
1     Y=X-I1
1 151 PROD = PROD *X /Y
  152 IF(I2.EQ.IS) GO TO 154
      III2 = IS-1-I2
      DO 153 IFAC =0,III2
1     X = IS-IFAC
1     Y = X-I2
1 153 PROD = PROD *X /Y
  154 IF(I3.EQ.IS) GO TO 156
      III3 = IS-1-I3
      DO 155 IFAC =0,III3
1     X = IS-IFAC
1     Y = X-I3
1 155 PROD = PROD *X /Y
  156 IF (I4.EQ.IS) GO TO 158
      III4 = IS-1-I4
      DO 157 IFAC =0,III4
1     X = IS-IFAC
1     Y = X-I4
1 157 PROD = PROD *X /Y
  158 A=A+PROD
      RETURN
      END
```

```
      SUBROUTINE  SUB6
      COMMON F(200),DA,I1,I2,I3,I4,I5,I6,I7,I8,I9,I10,I11,I12,I13,
     1     I14,A,IS
      DOUBLE PRECISION A,X,Y,PROD
      II1=I1+1
```

```
          II2=I2+1
          II3=I3+1
          II4=I4+1
          II5=I5+1
          II6=I6+1
          DAA = F(II1) +F(II2)+F(II3)+F(II4)+F(II5)+F(II6)
          IF(DAA.GE.DA) GO TO 300
          RETURN
      300 PROD = 1.0
          IF(I1.EQ.IS) GO TO 352
          III = IS-1-I1
          DO 351 IFAC = 0,III
1         X = IS-IFAC
1         Y = X-I1
1     351 PROD = PROD *X /Y
      352 IF(I2.EQ.IS) GO TO 354
          III2 = IS-1-I2
          DO 353 IFAC = 0,III2
1         X = IS-IFAC
1         Y = X-I2
1     353 PROD = PROD *X /Y
      354 IF (I3.EQ.IS) GO TO 356
          III3 = IS-1-I3
          DO 355 IFAC=0,III3
1         X = IS-IFAC
1         Y = X - I3
1     355 PROD = PROD *X /Y
      356 IF(I4.EQ.IS) GO TO 358
          III4 = IS-1-I4
          DO 357 IFAC=0,III4
1         X = IS-IFAC
1         Y = X-I4
1     357 PROD = PROD *X /Y
      358 IF(I5.EQ.IS) GO TO 360
          III5 = IS-1-I5
          DO 359 IFAC=0,III5
1         X = IS-IFAC
1         Y =X-I5
1     359 PROD = PROD *X /Y
      360 IF(I6.EQ.IS) GO TO 362
          III6 =IS-1-I6
          DO 361 IFAC=0,III6
1         X = IS-IFAC
1         Y = X-I6
1     361 PROD = PROD *X /Y
      362 A=A+PROD
          RETURN
          END
```

```
          SUBROUTINE  SUB8
          DOUBLE PRECISION A,X,Y,PROD
          COMMON F(200),DA,I1,I2,I3,I4,I5,I6,I7,I8,I9,I10,I11,I12,I13,
        1      I14,A,IS
          II1=I1+1
```

```
        II2=I2+1
        II3=I3+1
        II4=I4+1
        II5=I5+1
        II6=I6+1
        II7=I7+1
        II8=I8+1
        DAA = F(II1)+F(II2)+F(II3)+F(II4)+F(II5)+F(II6)+F(II7)+F(II8)
        IF(DAA.GE.DA) GO TO 500
        RETURN
    500 PROD = 1.0
        IF(I1.EQ.IS) GO TO 562
        III=IS-1-I1
        DO 561 IFAC =0,III
1       X = IS-IFAC
1       Y = X-I1
1   561 PROD = PROD *X /Y
    562 IF(I2.EQ.IS) GO TO 564
        III2 = IS-1-I2
        DO 563 IFAC =0,III2
1       X = IS-IFAC
1       Y = X-I2
1   563 PROD = PROD *X /Y
    564 IF(I3.EQ.IS) GO TO 566
        III3 = IS-1-I3
        DO 565 IFAC=0,III3
1       X = IS-IFAC
1       Y = X-I3
1   565 PROD = PROD*X/Y
    566 IF(I4.EQ.IS) GO TO 568
        III4 = IS-1-I4
        DO 567 IFAC=0,III4
1       X = IS-IFAC
1       Y = X-I4
1   567 PROD = PROD *X/Y
    568 IF(I5.EQ.IS) GO TO 570
        III5 = IS-1-I5
        DO 569 IFAC = 0,III5
1       X = IS-IFAC
1       Y = X-I5
1   569 PROD = PROD *X/Y
    570 IF(I6.EQ.IS) GO TO 572
        III6 = IS-1-I6
        DO 571 IFAC = 0,III6
1       X=IS-IFAC
1       Y=X-I6
1   571 PROD = PROD*X/Y
    572 IF(I7.EQ.IS) GO TO 574
        III7 = IS-1-I7
        DO 573 IFAC=0,III7
1       X=IS-IFAC
1       Y=X-I7
1   573 PROD=PROD*X/Y
    574 IF(I8.EQ.IS) GO TO 576
        III8 = IS-1-I8
        DO 575 IFAC=0,III8
1       X=IS-IFAC
1       Y=X-I8
1   575 PROD = PROD *X /Y
    576 A = A+PROD
        RETURN
```

```
      END

      SUBROUTINE  SUB10
      COMMON F(200),DA,I1,I2,I3,I4,I5,I6,I7,I8,I9,I10,I11,I12,I13,
     1     I14,A,IS
      DOUBLE PRECISION A,X,Y,PROD
      II1=I1+1
      II2=I2+1
      II3=I3+1
      II4=I4+1
      II5=I5+1
      II6=I6+1
      II7=I7+1
      II8=I8+1
      II9=I9+1
      II10=I10+1
      DAA = F(II1)+F(II2)+F(II3)+F(II4)+F(II5)+F(II6)+F(II7)+F(II8)
     1      +F(II9)+F(II10)
      IF(DAA.GE.DA) GO TO 781
      RETURN
  781 PROD = 1.0
      IF(I1.EQ.IS) GO TO 762
      III = IS-1-I1
      DO 761 IFAC =0,III
1     X = IS-IFAC
1     Y = X-I1
1 761 PROD = PROD*X/Y
  762 IF(I2.EQ.IS) GO TO 764
      III2 = IS-1-I2
      DO 763 IFAC =0,III2
1     X = IS-IFAC
1     Y = X-I2
1 763 PROD = PROD*X/Y
  764 IF(I3.EQ.IS) GO TO 766
      III3 = IS-1-I3
      DO 765 IFAC =0,III3
1     X = IS-IFAC
1     Y = X-I3
1 765 PROD = PROD*X/Y
  766 IF(I4.EQ.IS) GO TO 768
      III4 = IS-1-I4
      DO 767 IFAC =0,III4
1     X = IS-IFAC
1     Y = X-I4
1 767 PROD = PROD*X/Y
  768 IF(I5.EQ.IS) GO TO 770
      III5 = IS-1-I5
      DO 769 IFAC=0,III5
1     X = IS-IFAC
1     Y = X-I5
1 769 PROD = PROD*X/Y
  770 IF(I6.EQ.IS) GO TO 772
      III6 = IS-1-I6
      DO 771 IFAC=0,III6
```

```
1          X = IS-IFAC
1          Y = X-I6
1     771  PROD = PROD*X/Y
      772  IF(I7.EQ.IS) GO TO 774
           III7 = IS-1-I7
           DO 773 IFAC=0,III7
1          X = IS - IFAC
1          Y = X-I7
1     773  PROD = PROD*X/Y
      774  IF(I8.EQ.IS) GO TO 776
           III8 = IS-1-I8
           DO 775 IFAC=0,III8
1          X = IS-IFAC
1          Y = X-I8
1     775  PROD = PROD*X/Y
      776  IF(I9.EQ.IS) GO TO 778
           III9 = IS-1-I9
           DO 777 IFAC = 0,III9
1          X = IS-IFAC
1          Y = X-I9
1     777  PROD = PROD*X/Y
      778  IF(I10.EQ.IS) GO TO 780
           III10 = IS-1-I10
           DO 779 IFAC=0,III10
1          X = IS-IFAC
1          Y = X-I10
1     779  PROD = PROD*X/Y
      780  A = A + PROD
           RETURN
           END



           SUBROUTINE  SUB12
           COMMON F(200),DA,I1,I2,I3,I4,I5,I6,I7,I8,I9,I10,I11,I12,I13,
         1      I14,A,IS
           DOUBLE PRECISION A,X,Y,PROD
           II1=I1+1
           II2=I2+1
           II3=I3+1
           II4=I4+1
           II5=I5+1
           II6=I6+1
           II7=I7+1
           II8=I8+1
           II9=I9+1
           II10=I10+1
           II11=I11+1
           II12=I12+1
           DAA = F(II1)+F(II2)+F(II3)+F(II4)+F(II5)+F(II6)+F(II7)+F(II8)
         1      +F(II9)+F(II10)+F(II11)+F(II12)
           IF(DAA.GE.DA) GO TO 2613
           RETURN
     2613  PROD = 1.0
           IF(I1.EQ.2) GO TO 2562
           III = 1-I1
```

```
                DO 2561 IFAC =0,III
1               X = 2-IFAC
1               Y = X-I1
1         2561  PROD = PROD*X/Y
          2562  IF(I2.EQ.2) GO TO 2564
                III2 = 1-I2
                DO 2563 IFAC =0,III2
1               X = 2-IFAC
1               Y = X-I2
1         2563  PROD = PROD*X/Y
          2564  IF(I3.EQ.2) GO TO 2566
                III3 = 1-I3
                DO 2565 IFAC =0,III3
1               X = 2-IFAC
1               Y = X-I3
1         2565  PROD = PROD*X/Y
          2566  IF(I4.EQ.2) GO TO 2568
                III4 = 1-I4
                DO 2567 IFAC =0,III4
1               X = 2-IFAC
1               Y = X-I4
1         2567  PROD = PROD*X/Y
          2568  IF(I5.EQ.2) GO TO 2570
                III5 = 1-I5
                DO 2569 IFAC=0,III5
1               X = 2-IFAC
1               Y = X-I5
1         2569  PROD = PROD*X/Y
          2570   IF(I6.EQ.2) GO TO 2572
                III6 = 1-I6
                DO 2571 IFAC=0,III6
1               X = 2-IFAC
1               Y = X-I6
1         2571  PROD = PROD*X/Y
          2572  IF(I7.EQ.2) GO TO 2574
                III7 = 1-I7
                DO 2573 IFAC=0,III7
1               X = 2 - IFAC
1               Y = X-I7
1         2573  PROD = PROD*X/Y
          2574  IF(I8.EQ.2) GO TO 2576
                III8 = 1-I8
                DO 2575 IFAC=0,III8
1               X = 2-IFAC
1               Y = X-I8
1         2575  PROD = PROD*X/Y
          2576  IF(I9.EQ.2) GO TO 2578
                III9 = 1-I9
                DO 2577 IFAC = 0,III9
1               X = 2-IFAC
1               Y = X-I9
1         2577  PROD = PROD*X/Y
          2578  IF(I10.EQ.2) GO TO 2580
                III10 = 1-I10
                DO 2579 IFAC=0,III10
1               X = 2-IFAC
1               Y = X-I10
1         2579  PROD = PROD*X/Y
          2580  IF(I11.EQ.2) GO TO 2582
                 III11 = 1 - I11
                DO 2581 IFAC = 0,III11
```

```
1             X = 2 - IFAC
1             Y = X - I11
1      2581   PROD = PROD * X / Y
       2582   IF(I12.EQ.2) GO TO 2612
              III12 = 1 - I12
              DO 2611 IFAC = 0,III12
1             X = 2 - IFAC
1             Y = X - I12
1      2611   PROD = PROD * X / Y
       2612   A = A + PROD
              RETURN
              END
```

```
              SUBROUTINE  SUB14
              COMMON F(200),DA,I1,I2,I3,I4,I5,I6,I7,I8,I9,I10,I11,I12,I13,
          1        I14,A,IS
              DOUBLE PRECISION A,X,Y,PROD
              II1=I1+1
              II2=I2+1
              II3=I3+1
              II4=I4+1
              II5=I5+1
              II6=I6+1
              II7=I7+1
              II8=I8+1
              II9=I9+1
              II10=I10+1
              II11=I11+1
              II12=I12+1
              II13=I13+1
              II14=I14+1
              DAA = F(II1)+F(II2)+F(II3)+F(II4)+F(II5)+F(II6)+F(II7)+F(II8)
          1        +F(II9)+F(II10)+F(II11)+F(II12)+F(II13)+F(II14)
              IF(DAA.GE.DA) GO TO 3613
              RETURN
       3613   PROD = 1.0
              IF(I1.EQ.2) GO TO 3562
              III = 1-I1
              DO 3561 IFAC =0,III
1             X = 2-IFAC
1             Y = X-I1
1      3561   PROD = PROD*X/Y
       3562   IF(I2.EQ.2) GO TO 3564
              III2 = 1-I2
              DO 3563 IFAC =0,III2
1             X = 2-IFAC
1             Y = X-I2
1      3563   PROD = PROD*X/Y
       3564   IF(I3.EQ.2) GO TO 3566
              III3 = 1-I3
              DO 3565 IFAC =0,III3
1             X = 2-IFAC
1             Y = X-I3
1      3565   PROD = PROD*X/Y
```

```
      3566 IF(I4.EQ.2) GO TO 3568
           III4 = 1-I4
           DO 3567 IFAC =0,III4
1          X = 2-IFAC
1          Y = X-I4
1     3567 PROD = PROD*X/Y
      3568 IF(I5.EQ.2) GO TO 3570
           III5 = 1-I5
           DO 3569 IFAC=0,III5
1          X = 2-IFAC
1          Y = X-I5
1     3569 PROD = PROD*X/Y
      3570  IF(I6.EQ.2) GO TO 3572
           III6 = 1-I6
           DO 3571 IFAC=0,III6
1          X = 2-IFAC
1          Y = X-I6
1     3571 PROD = PROD*X/Y
      3572 IF(I7.EQ.2) GO TO 3574
           III7 = 1-I7
           DO 3573 IFAC=0,III7
1          X = 2 - IFAC
1          Y = X-I7
1     3573 PROD = PROD*X/Y
      3574 IF(I8.EQ.2) GO TO 3576
           III8 = 1-I8
           DO 3575 IFAC=0,III8
1          X = 2-IFAC
1          Y = X-I8
1     3575 PROD = PROD*X/Y
      3576 IF(I9.EQ.2) GO TO 3578
           III9 = 1-I9
           DO 3577 IFAC = 0,III9
1          X = 2-IFAC
1          Y = X-I9
1     3577 PROD = PROD*X/Y
      3578 IF(I10.EQ.2) GO TO 3580
           III10 = 1-I10
           DO 3579 IFAC=0,III10
1          X = 2-IFAC
1          Y = X-I10
1     3579 PROD = PROD*X/Y
      3580 IF(I11.EQ.2) GO TO 3582
            III11 = 1 - I11
           DO 3581 IFAC = 0,III11
1          X = 2 - IFAC
1          Y = X - I11
1     3581 PROD = PROD * X / Y
      3582 IF(I12.EQ.2) GO TO 3584
           III12 = 1 - I12
           DO 3583 IFAC = 0,III12
1          X = 2 - IFAC
1          Y = X - I12
1     3583 PROD = PROD * X / Y
      3584 IF(I13.EQ.2) GO TO 3586
           III13 = 1-I13
           DO 3585 IFAC=0,III13
1          X = 2-IFAC
1          Y = X-I13
1     3585 PROD = PROD*X/Y
      3586 IF(I14.EQ.2) GO TO 3612
```

```
                III14 = 1 - I14
                DO 3611 IFAC = 0,III14
1               X = 2 - IFAC
1               Y = X - I14
1          3611 PROD = PROD * X / Y
           3612 A = A + PROD
                RETURN
                END

END FTN 3097 IBANK 450 DBANK 218 COMMON
```

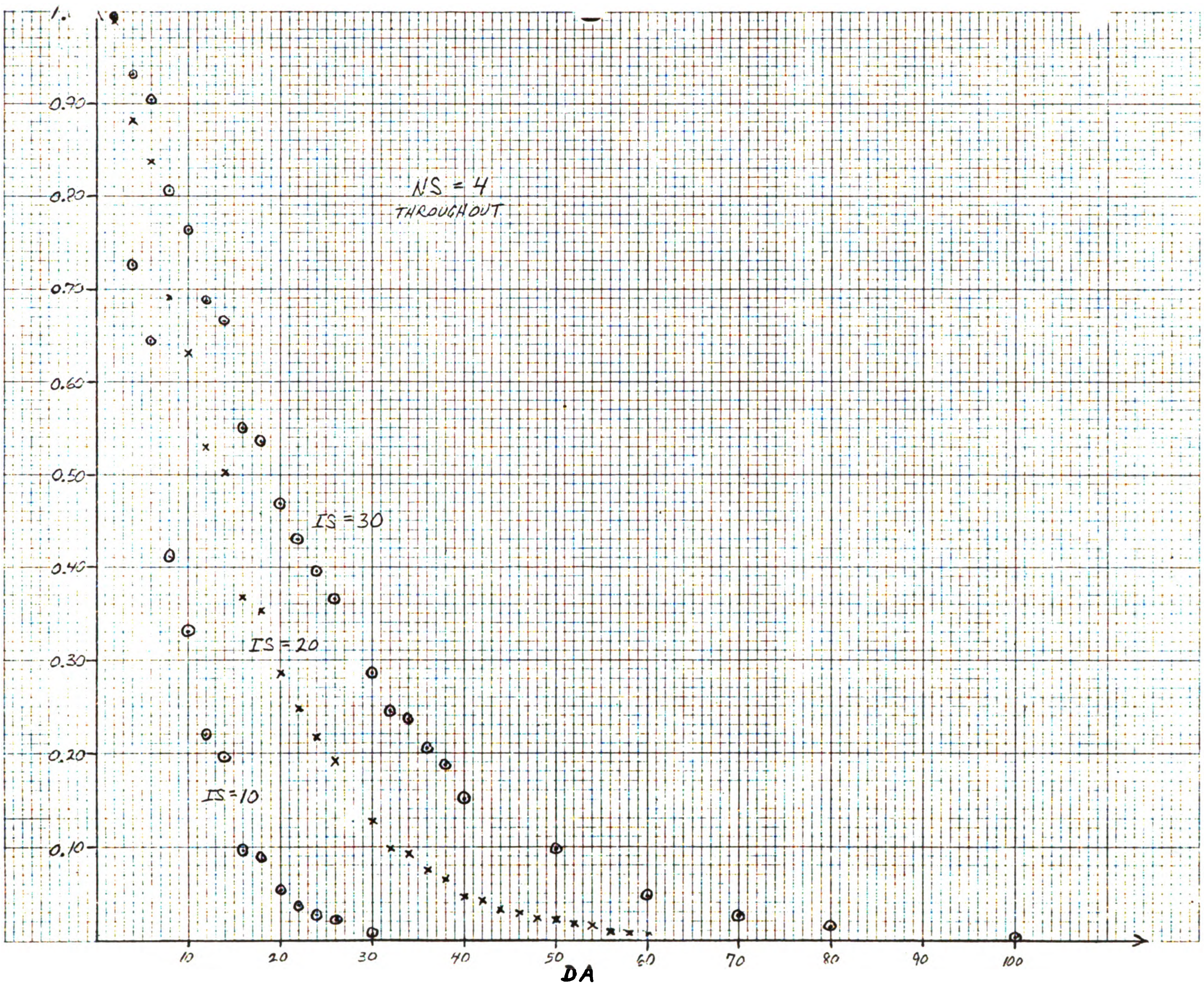
12345678901234567890123456789012345678901234567890123456789012345678901234567890123

```
@FTN,S KANOVA,WC,KEK
FTN 10R1   01/06/82-11:07(30,)
            WRITE (6,1100)
       1100 FORMAT(5X,'H0: EACH SAMPLE HAS AT LEAST  100XRA% CHANCE OF'
          1  ' HAVING'/9X,'ONE OBSERVATION INSIDE AND ONE'
          2  ' OBSERVATION OUTSIDE'/9X,'THE INTERQUARTILE'
          3  ' RANGE'//5X,'H1: EACH SAMPLE HAS AT LEAST  100XRB% CHANCE'
          4  ' OF HAVING'/9X,'BOTH OBSERVATIONS FALL TOGETHER'/9X,
          5  'EITHER INSIDE OR OUTSIDE'//////5X,'INPUTS ARE THE '
          6  'NUMBER OF SAMPLES'/5X,'THE NUMBER OF ZEROES'/5X,'RA AND '
          7  '  RB'//5X,'THE NUMBER OF ZEROES IS THE NUMBER OF SAMPLES'
          8  ' WHOSE TWO OBSERVATIONS'/5X, 'ARE FOUND TOGETHER'//5X,'THE'
          9  ' FORMAT IS       I3,1X,I3,1X,F4.2,1X,F4.2'/5X,'FOR '
          1  'NS,NZ,RA,RB'//5X,'IN THIS PROGRAM  NS  MUST BE EVEN '
          2  'AND GREATER THAN OR EQUAL TO  6 ')
       1000 READ (5,1,END = 2010) NS,NZ,RA,RB
          1 FORMAT (I3,1X,I3,1X,F4.2,1X,F4.2)
            I = NS/2
            J = NZ/2
            PA = 0
            PB = 0
            K = 0
          2 X = I
            Y = K
            W = 0.5*I
            IF(Y.GE.W) GO TO 100
          3 P = X
            P = P/(X - Y)
            T = 0
          5 T = T + 1
            P = P*(X - T)
            P = P/ (X - Y - T)
            L = X - Y - T
            IF (L.GT.1) GO TO 5
            IF (K.GT.J) GO TO 6
            PB = PB + P*((RB**K)*((1-RB)**(I-K)))
            K = K + 1
            IF(K.LE.J) GO TO 2
            K = K - 1
          6 PA = PA + P*(((1-RA)**K)*(RA**(I-K)))
            K = K + 1
            IF(K-I) 2,2,2000
        100 Y = X-Y
            GO TO 3
       2000 WRITE (6,2001) PA,PB
       2001 FORMAT (1X,F6.4,5X,F6.4)
            GO TO 1000
       2010 STOP
            END

END FTN 130 IBANK 216 DBANK

@BRKPT PRINT$
```

APPENDIX B

The graph here is of PA values for NS = 4 and IS = 10, 20, and 30. Note that DA = 28 is impossible, so a value of 26.01 entered into program WA or WA1 would yield the proper result for DA = 30.

PA



1.
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
NS = 4
THROUGHOUT
IS = 30
IS = 20
IS = 10
10
20
30
40
50
60
70
80
90
100
DA

APPENDIX C

R VALUES TO NORMALIZE DISTRIBUTION OF THE ALTERNATIVE HYPOTHESIS

NS	IS	W	R
4	20	0.40	0.500129
4	20	0.30	0.500553
4	20	0.25	0.500913
4	30	0.30	0.500367
4	14	0.25	0.501313
4	10	0.30	0.501122
4	10	0.25	0.501856
4	8	0.25	0.502339
3	60	0.40	0.500092
2	20	0.30	0.501122
2	20	0.25	0.501856
2	10	0.25	0.503840
2	40	0.40	0.500129

APPENDIX D

This is a comparison of the exact test results of program WA with the Chi-square results given by the program "CROSSTABS" found in the SPSS, and/or a program for Chi-square with three degrees of freedom found in Appendix E, written by Dr. J. V. Blowers. Dr. Blowers' program is applicable any time NS=4. However CROSSTABS is only applicable when calculating PA.

EXACT TEST

NS = 4

IS = 20

H_o:	10	10	10	10	
	10	10	10	10	
H_1:	5	5	15	15	
	15	15	5	5	(i.e., W = 0.25)
Observed:	7	8	12	13	
	13	12	8	7	

PA = 0.1906

PB = 0.0917

CHI-SQUARE

NS = 4

IS = 20

H_o:	10	10	10	10
	10	10	10	10
H_1:	5	5	15	15
	15	15	5	5
Observed:	7	8	12	13
	13	12	8	7

PA = 0.1577 (from "CROSSTABS")

PB = 0.0741 (from program by J. V. Blowers)

EXACT TEST

NS = 4

IS = 20

H_o:

10	10	10	10
10	10	10	10

H_1:

6	6	14	14
14	14	6	6

Observed:

8	8	12	12
12	12	8	8

PA = 0.3677

PB = 0.2827

CHI-SQUARE

NS = 4

IS = 20

H_o:

10	10	10	10
10	10	10	10

H_1:

6	6	14	14
14	14	6	6

Observed:

8	8	12	12
12	12	8	8

PA = 0.3618

PB = 0.2828

EXACT TEST

NS = 4

IS = 20

H_o:	10	10	10	10
	10	10	10	10
H_1:	8	8	12	12
	12	12	8	8
Observed:	9	9	11	11
	11	11	9	9

PA = 0.8817

PB = 0.8385

CHI-SQUARE

NS = 4

IS = 20

H_o:	10	10	10	10
	10	10	10	10
H_1:	8	8	12	12
	12	12	8	8
Observed:	9	9	11	11
	11	11	9	9

PA = 0.8495

PB = 0.8415

EXACT TEST					CHI-SQUARE				
NS = 4					NS = 4				
IS = 10					IS = 10				
H_o:	5	5	5	5	H_o:	5	5	5	5
	5	5	5	5		5	5	5	5
H_1:	2.5	2.5	7.5	7.5	H_1:	2.5	2.5	7.5	7.5
	7.5	7.5	2.5	2.5		7.5	7.5	2.5	2.5
Observed:	3	4	6	7	Observed:	3	4	6	7
	7	6	4	3		7	6	4	3
PA = 0.3305					PA = 0.2615				
PB = 0.6196					PB = 0.4459				

Observed:	4	4	6	6	Observed:	4	4	6	6
	6	6	4	4		6	6	4	4
PA = 0.7269					PA = 0.6594				
PB = 0.2912					PB = 0.1871				

EXACT TEST		CHI-SQUARE

EXACT TEST

NS = 4

IS = 30

H_o:

15	15	15	15
15	15	15	15

H_1:

9	9	21	21
21	21	9	9

Observed:

13	14	16	17
17	16	14	13

PA = 0.7626

PB = 0.0033

Observed:

11	11	19	19
19	19	11	11

PA = 0.0375

PB = 0.4600

CHI-SQUARE

NS = 4

IS = 30

H_o:

15	15	15	15
15	15	15	15

H_1:

9	9	21	21
21	21	9	9

Observed:

13	14	16	17
17	16	14	13

PA = 0.7212

PB = 0.0046

Observed:

11	11	19	19
19	19	11	11

PA = 0.0362

PB = 0.4682

APPENDIX E

Computer program for Chi-Square with three degrees of freedom (Blowers).

```
@FTN,S D1K.JVBCS3,KEK
FTN 10R1    01/05/82-11158(11,)
                    PROGRAM JVB
                    DIMENSION A(8),H(8)
              100   WRITE(6,1)
                    WRITE(6,44)
               44   FORMAT(3X,'AAAAAAABBBBBBBCCCCCCCDDDDDDDEEEEEEEFFFFFFF'
                1   'GGGGGGGHHHHHHH')
                1   FORMAT(1X,'INPUT DATA BY ROWS,F7.2 FORMAT')
                    READ(5,2)(A(I),I=1,8)
                    WRITE(6,3)
                3   FORMAT(1X,'INPUT HYPOTHESIS BY ROWS')
                    READ(5,2)(H(I),I=1,8)
                2   FORMAT(8F7.2)
                    SAM=0
                    DO 1000 I=1,8
             1000   SAM=SAM+(A(I)-H(I))*(A(I)-H(I))/H(I)
                    TH=CHIS3(SAM)
                    WRITE(6,10)SAM,1.-TH
               10   FORMAT(1X,'CHI-SQUARE=',F18.8,' SIGNIFICANCE = ',F13.5)
                    GOTO 100
                    END

                    FUNCTION CHIS3(XN)
                    H=XN/1000.
                    CHIS3=0.
                    DO 1 I=0,1000
                    IF(I.EQ.(I/2)*2)C=2.
                    IF(I.NE.(I/2)*2)C=4.
                    IF(I.EQ.0.OR.I.EQ.1000)C=1.
                    CHIS3=CHIS3+H/3*C*F(I*H)
                1   CONTINUE
                    RETURN
                    END

                    FUNCTION F(X)
                    F=SQRT(X/2./3.14159265358979)*EXP(-X/2)
                    RETURN
                    END

END FTN 138 IBANK 153 DBANK

@BRKPT PRINTS
```

APPENDIX F

This is a comparison of WB1 and WC program results designed to show when the null hypotheses in each case are equivalent. The last page of this appendix provides RA values that cause approximate equivalency for some NS and NZ (or DA) conditions. In projecting RA and RB values to use when NS values are too large to use program WB1, a high estimate will be best as that will result in a PA or PB value that is smaller than the result should be, and will help reject doubtful hypotheses.

WB1		WC
	NS= 6	
	DA = NZ = 2	
0.9307		0.9311 when RA = 0.41
	DA = NZ = 4	
0.4113		0.4104 when RA = 0.56
	DA = NZ = 6	
0.0216		0.0220 when RA = 0.72

WB1		WC
	NS = 8	
	DA = NZ = 2	
0.9801		0.9791 when RA = 0.38 (0.9813 when RA = 0.37)
	DA = NZ = 4	
0.7016		0.7023 when RA = 0.49
	DA = NZ = 6	
0.1795		0.1792 when RA = 0.60
	DA = NZ = 8	
0.0054		0.0053 when RA = 0.73

WB1		WC
	NS = 10	
	DA = NZ = 2	
0.9945		0.9947 when RA = 0.35
	DA = NZ = 4	
0.8698		0.8688 when RA = 0.45
	DA = NZ = 6	
0.4333		0.4253 when RA = 0.54 (0.4439 when RA = 0.53)
	DA = NZ = 8	
0.0696		0.0726 when RA = 0.62 (0.0660 when RA = 0.63)
	DA = NZ = 10	
0.0014		0.0014 when RA = 0.73

WB1		WC
	NS = 12	
	DA = NZ = 12	
0.9985		0.9985 when RA = 0.34
	DA = NZ = 4	
0.9485		0.9490 when RA = 0.42
	DA = NZ = 6	
0.6673		0.6748 when RA = 0.49 (0.6562 when RA = 0.50)
	DA = NZ = 8	
0.2300		0.2232 when RA = 0.57 (0.2390 when RA = 0.56)

		NS			
		6	8	10	12
	2	0.41	0.38-	0.35	0.34
NZ	4	0.56	0.49	0.45	0.42
or DA	6	0.72	0.60	0.54-	0.49+
	8		0.73	0.62+	0.57-
	10			0.73	E T C.

APPENDIX G

ACKNOWLEDGMENTS

Thanks are due to J. V. Blowers and others for helpful comments on computer programming and other areas and also to G. W. Barnard at White Sands Missile Range, and others, for their encouragement.

APPENDIX H

REFERENCES

H-1. Bradley, Distribution-Free Statistical Tests.

H-2. Fischer, Lankford and Knaub, "Plan for the Data Acquisition and Data Analysis of US Army Manpower Nonavailability Factors," Draft, Improved MACRIT Project, 19 August 1981, US Army Logistics Center, Fort Lee, Virginia.

H-3. Knaub, et.al., "Missile Round Sample Size Considerations for Test Planning and Reporting" and "Supplement," a methodology study for TECOM, 1980. (Note that the Wilcoxon Matched Pairs Signed Rank Test, more powerful than the Sign Test when appropriate, is not mentioned here, but is included in the SPSS.)

H-4. Lewis and Anderson, "Basic Planning Factors Used in Development of TOE MACRIT," Volume II, December 1975.

H-5. Siegel, Non-parametric Statistics for the Behavioral Sciences, pp. 159-160.

H-6. Southworth, unpublished notes at White Sands Missile Range.

H-7. Walpole and Myers, Probability and Statistics for Engineers and Scientists, pp. 365, 366, 474.

Addendum:

In Bradley, Distribution-Free Statistical Tests, pages 237-242, the Brown-Mood Multi-Sample Median Test turns out to be " . . . a generalization to the multi-sample (and multi-population) case of the two-sample Westenberg-Mood Median Test . . ." This is part of what has been suggested here although the interquartile range is of more interest in this study since there seems to be a greater selection of median oriented tests.*

Some modifications (corrections) can be made to the chi-square test to make it a better approximation. It has been stated (Bradley, page 239) that multiplying the X^2 test statistic by $(n-1)/n$ will improve the approximation (Mood) (where n is the combined sample size) and this appears to be true in the examples given here as applied to the null hypothesis, but not the alternative hypotheses. (Bradley's book does not specifically address such alternative hypotheses directly, but instead deals with asymptotic relative efficiencies (AREs).)

*In reference 4, it was noted that the Kruskal-Wallis Test indicated median K.P. times from questionnaires administered to various posts were probably identical when it was obvious that the dispersions were different. That is why dispersion was emphasized in this paper.

A DATA BASED RANDOM NUMBER GENERATOR FOR A MULTIVARIATE DISTRIBUTION*

James R. Thompson   Rice University
Malcolm S. Taylor   USA Ballistic Research Laboratory

ABSTRACT. Let X be a k-dimensional random variable serving as input for a system with output Y (not necessarily of dimension k). Given X, an outcome Y or a distribution of outcomes $G(Y|X)$ may be obtained either explicitly or implicitly. We consider here the situation in which we have a real world data set $\{X_j\}_{j=1}^n$ and a means of simulating an outcome Y. A method for empirical random number generation based on the sample of observations of the random variable X without estimating the underlying density is discussed.

INTRODUCTION. The manner of dealing with multivariate data depends upon the application at hand. For example, let us suppose that $\{X_j\}_{j=1}^n$ is a sample of size n of a k-dimensional random variable. We may be interested simply in estimating the mean μ. In such a case, we may complete our task by computing the sample mean $\bar{X}$. If we are interested in the interrelationships between the various vector components, we may find it desirable to compute the sample covariance matrix $\hat{\sigma}$.

At a greater level of complexity, we may be required to estimate the density of X nonparametrically [1,3]. Here, the representational difficulties are substantial--- particularly for $k > 2$, where our 3-dimensional intuitions are inadequate for graphing the density even if we knew it precisely on a discrete mesh. Indeed, it would appear that for increasing dimensionality, our estimation theoretic difficulties pale in comparison to those of representation.

*This research was supported in part by ARO Contract DAAG-29-82-K-0014 at Rice University.

Suppose we are given, for example, the task of estimating the density f at a point X_o in k-space, based on a sample of size n. The naive nearest neighbor estimator

$$\hat{f}(X_o) = \frac{p}{n} \frac{1}{V_k(X_o, d(X_o, p))},$$

where $d(X_o, p)$ is the Euclidean distance from X_o to the p^{th} nearest neighbor and $V_k(X_o, d(X_o, p))$ is the volume of the k-sphere centered at X_o with radius $d(X_o, p)$, is likely to be quite satisfactory. But a problem occurs when we are asked for a usable summary of the unknown density over the space of non-negligible mass. If we know the functional form of the density $f(X;\theta)$, then we have a relatively easy task--- the estimation of θ. But in the highly ubiquitous nonparametric situation, in which we do not know the functional form of f, we are not so fortunate. We might decide, for example, to tabulate $\hat{f}$ on a mesh of size 20 in each dimension. This would require 20^k pointwise estimations of f--- a tedious but manageable task. But how shall we scan this k-dimensional table to obtain a useful feel for the density? Other approaches, clearly are required. One of these is discussed in [2].

There are, happily, cases in which the density representational difficulties may be sidestepped when coping nonparametrically with data sets in higher dimensions. For example, let us suppose the k-dimensional random variable X is an input into a system with output Y (of whatever dimension). Given X, an outcome Y or a distribution of outcomes $G(Y|X)$ is obtained either explicitly or implicitly through an output data set. Let us suppose these outcomes fall into six categories: Very Good, Good, Fair, Poor, Very Bad, Catastrophically Bad. Suppose further that these sets are well-defined in the Y-space. We are given a real world data set $\{X_j\}_{j=1}^{n}$. We have a means of simulating an outcome Y given the input X. We wish to determine the probability of arriving in each of the six category sets.

One way to achieve this result might be, simply, to sample from the n data points $\{X_j\}_{j=1}^{n}$. In many cases this will prove quite satisfactory. But let us suppose that "Catastrophically Bad" happens for Y > 10,

$$\text{where } Y = 1/\sum_{i=1}^{4} x_i^2 \quad \text{with } X = (x_1,x_2,x_3,x_4).$$

Then, if the x_i's are (unbeknownst to us, but in actuality) independently distributed as N(0,1), the chance of a "Catastrophically Bad" event is .0012. Let us suppose the size (n) of our data set is 100. The chance of <u>none</u> of these observations being in the "Catastrophically Bad" region is .887. So, a simulation which used only the 100 data points would, with probability .887, give us the information that "Catastrophically Bad" occured with zero probability. We need to avoid this pitfall.

One procedure would be to estimate the density of X nonparametrically and then build a random number generator using the density. Such a scheme would run into the representational difficulties mentioned above. We can be much more efficient.

<u>THE ALGORITHM</u>. Let us consider the following situation: We have a random sample $\{X_j\}_{j=1}^{n}$ of size n from a multivariate distribution of dimension k, and we want to generate pseudorandom vectors from the underlying, but unknown, distribution that gave rise to the random sample. Since we do not know, and usually will never know, the form of this distribution, our attack

should be empirical. We shall endeavor to see to it that our pseudorandom vectors look very much like those in the original data set. In so doing, we will maintain the essential structural integrity of the problem.

We now direct our attention to the mechanics of the algorithm. After carrying out a rough rescaling to account for differing variances that may exist among the k variates, we select at random one of the n data points, say X_1, from the data base and then proceed to determine its m-1 nearest neighbors. The nearest neighbors are determined under the ordinary Euclidean metric and the value of m will depend upon the sample size n, the characteristics of the data, and can best be determined after perusal of the data. A conservative estimate would be to choose $m = n/20$.

The vectors $\{X_j\}_{j=1}^m$ are now coded about the sample mean $\bar{X} = 1/m \; \Sigma \; X_i$ to yield $\{X_j'\} = \{X_j - \bar{X}\}_{j=1}^m$, and an independent random sample of size m is generated from the uniform distribution $U(1/m - \sqrt{\frac{3(m-1)}{m^2}}, 1/m + \sqrt{\frac{3(m-1)}{m^2}})$.

Now the linear combination

$$X' = \sum_{\ell=1}^{m} u_\ell X_\ell'$$

is formed, where $\{u_\ell\}_{\ell=1}^m$ is the random sample from the $U(1/m - \sqrt{\cdot}, 1/m + \sqrt{\cdot}\,)$. Finally the translation

$$X = X' + \bar{X}$$

restores the relative magnitude, and X is a pseudorandom vector which we propose to be representative of the multivariate distribution that provided the $\{X_j\}_{j=1}^n$.

To obtain the next pseudorandom vector we randomly select another of the n data points and proceed as above.

We will now attempt to motivate the algorithm by considering the mathematics that suggests the mechanics that we have just outlined. Consider the distribution of X_1 and its m-1 nearest neighbors: $\{(x_{1\ell}, x_{2\ell}, \ldots, x_{k\ell})'\}_{\ell=1}^{m} = \{X_\ell\}_{\ell=1}^{m}$. Let us suppose that this "truncated set" of random observations has mean vector μ and covariance matrix σ. Let $\{u_\ell\}_{\ell=1}^{m}$ be an independent random sample from the uniform distribution $U(1/m - \sqrt{\cdot},\ 1/m + \sqrt{\cdot}\ )$. Then, $E(u_\ell) = 1/m$, $Var(u_\ell) = (m-1)/m^2$, and $Cov(u_i, u_j) = 0$, for $i \neq j$.

Forming the linear combination

$$Z = \sum_{\ell=1}^{m} u_\ell X_\ell$$

we have, for the r^{th} component $z_r = u_1 x_{r1} + u_2 x_{r2} + \ldots + u_m x_{rm}$, the following relations

$$E(z_r) = m \cdot 1/m \cdot \mu_r = \mu_r,$$

$$Var(z_r) = \sigma_r^2 + (m-1)/m \cdot \mu_r^2,$$

$$Cov(z_r, z_s) = \sigma_{rs} + (m-1)/m \cdot \mu_r \mu_s.$$

Clearly, if the mean vector of X was $\mu = (0,0, \ldots, 0)'$, then the mean vector and covariance matrix of Z would be identical to those of X. In the less idealized situation with which we are confronted, the translation to the sample mean of the nearest neighbor cloud should result in the pseudoobservation having very nearly the same mean and covariance structure as that of the (truncated) distribution of the points in the nearest neighbor cloud, a conjecture borne out in many actual cases that have been considered. For m moderately large, our algorithm essentially samples from n Gaussian distributions with the means and covariance matrices corresponding to those of the n m nearest neighbor clouds.

<u>EXAMPLES</u>. For a substantial test case, we considered a mixture of three bivariate normal distributions. The first (N_1) has mean vector $\binom{-1}{-2}$ and covariance matrix $\begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}$; the second ($N_2$) has mean vector $\binom{-2}{3}$ and covariance matrix $\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$; and the third ($N_3$) has mean vector $\binom{2}{3/2}$ and covariance matrix $\begin{pmatrix} 1 & 1/10 \\ 1/10 & 1 \end{pmatrix}$. The corresponding mixing scalars are $\alpha_1 = 1/2$, $\alpha_2 = 1/3$, and $\alpha_3 = 1/6$, respectively. Representative contours of equal density are illustrated in Figure 1. To establish a data base, a sample of eighty-five points was generated from this distribution via Monte Carlo simulation; a sample of eighty-five pseudorandom values was then produced by the algorithm, and the combined sample is shown in Figure 2.

Notice that the structure of the data is maintained in that the modes are preserved; the algorithm has not attempted to fill in gaps where gaps belong; the algorithm has, however, generated some points outside the boundary of the convex hull of the data base, all of which are desirable properties. These observations lend credence to the term "structural integrity" mentioned previously.

An application of the algorithm to a real world data set is summarized in Figures 3 and 4. In Figure 3, a two-dimensional marginal of a set of 973 four-dimensional behind armor debris measurements is portrayed; in Figure 4, 973 simulated data points produced by our procedure. Once again, the salient features of the data set are preserved.

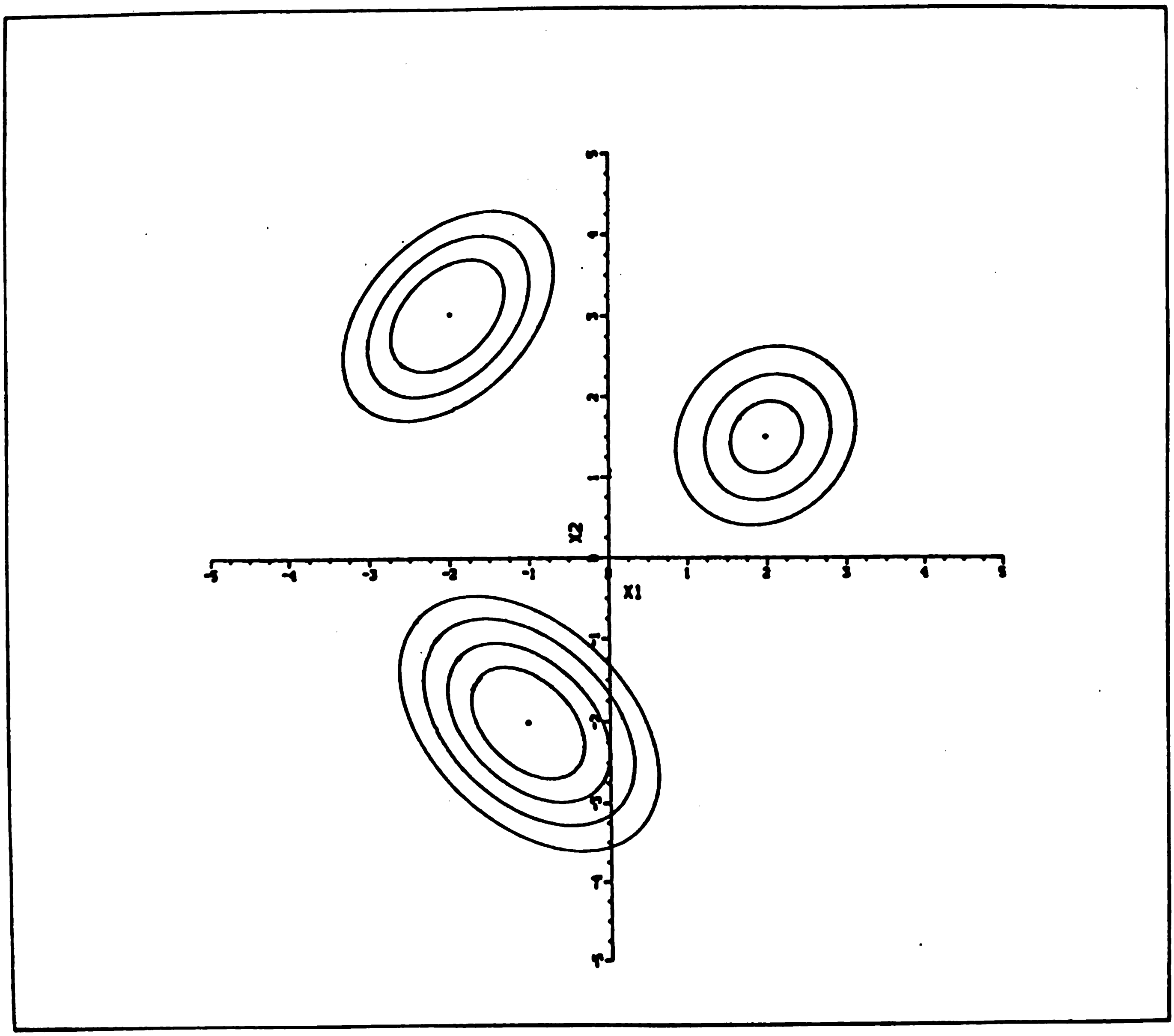


Fig. 1. Density contours for a mixture of three bivariate normal distributions.

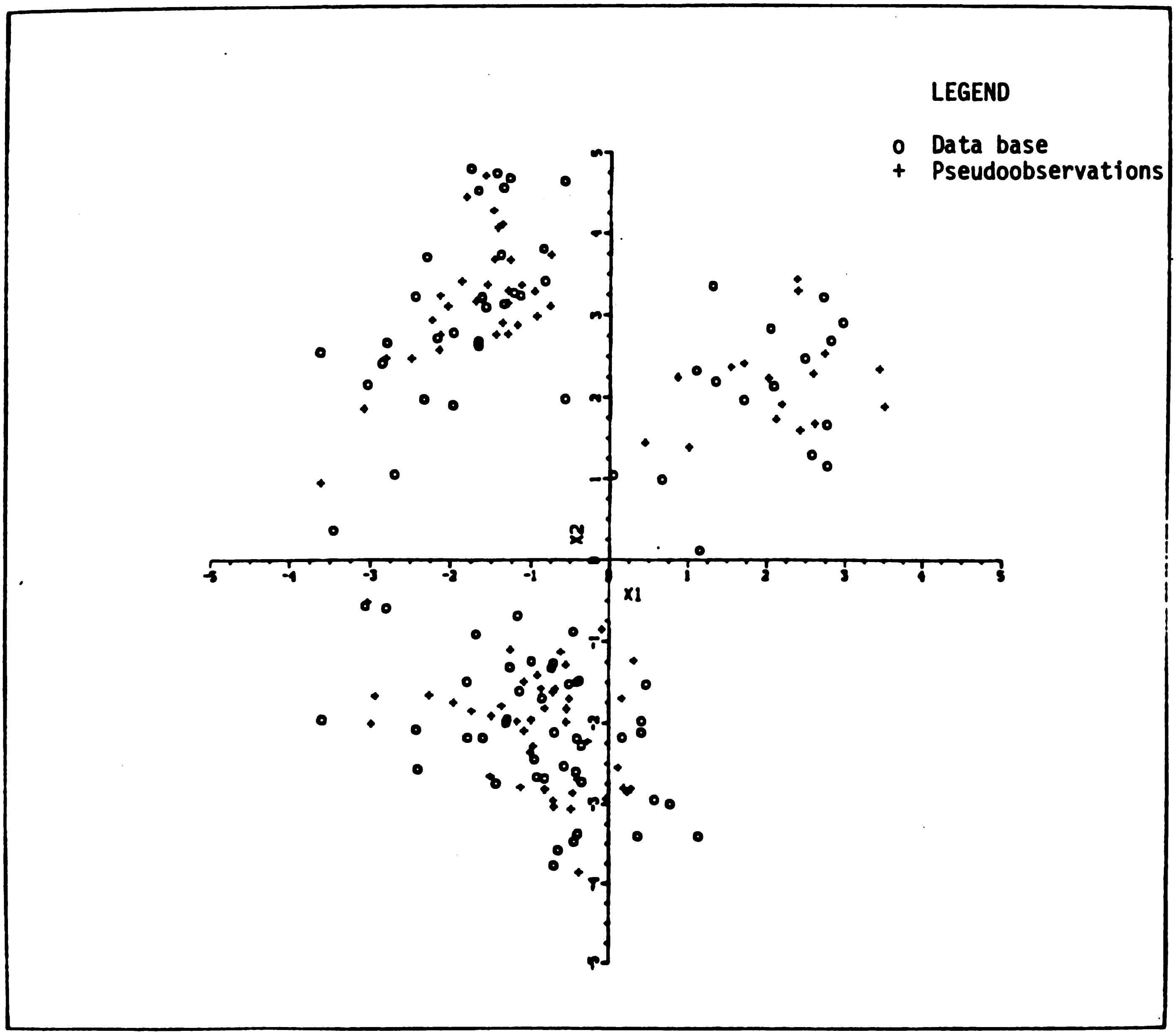


Fig. 2. Combined sample: Data base and Pseudoobservations.

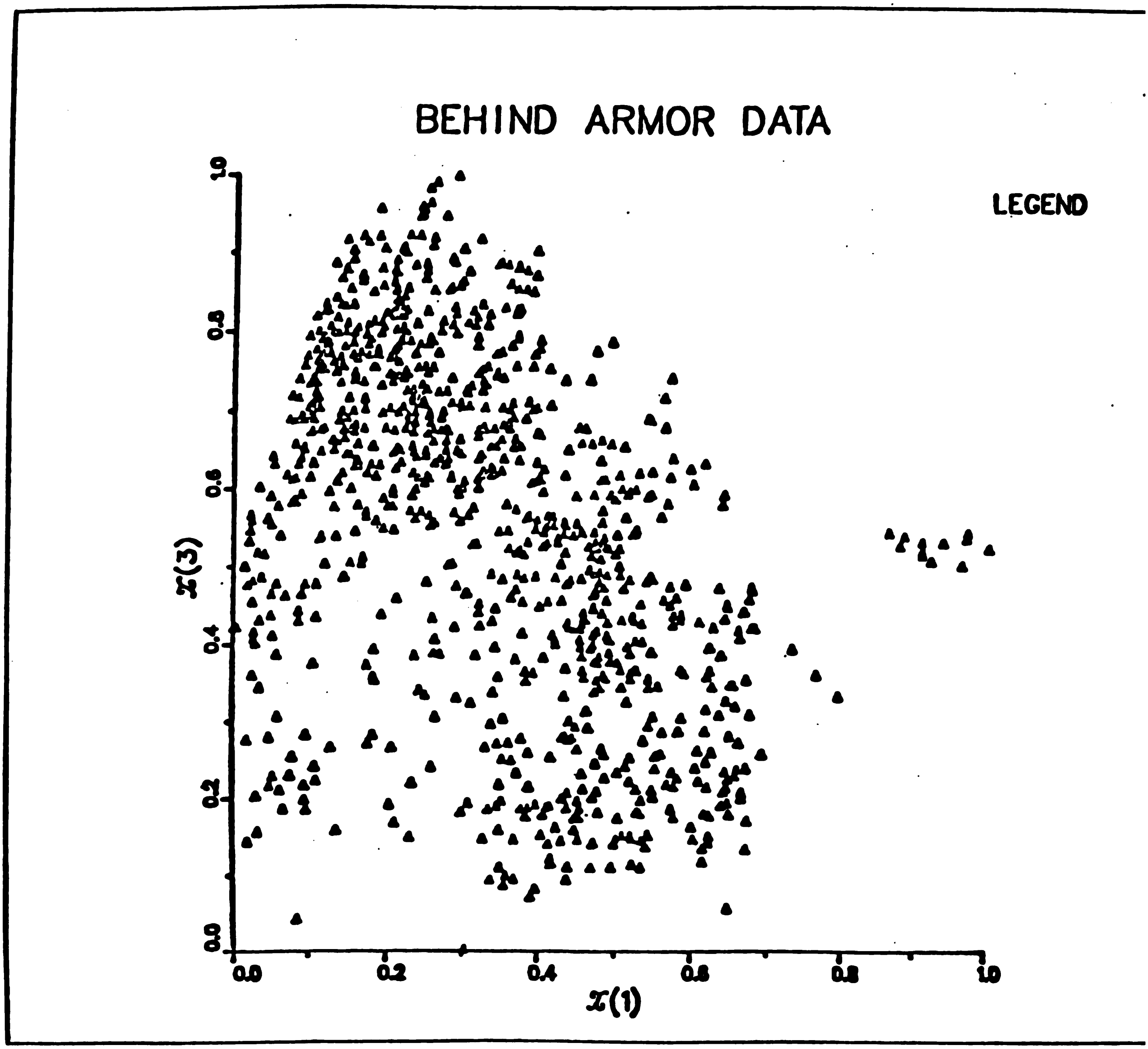


Fig. 3. Marginal data for 4-dimensional behind armor debris.

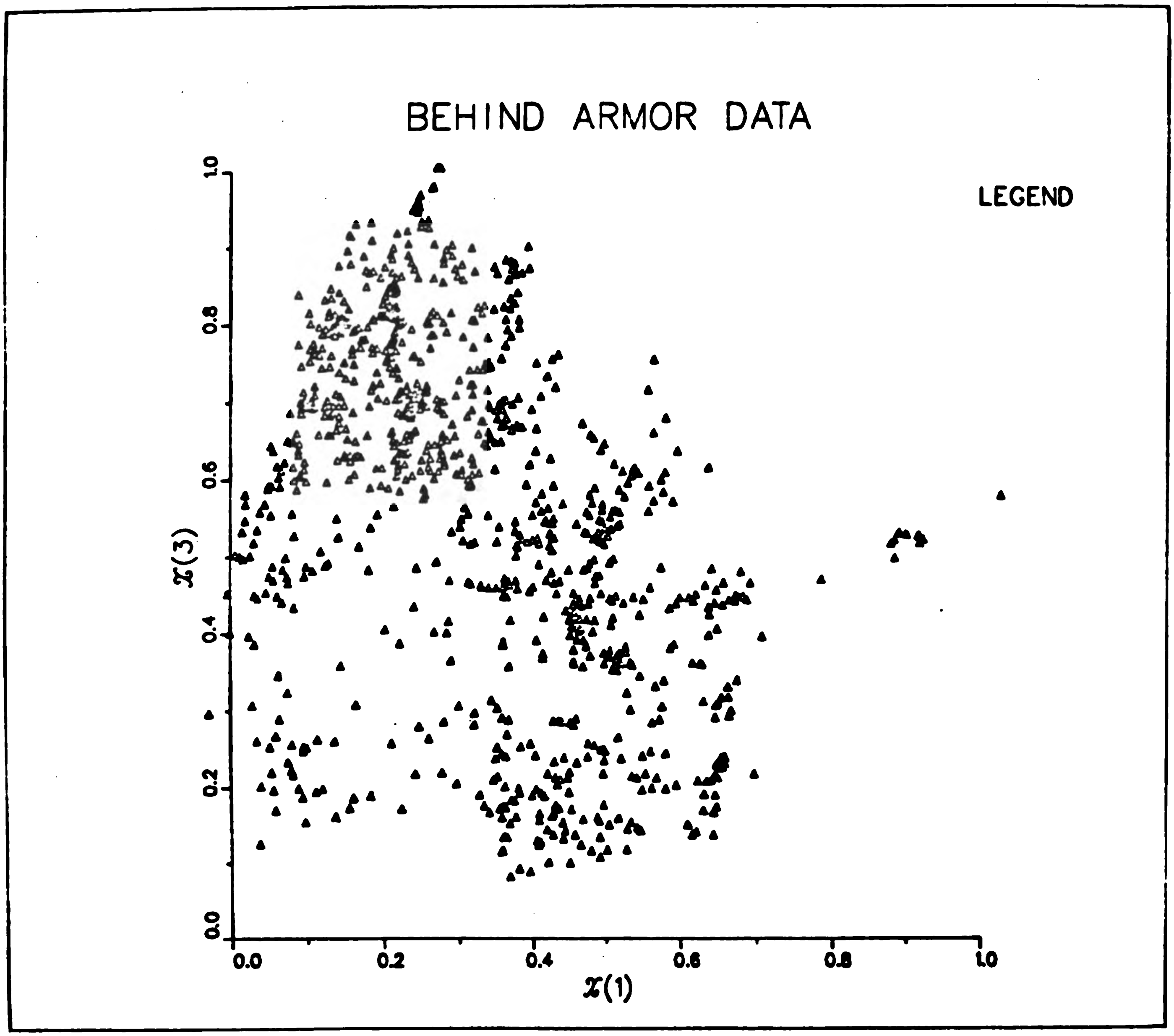


Fig. 4. Simulated behind armor debris.

CONCLUSIONS. We have demonstrated a means of empirical random number generation based on a sample of observations of a random variable X. No estimation of the underlying density is required. And, because of the local nature of the generation scheme, it is essentially free of assumptions on the underlying density of X. Naturally, any attempt to use this algorithm for generating bona fide new observations using the computer rather than producing real world data would be unwise. Rather, the algorithm operates somewhat like a smooth interpolator--- highly dependent on the quality of the data points on which it is based. It gives us a means of avoiding nonrobust conclusions due to "holes" in the data set at important points of the simulation model.

REFERENCES

1. Bean, Steven J. and Tsokos, Chris P. (1980). "Developments in nonparametric density estimation," International Statistical Review, v. 48, pp. 267-287.

2. Fwu, Chih-chy, Tapia, Richard A. and Thompson, James R. (1980). "The nonparametric estimation of probability densities in ballistics research," to appear in the Proceedings of the Twenty-Sixth Conference on the Design of Experiments in Army Research, Development and Testing.

3. Scott, David W., Tapia, Richard A. and Thompson, James R. (1978), "Multivariate density estimation by discrete maximum penalized likelihood methods," Graphical Representations of Multivariate Data, Peter C. Wang, ed., Academic Press, pp. 169-182.

IRREGULARITIES IN THE ERROR ANALYSIS OF A PIECE-WISE CONTINUOUS FUNCTION

PAUL H. THRASHER

QUALITY ASSURANCE OFFICE
US ARMY WHITE SANDS MISSILE RANGE
WHITE SANDS MISSILE RANGE, NM 88002

I. INTRODUCTION.

A. During a routine error propagation analysis, a continuous function with piece-wise continuous first derivatives was encountered. A couple of interesting effects were observed. These are described in this paper. The magnitude of these effects are quite small but they are quite distinctive. Their explanation is quite simple but they are observed so rarely that their explanations are not well known.

B. The source of the particular equation was nuclear radiation testing. In the equipment under study, the radiation source and detector are in air but may be separated by steel. The derivation of the function is outlined in Figure 1; I is the intensity, C is the number of photons emitted per second by the source, D is the distance that the radiation penetrates the steel, and μ is the attenuation coefficient of the steel. When the detector distance R is twice the source parameter A, the distance D is zero; thus, $R = 2A$ is the break-point between the two functions describing the intensity I. This piece-wise continuous function for I is plotted in Figure 2 when the product μB is taken to be unity.

II. ROUTINE ERROR ANALYSIS.

A. The routine random error analysis of the equation is outlined in Figure 3. Only the standard deviation in R is considered in the propagation of error to I because the random error instead of the total error is the item of interest and the bias error receives the contributions from the standard deviations in C, A, B, and μ. The result is presented graphically in Figure 4 with numerical values of one for μB, eight centimeters for A, and two millimeters for δR.

B. The discontinuity at the break-point (at R = 2A = 16 cm) results mathematically from the inequally of the absolute value of the slope of I versus R on the two sides of the break-point. An obvious question is: "Is there a standard error propagation procedure to eliminate this discontinuity and obtain a better estimate of the error at the break-point?" Perhaps, the answer is that the question isn't relevant since simulation can be used instead of error propagation.

III. SIMULATION.

A. The estimation of the error at the break-point can be done by simulation. All simulations presented in following Figures will be based on 500 random-normal samples for each plotted point of R. The standard deviation of R is retained at two millimeters.

B. The first effect noted in simulation is the appearance of horns at the break-point of graphs relating fractional standard deviations from simulations to the radius. Figure 5 presents a simulation that is too course in R values to tell how the standard deviation behaves as R passes through the break-point. Figure 6 presents a simulation of the fractional standard deviation of R values near the break-point. A horn appears at the break-point.

C. The second effect noted in simulation is the appearance of dual valleys in the graphs of fractional standard deviations. This occurs when the intensity function is adjusted to make the absolute value of dI/dR continuous at the break-point. As Figure 7 and Figure 8 indicate the attentuation coefficient can be set to make both $|dI/dR|$ and $\delta I/I$ from error propagation continuous. Even though Figure 9 shows that $\delta I/I$ is indeed continuous and Figure 10 shows that the simulation of σ_I/I generally follows $\delta I/I$, the detailed view of Figure 11 reveals an irregularity. At first glance, there appears to be a horn at the break-point, but closer examination shows that σ_I/I at the break-point has the expected value while σ_I/I is depressed on both sides of the break-point.

D. A deep, single valley is obtained on the simulation of σ_I/I when the sharp point at the break-point is rounded with a smoothing factor. Figure 12 shows the smoothing factor and parameters used. The resulting intensity curve is shown in Figure 13. The simulation of σ_I/I versus R yields a deep valley shown in Figure 14; Figure 15 shows that the bottom of this valley is flat.

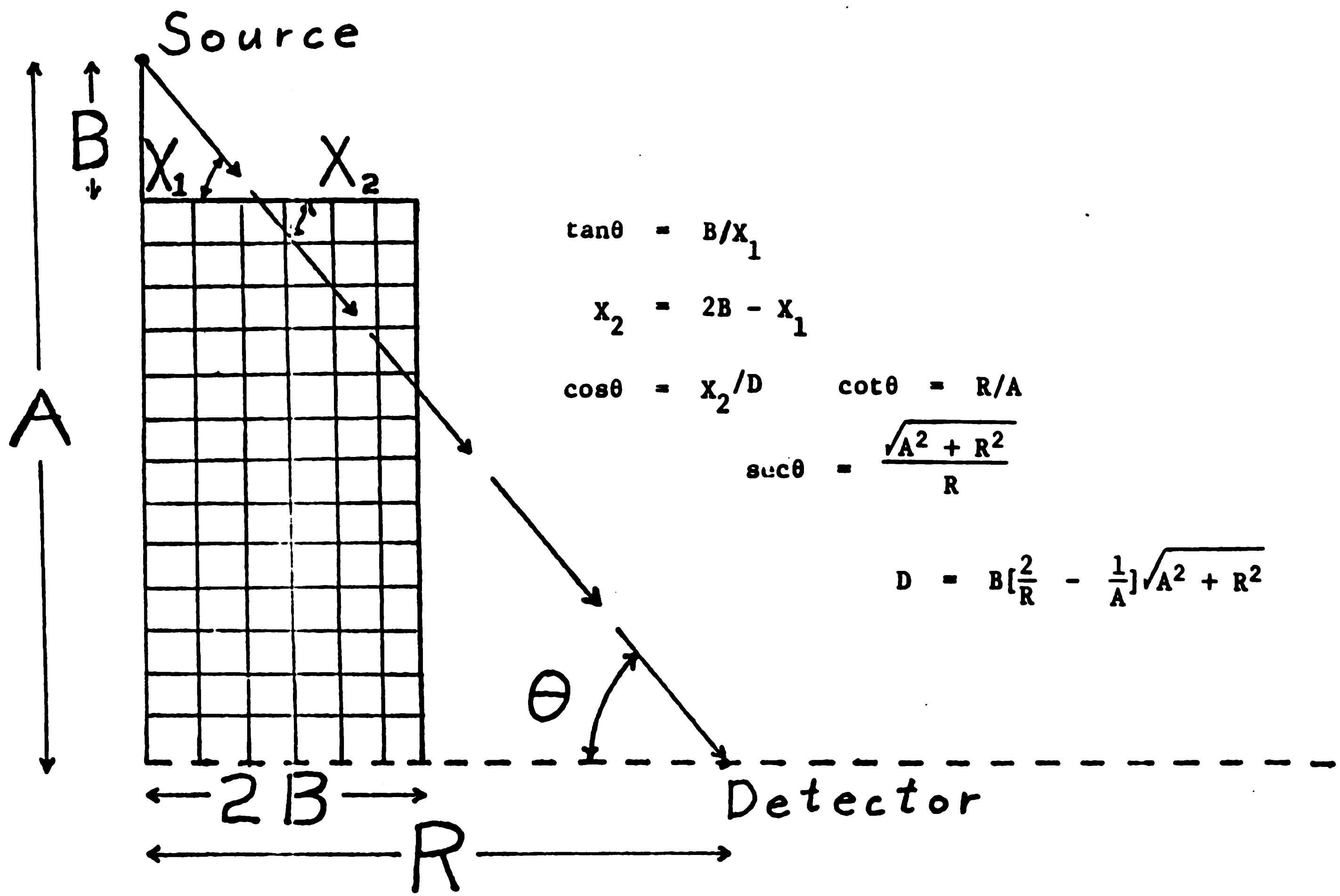
IV. EXPLANATION.

A. The origin of these irregularities is clearest when attention is focused on the valleys in σ_I/I. Probable error bars of magnitude $(2)(2/3)(\sigma_I) = 2\ 2/3$ millimeters may be used as a tool to aid in understanding the values of σ_I/I.

B. The easiest irregularity to describe is the deep valley of Figure 14 and Figure 15. Figure 16 shows an enlarged view of the rounded, symmetric point on I(R) versus R. The probable error bars on R are used to define the pseudo-error of I. This equation was used to calculate the curve on Figure 17. Figure 17 is in good qualitative agreement with Figure 14 and Figure 15.

C. Next, the irregularity of Figure 11 may be described. Figure 18 shows an enlarged view of the symmetric point on I(R) versus R. The same pseudo-error function that was used to describe the valley of Figure 14 and Figure 15 also generates the curve of Figure 19. This curve qualitatively describes the ω-shaped dual valleys of Figure 11.

D. Finally, the horn of Figure 6 is described by the same technique. The result, shown on Figure 20, not only reproduces the horn but shows that sides of horn are depressed instead of the point being raised.



$$\tan\theta = B/X_1$$

$$X_2 = 2B - X_1$$

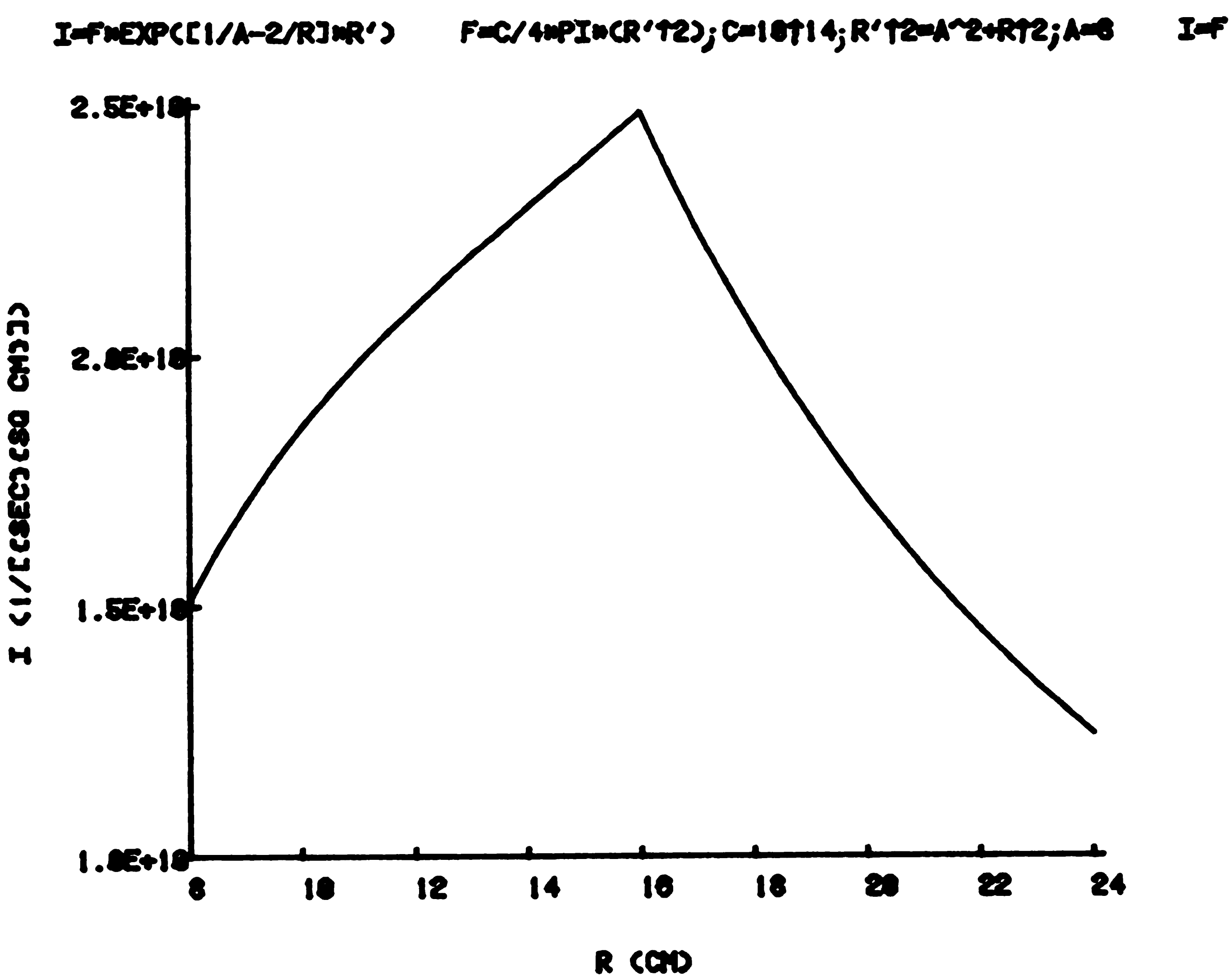
$$\cos\theta = X_2/D \qquad \cot\theta = R/A$$

$$\sec\theta = \frac{\sqrt{A^2 + R^2}}{R}$$

$$D = B\left[\frac{2}{R} - \frac{1}{A}\right]\sqrt{A^2 + R^2}$$

$$I = \begin{cases} \dfrac{C}{4\pi(A^2 + R^2)} \exp(-\mu D) & \text{if } R \leq 2A \\[2ex] \dfrac{C}{4\pi(A^2 + R^2)} & \text{if } R \geq 2A \end{cases}$$

Figure 1



I=F*EXP([1/A-2/R]*R')
F=C/4*PI*(R'↑2);C=10↑14;R'↑2=A^2+R↑2;A=6
I=F
2.5E+10
2.0E+10
1.5E+10
1.0E+10
I (1/[(SEC)(SQ CM)])
8
10
12
14
16
18
20
22
24
R (CM)

Figure 2

$$I = \begin{cases} \dfrac{C}{4\pi(A^2 + R^2)} \exp(-\mu B[\frac{2}{R} - \frac{1}{A}]\sqrt{A^2 + R^2}) & \text{if } R \leq 2A \\ \dfrac{C}{4\pi(A^2 + R^2)} & \text{if } R \geq 2A \end{cases}$$

$$\frac{\delta I}{I} = \frac{1}{I}\sqrt{\sum_{i=1}^{N}[\frac{\partial I}{\partial x_i}\delta x_i]^2} = \frac{1}{I}\left|\frac{\partial I}{\partial R}\right|\delta R$$

$$\frac{\delta I}{I} = \begin{cases} \delta R\left|\dfrac{2R}{A^2 + R^2} + \dfrac{\mu BR[\frac{2}{R} - \frac{1}{A}]}{\sqrt{A^2 + R^2}} - \dfrac{2\mu B\sqrt{A^2 + R^2}}{R^2}\right| & \text{if } R \leq 2A \\ \delta R\dfrac{2R}{A^2 + R^2} & \text{if } R \geq 2A \end{cases}$$

Figure 3

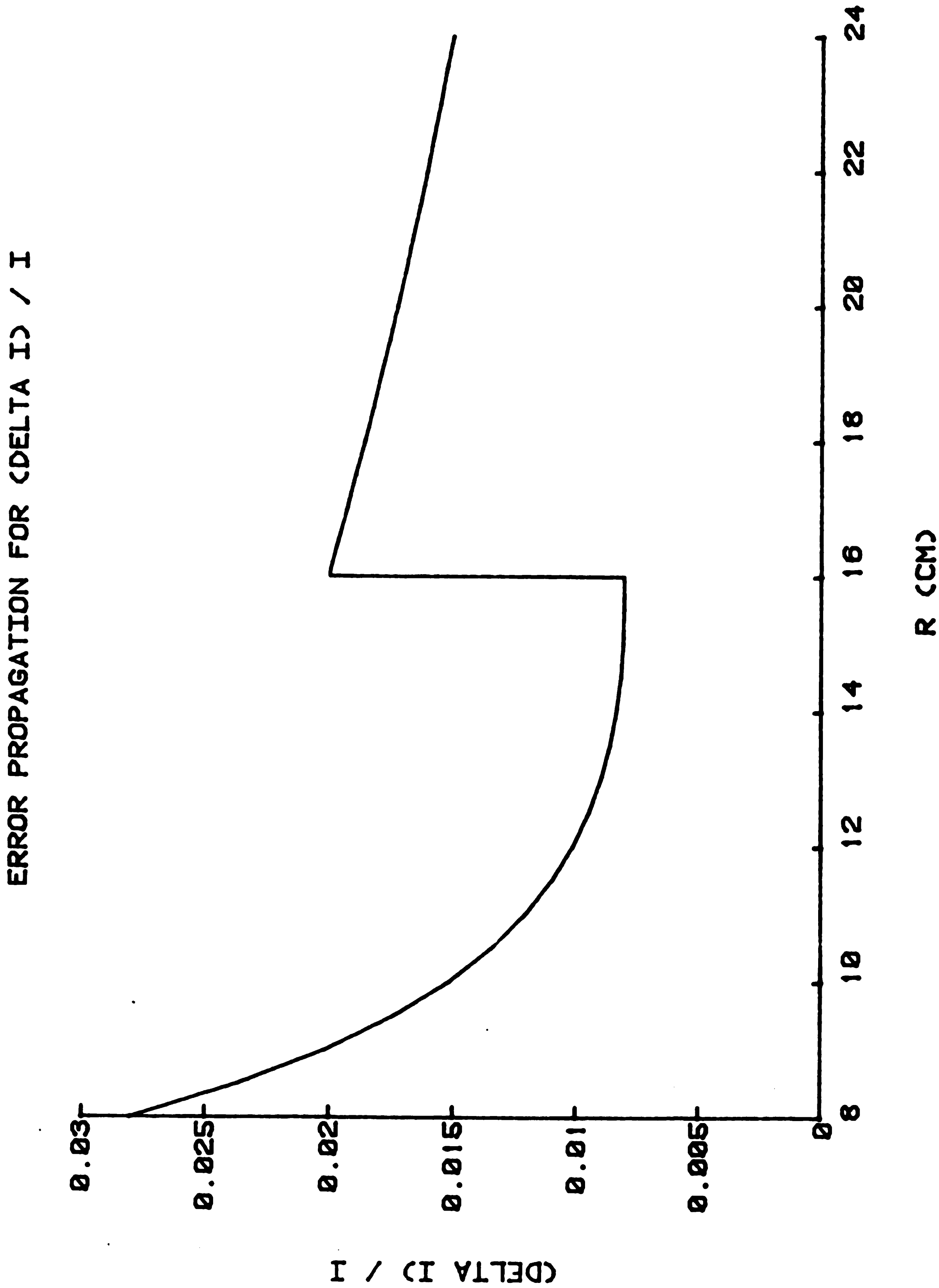

ERROR PROPAGATION FOR (DELTA I) / I
(DELTA I) / I
0
0.005
0.01
0.015
0.02
0.025
0.03
R (CM)
8
10
12
14
16
18
20
22
24

Figure 4

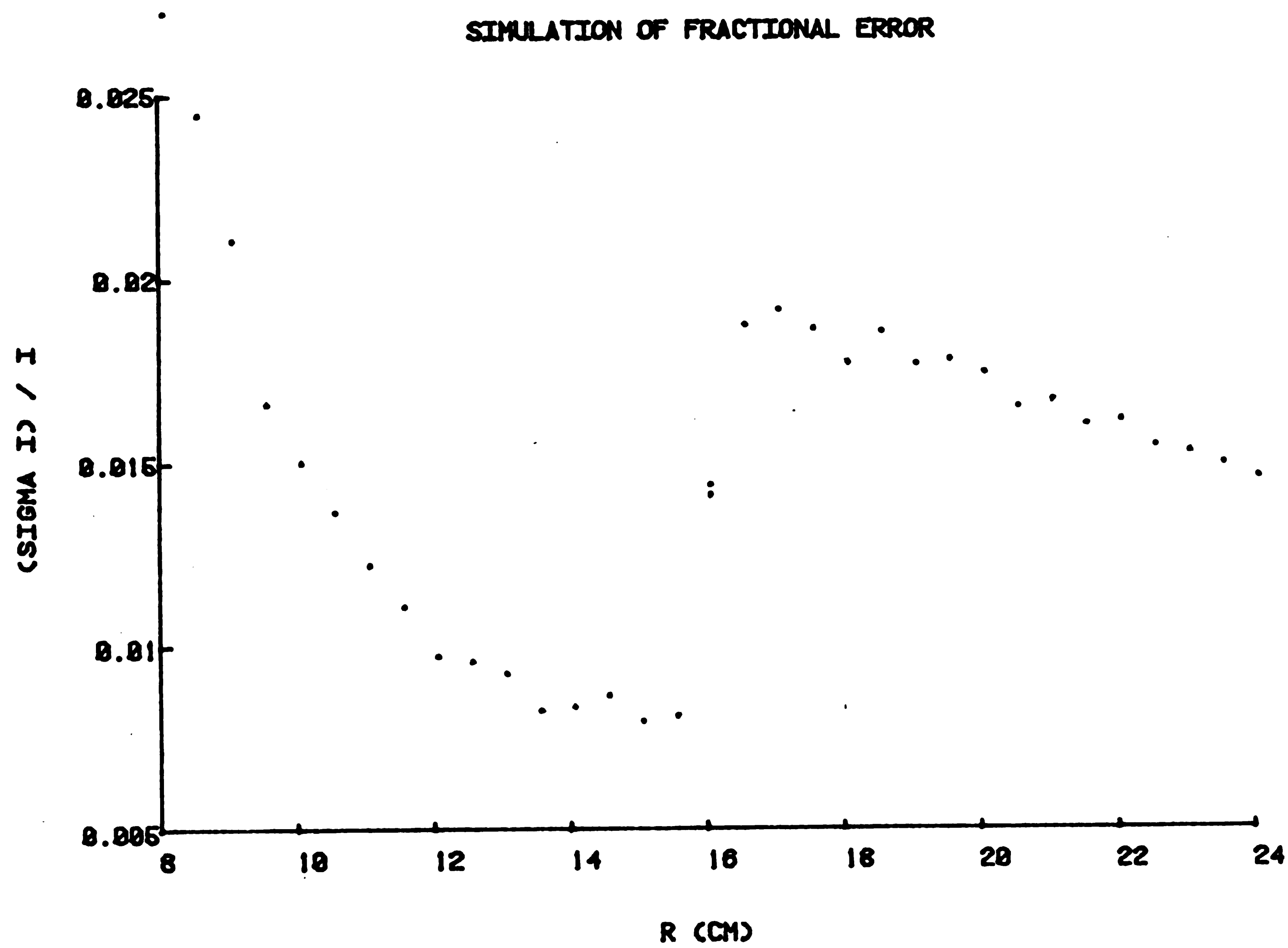

SIMULATION OF FRACTIONAL ERROR
(SIGMA I) / I
0.025
0.02
0.015
0.01
0.005
8
10
12
14
16
18
20
22
24
R (CM)

Figure 5

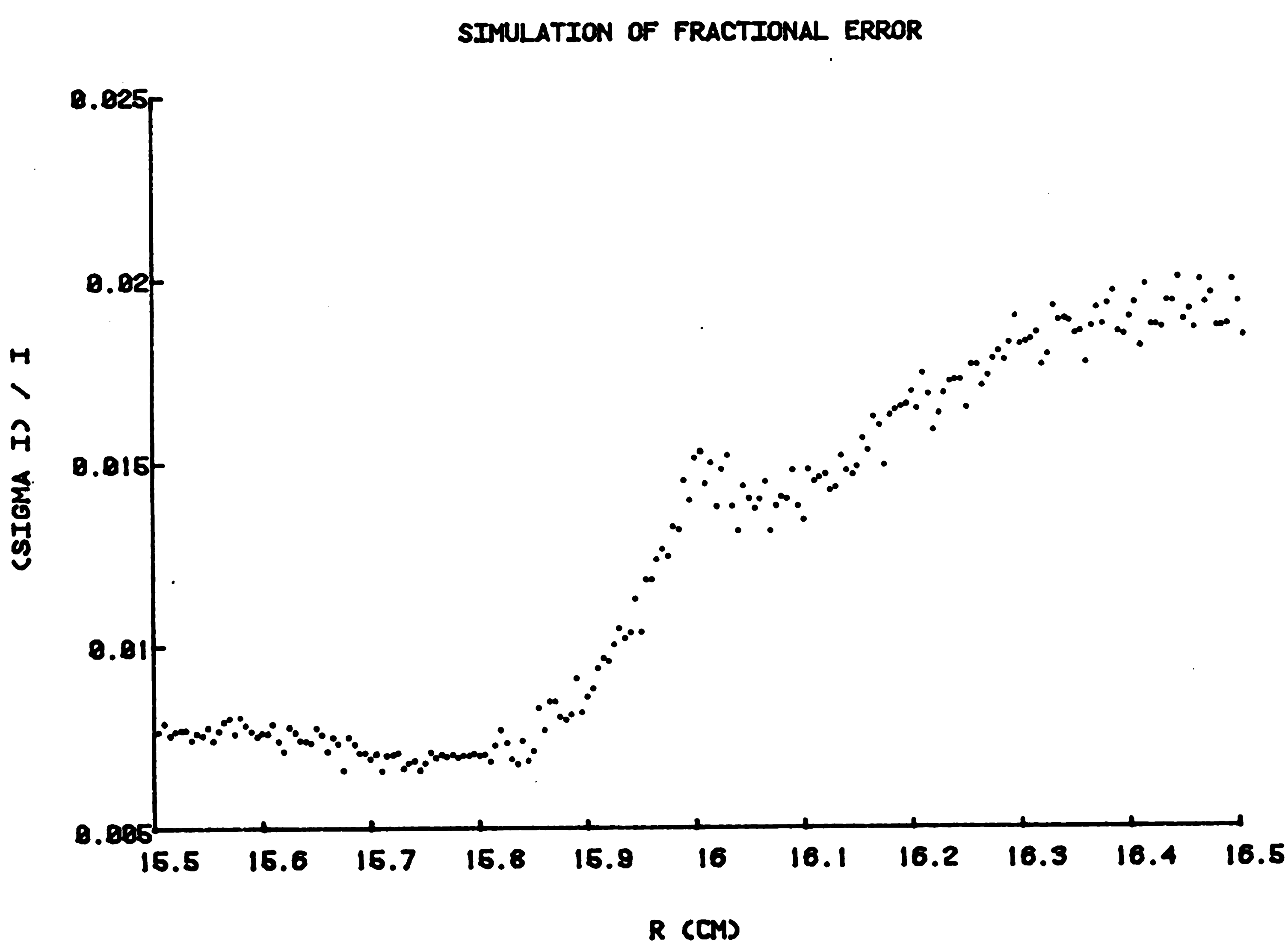

SIMULATION OF FRACTIONAL ERROR
(SIGMA I) / I
0.025
0.02
0.015
0.01
0.005
15.5
15.6
15.7
15.8
15.9
16
16.1
16.2
16.3
16.4
16.5
R (CM)

Figure 6

$$\lim_{\substack{R \to 2A \\ \text{from left}}} \left| \frac{\partial I}{\partial R} \right| = \lim_{\substack{R \to 2A \\ \text{from right}}} \left| \frac{\partial I}{\partial R} \right|$$

$$\mu B = \frac{16}{5\sqrt{5}}$$

$$\lim_{\substack{R \to 2A \\ \text{from left}}} \frac{\delta I}{I} = \lim_{\substack{R \to 2A \\ \text{from right}}} \frac{\delta I}{I}$$

Figure 7

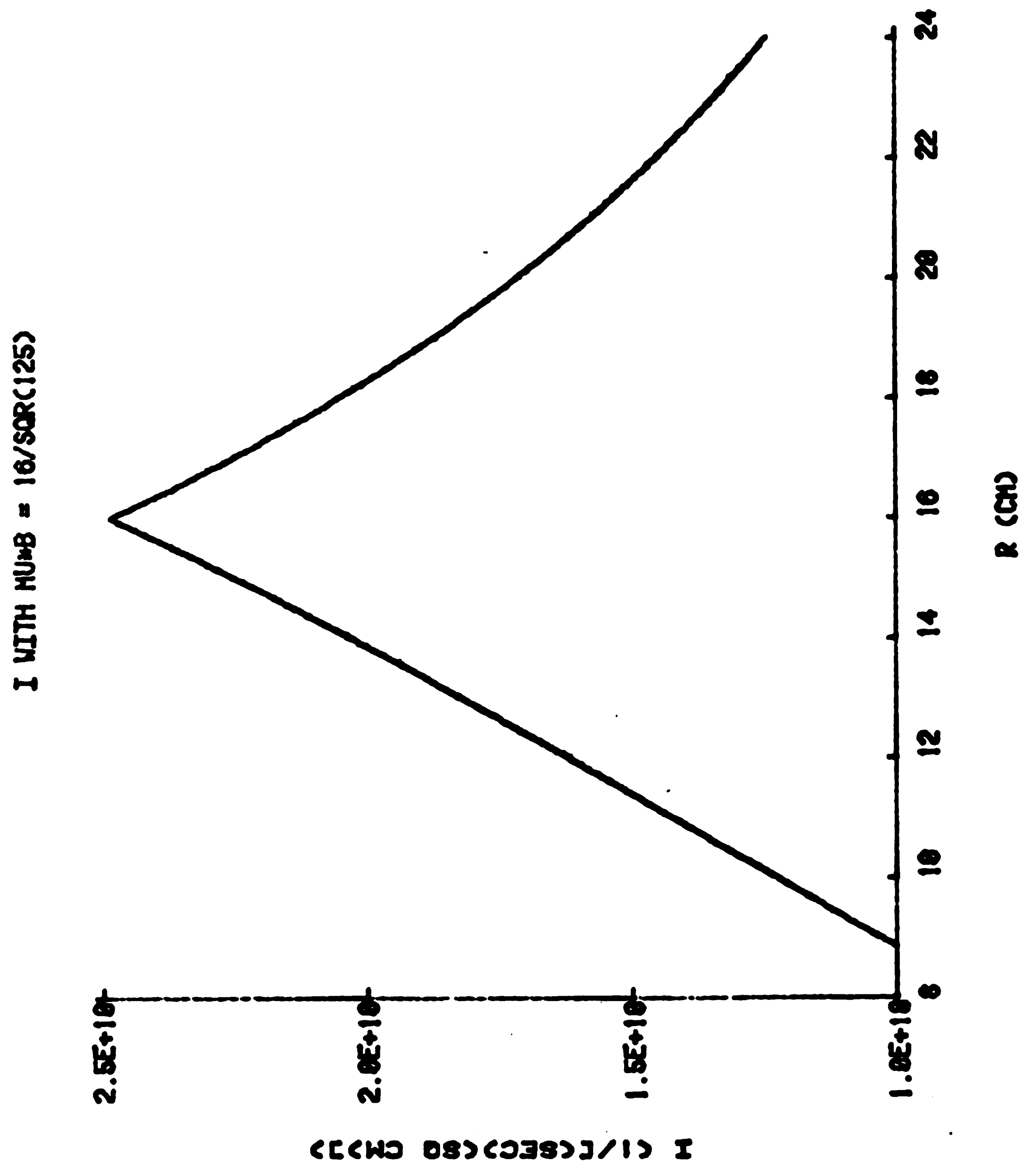

I WITH MU*B = 16/SQR(125)
2.5E+10
2.0E+10
1.5E+10
1.0E+10
I (1/[(SEC)(SQ CM)])
8
10
12
14
16
18
20
22
24
R (CM)

Figure 8

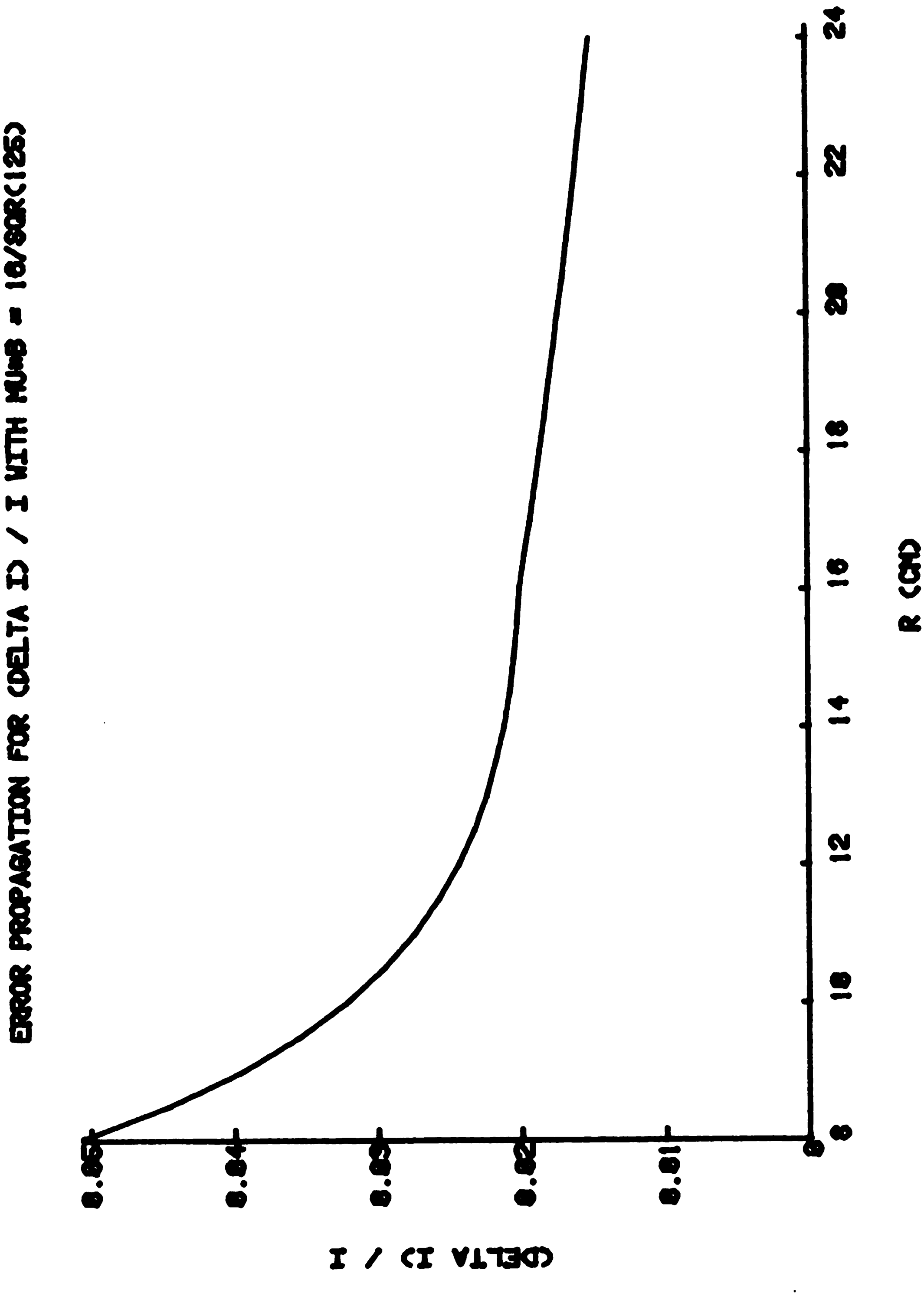

ERROR PROPAGATION FOR (DELTA I) / I WITH MU*B = 10/SQR(125)
(DELTA I) / I
0.05
0.04
0.03
0.02
0.01
0
0
10
12
14
16
18
20
22
24
R (CM)

Figure 9

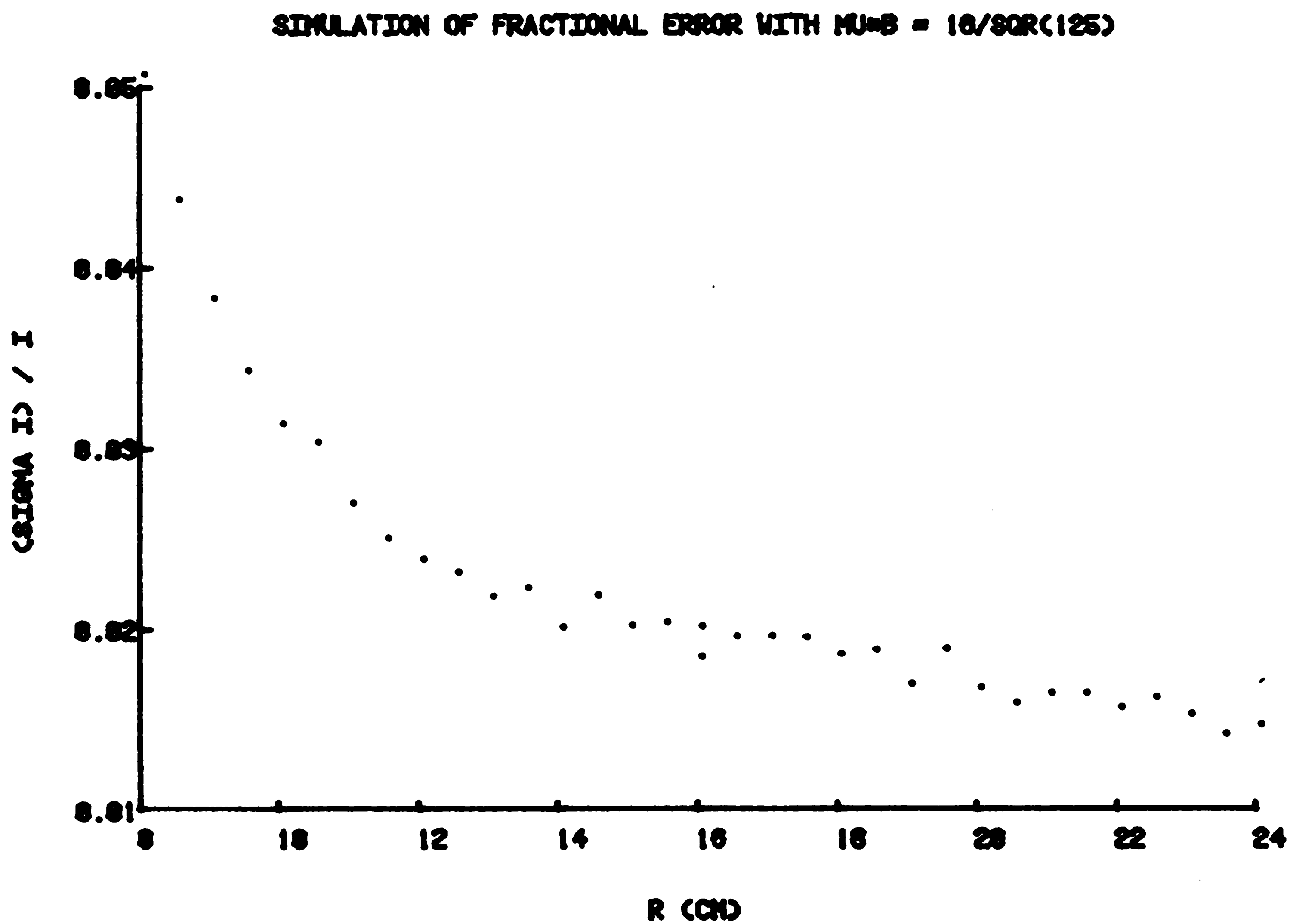

SIMULATION OF FRACTIONAL ERROR WITH MU*B = 10/SQR(125)
0.05
0.04
0.03
0.02
0.01
(SIGMA I) / I
8
10
12
14
16
18
20
22
24
R (CM)

Figure 10

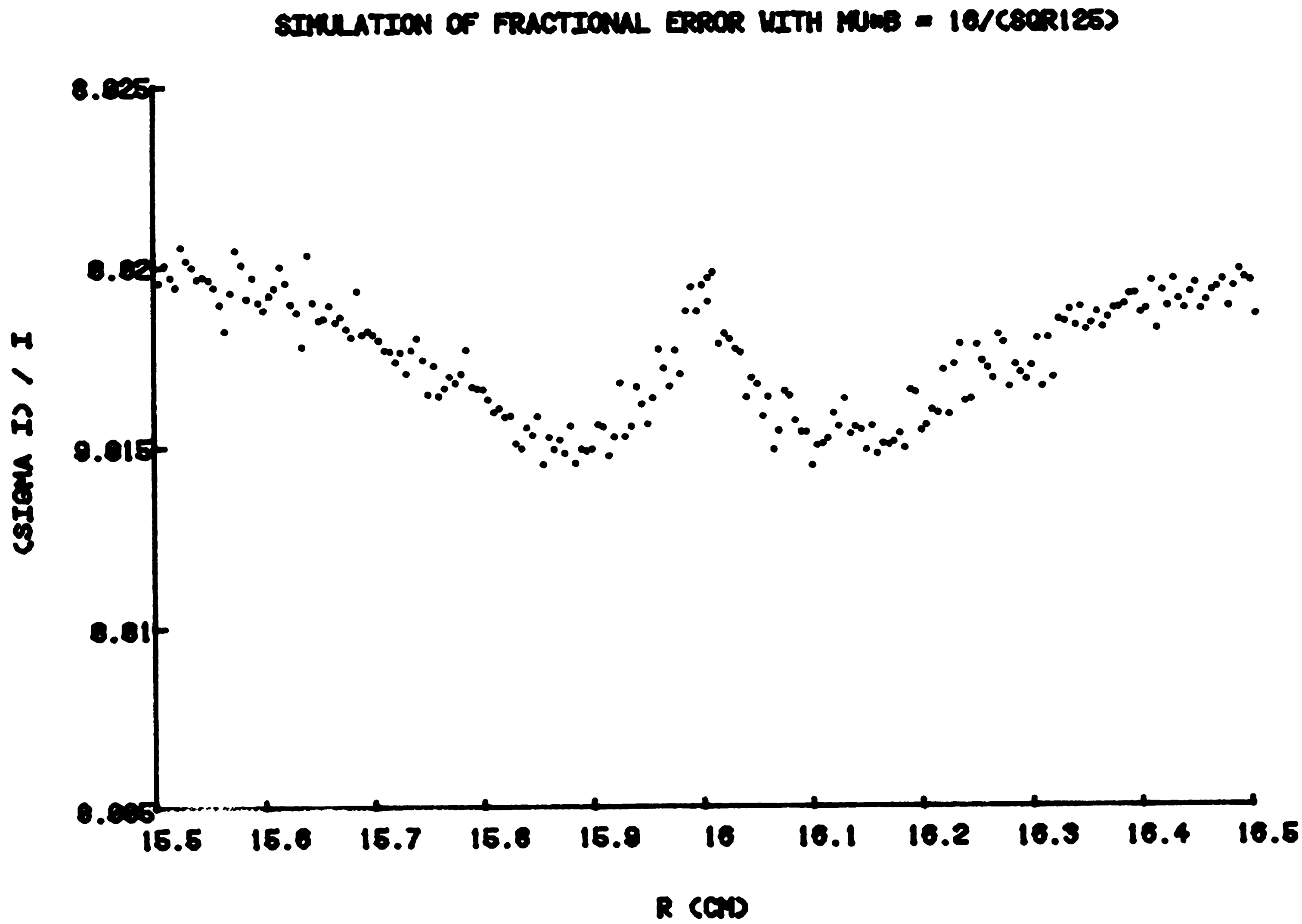

SIMULATION OF FRACTIONAL ERROR WITH MU*B = 16/(SQR125)
0.025
0.02
0.015
0.01
0.005
(SIGMA I) / I
15.5
15.6
15.7
15.8
15.9
16
16.1
16.2
16.3
16.4
16.5
R (CM)

Figure 11

SMOOTHING FACTOR

$$S(R) = \begin{cases} [1 - N_1 \exp\{-P_1(2A - R)\}]^{M_1} & \text{if } R \leq 2A \\ [1 - N_2 \exp\{-P_2(R - 2A)\}]^{M_2} & \text{if } R \geq 2A \end{cases}$$

$$I(R) = \begin{cases} \frac{C}{4\pi(A^2 + R^2)} \exp\{-\mu B(\frac{2}{R} - \frac{1}{A})\sqrt{A^2 + R^2}\} & \text{if } R \leq 2A \\ \frac{C}{4\pi(A^2 + R^2)} & \text{if } R \geq 2A \end{cases}$$

$$I_S \equiv [I][S]$$

$$\lim_{R \to 2A \text{ left}} I_S = \lim_{R \to 2A \text{ right}} I_S$$

$$0 = \lim_{R \to 2A \text{ left}} [\frac{dI_S}{dR}] \qquad 0 = \lim_{R \to 2A \text{ right}} [\frac{dI_S}{dR}]$$

$$M_2 = M_1 \ln(1 - N_1)/\ln(1 - N_2)$$

$$P_1 = \frac{\mu B5\sqrt{5} - 8}{(10)AM_1}(\frac{1}{N_1} - 1) \qquad P_2 = \frac{4}{5AM_2}(\frac{1}{N_2} - 1)$$

If $N_1 = 1/2$, $N_2 = 1/2$ and $M_1 = 0.1$, then $M_2 = 0.1$, $P_1 = 1$ and $P_2 = 1$

Figure 12

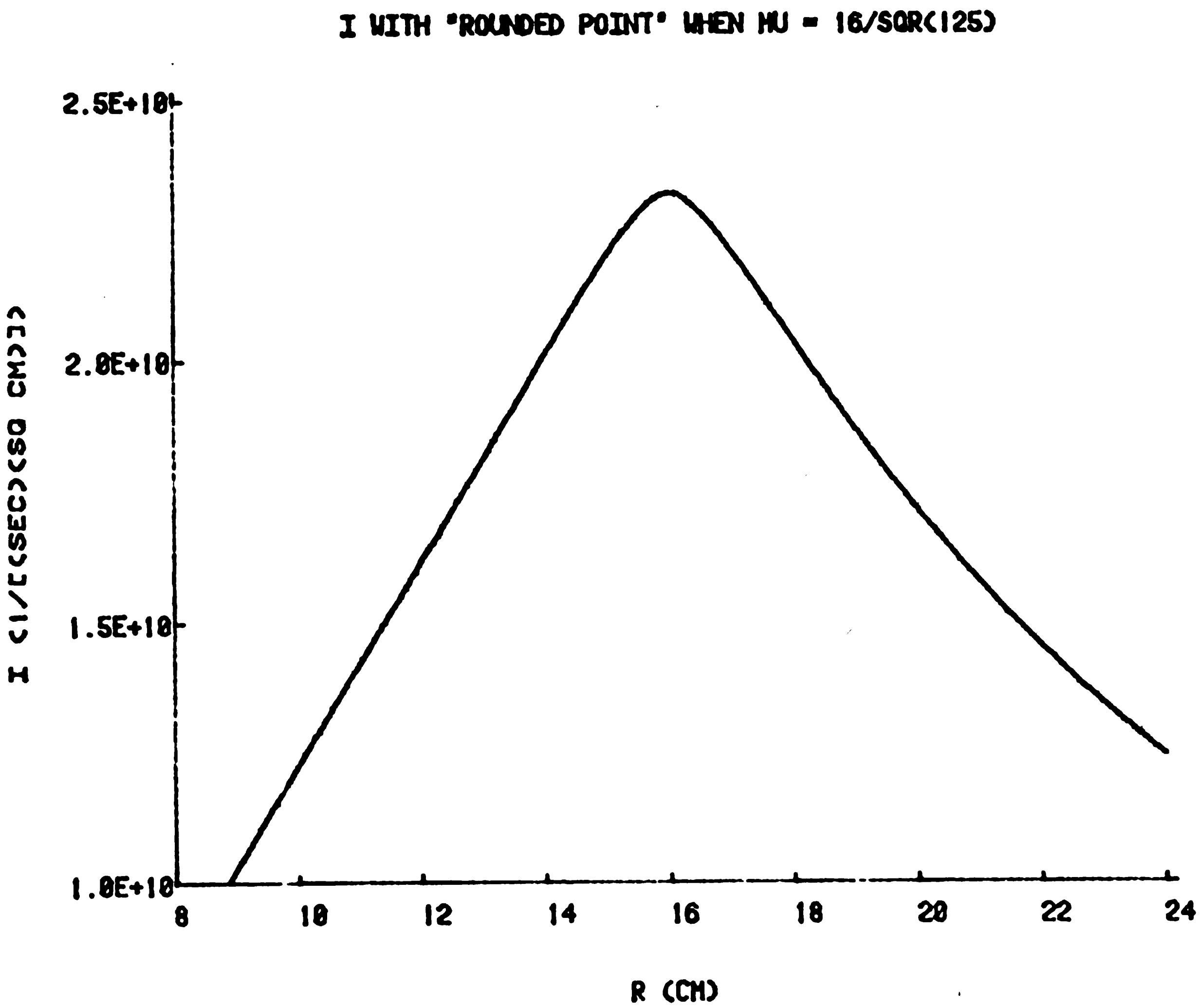

I WITH "ROUNDED POINT" WHEN MU = 16/SQR(125)
2.5E+10
2.0E+10
1.5E+10
1.0E+10
I (1/[(SEC)(SQ CM)])
8
10
12
14
16
18
20
22
24
R (CM)

Figure 13

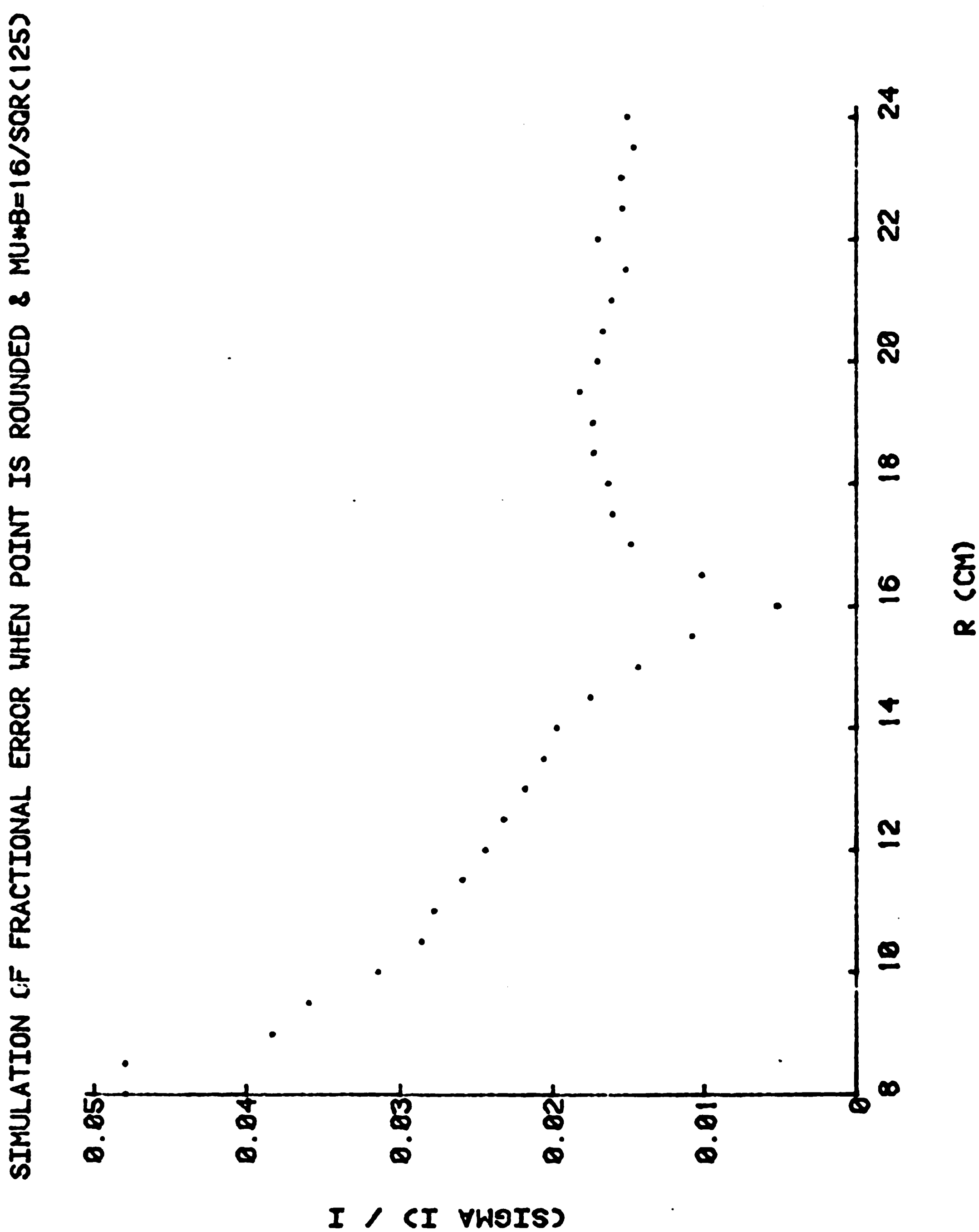

SIMULATION OF FRACTIONAL ERROR WHEN POINT IS ROUNDED & MU*B=16/SQR(125)
(SIGMA I) / I
0.05
0.04
0.03
0.02
0.01
0
8
10
12
14
16
18
20
22
24
R (CM)

Figure 14

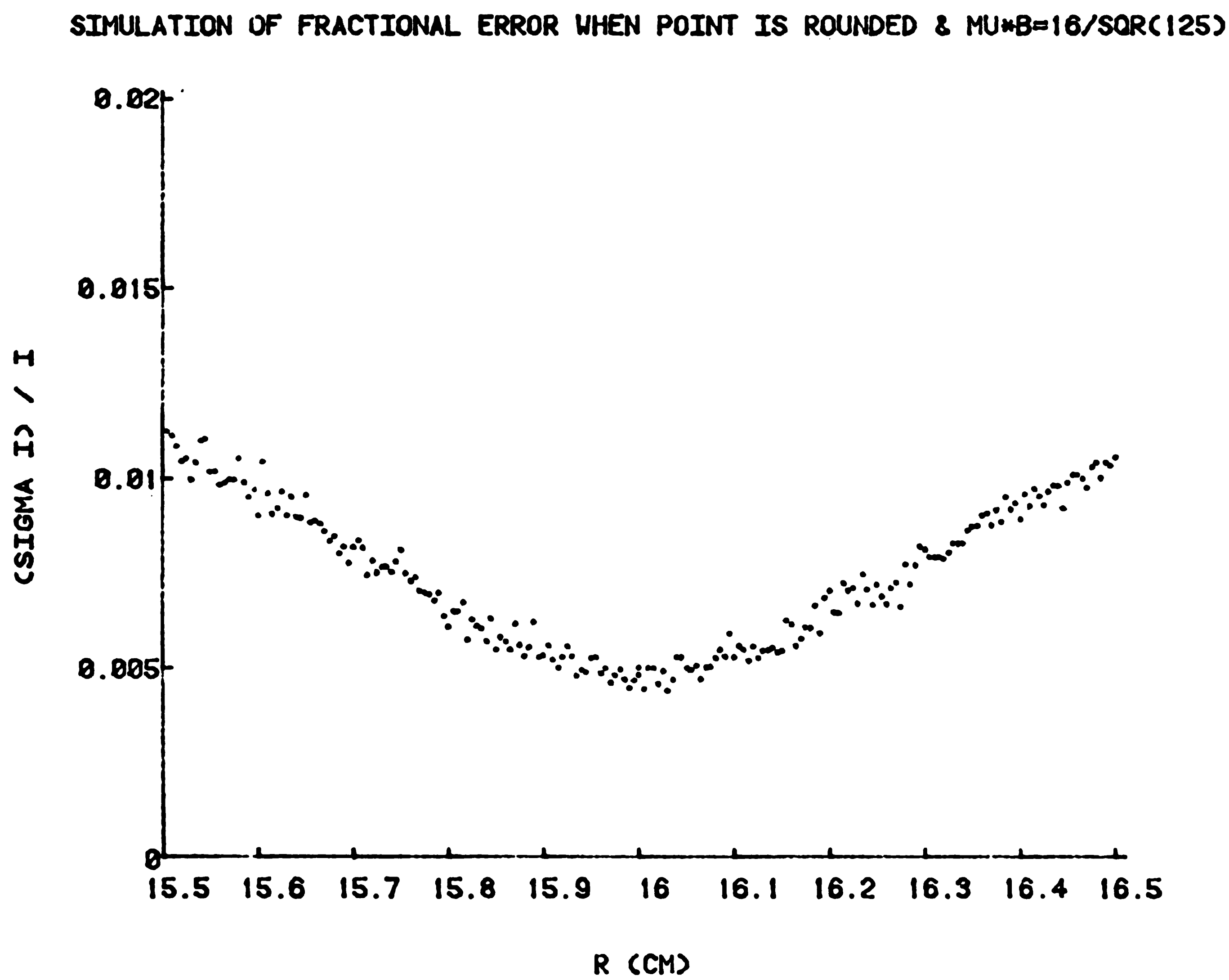

SIMULATION OF FRACTIONAL ERROR WHEN POINT IS ROUNDED & MU*B=16/SQR(125)
0.02
0.015
0.01
0.005
0
(SIGMA I) / I
15.5
15.6
15.7
15.8
15.9
16
16.1
16.2
16.3
16.4
16.5
R (CM)

Figure 15

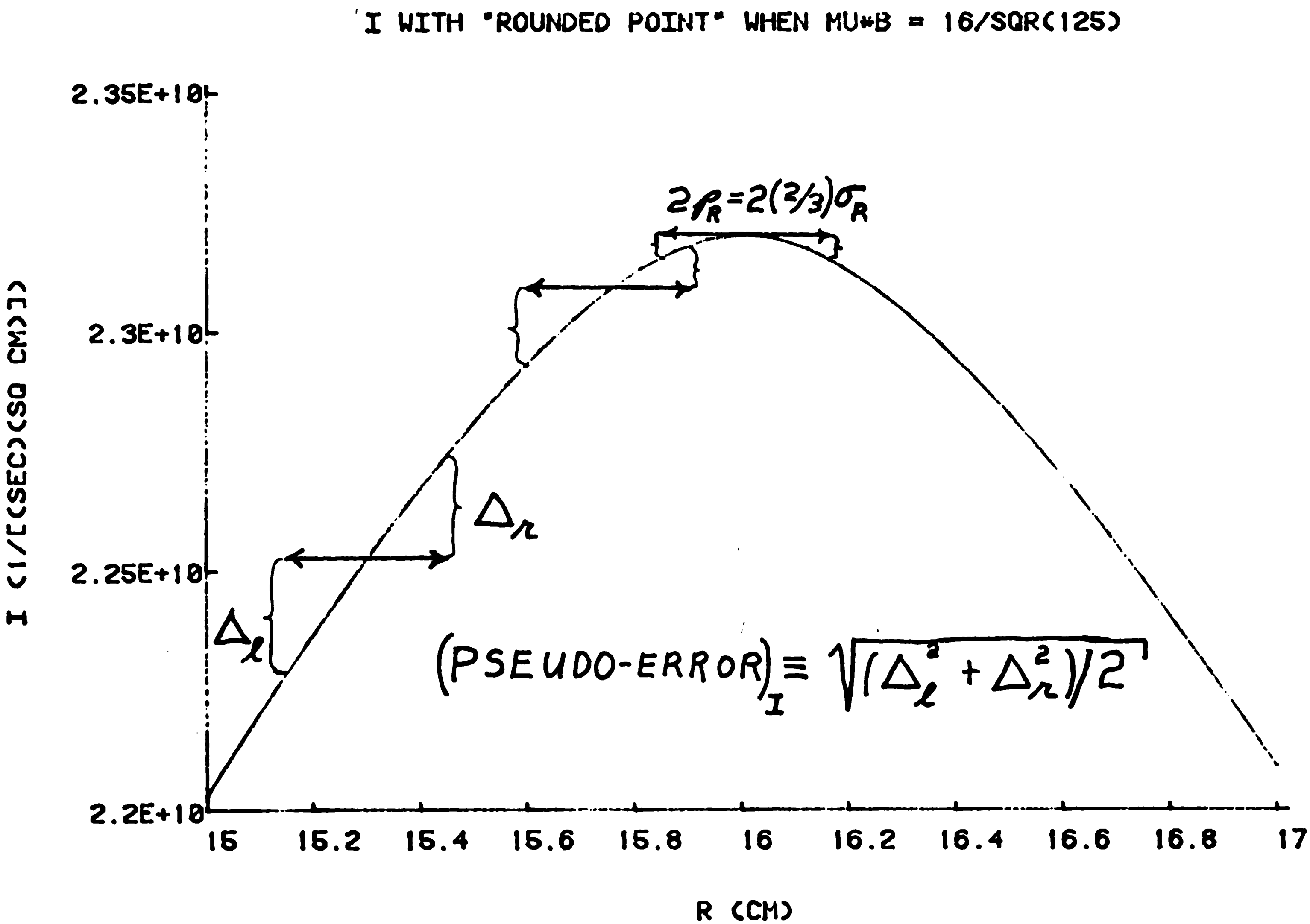

I WITH 'ROUNDED POINT' WHEN MU*B = 16/SQR(125)
2.35E+10
2.3E+10
2.25E+10
2.2E+10
I (1/[(SEC)(SQ CM)])
15
15.2
15.4
15.6
15.8
16
16.2
16.4
16.6
16.8
17
R (CM)
$2\rho_R = 2(2/3)\sigma_R$
Δ_r
Δ_ℓ
$(\text{PSEUDO-ERROR})_I \equiv \sqrt{(\Delta_\ell^2 + \Delta_r^2)/2}$

Figure 16

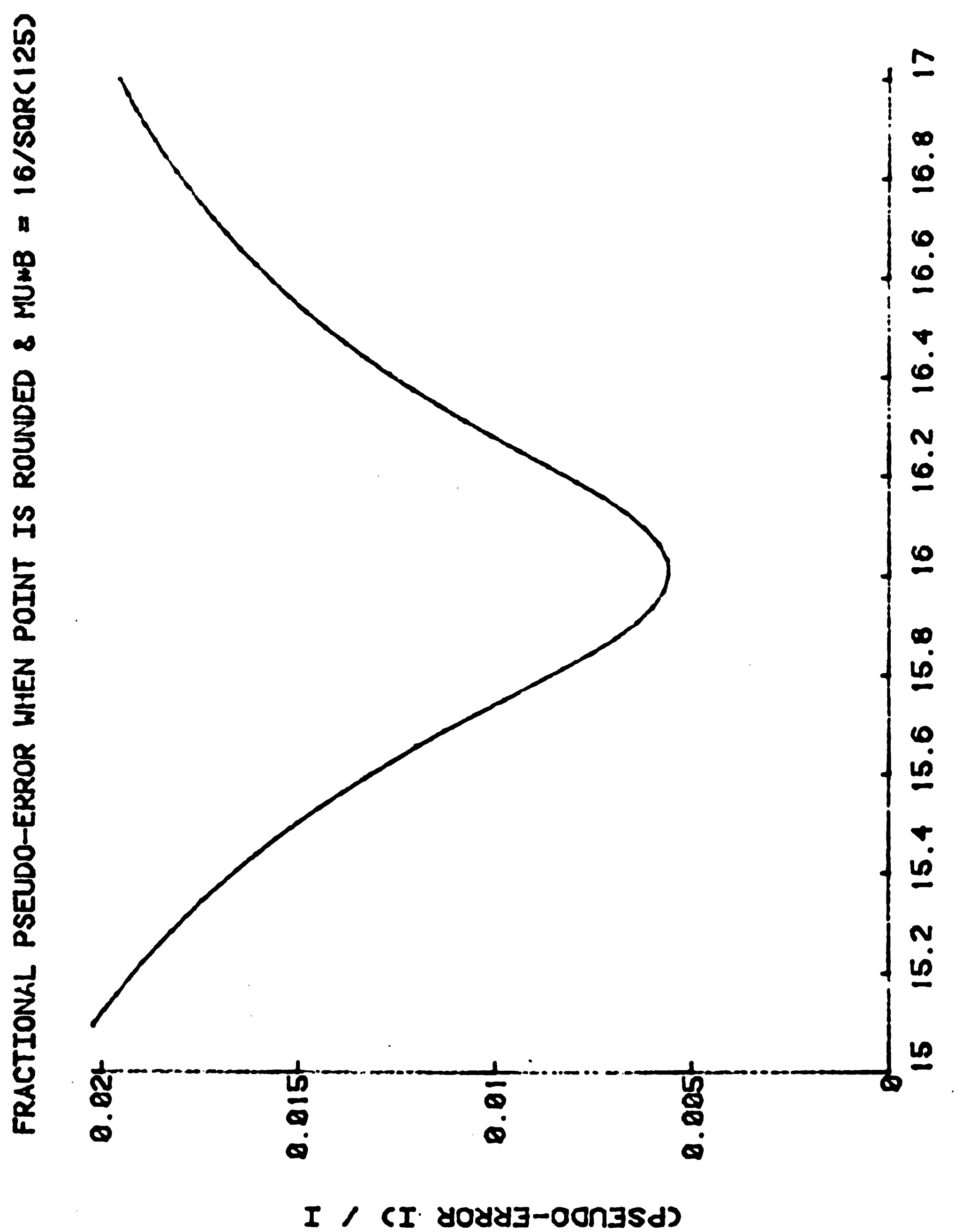

FRACTIONAL PSEUDO-ERROR WHEN POINT IS ROUNDED & MU*B = 16/SQR(125)
(PSEUDO-ERROR I) / I
0.02
0.015
0.01
0.005
0
15
15.2
15.4
15.6
15.8
16
16.2
16.4
16.6
16.8
17
R (CM)

Figure 17

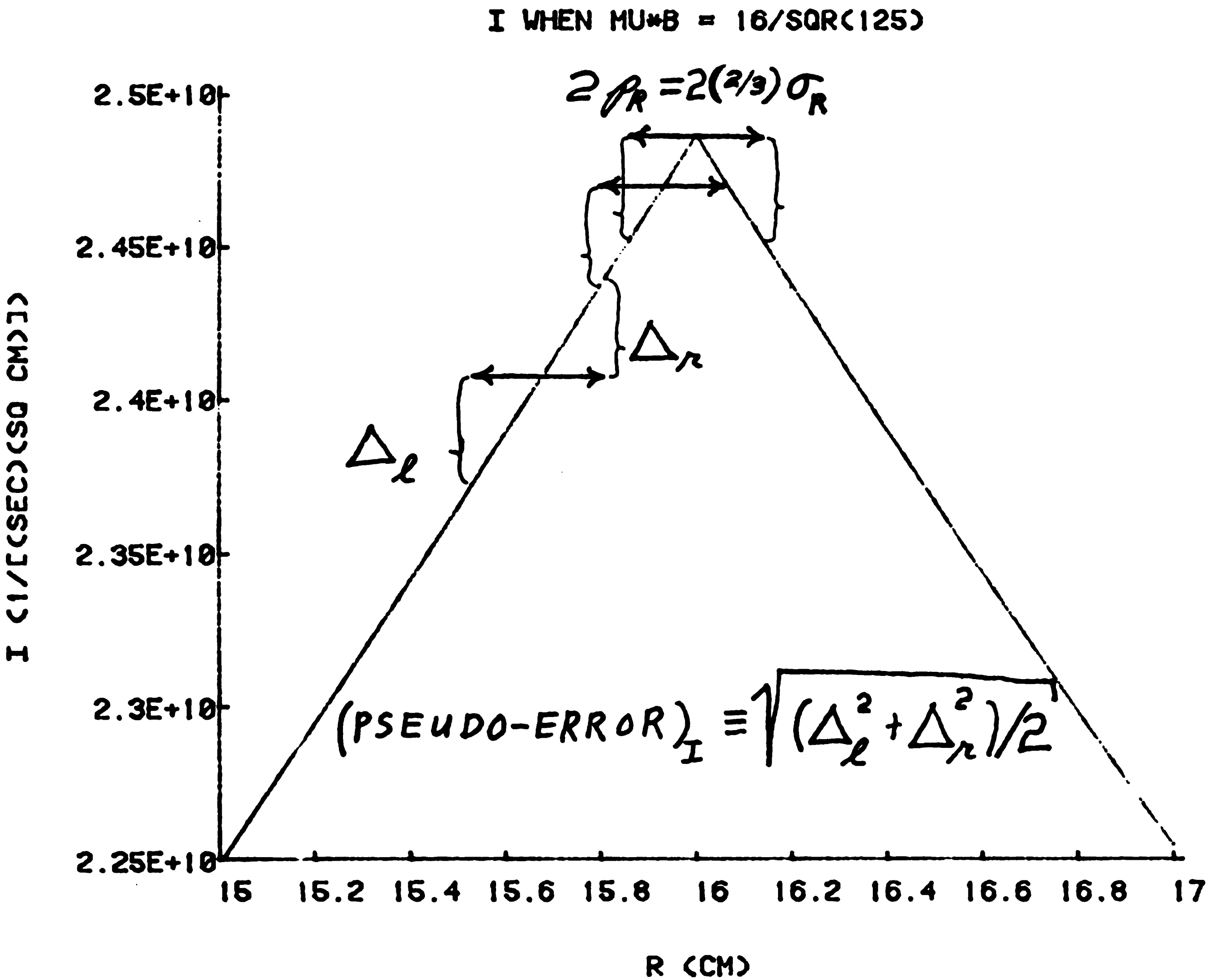

I WHEN MU*B = 16/SQR(125)
$2\rho_R = 2(2/3)\sigma_R$
Δ_r
Δ_ℓ
$(\text{PSEUDO-ERROR})_I \equiv \sqrt{(\Delta_\ell^2 + \Delta_r^2)/2}$
2.5E+10
2.45E+10
2.4E+10
2.35E+10
2.3E+10
2.25E+10
I (1/[(SEC)(SQ CM)])
15
15.2
15.4
15.6
15.8
16
16.2
16.4
16.6
16.8
17
R (CM)

Figure 18

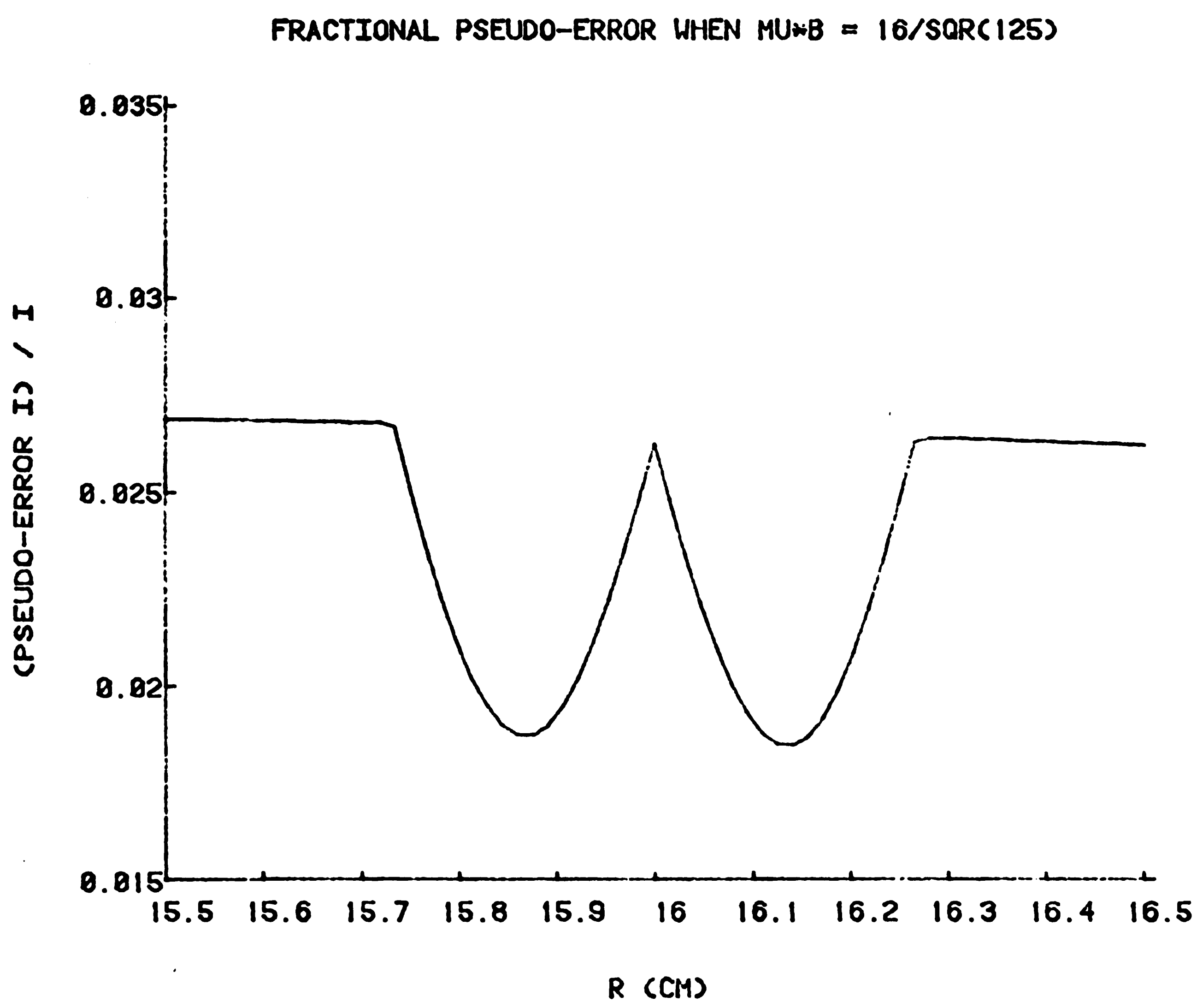

FRACTIONAL PSEUDO-ERROR WHEN MU*B = 16/SQR(125)
(PSEUDO-ERROR I) / I
0.035
0.03
0.025
0.02
0.015
15.5
15.6
15.7
15.8
15.9
16
16.1
16.2
16.3
16.4
16.5
R (CM)

Figure 19

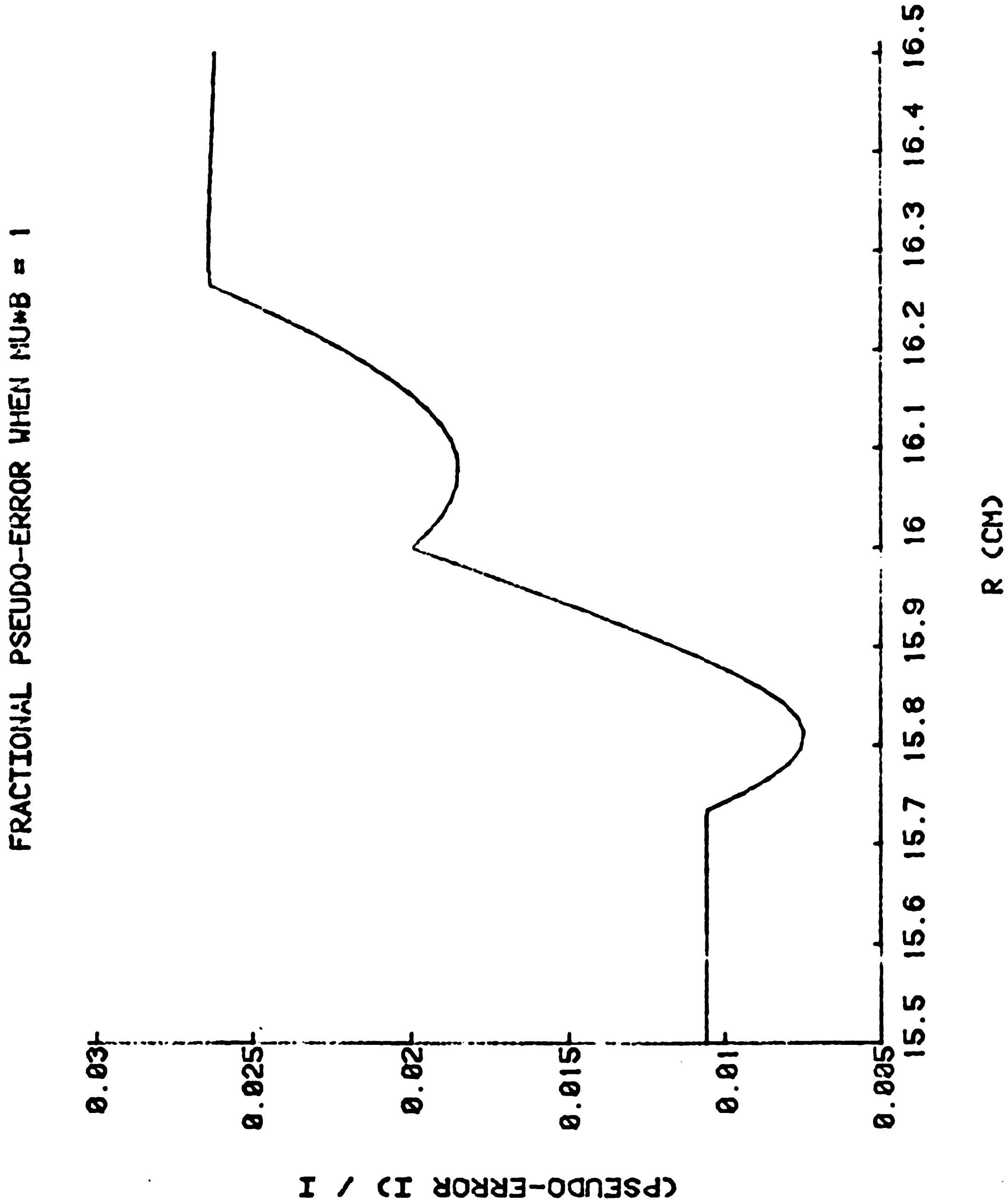

FRACTIONAL PSEUDO-ERROR WHEN MU*B = 1
(PSEUDO-ERROR I) / I
0.005
0.01
0.015
0.02
0.025
0.03
15.5
15.6
15.7
15.8
15.9
16
16.1
16.2
16.3
16.4
16.5
R (CM)

Figure 20

INFERENCE ON A FUTURE RELIABILITY PARAMETER

WITH THE WEIBULL PROCESS MODEL

Grady Miller
Methodology Office
Reliability, Availability, and Maintainability Division
U S Army Materiel Systems Analysis Activity
Aberdeen Proving Ground, Maryland

ABSTRACT

An inferential procedure is presented which provides confidence intervals for a future reliability parameter when reliability growth testing is only partially completed. Hypothesis tests based on this method are uniformly most powerful unbiased. These results are applicable if (1) the system failure rate can be modeled as the intensity function of a Weibull process and (2) efforts to improve reliability are assumed to continue at a steady rate throughout the intervening period of testing. The usefulness of this methodology is illustrated by evaluating the risk of not reaching some future reliability milestone. If such risk is unacceptably high, program management may have time to identify problem areas and take corrective action before testing has ended. As a consequence, a more reliable system may be developed without incurring overruns in the scheduling or cost of the development program.

1. INTRODUCTION

Reliability growth management is a critical function in the development programs of major defense systems.[1] It consists of planning, monitoring, and controlling the growth of reliability parameters throughout system development in order to achieve the reliability milestones for each test phase and for the overall program. A key factor in this process is the ability to assess the risk of not meeting a reliability requirement and to make such assessment at an early stage in the current test phase. If this risk is unacceptably high, the program manager may then have an opportunity to take remedial action before test time or other program resources are exhausted. The risks of failing to achieve program goals or contractual requirements can therefore be minimized. Instead of having to react to program shortcomings after the fact, management can exert positive control over the growth process to accomplish reliability objectives.

Reference 1 (pp. 10, 23, 28, 64-66, 75-78) discusses the use of reliability growth models to project reliability estimates beyond the present test time to some future time, such as the end of the current test phase. These projections are valid only if test conditions remain relatively constant and the development effort continues at its previous level. The projected reliability estimates are compared with future milestones in order to assess whether the reliability enhancement program is likely to reach a successful conclusion.

One of the problems with assessing a program by this method is how to evaluate the accuracy of the reliability projections. Such projections are only point estimates and do not reflect the uncertainties that accompany random sampling from a probabilistic model. In this paper we show how to quantify these uncertainties when the Weibull process is used to model and forecast reliability growth. The result is an objective appraisal of current program risks, and this appraisal can be factored into those management decisions which may impact on future reliability parameters.

The Weibull process model has been successfully applied to the reliability test results of many complex defense systems. It is introduced in Section 2 in a parametric form that is especially suited to the problem of forecasting. The basic features of this model are described in Appendix C of Reference 1, which includes confidence interval procedures for the reliability of the current system configuration. (See also References 2 and 3.) The theory developed in Section 3 extends these latter results to provide inferential

---

[1]Department of Defense, Reliability Growth Management, Military Handbook 189, Naval Publications and Forms Center, Philadelphia, PA, February 1981.

[2]Bain, L. J. and M. Engelhardt, "Inferences on the Parameters and Current System Reliability for a Time Truncated Weibull Process," Technometrics, Vol. 22, pp. 421-426, August 1980.

[3]Crow, L. H., Confidence Interval Procedures for Reliability Growth Analysis, Technical Report No. 197, U S Army Materiel Systems Analysis Activity, Aberdeen Proving Ground, MD, June 1977.

procedures for future reliability levels. These procedures are illustrated in Section 4, where confidence intervals are obtained for the reliability to be achieved at future points in a test phase which is still in progress. Also obtained by an equivalent technique is the risk of not achieving a certain reliability level at the end of the test phase.

2. SPECIFICATION OF MODEL

Consider a reliability growth test phase which has been underway for T units of testing. We shall hereinafter regard these test units as time, although they could equally well represent other units such as distance. Suppose that the test phase began at time 0, but is planned to continue for an additional S units of testing till test time T+S, at which point the system configuration will have failure rate R. Our objective is to make inferences about the parameter R.

A Weibull process is a nonhomogeneous Poisson process with an intensity function that can be expressed as a multiple of some power of the test time. For the particular test phase described above, an intensity function of the appropriate parametric form is

$$r(t) = R[t/(T+S)]^{\beta-1}, \tag{1}$$

where $R>0$, $\beta>0$, and $0<t\leq T+S$. As shown in Figure 1, the function $r(t)$ models the failure rate of the system configuration as it changes over a reliability growth test phase of length T+S, and the failure rate at the end of the (as yet uncompleted) test phase is given by $r(T+S) = R$.

The failure rate model in Figure 1 shows a decreasing trend during future testing from time T to time T+S. This trend reflects our previously stated intention to continue reliability improvements throughout this period. The case in which reliability is constant from T to T+S is treated in Reference 4.

According to the scenario of this paper, test results are available for the test period from time 0 to the (current) time T, but the system testing from time T to time T+S has not yet been accomplished. Let N be the number of failures that occur before time T and $T_1, \ldots, T_N$ the observed failure times $(0<T_1 < \ldots <T_N<T)$. Then the Poisson process with intensity function $r(t)$ has a sample function density given by

$$f_{N,T_1, \ldots, T_N}(n, t_1, \ldots, t_n)$$

[4]Miller, G., "Efficient Methods for Assessing Reliability," Proceedings of the Nineteenth Annual U S Army Operations Research Symposium, Part III, pp. 33-42, October 1980.

$$
= \begin{cases} \exp[-(RT/\beta)Q^{1-\beta}] & \text{if } N = 0, \quad (2.1) \\ R^n \prod_{i=1}^{n} [(T+S)/t_i]^{1-\beta} \exp[-(RT/\beta)Q^{1-\beta}] & \text{if } N = n>0, \quad (2.2) \end{cases}
$$

where $Q = (T+S)/T$; $n = 0, 1, \ldots$; and $0<t_1 <\ldots <t_n<T$. (See e.g., Reference 5.)

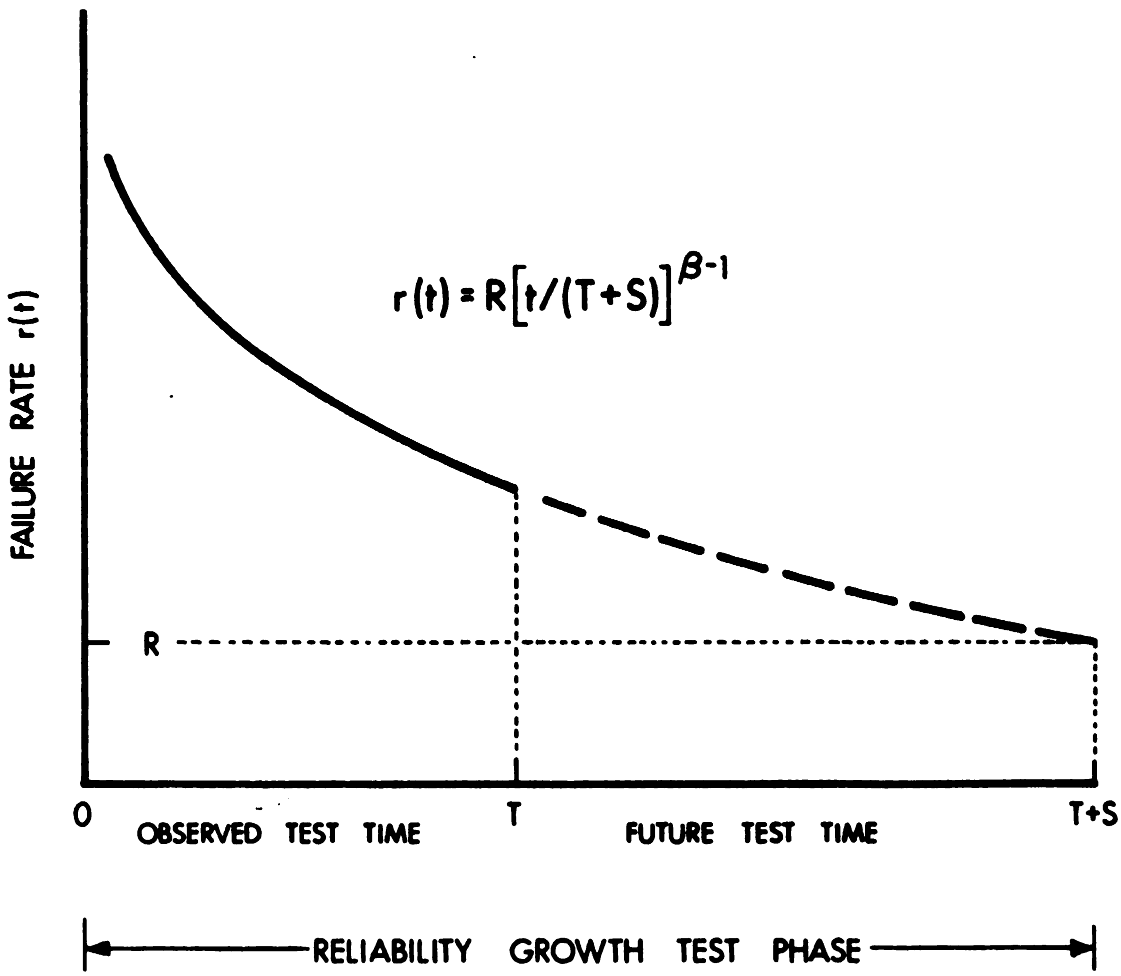


FIGURE 1. Intensity Function for the Case $\beta<1$.

[5]Snyder, D. L., Random Point Processes, John Wiley and Sons, New York, NY, 1975.

3. DERIVATION OF RESULTS

3.1 Point Estimators

The Weibull process model is used in applications where $Pr(N = 0)$ is quite small, and therefore the likelihood expression in Equation (2.2) can be maximized to obtain point estimators for β and R as follows:

$$\hat{\beta} = N/\sum_{i=1}^{N} \ln(T/T_i), \tag{3}$$

$$\hat{R} = N\hat{\beta}Q^{\hat{\beta}-1}/T. \tag{4}$$

As would be expected, the expression in (3) is identical to the estimator for β in Reference 2 (Equation (4)). The projected mean time between failures (MTBF) for the system configuration at the end of the test phase (time T+S) is estimated by $\hat{R}^{-1}$.

The point estimators in Equations (3) and (4) are convenient because of their simplicity, but were obtained without conditioning formally on the event $N>0$. As a practical matter, inferences on the two-parameter Weibull process are possible only when $N>0$, and we shall condition on this event in the sequel without further mention.

3.2 Reduction of the Parameter Space

Let $V = \sum_{i=1}^{N} \ln[(T+S)/T_i]$, and observe from Equation (2) that V is a sufficient statistic for β. It follows from Reference 6 (pp. 134-140) that uniformly most powerful unbiased (UMPU) hypothesis tests on the future failure rate R can be constructed by utilizing the conditional distribution of N given $V=v$. To obtain this distribution, we begin by determining the conditional distribution of V given $N=n$.

Given $N=n$, the random variables $T_1, \ldots, T_n$ are distributed as the order statistics from n independent distributions with cumulative distribution function

$$F(t) = \int_0^t r(x)dx/\int_0^T r(x)dx$$

$$= (t/T)^{\beta}, \tag{5}$$

[6]Lehmann, E. L., Testing Statistical Hypotheses, John Wiley and Sons, New York, NY, 1959.

where $0<t<T$. Let X be a random variable with distribution function F. Then a straightforward calculation shows that the random variable $\ln[(T+S)/X]$ is distributed over the interval $(\ln Q,\infty)$ according to

$$\Pr\{\ln[(T+S)/X] \leq y\} = 1-\exp[-(y-\ln Q)\beta], \tag{6}$$

where $\ln Q<y<\infty$. This latter function is a two-parameter exponential distribution function on the interval $(\ln Q,\infty)$. The conditional distribution of V given N=n is therefore the sum of n such distributions, all independent, and consequently is a three-parameter gamma distribution with density function

$$f_{V|N}(v|n)$$

$$= \beta^n(v-n\ln Q)^{n-1}\exp[-\beta(v-n\ln Q)]/(n-1)!, \tag{7}$$

where $n\ln Q<v<\infty$.

The random variable N is Poisson distributed with mean value $\theta\equiv(RT/\beta)Q^{1-\beta}$, so that (conditional on N>0)

$$\Pr(N=n) = [1-\exp(-\theta)]^{-1}\theta^n\exp(-\theta)/n!, \tag{8}$$

$n = 1, 2, \ldots$. Thus the joint density function of V and N is

$$f_{V,N}(v,n) = f_{V|N}(v|n)\ \Pr(N=n)$$

$$= \frac{\exp(-\theta-\beta v)}{1-\exp(-\theta)}\,(RTQ)^n\,\frac{(v-n\ln Q)^{n-1}}{n!(n-1)!}, \tag{9}$$

where $n = 1, 2, \ldots$ and $n\ln Q<v<\infty$.

In the case S=0 (forecasting zero time into the future), we see that $\ln Q=0$ and that the results in this paper generalize certain results in [3] and [2] on inferences for current system reliability. In the case S>0, the above inequality $n\ln Q<v<\infty$ implies that N has finite support, given V=v:

$$\Pr(0<N<v/\ln Q|V=v) = 1. \tag{10}$$

Given V=v, let G(v,S) be the greatest integer less than $v/\ln Q$ if S>0 and $G(v,S) = \infty$ if S = 0.

We can now write down the conditional distribution of N given V=v as

$$p(n;R) \equiv \Pr(N=n|V=v,\ N>0)$$

$$= \frac{(RTQ)^n(v-n\ln Q)^{n-1}/n!(n-1)!}{\sum_{k=1}^{G(v,S)} (RTQ)^k\,(v-k\ln Q)^{k-1}/k!(k-1)!}, \tag{11}$$

where $n = 1, 2, \ldots, G(v,S)$. This expression for $p(n;R)$ can be readily evaluated at minimal cost with an electronic computer.

3.3 Inferential Procedures

A conservative $1-\alpha$ confidence interval for R can be constructed by obtaining values R_1 and R_2 which satisfy $\sum_{k=n}^{G(v,S)} p(k;R_1) = \alpha_1$ and $\sum_{k=1}^{n} p(k;R_2) = \alpha_2$, where $\alpha_1 + \alpha_2 = \alpha$. The corresponding confidence bounds for R^{-1} (the MTBF at test time T+S) are R_2^{-1} and R_1^{-1}. Because N is a discrete random variable, construction of exact confidence intervals would require randomization. A UMPU test of $H_0: R \leq R_0$ versus $H_1: R > R_0$ at significance level α calls for rejection of H_0 if $\sum_{k=n}^{G(v,S)} p(k;R_0) \leq \alpha$. Other UMPU hypothesis tests can be constructed in a similar manner. If R_0^{-1} is the MTBF goal for the end of the test phase (time T+S), then the risk of not achieving this goal may be evaluated as $\sum_{k=1}^{n} p(k;\ R_0)$.

4. EXAMPLE

Suppose that a reliability growth test phase has been in progress for T=200 hours and is scheduled to continue for another S=200 hours. From the test data up to time 200, we wish to obtain an 80 percent confidence interval for the MTBF at time T+S = 400. The following failure times t_i were recorded (n=21): 2.2, 3.3, 4.5, 5.3, 5.8, 20.3, 27.4, 34.1, 55.2, 58.4, 61.4, 62.2, 78.3, 78.4, 91.9, 97.7, 112.4, 116.9, 142.4, 176.8, 181.5.

Equations (3) and (4) yield $\hat{\beta} = .591$ and $\hat{R}^{-1} = 21.4$, and it is also of interest to observe that $v/\ln Q = 72.3$. Thus $G(v,S) = 72$, so that the set of positive integers less than or equal to 72 is a support of the conditional distribution of N given V=v.

With Equation (11) we obtain by iteration the values $R_2^{-1} = 12.7$ and $R_1^{-1} = 38.6$ such that $\sum_{k=21}^{72} p(k;R_1) = .10$ and $\sum_{k=1}^{21} p(k;R_2) = .10$. The interval (12.7, 38.6) is therefore an 80 percent confidence interval for the MTBF at time 400.

By successively taking S=0 and S=100, we can obtain in a similar manner 80 percent confidence intervals (10.7, 26.0) and (11.9, 32.7) for the MTBF at times 200 and 300, respectively. All three confidence intervals are shown in Figure 2 for comparison purposes.

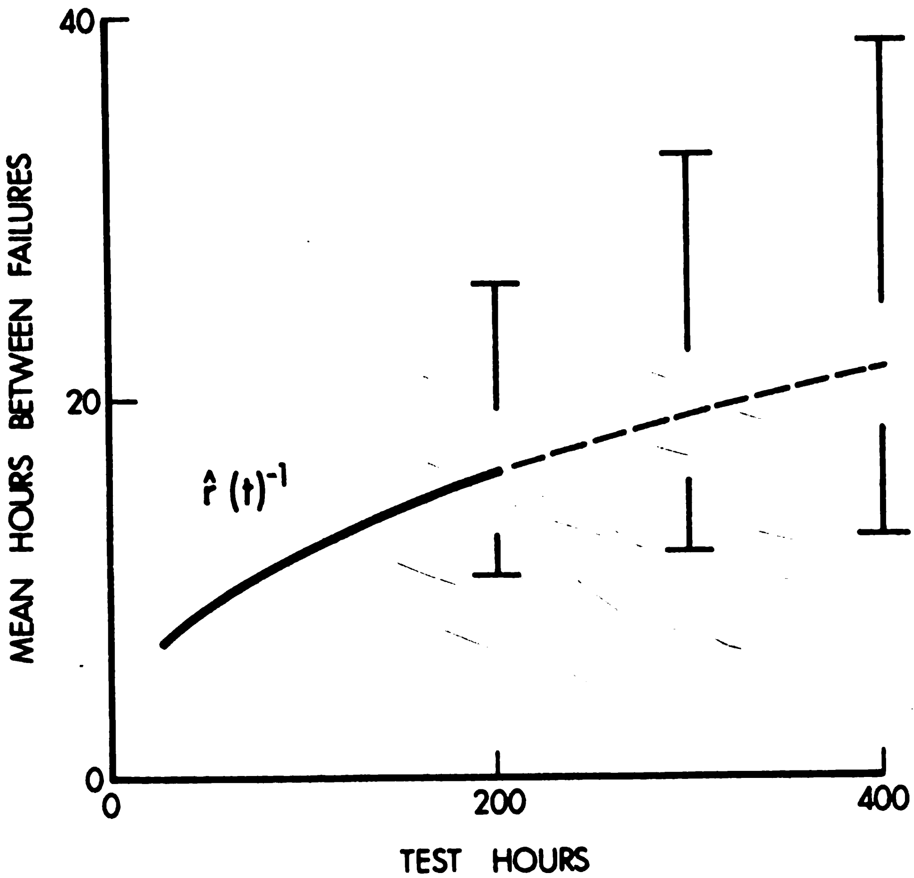


FIGURE 2. Eighty Percent Confidence Intervals for Current and Future MTBF Parameters.

Suppose further that an MTBF goal of 15.0 has been set as a milestone for the end of the current reliability growth test phase (T+S = 400). Based on the data up to time T = 200, the risk of not achieving this goal is $\sum_{k=1}^{21} p(k;15^{-1}) = .20$. In view of such a result, the program manager should feel optimistic about this aspect of the development program, but will probably want to avoid any actions which might adversely affect the overall reliability enhancement effort.

REFERENCES

1. Department of Defense, Reliability Growth Management, Military Handbook 189, Naval Publications and Forms Center, Philadelphia, PA, February 1981.

2. Bain, L. J. and M. Engelhardt, "Inferences on the Parameters and Current System Reliability for a Time Truncated Weibull Process, Technometrics, Vol. 22, pp. 421-426, August 1980.

3. Crow, L. H., Confidence Interval Procedures for Reliability Growth Analysis, Technical Report No. 197, U S Army Materiel Systems Analysis Activity, Aberdeen Proving Ground, Maryland, June 1977.

4. Miller, G., "Efficient Methods for Assessing Reliability," Proceedings of the Nineteenth Annual U S Army Operations Research Symposium, Part III, pp. 33-42, October 1980.

5. Snyder, D. L., Random Point Processes, John Wiley and Sons, New York, NY, 1975.

6. Lehmann, E. L., Testing Statistical Hypotheses, John Wiley and Sons, New York, NY, 1959.

SPECIFYING A DETECTABLE 3-FACTOR INTERACTION WITH THE NON-CENTRAL F

Walter D. Foster
Jack L. Wray
AFIP, Washington D.C. 20306

ABSTRACT. Just as the "pumpkin papers" typewriter was identified, it is plausible that the print-out can identify a small (home) computer. A typical print head consists of 7 (or 9) rods or pins in a vertical column which move in time to 7 successive positions in the character space. By high precision photography and magnification, density readings offer an approach towards identification of a specific printer from its dot matrix. Because ink depletion on a ribbon or a new ribbon on the same printer will cause changes in density, the variable of analysis was taken to be relative density, the density of a set of pins relative to a specific pin. Analysis of repeated characters per printer over a set of 9 printers gave rise to an analysis of variance in which a 3-factor interaction term, if large relative to the within term (for repeated determinations), was deemed to be capable of excluding printers. A candidate printer is identified as the putative originator when the 3-factor interaction is equal to the within mean square. Since identification is based upon accepting the null hypothesis, the non-central F distribution was used to set a value for the 3-factor interaction under the concept that if such a departure existed, the analysis would have $1 - \beta$ power to detect it.

1. INTRODUCTION. Tracing ransom notes to a specific typewriter primarily has depended upon finding a type defect in the document and tracing it to a particular machine. This tactic is successful when the defect is unusual or unique to that machine. The same technique has been applied to print-outs from small (home) computers. This paper is concerned with a statistical model for the identification of a computer printer that is operating properly by means of the density pattern from pin to pin in the dot-matrix print-head of the printer. A typical print-head consists of 7 pins or rods in a vertical column which move horizontally in a character space to create a given character. Because of curvature of the platen, the dots in the print matrix tend to have different densities. A well-worn ribbon leaves different densities than a new ribbon. Hence, density of itself is not a satisfactory measure for identification. However, the difference in density relative to a specified pin will show patterns characteristic of a print-head.

2. METHODS. Measurements of density were obtained by making high-precision, photographic enlargements of characters on a print-out. A precision densitometer with an absolute reference calibration was used to measure the densities of pin marks on the print-out. Vertical characters were represented by 1, B, D, E and slant characters by 7, X, Z, /. Determinations of relative density were made on each of five pins in each character for nine printers. The characters were printed five times by each printer to provide replication.

3. STATISTICAL MODEL. Main factors were defined as Exemplars (or print-outs from printers), Pins, Verticality, and Characters in verticality. Characters were nested in verticality; all factors were considered as fixed effects except characters. This balanced model is written as:

$$Y = u+E_e+P_p+V_v+C_{cv}+EP_{ep}+EV_{ev}+PV_{pv}+EPV_{epv}+EC_{ecv}+PC_{pcv}+EPC_{epcv}+e_{epcvr}.$$

Exemplars as a main effect had no relevant interpretation because the variable of analysis, relative density, included an arbitrary reference pin. Verticality was defined as a two-level factor comprised of the average of the vertical characters and the average of the slant characters. The pattern of interest was the Pin x Verticality interaction and how it varied from one exemplar to another, Figure 1. Thus, the 3-factor interaction, Exemplar x Pin x Verticality was taken as the criterion for identifying two print-outs as coming from the same print-head or conversely denying that they had a common source.

Denying that two print-outs could have come from the same print-head was achieved by finding a statistically significant EPV interaction. Thus the probability of making a Type I error was fixed at α.

Affirming that two print-outs could have come from the same print-head was equivalent to accepting the null hypothesis. In order to control the rate of Type II errors, β, it was necessary to specify a value for the EPV interaction that would be detected if it existed $1 - \beta$ proportion of the time. The non-central F parameter,

$$\phi^2 = \frac{\sum_1^k (u_i - \bar{u})^2/k}{\sigma^2/n}, \text{ where } k = \text{\# means and } n = \text{\# values per mean,}$$

is customarily used to specify a set of means under the alternate hypothesis. Because ϕ^2 has the F structure--the numerator is a mean square of the means under the alternate hypothesis and the denominator is the variance of a mean--it is clear that ϕ^2 can be used to specify an interaction term as well.

It is not immediately clear what value of k or degrees of freedom should be used for ϕ^2 in the case of interaction. We begin our argument by examining a 2-factor, 2 x C interaction, Figure 2b. Under the null hypothesis, the F-statistic tests whether the difference between two rows is the same from column to column. The alternative hypothesis would specify a pattern of differences. Therefore it is argued that the degrees of freedom for the 2 x C interaction would be C, one for each column.

The case for a 3-factor interaction composed of two categories, two rows/category, and C columns follows the same argument. Under the null hypothesis, the F-statistic tests whether the pattern of row differences across columns is the same from one category to the other, Figure 2c. The alternative hypothesis would specify how the pattern of row differences across columns would vary from category to category. In the special case of two categories and two rows/category, it is argued that the degrees of freedom for ϕ^2 under the alternative hypothesis are C, one for each difference (of the differences). This is to say that the value of k in the expression for ϕ^2 in the special case of a 2 x 2 x C interaction should be C.

Alternatively, the value of k may be approached by the conventional concept of sample size minus the number of constraints imposed by the assumptions and analysis. In Figure 2a, the row vector of means refers to a set of C population means. There are no constraints on the means so that $k = C - 0 = C$. In Figure 2b, the number of population means is 2C. Because C differences between two rows are specified under the alternative hypothesis, one row (or the total of both rows) must be fixed or constrained. Therefore $k = 2C - C = C$. In Figure 2c, the number of means is 4C. The quantity to be specified under H_a is $d_{Aj} - d_{Bj}$, where d_{Aj} is the difference between Row 1 and Row 2 for column j in the A category. The number of constraints on the d_{Aj} is C, according to the argument for 2b. Likewise, the number of constraints on the d_{Bj} is C. The differences, $d_{Aj} - d_{Bj}$, require that either the column totals for one category or the column totals over both categories be fixed, so that the number of constraints is C. Therefore, $k = 4C - C - C - C = C$.

To complete the statistical model using the non-central F statistic, the values of $\alpha = .05$, $\beta = .10$, and $k = 5$ (for the 5 pins) were used to specify a value of $\phi^2 = 2.72$. (In a court case, maybe β should be set at .01 with a corresponding $\phi^2 = 4.52$). The decision rule is now defined: it denies the same print-head for two print-outs with a significant[1/] EPV interaction and affirms the same print-head if the EPV interaction is less than 2.72 times the Within mean square.

4. <u>RESULTS</u>. We applied this decision rule to several comparisons. The analysis of variance for the nine exemplars or print-outs that are shown in Figure 1 is given in Table 1 under the heading, '9 Exemplars.' The F-statistic for the null hypothesis was 1.57, with P = .05. Of course, the exemplars are known to be different. So far so good.

An interested computernik friend offered to test our system. He brought in three print-outs and challenged us to identify correctly what (if any) print-heads were the same or different.

<u>1/</u> If statistical significance corresponds to a decision rule of $F = 2.72$, i.e. reject if $F \geq 2.72$, accept if $F < 2.72$, note that $\alpha \cong .03$ for F(4, 320).

The print-outs were photographed, enlarged, and read with the same densitometer using five repeated characters as before. The analysis of variance is given in Table 1 under the heading HWL345. The pattern for the EPV interaction is shown in Figure 3 for the set marked H3, H4, and H5. These patterns are seen to be remarkable similar but with some failure to be exactly the same. The analysis of variance gave an F-ratio of 1.82/5.67 for the EPV/Within ratio, strongly suggesting that the three print-outs were far more similar than expected according to experimental technique. The decision rule required us to say that the three print-outs all came from the same print-head. 1/ If statisticians on occasion refer to computerniks in less than flattering terms, this computernik had the last word, for after he had stopped laughing (some 300 x 10^9 nanoseconds later according to his computer clock), he revealed that the three test print-outs indeed came from three separate machines but of the same model. Of course we can't tell you whose mfg these machines were, but our code for it is HWL. At any rate, we had flunked this test.

We then designed our own test, taking two successive print-outs from the same machine. The patterns are shown in Figure 2, 7D and 7E, with the corresponding analysis of variance given in Table 1 with the same column heading. Again, the patterns can be seen to be very similar, in fact too similar according to expectation from the Within mean square. But this time when we concluded that the two print-outs were from the same machine, we were right.

Two other comparisons are shown Table 1. Exemplars 2 and 4 were chosen at random for comparison and found to have an F-ratio of 9.00/5.62 for the EPV/W test. It comes as something of a shock when making the visual comparison of exemplars 2 and 4 in Figure 1 to realize that the EPV mean square merely reflects whether the pattern of differences in exemplar #2 is equal to that of #4. Score-wise for this comparison, ouch.

The last comparison concerns 7AC, whose patterns are shown in Figure 3. The EPV/W ratio was found to be 16.83/9.8 , a result that affirms that the print-heads were the same. They were.

Two problems have surfaced. First is the failure of the within meansquare with its very large number of df per analysis to remain reasonably consistent. Secondly, it may well be that five repeat characters is not enough to meet the stringency this procedure may need.

1/ Technically, we would be required by the decision rule to make three, pair-wise comparisons of the exemplars and then to live with the problem of dependent comparisons.

FIGURE 1.
GRAPHS OF RELATIVE DENSITY FOR AVERAGE OF VERTICAL (I,B,D,E) AND SLANT (7,X,Z,/) CHARACTERS FOR EACH OF NINE PRINTERS

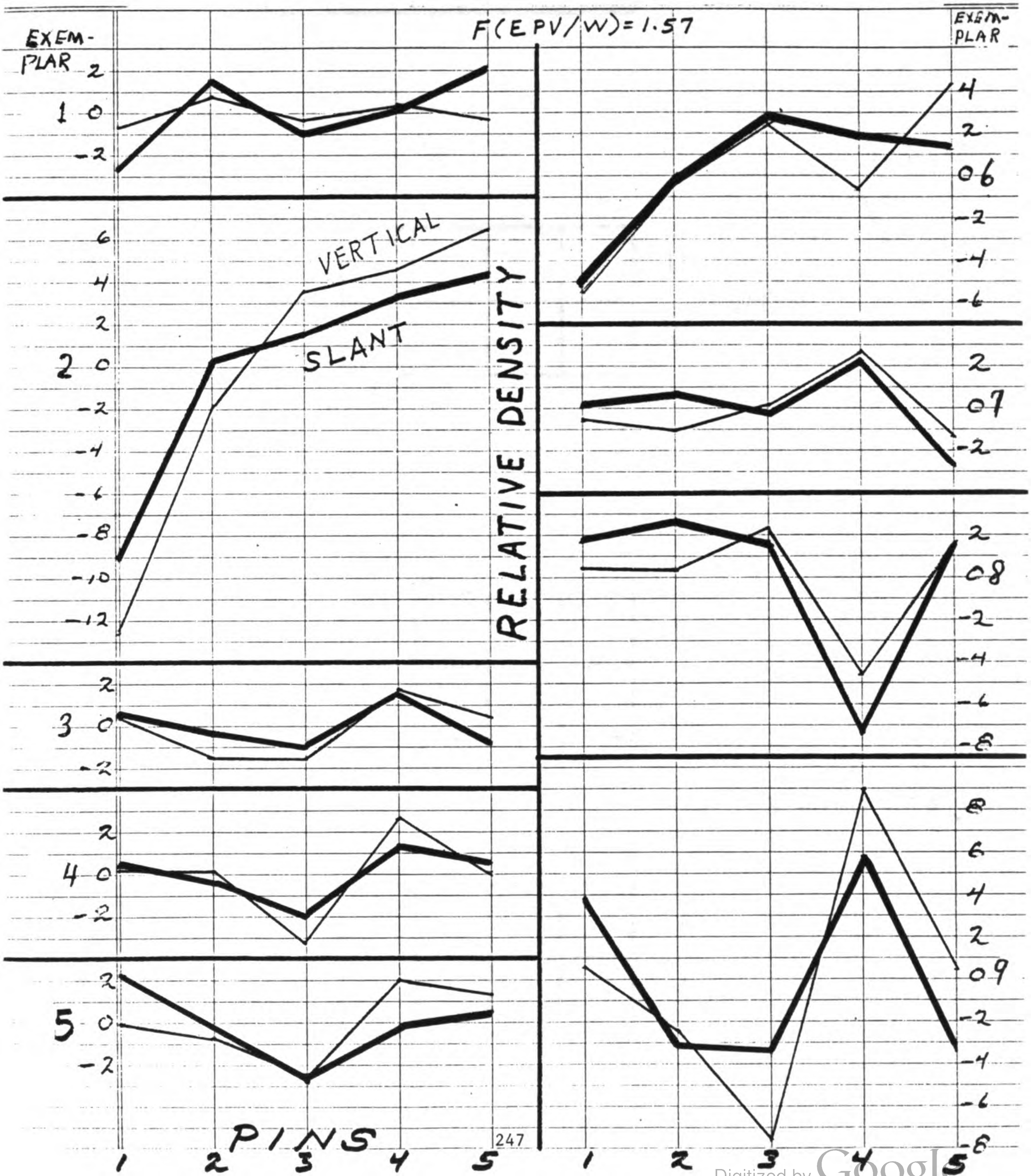


FIGURE 2.

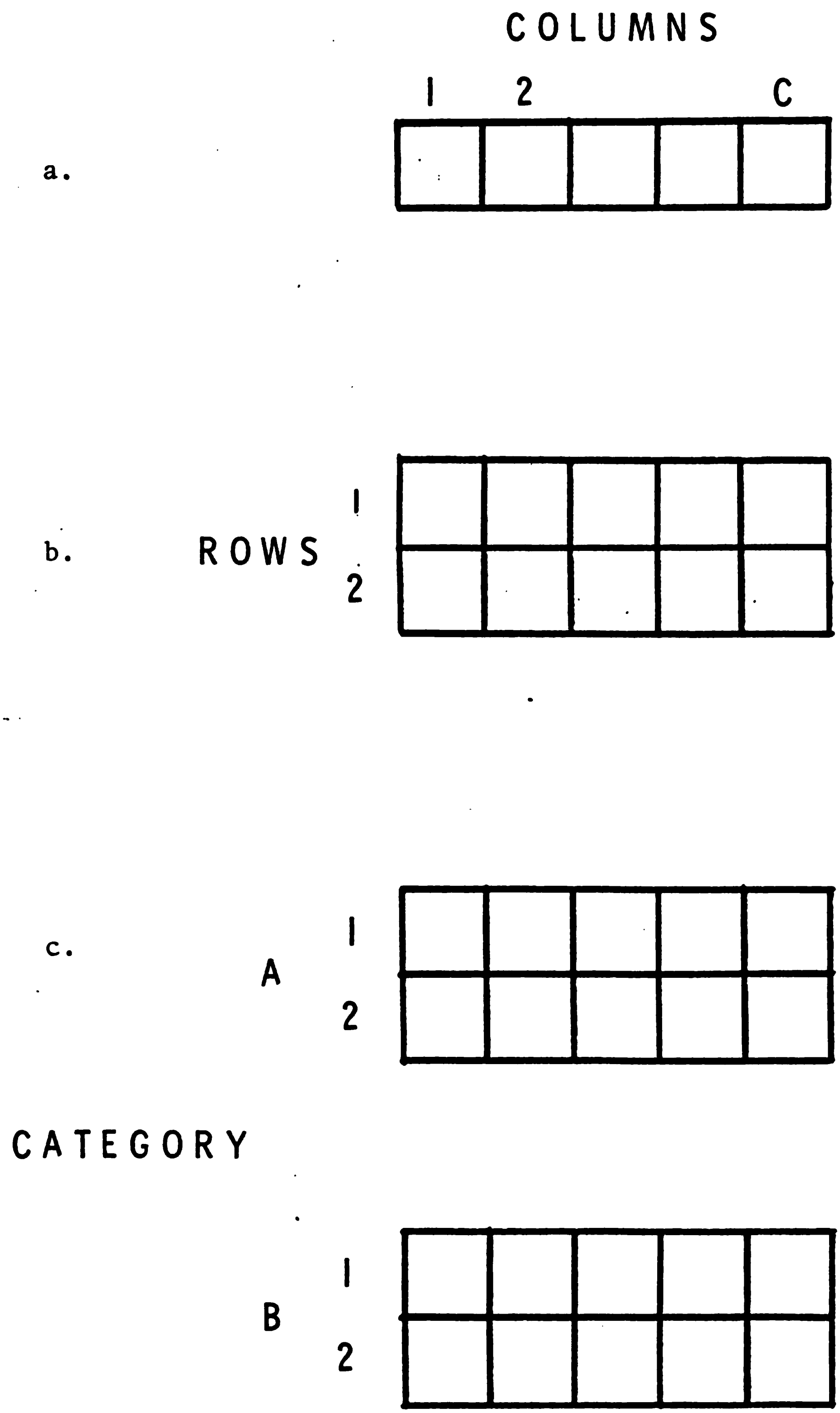


FIGURE 3.
GRAPHS OF RELATIVE DENSITY FOR AVERAGE OF VERTICAL (I,B,D,E) AND SLANT (7,X,Z,/) CHARACTERS FOR SPECIFIC SETS OF PRINTERS

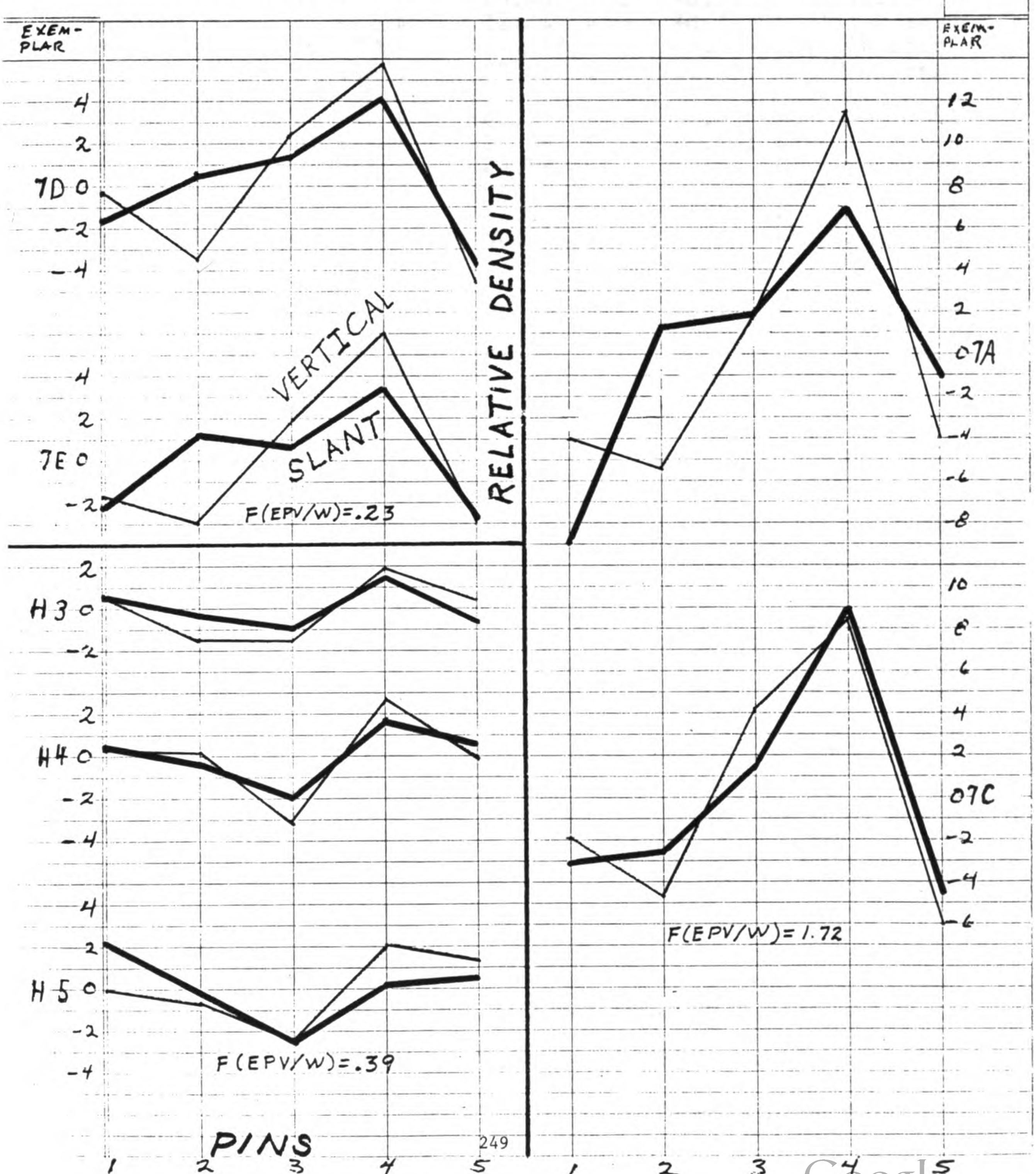


TABLE 1. ANALYSES OF VARIANCE ON DENSITY

SUBSET COMPARISONS

	7D vs 7E		HWL-345		7-AC		ALL NINE	
Source	df	MS	df	MS	df	MS	df	MS
Exemplar	1	.68	2	64.00	1	301.86	8	708.36
Pins	4	174.08	4	47.25	4	516.26	4	100.19
Verticality	1	777.50	1	536.79	1	2291.94	1	96.20
Char in V	6	34.33	6	50.79	6	115.16	6	34.52
EP	4	2.92	8	2.69	4	27.05	32	86.28
EV	1	199.08	2	49.01	1	44.70	8	86.86
EC	6	47.52	12	57.63	6	126.69	48	39.95
PV	4	26.65	4	3.88	4	39.95	4	17.99
PC	24	3.93	24	5.50	24	5.47	24	6.57
EPV	4	1.60	8	1.82	4	16.83	32	6.77
EPC	24	4.64	48	8.94	24	8.49	192	7.69
Within/S	320	7.00	480	5.67	56	9.80	1440	4.30

	2 vs 4	
Exemplar	1	3663.92
Pins	4	178.31
Verticality	1	161.88
Char in V	6	19.01
EP	4	177.14
EV	1	254.18
EC	6	33.09
PV	4	6.78
PC	24	6.99
EPV	4	9.00
EPC	24	9.97
Within/S	320	5.62

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

SHOULD CRITERIA FOR FIELD TESTS BE FORMULATED AS STATISTICAL HYPOTHESES?

Carl T. Russell
US Army Cold Regions Test Center
Fort Greely, Alaska

ABSTRACT. This paper presents for discussion an example typical of many faced by a statistician involved in the planning and analysis of Army field tests. The example raises philosophical and procedural questions concerning the degree to which statistical formalism (especially that of hypothesis testing) should be applied to field test planning and analysis in cases where numerical criteria are given. The author believes that although such criteria can serve as useful planning guides, it is usually foolish to behave as if such criteria can reduce test objectives to tests of statistical hypotheses. This belief has led the author to be quite casual in much of his statistical planning and analysis, as the example shows. The author solicits both a critique of his approach and suggested improvements.

I. INTRODUCTION AND STATEMENT OF PROBLEM. Winter temperatures below freezing are common in highly industrialized regions of Europe, Russia, Red China, and Korea. Since these are all regions where the U.S. Army could fight, the need for a test center for examining performance of Army personnel and materiel in the cold is obvious. The U.S. Army Cold Regions Test Center (CRTC) is located at Fort Greely, Alaska where temperatures are below 32°F more than 80 percent of the time during the winter months, averages 49 days per year below -25°F, averages 32 days per year below -40°F, and averages an annual low of -59°F. Although colder areas exist other than Fort Greely, no other accessible area in the United States is available to the U.S. Army for cold regions testing of military systems in the cold.

One system recently tested at CRTC was a pyrotechnic warning cartridge (aerial flare) consisting of a pyrotechnic whistle and three pyrotechnic stars. Previous experience with similar pyrotechnics had shown that storage, transportation, and firing at subzero temperatures could result in severe performance degradation. Two hundred cartridges from a total of approximately 2,500 prototype cartrides manufactured were provided for testing at CRTC. Late in test planning it was determined that each cartridge would be enclosed in a hermetically sealed can and that these cans would be packaged in cardboard cartons (unit packs) of eight. The cartridges arrived at CRTC in seven wooden boxes; six boxes contained four unit packs each and the seventh box contained one unit pack. The 25 unit packs were labeled "A" through "Y" where unit packs A through D came from the first box, E through H came from the second box, and so on. To avoid confusion during test execution (outside, at subzero temperatures) virtually all pretreatment and firing was planned in blocks of eight cartridges, and blocks were confounded with unit packs.

Among the criteria for test were numerical criteria for physical characteristics (cartridge size and weight) and numerical criteria for performance characterisitics (audible and visual signal timings and height of burst). CRTC's draft Detailed Test Plan (DTP) treated both types of criteria rather casually, translating them into decision rules (which could be interpreted as critical regions for appropriate hypothesis tests) rather than formulating them as hypotheses for test. Higher level review of the DTP produced recommendations for a more formal approach which were only cosmetically incorporated by CRTC.

In the author's view, data generated from the test indicated that a very informal approach to analysis was appropriate. In order to address the formal issues raised at higher level, however, a moderately thorough formal analysis was conducted and partially presented in the test report. But even the limited formality of the analysis presented in the test report seems (to the author) to detract from the simple test results by adding unnecessary statistical pedantry. Simple summary descriptions and graphical displays relating results to criteria would have sufficed.

What follows is a two-part example in which first the criteria for physical characteristics then the criteria for performance characteristics are examined in terms of the plans, comments, and analysis which they generated. The example is taken almost verbatim from test documentation, and it raises philosophical and procedural questions concerning the degree to which statistical formalism should be applied to field test planning and analysis in cases where numerical criteria are given. In a statistically simple world, criteria would accurately reflect all essential system characteristics, planning would direct testing at those criteria, and the only important test results would be whether or not the system met the criteria and at what level of statistical significance. The author does not deal with such a world, and to act as if he did would not only appear very foolish to the nonstatisticians with whom he works, but also adversely influence both test results and their presentation. However, the casual approach to criteria documented here has not proven to be acceptable throughout the Army statistical community, and the author solicits both a critique of his approach and suggested improvements.

II. PHYSICAL CHARACTERISTICS OF CARTRIDGES.

Criterion

The cartridge shall not exceed the following size and weight limitations: length - 30.5 centimeters (12 in); diameter - 40 millimeters (1.58 in); weight - 681 grams (1.5 lbs).

Planned Test Design/Procedure

One unit pack was to be selected at random and all eight cartridges removed from their sealed cans, weighed and measured. These cartridges were to be used in initial safety firings.

Planned Analysis

Physical dimensions and weights will be presented in tabular form. The criterion concerning physical characteristics will be considered met if the mean is less than the required limits.

Comments from Higher Level Review

Comparing the mean to the requirement is not an adequate analysis. The requirement states values which the item shall not exceed. The mean comparison would allow up to 50 percent of the items to exceed the requirement. The plan should include either of the following:

(a) A one-sided test of hypothesis at some α-risk level.

(b) Computation of a one-sided 90 or 95 percent tolerance limit at some 1-α confidence level and comparison of the limit to the required value.

Response to Higher Level Review

Add to planned analysis the sentence: "Confidence levels will be given where appropriate."

Actual Test Design/Procedure

As planned.

Results and Analysis

TABLE 1.--Cartridge Weights and Dimensions

Cartridge Number	Cap* Diameter (mm)	Cartridge Diameter (mm)	Length (cm)	Weight (g)
1	42.69	39.90	25.73	492.0
2	42.78	40.14*	25.72	494.5
3	41.92	39.98	25.73	492.5
4	42.04	39.96	25.70	491.0
5	41.40	39.90	25.75	493.0
6	41.92	40.06*	25.32	492.5
7	42.06	40.04*	25.73	492.0
8	41.78	40.02*	25.71	490.0
Mean	42.07	40.00	25.67	492.2
Std Deviation	0.46	0.08	0.14	1.3

*Exceeds criterion.

(a) The criterion concerning physical characteristics was considered met since all lengths and weights were well below the criterion values.

(b) Although the criterion did not specify whether cap diameter or cartridge diameter was to be less than 40 millimeters (1.5748 in), it was assumed that cartridge diameter was most relevant. The mean cartridge diameter was exactly 40.00 millimeters. Basing a confidence interval on the t-statistic, it can be stated with 99 percent confidence that the mean cartridge diameter is between 39.89 millimeters and 40.11 millimeters (1.570 in and 1.579 in).

III. PERFORMANCE CHARACTERISTICS OF CARTRIDGES.

Criteria

(a) Audible and visual signals of maximum intensity shall be produced within 7 seconds after the cartridge has been triggered to fire.

(b) When launched vertically upward, the cartridge shall have the capability of producing audible and visual signals at a minimum altitude above the launch site of 500 feet (152.4 meters).

(c) The duration of the audible signal and the duration of the visual signal shall each be at least 5 seconds. Signal durations greater than 7 seconds will not serve a useful purpose.

Planned Test Design/Procedure

Times were to be obtained by a ground observer using a stopwatch, and it was hoped that burst locations could be obtained using a video scoring system developed at CRTC. Experimental design complications arose since in addition to examining the performance criteria given above at varying temperatures and with or without transportation of rounds prior to firing, the visibility and audibility of signals at various distances from the launch site under various light and weather conditions were of concern, as was the ability of various personnel to fire the cartridge wearing various cold weather gear. The hope was to rotate seven observers through six observer positions and one firing position and to fire cartridges both day and night, both clear and snowing, and both calm and windy. In addition, safety considerations necessitated plans to fire at least 20 cartridges under any temperature condition prior to hand firing at that temperature. With 200 cartridges available for firing, -- essentially in blocks of eight -- no standard balanced design approach seemed feasible. The design approach taken was a traditional approach varying one selected factor at a time as summarized in table 2.

TABLE 2.--Planned Firing

Trial	Number[1] of Rounds	Temperature	Daylight	Snow	Wind (Kts)	Firer	Transportation	Human[3] Factors
1	4	>0°F	Y	N	<10	R[2]	N	Y
2	4	>0°F	N	N	<10	R[2]	N	Y
3	8	C1[5]	Y	N	<10	R[2]	Y	Y
4	16	C1[5]	Y	N	<10	R[2]	N	Y
5	8	C1[5]	N	N	<10	R[2]	Y	Y
6	8	C1[5]	N	N	<10	R[2]	N	Y
7	8	C1[5]	Y	Y	<10	R	N	Y
8	8	C1[5]	N	Y	<10	R	N	Y
9	8	C1[5]	Y	N	10-20	R	N	Y
10	8	C1[5]	Y	N	<10	A[4]	N	N
11	8	C1[5]	Y	N	<10	B[4]	N	N
12	8	C1[5]	Y	N	<10	C[4]	N	N
13	8	C1[5]	Y	N	<10	D[4]	N	N
14	8	C1[5]	Y	N	<10	E[4]	N	N
15	8	C1[5]	Y	N	<10	F[4]	N	N
16	8	C1[5]	Y	N	<10	G[4]	N	N
17	8	C2[5]	Y	N	<10	R[2]	Y	Y
18	16	C2[5]	Y	N	<10	R[2]	N	Y
19	8	C2[5]	N	N	<10	R[2]	Y	Y
20	8	C2[5]	N	N	<10	R[2]	N	Y
21	16	C3[5]	Y	N	<10	R[2]	N	Y
22	8	C3[5]	N	N	<10	R[2]	N	Y

[1] Eight cartridges available as replacement for any trial.
[2] Safety will be assessed for each firing; however remote firings are specifically to confirm the safety of firing at a new low temperature.
[3] Human factors evaluated as a part of each firing. Timing for CB protective ensemble will be recorded as well.
[4] Two rounds fired with chemical protective handgear, two rounds with trigger finger mittens, and four rounds fired with arctic mittens.
[5] Climatic design types as defined in AR 70-38:

C1 (basic cold) - -5°F to -25°F
C2 (cold) - -35°F to -50°F
C3 (severe cold)- below -60°F

Planned Analysis

(a) Criterion (a) will be met if 95 percent of the functioning cartridges provide audible and visual signals within 7 seconds after the cartridge has been triggered to fire.

(b) Criterion (b) will be met if 95 percent of the functioning cartridges produce audible and visual signals at a minimum altitude of 500 feet above the launch site.

(c) Criterion (c) will be met if 95 percent of the functioning cartridges produce audible and visual signals of a 5-second duration each.

Comment from Higher Level Review

The analysis should be more than "met and not met" decision rules. How will the data be analyzed and presented? The analysis should address such things as whistle duration vs. temperature and differences in cartridge performance after being transported.

Response to Higher Level Review

Add to planned analysis the paragraph: "The data will be examined for trends and exceptional values, and substantive findings will be discussed."

Actual Test Design/Procedure

Exceptionally warm weather after the test items arrived forced testing to be done quickly when cold temperatures were available. No firing was conducted in winds exceeding 5 knots. Observers were not rotated, but 25 different observers were used. Only sixteen cartridges were fired hand-held due to lateness of an appropriate safety release, and burst locations were not measured for the hand-held launches. The video scoring system was unavailable, so burst locations were determined from azimuth and elevation observations taken by qualified personnel at five ground observation points. Different personnel measured signal timings on different days. The actual design executed is summarized in table 3.

TABLE 3.--Actual Firing Matrix

Trial	Number of Rounds	Ambient Air Temperature (°F)[1]	Daylight	Remarks [2, 3]
1	4	26	Y	
2	4	26	N	9
3	8	-18	Y	10
4	16	-25	Y	
5	8	-12	N	9, 10
6	8	-10	N	
7	8	-10	Y	9
8	8	-10	N	
9	8	30	Y	4, 6, 8
10	8	-23	Y	
11	8	-22	Y	
12	8	-22	Y	
13	8	-20	Y	
14	8	-20	Y	
15	8	-16	Y	7
16	8	-17	Y	
17	8	-36	Y	11, 12
18	16	-39	Y	
19	8	-36	N	11
20	8	-39	N	
21	16	-22	Y	12, 13
22	8	-22	N	12, 13
23	8	-22	Y	5, 6, 12

1 Mean temperature.
2 Safety was assessed for each firing; however remote firings were specifically to confirm the safety of firing at low temperatures.
3 Observer human factors evaluated as a part of each firing.
4 Four rounds were fired hand-held with leather work gloves, two rounds were fired with trigger finger mittens, and two rounds were fired with arctic mittens.
5 Fired, hand-held, with leather work gloves.
6 Observers unwarned.
7 One round was "no test."
8 Rounds conditioned to -40°F.
9 Fired in falling snow.
10 Rounds transported 100 miles prior to firing.
11 Rounds transported 50 miles prior to firing.
12 Sound readings obtained.
13 Rounds conditioned to -55°F prior to firing.

Results and Analysis

(a) All cartridges exceeded 500 feet (152.4 meters) height of burst (HOB) except for one which had a 378 feet (155 meters) HOB. All functioned at peak trajectory and all residue extinguished prior to impact.

(b) Table 4 summarizes cartridge performance for the time to audible and visual signal initiation (maximum signal intensity was obtained almost simultaneously with signal initiation and could not be separately timed), audible and visual signal duration, HOB, and burst deviation from vertical. A round-by-round listing of performance data appears at appendix A.

TABLE 4.--Performance Data Summary

Statistic	Time to Signal Initiation (seconds)	Audible Signal Duration (seconds)	Visual Signal Duration (seconds)	HOB (feet)	Deviation from Vertical (degrees)
Minimum	3.9	6.1	6.0	378	0.9
Maximum	6.5	13.0	11.0	803	34.8[9]
Median	5.0	7.2	7.7	672	6.3
Mean	5.0	7.4	7.7	667	7.6
No. Observed	195[1]	195[2]	197[3]	183[4]	183[4]
No. Satisfactory	195[5]	195[6]	197[6]	182[7]	133[8], [9]
Percent Satisfactory	100.0	100.0	100.0	99.5	72.7

[1] Excludes rounds trial 7, Nos. 1 & 3; trial 9, Nos. 1 & 2; and trial 15, No. 8.
[2] Excludes rounds trial 9, No. 5; trial 11, Nos. 1, 4, 8; and trial 15, No. 8.
[3] Excludes rounds trial 2, No. 4; trial 9, No. 1; and trial 15, No. 8.
[4] Excludes 16 hand-held rounds of trials 9 and 23; as well as "no test" round of trial 15, No. 8.
[5] Number less than 7 seconds.
[6] Number greater than 5 seconds.
[7] Number greater than 500 feet. Low round trial 22, No. 2 was cold conditioned.
[8] Number less than 10 degrees from vertical.
[9] All but round trial 22, No. 2 were less than 19 degrees from vertical.

(c) Criteria (a), (b), and (c) were satisfied since the cartridge performed as designed. The data were examined for trends, but no substantial trends were found. A statistical discussion appears at appendix B.

APPENDIX A

Performance Characteristics

Trial	Round	Time to Signal Initiation (Sec)	Audible Signal Duration (Sec)	Visual Signal Duration (Sec)	Height of Burst[1] (Feet)	Deviation from Vertical[2] (Degrees)	Firing Order
1	5	4.0	7.0	7.8	755.3	2.0	1
1	6	4.0	7.0	7.0	803.4	2.7	2
1	7	4.0	6.2	8.0	752.0	2.4	3
1	8	4.5	6.3	7.8	750.8	4.0	4
2	3	5.0	6.9	7.0	771.5	3.0	5
2	4	5.0	8.0	NR	746.5	4.3	6
2	5	5.0	7.5	6.5	752.3	1.5	7
2	6	5.0	8.0	6.5	750.8	2.5	8
3	1	5.3	7.4	7.1	688.6	2.4	81
3	2	4.6	7.1	6.7	683.8	3.7	82
3	3	4.8	7.0	6.1	682.1	3.5	83
3	4	5.5	7.3	6.3	669.2	0.9	84
3	5	5.3	7.3	7.8	682.1	1.7	85
3	6	5.1	7.3	6.7	683.3	1.6	86
3	7	5.0	7.6	6.8	650.4	3.3	87
3	8	5.4	7.8	6.6	669.5	2.7	88
4	1	6.0	6.1	8.5	675.7	8.0	25
4	2	5.0	6.6	8.2	682.7	8.1	26
4	3	5.0	6.4	8.5	793.6	3.9	27
4	4	6.0	7.8	8.5	673.2	5.0	28
4	5	5.0	7.0	8.0	783.4	4.8	29
4	6	5.0	6.9	7.0	716.7	5.9	30
4	7	5.0	7.3	7.6	692.7	5.6	31
4	8	5.0	6.6	8.3	621.0	8.6	32
4	9	5.0	7.2	6.5	691.9	4.4	33
4	10	5.0	7.1	9.0	646.8	5.3	34
4	11	5.0	7.0	7.9	664.1	3.0	35
4	12	5.6	7.4	7.2	680.1	7.1	36
4	13	5.0	7.9	8.1	684.9	8.1	37
4	14	4.6	7.7	7.1	679.5	5.2	38
4	15	4.4	7.4	8.3	674.0	6.2	39
4	16	4.4	7.6	7.7	684.3	2.0	40

[1] Mean of values calculated for pairs of observers present (weighted by the number of observers involved). The mean was used because values for different pairs were in good agreement.

[2] Median of values calculated for pairs of observers present. The median was used because occasional large differences between pairs occurred in the data.

NR - Not Recorded

Performance Characteristics (Continued)

Trial	Round	Time to Signal Initiation (Sec)	Audible Signal Duration (Sec)	Visual Signal Duration (Sec)	Height of Burst[1] (Feet)	Deviation from Vertical[2] (Degrees)	Firing Order
5	1	4.8	7.0	7.8	710.4	1.5	105
5	2	4.9	7.0	7.3	654.8	6.0	106
5	3	5.2	7.1	8.1	650.8	5.5	107
5	4	5.5	6.9	7.6	699.9	6.3	108
5	5	4.6	8.3	7.6	717.0	1.8	109
5	6	5.0	7.7	7.0	722.4	9.2	110
5	7	5.2	7.8	7.8	686.0	8.4	111
5	8	5.0	6.6	7.0	717.3	1.6	112
6	1	5.5	6.6	6.0	662.1	3.2	9
6	2	5.0	7.2	8.1	687.3	4.9	10
6	3	5.0	6.6	8.1	656.4	3.5	11
6	4	5.5	7.0	7.2	676.0	1.3	12
6	5	5.5	7.5	8.0	716.6	2.4	13
6	6	5.0	7.0	7.8	737.9	6.8	14
6	7	5.2	7.0	8.0	715.7	5.4	15
6	8	4.2	7.5	7.5	672.9	6.1	16
7	1	NR	6.8	8.3	672.8	6.0	113
7	2	5.8	6.6	7.6	647.7	7.9	114
7	3	NR	6.8	7.6	686.5	5.0	115
7	4	4.5	7.3	8.0	663.0	7.5	116
7	5	4.7	7.0	7.8	546.9	9.0	117
7	6	4.7	7.8	7.7	691.2	2.6	118
7	7	4.5	7.8	7.6	634.1	5.4	119
7	8	5.3	6.1	7.8	677.1	1.9	120
8	1	5.0	6.6	8.5	707.9	5.9	17
8	2	5.0	6.7	7.8	690.8	2.7	18
8	3	5.2	6.9	7.2	695.4	2.8	19
8	4	5.0	6.7	8.0	693.3	5.0	20
8	5	5.2	6.9	7.8	651.2	6.2	21
8	6	5.8	7.5	7.2	703.3	3.5	22
8	7	5.0	7.0	8.0	706.4	3.7	23
8	8	5.0	6.7	7.2	700.4	5.3	24

[1] Mean of values calculated for pairs of observers present (weighted by the number of observers involved). The mean was used because values for different pairs were in good agreement.

[2] Median of values calculated for pairs of observers present. The median was used because occasional large differences between pairs occurred in the data.

NR - Not Recorded

Performance Characteristics (Continued)

Trial	Round	Time to Signal Initiation (Sec)	Audible Signal Duration (Sec)	Visual Signal Duration (Sec)	Height of Burst[1] (Feet)	Deviation from Vertical[2] (Degrees)	Firing Order
9	1	NR	7.0	NR	NR	NR	193
9	2	NR	6.5	7.0	NR	NR	194
9	3	6.5	7.0	11.0	NR	NR	195
9	4	5.5	7.0	7.0	NR	NR	196
9	5	6.0	NR	6.5	NR	NR	197
9	6	6.0	7.2	7.0	NR	NR	198
9	7	6.0	6.8	6.0	NR	NR	199
9	8	6.5	6.8	7.0	NR	NR	200
10	1	5.0	7.1	7.8	627.2	7.8	41
10	2	5.6	7.1	7.9	717.8	7.6	42
10	3	5.4	6.9	7.8	671.8	13.6	43
10	4	5.0	6.4	7.9	712.9	4.0	44
10	5	5.0	7.3	8.5	686.8	3.0	45
10	6	5.4	7.0	7.3	682.8	3.6	46
10	7	5.1	7.2	7.3	672.7	3.4	47
10	8	4.7	7.7	6.7	689.8	4.2	48
11	1	5.0	NR	8.1	683.9	4.7	49
11	2	4.5	7.1	7.3	628.9	8.1	50
11	3	5.5	6.9	7.2	673.7	9.4	51
11	4	5.5	NR	8.3	659.8	2.3	52
11	5	4.3	7.6	7.6	630.2	5.8	53
11	6	5.1	7.1	8.3	574.0	7.7	54
11	7	5.0	7.1	7.8	668.3	6.5	55
11	8	5.1	NR	8.4	662.5	4.3	56
12	1	4.5	6.7	7.6	629.3	4.9	57
12	2	5.3	6.9	7.7	696.4	3.4	58
12	3	5.1	6.7	7.1	686.5	5.2	59
12	4	5.0	7.2	7.3	694.0	1.8	60
12	5	5.1	7.3	7.3	649.4	2.5	61
12	6	6.1	6.6	7.3	662.3	12.2	62
12	7	5.1	7.7	8.3	656.8	4.6	63
12	8	4.6	6.5	8.1	677.1	4.0	64

[1] Mean of values calculated for pairs of observers present (weighted by the number of observers involved). The mean was used because values for different pairs were in good agreement.

[2] Median of values calculated for pairs of observers present. The median was used because occasional large differences between pairs occurred in the data.

NR - Not recorded

Performance Characteristics (Continued)

Trial	Round	Time to Signal Initiation (Sec)	Audible Signal Duration (Sec)	Visual Signal Duration (Sec)	Height of Burst[1] (Feet)	Deviation from Vertical[2] (Degrees)	Firing Order
13	1	4.7	7.3	7.8	649.1	5.5	65
13	2	5.5	7.0	7.3	662.9	6.5	66
13	3	5.0	8.1	7.5	662.0	11.3	67
13	4	5.2	6.8	8.5	594.2	11.1	68
13	5	4.5	7.5	7.9	595.9	8.8	69
13	6	4.5	8.0	6.9	661.4	4.7	70
13	7	4.7	7.5	7.5	695.2	5.8	71
13	8	5.0	6.9	8.1	643.1	2.1	72
14	1	5.6	6.3	7.1	528.4	9.4	73
14	2	4.2	7.5	7.3	605.0	2.7	74
14	3	5.2	7.7	7.3	623.7	3.0	75
14	4	4.9	7.9	6.9	670.9	2.8	76
14	5	4.2	7.3	7.0	605.7	5.0	77
14	6	4.4	7.3	6.9	654.0	2.1	78
14	7	4.9	8.0	7.4	635.9	6.6	79
14	8	4.9	7.7	8.3	670.1	6.2	80
15	1	4.9	8.0	6.1	655.7	1.2	89
15	2	5.4	7.3	6.7	688.7	3.6	90
15	3	5.4	8.1	7.1	704.0	10.0	91
15	4	5.2	7.7	6.8	622.3	5.2	92
15	5	5.0	7.3	6.7	703.0	3.1	93
15	6	4.8	7.0	6.7	653.3	4.8	94
15	7	5.6	7.8	6.8	608.4	6.8	95
15	8	NR	NR	NR	NR	NR	96
16	1	4.6	7.3	6.9	651.5	1.8	97
16	2	5.6	7.7	7.9	693.4	2.7	98
16	3	4.8	6.9	7.0	663.0	6.6	99
16	4	4.9	7.7	6.9	643.0	4.9	100
16	5	6.4	6.8	6.9	687.4	2.3	101
16	6	4.5	7.3	6.9	607.9	5.4	102
16	7	4.9	7.7	6.3	666.0	4.7	103
16	8	5.3	7.1	6.1	686.9	6.5	104

[1] Mean of values calculated for pairs of observers present (weighted by the number of observers involved). The mean was used because values for different pairs were in good agreement.

[2] Median of values calculated for pairs of observers present. The median was used because occasional large differences between pairs occurred in the data.

NR - Not recorded

Performance Characteristics (Continued)

Trial	Round	Time to Signal Initiation (Sec)	Audible Signal Duration (Sec)	Visual Signal Duration (Sec)	Height of Burst[1] (Feet)	Deviation from Vertical[2] (Degrees)	Firing Order
17	1	4.9	6.9	7.0	641.2	13.1	153
17	2	5.3	7.5	7.9	666.6	11.9	154
17	3	4.9	8.1	8.5	668.5	10.5	155
17	4	4.7	7.7	7.3	638.9	15.1	156
17	5	4.6	8.5	7.3	585.8	15.3	157
17	6	4.2	7.8	7.6	587.3	8.9	158
17	7	3.9	7.6	8.1	600.1	8.9	159
17	8	5.4	7.5	7.8	678.4	10.8	160
18	1	5.0	7.0	7.5	696.3	13.8	129
18	2	4.7	7.0	7.6	601.3	18.5	130
18	3	5.5	7.0	8.2	598.7	18.6	131
18	4	4.9	7.0	7.3	699.0	3.4	132
18	5	4.9	7.0	8.2	609.6	10.1	133
18	6	5.1	7.0	8.1	689.4	10.1	134
18	7	4.8	7.0	7.2	689.0	10.8	135
18	8	5.7	7.0	7.7	677.7	8.6	136
18	9	4.6	7.0	7.9	652.7	13.9	137
18	10	5.1	7.0	7.7	725.8	11.0	138
18	11	4.9	7.0	7.7	644.9	15.0	139
18	12	4.9	7.0	7.8	719.0	9.9	140
18	13	4.7	7.0	7.2	647.6	15.2	141
18	14	5.1	7.0	8.1	661.4	9.3	142
18	15	5.5	7.0	10.6	707.8	9.3	143
18	16	5.3	7.0	7.8	627.9	2.6	144
19	1	4.9	8.5	8.3	708.9	13.4	145
19	2	5.2	7.2	8.7	682.2	9.2	146
19	3	4.8	7.7	8.9	670.0	11.0	147
19	4	5.3	7.5	8.3	637.2	10.3	148
19	5	5.2	7.9	7.7	670.2	8.0	149
19	6	5.1	6.9	8.0	649.6	17.1	150
19	7	5.4	7.2	7.9	684.5	6.7	151
19	8	4.6	7.0	7.9	646.1	14.0	152

[1] Mean of values calculated for pairs of observers present (weighted by the number of observers involved). The mean was used because values for different pairs were in good agreement.

[2] Median of values calculated for pairs of observers present. The median was used because occasional large differences between pairs occurred in the data.

Performance Characteristics (Continued)

Trial	Round	Time to Signal Initiation (Sec)	Audible Signal Duration (Sec)	Visual Signal Duration (Sec)	Height of Burst[1] (Feet)	Deviation from Vertical[2] (Degrees)	Firing Order
20	1	4.8	8.0	7.4	722.2	10.5	121
20	2	4.8	7.0	7.0	684.5	14.8	122
20	3	5.0	7.0	7.1	646.1	16.7	123
20	4	4.9	8.0	8.4	707.5	13.3	124
20	5	5.7	7.0	8.0	671.9	12.3	125
20	6	5.0	7.0	8.0	698.1	11.7	126
20	7	4.6	7.0	7.8	753.2	11.3	127
20	8	5.3	7.0	7.0	699.8	15.4	128
21	1	4.6	8.1	8.6	638.1	8.9	177
21	2	4.9	7.9	8.7	634.7	9.5	178
21	3	5.5	7.3	8.0	686.1	7.9	179
21	4	5.5	7.9	8.0	659.9	8.8	180
21	5	4.5	8.1	7.9	663.5	12.8	181
21	6	4.9	7.9	7.7	647.2	15.4	182
21	7	5.0	7.7	8.5	673.7	12.3	183
21	8	5.3	8.0	8.0	679.4	9.4	184
21	9	4.8	7.6	8.7	578.7	15.4	185
21	10	4.0	7.9	7.5	607.5	15.2	186
21	11	4.9	7.7	8.0	628.9	14.8	187
21	12	4.8	7.3	8.3	599.9	16.5	188
21	13	4.9	7.7	8.0	545.5	18.1	189
21	14	4.8	8.5	8.9	657.4	14.9	190
21	15	4.6	8.0	7.8	618.2	9.2	191
21	16	4.9	7.3	8.3	639.4	14.5	192
22	1	5.6	7.0	8.0	644.9	14.7	169
22	2	4.7	7.4	8.1	378.5	34.8	170
22	3	4.0	7.5	9.0	576.8	16.6	171
22	4	4.8	8.0	9.4	649.2	10.3	172
22	5	4.4	7.7	8.9	617.5	14.9	173
22	6	4.6	8.0	8.4	656.9	9.6	174
22	7	5.1	7.5	8.9	616.1	12.7	175
22	8	5.2	7.5	9.1	649.1	7.7	176

[1] Mean of values calculated for pairs of observers present (weighted by the number of observers involved). The mean was used because values for different pairs were in good agreement.

[2] Median of values calculated for pairs of observers present. The median was used because occasional large differences between pairs occurred in the data.

Performance Characteristics (Continued)

Trial	Round	Time to Signal Initiation (Sec)	Audible Signal Duration (Sec)	Visual Signal Duration (Sec)	Height of Burst[1] (Feet)	Deviation from Vertical[2] (Degrees)	Firing Order
23	1	4.4	11.0	10.0	NR	NR	161
23	2	4.6	10.0	6.0	NR	NR	162
23	3	4.9	12.0	7.0	NR	NR	163
23	4	4.5	10.0	7.0	NR	NR	164
23	5	4.8	13.0	8.0	NR	NR	165
23	6	4.6	11.0	7.0	NR	NR	166
23	7	4.1	11.0	8.0	NR	NR	167
23	8	4.7	10.0	7.0	NR	NR	168

[1] Mean of values calculated for pairs of observers present (weighted by the number of observers involved). The mean was used because values for different pairs were in good agreement.

[2] Median of values calculated for pairs of observers present. The median was used because occasional large differences between pairs occurred in the data.

NR - Not recorded

APPENDIX B

Statistical Analysis of Performance Data

1. Summary. Except for one cartridge (trial 22, round 2) which burst below 500 feet, all cartridges performed as designed regardless of test conditions. In order to determine whether cartridge performance was substantially affected by changes in test conditions, however, performance data on each of the five variables were analyzed using analysis of variance. No substantial trends were discovered. Nevertheless, there were statistically significant differences for each variable between data taken under different test conditions. In particular, the cold conditioned rounds burst significantly lower (40 feet) and significantly farther from vertical (6 degrees) than the average round while rounds from the first box of cartridges fired burst significantly higher (45 feet) and significantly closer to vertical (3 degrees) than the average round.

2. Linear model used. Since (except for the first eight rounds) cartridges from the same unit pack of cartridges were always subjected to the same treatment (e.g., cold conditioning) and fired together under nearly identical test conditions, any trends due to changes in test conditions would be detectable only from unit pack to unit pack. Differences in performance within a unit pack could only be attributed to random variation. Thus, for each response variable, Y, the linear model:

$$Y_{ik} = \mu + \sum_{i=1}^{n} \beta_i + \varepsilon_{ik}$$

was used to estimate the difference, β_i, from the overall mean, μ, for each of the n unit packs. Least squares estimation was used, and the β's were subjected to the side condition:

$$\sum_{i=1}^{n} n_i \beta_i = 0$$

where n_i, was the number of available observations (for the reponse variable under consideration) on the i^{th} unit pack. This standard parameterization made the β-estimators into contrasts and enabled consideration of other selected contrasts (in particular, that for cold conditioned rounds) as linear combinations of the β's.

3. Results. Five response variables were analyzed in the context of this linear model: time to signal initiation in seconds, audible signal duration in seconds, visual signal duration in seconds, height of burst in feet, and deviation from vertical in degrees. In addition, height of burst and deviation from vertical were analyzed both with the low round (trial 22, round 2) included and with the low round excluded. An analysis of variance was performed on each variable (table 1), and the F-value for testing the null hypothesis of no difference in performance from unit pack to unit pack was

greater than the critical F-value for 0.005 significance in every case.[1] Estimates of the β's (unit pack effects) as well as estimates of selected contrasts found to be of interest appear in table 2 along with the calculated values of Student's t for testing whether the coefficients are zero.

4. Discussion. Burst locations for the 183 remotely fired rounds are shown to scale in figure 1. Although all but one round (a cold conditioned round) burst above 500 feet, the cold conditioned rounds (even excluding the low round) tended to burst significantly lower and significantly farther from vertical than the average round. However, the rounds from box 1 (including the initial 8 rounds and 24 additional rounds of no particular distinction) tended to burst significantly higher and significantly closer to vertical than the average round. This indicates that physical/chemical differences between unit packs or boxes may have been significant or that some other unmeasured test variables (and not cold conditioning) may have significance. However, none of the observed differences appeared to be substantial. There were no clear trends with temperature for any variable (figures 2a through 6a) but there were some indications of trends with firing order (figure 2b through 6b). The virtually constant audible signal durations measured on trial 23 (firing order 160 through 168; these were all of the rounds fired on 18 February) seem to show merely a difference in measurement technique. Likewise, the trends in burst location might be partially due to day to day variations in equipment or to meteorological conditions effecting line of sight.

[1] The assumptions of normality and homogeneity of variance for the error terms which are necessary for the calculated F-values to follow the F-distribution, were not satisfied. In particular, it is obvious from figures 2 through 6, that times measured were nearly constant for some unit packs, and the distribution of the residuals for degrees deviation from vertical was skewed. Variable transformation and sensitivity analyses on subsets of the data which excluded the nearly constant times could have been performed but were not; it was felt that any changes in the results obtained by using more elaborate analysis would be minimal and would not justify the cost of additional analysis.

TABLE 1.--Analysis of Variance Summary

Statistic	Time to Signal Initiation	Audible Signal Duration	Visual Signal Duration	Height of Burst: With Low Round	Height of Burst: No Low Round	Deviation From Vertical: With Low Round	Deviation From Vertical: No Low Round
Total Sum of Squares							
Degrees of Freedom	194	196	194	182	181	182	181
Value	41.15	113.04	158.54	438071	354329	4345	3600
Regression Sum of Squares							
Degrees of Freedom	24	24	24	22	22	22	22
Value	13.80	41.01	122.39	205604	177179	2690	2385
Error Sum of Squares							
Degrees of Freedom	170	172	170	160	159	160	159
Value	27.36	72.04	36.16	232467	177150	1654	1215
F-Value	3.57	4.08	23.98	6.43	7.23	11.83	14.19
R-Squared	0.34	0.36	0.77	0.47	0.50	0.62	0.66
Mean	5.00	7.66	7.42	667.08	668.66	7.55	7.40
Standard Error of Mean	0.03	0.05	0.03	2.82	2.47	0.24	0.20

TABLE 2.--Parameter Estimates With Corresponding t-Statistics

Factor Trial: Rounds; Unit Pack & Box Number	Variable Analyzed Time to Signal Initiation Coeff	t	Audible Signal Duration Coeff	t	Visual Signal Duration Coeff	t	Height of Burst With Low Round Coeff	t	Height of Burst No Low Round Coeff	t	Deviation From Vertical With Low Round Coeff	t	Deviation From Vertical No Low Round Coeff	t
1: 5-8, 2: 3-6; A1	-0.441	-3.17	-0.309	-1.93	-0.432	-1.60	93.2	7.07*	91.6	7.94*	-4.74	-4.26	-4.59	-4.80
6: 1-8; B1	0.109	0.79	-0.374	-2.34	-0.073	-0.33	23.5	1.79	22.0	1.90	-3.35	-3.01	-3.20	-3.34
8: 1-8; C1	0.147	1.05	-0.544	-3.41	-0.052	0.23	26.5	2.01	24.9	2.16	-3.17	-2.85	-3.02	-3.16
4: 1-8; D1	0.247	1.78	-0.586	-3.67	0.414	1.85	37.8	2.87	36.2	3.14	-1.32	-1.19	-1.17	-1.22
4: 9-16; E2	-0.126	-0.92	-0.011	-0.07	0.064	0.29	8.6	0.65	7.0	0.61	-2.40	-2.16	-2.25	-2.35
15: 1-8; F2	0.183	1.23	-0.176	1.03	-0.961	-4.00	-4.9	-0.34	-6.5	-0.52	-2.60	-2.18	-2.45	-2.39
10: 1-8; G2	0.147	1.05	-0.335	-2.11	-0.011	-0.05	15.6	1.19	18.9	1.61	-1.64	-1.47	2.31	1.73
11: 1-8; H2	-0.003	-0.02	-0.254	-1.30	0.214	0.96	-19.4	-1.47	-21.0	-1.82	-1.45	-1.30	-1.30	-1.36
12: 1-8; I3	0.097	0.70	-0.474	-2.97	-0.073	-0.33	1.9	0.15	0.3	0.03	-2.73	-2.46	-2.58	-2.70
13: 1-8; J3	-0.116	-0.83	-0.035	-0.23	-0.027	0.12	-21.6	-1.64	-23.2	-2.01	-0.57	-0.51	-0.42	-0.44
14: 1-8; K3	-0.216	-1.55	0.039	0.24	-0.386	-1.72	-41.9	-3.18	-43.5	-3.77	-2.83	-2.55	-2.68	-2.81
16: 1-8; L3	0.121	0.88	-0.111	-0.70	-0.798	-3.56	-4.7	-0.36	-6.3	-0.54	-3.18	-2.86	-3.03	-3.17
20: 1-8; M4	0.009	0.07	-0.174	-1.09	-0.073	-0.33	30.8	2.34	29.2	2.53	5.70	5.12	5.85	6.12*
18: 1-8; N4	0.072	0.52	-0.424	-2.65	0.054	0.29	-9.5	-0.72	-11.0	-0.96	4.18	3.76	4.33	4.53
18: 9-16; O4	0.008	0.07	-0.424	-2.65	0.439	1.96	6.3	0.48	4.7	0.41	3.22	2.90	3.37	3.53
17: 1-8; P4	-0.266	-1.91	0.276	1.73	0.027	0.12	-33.7	-2.56	-35.3	-3.06	4.27	3.84	4.42	4.62
7: 1-8; Q5	-0.086	-0.54	-0.399	-2.50	0.139	0.62	-14.7	-1.11	-16.3	-1.41	-1.88	-1.69	-1.73	-1.81
3: 1-8; R5	0.122	0.88	-0.074	-0.46	-0.898	-4.01	9.1	0.69	7.5	0.65	-5.07	-4.56	-4.92	-5.15
5: 1-8; S5	0.022	0.16	-0.124	-0.78	-0.136	-0.61	27.7	2.11	26.1	2.27	-2.51	2.26	-2.36	-2.47
19: 1-8; T5	0.060	0.43	0.064	0.40	0.552	2.46	1.5	0.11	-0.1	-0.01	3.67	3.30	3.82	4.00
23: 1-8; U6	-0.428	-3.08	3.576	22.40*	-0.161	-0.72	**	**	**	**	**	**	**	**
21: 9-16; V6	-0.291	-2.09	0.326	2.04	0.527	2.35	-57.6	-4.37	-59.2	-5.13	7.27	6.54*	7.42	7.77*
22: 1-8; W6	-0.203	-1.45	0.151	0.95	1.064	4.75	-68.6	-5.20	-38.7	-3.13	7.61	6.85*	4.96	4.84
21: 1-8; X6	0.022	0.16	0.439	2.75	0.514	2.29	-6.7	-0.51	-8.3	-0.72	3.18	2.86	3.33	3.49
9: 1-8; Y7	1.050	6.70*	-0.524	-3.05	-0.304	-1.26	**	**	**	**	**	**	**	**
Cold Conditioned	-0.157	-2.05	0.305	3.46	0.702	5.57	-44.3	6.11*	-35.3	-5.42	6.02	9.84*	5.25	9.75*
Box 1 (No A1)	0.168	2.19	-0.502	-5.69	0.131	1.05	29.3	4.04	27.7	4.36	-2.61	-4.27	-2.45	-4.68
Box 1	0.055	0.24	-0.483	-6.08	0.004	0.03	45.3	7.40*	43.7	8.16*	-3.14	-6.65*	-2.99	-6.74*

*Significant at α = .05 using Sheffe's S test.
**Hand-held rounds (no data).

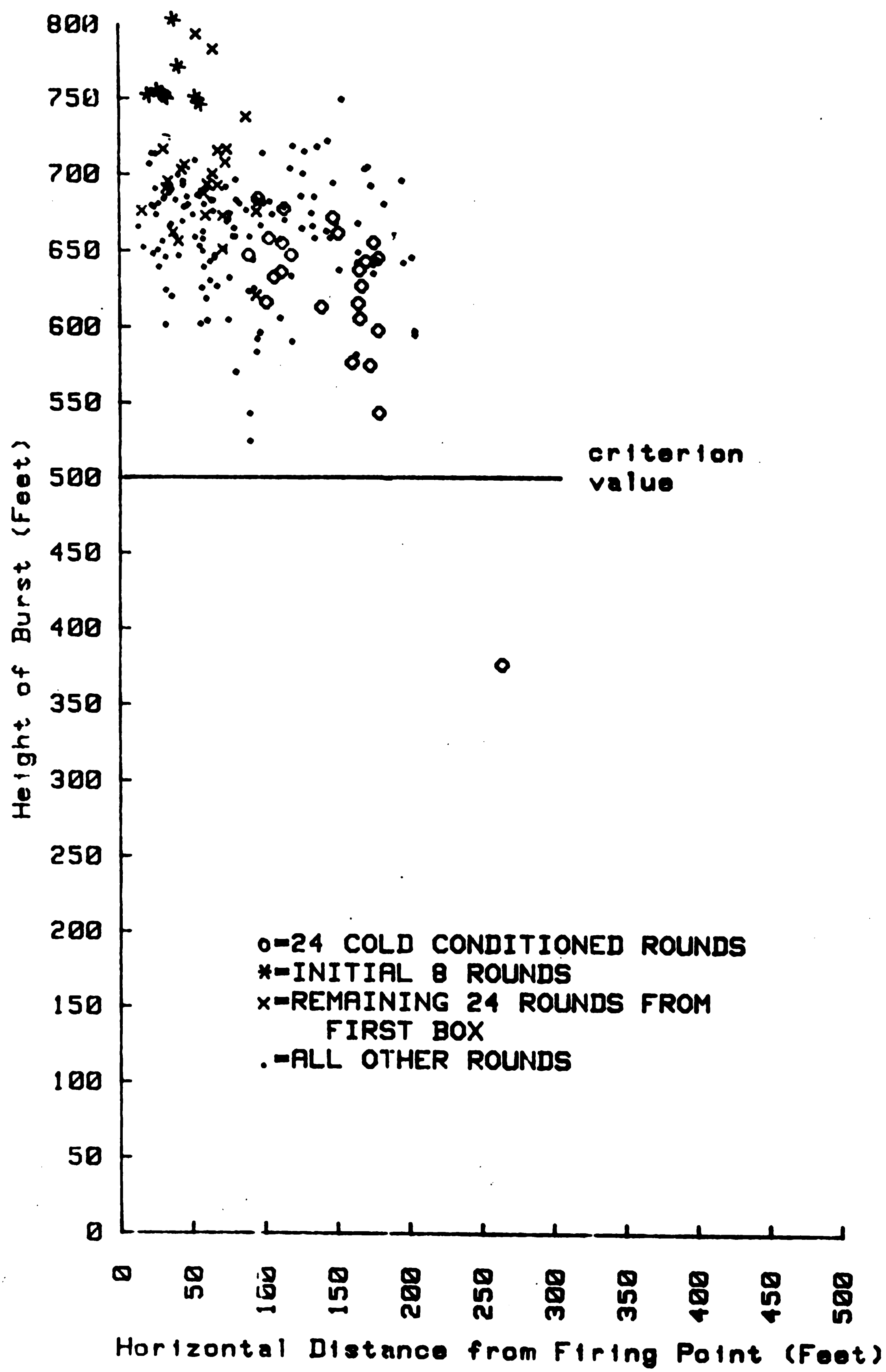


FIGURE 1.--Burst Locations

B-5

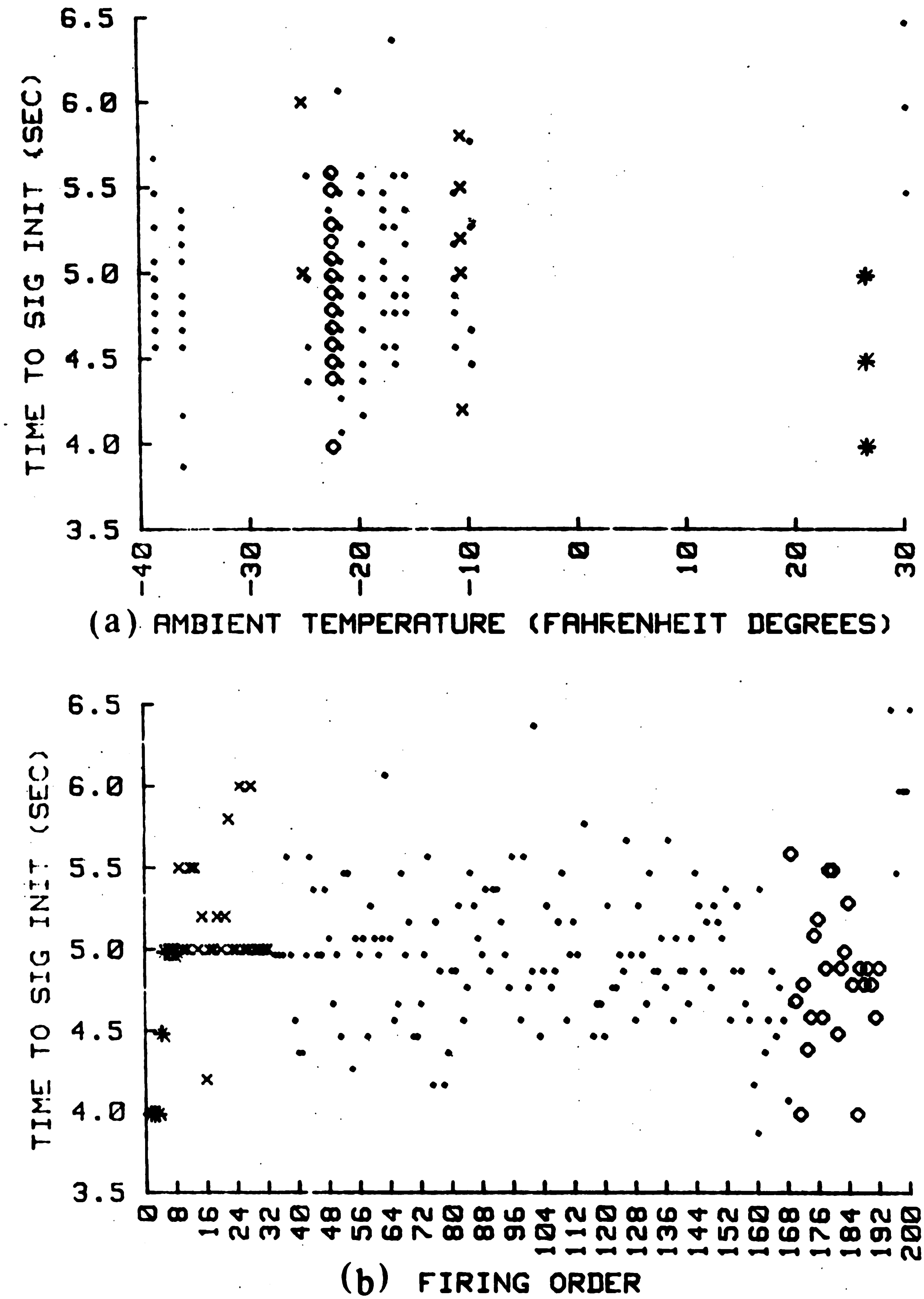


FIGURE 2.--Time to Signal Initiation Versus
(a) Ambient Temperature and
(b) Firing Order (for Legend, See Figure 1)

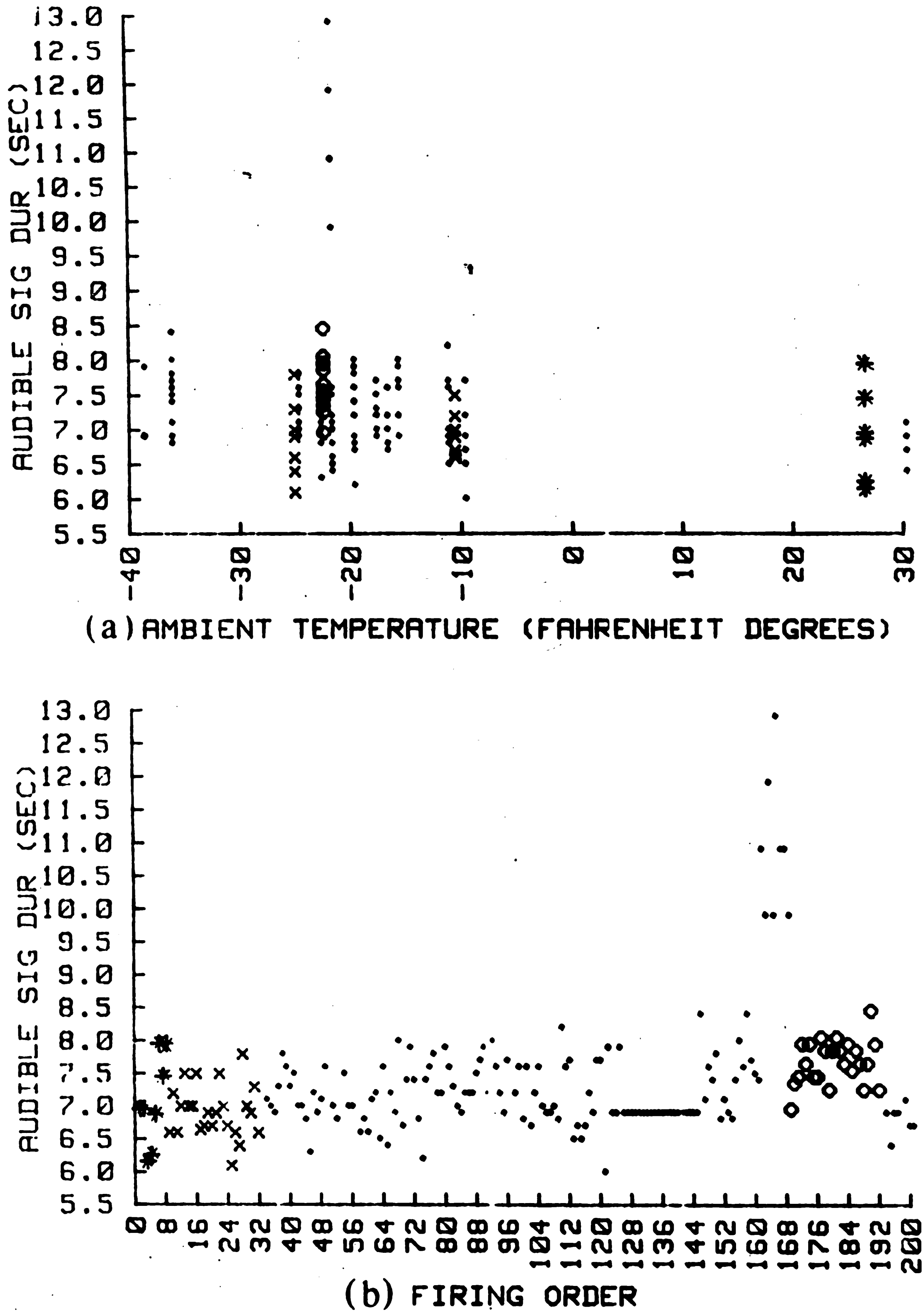
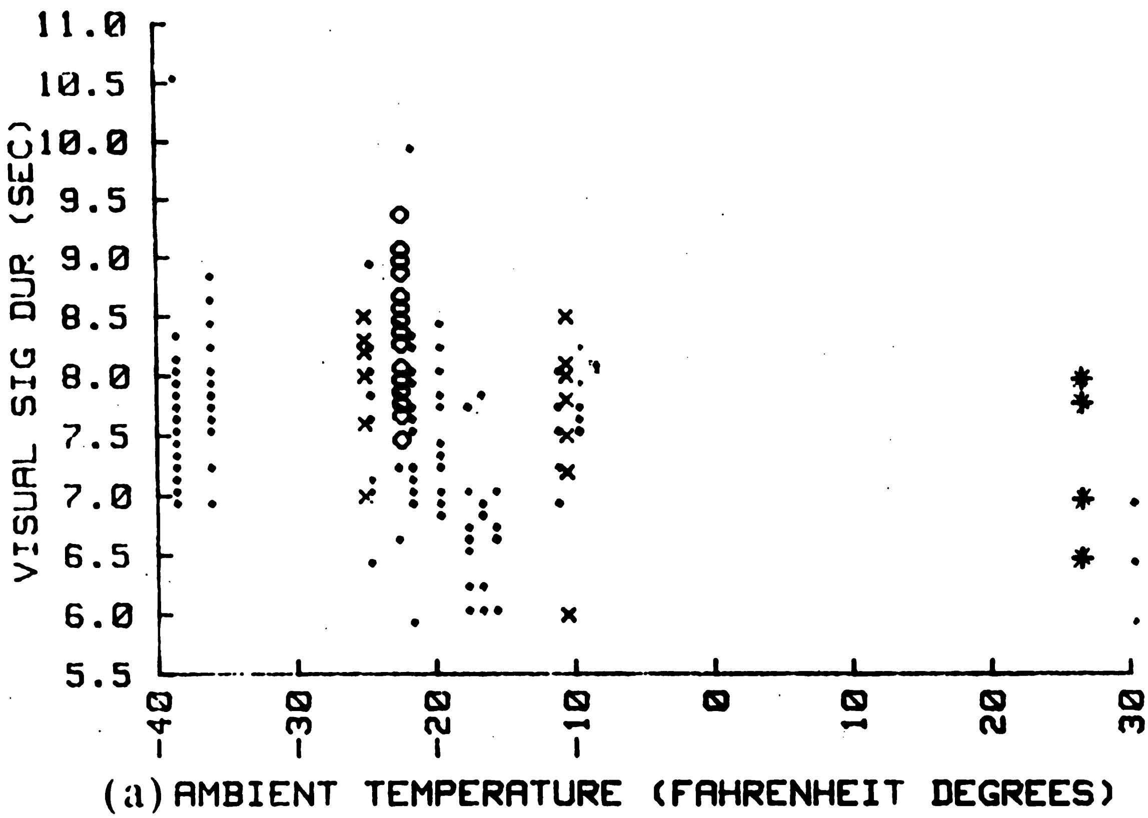


FIGURE 3.--Audible Signal Duration Versus
(a) Ambient Temperature and
(b) Firing Order (for Legend, See Figure 1)



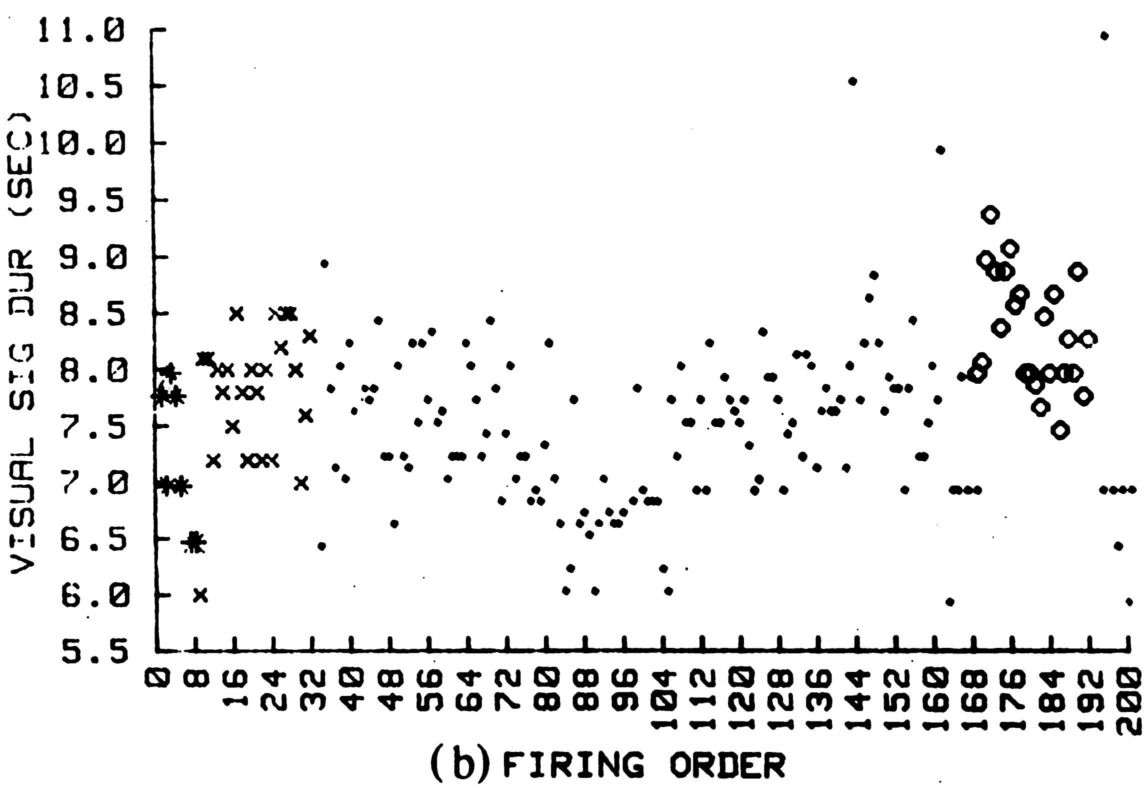
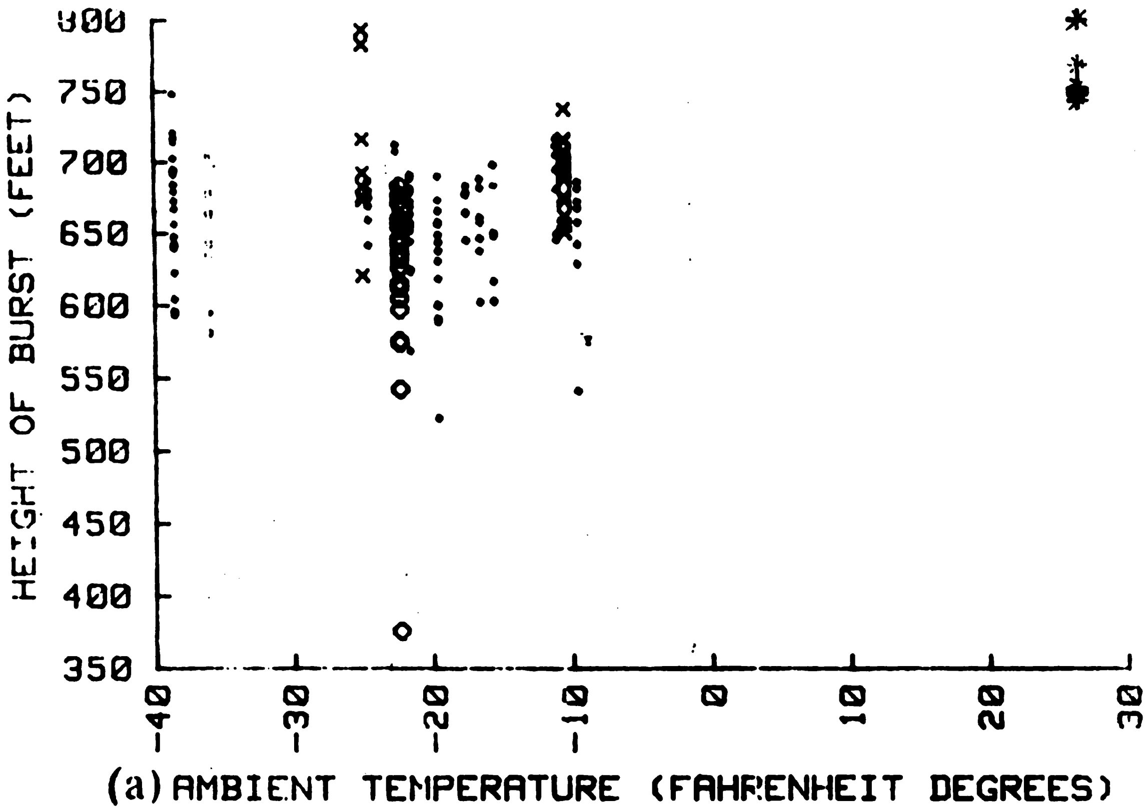


FIGURE 4.--Visual Signal Duration Versus
(a) Ambient Temperature and
(b) Firing Order (for Legend, See Figure 1)



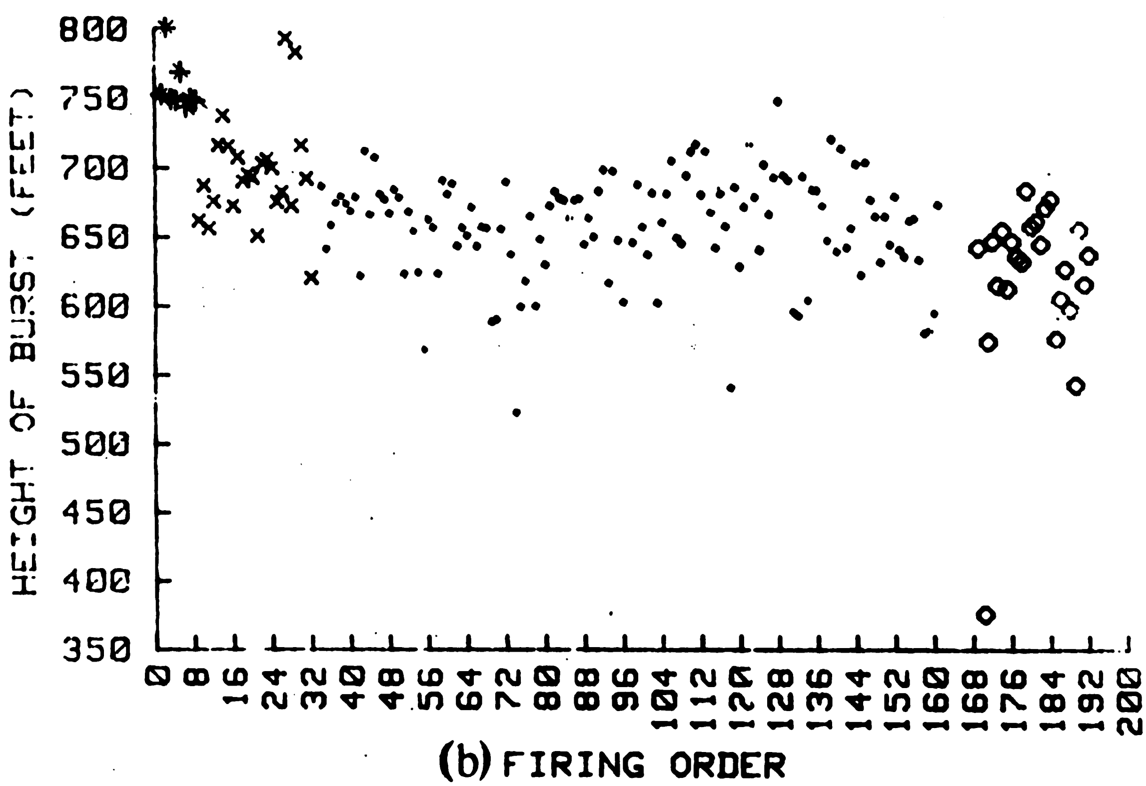


FIGURE 5.--Height of Burst Versus
(a) Ambient Temperature and
(b) Firing Order (for Legend, See Figure 1)

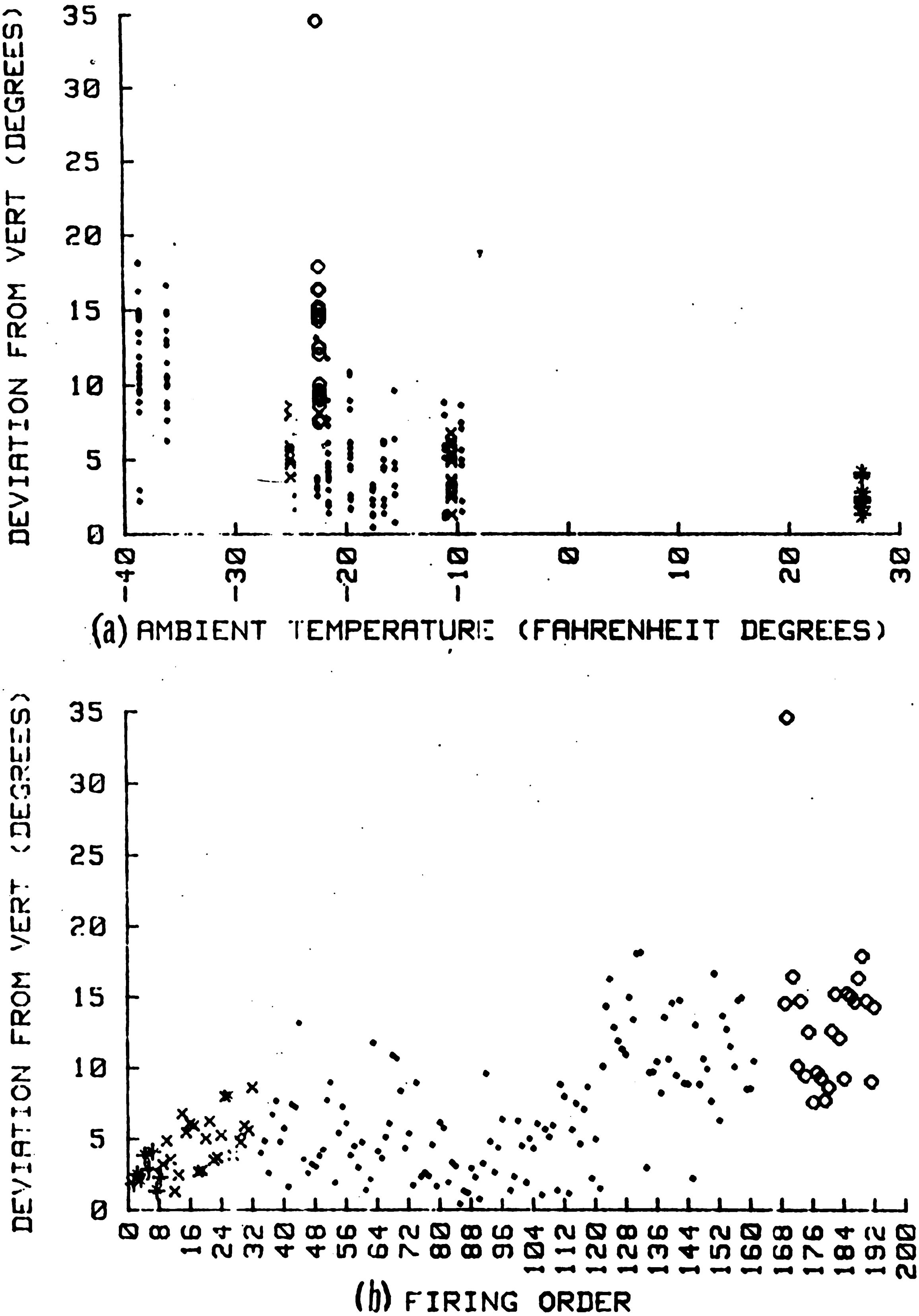


FIGURE 6.--Deviation From Vertical Versus
(a) Ambient Temperature and
(b) Firing Order (for Legend, See Figure 1)

STATISTICAL TESTING OF LARGE COMPLEX COMPUTER SIMULATION MODELS

Carl B. Bates

US Army Concepts Analysis Agency

Bethesda, Maryland 20814

ABSTRACT. In response to increasing requirements for communications/electronics (EW) analyses, the US Army Concepts Analysis Agency (CAA) undertook the development of a Divisional Electronic Warfare Combat (DEWCOM) Model. The model has been developed, and it contains approximately 20,000 lines of code and 250 input variables. It simulates two-sided play of combat, communications, and electronic warfare in conventional warfare with close air support. Before the model is committed to use in support of CAA studies, it is undergoing test and evaluation. The test and evaluation of the model is being conducted in two phases--data base development and verification. A part of the verification phase involved a sensitivity analysis of model output to changes in model input. A group-screening approach was applied, and a resolution V experimental design was employed. The experimental design and the analyses results are presented and discussed.

1. INTRODUCTION. Due to increasing needs for the performance of EW analyses, CAA undertook the development of an EW combat simulation model. Model development was contracted out; the developed model, Divisional Electronic Warfare Combat (DEWCOM) Model, was delivered to CAA during the summer of 1980. The model is a fully computerized, stochastic simulation model. It simulates conventional ground warfare with close air support. Two-sided play of combat, communication, and

electronic warfare is provided. The model has a variable force resolution capability; representation of up to five echelons, typically corps to company, can be simulated. The model is organized into three interacting modules as illustrated in Figure 1. One module performs communication operations, another module performs tactical operations, and the third module performs electronic warfare operations. The modules are driven by a list of events which specify actions to be taken at a scheduled time in the future. The list of events is initialized at the start of the simulation by the user. Thereafter, the list is continuously updated as a consequence of actions occurring in the simulation. This causes additional events to be scheduled for future combat, communication, and EW operations. The process continues until a user specified termination time. At time intervals specified by the user, model reports are produced summarizing the actions which have been simulated. Naturally, before the model could be approved for use in Agency studies, the model had to be tested. Consequently, a test and evaluation effort was initiated.

2. TEST METHODOLOGY

a. Objective. The objective of the test and evaluation was to establish that the model, given the appropriate inputs, accurately represents the performance of communications and EW systems in a tactical environment and portrays realistic combat outcomes.

The test and evaluation was conducted in two phases:

i. Data base development.

ii. Verification testing to determine (1) if the various functions in the model performed as intended and (2) if the model portrays an accurate representation of real-world systems.

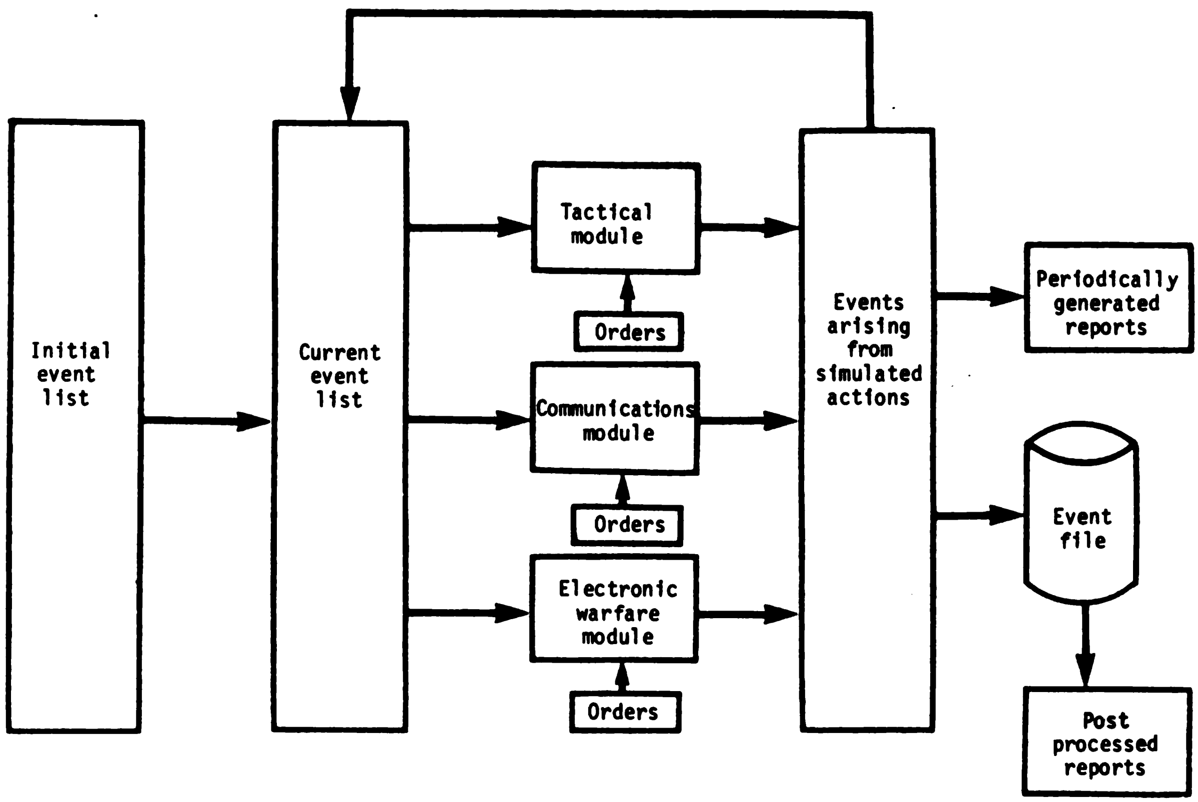


Figure 1. DEWCOM Model Organization

b. <u>Data Base</u>. The data base consisted of a real-world base case (Blue brigade versus a Red division) consistent with preselected equipment, doctrine, and tactics. The data base was an unclassified adaptation of the Force Electronic Warfare/Tactical SIGINT (FEWTS) Study classified data base. The FEWTS data were converted to the DEWCOM input format and additional data required by the DEWCOM were added such as equipment performance characteristics, quantity, deployment, units, sizes, and doctrine. The data base supports an 8-hour battle between a Blue armor brigade in a mobile defense and a Red tank division conducting a breakthrough operation. Both sides have organic and supporting artillery and EW assets. Variants of the data base were developed for the two-part verification testing.

c. <u>Verification, Part I</u>. The objective of the Part I verification test was to establish that the model accurately represents the performance of communication and electronic warfare systems during tactical operations. The forces used were representative of brigade/division unit types. The scenario involved a Blue brigade engaging a Red division in Central Europe. To simplify assessment of the simulation results, the scenario was sampled in a series of scenes, each of which was selected to exercise specific model processes. The combat units were represented by one Blue battalion, consisting of three companies, faced by a Red regiment of three battalions. The supporting units were represented by a direct support fire battery, a close air support element, EW units, and an intelligence section. Higher headquarters elements were included to incorporate command and control actions from brigade level down on the Blue side and from division level down on the Red side.

The verification proceeded in a series of tests, each focusing on one or more of the module operations. Hand-calculated results were compared with appropriate model output. Agreement with hand calculations constituted the Part I verification of the model operation. A lack of agreement was attributed to either input or code deficiency. Both possibilities were assessed, and input and/or code was modified and the model rerun if necessary. The process was repeated until either a satisfactory model output was achieved or until enough information was obtained that the problem could be addressed at a later date. A subset of the results of the Part I verification test is illustrated in Table 1.

The Part II verification test was to determine if the model gave an accurate representation of current and projected combat, communications, and electronic warfare environment. Part II consisted of three subparts. The part involving excursion runs from the base case and the part involving the comparison of DEWCOM with another model are not discussed in this paper. The third part (sensitivity test) of the Part II test is discussed in the remainder of this paper.

Table 1. Verification Results

Module/operation	Response to test objectives as of initial testing	Action taken	Response to test objectives as of record run (model realism)
TACTICAL			
Movement	Spurions movement by EW units' improper reference to unit coordinates in test for FEBA distance	Code corrected	S
Direct fire attrition	Low attrition - input data calibration needed	Alternate algorithms under consideration	S
Indirect fire attrition	Near zero attrition	Algorithm incorrect - replaced	S
Stop/restart	Model cannot be properly restarted if changes affect ongoing activity	Remains to be resolved	U
COMMUNICATIONS			
Message generation	Description of parameters controlling message generation not clear - otherwise satisfactory	Documentation updates prepared	S
Message processing	Message processed only by one net type - no routing over other net types - some redundant message generation	Remains to be resolved	M
ELECTRONIC WARFARE			
Locate operation	Improperly activated - otherwise satisfactory	Code corrected	S
Intercept operations	Improperly activated - otherwise satisfactory	Code corrected	S
Jam operations	Radar signal not jammed - otherwise satisfactory	Code corrected	S

NOTE: S = satisfactory
M = marginal
U = unsatisfactory

3. PROBLEM DESCRIPTION AND BACKGROUND

a. Description. The purpose of the sensitivity test was to identify those input factors which have the largest impact upon selected model output variables and to estimate the magnitude of the input factor effects.

b. Background. The Concepts Analysis Agency has conducted extensive computer simulation model sensitivity testing, e.g., Bates (1974), Thomas (1975), and Bates (1977). Past sensitivity testing has always been directed at, or in support of, a particular study; that is, a study involving the investigation of tradeoffs of particular combat parameters which could be associated with specific model input factors. The number of input factors to be investigated was always large and the total number of computer model simulation runs was always limited. In all cases, a decision had to be made between the number of factors and the number of factor levels. Invariably, the objective was to investigate as many input factors as possible. Consequently, ultimate experimental designs developed were 2^m and/or 3^n fractional factorial designs.

The DEWCOM sensitivity test was different from previous model sensitivity tests. The test was not for the investigation of particular input factors in order to assess the applicability of the model for a particular study's use. The test was a part of an overall test and evaluation of the model following its initial development. It was desired that the sensitivity test address as many of the 250 model input variables as possible. Ultimately, a group-screening approach was used in the experimental design development.

c. Group-screening Designs. Group-screening experimentation is not new. Watson (1961) discusses two-stage screening procedures. Patel (1962) and Li (1962) independently introduced multistage group-screening designs. Hunter and Mezaki (1964) illustrate the application of group-screening designs to chemical reaction experimentation. Kleijnen (1975a) gives a survey of screening designs, and Kleijnen (1975b) contains a more detailed discussion of screening designs. Mauro and Smith (1980) examine two-stage, group-screening methods, and Mauro and Smith (1981) examine a random balance/Plackett-Burman, two-stage strategy.

4. EXPERIMENTAL DESIGN. A detailed examination was made of each of the 250 model input variables. It was decided to consider only Blue input variables; therefore, the Red input variables were excluded. Also excluded were variables causing abrupt changes, e.g., threshold and switching variables. An attempt was made to include only those variables having a continuous, rather than discontinuous, effect upon model output. Also, an attempt was made to include variables which were expected to have a significant effect upon model output. Ultimately, approximately 50 input variables were selected for investigation. Nominal values were then selected for each of the input variables. Finally, a "high" and a "low" value was determined for each of the variables. Illustrations of the input variables and their values are given in Tables 2, 3, and 4. Table 2 contains tactical variables, Table 3 contains communication variables, and Table 4 contains EW variables. The high and low values were picked to be those values which were expected to contain the expected achievable within the 1990 timeframe. Care was taken to

defining lows and highs of the variables being grouped together so that their expected effect would be in the same direction. We did not want variable effects to inadvertently cancel each other.

Table 2. Tactical Variables

Test factor	Input variable	Input description	Nominal value[a]	Variation about nominal value	
				Low	High
Unit movement	TU.MOVE.RATE	Unit move rate	100 m/min	X 1/2	X 2
	TU.RADIUS	Circular area occupied by unit	(100-999)m	X 1/2	X 2
Direct fire	TU.WP.QUANTITY	Number weapons by type owned by unit	(2-44)	X 1/2	X 2
	WT.COMBAT.VALUE	Combat value with weapon type	(2-100)	X 4/5	X 6/5
	DAMAGE.CLASS	Coefficient modeling effects of attrition of weapon type	(50-100)	+25	-25
Indirect fire	SD.ARTY.RESET.TIME	Minimum time interval between artillery fire missions	2 min	X 2	X 1/2
	TU.SUPPRESSION.FACTOR	Percent decrease in unit effectiveness due to arty fire	(12-60) min	X 2	X 1/2
	TU.DURATION.OF.SUPPRESSION	Period of decrease in unit effectiveness due to arty fire	(4-15) min	X 2	X 1/2
	TU.ARTY.DURATION	Duration of arty fire against unit	15 min	X 1/2	X 2
	TU.ARTY.INTERVAL	Interval between artillery fires against unit	1 min	X 1/2	X 2

[a]Nominal values which show a range depict the spread for all types of equipment being modeled, e.g., the combat values assessed each type weapon varied from 2 to 100.

Table 3. Communication Variables

Test factor	Input variable	Input description	Nominal value[a]	Variation about nominal value	
				Low	High
Message generation	SD.COORDINATION.INTERVAL	Minimum time between messages for units	30 min	X 2	X 1/2
	CO.LENGTH	Transmission time of message	(30-120) sec	X 2	X 1/2
Message processing	BACKGROUND.TRAFFIC.UPDATE.TIME	Interval at which traffic delays are computed	15 min	X 1/3	X 3
	MAX.DELAY	Maximum time for background traffic delay	10 min	X 2	X 1/2
	CO.PROCESSING.TIME	Time before and after transmission needed to process a message	(1-3) min	X 3	X 1/3
Network maintenance	CE.FS.QUANTITY	Initial quantity of communication equipment	(1-6)	X 1/2	X 6/5
	DAMAGE.CLASS	Coefficient modifying effects of attrition on communication equipment	100	+25	-25
	CET.MTBF	Mean time between failures for communication equipment	(600-700) hrs	X 1/2	X 2
	TU.COMM.SETUP.TIME	Interval needed to set up wire communication for unit	(5-10) min	X 2	X 1/2
	TU.COMM.TEARDOWN.TIME	Interval needed to tear down wire communication for unit	(5-10) min	X 2	X 1/2

[a]Nominal values which show a range depict the spread for all types of equipment being modeled, e.g., the processing time for the various types of communication ordered being modeled varied from 1 to 3 minutes.

Table 4. Electronic Warfare Variables

Test factor	Input variable	Input description	Nominal value[a]	Variation about nominal value	
				Low	High
Intercept	EWE.FS.QUANTITY	Initial quantity of intercept equipment	1	X 1/2	X 2
	DAMAGE.CLASS	Coefficient modifying effects of attrition on intercept equipment	100	+25	-25
	EWT.MTBF	Mean time between failure for intercept equipment	70 hrs	X 1/2	X 2
	EWT.MTTR	Mean time to repair for intercept equipment	3 hrs	X 2	X 1/2
Locate	EWT.DF.TIME	Time period (in min) to perform DF function	(1-3) min	X 3	X 1/3
	SD.DF.RATE.1.UNIT.OUT	Percent decrease in intel rate for DF less 1 station	50%	+25	-25
	SD.DF.RATE.2.UNITS.OUT	Percent decrease in intel rate for DF less 2 stations	75%	+25	-25
	EWE.FS.QUANTITY	Initial quantity of locate equipment	1	X 1/2	X 2
	DAMAGE.CLASS	Coefficient modifying effects of attrition on locate equipment	100	+25	-25
	EWT.MTBF	Mean time between failures for locate equipment	70 hrs	X 1/2	X 2
	EWT.MTTR	Mean time to repair locate equipment	3 hrs	X 2	X 1/2
Ground surveillance	EWE.FS.QUANTITY	Initial quantity of surveillance equipment	1	X 1/2	X 2
	DAMAGE.CLASS	Coefficient modifying effects of attrition on surveillance equipment	100	+25	-25
	EWT.MTBF	Mean time between failures for surveillance equipment	70 hrs	X 1/2	X 2
	EWT.MTTR	Mean time to repair for surveillance equipment	3 hrs	X 2	X 1/2
	EWT.POWER	Power output of surveillance equipment	40 db	-20	+50

[a]Nominal values which show a range depict the spread for all types of equipment being modeled, e.g., the time to perform the DF function for the various EW equipment varied from 1 to 3 minutes.

Simultaneous to the above, the primary operations of each module were enumerated and basic military functions were associated with the module operations. The association between the module operations and military functions is shown in Table 5. The 11 module operations are hereafter termed model input factors. The experimental design, therefore, involved the 11 two-level input factors. All factors are completely crossed. It was suspected that the factors within a module may interact. Consequently, an experimental design which would permit assessment of main effects and first order (two-factor) interactions was desired. A $1/16 \times 2^{11}$ experiment was designed using

$$I = ABEFJL = CDEFKL = BCEGJKL = ABCDEFGH$$

as the defining contrast. The fractional factorial design required 128 model runs and permitted assessment of the 11 main effects and the $\binom{11}{2} = 55$ two-factor interaction effects.

The analysis of variance (ANOVA) model for the design is

$$y = \mu + A + B + \dots + L + AB + AC + \dots + KL,$$

where μ is a true but unknown effect; A, B, ..., L are factorial effects; and y is a particular model output variable.

Table 5. Test Factors

Test factor	DEWCOM operations	Military functions					
		Mvr	Ind fire	Cmd & con	Comm	Intel coll	EW
	TACTICAL MODULE						
A	Unit movement	X					
B	Direct fire	X					
C	Indirect fire		X				
	COMMUNICATIONS MODULE						
D	Message generation			X			
E	Message processing				X		
F	Network maintenance				X		
	EW MODULE						
G	Intercept					X	
H	Locate					X	
J	Ground surveillance					X	
K	Intel acquisition					X	
L	Jamming						X

The following four measures of effectiveness (MOE) were selected as model output variables for analysis:

Red personnel losses

Red weapons losses

Blue personnel losses

Blue weapons losses

The simulation experiment was executed in accordance with the experimental design, and the four MOE were analyzed in accordance with the analysis plan.

5. ANALYSIS. The analysis of variance model is a fixed effects model. Consequently, all 11 main effects and all 55 interaction effects in the ANOVA table are tested over the Mean Square (Residual) which has 61 degrees of freedom. For example, if MS(A)/MS(R) is equal to or greater than $F_{1,61,(1-\alpha)}$, input factor A is statistically significant at the α-level of significance. For each MOE, the marginal and two-way means were tabulated. ANOVAs were performed and the marginal means and significant interactions were plotted.

The ANOVA results are summarized in Table 6. First, considering the two Red MOE, we see that three interactions, DE, EF, and FK, are significant for Red personnel, and two interactions, DE and FK, are significant for Red weapons. Each MOE has the same significant main effects--A, B, F, and K. The decreasing order of the four significant main effects was B, K, F, and A for both Red MOE. Also, the direction of the effects was as expected--Red losses decrease as the input factors are changed from low to high levels. An examination of the Blue MOE results shows that the seven significant main effects are a subset of the significant interaction effects. In addition, the significant interactions contain factors F and K. Therefore, all input factors except G and J have a significant influence upon both Blue MOE. Expectations were that changing the input factors from low to high would have an increasing effect upon both Blue MOE; however, the change from low to high of factors B, D, and E had a decreasing effect. This apparent anomaly was explained after subsequent study of the model and input.

Table 6. ANOVA Tests Significance Levels

Source	Measures of effectiveness			
	Red		Blue	
	Personnel losses	Weapons losses	Personnel losses	Weapons losses
TACTICAL				
A - Unit movement	0.05	0.05	0.001	0.001
B - Direct fire	0.001	0.001	0.001	0.001
C - Indirect fire			0.001	0.001
COMMUNICATIONS				
D - Message generation			0.05	0.05
E - Message processing			0.001	0.001
F - Network maintenance	0.001	0.001		
ELECTRONIC WARFARE				
G - Intercepting				
H - Locating			0.05	0.05
J - Intel seeking				
K - Intel evaluation	0.001	0.001		
L - Jamming			0.01	0.001
AB -			0.001	0.001
AC -			0.001	0.001
AL -			0.01	0.001
BC -			0.001	0.001
BE -				0.05
DE -	0.05	0.05	0.05	0.05
DF -				0.05
EF -	0.05		0.05	0.05
EK -				0.05
FK -	0.001	0.01		
KL -			0.05	0.05

6. SUMMARY. Group-screening designs have potential application for the statistical testing of large complex computer simulation models. However, to date, the literature seems void of illustrations of group-screening designs applied to large real-world simulations. The above illustrates only the first stage in the application of group screening. Subsequent stages are essential for the usefulness of group-screening designs to be realized. For the above described problem, possibly resolution III designs should be used rather than resolution V designs. Because of the large number of interactions in the above illustration, however, it appears that resolution III designs would be inappropriate even for early stages. Too much care cannot be taken in the grouping of factors to ensure that effects do not cancel each other. Also, the direction of the effect of grouped factors must be known to be the same.

REFERENCES

1. Bates, C. B. (1974), Sensitivity Analysis of Selected Variables in the CEM, US Army Concepts Analysis Agency Technical Paper (TP)-74-2, Bethesda, Maryland.

2. Bates, C. B. (1977), Sensitivity Analysis of Selected DIVOPS Input Factors, US Army Concepts Analysis Agency Technical Paper (TP)-77-9, Bethesda, Maryland.

3. Hunter, W. G. and Mezaki, R. (1964), Catalyst Selection by Group Screening, Industrial and Engineering Chemistry, Vol. 56, No. 3, pp. 38-40.

4. Kleijnen, J. P. C. (1975a), Screening Designs for Poly-factor Experimentation, Technometrics, Vol. 17, No. 4, pp. 487-493.

5. Kleijnen, J. P. C. (1975b), Statistical Techniques in Simulation, Part II, Marcel Dekker, Inc., New York.

6. Li, C. H. (1962), A Sequential Method for Screening Experimental Variables, JASA, Vol. 57, pp. 455-477.

7. Mauro, C. A. and Smith, D. E. (1980), An Examination of the Performance of Two-Stage Group Screening for Use in Computer Simulation, Technical Report No. 113-3, Desmatics, Inc.

8. Mauro, C. A. and Smith, D. E. (1981), Factor Screening in Simulation: Performance of a Two-Stage Random Balance/Plackett-Burman Procedure, Technical Report No. 113-5, Desmatics, Inc.

9. Patel, M. S. (1962), Group-screening with More than Two Stages, Technometrics, Vol. 4, No. 2, pp. 209-217.

10. Thomas, J. T. (1975), TARTARUS Model Sensitivity Experiment, US Army Concepts Analysis Agency Technical Paper (TP)-75-9, Bethesda, Maryland.

11. Watson, G. S. (1961), A Study of the Group Screening Method, Technometrics, Vol. 3, No. 3, pp. 371-388.

SELLING A COMPLICATED EXPERIMENTAL DESIGN TO THE FIELD TEST OPERATOR

Carl T. Russell
US Army Cold Regions Test Center
Fort Greely, Alaska

ABSTRACT. After studying the objectives and planning constraints for cold regions performance testing with the main gun of the XM1 tank, the author determined that a rather complicated experimental design seemed appropriate. In particular, the blocking scheme required that carefully chosen quarter replicates of a 2^4 design be conducted sequentially. Such a design lays out a firing schedule almost round by round and, on the surface, presents the test operator with insurmountable problems in execution. A test operator generally expects and receives only instructions to complete a prescribed number of "replications" under each combination of test conditions, and he deeply resents the intrusion of a statistician into detailed test scheduling. By carefully emphasizing the practical and intuitively advantageous aspects of the experimental design, however, the author was able to sell the design to the test operator, and the sales pitch is the topic of this paper.

I. INTRODUCTION AND STATEMENT OF THE PROBLEM. Winter temperatures below freezing are common in highly industrialized regions of Europe, Russia, Red China, and Korea. Since these are all regions where the U.S. Army could fight, the need for a test center to examine the performance of Army personnel and materiel in the cold is obvious. The U.S. Army Cold Regions Test Center (CRTC) is located at Fort Greely, Alaska, where temperatures are below 32°F more than 80 percent of the time during the winter months, average 49 days per year below -25°F, and average an annual low of -59°F. Although colder areas exist than Fort Greely, no other accessible area in the United States is available to the U.S. Army for cold regions testing of military systems.

As part of Development Test II of the XM1 tank (M1 Abrams tank), cold regions testing was conducted at CRTC. Main gun firing performance at temperatures below 0°F was one of the many issues to be addressed during this testing. Criterion values for probability of hit against targets of prescribed size were given in a matrix for each combination of four test factors:

- Tank Mode--stationary or moving,
- Target Mode--stationary or moving,
- Nominal Range--short (approximately 1500 meters) or long (approximately 2500 meters),
- Round Type--high explosive anti-tank (HEAT) or kinetic energy (KE).

In addition, there were requirements to compare firing performance of the XM1 with the standard M60 tank and to determine whether XM1 firing performance degraded at lower temperatures.

There were numerous constraints:

- Rounds were to be fired in five-round shot groups at panel targets with hit probabilities to be estimated from impact coordinates for each five-round group.

- Around-the-clock testing was planned, with firing periods sandwiched between mobility exercises. Four shot groups per firing period appeared reasonable and feasible, and three tank crews were available for test conduct.

- Weather conditions were uncontrolled. Temperature was a factor of direct interest, but other factors (such as visibility) were regarded primarily as nuisance factors.

- An important decision point was scheduled before test termination, so partial data had to be interpretable.

- Although the criterion addressed only HEAT and KE rounds, two types of KE rounds--armor piercing descarding sabot (APDS) and armor piercing fin stabilized discarding sabot (APFSDS)--were provided for test.

- A few high explosive projectile (HEP) rounds were also provided for test, and there was some interest in ranges other than 1500 and 2500 meters.

Previous test planning had identified the number of rounds to be fired, and an unbalanced test matrix (Table 1) had been formulated to spread the rounds over the test conditions. This matrix--seldom differentiated from the test design--is typical of those usually proposed for field tests, and it would typically be analyzed as if it were conducted as a completely randomized experimental design. But its conduct would almost certainly have been dictated by efficiency, resulting in little actual randomization.

Russell (2) argued that because completely randomized designs are inimical to efficient test conduct in a field environment (they require overall conduct by chance rather than by organization), they should be replaced wherever possible by designs requiring only small-scale randomization easily generated during day-to-day conduct. An obvious approach was advocated: design in blocks compatible with test constraints and executable within a relatively short time period, repeating similar blocks throughout the test.

TABLE 1: Tentative Test Matrix (Tabulated Values Are Number of Five-Round Shot Groups at Each Combination of Test Conditions)

Tank/Target Mode	Round Type	XM1 Below -25°F		XM1 Above -25°F		M60 Above -25°F	
		1500	2500	1500	2500	1500	2500
S/S	HEAT	4	4	5	5	3	3
	APDS	5	5	7	7	4	4
	APFSDS	4	4	6	6	3	3
	HEP	2	2	4	4	2	2
S/M	HEAT	4	4	4	4	3	3
	APDS	5	5	5	5	4	4
	APFSDS	4	4	4	4	3	3
M/S	HEAT	4	4	4	4	3	3
	APDS	5	5	5	5	4	4
	APFSDS	4	4	4	4	3	3
M/M	HEAT	4	4	4	4	3	3
	APDS	5	5	5	5	4	4
	APFSDS	4	4	4	4	3	3

I used this "basic matrix approach" to devise a rather complicated revision to the tentative design for XM1 firing performance which essentially scheduled the crew and order for every shot group fired. Fundamentally, the complication was that inherent in any statistically sound field test design: instead of emphasizing sample size in terms of requisite "replications" in each cell (with an unrealistic request for complete randomization), the statistical advice concentrated on a method of detailed test conduct directed at obtaining a data set amenable to thorough statistical analysis. From the point of view of a test operator (at CRTC, these are usually 0-3's), such detailed statistical advice is inherently unwelcome. Instead of setting a clear objective (get so many observations per cell) with what is preceived as minimal guidance (randomize), the advice appears to set a vague objective (get a good data set) with strangling guidance (do it just this way). Thus my problem evolved from creating a sound design to selling it. How could I convince the test operator that it was possible and advantageous to execute my proposed design rather than simply obtain required "replications" of cells in some matrix?

My solution was to prepare and present a briefing designed to show not only that my proposed design was executable but also that it provided organized solutions to potential problems of test conduct while leaving a great deal of flexibility for the test operator and relieving him of awkward planning details. The next section of this paper presents this briefing in narrative form. A brief technical discussion of the design follows, and the paper concludes with a few final comments and a summary of the test outcome.

II. THE SALES PITCH. The briefing consisted of six parts: an introduction stating the design goals, an overview of the design which described the design in terms of four prioritized test matrices, and discussions of each test matrix in order. Because the briefing was a sales pitch, it emphasized in nontechnical terms why my proposed design should be conducted, how it could be conducted, and how it would provide advantages to the test operator which at least offset its disadvantages.

A. Experimental Design Goals. This portion of the briefing told what I was trying to accomplish:

- Compare
  - -stationary versus moving tank
  - -stationary versus moving target
  - -HEAT versus KE (APDS and APFSDS) rounds
  - -1500 meter versus 2500 meter tank/target range
  - -XM1 versus M60

  under test conditions as similar as possible.

- Make these comparsions over as wide a variety of test conditions as possible, but do so in such a way that the effects of selected test conditions (in particular temperature) can probably be isolated.

- Preplan order of trials in such a way that as much balance as possible is maintained on a day-to-day basis.
  - -to increase the likelihood that reasonably accurate partial results will be available quickly.
  - -to minimize the impact of unforeseen delays.

- Allow sufficient flexibility that, with a reasonable amount of good luck, the design can be executed.

B. Overview of Proposed Design. This portion of the briefing described the overall test in terms of four test matrices (Figure 1): two test matrices for XM1 alone (one for each temperature range of interest), a matrix for the XM1 versus M60 comparison, and a matrix for side tests and make-up. The emphasis here was on overall resource distribution rather than detailed test structure. Together with Table 2, which compared the test matrix associated with the proposed design to the tentative matrix of Table 1, Figure 1 was meant to reassure the audience that no radical departure from the status quo was being advocated. But both Figure 1 and Table 2 were also used to point out two inherent advantages of the proposed design, namely its balance (in my experience balance appears to almost anyone as

TABLE 2: Comparison of the Proposed Test Matrix to the Tentative Test Matrix of Table 1
(Tabulated Values Are Number of Five-Round Shot Groups at Each Combination of Test Conditions)

Tank/Target Mode	Round Type	Range (Meters)	XM1-Alone Above -25°F		XM1-Alone Below -25°F		XM1 Versus M60 Above -25°F				XM1 Side Test
							XM1		M60		
			Prop	Tent	Prop	Tent	Prop	Tent	Prop	Tent	(Prop)
S/S	HEAT	1500	3	2	3	4	3	3	3	3	
		2500	3	2	3	4	3	3	3	3	
	APDS	1500	3	3	3	5	3	4	3	4	8
		2000									8
		2500	3	3	3	5	3	4	3	4	8
		3000									8
	APFSDS	1500	3	3	3	4	3	3	3	3	
		2000									4
		2500	3	3	3	4	3	3	3	3	
		3000									4
	HEP	1500		2		2		2		2	8
		2000									8
		2500		2		2		2		2	
S/M	HEAT	1500	3	2	3	4	3	3	3	3	
		2500	3	2	3	4	3	3	3	3	
	APDS	1500	3	3	3	5	3	4	3	4	
		2500	3	3	3	5	3	4	3	4	
	APFSDS	1500	3	3	3	4	3	3	3	3	
		2500	3	3	3	4	3	3	3	3	
M/S	HEAT	1500	3	2	3	4	3	3	3	3	
		2500	3	2	3	4	3	3	3	3	
	APDS	1500	3	3	3	5	3	4	3	4	
		2500	3	3	3	5	3	4	3	4	
	APFSDS	1500	3	3	3	4	3	3	3	3	
		2500	3	3	3	4	3	3	3	3	
M/M	HEAT	1500	3	2	3	4	3	3	3	3	
		2500	3	2	3	4	3	3	3	3	
	APDS	1500	3	3	3	5	3	4	3	4	
		2500	3	3	3	5	3	4	3	4	
	APFSDS	1500	3	3	3	4	3	3	3	3	
		2500	3	3	3	4	3	3	3	3	

XM1 ABOVE -25°F

	120 HEAT
360 RDS:	120 APDS
	120 APFSDS

XM1 BELOW -25°F

	120 HEAT
360 RDS:	120 APDS
	120 APFSDS

XM1 VERSUS M60 COMPARISON (ABOVE -25°F)

	120 HEAT		120 HEAT
XM1, 360 RDS:	120 APDS	M60, 360 RDS:	120 APDS
	120 APFSDS		120 APFSDS

XM1 SIDETESTS & MAKE-UP

	160 APDS
280 RDS:	40 APFSDS
	80 HEP

Figure 1. Overall Structure of Proposed Design in Terms of Resources Distributed Among Four Test Matrices.

intuitively advantageous) and its formal distinction between highest priority testing (XM1 alone, especially above -25°F), secondary testing (XM1 versus M60), and testing to be done if possible. A problem with three crews was also discussed connection with balance. It would clearly be desirable to have each crew fire the same number of five-round shot groups under similar conditions, but three crews cannot possibly fire cell totals of two, four, five, or seven shot groups (from the tentative matrix) in a balanced fashion. Conveniently, my revised design requires exactly three shot groups in every cell except those in the XM1 side test.

C. XM1 Trials Above -25°F. This portion of the briefing discussed the details of the test design for the highest priority test matrix in depth.

A "basic matrix" of test conditions (Table 3) was introduced and terms were defined. Each combination of conditions in the matrix (cell) was to be executed three times, once by each crew. One execution of a cell (mission) was to consist of a crew firing a five-round shot group under the stated conditions. A trial was to consist of four prescribed missions by the same crew during one firing period.

The key to the design was the typical trial. A specific example of the typical trial was given for crew 3:

S/S, HEAT, 1500
S/M, HEAT, 2500
M/S, APDS, 2500
M/M, APDS, 1500.

TABLE 3: Basic Test Matrix (Cells Numbered for Reference)

Tank/Target Mode	HEAT 1500	HEAT 2500	APDS 1500	APDS 2500	APDSDS 1500	APDSDS 2500
S/S	1*	2	3	4	5	6
S/M	7	8*	9	10	11	12
M/S	13	14	15	16*	17	18
M/M	19	20	21*	22	23	24

*This is the specific example of a "typical trial".

The typical trial was then discussed both in terms of this specific example and a general version which specified that a crew was to execute four missions:

- One at each tank/target mode,
- Two with each of two round types,
- Each round type at both ranges.

(The order of missions within each trial was to be randomized as much as possible within test constraints.)

The disadvantage of trials of this sort was openly admitted: test conditions within every trial were to be totally mixed with nothing fixed. Crews would not be allowed to fire easier targets (that is, stationary tank or target, short range) first, the moving target would have to be available for every trial, and careful instruction of crews would be necessary to prevent round types being fired at the wrong tank/target modes or the wrong ranges. On the surface, for instance, using the same round type and firing only one target throughout a trial would be more efficient and less prone to error than trials with the proposed structure. But the great advantage of trials with the proposed structure was also pointed out: the test operator's tally sheet would be nearly balanced after each trial with the same number of missions at each range, at each target/tank mode, and (as much as one could hope) at each round type. Furthermore, although firing of easier targets first was not universally permitted, it was stressed that no great harm would ensue if in some of the earlier trials easier targets were fired first. Likewise a certain amount of systematic manipulation of ranges was permissible so long as it was not always done the same way: for instance, in the specific example, the tank could fire first on the move five APDS rounds against a 2500 meter stationary target, then fire stationary HEAT rounds against a 2500 meter moving target, followed by the missions against 1500 meter targets.†

After this somewhat lengthy discussion of the typical trial, the proposed order for conduct of trials was given in terms of prioritized lists of trials for each crew (Table 4). The specific typical trial discussed previously was identified in the table, and a brief examination of other trials showed that they are indeed very similar.

TABLE 4: Prioritized Lists of Trials for Each Crew, XM1 Alone, Above -25°F (Numbering from Table 3)

Crew 1	Crew 2	Crew 3
2, 7,15,22	3,10,14,19	1, 8,16,21*
6,11,13,20	3, 8,18,23	5,12,14,19
1, 8,18,23	6,11,13,20	2, 7,17,24
3,10,14,19	2, 7,15,22	4, 9,13,20
4, 9,17,24	5,12,16,21	6,11,15,22
5,12,16,21	4, 9,17,24	3,10,18,23

*This is the specific example of a "typical trial".

The idea behind the lists was as follows. If after a mobility run the temperature were between 0°F and -25°F, whatever crew was in the tank would

†As anticipated, this cavalier attitude towards randomization raised comments following the paper's formal presentation. What I was trying to give the test operator, however, was an executable design with reasonable flexibility. In fact, I eventually provided the test operator with preferred order of conduct in the form of mission-by-mission schedule lists, and mission order in these lists was formally randomized within trial. But I do not believe that minor nonsystematic deviations from a formal randomization scheme could markedly effect the experimental results in an experiment of this size, and I doubt that even naively systematic deviations from within trial randomization could have overwhelming effects. Randomization within trials would protect primarily against possible tendancies for crew performance to vary consistently within trials rather than from trial-to-trial. Although attempting such randomization is worthwhile, its benefits should be small unless consistent within-trial trends are substantial relative to effects of primary interest--which I believe to be highly unlikely in the present case. Thus even naively systematic within-trial ordering should produce only small bias under reasonable assumptions about operational performance. In fact, conducting a test in small blocks of time and space forces any potential damage from nonrandomization to be small (under mild assumptions), permitting the test operator substantial deviations from formal randomization within blocks. To me, the risk of bias from lack of within-trial randomization is small compared to the risk of losing influence over test conduct by pedantically restraining the test operator.

fire its next scheduled trial. (If the temperature were between -25°F and -50°F, the crew would fire its next scheduled trial from a similar list for trials below -25°F, and if the tamperature were above 0°F or below -50°F, no firing would take place.) Crews would then change, and another mobility run would begin.

With reasonable attention to crew scheduling and some luck, it should be possible to conduct at least three or four trials by each crew according to the proposed order. Toward the end of the test, however, instances could be expected when the crew in the tank had already completed all missions in the temperature range present at the end of a mobility run. In such instances, the test operator was advised simply to have the crew fire (from bottom/up in the list) any available mission for another crew in the correct temperature range. If no such trials in the correct temperature range were available, then based on test time remaining, available trials, and forecasted weather, the test operator could opt not to fire or opt to fire an available trial from a list for the incorrect temperature range (working from bottom/up in the list, preferably using the correct crew). Likewise, if the moving target array should break down, the test operator was advised simply to fire (from bottom/up) the first available trial, ignoring the requirement for moving targets; that is, fire all four missions at stationary targets but use the tank mode, range, and round type specified. Other operational problems were portrayed similarily:

- If all goes well, conduct the next available trial from top/down in the prioritized list for the crew in the tank.

- If problems arise but a decision to fire anyway is prudent, conduct the lowest priority available trial in as close accordance with the prescribed conditions as possible.

- The statistician would be available at any time to provide advice.

A field test is a moving train, and the engineer deserves advice which will help him be on time.

The proposed ordering of trials, if executed as just described, provides the statistician with a usable data set even if many trials cannot be completed according to plan. In fact, provided there were no great difference in firing performance between APDS and APFSDS rounds (none was expected):

- Once any three of the first four trials with any one crew were completed, the data set would be usable.

- Once the first, second, third, or fourth trials on all three crews were completed, the data set would be usable.

If only half the data were obtained in accordance with the prescribed plan, the statistician would be in pretty good shape for analysis (but his statistical statements could not be as precise as with a complete data set). Some intuitive understanding of why this is so can be seen by examining how the basic matrix fills up trial-by-trial (Figure 2). All cells of the HEAT versus KE matrix fill up in an organized way as trials progress, with at least one observation per cell after the first two proposed trials for each crew and with three observations per cell (one for each crew) after the first four proposed trials for each crew. The last two trials for each crew compare only the two KE round types, filling in the holes left after the first four trials. (The demonstration in Figure 2 was accomplished with overlaid vu-graphs in the actual briefing.)

The test operator would have to devise some sort of organized schedule even to fill the tentative matrix of Table 1. What the statistician has done here is to relieve the test operator of a tedious task by providing him with a balanced version of Table 1 together with a flexible schedule which incorporates sound statistical advice directed towards obtaining as much information as possible from firing performance data.

D. XM1 Trials Below -25°F. This portion of the briefing quickly described the second test matrix, a matrix with slightly lower priority than the first. Table 5 gives the firing lists with the numbering of Table 3.

TABLE 5: Prioritized Lists of Trials for Each Crew, XM1 Alone, Below -25°F (Numbering from Table 3)

Crew 1	Crew 2	Crew 3
5,12,14,19	2, 7,17,24	6,11,13,20
1, 8,16,21	4, 9,13,20	2, 7,15,22
4, 9,13,20	1, 8,16,21	3,10,14,19
2, 7,17,24	5,12,14,19	1, 8,18,23
6,11,15,22	3,10,18,23	5,12,16,21
3,10,18,23	6,11,15,22	4, 9,17,24

The lists in Table 5 are very similar to those in Table 4. In fact the only difference in the two sets of lists is that in the lists of Table 5:

- Each crew fires its HEAT rounds in the opposite order from that of Table 4.

- The APDS and APFSDS rounds fired in HEAT versus KE trials are those fired in APDS versus APFSDS trials in Table 4.

E. XM1 versus M60 Trials. This portion of the briefing introduced the new problem with comparison firing and described the proposed solution, emphasizing the similarity of XM1 versus M60 trials to those with XM1 alone. Comparison trials were to have substantially lower priority than trials for XM1 alone.

AFTER FIRST TRAILS FOR ALL CREWS

TNK/TGT	HEAT		KE(TOTAL)		APDS		APFSDS	
MODE	1500	2500	1500	2500	1500	2500	1500	2500
S/S	3	1	2		2			
S/M	1	3		2		2		
M/S		2	1	3	1	3		
M/M	2		3	1	3	1		

AFTER SECOND TRIALS FOR ALL CREWS

TNK/TGT	HEAT		KE(TOTAL)		APDS		APFSDS	
MODE	1500	2500	1500	2500	1500	2500	1500	2500
S/S	23	1	23	1	2		3	1
S/M	1	23	1	23		2	1	3
M/S	1	23	1	23	1	3		2
M/M	23	1	23	1	3	1	2	

AFTER THIRD TRIALS FOR ALL CREWS

TNK/TGT	HEAT		KE(TOTAL)		APDS		APFSDS	
MODE	1500	2500	1500	2500	1500	2500	1500	2500
S/S	123	1 3	23	12	2		3	12
S/M	1 3	123	1 3	23		2	12	3
M/S	12	23	1 3	123	1	3	3	12
M/M	23	12	123	1 3	3	1	12	3

AFTER FOURTH TRIALS FOR ALL CREWS

TNK/TGT	HEAT		KE(TOTAL)		APDS		APFSDS	
MODE	1500	2500	1500	2500	1500	2500	1500	2500
S/S	123	123	123	123	12	3	3	12
S/M	123	123	123	123	3	12	12	3
M/S	123	123	123	123	12	3	3	12
M/M	123	123	123	123	3	12	12	3

AFTER SIXTH TRIALS FOR ALL CREWS

TNK/TGT	HEAT		KE(TOTAL)		APDS		APFSDS	
MODE	1500	2500	1500	2500	1500	2500	1500	2500
S/S	123	123	123	123	123	123	123	123
S/M	123	123	123	123	123	123	123	123
M/S	123	123	123	123	123	123	123	123
M/M	123	123	123	123	123	123	123	123

Figure 2. Cumulative Completion of Cells in the Basic Matrix. (Crew Numbers of Crews Firing Each Cell Are Shown.)

The new problem was that for comparison firing, two tanks and two crews must be present for every trial. The proposal was that each crew fire four missions during a trial, two from each tank, so that a trial would consist of eight missions rather than four missions as in XM1-alone trials. Each crew would fire at both ranges and at both stationary and moving targets, but one crew would fire only from a moving tank, and the other crew would fire only from a stationary tank. Each crew would fire only one round type, but each crew would fire the same combinations of tank/target mode, range, and round type for both tanks. In a typical trial, the crews could be crew 1 and crew 2, say, and execute the following missions during one firing period:

Crew 1	Crew 2
XM1, S/S, HEAT, 1500	XM1, M/S, APDS, 2500
XM1, S/M, HEAT, 2500	XM1, M/M, APDS, 1500
M60, S/S, HEAT, 1500	M60, M/S, APDS, 2500
M60, S/M, HEAT, 2500	M60, M/M, APDS, 1500

These trials are actually very similar to those for XM1 alone. In fact, the pattern for each tank is exactly the pattern for the specific example of a typical trial for XM1 alone above -25°F:

S/S, HEAT, 1500
S/M, HEAT, 2500
M/S, APDS, 2500
M/M, APDS, 1500

(This ongoing similarity should be comforting to the test operator: it shows that the proposed design presents essentially one obstacle to control of trials, not many.)

The order of mission conduct within each trial should be randomized as much as possible within test constraints, but only limited randomization would probably be possible. In the typical trial, crew 1 might be in the XM1 on a mobility run, and crew 2 would be due to replace crew 1 in the XM1. It would be sensible for the test operator to have crew 2 fire its missions in the M60 before the XM1 arrives, then have crew 1 fire its XM1 missions, change crews, have crew 2 fire its missions (which frees the XM1 for another mobility run), and finally have crew 1 fire its M60 missions.

Detailed scheduling lists are given in Table 6. As with XM1-alone trails, they should be conducted top/down if all goes well, and bottom/up if problems arise. The ordering and its benefits is also similar to XM1-alone trials. The obvious problem with these lists, however, is that they require specific pairs of crews to be present for each trial, which imposes awkward scheduling difficulties on the test operator. These difficulties are unavoidable since, for instance, if crew 1 and crew 2 were to fire eight trials together rather than the planned six, then no crew would be available as a partner for crew 3 on two trials. For all crews to fire the same number of trials when two crews are necessary for each trial, each pair of crews must fire the same number of trials together.

TABLE 6: Prioritized Lists of Trials for Each Pair of Crews, XM1 versus M60, Above -25°F (Numbering from Table 3)

Crew 1/Crew 2 (XM1; M60/XM1; M60)	Crew 1/Crew 2 (XM1; M60/XM1; M60)	Crew 1/Crew 2 (XM1; M60/XM1; M60)
4, 9; 4, 9/13,20;13,20	2, 7; 2, 7/17,24;17,24	14,19;14,19/ 5,12; 5,12
17,24;17,24/ 2, 7; 2, 7	13,20;13,20/ 4, 9; 4, 9	1, 8; 1, 8/16,21;16,21
14,19;14,19/ 5,12; 5,12	16,21;16,21/ 1, 8; 1, 8	4, 9; 4, 9/13,20;13,20
1, 8; 1, 8/16,21;16,21*	5,12; 5,12/14,19;14,19	17,24;17,24/ 2, 7; 2, 7
15,22;15,22/ 6,11; 6,11	18,23;18,23/ 3,10; 3,10	3,10; 3,10/18,23;18,23
3,10; 3,10/18,23;18,23	6,11; 6,11/15,22;15,22	15,22;15,22/ 6,11; 6,11

*This is the specific example of a typical trial.

F. XM1 Side Tests and Make-up. This portion of the briefing described how any rounds left over from the main design could be used.

In the unlikely event that the main design could be conducted quickly without major deviations from the plan, the remaining rounds could be fired in eight trials, each trial fired by one crew during one firing period and consisting of seven missions (five-round shot groups) from a stationary tank against stationary targets:

HEP at 1500 meters,
HEP at 2000 meters,
APDS at 1500 meters,
APDS at 2000 meters,
APDS at 2500 meters,
APDS at 3000 meters,
APFSDS at either 2000 meters or 3000 meters but not both.

Trials should be balanced over crews and temperatures as much as possible, and half of the APFSDS missions should be fired at each of 2000 meter and 3000 meter ranges. Order of missions within each trial should be randomized as much as test conditions permit.

If conducted, these trials could provide some insight to HEP performance in the cold and to KE performance at ranges not addressed in the main design. In the more likely event that during the conduct of the main design extra rounds were needed for zeroing, diagnostic testing, or re-executing partially completed trials, this last matrix provides a store of low priority rounds for use.

III. TECHNICAL ASPECTS OF THE DESIGN. The previous section of this paper described the proposed design and its advantages in nontechnical terms. This section discusses briefly how the design was constructed and sketches its analytic properties.

The key to the design is its blocking scheme, which is based on P. W. M. John's three-quarter replicates (1). Ignoring for the moment the two KE round types, there are four primary factors of interest, each at two levels:

Factor A, tank mode (stationary, moving),
Factor B, target mode (stationary, moving),
Factor C, range (1500 meters, 2500 meters),
Factor D, round type (HEAT, KE).

For firing of the XM1 alone, at each temperature range, each crew was to execute the resulting 2^4 design in blocks (trials) of four missions with defining contrasts:

(i) $I = +AD = +ABC = +BCD$,
(ii) $I = -AD = +ABC = -BCD$,
(iii) $I = +AD = -ABC = -BCD$,
(iv) $I = -AD = -ABC = +BCD$.

With this blocking scheme, all main effects and all 2-factor interactions (except AD) can be estimated free from blocks and other 2-factor interactions. Moreover, if one of the blocks is missing, the remaining 12-point design is a saturated fraction with all main effects and 2-factor interactions (except AD) still estimable free from blocks. (After thinking about tank gunnery, it was felt that AD and BD were likely to be the least important 2-factor interactions).

The problem of two KE round types was solved by letting KE round be one type in the blocks having +BCD in their defining contrast and the other type in blocks having -BCD in their defining contrast, and running two more blocks with only KE rounds. Using the coding from Table 4, this yielded two possible blocking schemes for conducting the 24-point design in 6 blocks:

Block	Scheme 1	Scheme 2
1	2, 7,15,22	2, 7,17,24
2	6,11,13,20	4, 9,13,20
3	1, 8,18,23	1, 8,16,21
4	3,10,14,19	5,12,14,19
5	4, 9,17,24	3,10,18,23
6	5,12,16,20	6,11,15,22

Ignoring KE round type, blocks 1-4 correspond to the defining contrasts (i)-(iv) in the previous paragraph. An examination of Tables 4 and 5 shows that Scheme 1 was used to construct the lists for crews 1 and 2 above -25°F and for crew 3 below -25°F, while Scheme 2 was used to construct the lists for crew 3 above -25°F and for crews 1 and 2 below -25°F. The introduction of two KE round types in this manner gives the design an incomplete blocks aspect: the effect of round type is partially confounded with the BC interaction, and additional information about the effect of round type can be obtained from an interblock analysis. All effects of interest can still be estimated with any one block missing or one of the HEAT-versus-KE blocks and an appropriate KE-only block missing. (Since any difference between KE round types measurable from this experiment was likely to be negligible, the analysis in practice was likely to proceed as if there were only one KE round type.)

For XM1-alone trials in either temperature range, the planned trial order was chosen so that the first four trials for any crew were KE-versus-HEAT trials and so that the Nth trials on all three crews represented three of the four different blocks with the same two round types. Thus at either temperature range, any three of the first four trials on any crew constituted a three-quarter replicate if KE round type were ignored, and the Nth trial on all three crews constituted a three-quarter replicate confounding crew with blocks.

Taking both XM1-alone firing matrices together, the design is a split-plot design. The subplots are missions (shot groups) treated by tank/target mode, range, and round type using a factorial scheme. Even with substantial data loss, clear inference concerning subplot factors should be possible since presence of even one three-quarter replicate guarantees estimability of interesting effects. The main plots are trials (blocks) treated by round combination, crew, order, temperature range, temperature, and additional random error. With some luck, inference concerning main plot factors should be possible. In the unlikely event that the entire XM1-alone design could be run as planned, quite elaborate analyses would be possible, one of which is indicated in Table 7. With only moderate data loss, analysis along the lines of that in Table 7 could probably still be conducted with some success. As in all split-plot designs, however, care must be taken with the error terms.

The designs corresponding to the remaining two test matrices were not as neatly structured as the design corresponding to the XM1-alone matrices. The lowest priority XM1 firing subtest was essentially a nonstatistical demonstration subject to cannibalization for rounds. The XM1 versus M60 comparison had lower priority than XM1-alone testing, and by emphasizing the comparison between tanks, the design lost much of its analytic potential concerning other effects. For any particular pair of crews and either tank type, the same blocking scheme used previously was exploited by confounding crew effect with the ABCD interaction, which confounds (crew)x(tank mode) and (crew)x(round type) with blocks. Three-quarter replicates were still preserved, but with lower resolution for A, B, C, and D (main effects only). By crossing tank type with the design in the other factors, however, maximum information about tank effects was obtained and potential for simple and easily presented paired-comparison analysis was introduced.

TABLE 7: Possible Analysis of Variance for Trials Involving XM1 Alone, Assuming All Trials Run Successfully

	Source	DF		Source	DF
1.	Total Sum of Squares	143			
2.	Blocks (Whole Plots)	35	3.	Treatments (Subplots)	19
2.1	Round Combination	2	3.1	A (Tank Mode)	1
2.2	Crew	2	3.2	B (Target Mode)	1
2.3	Order	1	3.3	A x B	1
2.4	Crew x Order	2	3.4	C (Range)	1
2.5	Temperature Range	1	3.5	A x C	1
2.6	Temperature	1	3.6	B x C	1
2.7	$(\text{Temperature})^2$	1	3.7	D (Round Type)	2
2.8	Whole Plot Error		3.8	C x D	2
	[2-(2.1+...+2.7)]	25	3.9	A x Crew	2
			3.10	B x Crew	2
			3.11	C x Crew	2
			3.12	A x Temperature Range	1
			3.13	B x Temperature Range	1
			3.14	C x Temperature Range	1
				Subplot Error	
				[1-2-(3.1+...+3.14)]	89

IV. SUMMARY AND CONCLUSION. The goals for cold regions testing of XM1 firing performance were ambitious, and a large number of rounds were available for test. I can justify neither a statistically naive nor a statistically pure approach to such testing. By working from basic statistical principles tempered by a concern for operational constraints, I was able to devise what I believe was a statistically sound and operationally executable design for this particular test. By suppressing some statistical niceties and most technical jargon, I was able to sell the design to the test operator in the sense that he agreed to attempt it along the proposed lines. Once this agreement was reached, I was able to gain control of certain detailed planning tasks through which I made actual execution of the proposed design more likely. The design proved flexible in that modifications could be made easily as test planning progressed. In particular, concern with comparison of two KE round types was eventually dropped (with the obvious design modification), and an eventual reduction to two crews was easy to accommodate. Unfortunately, temperatures during the test season were exceptionally warm, and when a few days of appropriately cold conditions finally arrived, tank malfunctions precluded firing performance testing. No rounds were fired for record. Nevertheless, I believe this paper shows that sophisticated designs for field tests are not only feasible but also marketable to the testing community if technical scruples are not allowed to dominate potential bottom-line results.

V. REFERENCES.

1. John, Peter W. M. (1971), Statistical Design and Analysis of Experiments, 161-165. MacMillan, New York.

2. Russell, Carl T. (1981). "The Potential Utility of Crossing a Fractional Factorial with a Full Factorial in the Design of Field Tests." U.S. Army Research Office Report No. 81-2, Proceedings of the Twenty-Sixth Conference on the Design of Experiments, 335-346.

LONG-TERM STORAGE OF ARMY RATIONS

Edward W. Ross, Jr.

US Army Natick Research and Development Laboratories
Natick, Massachusetts 01760

ABSTRACT. An important part of the Army's efforts to improve the food available to soldiers is a program of testing the ability of rations to maintain their acceptability when stored. This paper describes an experiment now underway on one of the Army combat rations, the Meal, Ready-To-Eat (MRE). In this paper we are mainly concerned with the statistical methods used to analyze data of the type obtained in this experiment in order to estimate the shelf-life of the food involved.

I. INTRODUCTION. This report is about the treatment of data from a large-scale, long-term storage experiment on a certain type of Army ration, called the Meal, Ready-To-Eat (MRE). The main purpose of the test is to find the effect of storage at various temperatures on the acceptance of this ration, i.e., how well it is liked by its consumers. There are a number of interesting questions involved in gathering and analyzing this data as well as interpreting the results to potential users.

II. THE EXPERIMENT AND THE DATA. The experiment consists of purchasing the rations, testing a sample of each, then storing the remainder at four different temperatures, 4, 21, 30 and 38°C. After specified time intervals, more samples are withdrawn from storage and tested, and the results compared with those obtained earlier. The schedule of storage times and temperatures is shown in Table 1.

The rations consist of twelve menus, each comprising four or five items. The total number of items is 52, provided the same food in two different menus is viewed as two different items. When the ration is tested, all the items in it are presented to each of 36 people. Each person rates each item in the menu on a scale of 1 through 9, where

9 means "like extremely"
⋮
5 means "neither like nor dislike"
⋮
1 means "dislike extremely."

Thus, for each of the 29 combinations of storage duration and temperature, there are 52 x 36 = 1872 scores to be analyzed.

There are a number of easily perceptible difficulties with this test plan:

(a) The amount of data that will eventually be accumulated may be quite large.

(b) The pool of test subjects is essentially the work force at Natick Laboratories, which perhaps differs somewhat in composition from the consumer population for whom the meal is intended, i.e. the Armed Forces.

(c) The long duration of the test may cause various kinds of trouble. E.g., the tastes of the test pool or population may drift, and there may be changes in the people giving the test and analyzing the data, caused by death, retirement or job actions.

(d) The form of the data is a source of some uncertainty as to the appropriate method of analysis. Most statistical data is either continuous or categorical with a few categories (typically ≤ 4). Our data are ordinal and categorical with nine categories, which places it in an intermediate situation where neither kind of analysis is wholly satisfactory.

Some of these issues will be discussed in the subsequent sections.

III. <u>ANALYSIS OF DATA</u>. For each food at each combination of time and temperature the data form a histogram for the frequency of each integer in the range 1 through 9, the total count being 36. From this information we wish to characterize the acceptability of the food when stored for various times and temperatures.

There are many ways of characterizing the storage stability of a food with regard to consumer acceptance. A common ingredient in most such descriptions is the definition of a critical score, with the property that the food is pronounced unacceptable if its score falls below the critical score. Usually the critical score is taken as 5. Given this, we shall regard as basic the idea of shelf-life, λ_H, at storage temperature H. λ_H is the time in months required for food stored at temperature H to obtain a score of 5. We assume that the initial score > 5, for otherwise the food would not be in the system.

Other parameters which characterize storage stability are, e.g., the average score after a fixed storage period (say 12 months). This is less useful than λ_H but easier to estimate. If more were known about the relationship among score, temperature and storage time, it might be possible to define a single parameter which would predict all combinations of time and temperature that cause a score of 5 for a food. The present data may lead to such a description, but we shall not pursue it further here.

We have already mentioned the fact that neither a categorical treatment (i.e. via contingency tables) nor a continuous approach (e.g., linear regression) is wholly satisfactory in analyzing this data. The categorical treatment does not lead easily to a prediction of shelf life, still less to estimates of its variance. The continuous methods assume a Gaussian distribution of scores, which is not satisfied.

The scheme we adopt uses both methods in an attempt to avoid the pitfalls faced by each separately. Moreover, it carries out the analysis at two levels of intensity on the data for each food and storage temperature up to the current time. First a coarse computation is done to determine whether the scores have changed during the time of the test. If they have not, we record the histogram of scores up to the current time, calculate the mean and standard deviation of those scores, but do no computation of shelf life (which is effectively infinite in this eventuality). If the coarse computation shows significant change in scores, we do two, more elaborate analyses in order to predict shelf life. We describe both the coarse and elaborate analyses in the following paragraphs.

The coarse analysis uses two methods, a contingency table analysis and a linear least squares calculation. The contingency table analysis is done twice, once with all non-empty columns and then with only 2 columns, usually obtained by pooling scores 1 through 6 and 7 through 9. The tail probabilities associated with the chi-square test are recorded for both. Also, the tail probability associated with the F-test of the hypothesis that the slope is zero is recorded from the regression. If any of these three tail probabilities is small enough (usually $< .10$ or even $< .20$), the more elaborate analysis is done. In addition, estimates of the shelf-life and a 90% lower confidence limit are recorded if the regression slope is non-zero.

The elaborate analyses apply a non-linear least squares (NLLS) and a multinomial logit method to the data. The non-linear least squares procedure is based on the model

$$y = x_1 + \varepsilon, \qquad t < x_3$$

$$y = x_1 - \frac{(x_1 - 5)(t - x_3)}{x_2 - x_3} + \varepsilon, \quad t \geq x_3$$

$$X^T = [x_1, x_2, x_3]$$

where y is the score, t the storage duration and X is the parameter vector fitted by the least-squares process. The X-components have the meanings

x_1 = initial score

x_2 = shelf life (months)

x_3 = lag or induction period (months) see Figure 1. ε is an i.i.d Gaussian $(0,S^2)$ random variable. The non-linear least-squares program NL2SOL was used to estimate X and its Hessian, from which confidence limits were obtained.

The multinomial logit method estimates the histogram of score probabilities

$$\Theta^T(t) = [\theta_1(t),\theta_2(t),\cdots,\theta_q(t)]$$

where $\theta_j(t)$ is the probability of score j at time t. We assume the logit model

$$\theta_j(t) = \exp(-\gamma_j(t))/\sum_{k=1}^{q}\exp(-\gamma_k(t))$$

$$\gamma_j(t) = \sum_{m=1}^{6}u_m\phi_{mj}(t)$$

$$\phi_{1j} = j - 5 \quad , \quad \phi_{2j} = \phi_{1j}^2$$

$$\phi_{3j} = \phi_{1j}\cdot t/12 \quad , \quad \phi_{4j} = \phi_{2j}\cdot t/12$$

$$\phi_{5j} = \phi_{1j}(\phi_{2j} - 10)/20 \quad , \quad \phi_{6j} = \phi_{5j}\cdot t/12 \ .$$

and U is the vector of parameters to be fitted,

$$U^T = [u_1,u_2,\cdots,u_6] \ .$$

The estimation of U is done by minimizing the negative logarithm of the likelihood of getting the observed counts, $r_j(t)$, $j = 1,\cdots,q$ which leads to minimizing

$$F = \sum_{i=1}^{N_T}\{\sum_{j=1}^{q}r_j(t_i)\gamma_j(t_i) + 36\ell n\sum_{j=1}^{q}\exp(-\gamma_j(t_i))\} \ .$$

N_T is the number of times at which we have data. Having solved for U, the shelf life is found by solving

$$\psi(t,U) = \sum_{j=1}^{q} (j - 5)\exp(-\gamma_j(t)) = 0$$

for t. A linearized estimate of the variance in shelf life is also obtained from related formulas. A special purpose minimizer of Newton type was written to solve for U and the IMSL version of the Brent algorithm furnished the shelf-life estimate.

These elaborate analyses require considerable computation but provide a good deal of information about the food. In particular, we obtain estimates of shelf life and 90% lower confidence limits from both methods. Also, each method allows us to predict the average score for any time, and the logit method predicts the complete histogram for any time. Obviously, prediction too far into the future by either method is risky.

Both methods require non-linear minimization to solve for their unknown parameters and can, therefore, encounter a variety of difficulties, e.g., non-convergence, convergence to a local but not global minimum, singularity of the Hessian, etc. Both minimizers contain some guards against these perils, and an additional check is furnished by comparing the resulting shelf-lives, but absolute certainty is not possible. The two methods need not lead to the same shelf-life estimates, though we expect them to be reasonably close if both converge well.

Most of the computations were done by means of the IMSL subroutines for forming and analyzing contingency tables and doing least-squares. The LINPACK subroutines were used extensively in the minimizer for the logit method.

IV. INTERPRETATION. An interesting aspect of the present problem concerns the reporting of results, i.e., how much of what kind of information should be relayed to the food technologists and thence to the logistical planners and purchasing agents. It is clear that a lot of information is produced at each stage by the computations, some of which is not directly useful to the food technologist.

At the current time, after 12 months of storage, the information reported to the technologist is shown in Table II and Figure 2. The first is the more important. It is a table of foods and storage temperatures whose entries are the shelf-lives of foods estimated to have shelf-lives $\leq$ 24 months. Estimates are listed only if they have some credibility, i.e., in this case a 90% lower confidence limit which is positive. Foods with shelf lives $>$ 24 months are currently estimated with poor accuracy since all the data is for $t \leq$ 12 months.

Food technologists are occasionally asked about the mean scores of various foods during the test. Figure 2 gives the mean food score at the most recent time of test, i.e., 12 months, in the form of a histogram. Each food is represented by a 3-character plotting symbol, of which the first two characters are the food number, and the third is either S or D, S if the food score was judged to be stable and D if deteriorating. For foods labelled S the mean score was calculated over all times up to the current test, but for D the mean is taken for the most recent time of test only.

For example, Table 2 shows that food number 2, ham-chicken loaf, had a shelf-life estimated as 13 months at 38^o storage temperature. In Figure 2 the histogram for 38^o shows that food number 02 had a mean score of 5.1 and was deteriorating at 12 months.

For the sake of ready reference, a simple table of mean scores of all foods at the four storage temperatures was also given to the technologists. However, the histogram is in most respects a more useful form for this information.

V. EXAMPLE. In Figures3 thru 9 we show examples of input and output produced by the programs in the course of analyzing the data for food number 34, fruit mix, after updating with the new scores at 12 months.

Figure 3 shows the input, consisting of ten lines of data. At this time the data comprises two lines for 4^o and 21^o (at 0 and 12 months) and three lines for 30^o and 38^o (at 0, 6 and 12 months).

Figures 4 and 5 present the output from the coarse analysis. For 4^o and 21^o only two times are available so the contingency tables consist of only two rows plus a sum row. The program does a t-test for a difference in means instead of the linear regression. At 30^o and 38^o there are three times, hence 3 rows in the contingency table, and the linear regression is done. The last line writes the smaller of the two tail probabilities from the contingency table and the tail probability from the t-test or linear regression.

We see that for temperatures 4, 21 and 30 degrees, none of the tail probabilities were less than .10, the critical value used here. At 38 degrees, however, both contingency table methods and the linear least squares analysis gave probabilities $< .10$, i.e., both methods agreed that the food scores had changed. For many foods and temperatures the results were not as clear as in this example.

Figures 6-9 show the results of the more elaborate analysis, done only for the 38^o storage case. Figure 6 is the printout from the NLLS analysis. We see the process converged after 6 iterations because CONV=4, i.e., the gradient became small, and the solutions were

X_1 = initial score = 7.19

$X_2 = \lambda_{38} = 20.7$ months

X_3 = lag or induction period = 5.4 months,

the lower 90% confidence limit for the shelf life being 16.5 months.

The printout and plot of the data and the fitted function show that at the next withdrawal (18 months) the average score is predicted to be about 5.39. This plot shows the upper and lower quartiles of the data at each storage time as a "U" and "L", respectively. The mean at each time is plotted as "O", but an "M" is printed wherever two plotted points coincide, as the theoretical and data means do here.

The logit results depicted in Figures 8 and 9, give information first about the minimization process, including estimates of the parameters U_1 to U_6 and comparisons of the experimental and predicted score counts. The results of the shelf-life calculation are then stated, followed by the predicted counts and mean score at the next withdrawal period (18 months). We see that the shelf-life is estimated as 20.3 months, which agrees quite well with the estimate of the NLLS calculation. The lower confidence limit, 10.8 months is appreciably lower than the 16.5 months given by NLLS method. The model predicts at 18 months a mean score 5.42, not much higher than the 5.39 value obtained from NLLS.

The comparison between experimental and theoretical histograms is shown in Figure 9, where "+" signified (experimental counts) > (theoretical counts) and "-" the reverse. It is clear that the model does not reproduce the details of the experimental counts very well in this case.

Sometimes the difference between the logit and NLLS calculations was greater than in this example but seldom exceeded 4 months when both procedures converged. However, there were many cases where only one converged, or both converged,but one or both had poorly conditioned Hessian matrices. Generally, the agreement was good when both shelf-life estimates were < 20 months, but became poorer as the estimates increased.

VI. <u>CONCLUSIONS.</u> The results up through the present time, 12 months of storage, show that only two of the 52 foods have failed. These are frankfurters (#6) and brownies (#13). Neither result is certain. Frankfurters had scores averaging < 5 on the initial test and have scored > 5 on subsequent tests. This suggests that the initial lot may have been uncharacteristically bad, but it also calls for closer scrutiny of that product. Brownies occur in two menus, as foods number 13 and 14. As #13, it had a shelf-life of only 11 months at 38°, but as #14 it showed no change at all! This suggests that there may be an interaction between this food and some of the other foods in these menus, but this too requires detailed study.

The methods used in this statistical treatment are apparently adequate though not the only ones possible. E.g., a different approach, via reliability procedures, is possible, and the methods based on information theory (see Kullback [1]) are also available. The procedures used here were chosen because they could be carried out with available computer programs and had substantially different viewpoints toward the data.

Again, the reporting tools used in Section IV appear to be satisfactory to the food technologists and are not too time-consuming to execute.

Clearly, the results will change with the passage of time and some changes in method may become necessary. It is hoped that the predictive capabilities of these methods will enable the food technologists to avoid unnecessary testing, but this remains to be seen.

REFERENCE

1. Kullback, S., Information Theory and Statistics, Gloucester, MA Peter Smith (1978).

TABLE 1: TEST PLAN

STORAGE TEMP	STORAGE DURATION (MONTHS)
4 DEG C	0,12,30,36,48,60,108
21 DEG C	0,12,18,24,30,36,48,60,126
30 DEG C	0,6,12,18,24,30,36
38 DEG C	0,6,12,18,24,114

TABLE 2: MRE SUMMARY AT 12 MONTHS
****

ESTIMATED SHELF-LIVES (IN MONTHS) OF SHORT-LIVED FOODS

FOOD NO.	NAME	IN MENU	ALSO MENU	AT TEMP 4 DEG	21 DEG	30 DEG	38 DEG	NOTE
**	******	****	****	******	******	******	******	****
1	PK SSG PTY	1	-	16	-	-	19	$
2	HAM CHIC LF	2	-	-	-	-	13	
4	BF SLI BBQ	4	-	-	-	22	17	
5	BEEF STEW	5	-	-	-	17	-	$
6	FRANKS	6	-	0	0	0	0	
11	HAM SLICES	11	-	-	-	-	22	
13	BROWNIES	3	8	-	22	-	11	$&
19	CHRY NT CK	5	-	-	-	18	-	
20	MAPL NT CK	10	-	-	-	-	20	
27	PTO PATTY	11	10	-	-	13	21	$&
30	PEACHES	1	4	-	23	19	16	$&
31	STRWBRIES	2	12	-	-	-	13	
32	STRWBRIES	12	2	-	-	-	17	
34	FRUIT MIX	5	-	-	-	-	20	
38	COCOA	1	7,9	-	-	16	18	&
40	COCOA	9	1,7	-	-	-	21	&
47	CRACKERS	2	11	-	-	20	-	$
48	CRACKERS	11	2	-	-	-	13	

INTERPRETATION: THE SYMBOLS IN THE COLUMN HEADED NOTE HAVE THE FOLLOWING MEANINGS. COMMON SENSE TELLS US THAT IF A FOOD HAS A SHORT LIFE AT ANY TEMPERATURE, IT SHOULD ALSO HAVE A SHORTER LIFE AT ALL HIGHER TEMPERATURES. THE SYMBOL $ MARKS THOSE CASES WHERE THIS DICTUM IS VIOLATED. COMMON SENSE ALSO TELLS US THAT IF A FOOD HAS A SHORT LIFE IN ONE MENU, IT SHOULD HAVE A SHORT LIFE IN ANY OTHER MENUS WHERE IT OCCURS. THE SYMBOL & MARKS THOSE CASES WHERE THIS TENET IS CONTRADICTED.

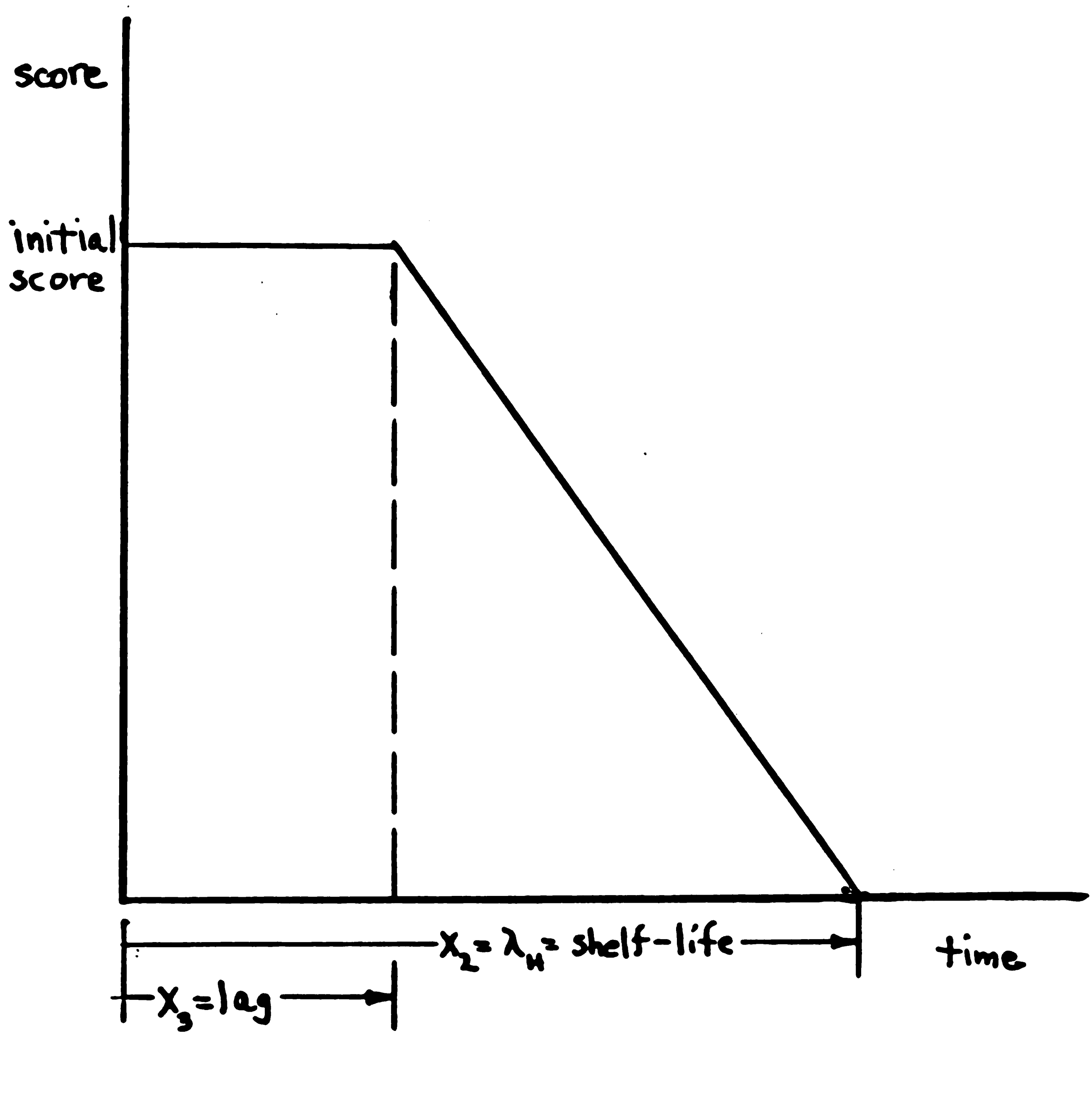


Figure 1

HISTOGRAMS OF MEAN SCORES AT 12 MONTHS

MRE CONSUMER ACCEPTANCE RESULTS

SCORE	TEMP= 4 DEG	21 DEG	30 DEG	38 DEG
8.0				
7.9				
7.8				
7.7	40D			
7.6			33D	17D
7.5	31D39S	38S39S	16S39S	
7.4		16S33S15D		
7.3	16D38D	40D36S17S	15S	16D33S
7.2	36S17S			
7.1	27D21S33D		40D17S	38D36S
7.0	37S23S		11D	40D
6.9	49S	47D23S49S21S	49S32S	39D
6.8	15D50D43S	19S37S	23S21S36D	23S24S15D
6.7	34D18S47S09S20S52S19S32S	48S22S26S24S34D32S52S	42D31D20S22S	26S
6.6	51S	31D20S09S27S50D	34D38D01S24S48S26S50D	
6.5		51S18S05S	28D09D51D52S	44S22S18S21D
6.4	22S05S04D24D	25S29S	37D25S44D18D	09S49D05S45D
6.3	23S11D48D25S		07S13D29S	11D19D42S
6.2		11D07S	03S	34D51S25S50D
6.1	42S	10S01D	45S	28D47D37D
6.0	10S26D	28D04D13D	05D19D	07S41S
5.9	30D07D		47D10D12D	20D29D27D10S
5.8	03D02S29D14S	43D30D	14S02S	52D01D
5.7	13D	14S02S	04D30D	12S03D32D08S
5.6		41D		30D04D
5.5	01D		41D08S	14S
5.4	41D	08S12D		31D
5.3	46S08S	03D42D		48D
5.2	12D	06D	06D27D	
5.1	06D	46S		02D
5.0				
4.9				13D
4.8				
4.7				06D
4.6				
4.5				
4.4				
4.3				
4.2				
4.1				
4.0				

Figure 2

```
FOODNAME        T  H  N
********       ** ** **
FRUIT MIX       0  4 34 789677446976998789787368776678797867
FRUIT MIX      12  4 34 687688467578679787967647787765773778
FRUIT MIX       0 21 34 789677446976998789787868776678797867
FRUIT MIX      12 21 34 764868538848274877698768868978667777
FRUIT MIX       0 30 34 789677446976998789787868776678797867
FRUIT MIX       6 30 34 798976868788527856886778867885448378
FRUIT MIX      12 30 34 874848798786977786837777765286567776
FRUIT MIX       0 38 34 789677446976998789787868776678797867
FRUIT MIX       6 38 34 587587867478788478858788878787877788
FRUIT MIX      12 38 34 487447766637778888865434477787486958

     T = STORAGE TIME, MONTHS
     H = STORAGE TEMP.,DEG C
     N = FOOD ITEM NUMBER
```

Figure 3

```
  FOOD NO. 34 FRUIT MIX     AT TEMP  4 DEG C
CONTINGENCY TABLES: CHISQ=   5.2095 DF=  6. PROB= .51723
 3      4      5      6      7      8      9
  .00   2.00    .00   7.00  13.00   8.00   6.00  36.00
 1.00   2.00   2.00   7.00  15.00   7.00   2.00  36.00
 1.00   4.00   2.00  14.00  28.00  15.00   8.00  72.00


  .50    .00   1.00    .00    .07    .03   1.00   2.60
  .50    .00   1.00    .00    .07    .03   1.00   2.60
 1.00    .00   2.00    .00    .14    .07   2.00   5.21


CONTINGENCY TABLES: CHISQ=    .6050 DF=  1. PROB= .43666

LIN. REGR. AT TIME 12 MONTHS, USING 2 TIMES
FROM TTEST: DIFF OF MEANS=   .4722 TAIL PROB=   .1254
TAIL-PROBS: FROM CT   .4367 FROM TTEST   .1254

  FOOD NO. 34 FRUIT MIX     AT TEMP 21 DEG C
CONTINGENCY TABLES: CHISQ=   6.6050 DF=  7. PROB= .53218
 2      3      4      5      6      7      8      9
  .00    .00   2.00    .00   7.00  13.00   8.00   6.00  36.00
 1.00   1.00   3.00   1.00   7.00  10.00  11.00   2.00  36.00
 1.00   1.00   5.00   1.00  14.00  23.00  19.00   8.00  72.00


  .50    .50    .10    .50    .00    .20    .24   1.00   3.03
  .50    .50    .10    .50    .00    .20    .24   1.00   3.03
 1.00   1.00    .20   1.00    .00    .39    .47   2.00   6.06


CONTINGENCY TABLES: CHISQ=   1.0473 DF=  1. PROB= .30614

LIN. REGR. AT TIME 12 MONTHS, USING 2 TIMES
FROM TTEST: DIFF OF MEANS=   .5278 TAIL PROB=   .1303
TAIL-PROBS: FROM CT   .3061 FROM TTEST   .1303
```

Figure 4

```
  FOOD NO. 34 FRUIT MIX     AT TEMP 30 DEG C
CONTINGENCY TABLES: CHISQ=   12.5048 DF= 14. PROB= .56583
  2      3      4      5      6      7      8      9
  .00    .00   2.00    .00   7.00  13.00   8.00   6.00  36.00
 1.00   1.00   2.00   3.00   5.00   8.00  14.00   2.00  36.00
 1.00   1.00   2.00   2.00   6.00  14.00   8.00   2.00  36.00
 2.00   2.00   6.00   5.00  18.00  35.00  30.00  10.00 108.00


  .67    .67    .00   1.67    .17    .15    .40   2.13   5.85
  .17    .17    .00   1.07    .17   1.15   1.60    .53   4.85
  .17    .17    .00    .07    .00    .47    .40    .53   1.80
 1.00   1.00    .00   2.80    .33   1.77   2.40   3.20  12.50

CONTINGENCY TABLES: CHISQ=     .7855 DF=  2. PROB= .67521

LIN. REGR. AT TIME 12 MONTHS, USING 3 TIMES
PROB= .11765,PCT VAR EXPL=  2.20,RESID SD= 1.494
SLOPE= -.04630,ST.ERR=   .02935, INCPT= 7.148,ST.ERR= .227
PRED SHELF-LIFE= .4640+02,90% LOW CONF-LIM= .2883+02
TAIL-PROBS: FROM CT  .5658 FROM TTEST  .1176

  FOOD NO. 34 FRUIT MIX     AT TEMP 38 DEG C
CONTINGENCY TABLES: CHISQ=  29.3044 DF= 12. PROB= .00355
  3      4      5      6      7      8      9
  .00   2.00    .00   7.00  13.00   8.00   6.00  36.00
  .00   2.00   3.00   1.00  13.00  17.00    .00  36.00
 2.00   7.00   2.00   5.00  10.00   9.00   1.00  36.00
 2.00  11.00   5.00  13.00  36.00  34.00   7.00 108.00


  .67    .76   1.67   1.64    .08    .98   5.76  11.56
  .67    .76   1.07   2.56    .08   2.83   2.33  10.31
 2.67   3.03    .07    .10    .33    .48    .76   7.44
 4.00   4.55   2.80   4.31    .50   4.29   8.86  29.30

CONTINGENCY TABLES: CHISQ=   7.1467 DF=  2. PROB= .02803

LIN. REGR. AT TIME 12 MONTHS, USING 3 TIMES
PROB= .00505,PCT VAR EXPL=  7.18,RESID SD= 1.380
SLOPE= -.07870,ST.ERR=   .02749, INCPT= 7.324,ST.ERR= .213
PRED SHELF-LIFE= .2963+02,90% LOW CONF-LIM= .2202+02
TAIL-PROBS: FROM CT  .0036 FROM TTEST  .0051
```

Figure 5

```
**** NON-LINEAR LEAST SQUARES RESULTS, USING NL2SOL ****

  FOOD NO. 34 FRUIT MIX    AT TEMP 38 DEG C
NON-LIN. REGR., TIME 12 MONTHS, TIME-STEPS= 3

CONV= 4 ITNS= 6 S USE= 1 F= .1020+03 MAXCOS= .1623-05
EST. STD DEVN= .1394+01 RECIP COND OF HESSIAN= .1023-02

       SOLN       GRAD       SD OF X     T-VAL.     PROB
  1 .71944+01  .1549-03   .1667+00   .4317+02   .0000
  2 .20710+02 -.5393-05   .3289+01   .6297+01   .7190-08
  3 .54194+01  .2231-04   .1720+01   .3152+01   .2117-02
PRED. SHELF-LIFE=   20.71, 90% LOW CONFLIM=   16.50

PRED SCORES, MO.,SCORE:   18. 5.39
```

Figure 6

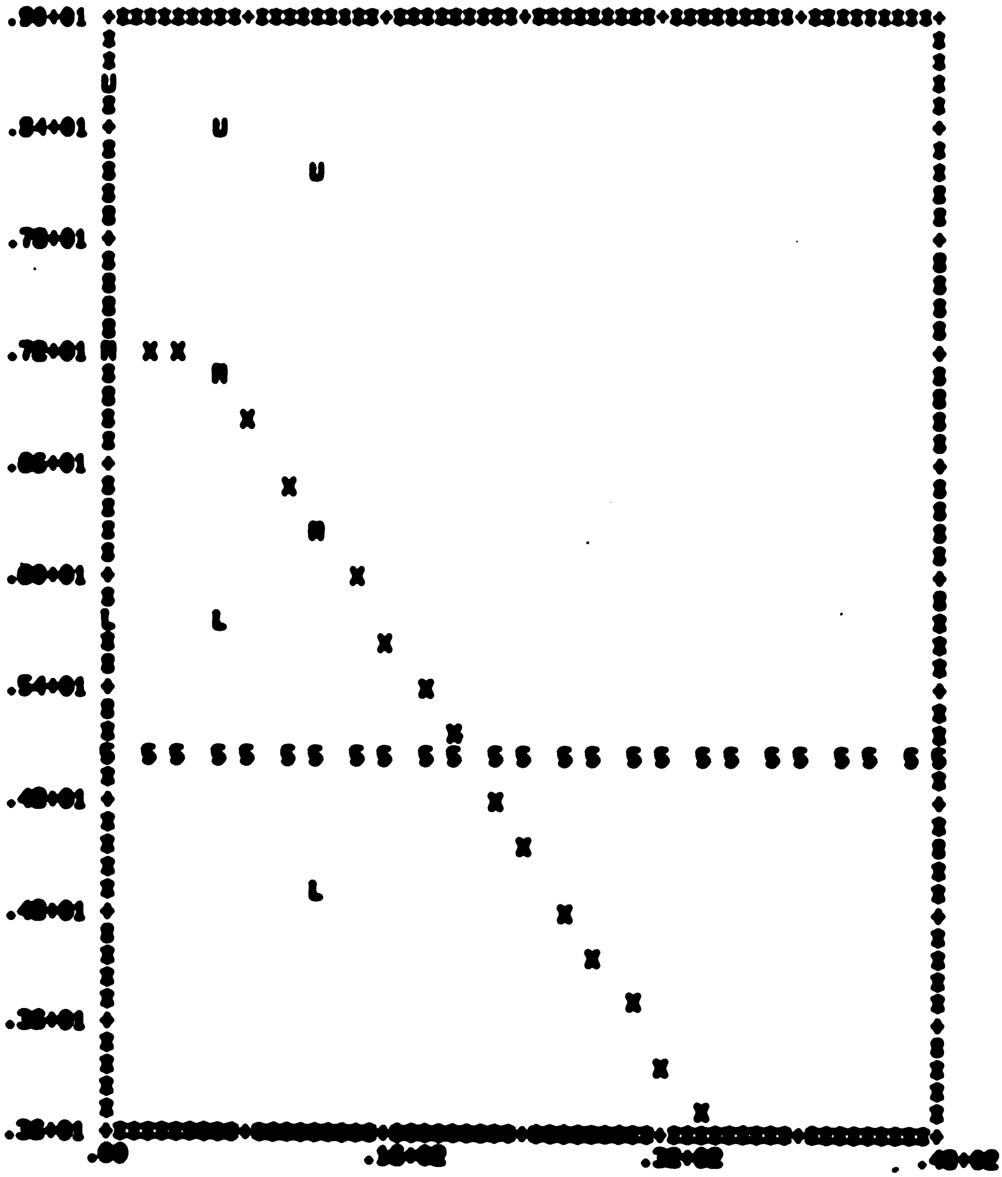


Figure 7

```
***** MULTINOMIAL LOGIT ANALYSIS ******


  FOOD NO. 34 FRUIT MIX    AT TEMP 38 DEG C

F=  180.5 ITNS= 6
GR NORM= .6912-02 REL FCH= .1185-04 ST NORM= .2104-01
   U(J)      GRAD(J)
-.94502+00 .59987-02
 .10875+00 .31998-02
 .56302+00 .11794-02
-.37998-02-.12612-03
 .86026+00-.38032-03
-.39885+00-.65163-04
HESSIAN: RECIP COND NO.= .2555-03 INFO=  0

OBSERVED AND ESTIMATED COUNTS:
T=  0. MOS:    .00   .00   .00  2.00   .00  7.00 13.00  8.00  6.00
               .02   .03   .10   .39  1.64  5.58 11.79 11.96  4.50
T=  6. MOS:    .00   .00   .00  2.00  3.00  1.00 13.00 17.00   .00
               .09   .14   .36  1.05  3.05  7.17 11.16  9.43  3.55
T= 12. MOS:    .00   .00  2.00  7.00  2.00  5.00 10.00  9.00  1.00
               .35   .58  1.16  2.49  5.00  8.11  9.30  6.55  2.47
SHELF-LIFE=   20.3 90% LOW CONFLIM=   10.8
STD ERR= .74194+01 ITNS= 10

PREDICTED COUNTS AT FUTURE TIMES
T= 18. MOS:   1.16  1.25  3.04  4.76  6.57  7.37  6.22  3.65  1.38
 MEAN SCORE= 5.42
```

Figure 8

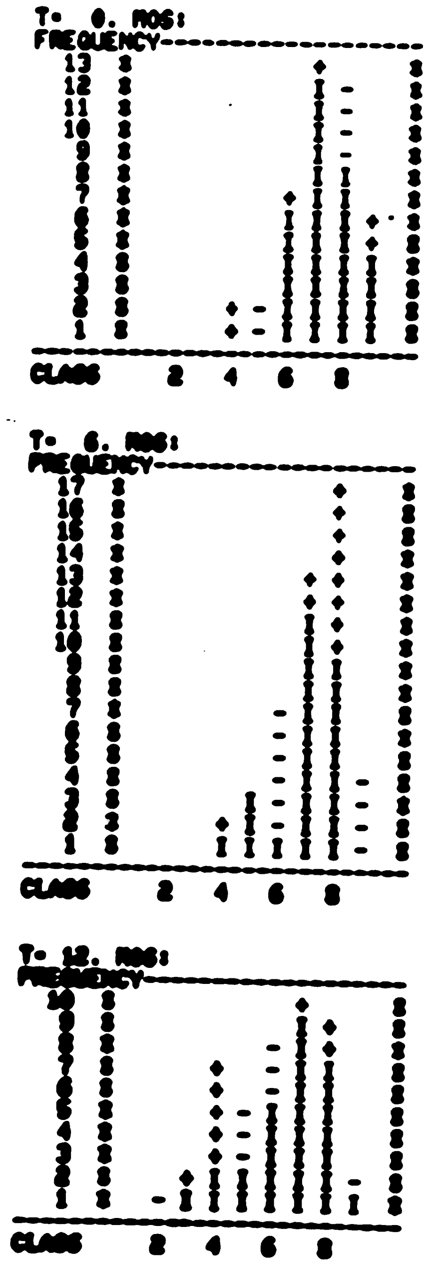

T= 0. MOS:
FREQUENCY
CLASS 2 4 6 8
T= 6. MOS:
FREQUENCY
CLASS 2 4 6 8
T= 12. MOS:
FREQUENCY
CLASS 2 4 6 8

Figure 9

ASSURING QUALITY THROUGH
BALLISTIC TESTING

Michael P. McMiller
Product Assurance Directorate
US Army Aviation Research and
Development Command
St. Louis, MO

ABSTRACT. The V_{50} test is a standard test used to measure the ballistic tolerance of armor material. This test is based on the assumption that the armor material lot is homogeneous. This paper poses the question: "If the lot of armor to be tested is not homogeneous is the V_{50} test still appropriate?" Also, how should the Operating Characteristics (OC) Curve be determined given the constraints detailed later in this paper?

I. INTRODUCTION.

In late March 1981, personnel from the Army Aviation Research and Development Command (AVRADCOM) witnessed a ballistic tolerance test. This test was the culmination of an effort to substitute test coupons in lieu of helicopter pilot and copilot crew seats. These seats are required to be ballistically tolerant to a 7.62 mm projectile at 2700 feet per second (fps). Because of the high cost of shooting the seats for production acceptance, a coupon test using representative materials was developed. The test demonstrated that the coupon failed to provide the required ballistic tolerance.

The test results were contradictory to all of the ballistic experience accumulated up to that time and an investigation into the reason for failure was undertaken. The seat and coupons consisted of a boron-carbide ceramic plate bonded to a Kevlar frame. The first step was to determine whether the Kevlar or the ceramic was at fault. Since there is no mechanical property or combination of properties which can accurately predict ballistic tolerance, coupons must be subjected to actual weapons fire to determine the ballistic tolerance.

II. THE V_{50} TEST.

In order to measure the ballistic protection afforded by an armor material, a V_{50} test is performed. V_{50} is defined as that projectile velocity which results in complete penetration 50% of the time and partial penetration 50% of the time. The definitions of complete and partial penetration depend on the ballistic protection criteria being used. During the testing the Protection Ballistic Limit (PBL) criterion was used. Using this criterion, a complete penetration occurs whenever a fragment or fragments are ejected from the back of the armor with sufficient remaining energy to pierce a "witness plate". The "witness plate" is a thin sheet of aluminum alloy placed 6 inches behind and parallel to the armor plate. If light can be seen through punctures in the witness plate, the penetration is complete. If not, the penetration is partial.

The V_{50} definition results from the fact that a bullet velocity sufficent to penetrate one coupon may result in only a partial penetration in another coupon. The relationship between projectile velocity and the ballistic tolerance satisfies the mathematical conditions of a probability distribution. For low projectile velocities the probability of a complete penetration approaches zero, for high velocities the probability of a complete penetration approaches one. Between those extremes of velocity the probability increases with increasing velocity. When that general model describes the physical events, probability of penetration can be treated as a probability distribution and is usually described as a Gaussian or normal distribution.

The procedure to experimentally determine V_{50} is as follows: The first round shall be loaded with an amount of propellant calculated to give the projectile a velocity of 2750 (fps) for the specified ballistic limit of 2700 (fps). Each succeeding round shall be loaded with an amount of propellant calculated to produce a velocity change of 25 to 50 fps. The criteria to determine whether an increase or decrease of velocity is required is as follows:

a. If the preceeding velocity resulted in a partial penetration, the charge will be increased to produce a velocity increase.

b. If the preceeding velocity resulted in a complete penetration, the charge will be decreased to produce a velocity decrease.

c. A minimum of six shots will be required to determine each V_{50}. The V_{50} is equal to the average of six impact velocities comprising the three lowest velocities resulting in complete penetration and the three highest velocities resulting in partial penetration. Additional shots are permitted if after six impacts three complete penetrations and three partial penetrations have not been achieved.

III. PROBLEM WITH THE V_{50} TEST.

Table I summarizes the evidence which suggests that the assumption of homogeneity may not be correct.

TABLE I

TEST 1		TEST 2	
2742	Partial	2740	Partial
2776	Partial	2779	Complete
2792	Partial	2755	Complete
2849	Complete	2741	Complete
2822	Complete	2702	Complete
2804	Complete	2687	Complete

Both tests were drawn from a single lot of ceramic and Kevlar. Thus the two tests should have yielded near identical V_{50} values. However, they showed significantly different ballistic tolerance levels. The test apparatus was checked and rechecked to make certain that the fault did not lie with the test equipment. No fault with the equipment could be found.

Throughout the testing similar incidents occurred, incidents which suggested that there may be some defective panels either in a single lot of Kevlar or a single lot of ceramic. Put another way, the testing suggested that a single lot of Kevlar or ceramic was not homogenous. There could be panels in a lot which have extremely poor ballistic tolerance, much lower than had been evident previous to the March test.

It is interesting to note that the nature of information that a complete penetration provides is not the same as that provided by a strength test where the stress at rupture is of concern. In the strength test a defective specimen is noticeable by its unusually low rupture stress. The stress at rupture can be plotted and those specimens with very low rupture strength can be grouped and investigated further or eliminated from the data base. However, a velocity of a complete penetration does not provide the same type of data. A complete penetration at a specified velocity only shows that the coupon will not stop a bullet at that velocity. It does not provide any information on the bullet velocity the coupon will stop. Thus, it is necessary to take like coupons and continue to shoot until the bullet is stopped. Like coupons are assumed to be coupons made from a single lot of Kevlar and a single lot of ceramic.

If there are some defects within a lot of Kevlar or ceramic as suggested by the investigation then the question is, "Is the V_{50} test still a reasonable test to use during production acceptance testing?".

IV. THE OPERATING CHARACTERISTICS (OC) CURVE.

It became obvious during the course of the investigation that an OC Curve had to be developed for this armor material. But again, the fact that there may be defects in a lot will affect the data for the OC Curve. To complicate matters the lots of each material are of limited size. A lot of Kevlar will make only 12 test coupons while a lot of Boron-Carbide will make 28. How then should the test be designed to determine the OC Curve that will screen out the effect of the defectives?

V. CONCLUSION.

This author poses the following two questions to the 27th Conference on the Design of Experiments in Army research:

a. Given that a lot of coupons may contain some defectives is a V_{50} test appropriate for production acceptance?

b. How can a test be designed to determine the OC Curve that minimizes the presence of defects?

COMPARING SEVERAL GROUPS IN A TWO-WAY LAYOUT SETTING

DOUGLAS A. WOLFE

Department of Statistics
The Ohio State University
Columbus, Ohio 43210

ABSTRACT. Consider the problem of evaluating the relative merits of several drugs that hold promise for treating a certain disease. (Of course, one or more of these drugs could be placebos or currently standard controls.) Suppose that these drugs are administered to a set of patients similarly affected by the disease of interest, and that the effects of the drugs are evaluated by recording repeated observations on the patients over a fixed period of time. Within this framework there are many questions of interest. For example, do the patient groups react to the several drugs in a similar fashion over time? If the drugs' effects do differ, which of the drugs achieve the greatest degree of control over the disease? Which of the drugs reach their peak effectiveness most rapidly? The purpose of this talk is not to discuss the relative merits of how to rank for such two-way layout problems (i.e., within-blocks versus between-blocks rankings), but to instead consider the more basic question of what to rank to best answer the questions of interest. Several rank-like approaches to some of the mentioned problems are discussed.

1. INTRODUCTION. Since the pioneering work of Friedman (1937), Kendall and Babington-Smith (1939), and Kruskal and Wallis (1952) there has been a steady flow of research activity in the area of nonparametric procedures for the one- and two-way layout settings. However, there has been very little progress in the development of satisfactory nonparametric procedures for analyzing data involving three or more factors. In this paper we discuss a direction for future research that could help alleviate at least one of the impediments to such progress. Our approach hinges on the fact that the intuitive criteria utilized by nonparametric researchers differ between the one- and two-way layout settings, and that this should suggest an even different set of criteria is necessary for three or more factor experiments. We begin by considering the most common concerns put forth in the development of nonparametric procedures for the one- and two-way layout settings.

2. ONE-WAY LAYOUT. Let X_{ij}, $i=1,\ldots,n_j$ and $j=1,\ldots,k$, be k independent random samples from continuous distributions with distribution functions $F_j(x) = F(x-\tau_j)$, $j=1,\ldots,k$, where $F(\cdot)$ is unspecified. Most of the interest for this setting has been with developing effective hypothesis tests of H_0: $[\tau_1=\ldots=\tau_k]$, where the word effective has generally been related to power considerations for such tests.

What has been the primary motivation behind the different approaches to providing "effective" tests for these one-way layout data? It has certainly not been an inclination to use different ways to "nonparameterize" the information in the k samples. I think that it is safe to say that virtually every nonparametric one-way layout test has as its first step the replacement of the original observations by their combined samples ranks from least to greatest. That is, there is no distinction between competing procedures on the basis of what ranking scheme is used in extracting the "nonparametric" information from the data--they all use combined samples ranks of the original observations.

What then distinguishes between the procedures? Of primary importance has been the desire to design a test that would be especially powerful against either (i) a particular parametric model or (ii) a general class of alternative hypotheses of interest. Thus, for example, the parametric motivation in (i) led researchers to develop such criteria as locally most powerful or asymptotically most powerful rank tests as discussed, for instance, in Randles and Wolfe (1979). Examples of such tests are the one-way layout normal scores procedures.

Interest in the class of alternatives approach in (ii) has resulted in, among others, the test procedures proposed by: Kruskal and Wallis (1952) for general alternatives of the form H_1: [$\tau_i \neq \tau_j$ for at least one $i \neq j$]; Jonckheere (1954) and Terpstra (1952) for ordered alternatives of the form H_1: [$\tau_1 \leq \ldots \leq \tau_k$, with at least one strict inequality]; and, most recently, Mack and Wolfe (1981) for umbrella (quadratic) alternatives of the form H_1: [$\tau_1 \leq \ldots \leq \tau_{\ell-1} \leq \tau_\ell \geq \tau_{\ell+1} \geq \ldots \geq \tau_k$, with at least one strict inequality], where ℓ, the peak of the umbrella, is either known or unknown. Other applications of criterion (ii) have been carried out for the k-sample slippage problem and in treatments versus control settings. For the latter model, attention has also been given in Costello and Wolfe (1980) to collecting the treatments observations in a partially sequential manner.

3. <u>TWO-WAY LAYOUT</u>. In complicating the model by adding a second factor of interest, the two-way layout also introduces another area of variability in the proposed approaches to related problems. Both criteria

that were important in differentiating between one-way layout procedures remain important here as well. Thus, for example, we have the general alternatives procedure due to Friedman (1937) and Kendall and Babington-Smith (1939), as well as the ordered alternatives approach suggested by Page (1963), all of these dealing with the situation where there is one (no missing data) and only one (no replications) observation collected for each combination of the two factors. General alternatives procedures for the case of zero or one observation for each factor combination have been proposed by Durbin (1951) when we have a balanced incomplete block design and by Skillings and Mack (1981) for a more general case of arbitrarily missing data. For replications within the factor combinations, general alternatives procedures have been considered by Mack and Skillings (1980) and Mack (1981), while a similar procedure for ordered alternatives was studied by Skillings and Wolfe (1977, 1978) and Skillings (1980).

However, discussion over the appropriate procedure to use in a two-way layout setting has not been limited to the factors of (i) parametric model or (ii) alternatives of interest, as has been the case in the one-way layout. For two-way layout data, we also see considerable discussion on a very basic third factor, namely, (iii) <u>how</u> to rank the collected observations. To briefly describe this discussion, let X_{ij}, $i=1,\ldots,n$ and $j=1,\ldots,k$, be mutually independent, continuous random variables with X_{ij} having distribution function $F_{ij}(x) = F(x-\tau_j-\beta_i)$, where $F(\cdot)$ is unspecified. Thus the τ's represent the effects of the various levels of one of the factors and the β's represent the effects of the various levels of the second factor. (Note that an additive model is usually assumed. The problem of interaction has been particularly thorny in nonparametric statistics.) In discussing tests of H_0: $[\tau_1=\ldots=\tau_k]$ against a particular alternative of interest in this two-way layout setting we are also faced with the problem of how to rank the data: Do we rank observations only within levels of the second factor (i.e., rank $X_{i1},\ldots,X_{ik}$ separately for each $i=1,\ldots,n$) or is there some (appropriate) way that we can effectively rank all kn of the X_{ij} values together, as is done in the one-way layout setting? If this joint ranking can be legitimately (without comparing apples and oranges) accomplished without undue complication, it should produce reasonable competitors to those based on ranking only within the levels of the second factor.

All of the two-way layout procedures previously mentioned in this paper utilize the within-levels ranking scheme. Hodges and Lehmann (1962) were pioneers in the area of jointly ranking all of the observations when they suggested using aligned ranks in constructing appropriate conditional test procedures. Doksum (1967) and Hollander (1967) considered other ways to use between block information and still obtain at least asymptotically distribution-free tests. Mehra and Sarangi (1967) studied the power properties of some of the within-levels ranking procedures relative to those based on joint ranking schemes. The verdict on how to rank is not unanimous.

4. SEVERAL TWO-WAY LAYOUT GROUPS. This brings us to the actual title of the talk, namely, comparing several groups in a two-way layout setting. For $j=1,\ldots,k$ and $i=1,\ldots,n_j$, let $(X_{ij1},\ldots,X_{ijm})$ be mutually independent, continuous random vectors such that, for each fixed $j\varepsilon\{1,\ldots,k\}$, the n_j vectors $(X_{1j1},\ldots,X_{1jm}),\ldots,(X_{n_j j1},\ldots,X_{n_j jm})$ are identically distributed with joint distribution function $F_j(x_1,\ldots,x_m)$ and median vector $(\tau_{j1},\ldots,\tau_{jm})$.

In this section we consider general distribution-free approaches to constructing hypothesis tests about the τ vectors. First, however, it might be helpful to discuss the applied setting that led to our interest in such problems. Consider k different drugs that are potentially useful for treating a certain illness. (One or more of the drugs could certainly be control-standards or control-placebos.) These drugs are administered to patients with the prescribed illness and the effects of the drugs are recorded over a specified period of time. That is, each administration of one of the drugs to a patient results in repeated (dependent) measurements on the same subject over time. We are, of course, interested in potential treatment effect differences among the k drugs over the involved time period.

To set this problem in our stated model, we take X_{ijs} to be the measurement at the sth time point for the ith subject being treated with the jth drug. We thus have n_j patients taking the jth drug, $j=1,\ldots,k$, and being evaluated at m distinct time points, and our interest is in making inferences about the relative treatment-time effects of the k drugs.

Similar problems in the context of testing for agreement between two groups of judges have been considered by Schucany and Frawley (1973), Li and Schucany (1975), Schucany and Beckett (1976), and Hollander and Sethuraman (1978). The Schucany-Frawley-Li test is based on the average value of an appropriate series of Spearman correlations between rankings from one group of judges and rankings from the other group of judges. However, Hollander and Sethuraman suggested some possible problems in the consistency class and designated null hypothesis for the Schucany-Frawley-Li procedure, and they proposed a solution based on a conditionally distribution-free permutation test utilizing the Mahalanobis D^2 statistic.

Regardless of the relative merits of these competing procedures for the problem of two groups of judges, it is not obvious how either of them would be naturally extended to either the more general problems posed by competing drug studies or, even in the context of their problem, to more than two groups of judges. For these reasons we approach the drug evaluation problems from a different viewpoint in this paper.

We have previously noted that power considerations against certain classes of alternatives have been the primary motivations behind the development of most distribution-free one-way layout test procedures; that is, how to extract the important information from the agreed-upon ranking method has been paramount. To this criterion is added the problem of how to rank the collected data in a more complicated two-way layout setting. When we extend this one step more to the consideration of several groups in a two-way layout setting, we suggest that a third criterion, namely that of what to rank, should be given at least as much (and probably most) attention in developing appropriate test procedures. Thus instead of automatically presuming that our tests should be based on some function of some method for ranking the sample observations themselves, perhaps it would be beneficial to at least consider if there are other quantities that could be effectively ranked to address our questions. Such rank-like (i.e., ranking of quantities other than the collected data values) techniques have been proposed by Fligner and Killeen (1976), Fligner, Hogg and Killeen (1976), Brofitt, Randles and Hogg (1976), and Smith and Wolfe (1977), and are discussed in Randles and Wolfe (1979). Similar ideas have also been utilized by Koch (1972) in dealing with the use of nonparametric methods in a two-period change-over design.

Now, returning to our problem of evaluating several drug treatment groups in a two-way layout setting, we demonstrate this idea of the importance of what to rank through a series of examples dealing with different alternatives of interest.

Example 4.1. Suppose we wish to know which of the drugs under consideration achieves the greatest peak effectiveness. This could correspond to either highest or lowest measurement values, depending on the nature of the data being collected. For purposes of this paper, we will take large values to mean good effectiveness of a drug. Letting $X^*_{ij} = \underset{1 \leq s \leq m}{\text{maximum}} (X_{ijs})$, for $i=1,\ldots,n_j$ and $j=1,\ldots,k$, represent the maximum measurement value achieved by the ith subject on the jth treatment drug, we see that the null hypothesis of interest here could be taken to be H^*_0: $[\theta^*_1=\ldots=\theta^*_k]$, where θ^*_j represents the median of the distribution of X^*_{ij}, for $i=1,\ldots,n_j$. [Note that θ^*_j is analogous to $\max_{1 \leq s \leq m} \tau_{js}$.] Appropriate procedures for testing H^*_0: $[\theta^*_1=\ldots=\theta^*_k]$ can be based on the joint rankings of the $\sum_{j=1}^{k} n_j$ X^*_{ij} values. What particular method of evaluating this ranking information should be used will still depend on the alternative to H^*_0 that is of interest. For example, if the treatment drugs are such that H^*_1: $[\theta^*_1 \leq \ldots \leq \theta^*_k$, with at least one strict inequality] is appropriate, then we could apply the Jonckheere (1954) procedure to the X^*_{ij}'s, while for general alternatives the Kruskal-Wallis (1952) would be preferred. The main point is that such procedures would be applied to the X^*_{ij}'s, not the original data.

Example 4.2. In this example we would like to evaluate which of the drugs is quickest to achieve its peak effectiveness. Thus, letting $\xi_j = \underset{1 \leq s \leq m}{\text{maximum}} \tau_{js}$ and taking t_j to be the time point for which $\tau_{jt_j} = \xi_j$, we are here interested in testing H_0: $[t_1=\ldots=t_k]$ against general or ordered alternatives, for example. To construct a distribution-free test for this setting, we can again use the rank-like idea. Let X^*_{ij} be as defined in Example 4.1 and consider the sample time points where the various drugs reach their greatest effects (as measured by the X_{ij}^* values) on the patients to which they were given. Setting

$$N_{js} = [\text{number of patients given the jth drug that achieved their maximum measurement value at the sth time point}],$$

for $j=1,\ldots,k$ and $s=1,\ldots,m$, we could then test H_0: $[t_1=\ldots=t_k]$ against either general or ordered alternatives by applying an appropriate procedure for testing equality of multinomials to the k sets of counts $(N_{j1},\ldots,N_{jm})$, for $j=1,\ldots,k$. (Note that statistics other than the $\underset{1\leq s\leq m}{\text{maximum}}\,(X_{ijs})$'s could be used to indicate when the peak effectiveness of a drug is reached. For example, the peak-picker employed by Mack and Wolfe (1981) could be used effectively here as well.)

<u>Example 4.3</u>. Consider the problem discussed in Example 4.2 but with the further assumption that we know the magnitude of the peak effectiveness is the same for the k drugs. Thus, in this example we want to test H_0: $[t_1=\ldots=t_k]$ under the additional information that $\xi_1=\ldots=\xi_k$. In such a situation we are able to describe an exact distribution-free rank-like test that is a competitor to the approximate multinomial test discussed in Example 4.2. Let $\hat{\beta}_{ij}$ represent the slope of the line connecting X_{ij1} and $X_{ij}^* = \underset{1\leq s\leq m}{\text{maximum}}\, X_{ijs}$, for $i=1,\ldots,n_j$ and $j=1,\ldots,k$. (That is, $\hat{\beta}_{ij} = \dfrac{X_{ij}^*-X_{ij1}}{T_{ij}-T_0}$, where T_{ij} is the time point corresponding to X_{ij}^* and T_0 is the time point for the initial measurement.) An appropriate (depending on alternatives of interest) nonparametric distribution-free one-way analysis of variance procedure could then be applied to the k sets of estimated slopes. Such a test would be exactly distribution-free under the hypothesis of no difference in time effectiveness for the k drugs and would be especially powerful at detecting differences in the peak time points $t_1,\ldots,t_k$.

Actually the rank-like technique discussed in Example 4.3 could be useful even if we do not know that all the drugs have a common peak effectiveness. However, we would then have to be willing to accept both early peaking and larger, but slower-achieved peaks, as indicative of an effective drug, since either of these occurrences in the data would lead to large estimated $\hat{\beta}_{ij}$ slope values.

Example 4.4. As a final example consider the problem of evaluating whether the overall time-collected reactions of patients are similar for the k drugs in the study. (The settings discussed in Examples 4.1, 4.2, and 4.3 address particular aspects of this problem.) To use a rank-like procedure for this general question, we must first settle on a within-subjects statistic that is representative of our interest in the time-collected data. For example, we might wish to assume a straight line regression relationship between the measurements being collected and the times at which the data are obtained. If so, then a statistic such as an estimator for the slope of the regression line would be a logical candidate for comparisons between the k drug groups. That is, we would obtain estimates, $\hat{\beta}^*_{ij}$, of the slopes associated wish each of the $\sum_{j=1}^{k} n_j$ individuals in the study. (What method of estimation (e.g., least squares, median of all sample slopes, as discussed in Section 9.3 of Hollander and Wolfe (1973), etc.) is used to obtain the $\hat{\beta}^*_{ij}$ values is not important for maintaining the distribution-free property of the process. It is only necessary that the same method be used for all the individuals. Of course, the choice of estimation criterion could indeed have an effect on the power properties of the resulting distribution-free test.) After obtaining these individual slope estimators, we would then proceed as in Example 4.3 by applying an appropriate (depending on alternatives of interest) nonparametric distribution-free one-way analysis of variance procedure to the k groups of condensed data.

If a regression model more complex than a straight line is necessary to relate the sample observations and the time points at which the data are collected, a similar approach can be used to develop an appropriate distribution-free test. All that is required is some summary measure to represent this regression model for each of the individual subjects. So long as the <u>same</u> statistic is computed for each of the individuals, the resulting analysis of variance test will be nonparametric distribution-free for the null hypothesis of no difference in regression time effects for the k drugs. This will remain the case no matter how complex we make either the regression model or the summary statistic. (Similar approaches could also be taken even if there were multiple or missing observations for each individual at some time points.)

<u>5. DISCUSSION</u>. The primary intent of this paper has not been to propose and study a new test for a given problem, but rather to re-emphasize the flexibility that is available in constructing nonparametric distribution-free tests of hypotheses. Keeping in mind the well-established advantages of tailoring tests to alternatives of interests and the potential gains from consideration of different methods of ranking the actual collected data, we have suggested that an even more basic question of <u>what to rank</u> can play an important role in more complicated problems such as comparing several groups in a two-way layout setting. The use of such rank-like (i.e., ranking something other than the original data points) would also seem to have both appeal and merit for other problems, such as with multivariate data, where the usual nonparametric approaches have proven to be less than totally effective.

REFERENCES

Brofitt, J. D., Randles, R. H., and Hogg, R. V. (1976). Distribution-free partial discriminant analysis. J. Amer. Statist. Assoc. 71, 934-9.

Costello, P. and Wolfe, D. A. (1980). Partially sequential treatments versus control multiple comparisons. Biometrika 67, 403-12.

Doksum, K. (1967). Robust procedures for some linear models with one observation per cell. Ann. Math. Statist. 38, 878-83.

Durbin, J. (1951). Incomplete blocks in ranking experiments. Brit. J. Statist. Psychol. 4, 85-90.

Fligner, M. A., Hogg, R. V., and Killeen, T. J. (1976). Some distribution-free rank-like statistics having the Mann-Whitney-Wilcoxon null distribution. Comm. Statist.-Theor. Methods A5, 573-6.

Fligner, M. A. and Killeen, T. J. (1976). Distribution-free two-sample tests for scale. J. Amer. Statist. Assoc. 71, 210-3.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Amer. Statist. Assoc. 32, 675-701.

Hodges, J. L., Jr. and Lehmann, E. L. (1962). Rank methods for combination of independent experiments in analysis of variance. Ann. Math. Statist. 33, 482-97.

Hollander, M. (1967). Rank tests for randomized blocks when the alternatives have an *a priori* ordering. Ann. Math. Statist. 38, 867-77.

Hollander, M. and Sethuraman, J. (1978). Testing for agreement between two groups of judges. Biometrika 65, 403-11.

Hollander, M. and Wolfe, D. A. (1973). Nonparametric Statistical Methods. Wiley, New York.

Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. Biometrika 41, 133-45.

Kendall, M. G. and Babington Smith, B. (1939). The problem of m rankings. Ann. Math. Statist. 10, 275-87.

Koch, G. G. (1972). The use of non-parametric methods in the statistical analysis of the two-period change-over design. Biometrics 28, 577-84.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. J. Amer. Statist. Assoc. 47, 583-621.

Li, L. and Schucany, W. R. (1975). Some properties of a test for concordance of two groups of rankings. Biometrika 62, 417-23.

Mack, G. A. (1981). A quick and easy distribution-free test for main effects in a two-way ANOVA. To appear in Comm. Statist.-Sim. Comp.

Mack, G. A. and Skillings, J. H. (1980). A Friedman-type rank test for main effects in a two-factor ANOVA. J. Amer. Statist. Assoc. 75, 947-51.

Mack, G. A. and Wolfe, D. A. (1981). K-sample rank tests for umbrella alternatives. J. Amer. Statist. Assoc. 76, 175-81.

Mehra, K. L. and Sarangi, J. (1967). Asymptotic efficiency of certain rank tests for comparative experiments. Ann. Math. Statist. 38, 90-107.

Page, E. B. (1963). Ordered hypotheses for multiple treatments: a significance test for linear ranks. J. Amer. Statist. Assoc. 58, 216-30.

Randles, R. H. and Wolfe, D. A. (1979). Introduction to the Theory of Nonparametric Statistics. Wiley, New York.

Schucany, W. R. and Beckett, J. (1976). Analysis of multiple sets of incomplete rankings. Comm. Statist.-Theor. Methods A5, 1327-34.

Schucany, W. R. and Frawley, W. H. (1973). A rank test for two group concordance. Psychometrika 38, 249-58.

Skillings, J. H. (1980). On the null distribution of Jonckheere's statistic used in two-way models for ordered alternatives. Technometrics 22, 431-6.

Skillings, J. H. and Mack, G. A. (1981). On the use of a Friedman-type statistic in balanced and unbalanced block designs. Technometrics 23, 171-7.

Skillings, J. H. and Wolfe, D. A. (1977). Testing for ordered alternatives by combining independent distribution-free block statistics. Comm. Statist.-Theor. Method A6, 1453-63.

Skillings, J. H. and Wolfe, D. A. (1978). Distribution-free tests for ordered alternatives in a randomized block design. J. Amer. Statist. Assoc. 73, 427-31.

Smith, T. M. and Wolfe, D. A. (1977). Distribution-free tests for equality of several regression equations. Tech. Rept., Department of Statistics, Ohio State University, Columbus, Ohio.

Terpstra, T. J. (1952). The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. Indag. Math. 14, 327-33.

A MATHEMATICAL BASIS FOR TRACKING MANEUVERING AIRCRAFT WITH DOPPLER RADAR

Donald W. Rankin
Army Materiel Test and Evaluation Directorate
US Army White Sands Missile Range
White Sands Missile Range, NM 88002

ABSTRACT. Continuous Wave Doppler Acquisition Radars have several distinct advantages when employed against aircraft. Among them are (1) illumination of the target with great power and (2) elimination of unwanted returns from stationary objects, achieved by means of a suitable band-stop filter.

However, information arrives as azimuth and radial velocity, a form unsuited to coordinate transformation. First, range must be computed. Essentially, this is a problem in numerical integration.

Traditionally, tracking algorithms simplify this integration problem by assuming that the acceleration and velocity vectors are collinear. The assumption holds quite well in the case of ballistic missiles (computing in an inertial coordinate system), but for maneuvering aircraft is manifestly erroneous.

This paper reexamines the equations of motion of maneuvering aircraft with a view toward finding assumptions upon which to base a mathematical model for an efficient tracking algorithm.

An algorithm is developed, based upon the assumption of constant aircraft turn rate. Preliminary results against synthesized targets are most satisfactory.

I. INTRODUCTION

During a recent test, the author was assigned to evaluate the software associated with the range tracking algorithm of a particular continuous wave (Doppler) acquisition radar, which was exhibiting unexplained anomalies. In the course of the investigation, it became increasingly apparent that the subject of tracking algorithms is not well-documented and that sound, basic reference material is not always readily available to the composer.

The evaluation led to the identification of a single critical target performance parameter, hitherto apparently unnoticed, upon which a mathematical development properly could be based. When this critical parameter is assumed to be constant, the ensuing development provides a rational basis for selecting an efficient mathematical model which has demonstrated excellent tracking performance during computer simulations.

In view of all this, it was decided to include a discussion of certain fundamental principles. It is hoped that some use may be found by others when composing or evaluating tracking algorithms.

a. REMARKS CONCERNING DATA HANDLING

Let us define two terms for which we shall find frequent use:

(1) Track file data -- left-over data, processed or not, still available in the computer

(2) Raw data -- incoming data, not yet processed

Now, if we place great confidence in the accuracy of the raw data, we simply use it for whatever purpose we wish and pay little attention to the track file data. Unfortunately, this is rarely the case. Usually, the raw data is contaminated by noise and possibly other errors.

In the latter event, the only criterion for evaluating the raw data lies within the track file data. Two methods commonly are use -- often combined:

(1) The "window." From considering performance characteristics of the target being tracked, it may be possible to say that the raw data cannot vary from the track file data by more than a specified amount, say w. If it does, it is presumed to come from a different source.

(2) The "filter." It is assumed that some value which lies between the raw data and the track file data is a better estimate than either of them alone. As an example, let us suppose the track file contains computed radial velocity and acceleration, $\dot{R}$ and $\ddot{R}$. Some Δt seconds later, raw radial velocity, $\dot{R}_m$, arrives. On the basis of track file information alone, we could have predicted a value for radial velocity at this time:

$$\dot{R}_p = \dot{R} + \Delta t \ddot{R}$$

Now, if

$$|\dot{R}_m - \dot{R}_p| > w$$

it is said that $\dot{R}_m$ falls outside the window and therefore is not associated with this track. If, however, $\dot{R}_m$ falls within the window, a filtered (or smoothed or updated) value of the radial velocity is computed by

$$\dot{R} = k\dot{R}_m + (1 - k)\dot{R}_p$$

An exactly equivalent expression is

$$\mathring{R} = \mathring{R}_p + k(\mathring{R}_m - \mathring{R}_p) \tag{1}$$

Note that in order to insure that the new value of $\mathring{R}$ lies between $\mathring{R}_m$ and $\mathring{R}_p$, it is necessary that $0 < k < 1$. (It should be noted in passing that some command and control systems may employ a value of $k > 1$. But in every instance of this, the filter is being used for some purpose other than smoothing.)

b. MISSILE OR AIRCRAFT? THE PROPER USE OF ACCELERATION

Probably there exists no such thing as a satisfactory universal tracking algorithm. It seems self-evident that a procedure which accurately tracks ballistic missiles might be worthless against helicopters. Each algorithm must be tailored to a specific purpose, ignoring those problem areas whose probability of occurrence must be quite small.

One indispensable step is to define areas and to simplify procedures by making certain arbitrary a priori assumptions, based upon knowledge of performance. Thus, an algorithm to track helicopters might well assume that no velocities will be encountered in excess of 100 m/sec. Such an assumption in a missile tracker, however, would be a serious error.

Suppose we wish to compose an algorithm for tracking ballistic missiles. Assuming acquisition outside the atmosphere after thrust is spent, acceleration will be zero (except for gravity, the effect of which can be removed by computing in a suitable moving coordinate system). That is, velocity will be virtually constant and the acquisition arithmetic becomes quite easy. At atmospheric re-entry, we assume that drag produces a negative acceleration which changes very slowly and is directed along the longitudinal axis of the missile (i.e., parallel to the velocity vector). Acceleration is obtained by differencing; hence, the value is "old," but easily can be updated. This updated value is used to predict velocity (and, in turn, position). This results in a very simple and satisfactory algorithm which should produce accurate results.

Passing to the problem of tracking aircraft, it can be shown that the foregoing missile-tracking algorithm is unsuitable. Let us examine the assumptions. First, the assumption of zero acceleration limits acquisition to aircraft in virtually straight, unaccelerated flight. For acquisition, the assumption of <u>constant</u> acceleration is far less restrictive and therefore preferable.

Second, to assume that the acceleration and velocity vectors are parallel is completely untenable. The aircraft maneuver most likely to occur, which produces measurable acceleration, is the simple turn. In a turn of constant angular rate (which all pilots try hard to achieve), the aircraft flies in a circle with the acceleration vector directed towards its center. Thus, the acceleration and velocity vectors are directed 90 degrees apart and are said to be out of phase. A knowledge of one is of little use in predicting the other. We are not left helpless in this situation, however.

Before proceeding further, let us pause to coin a new word and to define it. Let the time derivative of the acceleration be called the ACC'RATE. It will have the dimension $m.sec^{-3}$ or some equivalent. Suppose we imagine that an aircraft is flying in a circle at a constant speed and that we are observing it from a point sufficiently far away that azimuth can be ignored. Obviously, the relative velocity varies as cos θ, with θ being the central angle. Now the derivative of the cosine is just the negative of the sine, so that the relative acceleration varies as -sin θ. Extending the process one more step, it is seen that the relative acc'rate varies as -cos θ. What this means is that the ratio

$$\frac{\text{acc'rate}}{\text{velocity}}$$

is constant and forever negative (or zero)!

In the computer, derivatives are estimated by taking differences. Letting triple dots indicate acc'rate, we find that, with suitable scaling,

$$\dddot{R}_{n-1} \approx \dot{R}_n - 2\dot{R}_{n-1} + \dot{R}_{n-2} \qquad (2)$$

Letting q denote the estimate of the acc'rate ratio,

$$\frac{\dddot{R}_{n-1}}{\dot{R}_{n-1}} \approx \frac{\dot{R}_n - 2\dot{R}_{n-1} + \dot{R}_{n-2}}{\dot{R}_{n-1}} = q_{n-1}$$

Now $\frac{\dddot{R}_{n-1}}{\dot{R}_{n-1}}$ is constant and therefore q nearly so.

Treating q as a constant enables us to drop the subscript and write

$$\dot{R}_n + \dot{R}_{n-2} = (2 + q)\, \dot{R}_{n-1} \qquad (3)$$

Equation (3) can be used to compute q. Or, if q is known, it becomes a three-term recurrence relation for the successive values of $\dot{R}$. It should be observed that equation (3) is independent of the units in which Δt, R, and $\dot{R}$ are expressed.

c. COORDINATE SYSTEMS

It seems desirable to write math models and algorithms which will operate in two- or three-dimensional cartesian or inertial systems whenever possible. The data, of course, must be amenable to transformation. In the case of Doppler radar, it is not. The data are available as radial velocity and azimuth, essentially. Before a transformation can be effected, range must be determined -- and this very range reduction requires that a track file be established in that coordinate system in which we find ourselves.

The fact that we necessarily are operating in polar coordinates produces an apparent outward acceleration which depends upon both the range of the target and its relative angle of approach. The effect on a crossing track at short range can be appreciable. For example, an aircraft at a range of 6 km flying a straight-line path with a constant speed of 280 $m.sec^{-1}$ at a relative radial angle of 60 degrees appears to have an outward acceleration of 9.8 $m.sec^{-2}$ when in fact there is none. This phenomenon is discussed further in section V.

d. SOME PRINCIPLES OF RADAR

By far the most common radar, and the one with which many readers will be familiar, is the pulsed radar. A short burst of energy is transmitted, followed by a long period of silence. If during this listening period the receiver detects a reflected signal, its time delay is measured from which range can be computed.

The data is reported as range and azimuth, which is already a perfectly good two-dimensional polar coordinate system. Transforming the data into any other desired coordinate system is very easy.

However, when the radar beam is directed at a very low angle, many reflections are received from stationary, ground-based objects. Returns from low-flying aircraft are lost in a welter of unwanted targets. In fact, a pilot's standard radar avoidance technique has always been to fly at tree-top level.

In an effort to detect these low-flying aircraft, the continuous wave (CW) radar was developed. The CW radar operates on the following principle. There is neither pulse nor listening period. Instead, a continuous unmodulated carrier wave is radiated. If any reflected signals are detected, they are "beat" against the transmitted wave. Reflection from a moving aircraft will alter the frequency, producing the well-known Doppler effect, and yielding a measurable beat-frequency. A band-stop filter can be applied near beat-frequency zero, eliminating unwanted reflections from stationary targets. (Also eliminated are returns from laterally-flying aircraft, since their radial velocity will be near zero, but these are of less interest than those which are approaching!)

Data arrives in natural units of velocity and azimuth -- not suitable for transformation. Range first must be computed. Early CW radars attempted to recover the range by integration. But from the very beginning, integral calculus teaches us that an arbitrary constant may be added to any integral without affecting the validity of the solution. Evaluation of this constant of integration is the whole burden of the subject of definite integrals and often proves to be difficult, or even impossible for mid-flight acquisition, the problem we are presently faced with.

Clearly we must get some bit of range information from somewhere. Yet the Doppler radar's sole raison d'être is the inability of other radars to detect under the stated conditions. Some device is needed, which is internal to the radar itself, that allows range to be determined.

e. RANGE DETERMINATION WITH THE DOPPLER RADAR

We define a scan as that period during which the radar antenna rotates through 360 degrees, beginning and ending at some known reference point.

During alternate scans (odd-numbered ones, let us say), an unmodulated carrier wave of constant frequency is continuously radiated. The Doppler velocity is measured in the standard way. We may call these scans CW-scans.

During the remaining scans (even-numbered, of course), the carrier wave is modulated by increasing the frequency at a known linear rate. Thus the frequency can be said to have a constant ramp. Note that the reflected signal will be beat against a frequency farther advanced along the ramp, causing an apparent reduction in Doppler velocity (for approaching targets). How much farther along the ramp is a function of elapsed time, and hence of range. Stationary targets will appear to recede, and hence can be filtered out. We may call these scans FM-scans.

If it can be assumed that the radial velocity is constant, then any two consecutive scans will produce sufficient data that range (at the time of the FM scan) can be determined.

However, the ability of an aircraft to execute simple turns enables it to produce large accelerations, the direction of which with respect to the radius vector (from radar to aircraft) can be quite random. Therefore, any assumption of constant radial velocity is completely untenable.

Thus is strongly highlighted the basic problem which any algorithm must solve. It is the computation or prediction of radial velocity for the time of an FM scan. Only when a satisfactory solution has been found can range be computed.

Equation (3) is well-suited to this purpose.

The concept of frequency modulation, however, introduces a new problem which, for want of a better name, we shall call "FM drop-out." If the returned signal is designated alternately as CW and FM, and if R and $\mathring{R}$ indicate the actual range and radial velocity, then, with suitable scaling,

$$CW = \mathring{R} \quad \text{and} \quad FM = \mathring{R} - hR \tag{4}$$

The constant h depends upon the frequency ramp and has the dimension $\sec^{-1}$. A representative value is $h = 0.002 \sec^{-1}$.

When R is too great, or $\mathring{R}$ too small, the resulting FM signal will not pass the band-stop filter, and FM drop-out occurs. The situation arises at very great ranges, with wide crossing angles, and for slow-moving aircraft. All these cases tend to fall in an area of lesser interest.

Too great a value of h also aggravates FM drop-out but that, of course, is a design problem.

II. MATHEMATICAL PRINCIPLES AND DEVELOPMENTS

a. THE DEVELOPMENT OF q

The following assumptions are made:

An aircraft is in a turn of constant rate, $d\theta/dt$. The value is taken as positive when the aircraft is turning to its right.

Range is sufficiently great that small changes in azimuth or altitude can be ignored without serious error.

Central angle $\theta = 0$ when aircraft is approaching radar head-on.

Aircraft is moving at a constant speed v.

Under these conditions, if R denotes range and $\mathring{R} = -dR/dt$ radial velocity,

$$\mathring{R} = v \cos \theta$$

$$d\mathring{R}/dt = -v \sin \theta \frac{d\theta}{dt}$$

$$d^2\mathring{R}/dt^2 = -v\left\{\sin \theta \frac{d^2\theta}{dt^2} + \cos \theta \, (d\theta/dt)^2\right\}$$

But $\frac{d^2\theta}{dt^2} = 0$, since $\frac{d\theta}{dt}$ is constant.

$$\therefore \frac{d^2\mathring{R}}{dt^2} = -v \cos\theta \, (d\theta/dt)^2$$

Dividing both sides by $\mathring{R} = v \cos\theta$,

$$\frac{d^2\mathring{R}/dt^2}{\mathring{R}} = -(d\theta/dt)^2 \qquad (5)$$

This quantity is called the "acc'rate ratio," and the estimate of it is designated by the symbol q. ("Acc'rate" refers to the derivative $d^2\mathring{R}/dt^2$.)

It can be observed that, under the stated conditions, the "acc'rate ratio"

(1) is constant

(2) is proportional to the square of the aircraft's turn rate

(3) is independent of the direction in which the aircraft is turning

(4) cannot take on positive values

In any computer solution, derivatives are estimated by a process of differencing. Let u_0, u_1, u_2, u_3,...be values of a function equally spaced in time. If the unit of time be taken as the increment between two successive arguments, then $u_0 - 2u_1 + u_2$ becomes an estimate of the second derivative at the time of u_1. It does not matter in which direction time is increasing, the value of the estimate remains the same.

For simplicity and clarity, the following conventions are adopted:

The period of one radar antenna rotation (e.g., three seconds) is taken as the unit of time.

Subscripts are expressed in multiples of this time unit and denote "age." Thus R_0 is present value of range, R_6 was the range value 6 seconds ago.

The subscript "p" (for "predicted") is used in place of the subscript "-3."

$\mathring{R}$ is used for radial velocity and is taken in a positive sense for approaching aircraft. Thus $\mathring{R} = -dR/dt$.

Applying the foregoing, it is found that

$$\mathring{R}_6 - 2\mathring{R}_3 + \mathring{R}_0 = q_3\mathring{R}_3$$

which equation yields a point estimate of q_3. But since q is presumed constant, the subscript can be dropped. Having now an estimate of q, we can advance the subscripts and write

$$\dot{R}_3 - 2\dot{R}_0 + \dot{R}_p = q\dot{R}_0$$

Solving for $\dot{R}_p$,

$$\dot{R}_p = (2 + q)\,\dot{R}_0 - \dot{R}_3 \tag{6}$$

which is simply equation (3) in slightly altered form.

This is the basic prediction equation and is fundamental to the algorithm. It solves the problem previously posed, and, provided turn-reversal has not just occurred, it is used for range computation.

Now at an FM scan, the received data is equivalent, not to $\dot{R}$, but to the quantity $\dot{R}$ - hR, where h is a constant whose value depends upon the frequency modulation rate. Therefore, there exists (at each FM scan) a corollary requirement to predict range as well as radial velocity. Since range prediction is required only on alternate scans, integration by Simpson's rule is well suited to the purpose. Simply stated,

$$R_p = R_3 - \frac{\Delta t}{3}\left(\dot{R}_3 + 4\dot{R}_0 + \dot{R}_p\right)$$

Using as an example an antenna rotation period of three seconds, ($\Delta t = 3$), and substituting the derived expression for $\dot{R}_p$, there is achieved the remarkably simple expression

$$R_p = R_3 - (6 + q)\,\dot{R}_0 \tag{7}$$

Referring to the prediction equations it is apparent that, relative to the time for which the prediction is made, data is used which is up to two scans "old" (6 seconds old in the example used). During this time frame, should the aircraft markedly alter its maneuver -- by turn reversal, for example -- the data may be so greatly perturbed that the equations cannot be used. The condition is temporary, lasting only until the aircraft has persisted in its new maneuver for a period of two scans or more, but must be identified, since a momentary change in procedure is required.

b. ACQUIRING THE TARGET

Suppose a CW radar with an FM coefficient h = 0.002 is required to operate effectively between the ranges of 5 and 60 statute miles. This operating range is called the "information band," and the values 5 and 60 the band edges. All information outside this band is considered suspect. The difference, 55 statute miles, is of course, the band width. Converting to metric units, the band edges are 8,047m and 96,561m, respectively, with a band width of 88,514m. Multiplying by $h = 0.002$ yields the "velocity band," with band edges of 16.1 m/sec and 193.1 m/sec. If the difference between the FM data and $\dot{R}$ does not fall within this band, the point estimate of the range also will be out-of-band. Note that velocity band width is 177 m/sec.

It turns out that if a suitable value of q is assumed ($q = -0.05$, say), track can be initiated from three sets of data, provided they are not all of the same kind. (If there is no FM data, range cannot be determined. If there is no CW data, radial velocity cannot be determined, which in turn precludes range determination.)

(1) <u>Track Initiation When the Data Comes From Two CW Scans and One FM Scan</u>

The value of $\dot{R}$ is given directly by the CW data, but is unknown at the FM scan, and at all scans where data is missing. The data on hand will be designated CW_i and FM_j. From the basic properties of the radar, we have

$$FM_j = \dot{R}_j - hR_j \tag{4}$$

from which

$$R_j = \frac{1}{h}(\dot{R}_j - FM_j) \tag{8}$$

To estimate $\dot{R}_j$, a value of q is arbitrarily assigned (e.g., $q = -0.05$) and the following equations are written

$$\dot{R}_6 + \dot{R}_0 = (2 + q)\ \dot{R}_3$$

$$\dot{R}_9 + \dot{R}_3 = (2 + q)\ \dot{R}_6$$

$$\dot{R}_{12} + \dot{R}_6 = (2 + q)\ \dot{R}_9$$

etc.

until a system of n simultaneous equations in n unknowns is obtained, which set can be solved for $\dot{R}_j$. If the difference $\dot{R}_j - FM_j$ falls within the proper velocity band, track is initiated. If not, the oldest data is discarded and another attempt is made at the next scan with suitable data.

EXAMPLE 1. Available data is CW_9, CW_3, and FM_0.

$$\dot{R}_6 + \dot{R}_0 = (2 + q)\ CW_3$$

$$CW_9 + CW_3 = (2 + q)\ \dot{R}_6$$

These two equations in two unknowns ($\dot{R}_6$ and $\dot{R}_0$) can be solved for both unknowns. However, $\dot{R}_6$ is not needed and can be simply eliminated, leaving

$$CW_9 + CW_3 = (2 + q) \{(2 + q) CW_3 - \dot{R}_0\}$$

from which

$$\dot{R}_0 = (2 + q) CW_3 - \frac{CW_9 + CW_3}{2 + q} \qquad (9)$$

and of course, from equation (8),

$$R_0 = \frac{1}{h} (\dot{R}_0 - FM_0)$$

These parameters possess dimension, and hence proper attention must be paid to scaling in the computer.

EXAMPLE 2. Available data is CW_6, FM_3, and CW_0.

$$CW_6 + CW_0 = (2 + q) \dot{R}_3$$

can be solved for the only unknown.

$$\dot{R}_3 = \frac{CW_6 + CW_0}{2 + q} \qquad (10)$$

and

$$R_3 = \frac{1}{h} (\dot{R}_3 - FM_3)$$

Note that this range reduction is valid for the time of the FM data. Equation (7) now yields R_p directly.

$$R_p = R_3 - (6 + q) CW_0$$

(2) Track Initiation When the Data Comes From One CW Scan and Two FM Scans

Since the value of $\dot{R}$ is unknown at an additional FM scan, another equation must be developed to complete the simultaneous set. This is done by evaluating the definite range integral between the FM scans, utilizing Simpson's rule. Continuing to illustrate by example, suppose the available data to be FM_6, CW_3, and FM_0. By Simpson's rule

$$R_0 = R_6 - \frac{\Delta t}{3} (\dot{R}_6 + 4\dot{R}_3 + \dot{R}_0)$$

If Δt = 3 sec, this simplifies to

$$R_0 = R_6 - \left(\dot{R}_6 + 4\dot{R}_3 + \dot{R}_0\right)$$

Now $\dot{R}_3 = CW_3$ and applying equation (3) yields

$$\dot{R}_6 + \dot{R}_0 = (2 + q)\ CW_3 \tag{11}$$

hence

$$R_0 = R_6 - (6 + q)\ CW_3 \tag{12}$$

From equation (4),

$$\dot{R}_6 - hR_6 = FM_6$$

and

$$\dot{R}_0 - hR_0 = FM_0$$

Subtracting and transposing,

$$h(R_0 - R_6) = \dot{R}_0 - \dot{R}_6 - FM_0 + FM_6$$

From equation (12)

$$h(R_0 - R_6) = -h(6 + q)\ CW_3$$

Consequently,

$$\dot{R}_0 - \dot{R}_6 = FM_0 - FM_6 - h(6 + q)\ CW_3 \tag{13}$$

the sought additional equation in the proper unknowns. Together with equation (11) it forms the required simultaneous pair. Again, $\dot{R}_6$ is not needed and can be eliminated, this time by simple addition.

$$2\dot{R}_0 = FM_0 - FM_6 + \left\{2 + q - h(6 + q)\right\}\ CW_3 \tag{14}$$

and

$$R_0 = \frac{1}{h}\left(\dot{R}_0 - FM_0\right)$$

(3) Further Remarks on Track Initiation

If it could be known that an aircraft had altered its maneuver during the period when data for track initiation was being collected, it would become mandatory to delay until data from three scans reflecting the new maneuver became available. Probably prospects are not as hopeless as they may at first seem.

If the first attempted range computation falls out of band, it is assumed that some change has taken place. The "oldest" data is discarded and the attempt repeated when suitable additional information is obtained. In this connection it is wise to keep the computational band width as narrow as practicable. Just because the radar can "see" a target is not of itself sufficient reason to compute its parameters.

Consider a Doppler radar being used in an anti-aircraft battery configuration. Ranges so great that the missile would be spent before reaching the aircraft need not be included in the computational band. The same can be said for ranges so small that insufficient time to permit launching remains.

When an isolated case of FM drop-out occurs, it usually is accompanied by turn reversal within less than two scans (whether before or after is unknown). Thus in the sequence CW_9 - miss - CW_3 - FM_0, the data at CW_9 is sometimes valid, sometimes suspect. However, if $CW_9 >> CW_3$, turn reversal is almost certainly found between these two scans. To limit the application of this sequence to the case $CW_9 < CW_6$ thus is indicated.

The above procedures will identify much (but not all) of the suspect data.

The three examples given in (1) and (2) above doubtless represent the only practical initiation sequences.

c. FOUR-SCAN ACQUISITION

It is natural to inquire why a four-scan acquisition procedure is not used, since the additional information would make it possible to compute the value of q, rather than arbitrarily assume it.

There is no theoretical reason why a four-scan procedure could not be employed. There are, however, two sound, practical ones.

(1) Acquisition will certainly be delayed one scan, merely to obtain the additional data.

(2) The probability that four successive pieces of information will not include a maneuver change obviously is less than the probability associated with only three.

Four-scan acquisition increases both the risk and the time required. In a critical situation, the price may be prohibitively high.

There can be no objection to a four-scan supplemental procedure, provided the additional data is compatible (i.e., indicates no perturbation such as turn-reversal). In fact, there is one instance where four-scan may be superior to three-scan. We discuss that case forthwith.

CASE 1. CW_{12} - miss - CW_6 - FM_3 - CW_0.

The corresponding three-scan case is the least accurate of the procedures discussed so far, due to the fact that velocity is estimated by extrapolation. Range errors can be quite large, and tend to be corrected very slowly at first, apparently because range and "q" err in the opposite sense (logically) with respect to the quantity $\dot{R}_m - \dot{R}_p$.

The greatly increased accuracy of the four-scan acquisition procedure recommends it. The mathematics is relatively simple. We have

$$CW_{12} + CW_6 = (2 + q)\ \dot{R}_9$$

$$CW_6 + CW_0 = (2 + q)\ \dot{R}_3$$

$$\dot{R}_9 + \dot{R}_3 = (2 + q)\ CW_6$$

Multiplying the last equation by (2 + q) and substituting the result into the sum of the first two equations,

$$CW_{12} + 2CW_6 + CW_0 = (2 + q)^2\ CW_6$$

from which

$$\frac{CW_{12} + CW_0}{CW_6} + 2 = (2 + q)^2$$

Essentially, we have the three-scan case of paragraph b(1), Example 2, except that the value of "q" is computed rather than assumed. If, however, $(2 + q)^2 > 4$, turn-reversal apparently has occurred. In this event, we merely discard CW_{12} and revert to the appropriate three-scan procedure.

CASE 2. CW_9 - FM_6 - CW_3 - FM_0

In this case the target already has been acquired (else CW_9 would have been discarded). The four-scan procedure is tentatively substituted for the usual update, provided a valid value of "q" is returned. The basic equations are:

$$\dot{R}_6 = \frac{CW_9 + CW_3}{2 + q}$$

$$CW_3(2 + q) = \dot{R}_6 + \dot{R}_0$$

$$h(R_6 - R_0) = (\dot{R}_6 - FM_6) - (\dot{R}_0 - FM_0)$$

$$R_6 - R_0 = (6 + q)\ CW_3 = 4CW_3 + (2 + q)\ CW_3$$

Solving for $(2 + q)$, the larger root of the equation

$$(1 + h)\ CW_3(2 + q)^2 + (FM_6 - FM_0 + 4hCW_3)\ (2 + q) - 2(CW_9 + CW_3) = 0$$

is chosen. If $(2 + q) > 2$, the four-scan procedure is not used. Probably a Class 2 fit* has occurred and the range should be "coasted" (see para d(3)).

If $(2 + q) < 2$ (i.e., "q" is negative), we proceed

$$\dot{R}_0 = (2 + q)\ CW_3 - \frac{CW_9 + CW_3}{2 + q}$$

$$R_0 = \frac{1}{h}(\dot{R}_0 - FM_0)$$

CASE 3. $FM_9 - CW_6 - FM_3 - CW_0$

This case is strikingly similar to the preceding one. A parallel development yields the following quadratic equation in $(2 + q)$.

$$(1 - h)\ CW_6(2 + q)^2 + (FM_3 - FM_9 - 4hCW_6)\ (2 + q) - 2(CW_6 + CW_0) = 0$$

Notice the reversed sign of the terms involving h.

Continuing, provided $(2 + q) < 2$,

$$\dot{R}_3 = \frac{CW_6 + CW_0}{2 + q}$$

$$R_3 = \frac{1}{h}(\dot{R}_3 - FM_3)$$

$$R_p = R_3 - (6 + q)\ CW_0$$

Again, if $(2 + q) > 2$, it is likely that a Class 2 fit is present.

d. UPDATING THE RANGE TRACK FILE SUBSEQUENT TO INITIATION

For a maneuvering aircraft, the radial velocity can change markedly from scan to scan. In fact, there is some justification for using the raw incoming data, $\dot{R}_m$, without smoothing. However, in most applications, it will be suspected that the $\dot{R}_m$ data contains noise, and that a small amount of smoothing will be beneficial.

---

*See section III for definition of "Class 2 fit."

Letting the subscripts "age" (i.e., the most recent track file value becomes $\dot{R}_3$), the basic updating equation becomes

$$\dot{R}_0 = \dot{R}_p + k(\dot{R}_m - \dot{R}_p) \qquad (1)$$

Radial velocity will not tend to seek any particular value with passage of time.

We have seen that at a CW-scan

$$\dot{R}_m = CW$$

However, at an FM-scan, the best we can do is to employ equation (4) to <u>estimate</u> $\dot{R}_m$; viz.

$$\dot{R}_m = FM + hR_p \qquad (15)$$

Since, in the two cases, the value of $\dot{R}_m$ is differently arrived at, there is no reason to suppose that the coefficient k might not also be different. Therefore, we shall use k' to denote the velocity filter coefficient at a CW-scan.

The smoothed value ($\dot{R}_0$) having been computed, it is employed to make a new velocity prediction for use at the following scan. From equation (3) we obtain immediately

$$\dot{R}_p(\text{new}) = (2 + q)\,\dot{R}_0 - \dot{R}_3$$

(1) <u>Range Reduction at an FM-Scan</u>

Range information is available only at FM scans. Provided there is suitable* data, equation (8) will yield a point estimate of the range

$$R_m = \frac{1}{h}(\dot{R}_p - FM)$$

It is desirable to smooth this estimate through a proper filter.

$$R_0 = \psi R_m + (1 - \psi)\,R_p$$

or equivalently

$$R_0 = R_p + \psi(R_m - R_p) \qquad (16)$$

*How to determine whether or not the data is "suitable" is the subject of a later section.

For a smoothing filter, $0 < \psi < 1$, of course.

Usually, it will be found that the track file value of the range tends to be more stable than the point estimate, whereupon a fairly small value of ψ is indicated.

Equation (16) can be developed into a form more useful for computation as follows: From equation (15) we have

$$\dot{R}_m = FM + hR_p \tag{15}$$

Subtracting $\dot{R}_p$ from both sides, it is found

$$\dot{R}_m - \dot{R}_p = hR_p - (\dot{R}_p - FM)$$

$$\dot{R}_m - \dot{R}_p = hR_p - hR_m$$

from which

$$R_m - R_p = -\frac{1}{h}(\dot{R}_m - \dot{R}_p)$$

Substituting this expression into equation (16) yields

$$R_0 = R_p - \frac{\psi}{h}(\dot{R}_m - \dot{R}_p) \tag{17}$$

which computes range directly from the basic quantity $\dot{R}_m - \dot{R}_p$.

(2) Range Prediction at a CW-Scan

There is no range information at a CW-scan. Range is predicted by integrating between FM scans, utilizing Simpson's rule. This is effected by equation (7), repeated here for convenience.

$$R_p = R_3 - (6 + q)\dot{R}_0 \tag{7}$$

Since R_p and R_3 refer to FM scans, $\dot{R}_0$ necessarily is the smoothed velocity at the intervening CW-scan.

(3) "Coasting" the Range

At an FM-scan, when there is no suitable data, equation (17) cannot be used to update the range track file. However, there is still a requirement to take aircraft movement into account. This is done by a process variously known as "coasting" or "dead reckoning."

First, it is necessary to compute, estimate, or arbitrarily assign a value to $\dot{R}_0$. Then the range can be "coasted" by Simpson's rule. In expanded form, after "aging" subscripts,

$$R_0 = R_6 - \frac{\Delta t}{3} (\dot{R}_6 + 4\dot{R}_3 + \dot{R}_0)$$

But we already find in the track file (still employing expanded form)

$$R_p = R_6 - \frac{\Delta t}{3} (\dot{R}_6 + 4\dot{R}_3 + \dot{R}_p)$$

Simple subtraction produces

$$R_0 = R_p - \frac{\Delta t}{3} (\dot{R}_0 - \dot{R}_p) \tag{18}$$

The simplification when $\Delta t = 3$ seconds is obvious.

e. UPDATING THE VALUE OF q

Aircraft maneuvers are subject to change. Further, an assumed value of q almost certainly is not without error. It is therefore worthwhile to attempt to measure the acc'rate ratio for the purpose of updating the value of q carried in the track file.

From the original definition of q, it is seen that a point estimate can be obtained from

$$q_m = \frac{\dot{R}_6 - 2\dot{R}_3 + \dot{R}_m}{\dot{R}_3}$$

Multiplying by $\dot{R}_3$ and transposing

$$(2 + q_m)\, \dot{R}_3 - \dot{R}_6 = \dot{R}_m$$

But we already find in the track file (after "aging" the subscripts)

$$(2 + q)\, \dot{R}_3 - \dot{R}_6 = \dot{R}_p(\text{old})$$

Subtracting,

$$(q_m - q)\, \dot{R}_3 = \dot{R}_m - \dot{R}_p \tag{19}$$

Since q is presumed constant, there is no need for a prediction; a suitable filter for updating is

$$q(\text{new}) = q + \xi(q_m - q)$$

Substituting equation (19) into this expression gives

$$q(\text{new}) = q + \frac{\xi}{\dot{R}_3}(\dot{R}_m - \dot{R}_p) \tag{20}$$

in terms of the basic quantity $(\dot{R}_m - \dot{R}_p)$. The track file value of q is simply augmented by an amount

$$\frac{\xi}{\dot{R}_3}(\dot{R}_m - \dot{R}_p)$$

It is known that q cannot take on positive values. Should the update produce a result $q(\text{new}) > 0$, the track file value -- call it q_{TF} here -- should be reset to zero. This is easily accomplished by passing all updates through the filter

$$q_{TF} = \frac{1}{2}\{q(\text{new}) - |q(\text{new})|\} \tag{21}$$

The process of differencing in the computer is inherently noisy. The estimate of the acc'rate ratio, being based on a second order difference, is doubly noisy. A quite small value of ξ therefore is required to achieve sufficient smoothing.

f. OPTIMUM FILTER COEFFICIENTS

It has been seen that the updating equations for range, velocity, and turn-rate (q) all depend upon $\dot{R}_m$ which, after all, is the sole source of new information. In each case, the quantity $(\dot{R}_m - \dot{R}_p)$ is multiplied by an appropriate (k, ψ, ξ) weighting coefficient. Now $\dot{R}_m - \dot{R}_p$ can be thought of as a residual or as an error but, regardless of concept, unarguably contains "noise." If the sum of the weights exceeds unity, i.e., if

$$k + \psi + \xi > 1$$

the algorithm will act as a noise amplifier and the solution can be expected to diverge. This results in the following limits being imposed:

$$k + \psi + \xi < 1 \tag{22}$$

and

$$k' + \xi < 1$$

Assuming a constant value of q, the algorithm equations are:

(1) At a CW-scan

$$\dot{R}_m = CW$$

$$\dot{R}_0 = k' \dot{R}_m + (1 - k') \dot{R}_p(old)$$

$$R_p = R_3 - (6 + q) \dot{R}_0$$

$$\dot{R}_p(new) = (2 + q) \dot{R}_0 - \dot{R}_3$$

(2) At an FM-scan

$$\dot{R}_m = FM + hR_p$$

$$\dot{R}_0 = k\dot{R}_m + (1 - k) \dot{R}_p(old)$$

$$R_0 = R_p - \frac{\psi}{h} \{\dot{R}_m - \dot{R}_p(old)\}$$

$$\dot{R}_p(new) = (2 + q) \dot{R}_0 - \dot{R}_3$$

Supposing there is an error ε in the data at a CW-scan. Then

$$k'(\dot{R}_m + \varepsilon) + (1 - k') \dot{R}_p = \dot{R}_0 + k'\varepsilon$$

$$R_3 - (6 + q) (\dot{R}_0 + k'\varepsilon) = R_p - (6 + q) k'\varepsilon$$

and

$$(2 + q) (\dot{R}_0 + k'\varepsilon) - \dot{R}_3 = \dot{R}_p(new) + (2 + q) k'\varepsilon$$

At the following FM-scan we then find

$$FM + h\{R_p - (6 + q) k'\varepsilon\} = \dot{R}_m - h(6 + q) k'\varepsilon$$

$$k\{\dot{R}_m - h(6 + q) k'\varepsilon\} + (1 - k) \{\dot{R}_p + (2 + q) k'\varepsilon\}$$
$$= \dot{R}_0 - k\{2 + q + h(6 + q)\} k'\varepsilon + (2 + q) k'\varepsilon$$

$$(2 + q)\,\dot{R}_0 - k(2 + q)\,\{2 + q + h(6 + q)\}\,k'\varepsilon + (2 + q)^2\,k'\varepsilon - (\dot{R}_3 + k'\varepsilon)$$

$$= \dot{R}_p(\text{new}) + \{(2 + q)^2 - 1\}\,k'\varepsilon - k(2 + q)\,\{2 + q + h(6 + q)\}\,k'\varepsilon$$

The error in $\dot{R}_p$(new) <u>VANISHES</u> <u>NON-TRIVIALLY</u> when

$$k(2 + q)\,\{2 + q + h(6 + q)\} = (2 + q)^2 - 1$$

Thus we can select a value of k, the FM-scan velocity filter coefficient, which prevents an error in the CW data from being propagated forward into the next CW-scan. Notice that k, thus selected, is a function of q alone.

A form more suitable for computation is

$$k\{2 + q + h(6 + q)\} = 2 + q - \frac{1}{2 + q} \tag{23}$$

For the values of q commonly encountered, k usually falls between 0.69 and 0.75. Its maximum value occurs when q = 0, hence is designated k_0. From equation (23)

$$4k_0(1 + 3h) = 3$$

Equation (22) shows that both ψ and ξ must be small in comparison with k. Any computation of them which involves subtraction thus will tend to be noisy. There seems to be little case for other than a fixed value of either ψ or ξ. Having established this point, we can recover from equation (22) the restraint

$$\psi + \xi < 1 - k_0$$

Dropping the inequality sign, let us adopt as one of the two defining equations for ξ and ψ the expression

$$\psi + \xi = 1 - k_0$$

From the earlier expression

$$q(\text{new}) = q + \frac{\xi}{\dot{R}_3}\,(\dot{R}_m - \dot{R}_p) \tag{20}$$

it can be seen that for very high speed aircraft, the update of q will become too sluggish unless ξ is increased. A procedure such as the following is suggested:

Let v_{max} denote the maximum expected airspeed (or possibly that airspeed most likely to be encountered) stated in m.sec^{-1}. Then

$$\frac{\xi}{\psi} = \frac{v_{max}}{450}$$

This becomes the second defining equation for ξ and ψ. The two expressions are solved simultaneously.

The optimum values of ξ and ψ are not further pursued in this paper. Obviously, they remain in an area worthy of investigation.

In the event that the error ε occurs at an FM-scan, an analogous procedure is followed. However, since both range and radial velocity are computed at an FM-scan, an additional coefficient (ψ) appears and the error expression becomes more complicated (as does the logic). Proceeding with the development,

$$FM + \varepsilon + hR_p = \dot{R}_m + \varepsilon$$

$$k(\dot{R}_m + \varepsilon) + (1 - k)\ \dot{R}_p(\text{old}) = \dot{R}_0 + k\varepsilon$$

(k is now a known function of q.)

$$R_p - \frac{\psi}{h}\ \{\dot{R}_m + \varepsilon - \dot{R}_p(\text{old})\} = R_0 - \frac{\psi\varepsilon}{h}$$

$$(2 + q)\ (\dot{R}_0 + k\varepsilon) - \dot{R}_3 = \dot{R}_p(\text{new}) + (2 + q)\ k\varepsilon$$

At the following CW-scan we find

$$k'\dot{R}_m + (1 - k')\ \{\dot{R}_p + (2 + q)\ k\varepsilon\} = \dot{R}_0 + (1 - k')\ (2 + q)\ k\varepsilon$$

$$R_3 - \frac{\psi\varepsilon}{h} - (6 + q)\ \{\dot{R}_0 + (1 - k')\ (2 + q)\ k\varepsilon\}$$

$$= R_p - \frac{\psi\varepsilon}{h} - (6 + q)\ (1 - k')\ (2 + q)\ k\varepsilon$$

$$(2 + q)\ \dot{R}_0 + (2 + q)^2\ (1 - k')\ k\varepsilon - (\dot{R}_3 + k\varepsilon)$$

$$= \dot{R}_p(\text{new}) + \{(2 + q)^2\ (1 - k') - 1\}\ k\varepsilon$$

Since both predictions, range and radial velocity, will be employed at the next (FM) scan, the equations there must be investigated. Continuing,

$$FM + h\{R_p - (6 + q)\ (1 - k')\ (2 + q)\ k\varepsilon\} - \psi\varepsilon$$

$$= \dot{R}_m - \psi\varepsilon - h(6 + q)\ (1 - k')\ (2 + q)\ k\varepsilon$$

The quantity $(\dot{R}_m - \dot{R}_p)$ is used to update all parameters in the track file. Since at an FM-scan both $\dot{R}_m$ and $\dot{R}_p$ are computed values, it could happen that $(\dot{R}_m - \dot{R}_p)$ is without error, even though there are errors in the track file. In this event, the track file errors remain uncorrected. The filter coefficients which produce this effect (no correction) can be taken as limiting values, beyond which the algorithm diverges. Obviously, $(\dot{R}_m - \dot{R}_p)$ contains no error when

$$-\psi\varepsilon - h(6 + q)\ (1 - k')\ (2 + q)\ k\varepsilon \quad = \quad \{(2 + q)^2\ (1 - k') - 1\}\ k\varepsilon$$

Dividing by $k\varepsilon$ and transposing,

$$1 - \frac{\psi}{k} \quad = \quad (2 + q)\ (1 - k')\ \{2 + q + h(6 + q)\} \tag{24}$$

Multiplying by k and substituting from equation (23),

$$k - \psi \quad = \quad (1 - k')\ \{(2 + q)^2 - 1\}$$

This expression gives an upper limit for k'. For $\xi = 0.1$ and $h = 0.002$ its value falls between 0.76 and 0.81. The restriction $k' < 1 - \xi$ also applies, of course.

Another (lower) limiting value for k' can be found by supposing the track file range error to be completely corrected at each FM-scan. Values of k' below this limit will produce a range track file which tends to be both oscillatory and divergent. The development proceeds

$$\begin{aligned} R_p - \frac{\psi\varepsilon}{h} - (6 + q)\ (1 - k')\ (2 + q)\ k\varepsilon - \frac{\psi}{h}\ \{\dot{R}_m - \psi\varepsilon \\ - h(6 + q)\ (1 - k')\ (2 + q)\ k\varepsilon\} + \frac{\psi}{h}\ (\dot{R}_p + [(2 + q)^2\ (1 - k') - 1]\ k\varepsilon\} \\ = R_0 - (1 - \psi)\ \frac{\psi\varepsilon}{h} - (1 - \psi)\ (6 + q)\ (1 - k')\ (2 + q)\ k\varepsilon \\ + \frac{\psi}{h}\ \{(2 + q)^2\ (1 - k') - 1\}\ k\varepsilon \end{aligned}$$

The error in R_0 vanishes when

$$(1 - \psi)\ \{\psi + (6 + q)\ (1 - k')\ (2 + q)\ hk\} \quad = \quad \{(1 - k')\ (2 + q)^2 - 1\}\ \psi k$$

Substituting from equation (23) and rearranging

$$\psi(1 - \psi + k) \quad = \quad (1 - k')\ \{(1 - \psi) - (1 - \psi - k)\ (2 + q)^2\} \tag{25}$$

from which the lower limit of k' can be computed. For $\xi = 0.1$ and $h = 0.002$, the value falls between 0.30 and 0.45.

A value of k' about midway between the above limits should achieve near-optimum smooth range tracking. Therefore let us require that the error in R_0 be roughly half that in R_6 (i.e., at the previous FM-scan). Immediately, we can write

$$-\frac{1}{2}\psi = -(1-\psi)\{\psi + (6+q)(1-k')(2+q)hk\}$$

$$+ \{(1-k')(2+q)^2 - 1\}\psi k$$

A development exactly similar to the preceding one yields

$$\psi\left(\frac{1}{2} - \psi + k\right) = (1-k')\{(1-\psi) - (1-\psi-k)(2+q)^2\} \qquad (26)$$

The value of k' supplied by equation (26) should be near optimum. For $\xi = 0.1$ and $h = 0.002$, it varies between 0.53 and 0.62.

III. SELECTION, CORRELATION, AND CLASSIFICATION OF RANGE AND VELOCITY DATA

Up to this point, the main purpose has been to develop mathematical formulae for acquiring and tracking a single aircraft in a turn of constant rate (which rate may be zero). It may happen that the expected sequence of events is altered so drastically that the standard formulae cannot be employed to treat the incoming data. Whether these perturbations are apparent or real, it is necessary to detect their occurrence, so that alternate steps can be taken.

In an attempt to perform this detection, a trichotomous device called "double-gating"* is employed, by which the data is evaluated and placed into one of three categories, as follows:

(1) Class 1 fit. The incoming data is consistent with the predicted values, and thus can be used to update the track file parameters.

(2) Class 2 fit. The incoming data is inconsistent with the predicted values, but is not inconsistent with the most recent track file values. It is assumed that some change has taken place, and therefore only certain of the track file parameters can be updated. (For example, q and R are not updated.) At a Class 2 fit, k = k' = 0.5 probably suffices.

*In this paper, "gate" is synonymous with "window."

(3) Class 0 fit - a miss. The incoming data is either erroneous or from an extraneous source, and hence cannot be used. This class includes the case of missing data (e.g., FM drop-out).

a. TURN REVERSAL

Perhaps the data are perturbed most violently when a high-speed maneuvering aircraft, crossing the radar line-of-sight at a wide angle, suddenly reverses its direction of turn. In fact, a suitable criterion for setting the "outer" gate is the desired maximum detectable perturbation of this type.

When a flight path is convex (as viewed from the radar), the aircraft is turning toward a receding aspect, and thus is an item of diminishing interest. Conversely, when the flight path is concave toward the radar, the aircraft is turning toward a more direct approach. When turn reversal (at a sufficiently high rate of turn) has produced the latter state of affairs, the data at the following scan invariably will exhibit certain symptoms. They are:

(1) Acc'rate will appear to be very large and POSITIVE.

(2) The absolute value of the difference $|\dot{R}_m - \dot{R}_p|$ will be large -- perhaps even beyond the outer gate (i.e., a miss).

(3) When the predicted velocity is replaced by the old track file velocity, the absolute value of the difference always is reduced. In other words

$$|\dot{R}_m - \dot{R}_3| < |\dot{R}_m - \dot{R}_p|$$

Moreover, this reduced difference rarely falls outside the window. Thus, the substitution (of $\dot{R}_3$ for $\dot{R}_p$) avoids declaring a miss with good data present.

(4) Because acc'rate is based upon a <u>second</u> difference, the "large and POSITIVE" symptom will persist for an <u>additional</u> scan.

(5) So, too, will persist the SIGN of the residual $(\dot{R}_m - \dot{R}_p)$. Thus, for the two scans following turn reversal

$$\frac{(\dot{R}_m - \dot{R}_p)\text{new}}{(\dot{R}_m - \dot{R}_p)\text{old}} > 0$$

b. TURBINE RETURNS AND OTHER LARGE TRANSIENTS

Sometimes perturbations in the data are not accompanied by corresponding changes in the flight pattern of the aircraft. Such a case might occur when the aircraft is approaching head-on, so that the radar "sees" a rotating turbine instead of the aircraft skin, with resulting frequency shift. The minimum expected frequency shift of this type is a criterion for setting the "inner" gate.

These perturbations rarely are persistent, enabling us to identify them merely by inspecting the data at the following scan. Applying the yardstick of the preceding subsection, we find

(1) The point estimate of the acc'rate ratio will be very large, but is equally likely to be positive or negative.

(2) The difference $(\dot{R}_m - \dot{R}_p)$ usually is large -- but can be of either sign.

(3) Substitution of the old track file velocity for the predicted velocity does not necessarily reduce the difference.

(4) At the following scan, the point estimate of the acc'rate ratio will tend to be large, but of opposite sign.

(5) The residual $(\dot{R}_m - \dot{R}_p)$ will change sign at the following scan. In fact, since $\dot{R}_p$(new) presumably is based upon erroneous data, whereas $\dot{R}_p$(old) is not, we expect

$$\frac{(\dot{R}_m - \dot{R}_p)\text{new}}{(\dot{R}_m - \dot{R}_p)\text{old}} < -1$$

c. DATA SELECTION

The correlation process begins as follows: Does the measured radial velocity differ from <u>both</u> the predicted value and the latest track file value by more than the outer gate? If so, a miss is declared. An exactly similar test is performed upon the azimuth measurement. (NOTE: Azimuth gate should vary inversely with range.)

If no miss is declared, the following tests are performed:

(1) Does the measured radial velocity differ from the predicted value by less than the inner gate?

(2) Is the point estimate of q less than or equal to zero? (To accommodate noise in the data, small positive values should be considered as zero.)

If the answer to either of these questions is "no," a Class 2 fit is declared. But if both questions are answered "yes," the result is a Class 1 fit. At the second of consecutive Class 2 fits, the ratio

$$\rho = \frac{(\dot{R}_m - \dot{R}_p)\text{new}}{(\dot{R}_m - \dot{R}_p)\text{old}}$$

is examined. This also produces three categories:

(1) $\rho < -1$. This implies a turbine return or other large transient at the previous scan. Since only $\dot{R}_m$ at that scan is suspect, substitution of the appropriate $\dot{R}_p$ for $\dot{R}_3$ is indicated. Certain derived values may have to be recomputed.

(2) $\rho > 0$. This implies a turn reversal before the previous scan. As a result, several actions are required:

(a) Do not update the acc'rate ratio. In effect, $\dot{R}_6$ is missing.

(b) Do not update the range at the FM-scan -- dead reckon it instead employing Simpson's rule. Effectively, $\dot{R}_p$ is missing, leaving nothing upon which to base a range reduction.

(c) Do not compute the radial velocity. The required information will be found to lie in disparate sets. In this case, however, there is sufficient justification for substituting the appropriate values of $\dot{R}_m$ for $\dot{R}_3$ and $\dot{R}_0$, respectively.

(3) $-1 < \rho < 0$. No clear implication. Compute the radial velocity and coast the range. Nothing better can be offered.

d. GATE SIZE

The purpose of any "gate" or "window" is to admit all truthful evidence while screening out all that is false. Unfortunately, there is no perfect gate size. A gate which admits all the desirable data may also admit much that is not wanted. And the converse is true. A gate which screens out all erroneous data may reduce to a trickle the flow of useful information.

All that ever can be said is that a gate accomplishes each part of its two-fold mission to some statistical probability. Clearly, such probabilities depend upon not only equipment design, but upon the use to which the equipment is put, as well as the performance characteristics of the expected targets. Lacking specifics, only general suggestions can be made.

(1) The Outer Gate. Obviously, the principal purpose of the outer gate is to screen out unwanted returns from extraneous sources. It must, however, be large enough to permit detection of certain specified maneuvers. For example, if $\Delta t = 3$ sec, and it is desired to detect turn-reversal on a target undergoing 6 g's acceleration (58.8 m/sec/sec), an outer gate of at least 176.4 m/sec will be required. See also paragraph (3) below.

(2) The Azimuth Gate is used in conjunction with the outer gate, and for the same primary purpose. It must be (in radians) at least $v\Delta t/R$. After a

miss, both the outer gate and azimuth gate should be increased, though perhaps not doubled. A multiplier of about 1.6 or 1.7 is suggested.

(3) The Inner Gate is NOT used primarily to eliminate unwanted data, but rather to identify perturbations in the desired data, so that some special action can be taken. To cite another example, suppose the 2σ noise level to be 25 m/sec. Further suppose that a turbine return causes a perturbation of 110 m/sec. The inner gate should be set to some value between 25 and 85 m/sec. If ν denotes noise level and τ the turbine or transient, the somewhat arbitrary formula

$$\text{inner gate} \quad = \quad \nu + \frac{(\nu + \tau)\, v_{max}}{2(450 + v_{max})}$$

yields a satisfactory value. The outer gate should be at least double the inner gate.

(4) The Q-Gate. In paragraph IIIc(2) above, it was established that a Class 2 fit should be declared whenever the point estimate of q exceeds a certain positive value. This effectively creates another window. Lest it superficially appear that the Q-gate duplicates the function of the inner gate and therefore is redundant, let us point out two essential differences. First, the Q-gate is one-sided, leaving negative values undisturbed. Second, the Q-gate is most sensitive to aircraft of low radial velocity. From equation (19), it is seen that

$$q_m \quad = \quad q + \frac{\dot{R}_m - \dot{R}_p}{\dot{R}_3}$$

The presence of $\dot{R}_3$ in the denominator is responsible for the essential difference between the Q-gate and the inner gate. The Q-gate is quite effective against aircraft which, crossing at a wide angle, begin a turn toward the radar site. A word of extreme caution is necessary however.

At successive scans, the roles of CW and FM information are interchanged. If the track file value of q is greatly in error, it will cause the value of q_m, the point estimate, to oscillate. If the Q-gate is too small, q_m might fall alternately in and out of the window, with catastrophic result. No appreciably large probability of this occurrence can be tolerated, since the condition can arise if a maneuvering aircraft abandons its maneuver for a direct attack upon the radar site itself.

A Q-gate of 0.15 to 0.2 is suggested.

IV. THE AZIMUTH PROBLEM

a. MEASUREMENT AND PREDICTION OF AZIMUTH

In general, CW radars can determine radial velocity by a direct phase comparison in the receiver, but can determine azimuth only by locating the region of maximum signal strength, or by detecting when that signal strength exceeds a certain threshold.

Evidently, then, the measurement of azimuth is less precise (i.e., is "noisier") than that of radial velocity. To predict azimuth from track file history alone is to be subject to occasional extrapolation errors too large to be tolerated. In an attempt to control this, two devices are used. It has been seen that

$$\dot{R} = v \cos \theta$$

and

$$d\dot{R}/dt = -v \sin \theta \frac{d\theta}{dt}$$

Letting β denote the azimuth and noting that under the assumed conditions a sufficiently close approximation to $R \frac{d\beta}{dt}$ is given by

$$R \frac{d\beta}{dt} = v \sin \theta$$

then

$$d\dot{R}/dt = - R \frac{d\beta}{dt} \frac{d\theta}{dt}$$

from which

$$d\beta/dt = - \frac{1}{R} \cdot \frac{d\dot{R}/dt}{d\theta/dt}$$

which allows azimuth prediction to be based upon the more accurately known parameters R, $\dot{R}$, and $\sqrt{-q}$. The sign of $d\theta/dt$ is as yet unknown, but by restricting use of the formula to the case of a Class 1 fit, it insures that the sign will not have changed during the last two scans, and hence, can be determined by comparison of recent changes in $\dot{R}$ and β. If $\beta_0 = \beta_3$, simply set $\beta_p = \beta_0 = \beta_3$.

If a Class 1 fit does not exist, or if the sign of $\sqrt{-q}$ cannot be determined (e.g., when $\dot{R}_0 - \dot{R}_3 = 0$), or if q is, numerically, very small ($q > -0.01$, say), azimuth must be extrapolated. The formula

$$\beta_p = \beta_0 + \mu(\beta_0 - \beta_3)$$

is recommended. If the coefficient μ is assigned a value less than unity, the cumulative azimuth correction (in the case of several iterations) will be limited, thereby preventing "runaway."

b. AZIMUTH CORRELATION

At track initiation, a fixed azimuth "window" must be used, since range is not known. For subsonic velocities, a window of 0.012 to 0.020 radians per second appears suitable. A fixed window tends to inhibit track initiation on laterally flying targets at short ranges, and therefore its value should be chosen according to the desires in that area.

After initiation, a window of the form a + b/R is used, where the first term accommodates the noise, the second term the range. If a is expressed in radians per second, then b should be expressed in meters per second, R in meters. The variable window should be equal to the fixed window at some selected midrange.

The incoming data, β_m, is compared both to β_p and to β_3 (subscript "aged"). If both differences exceed the window, a "miss" is declared. If neither azimuth data nor radial velocity data produces a miss, azimuth is updated by

$$\beta_0 = \eta\beta_m + (1 - \eta)\,\beta_x$$

where β_x is either β_p or β_3, whichever lies closer to β_m. The coefficient η probably should be chosen between 0.5 and 0.8 in order to given some weight to the track file data. (Some weight already had been given β_m by the selection of β_x.)

In the event of a miss, simply set $\beta_0 = \beta_p$(old).

V. ERRORS INDUCED BY THE POLAR COORDINATE SYSTEM

So far, the development has assumed that azimuth changes produce negligibly small errors. Strictly speaking, that is not always true. Let us examine Figure 1, following.

Let an aircraft fly a curved flight path (center of curvature at O) from A to B at speed v during the period of one scan. Let θ be a measure of aircraft heading, referenced to south. For a radar positioned at C, let α be a measure of aircraft azimuth, referenced to north. With these conventions, $\theta + \alpha$ is the aspect angle and $v \cos(\theta + \alpha) = \dot{R}$ gives the radial velocity (positive for approaching aircraft).

As the aircraft moves from A to B, the flight path is concave toward the radar, and the aspect angle changes more slowly than does the aircraft heading. Thus the apparent turn rate is less than the actual, and there is an apparent outward acceleration. Should the flight path be convex toward the radar, the aspect angle changes more rapidly than does the aircraft heading, so that the apparent turn rate is greater than the actual. But the apparent acceleration is still outward. This apparent outward acceleration is a function of range, turn rate, aircraft speed, and aspect angle. It typically

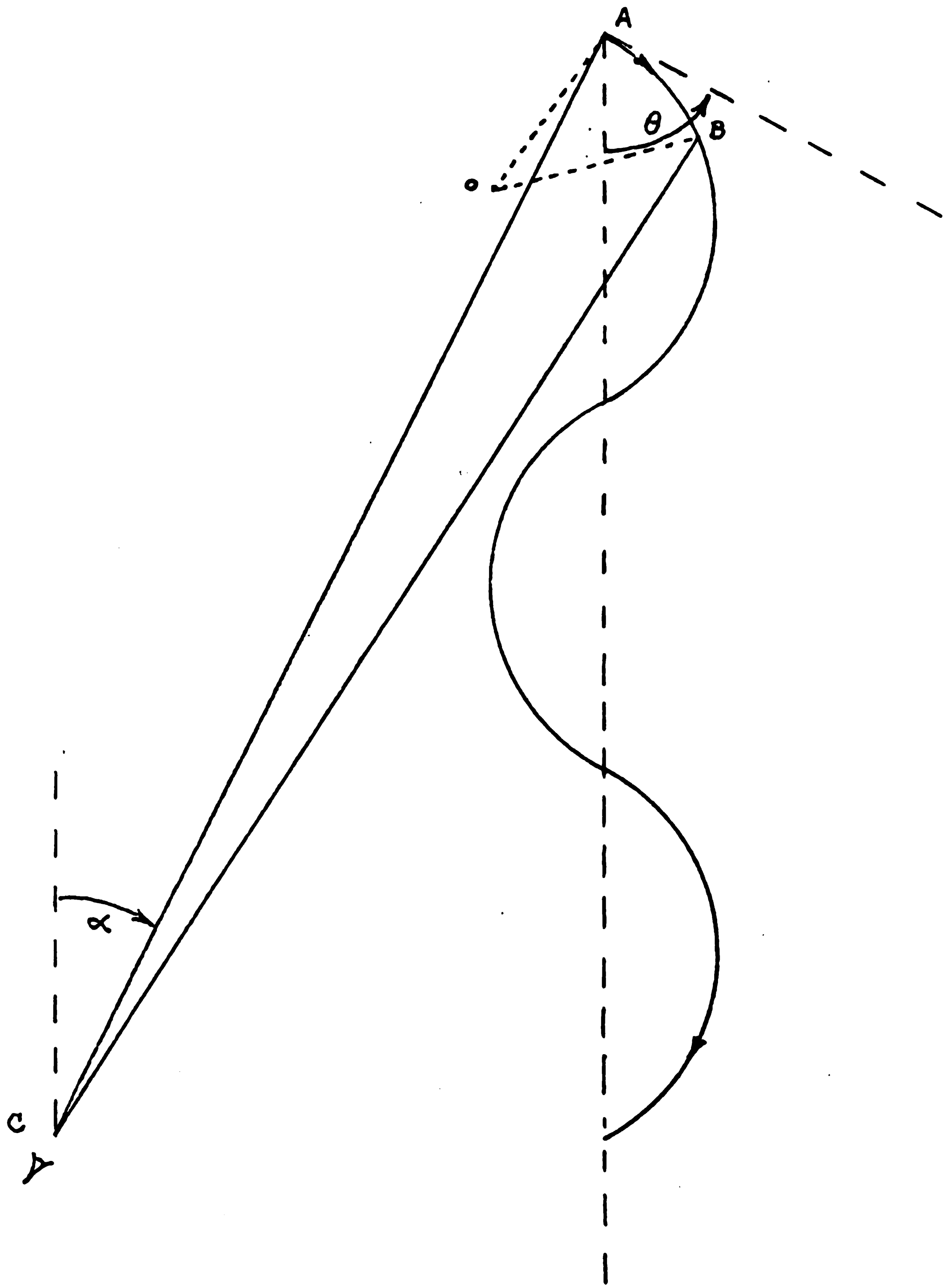


FIGURE 1. Flight Path of a Maneuvering Aircraft

appears in the equations of motion in a polar coordinate system. In so far as it arises solely because of the coordinate system being used, it is reminiscent of Coriolis' acceleration.

In designing a Doppler radar of the type with which we are here concerned, it is arbitrarily decided to observe only approaching aircraft whose radial velocity exceeds a specified threshold. Were the aircraft to fly a complete circle, the effect can be likened to observing the sun during a 24-hour period. The aircraft "rises" out of the band-stop filter, reaches maximum radial velocity at "noon," after which the radial velocity decreases at a faster and faster rate until it disappears behind the band-stop filter.

When the aircraft "rises," the square of the turn rate appears to be too small. As the aircraft "sets," the square of the turn rate appears to be too large. For the observable sector, then, the square of the apparent turn rate increases monotonically.

Now the observed value of q, the acc'rate ratio, being computed from actual observations, would seem to be the proper parameter for predicting future observations (of radial velocity) -- and so it is. However, from the original definition of q

$$\frac{\dot{R}_n - 2\dot{R}_{n-1} + \dot{R}_{n-2}}{\dot{R}_{n-1}} = q_{n-1}$$

It is seen that the estimate of q, being based upon the second difference, is valid only at the previous scan -- and the value of q is known to be changing monotonically.

The trouble with the estimate of q is not that it is in error, but that it is too "old." As a result, the prediction multiplier (2 + q) always will be too large, as will $\dot{R}_p$ and track file range. Under extreme conditions, this "over" range bias can exceed tolerances. A more up-to-date estimate of q is needed. Let us call the required estimate q_p.

The basic updating equation is

$$q(\text{new}) = q + \frac{\xi}{\dot{R}_3}(\dot{R}_m - \dot{R}_p)$$

and should be used in normal fashion. Now it is known that q will decrease monotonically as long as the present aircraft maneuver persists. Hence, the corrective equation

$$q_p = q(\text{new}) + \frac{\xi}{\dot{R}_0}(\dot{R}_m - \dot{R}_p)$$

suggests itself, with the proviso that

$$\frac{\xi}{\mathring{R}_0}(\mathring{R}_m - \mathring{R}_p) < 0$$

This latter is easily accomplished by employing the filter

$$q_p = q(\text{new}) + \frac{\xi}{2\mathring{R}_0}\{(\mathring{R}_m - \mathring{R}_p) - |\mathring{R}_m - \mathring{R}_p|\}$$

The value of q_p is not carried in the track file. It is used to predict $\mathring{R}_p(\text{new})$ and, at CW-scans, R_p, then discarded.

When q_p is thus used for prediction, an unexpected dividend accrues. Repeating for convenience

$$\mathring{R}_p(\text{new}) = (2 + q)\,\mathring{R}_0 - \mathring{R}_3 \tag{6}$$

and

$$R_p = -(6 + q)\,\mathring{R}_0 + R_3 \tag{7}$$

it is seen that when q_p is used for q, $\mathring{R}_0$ can be cancelled out of the correction term, leaving

$$\mathring{R}_p(\text{new}) = (2 + q_{new})\,\mathring{R}_0 - \mathring{R}_3 + \xi(\mathring{R}_m - \mathring{R}_p)$$

and

$$R_p = R_3 - (6 + q_{new})\,\mathring{R}_0 - \xi(\mathring{R}_m - \mathring{R}_p)$$

still subject to the proviso that $(\mathring{R}_m - \mathring{R}_p) < 0$.

EXAMINATION OF SIZE EFFECTS IN THE

FAILURE PREDICTION OF CERAMIC MATERIAL

D. M. Neal and E. M. Lenoe

Army Materials and Mechanics Research Center
Watertown, Massachusetts 02172

ABSTRACT

Probability of structural failure of a large hydro-burst ceramic Slip Cast Fused Silica (S.C.F.S.) ring was predicted from a data base of four point bend specimens. Both statistical surface and volumetric flaw distribution theory were considered. The major contributions of this paper are determination of an acceptable size effect relationship in predicting failure of S.C.F.S. and the recognition that uncontaminated test data is necessary in the prediction process. The removal of outliers and multi-modality from the test data will define the uncontaminated data. There was excellent agreement (less than 11% error) between predicted and actual experimental results when surface flaw theory is applied to uncontaminated data. The mini-max principle in conjunction with the maximum likelihood (M.L.) method is used to determine the outliers. The Quantile Box Plot method is applied when examining for multi-modality and outliers.

By considering only non-contaminated data, the essential need for equality of coefficient of variations between specimen test data and the structural component materials is satisfied in the predictive process. Removing outliers and multi-modality (data values resulting from errors in manufacturing or testing) insures the same basic properties of the ring and flexure specimen materials. Therefore, application of acceptable size effects relations in the prediction process can be more successful if types and distributions of flaws are similar in both components and only size governs differences in their mean strengths.

INTRODUCTION

Ceramic materials often exhibit a size effect relationship with respect to failure; that is a small test specimen will fail at higher stress levels than a larger one subjected to similar loading. The rationale: since the strength of a component is governed by chance that a severe stress concentration (c) will be subjected to a stress (s) such that the local stress (sc) exceeds the material strength and the chances of a more severe c value exists in the larger component thereby resulting in a lower failure stress. Since ceramic materials tested at ambient temperatures do not deform plastically and relieve these stresses, provides a reasonable verification for the above argument. It should also be noted that relatively low strength values could result if no severe flaws coincide with the maximum stress and failure occurs at a severe flaw subjected to a lower stress at a position where (sc) is maximum.

The primary objective of this paper is to determine a desirable method for predicting the mean failure stress of a relatively large ring from the test results of a small size flexure specimen (see Figure 1), where both materials are made from identical S.C.F.S material. The ring and flexure specimens were obtained from missile radomes. The rings were subjected to hydroburst tests and the flexure bars were tested as shown in Figure 1. The tests were conducted in order to establish quality control of the radome. If an acceptable size effect relationship can be established between the two components then the need for continued relatively expensive test of rings could be eliminated. More importantly a possible failure prediction methodology for S.C.F.S. as related to size effects will be made available. The authors have been fortunate in that considerable amount of data has been made available by the Raytheon Co. of Bedford, Mass. The data was separated into eight billets containing both ring and flexure test results providing combined total of 1300 specimens.

The authors examined the merits of using volume versus surface flaw theory in the failure prediction process. The conventional Weibull[1] method was applied in the failure prediction process with some success if the maximum stressed regions are considered. Another procedure[2] using a surface flaw distribution theory was also applied. With the sample sizes available, the opportunities existed for the authors to systematically establish an acceptable size dependent failure prediction method for S.C.F.S., with a reasonable degree of certainty.

In order to eliminate both parasitic stress[2] (producing low tensile stresses) and high tensile stresses (resulting from failures remote from the maximum stressed region), both robust estimating procedure[3] and the Quantile Box Plot[4] was applied to the data. These procedures will rocognize the outliers and establish uni-modal distributions in formal manner. Application of these procedures will result in data that represents the essence of the material strength capabilities in addition to providing a more acceptable representation of the data. This is particularly important when attempting to represent flexure test results adequately since this size specimen is often vulnerable to manufacturing or testing errors.

Robustness Method

The outliers are determined in a formal manner by applying a robust method involving application of the maximum likelihood (M.L.) estimation where the residuals are weighted in a systematic manner. The computed weights describe the relative importance of the data points. For example, a zero weight should indicate exclusion of a point. It should be emphasized that removing outliers without a valid reason is poor practice. Outliers should be examined for errors in testing or possible material defects. The removal of outliers (bad data) will essentially define robust data. The robust procedures applied in this paper involves using both the M-estimating technique of Huber[5] and Andrews[6]. Initially the Huber technique is applied in order to determine a robust location parameter (weighted mean). The Andrew's function is then applied using location parameter estimated from the Huber result. It should be noted that this robust method requires a uni-modal distribution of the data, therefore initial application of the Quantile Box Plot should be made.

The Huber m-estimation technique which involves defining the likelihood function

$$L(\theta) = \prod_{i=1}^{N} f(X_i-\theta), \quad -\infty < \theta < \infty \tag{1}$$

where f is a contaminated normal distribution,

X_i = data,

θ = location parameter and

N = sample size

By maximizing log $L(\theta)$ such that

$$\Sigma\psi(X_i-\theta) = 0, \tag{2}$$

where $\psi = f'/f$

then the solution of (2) is M.L. estimate of θ designated as $\hat{\theta}$. In order to represent ψ in scale invariant form, equation (2) can be rewritten as

$$\sum\psi\left(\frac{(X_i-\theta)}{d}\right)=0 \tag{3}$$

with d equal to the estimate of scale. The scale is often defined as

$$d = \text{median}|X_i - \text{median}(X_i)| \,/.6745$$

$$\text{or simple M.A.D.}/.6745 \tag{4}$$

This estimate is considerably more robust than using the complete samples which could result in poor representation of the actual scale.

By solving

$$\sum_{i=1}^{N} W_i(X_i-\theta) = 0 \tag{5}$$

where

$$W_i=\psi\left(\frac{X_i-\theta}{d}\right)\Big/\left(\frac{X_i-\theta}{d}\right)$$

$$\Psi=\begin{cases} r & |r|\le c_1 \\ c_1 \text{ sign } (r) & |r|>c_1 \end{cases},$$

c_1 is defined as the tuning constant and

$$r=\left(\frac{X_i-\theta}{d}\right)$$

An iterative process is then used in the solution of (5) such that when the differences in W_i become negligible provides the necessary criteria for an acceptable solution for the θ and W_i values. For c_1=1.345 the Huber's ψ function provides a 95% efficiency[3].

With estimate of $\hat{\theta}$ determined from the solution of (2) the iteration is continued where the Ψ function is now defined as

$$\Psi(r)=\begin{cases} c_1\sin(r/c_1), & |r|\le\pi c_1 \\ 0, & |r|>\pi c_1 \end{cases} \tag{6}$$

This new function is called the Andrew's wave equation. In order to obtain the desired robust data for the ψ function, the tuning was adjusted to c_1=1.34 and the scale defined as in equation (4).

It should be noted that Andrew's function was selected for its ability to describe outliers as data with essentially zero weights.

Quantile Box Plot

A general description of the Quantile Box Plot is shown in Figure 2. Where the quantile function is defined as

$$Q\,(u) = F^{-1}\,(u),\ 0 \leq u \leq 1 \tag{7}$$

that is, if the random variable x with distribution function given by F(x), then the root of F(x) = u, $0 \leq u \leq 1$ is the p^{th} quantile of F(x). From the ordered statistic $x_1 \leq x_2 \leq \ldots , x_n$, Q is defined as piece wise linear function with interval (0,1) divided into 2n subintervals. Therefore representing Q as

$$Q\left(\frac{2j-1}{2n}\right) = x_j\ ,\quad j = 1, 2, \ldots , n. \tag{8}$$

In order to interpolate

$$u \in \left(\frac{2j-1}{2n}\ ,\ \frac{2j+1}{2n}\right)$$

$$Q(u) = n\left(u - \frac{2j-1}{2n}\right) x_{j+1} + n\left(\frac{2j+1}{2n} - u\right) x_j, \tag{9}$$

where n equals the sample size .

The box boundaries are defined as

Q (.25) to Q (.75)

Q (.125) to Q (.875)

Q (.0625) to Q (.9375)

The Quantile function Q(u) is useful for detecting the presence of outliers, modes and the existence of two populations. Flat slots in Q(u) indicate modes. Sharp rises in Q(u) for u near 0 or 1 suggest outliers; sharp rises in Q(u) within the boxes indicate the existence of two (or more) populations. In obvious bimodality shown in Figure 3 is represented by the Quantile Box Plot displayed in Figure 4.

Weibull Distribution Function

The M.L. method is applied in order to obtain the two parameters of the Weibull function'

$$f(x) = \frac{m}{\mu} \left(\frac{x}{\mu}\right)^{m-1} \exp\left[-\left(\frac{x}{\mu}\right)^{m}\right] \qquad (10)$$

The method requires defining the likelihood function[7]

$$L = N! \prod_{i=1}^{N} \left\{ \frac{m}{\mu} \left(\frac{X_i}{\mu}\right)^{m-1} \exp\left[-\left(\frac{X_i}{\mu}\right)\right]^{m-1} \right\} \qquad (11)$$

where X_i = data,

m, μ = shape and normalizing parameter and

N = sample size,

By solving the following log likelihood equations

$$\frac{\partial \ell nL}{\partial \mu} = 0 \quad \text{and}$$

$$\frac{\partial \ell nL}{\partial m} = 0 \qquad (12)$$

determines the $\hat{m}$ and $\hat{\mu}$ values.

Equation (12) must be solved in an iterative manner where the initial estimates are obtained from the method of moments. The unbiased m and μ and their corresponding confidence intervals are obtained from Tables by [8].

Weibull Size and Stress Distribution Relationship

The basic equation for predicting mean failure stress $\bar{\sigma}_p$ of the ring from mean failure stress $\bar{\sigma}_1$, of flexure tests is

$$\frac{\bar{\sigma}_p}{\bar{\sigma}_1} = \left[\frac{K_1\ \Delta_1}{K_2\ \Delta_2}\right]^{1/m_1} \tag{13}$$

where a Δ_i dependence is assumed to regulate change in the failure stress. m_1 is the Weibull shape parameter obtained from the flexure data.

The Δ_2 and Δ_1 are the volumes or areas of the ring and flexural bar respectively, depending on whether surface or volume flaw theory is desired. Determination of K_2 and K_1 depends on the results from integrating the risk of rupture[9] relation

$$R = \int_{\Delta_i} \left(\frac{x}{\mu^*}\right)^m d\,\Delta_i \tag{14}$$

such that

$$R = K\Delta_i \left(\frac{x}{\mu^*}\right)^m$$

Note, $\mu^* = \mu(K\Delta_i)^{1/m}$, where μ is defined in equation (12) and that $K_2=1$ for ring since it is assumed to be a negligible stress gradient through ring thickness.

Statistical Flaw Distribution Theory (Alternative Method)

An alternative surface dependent relation was applied to the flexure data in order to predict mean failure stress of the ring. This method in a prior application[2] successfully predicted tensile failure of 96 percent alumina cylindrical rods from similar material and geometry using four point flexure test. The maximum stress region was considered in this application. The size dependent relation for mean failure stress $\bar{\sigma}_p$ prediction of ring is

$$\frac{\bar{\sigma}_p}{\sigma_1} = \left(\frac{A_1}{A_2}\right)^{1/K_1+1}, \tag{15}$$

where K_1 is determined from solution of

$$\frac{S_1}{\sigma_1} = \frac{\left[(K_1+1)^{2/K_1+1}\ \Gamma\left(\frac{K_1+3}{K_1+1}\right) - \left((K_1+1)^{1/K_1+1}\ \Gamma\left(\frac{K_1+2}{K_1+1}\right)\right)^2\right]^{1/2}}{(K_1+1)^{1/K_1+1}\ \Gamma\left(\frac{K_1+2}{K_1+1}\right)} \tag{16}$$

The ratio $\frac{S_1}{\sigma_1}$ is the coefficient of variation (C.V.) from flexure tests and σ_1 and S_1 are mean failure strength and standard deviation of flexure test results respectively.

A_1 and A_2 are areas of components subjected to maximum stress. This is a reasonable assumption for the requirements in the ring prediction process since elementary fracture mechanics evaluation[10] indicates that a flaw on outer ring surface would have to be 23 percent greater than on the inner ring surface for an equal chance of failure. Note with sizable increases in K, the difference in surface areas of ring and flexure bar will result in minimum effect for predicting failure.

RESULTS AND DISCUSSIONS

In Figure 5, P_f (probability of failure) vs. strength results for billet number 2001 are shown for both the ring (I) and flexure test (II) results. Application of the Quantile Box Plot shown in Figure 6 indicates bimodal distribution for the flexure test results. The mode representing larger strength values were removed. This is justified since ceramic material strength data is usually represented by an extreme value distribution (e.g. Weibull) which is skewed to the left. An additional reason for excluding the second mode resulted from the materials/test laboratory responsible for the data indicated that difficulties existed with regard to the test fixture used in obtaining the flexure results. Note the ring results (I) were generally represented by a relatively smooth curve with either no outliers or very few as shown in Figure 5. The type of test and the component geometry probably contributed to this situation. Note, the Weibull shape parameters (Modulus M) of 5.85 and 12.2 respectively for the flexure and ring data respectively, differed considerably when comparing the two test results. These results should correlate reasonably well otherwise application of the size effect relations described previously is not valid. That is, if materials are similar their dispersion values resulting from strength tests should be similar.

In Figure 7, the results from removal of second failure mode are shown as II^R. This uni-modal representation of flexure data provides a relatively smooth curve consistent in slope and appearance with the probability ranked ring stress failures. The agreement between the dispersion constants (m) is quite acceptable, therefore, allowing application of statistical flaw distribution theory, where it is assumed that material and uniformity of failure locations are similar for the two specimen geometries.

In Figure 8, the P_f results for billet No. 4001 is shown. In this case, outliers contaminated the data as shown in Quantile Box Plot (see Figure 9). Although the Huber and Andrews robust methods for determining outliers is usually acceptable, the authors considered the Quantile Box Plot more desirable. The arbitrariness in selecting tuning constants and scale parameters when applying the Huber Technique are the primary reasons for relying on this simpler and more efficient Quantile Box Plot procedure.

Since the Quantile function must initially be considered in determining multi-modality, then it is a simple matter to complete the investigation using the plot results. Additional results (Billets 1001 and 3001) are shown in Figure 10 and 11.

The results of predicting failure of ring from flexure test results are shown in Tables 1, 2 and 3. In Table 1, a comparison of m values indicates general agreement between flexure and ring results when the robust scheme is applied. In the conventional Weibull prediction process the predicted mean failure stress of ring using robust data ($\bar{\sigma}_p^R$) agreed somewhat better with the actual ring failure results ($\bar{\sigma}_2^T$) than the corresponding nonrobust case ($\bar{\sigma}_p$). If the maximum stressed region is considered then the absolute percent difference E^R_{MSA} is reduced considerably when applying the robust method as compared to nonrobust considerations (E_{MSA}). E^R_{TA} (total areas with robust data) indicates poor correlation for all billets where minimum of 20 percent is noted for 2001 series. From these results, it appears obvious that total area consideration is a poor choice in size effect failure prediction even with robustness. The results shown in Table 2 indicate Weibull volume dependency relationships are not applicable in the predictive process since a minimum percent difference of 62 exists when considering any of the four billets.

The results from applying the alternative method[2] with robust data are shown in Table 3. In general the method provided slightly better predictive qualities (E^R) than the Weibull approach when considering predicted and actual mean ring failure stress. The C.V. values did not correlate well for the 3001 series, predictive results were fair (11% error) indicating that actual equality of C.V. or m are necessary in applying these predictive methods if robust data is used.

CONCLUSIONS

This paper has described a predictive process that successfully determined mean failure of a large S.C.F.S. component (Hydroburst Ring) from flexure tests on small specimens. The statistically size dependency relations of the Alternative Method provided acceptable results with the Weibull method less desirable. The primary requirements are: the removal of outliers (bad data) or bi-modal distributions by formal statistical procedures and considering only maximum stressed surfaces in the predictive process for this particular material and manufacturing process. Results have indicated that S.C.F.S. is not a volume or total area sensitive material in predicting its failure. It has also been demonstrated that application of the robustness methods determines more realistic material characteristics in regards to the failure stresses. Bad data (outliers) or multi-modality tend to distort the statistical model in providing satisfactory failure prediction methods. The Quantile Box Plot provided a more desirable method for obtaining outliers and multi-modality in the data. The Huber and Andrews procedures were less desirable because the arbitrariness in selecting the tuning constants and scale measure. Symmetry assumptions prevented recognition of a separate set of outliers from either of the tail regions of the distributed data in addition to not recognizing multi-modality in the data.

ACKNOWLEDGEMENTS

The authors acknowledges AMMRC personnel, Sheila Evans and Lucy Ohannesian for typing and preparing the manuscript and William Kerr for developing the necessary computer codes in determination of robust modelling and size effect relationships.

The authors also recognize Dr. Bruce Center of the Raytheon Co., Bedford, Mass. for providing the necessary test data and support for this project.

REFERENCES

1. Weil, N. A. and Daniel, I. M., "Analysis of Fracture Probabilities in Nonuniformly Stressed Brittle Materials", Presented at the Sixty-Fifth Annual Meeting, The American Ceramic Society, Pittsburgh, PA, April 30, 1963 (Symposium on Mechanical Properties of Ceramics, No. 1-15 (63).
2. Platts, D. R., Kirchner, H. P., "Comparing Tensile and Flexural Strengths of a Brittle Material", Journal of Materials, JMLSA. Vol. 6, No. 1, March 1971, pp. 48-59.
3. Hogg, R. V., "Statistical Robustness: One View of Its Use In Applications Today", Technical Report No. 60, December 1977, Department of Statistics, Univ. of Iowa, Iowa City, Iowa.
4. Parzen, Emanuel "A Density-Quantile Function Perspective of Robust Estimation", Robustness in Statistics, Academic Press, New York, 1979.
5. Huber, P. J., "Robust Estimation of a Location Parameter", Annals of Mathematical Statistics, V. 35, 1964, pp. 73-101.
6. Andrews, D. F., Bickel, P. J., Hampel, F. K., Huber, P. J., Rogers, W. H. and Tukey, J. W., (1972), "Robust Estimates of Location, Survey and Advances", Princeton, NJ, Princeton Univ. Press.
7. Harter, L. H. and Moore, A. H., "Maximum Likelihood Estimation of the Parameter of the Gamma and Weibull Populations From Complete and From Censored Samples", Journal of the American Statistical Association, V. 61, 1966 pp. 842-851.
8. Thoman, D. R., Bain L. J. and Antle, C. E., "Inference on the Parameters of the Weibull Distributions", Technometrics, Vol. 11, No. 3, August 1969.
9. Dukes, W. H., "Handbook of Brittle Material Design Technology", AD719-712, Published by Advisory Group for Aerospace Research and Development, Paris, France (1971).
10. Parker, A. P., Private Conversation, Royal Military College of Science, Shrivenham, UK.

Figure 1 - Ring and Flexure Test and Geometry

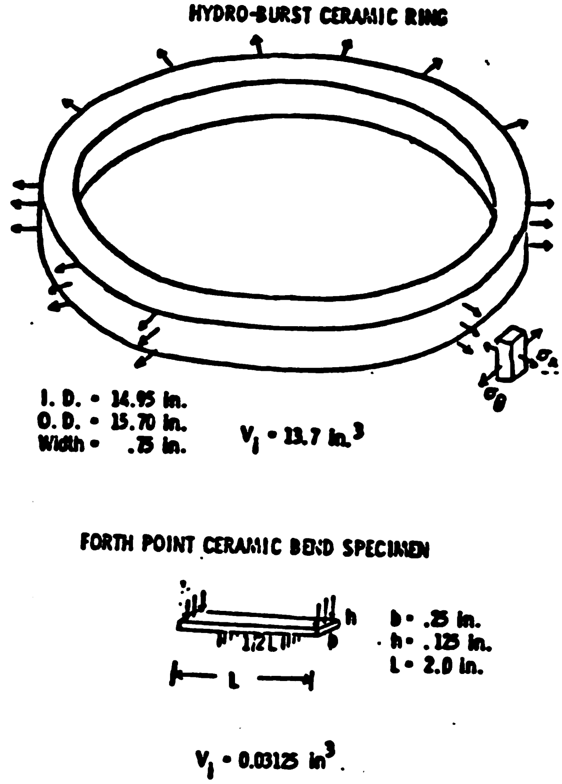


Figure 2. QUANTILE BOX PLOT

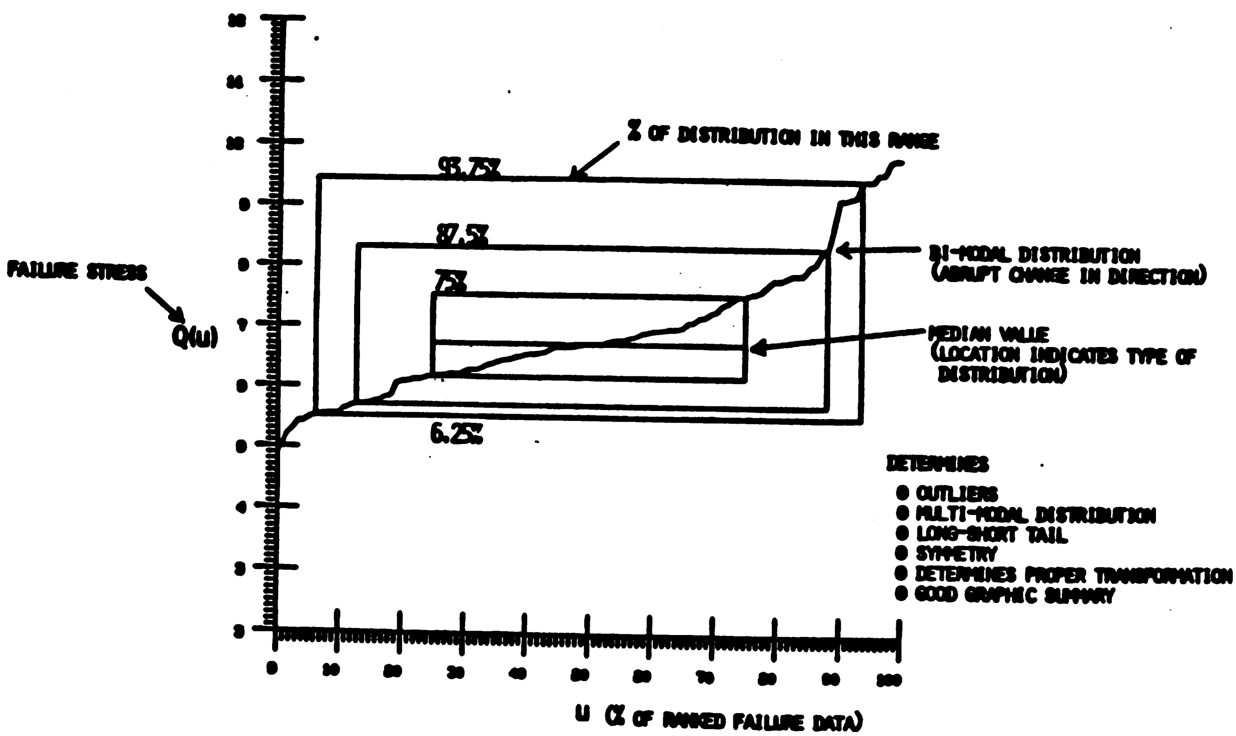


Figure 3. Probability Rank Data

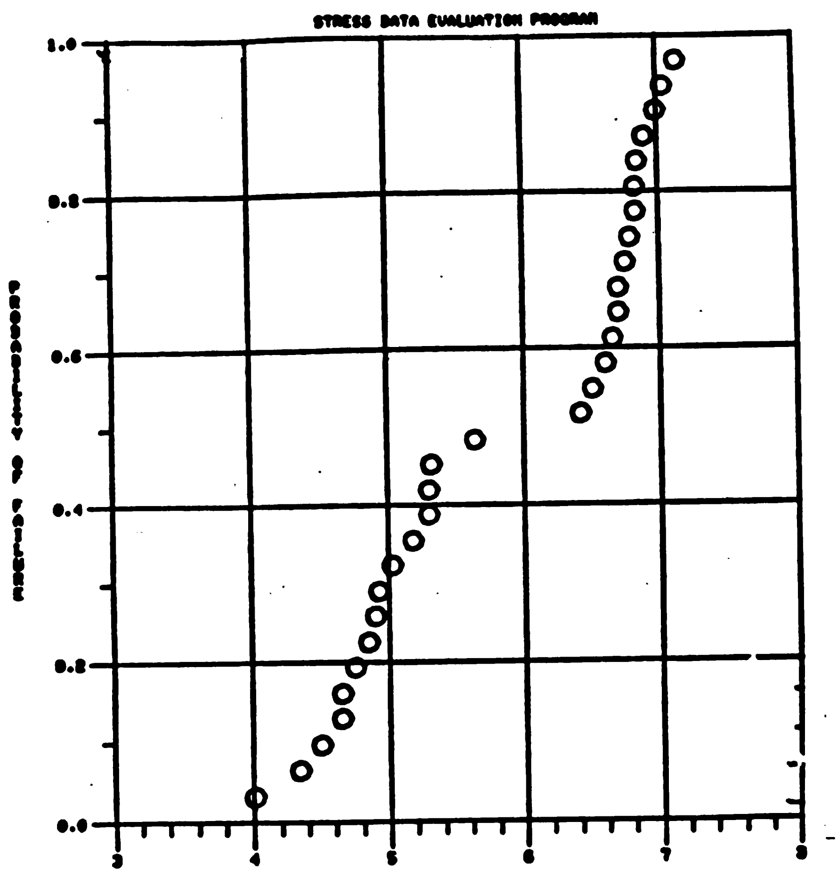


Figure 4. Quantile Box Plot

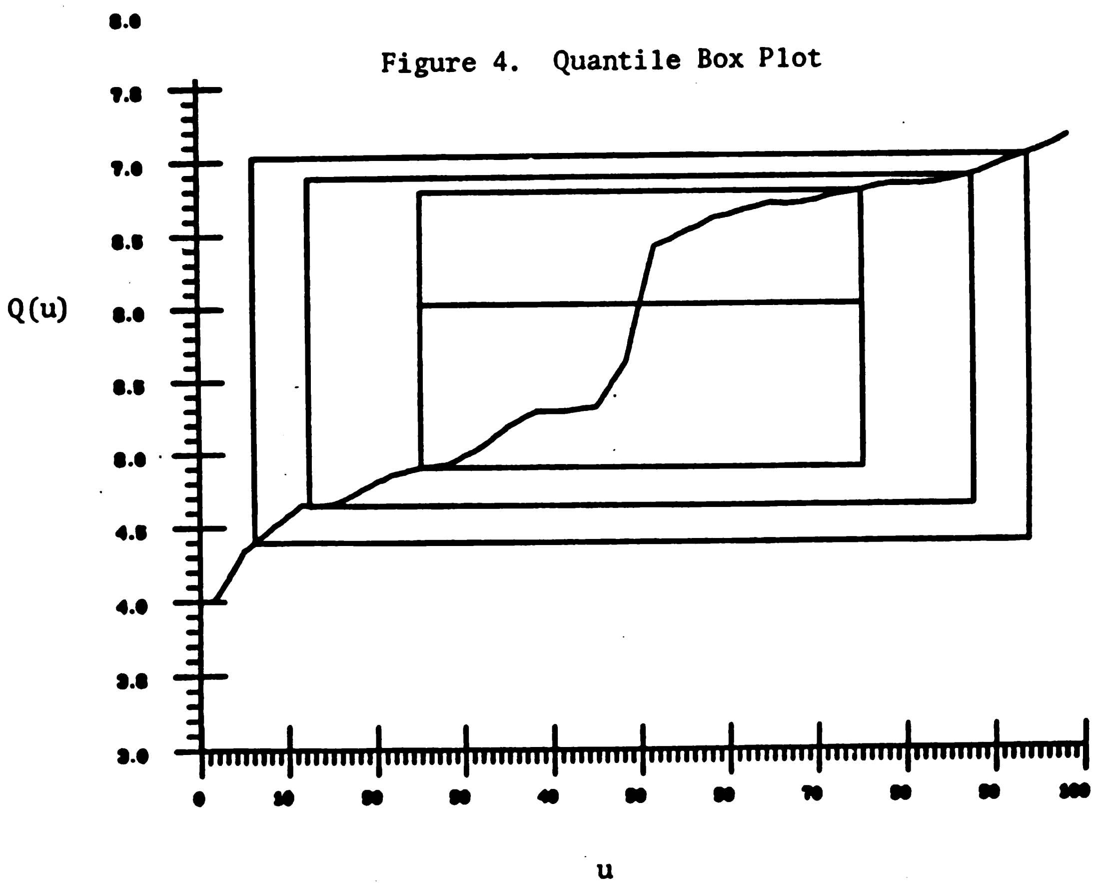
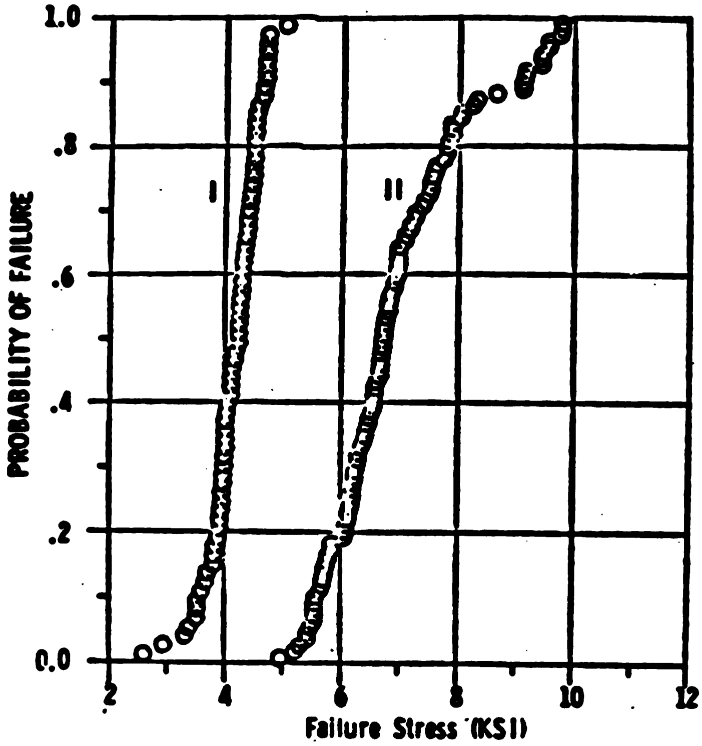


Figure 5 - Probability of Failure Vs. Stress Ring and Flexure Data



Billet Number 2001

I (Ring Test Results)

Mean Stress = 4.145 KSI
Standard Dev.= .419 KSI
Modulus M = 12.2
C.V.= .101

II (Flexure Test Results)
Non-Robust

Mean Stress = 6.952
Standard Dev.= 1.162
Modulus M = 5.858
C.V.= .167

Figure 6. Quantile Box Plot(Billet No. 2001)

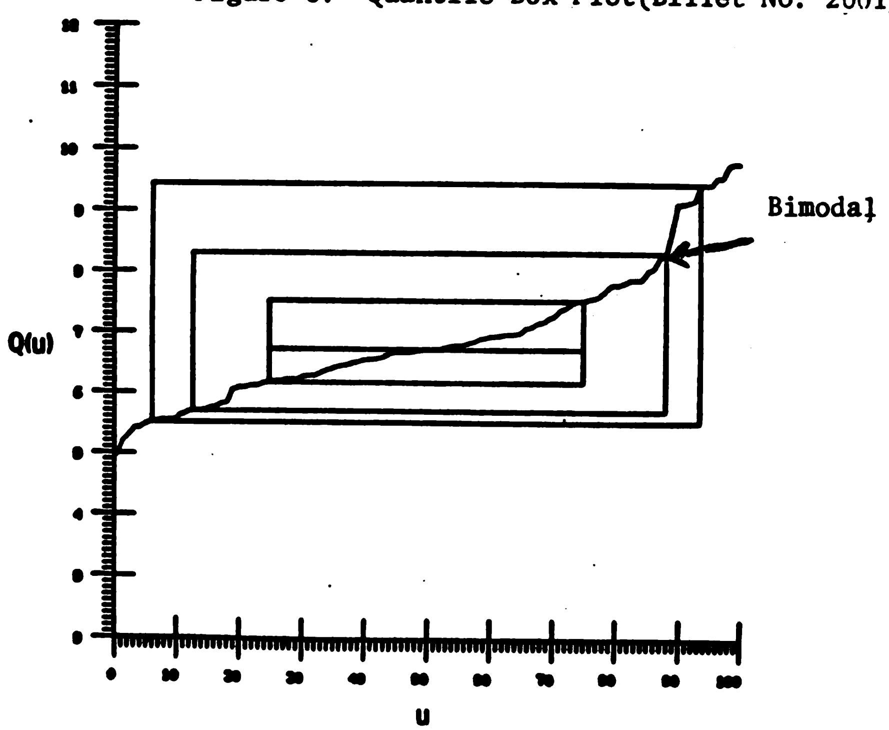
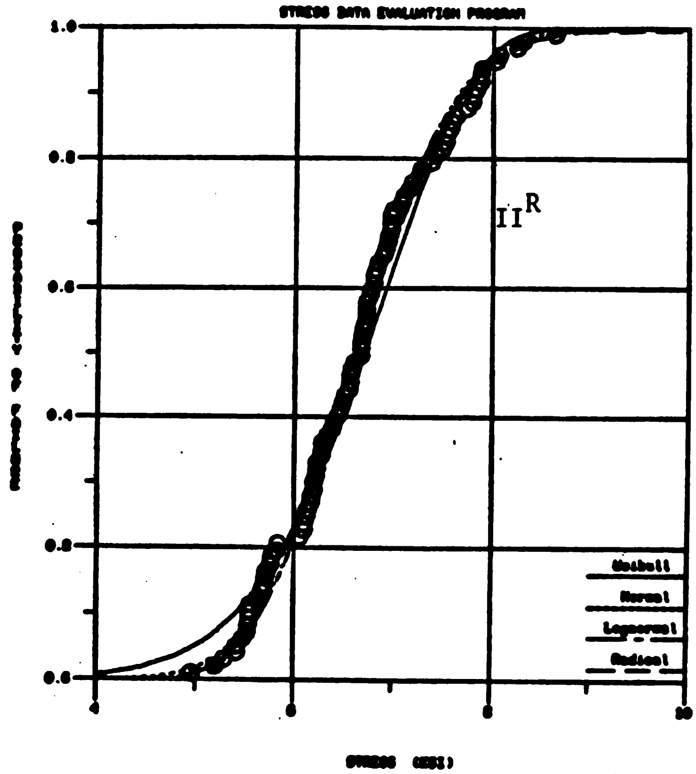


Figure 7. Probability of Failure vs. Robust Flexure Data (Billet No. 2001)



II^R Flexure Test Results

mean stress = 6.641

standard dev. = .793

modulus m = 8.77

c.v. = .119

Figure 7a Quantile Box Plot (Billet No. 2001)

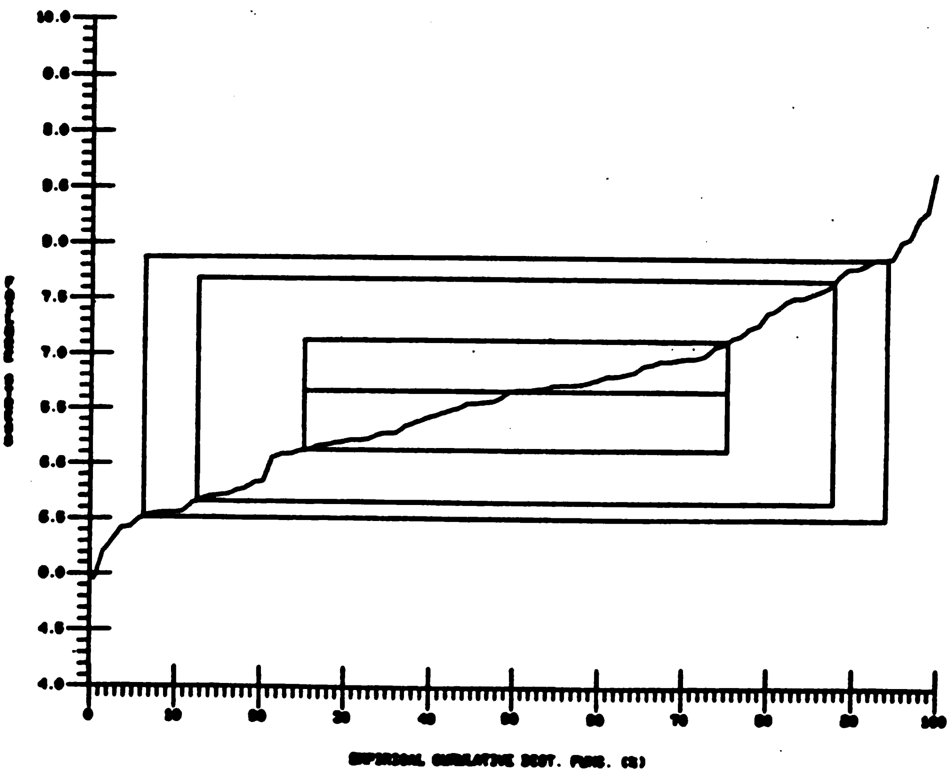


Figure 8. P_f vs. Failure Stress (Billet No. 4001)

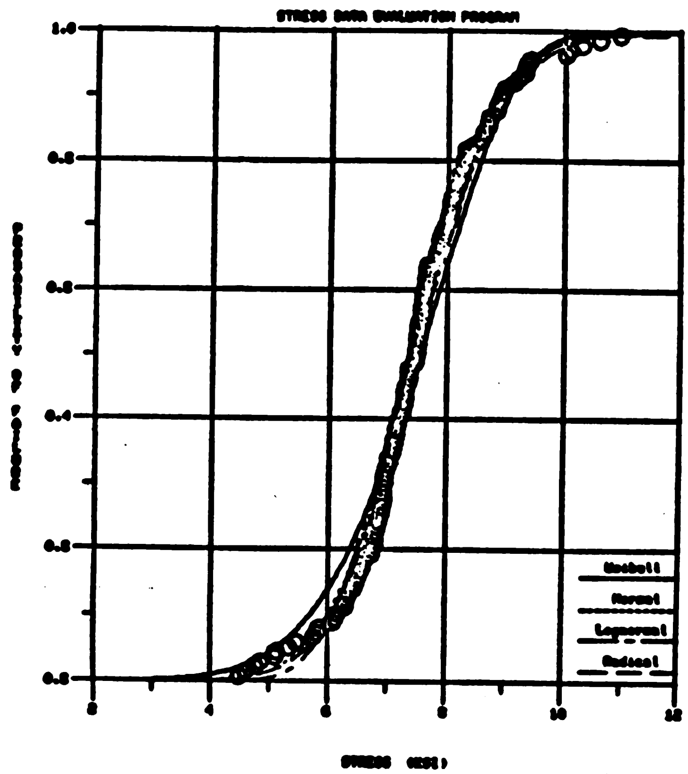
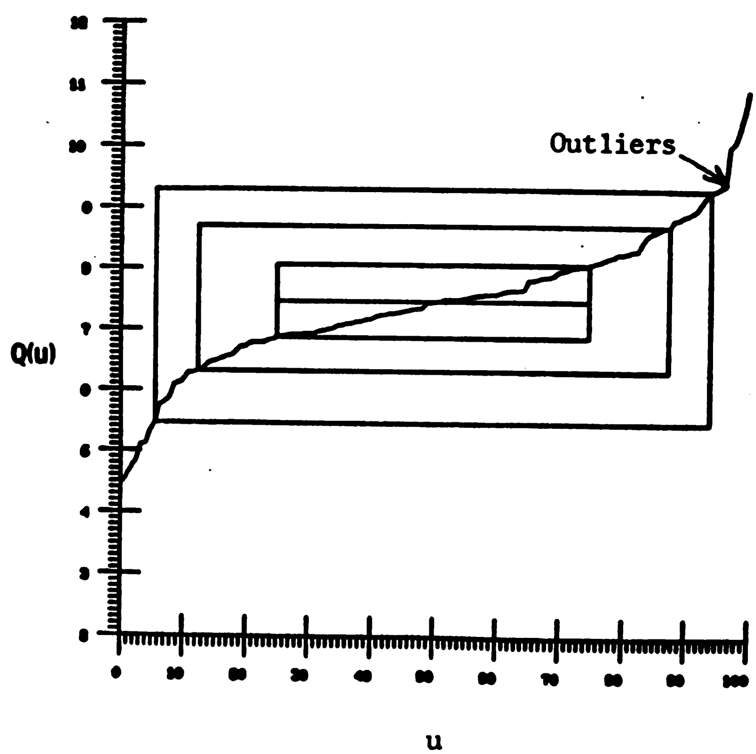


Figure 9. Quantile Box Plot (Billet No. 4001)



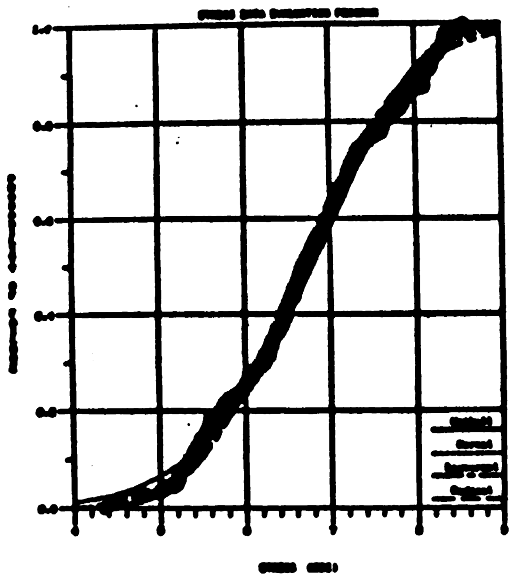


Figure 10.

P_f vs. Robust Failure Data
(Billet No. 1001)

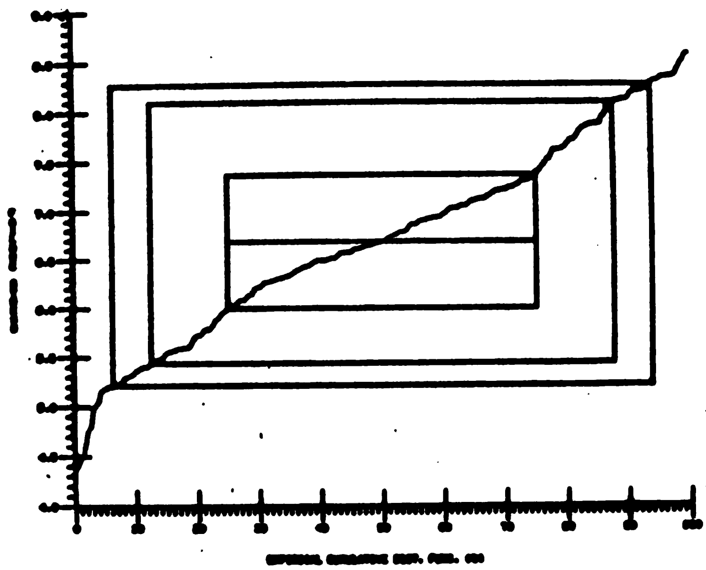


Figure 10a.
Quantile Box Plot
(Billet No. 1001)
Robust Data

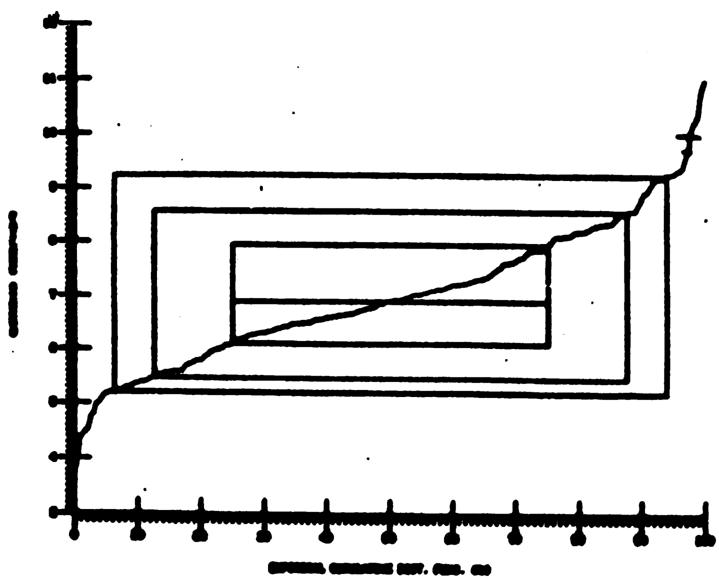


Figure 10b.
Quantile Box Plot
(Billet No. 1001)

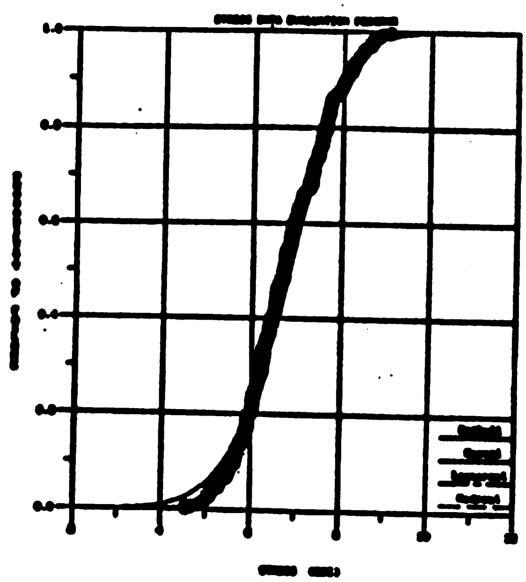


Figure 11.
P_f **vs. Robust Failure Data**
(Billet No. 3001)

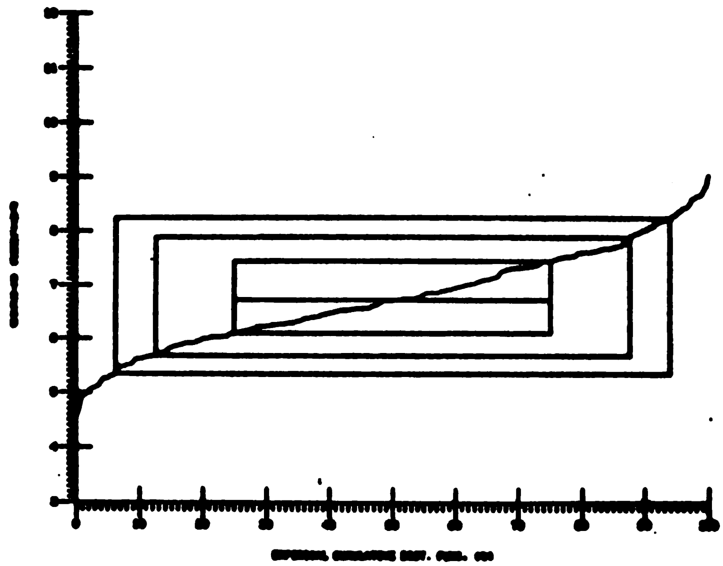


Figure 11a.
Quantile Box Plot
(Billet No. 3001)
Robust Data

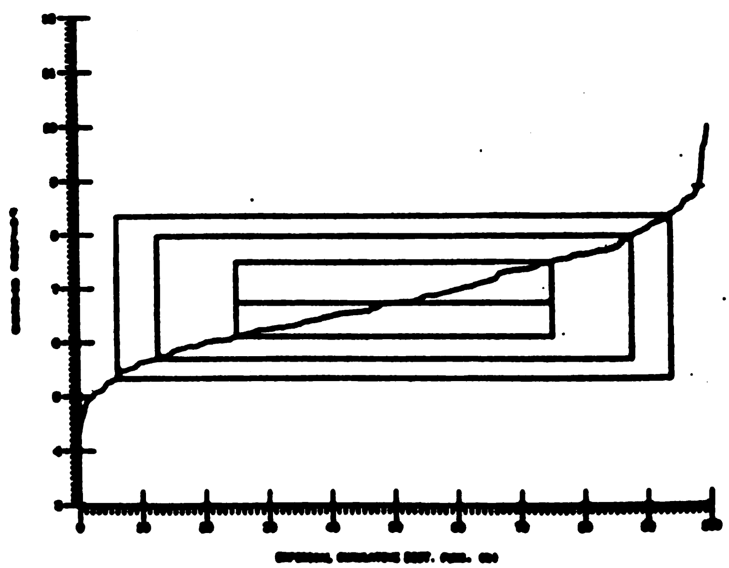


Figure 11b.
Quantile Box Plot
(Billet No. 3001)

TABLE 1 - COMPARISON OF MEAN FAILURE STRESS PREDICTED VS. ACTUAL APPLICATION OF WEIBULL SIZE EFFECT EQUATION (AREA DEP.)

Billet Identification Number and Number of Flex. Tests (N)	Weibull Shape Parameter			Predicted Mean Failure Stress (KSI) and Actual Test Result			Absolute Percent Diff. Pred. Vs. Actual		
	M_1	M_1^R	M_2	$\bar{\sigma}_P$	$\bar{\sigma}_P^R$	$\bar{\sigma}_2^T$	E_{TA}^R	E_{MSA}	E_{MSA}^R
1001 - (273)	5.48	7.59	7.05	2.06	2.82	3.78	34	32	8
2001 - (108)	5.86	8.77	12.2	2.14	3.44	4.14	20	39	9
3001 - (270)	7.13	7.87	6.78	2.65	3.43	4.29	25	26	19
4001 - (135)	6.68	8.25	7.60	3.14	3.32	4.06	22	6	.3

1 - Flexure Data

2 - Ring Data

R - Robustness

T - Tested Ring Results

M - Weibull Modulus

TA - Total Area of Spec.

MSA - Max. Stress Area

E - Error (%)

P - Predicted Result

σ - Failure Stress

TABLE 2 - COMPARISON MEAN FAILURE STRESS PREDICTED VS. ACTUAL (WEIBULL ANALYSIS - VOL. DEP.)

Billet Identification	Predicted Mean Failure Stress (KSI) and Actual			Absolute Percent Diff.	
	$\bar{\sigma}_P$	$\bar{\sigma}_P^R$	$\bar{\sigma}_2^T$	E^R (%)	E (%)
1001	1.28	1.91	3.78	98	195
2001	1.35	2.52	4.14	64	206
3001	1.78	2.51	4.29	71	141
4001	2.14	2.14	4.06	90	90

2 - Ring Data

R - Robustness

P - Predicted Result

T - Test Results (Ring)

TABLE 3 - COMPARISON MEAN FAILURE STRESS PREDICTED VS. ACTUAL (ALTERNATIVE METHOD)

Billet Identification	Predicted Mean Failure Stress (KSI) and Actual Ring Test Results			Absolute Percent Diff.		Coefficient of Variation	
	$\bar{\sigma}_P$	$\bar{\sigma}_P^R$	σ_2^T	$E^R(\%)$	$E(\%)$	$(C.V.)_2$	$(C.V.)_1^R$
1001	3.13	3.60	3.78	5.0	21.0	.160	.149
2001	3.44	4.07	4.14	1.7	20.0	.101	.119
3001	3.73	3.85	4.29	11.0	15.0	.160	.136
4001	3.93	4.15	4.06	2.2	3.2	.146	.138

1 - Flexure Data
2 - Ring Data
R - Robustness
P - Predicted Ring Failure

T - Ring Test Results
E - Error Percent
C.V. - Coefficient of Variation
σ - Failure Stress

THE TRASANA TERRAIN RESEARCH PROGRAM

Warren K. Olson
D. Hue McCoy

US Army TRADOC Systems Analysis Activity, WSMR, NM

1. Introduction. During the last decade, the US Army has grown increasingly aware of a need to fight outnumbered and win. Its field manuals and training literature are filled with suggestions regarding methods of generating "combat multipliers" (Ref. 1). One combat multiplier often discussed involves taking maximum advantage of battlefield terrain in the deployment of defensive and offensive forces. The extension of this concept involves the denial of advantageous terrain to the enemy. The TRASANA Terrain Research Program embraces this concept through the development of computer algorithms, terrain data bases, and operations research methodology which can be applied in the analysis of tactical alternatives and weapons systems effectiveness. The program also seeks to improve the the generalized understanding of the operational limitations imposed by the battlefield environment.

This paper discusses various computer programs, color graphics aids, and digital terrain data bases which are used in TRASANA's day-to-day analysis activities. New methods of developing and digitizing scenarios for insertion in Battalion combat models are presented. Case studies of the effects of terrain on battlefield activities are also provided. Finally, current research efforts aimed at classifying terrain and generalizing its effects on intervisibility and mobility are discussed.

2. Terrain Analysis Programs. Most of TRASANA's terrain analysis requirements center around the need to understand intervisibility conditions connected with the Army's current and future target acquisition and weapons delivery systems. These analyses are made more challenging by Army requirements to operate in many theaters around the globe, and use both airborne and ground-mounted surveillance and weapons systems in accomplishing its missions. The computer algorithms most frequently used in defining line-of-sight (LOS) conditions and in visualizing battlefield terrain are described below. Many of these programs are documented in Reference 2. However, other specialized, less frequently-used algorithms exist for digitizing information, preparing and collecting data, processing information, and in providing rapid analysis for study support.

a. Optimum Vantage Point LOS Algorithm (OPT1)

This program uses Defense Mapping Agency (DMA) digitized terrain with vegetation and urban development codes as a computational data base. The program generates line printer contour maps, vegetation/urban code maps, and line-of-sight (LOS) maps in a UTM grid reference system. The user may specify the sensor/target location and altitude for three different sensor/target

---

This article also appears in the proceedings of the 21st Army Operations Research Symposium.

platforms (ground, helicopter/nap-of-the-earth, and fixed wing aircraft/constant altitude). The LOS data is stored for use in a more sophisticated program (OPT2) which examines the unions and/or intersections of coverage from one or more sensors.

b. Optimum Vantage Point Pass 2(OPT2)

OPT2 uses the output from OPT1 as input. The program is capable of producing LOS maps for any or all of the sensors preprocessed by OPT1. OPT2 can examine the union and intersections of any set of sensors with any second set of sensors. This is particularly useful for examining the coverage of a set of defender weapons against an expected area threat. The program can be used to restrict the view of a given sensor to any direction with a limited field of view. All outputs from this program can be directed to the CALCOMP plotter or Color Graphics System and scaled to fit 1:50,000 maps (see Figure 1).

c. Point to Point LOS (PT2PT)

Point to Point calculates LOS from a discrete point to a set of target points. This program is used in the calculation of advance route/flight path exposure lengths and associated time durations from any given sensor. The output can be directed to CALCOMP plotter to give a profile of the terrain between the sensor and the target. This profile will graphically show the presence of vegetation and urban features (see Figure 2).

d. SEEFAR Model (AMSAA)

SEEFAR is an improved model for producing line-of-sight maps. Many models determine whether a target is within view by mathematically constructing a terrain profile between observer and target and examining it to see if it interferes with line-of-sight; this requires generating a completely new profile for each target position. SEEFAR avoids this time-consuming profile generation by dynamically recording the characteristics of a "running horizon" as computations are made for points further away from the observer. For each target point, a check is made to determine whether the target is behind the "horizon". This new approach results in a dramatic savings in both storage and computing time requirements. SEEFAR is ideally suited for air/ground, large area analyses (see Figure 3).

e. Plot Contour.

This program plots topographic contour maps from digital data, and includes vegetation and urban cover. The program is useful in both quality assurance checks of the loaded data base, and in the simulation of battlefield activities (see Figure 4).

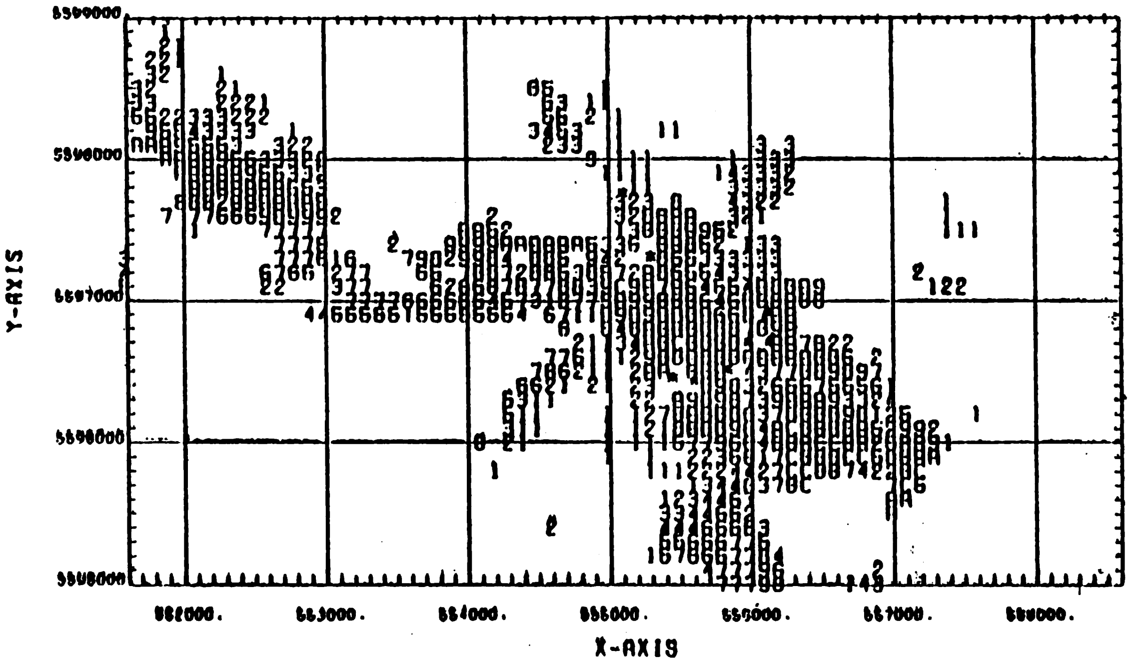


Figure 1 OPT2 Program Output

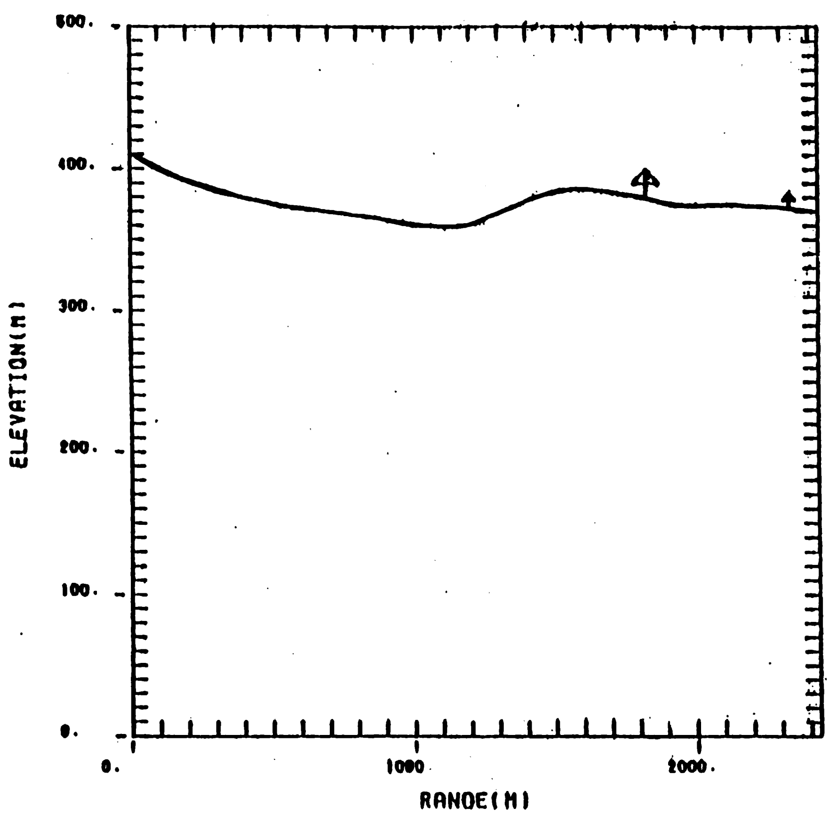


Figure 2 Point-to-Point Output

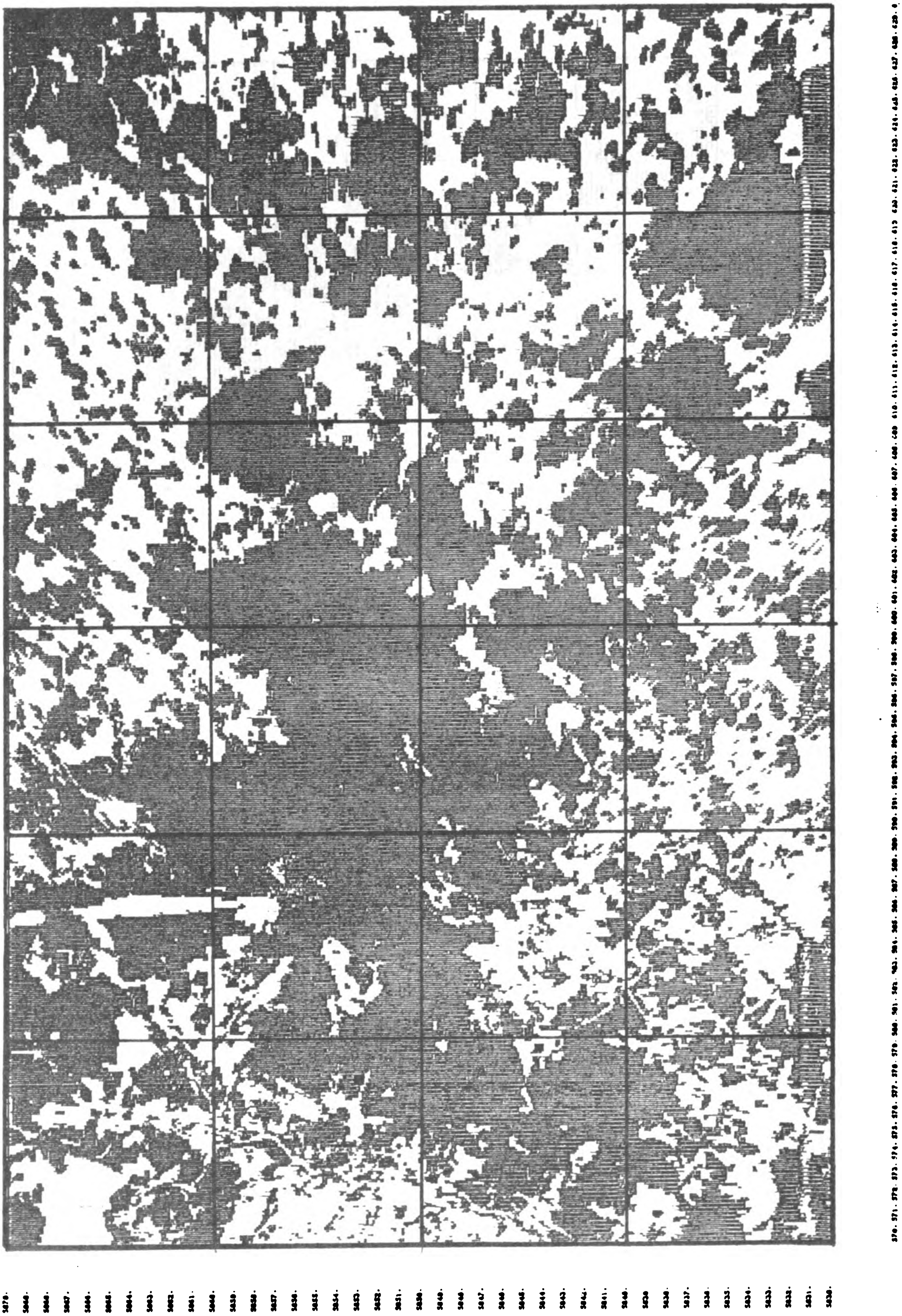


Figure 3. SEEFAR Model LOS Display

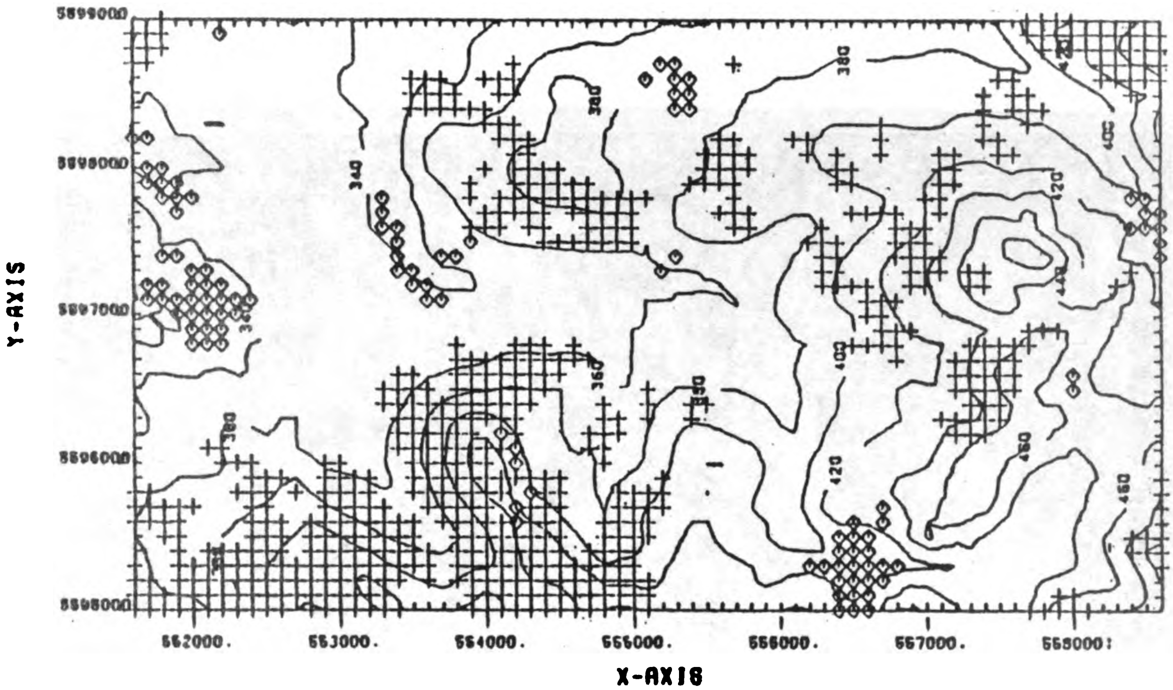


Figure 4 PLOT CONTOUR Topographic Map

f. TRPGRFX.

This Color Graphics program displays elevation intervals in colored bands; displays vegetation and urban features in appropriate colors; draws observer (defender) locations and attacker routes; generates LOS to an area or along routes only; provides a zoom-in capability for high resolution work; generates LOS statistics (to include: probability of LOS, in-view/out-of-view distributions, first sighting range, expected opening range, and average in-view/out-of-view segments per path); generates multiple attacker routes (using formations), and develops attacker position vs. time, with an LOS indicator for post processing other statistics (see Figures 5-9).

g. CALCOMP 3-D Package.

The program draws 3-D gridded representations of the earth's surface with or without perspective, viewed from the surface itself or from a defined altitude (see Figure 10).

h. 3-D Terrain Gray Scaling Progam.

This program produces a 3-D perspective image of digital terrain data by defining both a sun angle and viewing angle. The view on the Color Graphics resembles an oblique aerial photograph of an actual terrain surface (see Figure 11).

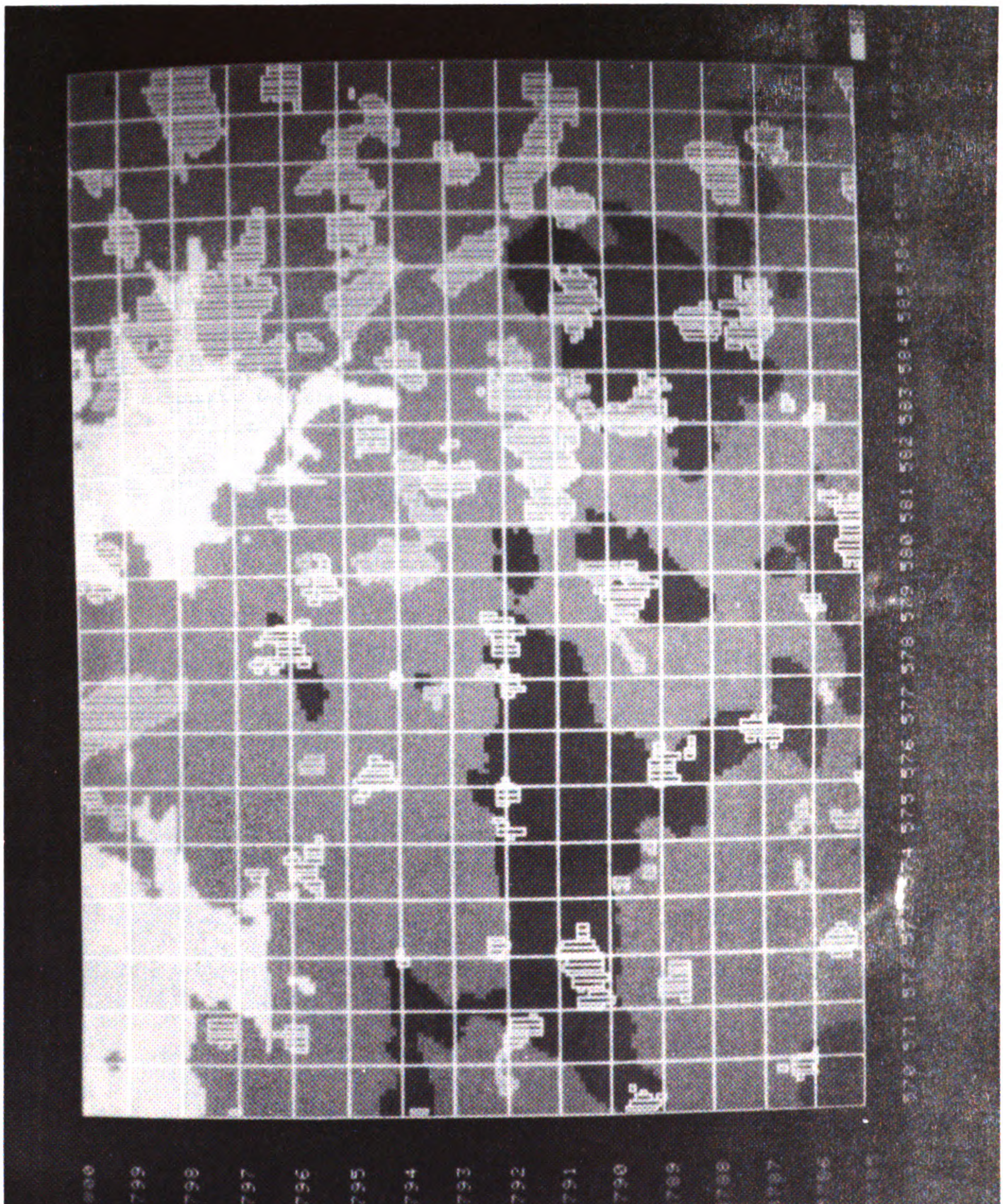


Figure 5 TRPGRFX Elevation, Vegetation, Urban Display

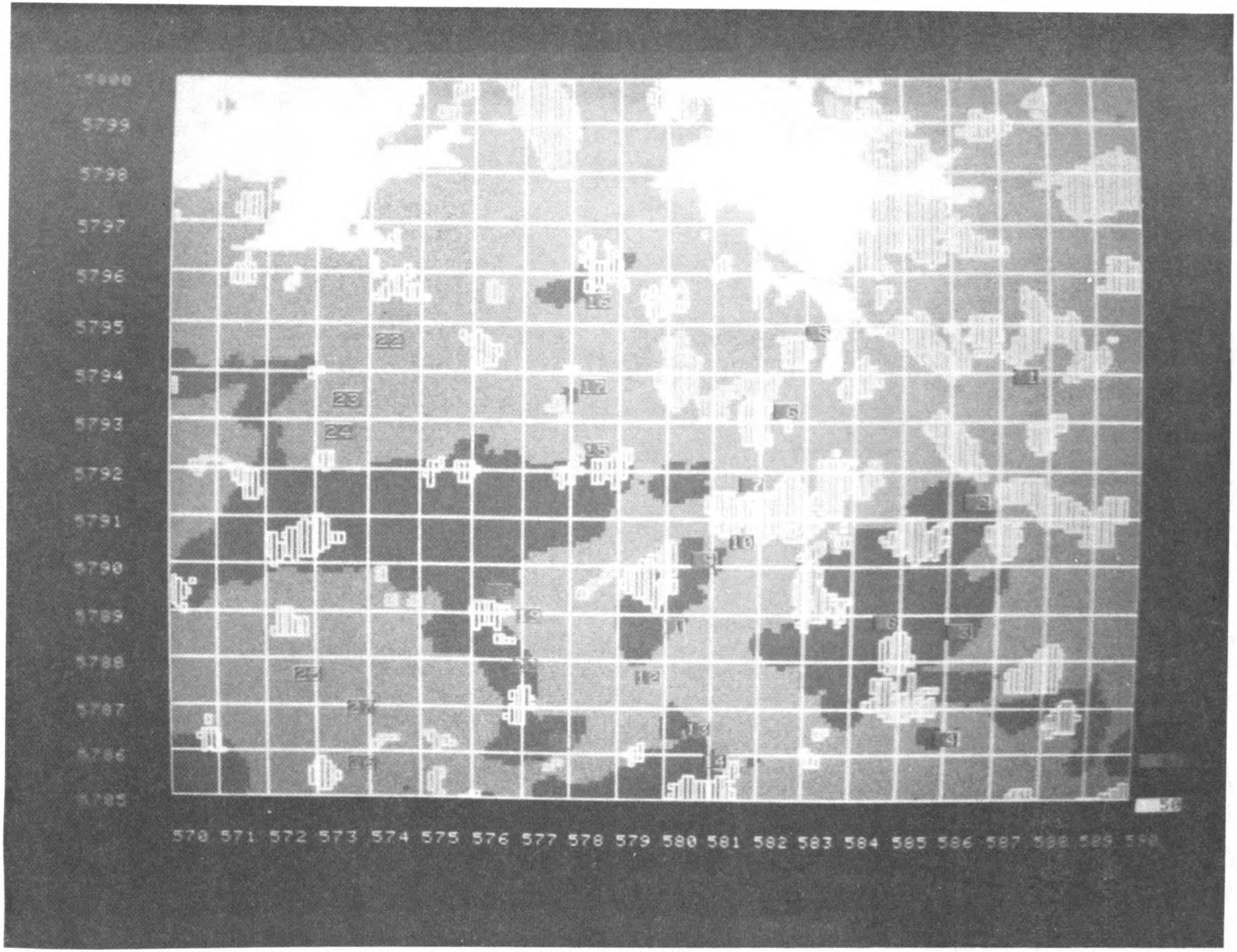


Figure 6 TRPGRFX Observer Position Display

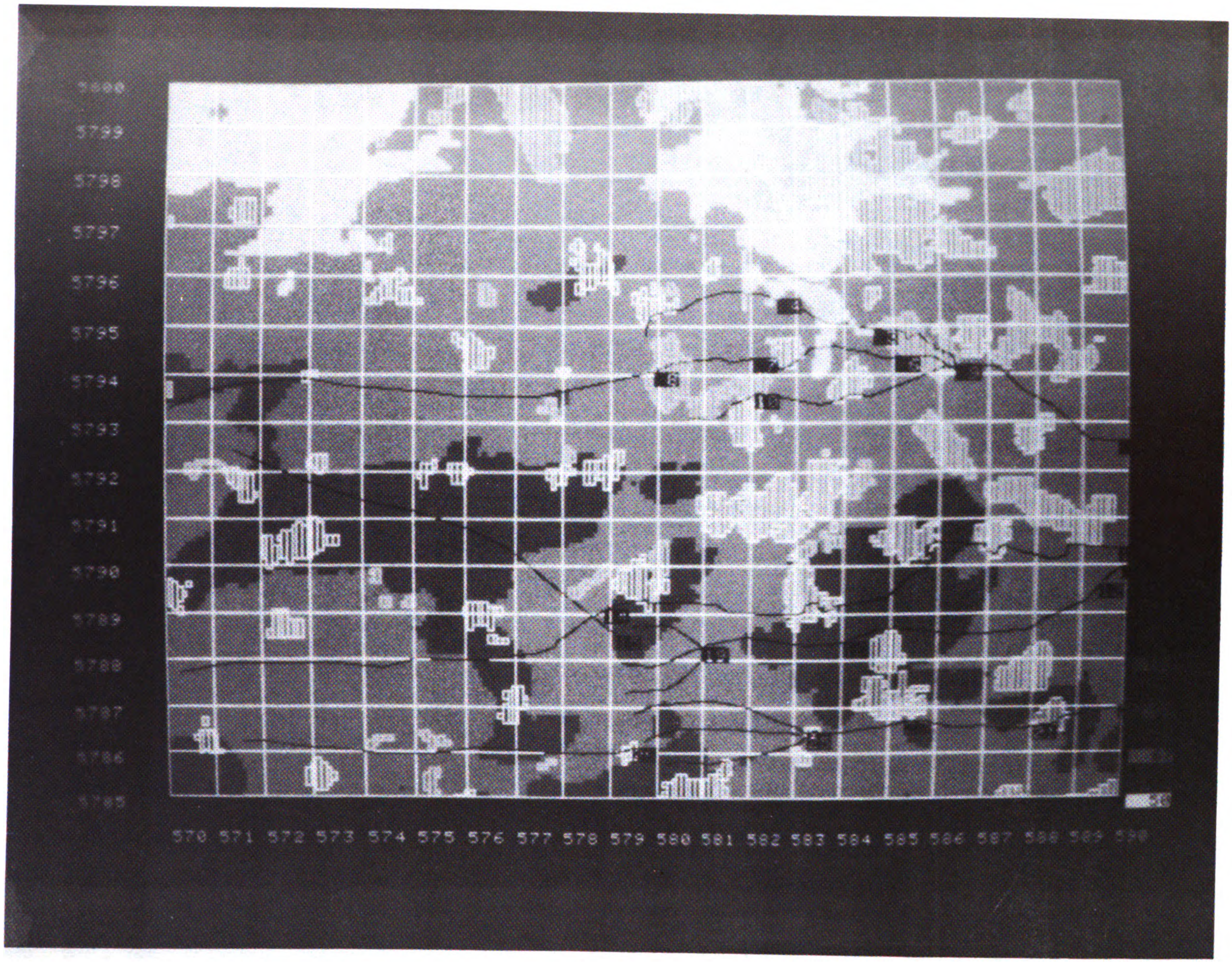


Figure 7 TRPGRFX Attacker Route Display

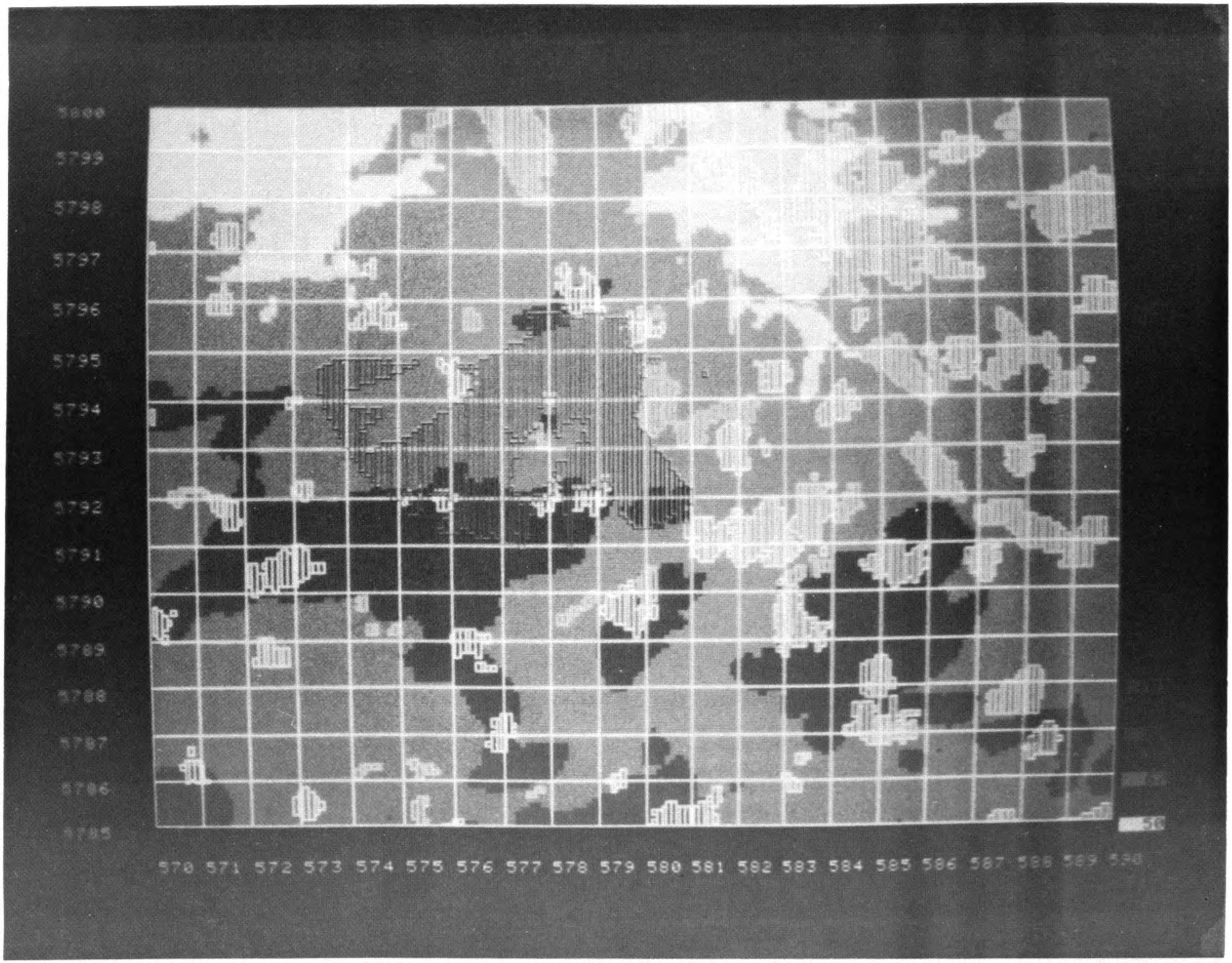


Figure 8 TRPGRFX Line-of-Sight Display

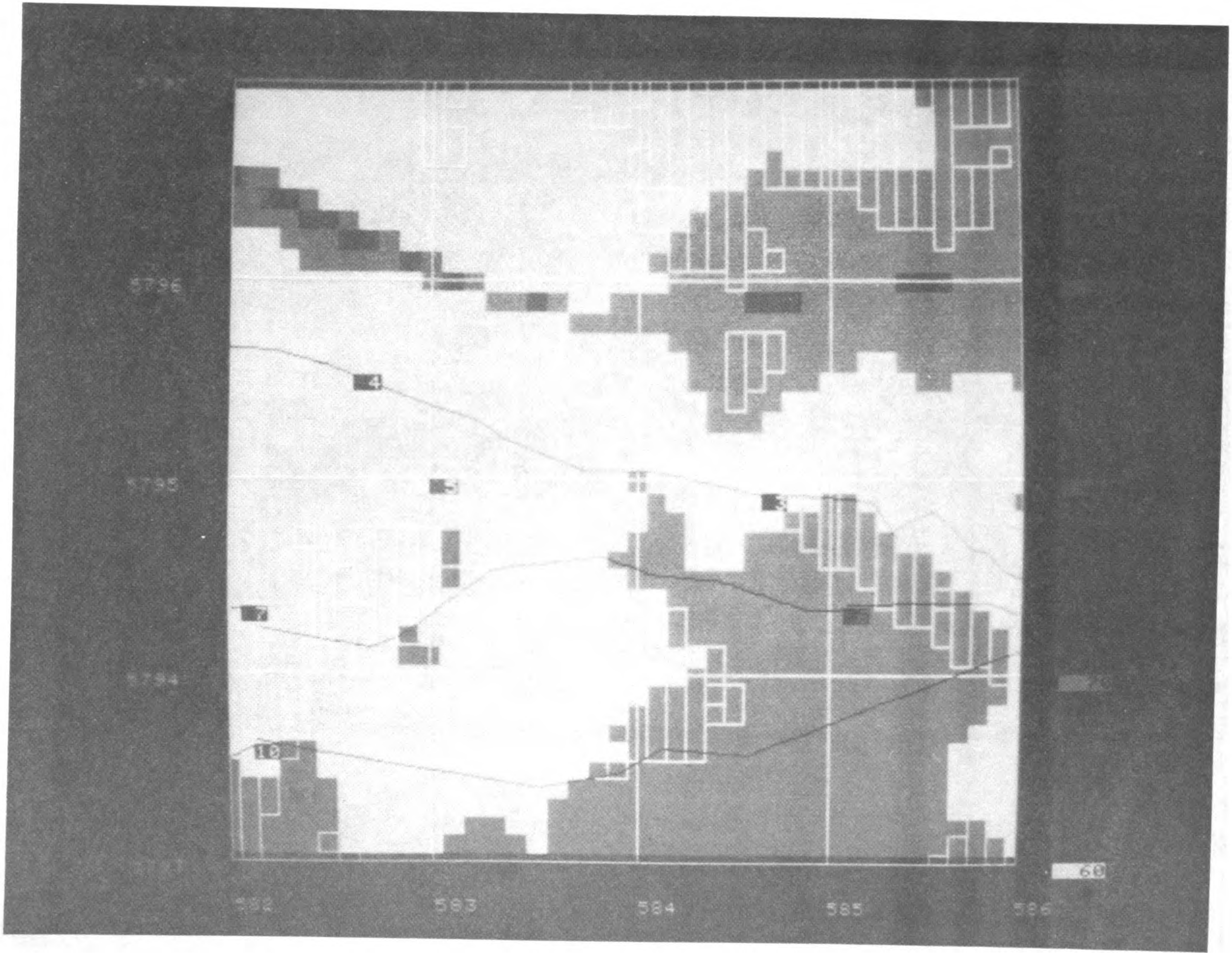


Figure 9 TRPGRFX Zoom-In Display

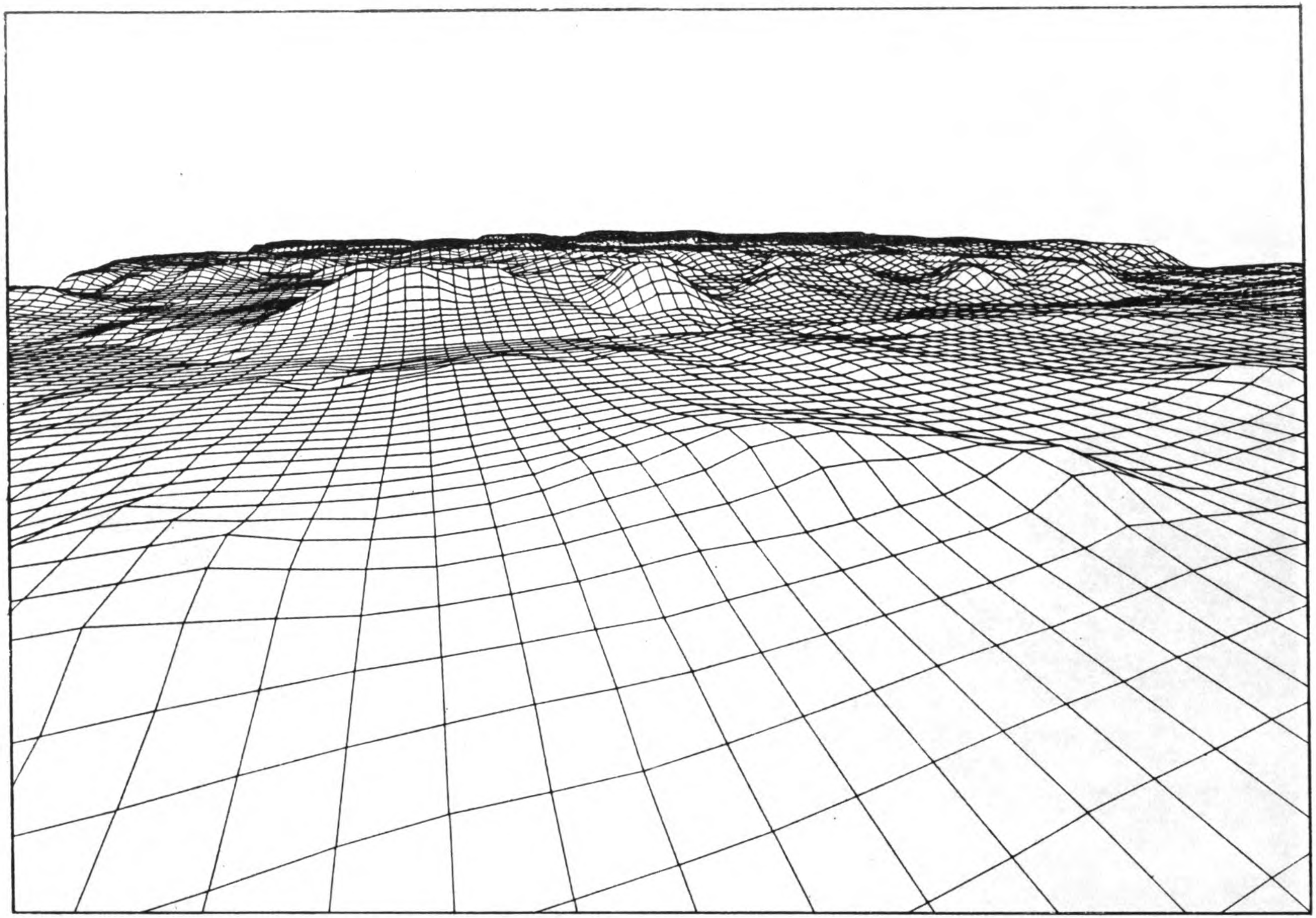


Figure 10 CALCOMP 3-D Perspective View of Fort Hood

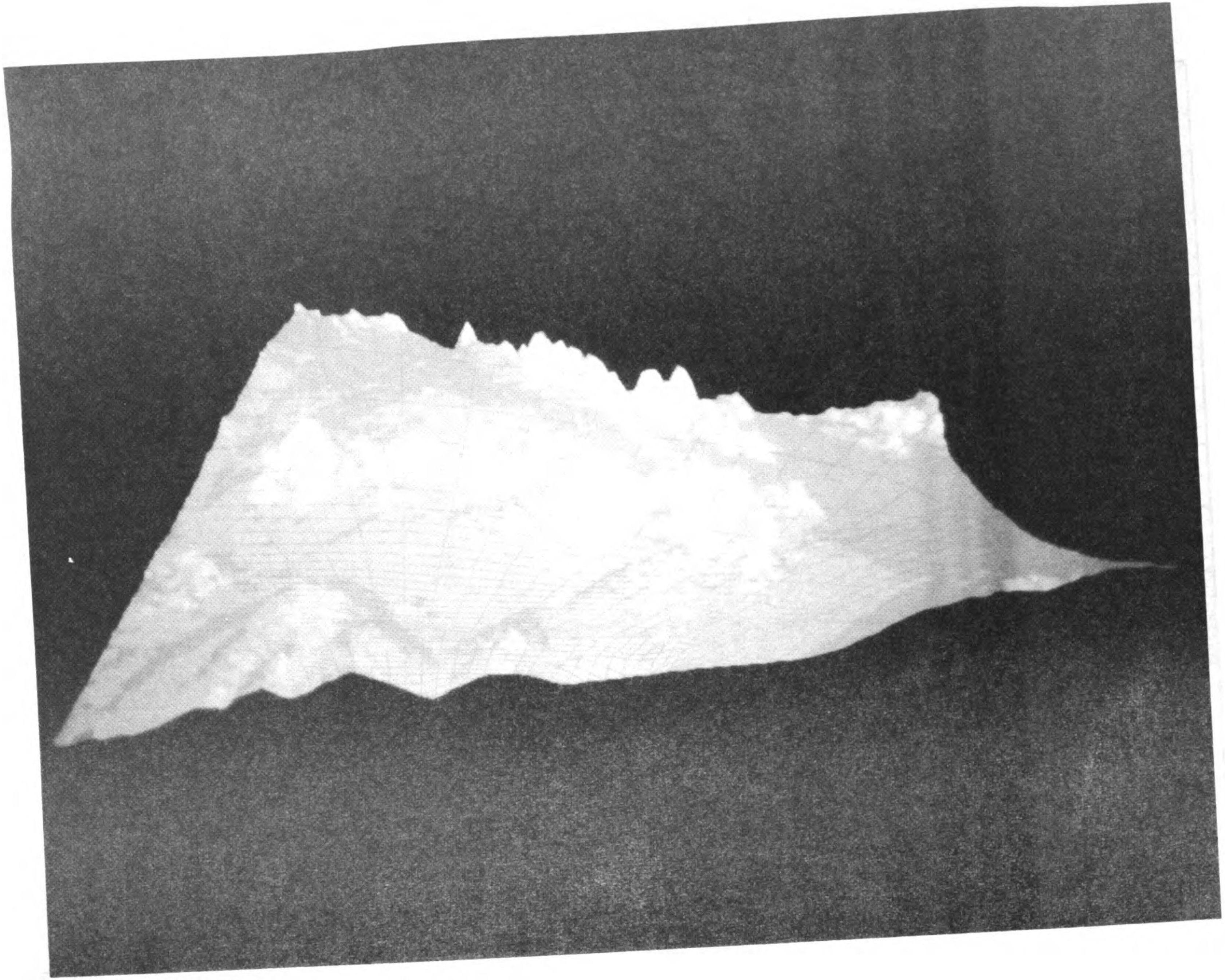


Figure 11   3-D Terrain Grey Scaling Program (Ft Irwin)

i. LOS Graphics.

This program uses digitized terrain data with vegetation and urban development heights as a computational data base. The program produces contour maps, vegetation plots, and road nets, within a UTM grid reference, using OPT1 as a driver model. LOS maps are produced almost instantaneously. The effects of smoke on visual LOS can be readily displayed. The program also produces composite LOS maps which show the union and intersection of LOS from different sensor locations (see Figure 12).

j. CARMONETTE History Display Program.

This real time, interactive graphics program displays the terrain, position of fighting units, firing events, and kills from a CARMONETTE history file. It can also be used to generate line-of-sight maps using the observer locations and the CARMONETTE land-deck. The program is used to develop scenarios and analyze battle outcomes (see Figures 13-18).

k. Data Base Display/DMA Merge.

This interactive program package "splices" multiple DMA terrain data tapes, by allowing an analyst to see the entire data library while seated at the Color Graphics console. The analyst, through joy stick or keyboard entry, specifies the region desired for analysis purposes. The package then automatically prepares the data file for use with other programs.

In addition to these programs, the Agency maintains a limited capability to do mobility processing with the Army Mobility Model (AMM76). However, most of the required mobility information is processed for TRASANA by the Waterways Experiment Station, Vicksburg, Mississippi.

3. <u>Computer Color Graphics</u>. A RAMTEK Color Graphics System (CGS) installed in 1977 interfaces with the TRASANA UNIVAC 1100/82 Mainframe computer. The TRASANA Computer Graphics Facility (CGF) is an applications/user oriented ineractive color graphics system. The hardware for the system includes a RAMTEK 9300 color raster graphics system with associated devices (trackballs, joysticks) including four display units, a SAC sonic digitizer, and a VARIAN V-77 minicomputer. The software includes a CGF applications library, plus an interface that permits use of the CALCOMP graphics library. The user programs reside on the 1100/82 and the graphics facility programs reside on the V-77 in an effort to distribute the workload for the CGF and therefore enhance the rate of response for user applications. This system is used for terrain analysis, detailed scenario preparation and many other analysis applications. The system components are shown in Figures 19 and 20.

4. <u>Digital Terrain Library</u>. Most of the digitized terrain data used for study purposes comes from two sources. The topographic information (to include elevation data at 12.5m intervals, vegetation, and urban features data) are provided by the Defense Mapping Agency, Washington, D.C. For some applications, including detailed combat modeling, additional information

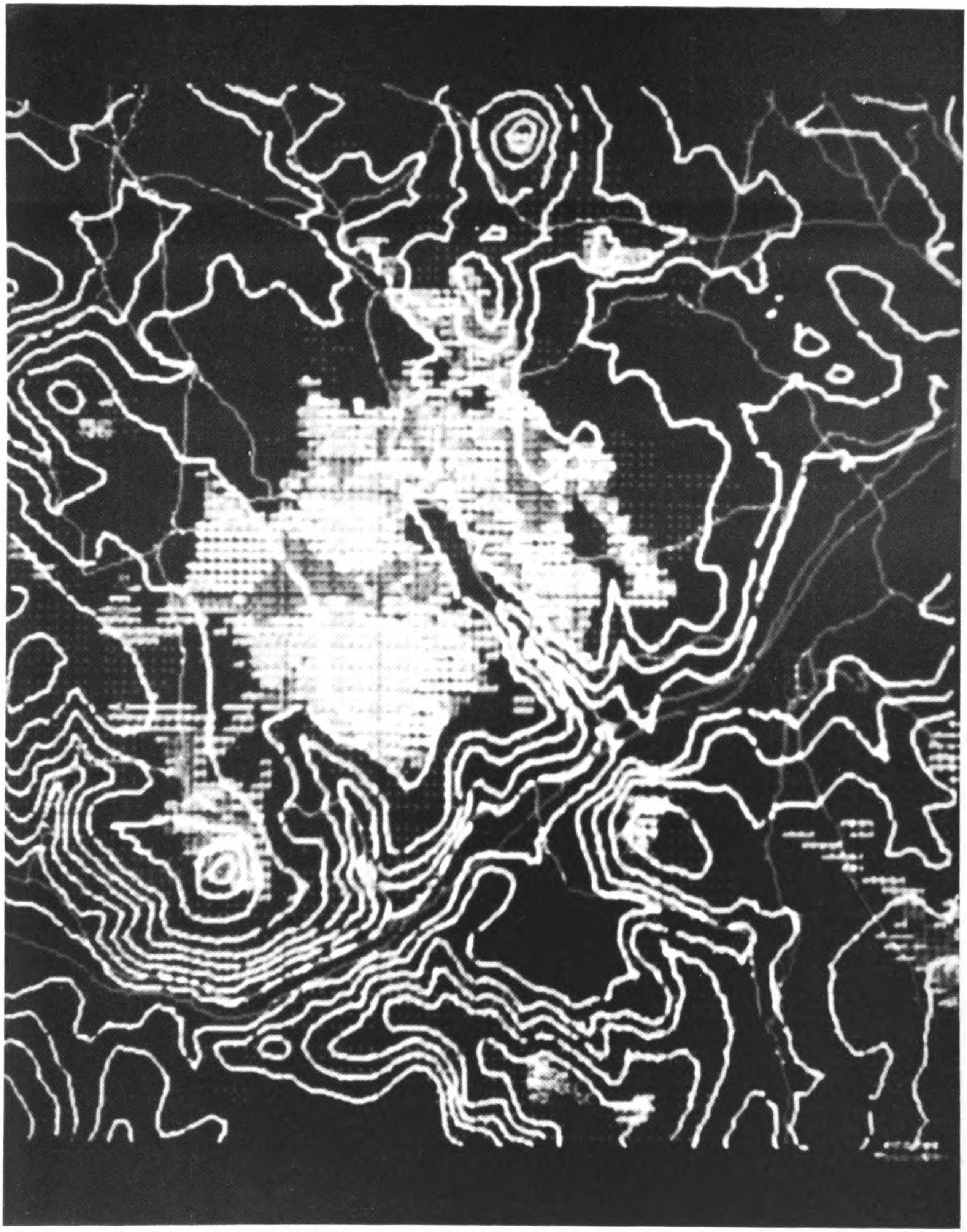


Figure 12 Line-of-Sight Graphics Display

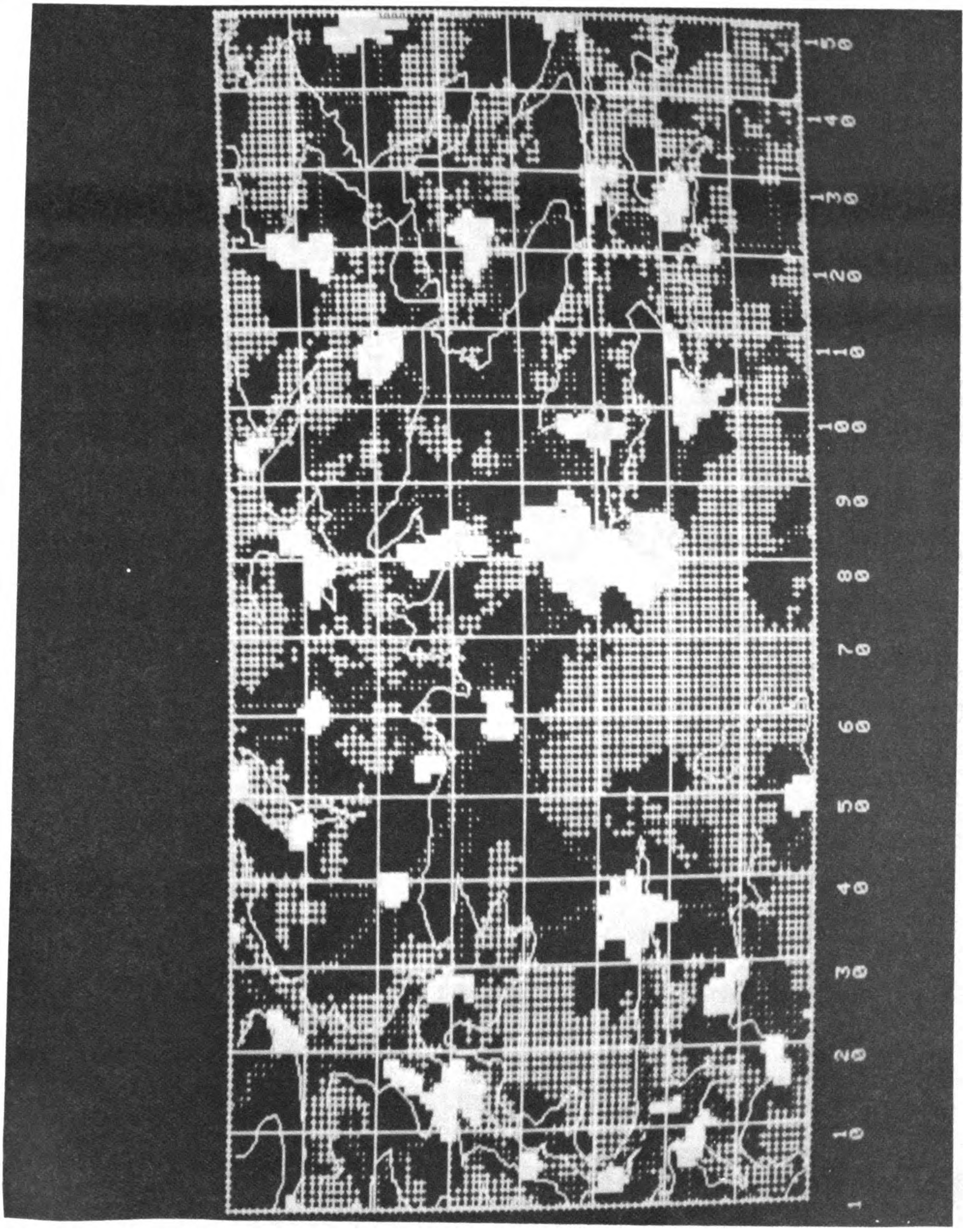


Figure 13 CARMONETTE Color Graphics Terrain Display

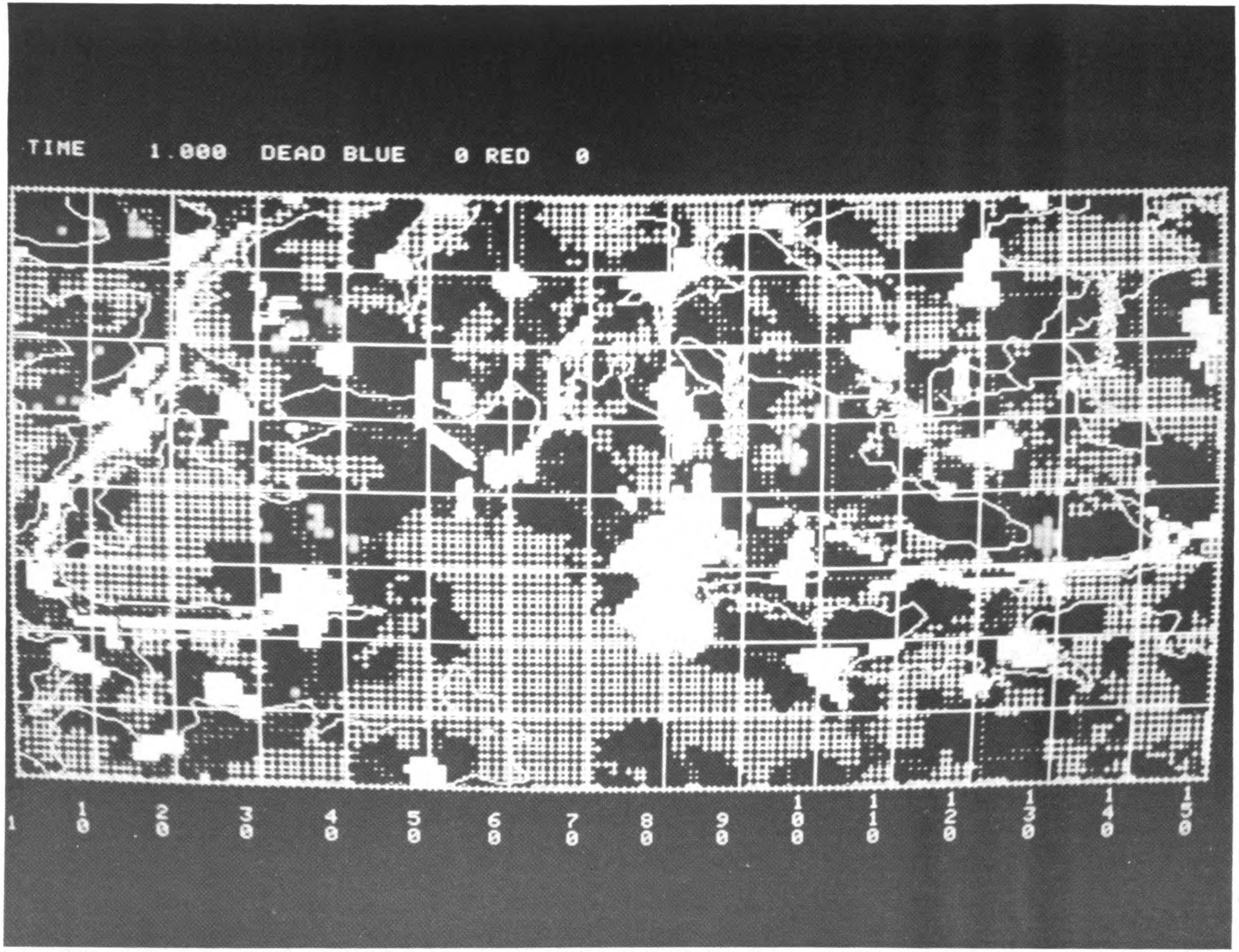


Figure 14 CARMONETTE Scenario — BLUE Positions and Minefields

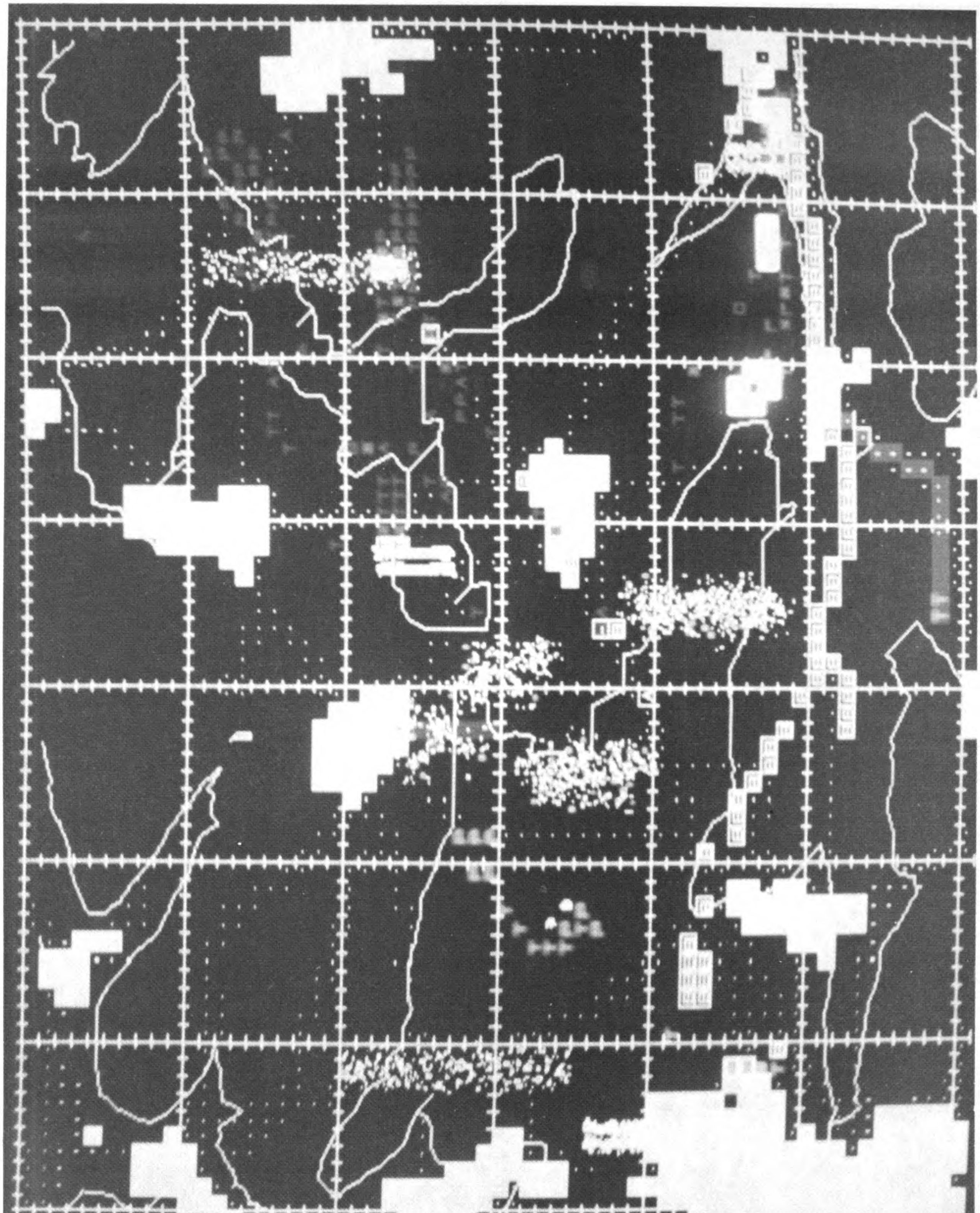


Figure 15 CARMONETTE GRAPHICS - BLUE Barrier Plan (Zoom-In)

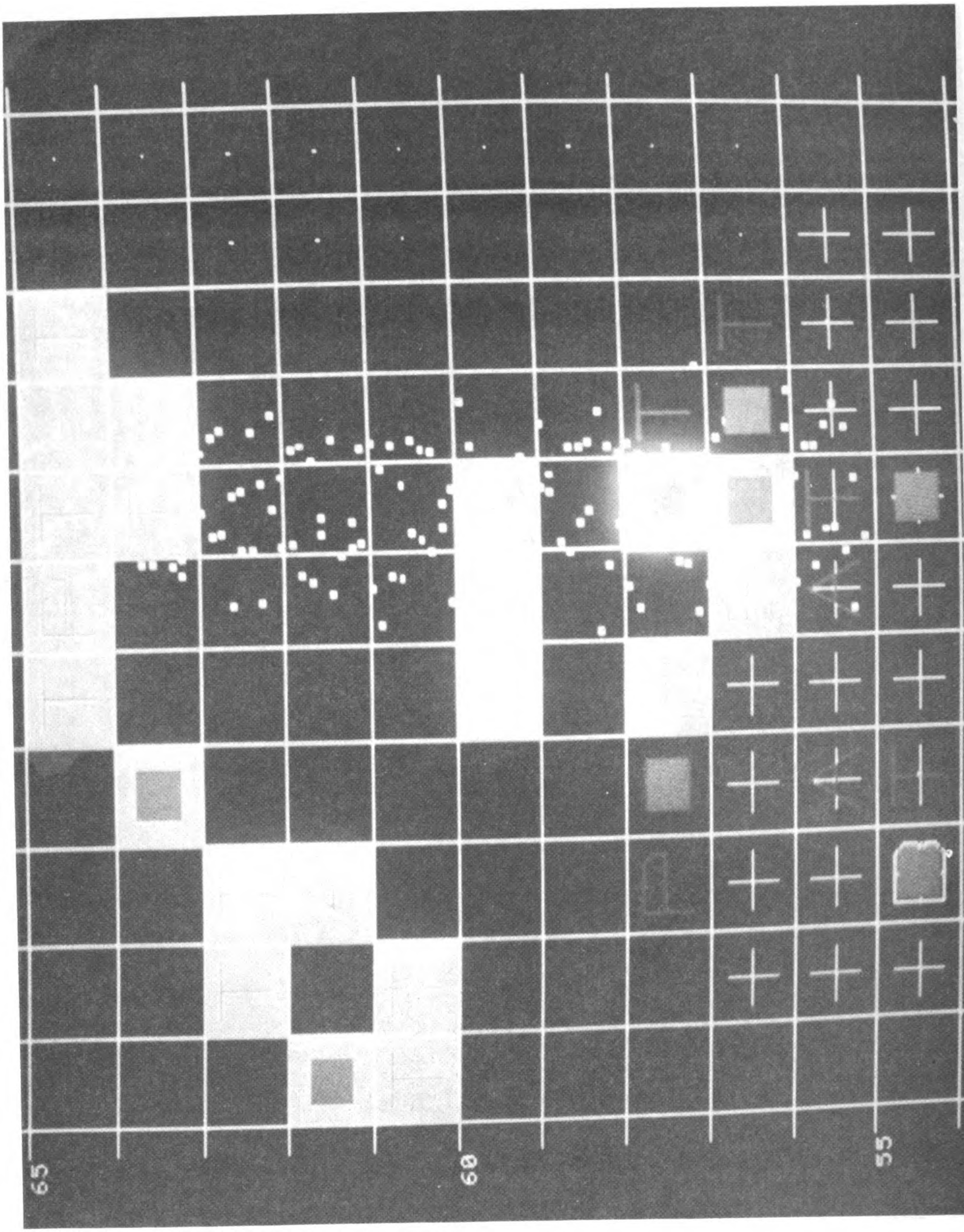


Figure 16 CARMONETTE LOS Check - Tanks (T) And APCs (A)

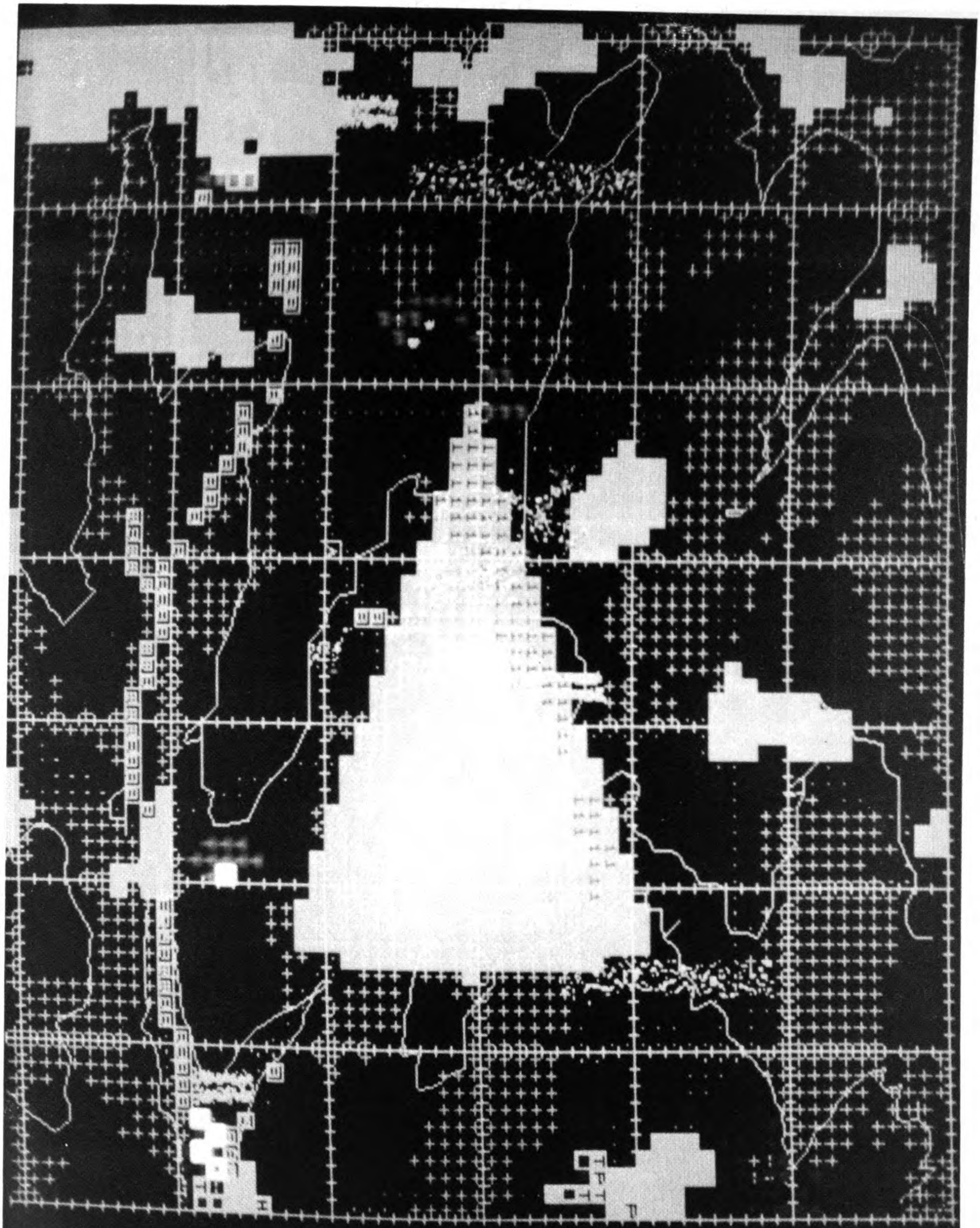


Figure 17 CARMONETTE Graphics - LOS Fan

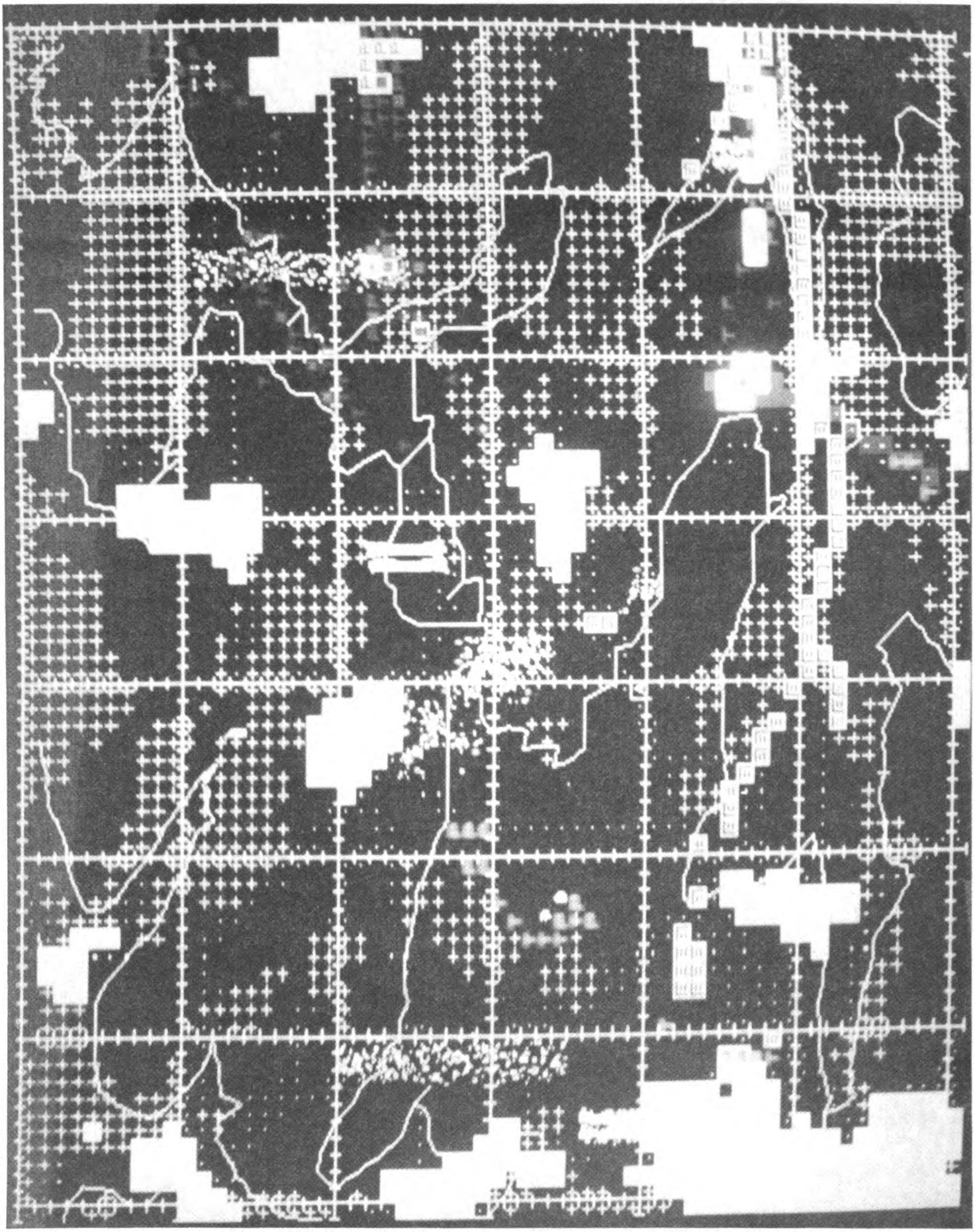


Figure 18 CARMONETTE Graphics - RED Mine Breach (5 Minutes)

EQUIPMENT

1 UNIVAC 1108 HOST COMPUTER

1 VARIAN 77-613 MINI COMPUTER

5 VARIAN 620/i MINI COMPUTERS

4 MONOCHROMATIC DISPLAYS

4 RAMTEK COLOR DISPLAYS

4 TRACK BALL/JOY STICK

1 GRAPH PEN-3 DATA TABLET (36 x 36")

1 ELECTROSTATIC PRINTER/PLOTTER

Figure 19 TRASANA RAMTEK Color Graphics System

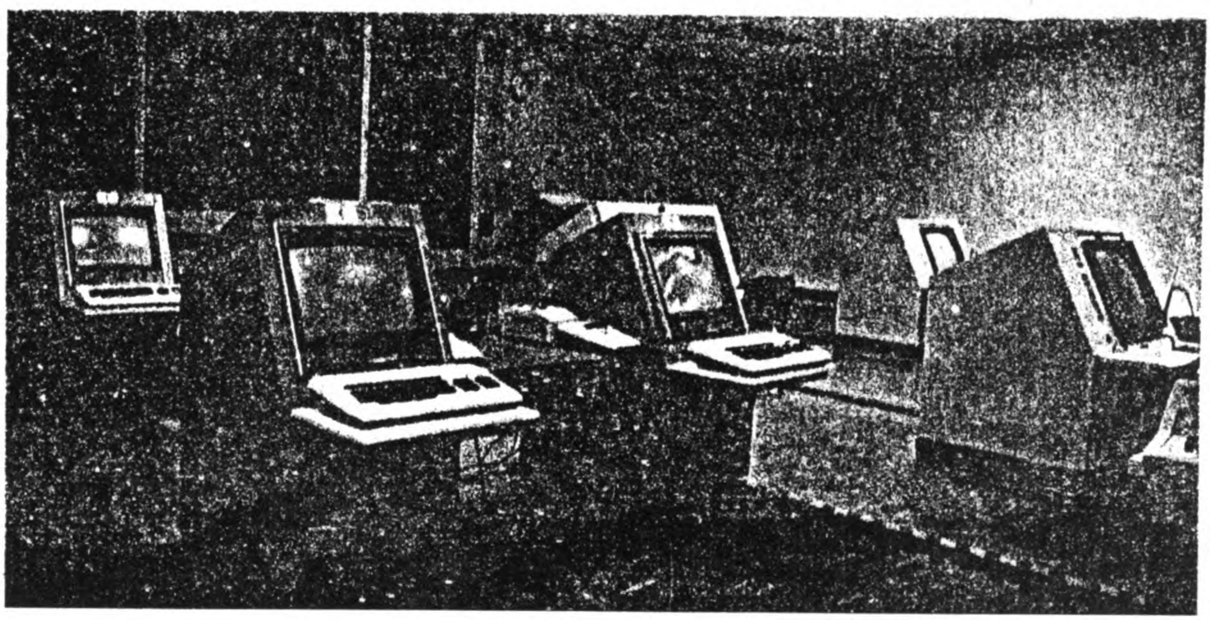


Figure 20 Color Graphics Facility

concerning cover, concealment, and on-and off-road trafficability are required. This data is purchased under contract from the Waterways Experiment Station. Detailed data are available for several regions around the world (Ref. 3). Some of the available regions are shown in Figure 21. Other DoD agencies and contractors are currently involved in the development of data on the structure of cities for use in studies of military operations in urban terrain (MOUT). The Defense Mapping Agency will shortly develop a new prototype land combat data base for evaluation (see Appendix A). If approved, the new data base would provide analysts with heretofore unavailable digital data, addressing surface configuration, surface features, surface materials, hydrography, movement and other features (approximately 200 bits of information would be stored per grid point). A DMA prototype land combat data base for Ft. Lewis, WA will be available for testing in November 1981.

5. Test Applications. With these tools and data bases, it is possible to investigate many aspects of the intervisibility "combat multiplier" without programming expensive field tests, scheduling manpower and equipment, experiencing delays due to weather, and facing a host of other problems. Rather, in the comfort of a computing laboratory, the analyst can easily duplicate many of the measurements made in previous field tests such as the NATO Range Study, Lost Horizons, TETAM, HELAST, and CHINESE EYE. These measurements and studies can be completed in days or weeks, as opposed to the prevous yard stick of months or years. Also, it is possible to analyze regions of the earth's surface which are not readily accessible for field testing, either due to distance, land use, or political considerations.

In order to feel comfortable with the results of these simulations, however it must be demonstrated that the computer algorithms and data bases are capable of closely matching field test results. In the late 1960's and early 1970's some effort was spent comparing computer predictions of intervisibility with major field test results. The findings were somewhat mixed, but held promise. Difficulties with the documentation of the field tests and digitized terrain resolution were noted, and modifications to some of the computer algorithms appeared warranted.

Recently, TRASANA had the opportunity to access some high resolution field test results developed by the Combat Development Experimentation Command (CDEC) during a telemetry test at the Army's new training facility at Ft. Irwin, California (Ref. 4). In the referenced study, while using the terrain analysis programs to optimize an RF position/location system, TRASANA was also able to validate these models with actual telemetry data (See Figure 22). This study is reported in detail in a separate 1981 AORS paper. It is sufficient here to note that 88 percent agreement between measured and simulated results was demonstrated. This is a remarkable achievement, considering the nature of the terrain and the fact that the telemetry system which produced the test data has a slight (but unquantified) capability to "see" beyond the horizon and through vegetation. In the Ft. Irwin study and a companion study just completed for the Ft. Sill HELBAT 8 Test (Ref. 5), the LOS programs have also been used to grade coverage quality, to "orthogonalize" positions in order to minimize target location error, and to optimize coverage with a minimum investment in equipment (see Figure 23).

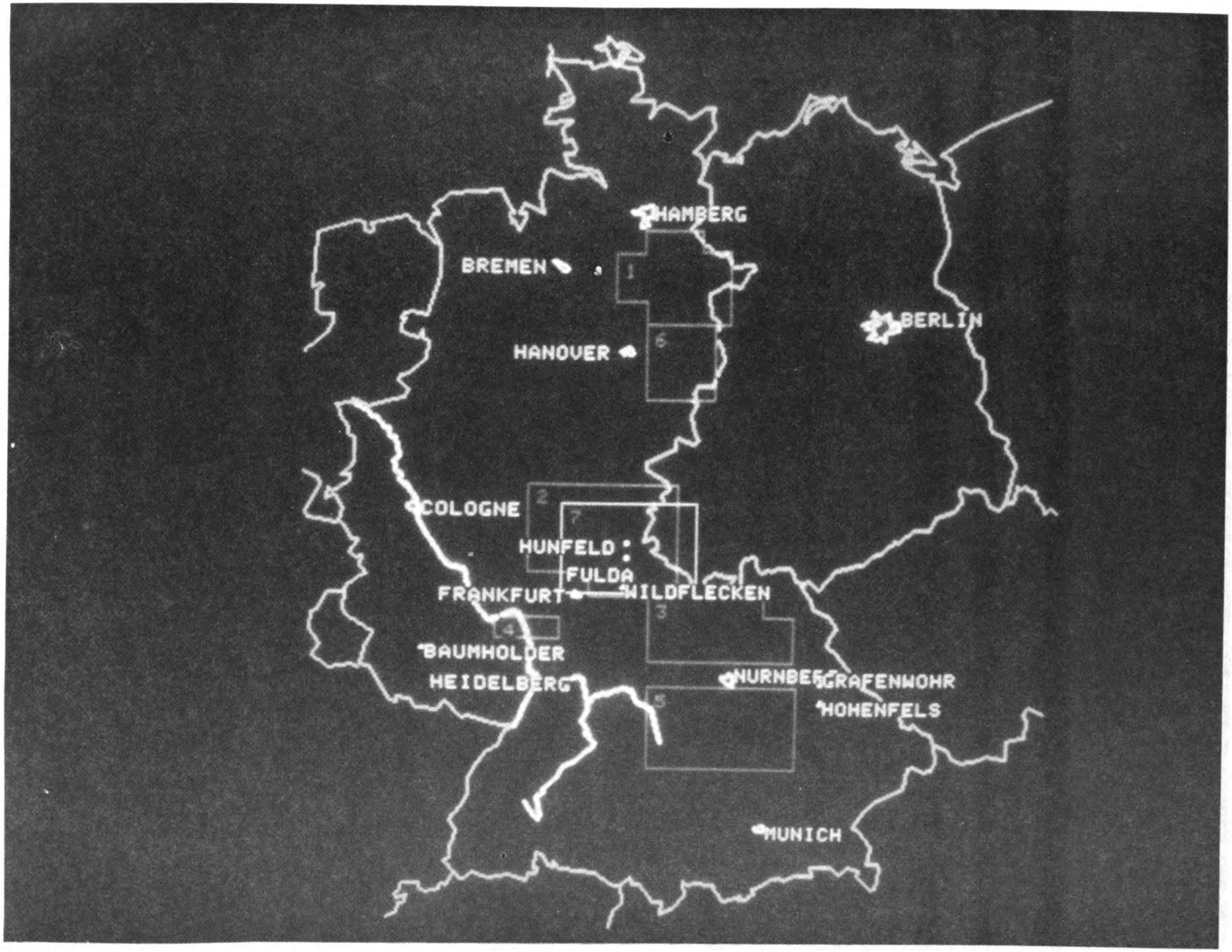


Figure 21 Digitized Terrain Display — West Germany

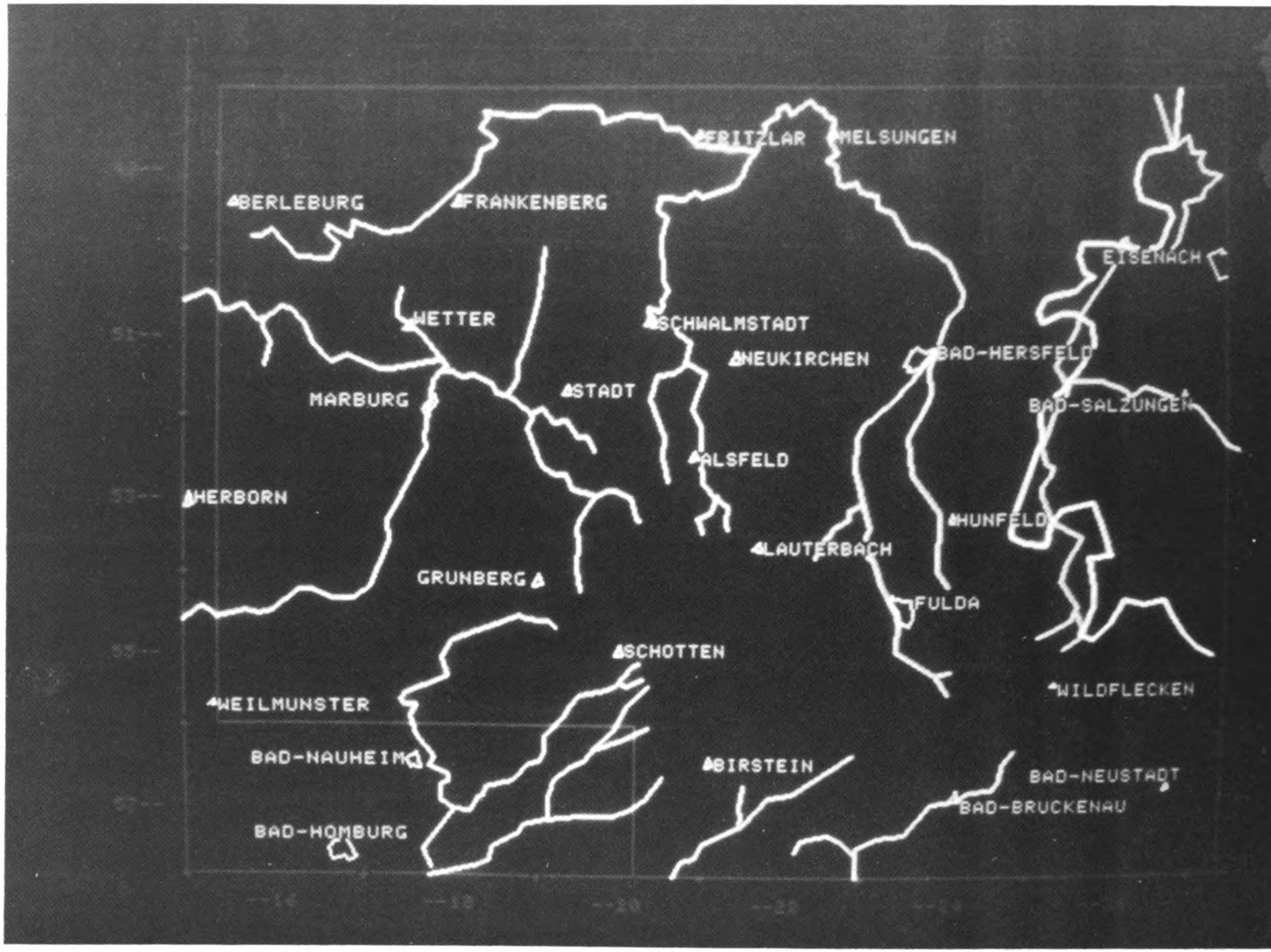


Figure 21a Digitized Terrain Display — Fulda Area

A-Station Number	Agree (%)		Disagree (%)		Total Percent Agreement
	Seen	Not Seen	Seen By Computer Only	Seen By A-Station Only	
1	26	67	1	6	93
2	14	83	2	1	97
3	46	47	2	5	93
4	32	56	11	1	88
5	24	69	3	4	93
6	43	48	1	8	91
7	26	60	1	13	86
8	21	60	18	1	81
9	39	48	6	7	87
10	28	60	4	8	88
11	47	41	6	6	88
12	53	35	5	7	88
13	38	47	5	10	85
14	15	77	3	5	92
15	30	57	6	7	87
16	39	50	2	9	89
17	30	54	16	0	84
18	48	39	12	1	87
19	11	70	14	5	81
AVERAGE	32	56	6	6	88

Figure 22 Fort Irwin Computer Simulation/Field Test Comparison

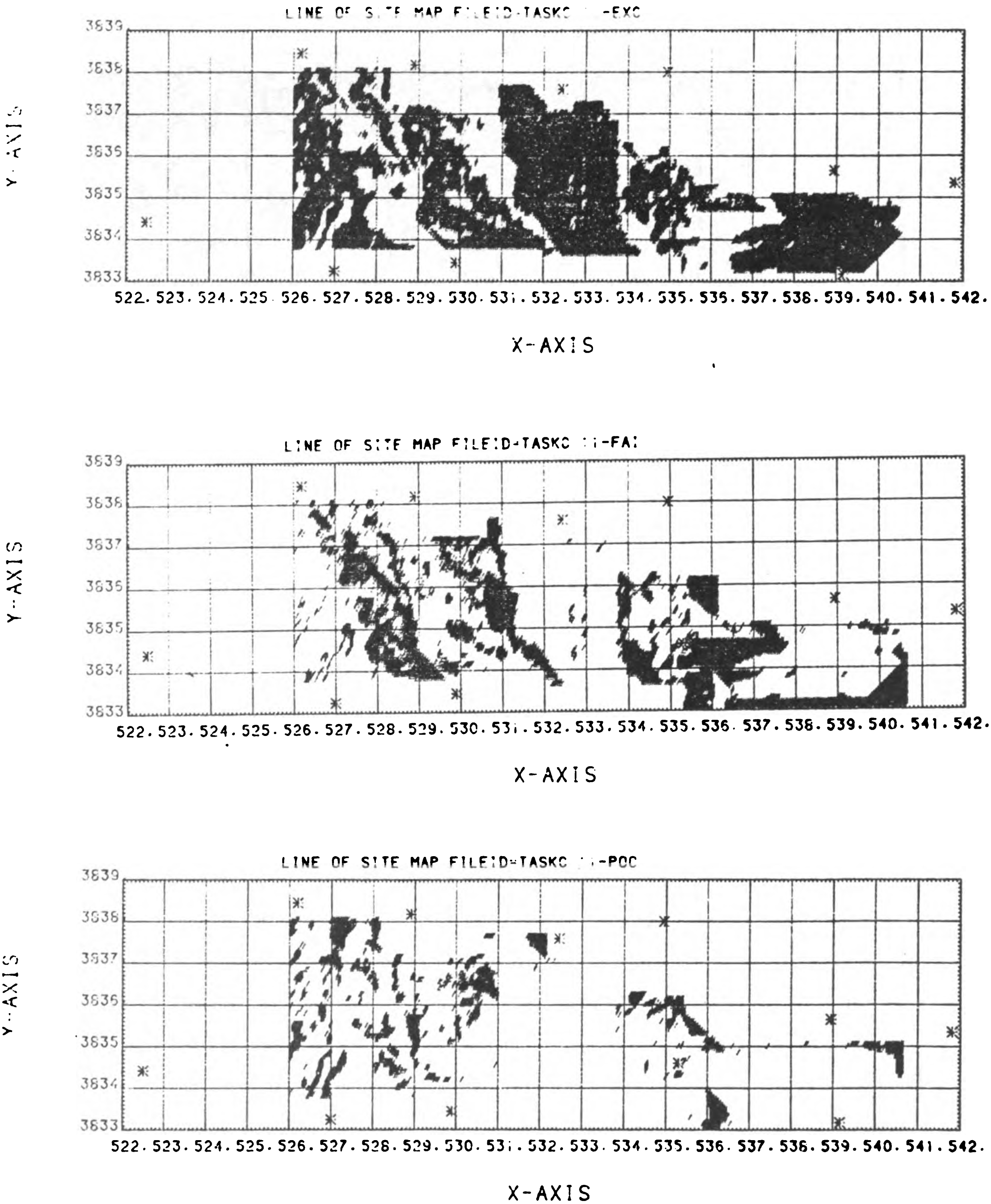


Figure 23 Fort Sill Telemetry Grading Analysis (11 Sites)

6. Terrain Analysis Studies. The TRASANA Terrain Research Program is also used to support a wide variety of Army studies, ranging from weapon systems design, to studies of tactics, reliability, and trade-off analyses between competing systems. Preprocessed information is also developed for use in higher-level models. Several examples are provided below.

a. ATGM Design.

One class of problem which occurs periodically concerns the appropriate range capability for a given type of antitank guided missile (ATGM). In a recent analysis, 16 scenarios were evaluated with regard to LOS and firing opportunity (Figure 24). Distributions of engagement ranges and intervisibility lengths were developed for European and Middle Eastern scenarios. The results show that increased engagement opportunities could be achieved by extending the range of the medium range ATGM from 1000 to 2500 meters; however, the value of increasing the range of long-range (heavy) ATGMs beyond 4000m, especially in the presence of smoke is doubtful (Ref. 6).

b. Main Battle Tank Reliability.

Addressal of the significance of M1 tank reliability test data recently required the analysis of distance moved on the battlefield during combat (see Figure 25). Analysis of fourteen European scenarios shows that the distribution of movement distances is log normal with a mean of 10.0km per move (Ref. 7).

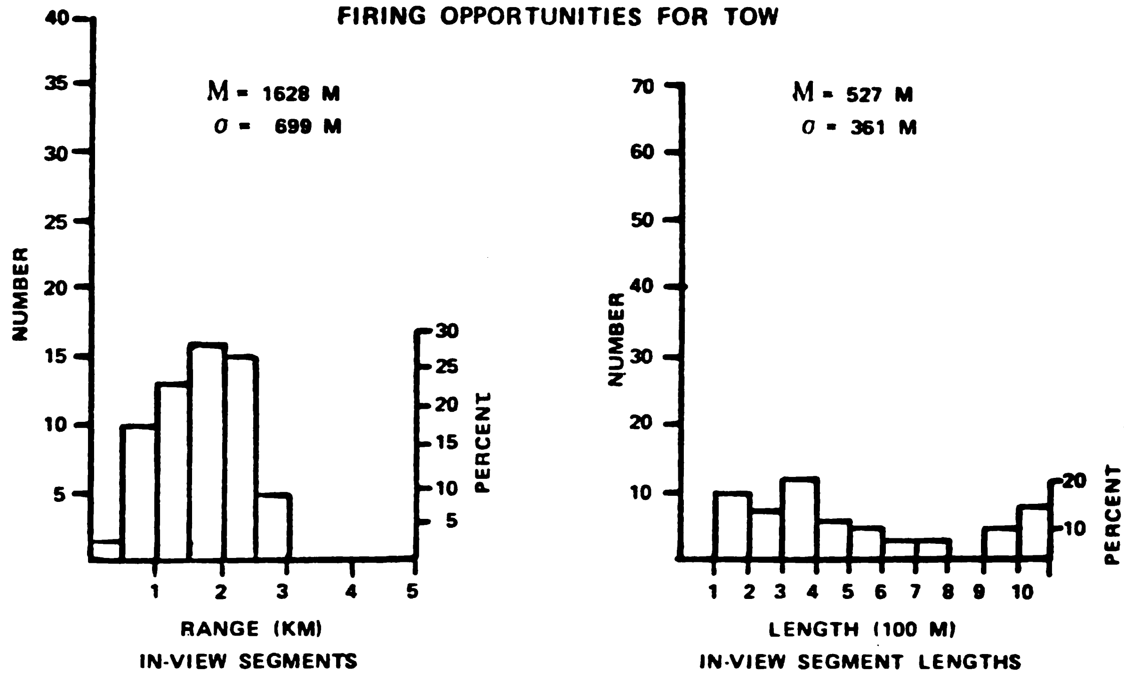
c. IFV/CFV Fighting Tactics.

Studies have been done which address the advantages of using pop-up tactics to limit exposure of the new class of fighting vehicles. It has been found that considerable natural defilade exists for use as fighting positions, that the pop-up tactic enhances survivability without a major degradation in target-servicing capability, and that based on the typical size of the killing zones, a four-missile launcher is desirable to minimize crew exposure and time away from position during reload (Ref. 8).

d. SOTAS/PAVE MOVER Intervisibility Modeling.

Questions concerning the level of detail required in air-to-ground intervisibility analyses in northern Europe and the need to provide intervisibility inputs to Division- and Corps-level wargames prompted a research effort which in part analyzed the level of correlated intervisibility from laterally separated aerial target acquisition positions (see Figure 26). It was concluded that in the NORTHAG area, medium-altitude aerial surveillance can be modeled reasonably well by selecting one general platform position which represents the position of the device within a Division area. Multiple positions separated from each other laterally by as much as 20km still exhibit LOS probabilities with correlation coefficients of 0.72 - 0.83, with a mean of 0.78. Reasonable surveillance is possible in this area using a mission

L 5122 SCENARIO NO. 11



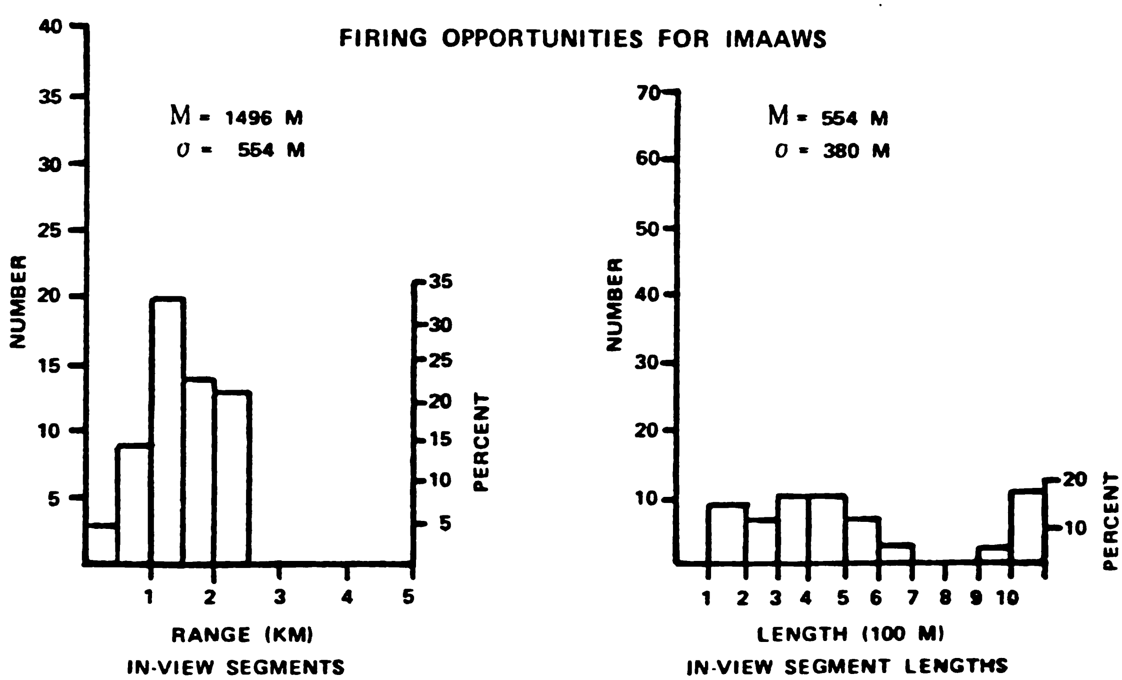
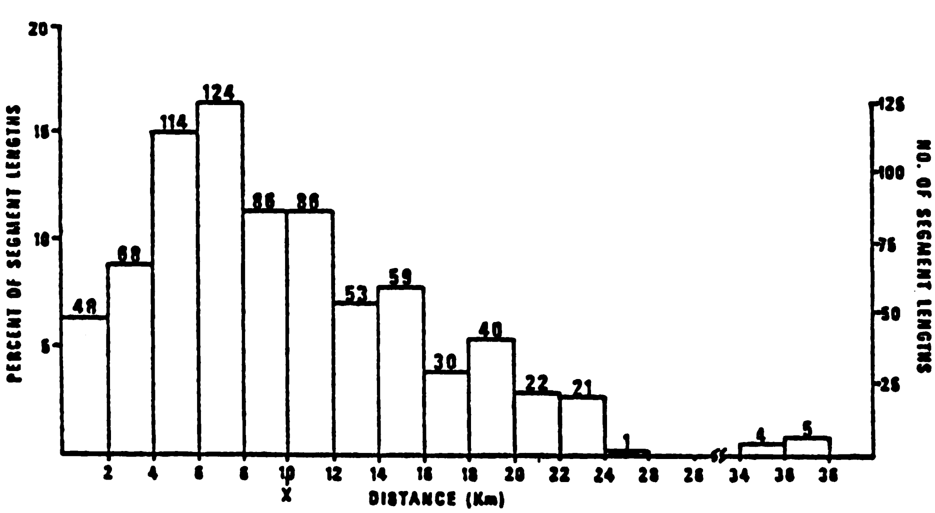


Figure 24 ATGM Firing Opportunities

altitude of approximately 1500m, AGL. It was found that target exposure is sensitive to the type of vehicle movement -- cross-country or on-road. High-altitude surveillance missions of the PAVE MOVER variety can be modeled reasonably well using the SOTAS results for look-down angles between 1.0 and 9.0 degrees. This study is now being extended into the CENTAG region, since it is expected that the rougher terrain in the south may change statistics and operating altitudes considerably (Ref. 9).

7. Terrain Classification Research. The above discussion captures the varied flavor of the terrain analyses often required to support US Army studies, and points to a need to generalize results whenever possible. Because of this requirement, an international working party was formed in 1978 to share information and then to seek ways of classifying terrain and its effects on military operations. The present effort emphasizes the causal relationships which drive the results of intervisibility and mobility analyses. In order to assist the development of a workable intervisibility classification system, TRASANA has embarked on a 30 man-month research program designed to investigate the relationships among terrain geometry, LOS statistics and predicted battle outcome. The TRASANA study (see Figure 27) will utilize digital topography, intervisibility computer programs, and the CARMONETTE battalion combat simulation in an attempt to establish intervisibility relationships within the context of military scenarios for selected regions of West Germany. This approach was developed based on experience gained in prior US studies (see Refs. 10-14). The goal of the study is to identify a classification scheme based on topography and surface clutter which can be used to predict intervisibility conditions for a wide range of military systems.

LENGTH OF TANK-TRAVERSED SEGMENTS X FREQUENCY OF LENGTH OCCURANCES



NUMBER OF SCENARIOS = 14
AVERAGE NUMBER OF MOVES PER SCENARIO = 54.4
AVERAGE NUMBER OF TANKS PER SCENARIO = 25
AVERAGE DISTANCE PER MOVE = 10.5 Km(S.d.=6.3)

AVERAGE MOVES TIMES:
10 Kph – 60 min
15 Kph – 40 min
20 Kph – 30 min

Figure 25 Distribution of Tank-Traversed Path Lengths

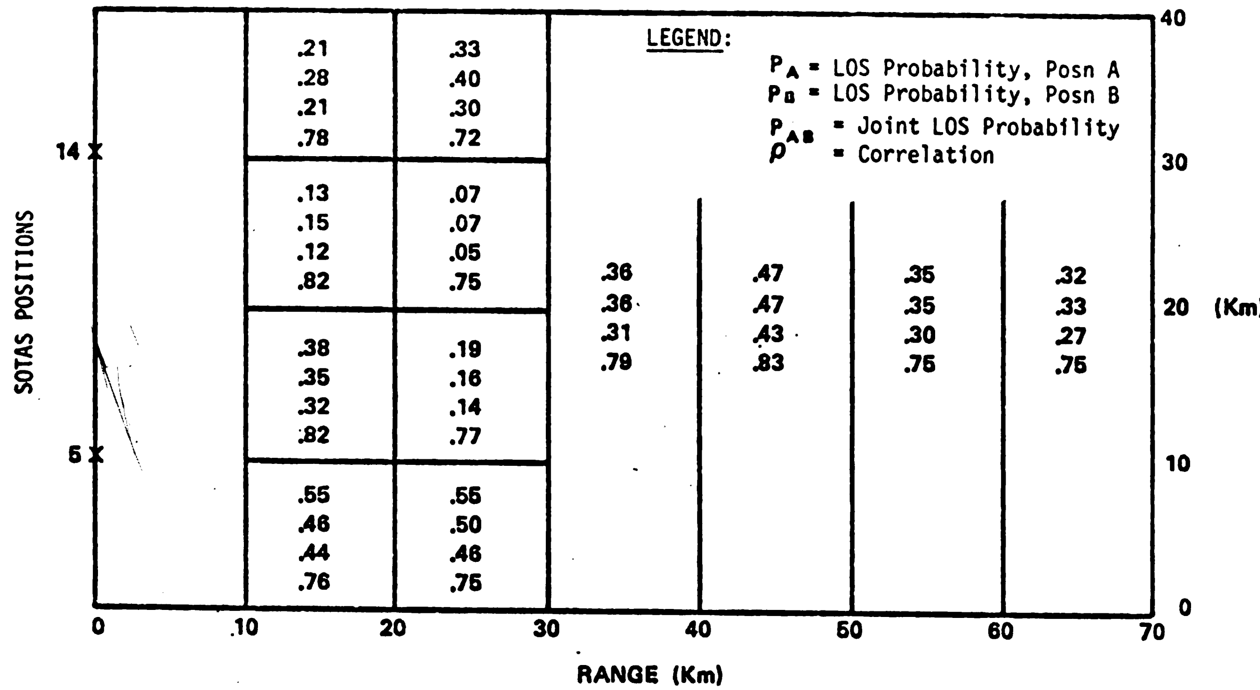


Figure 26 SOTAS/PAVE MOVER LOS Correlation Analysis

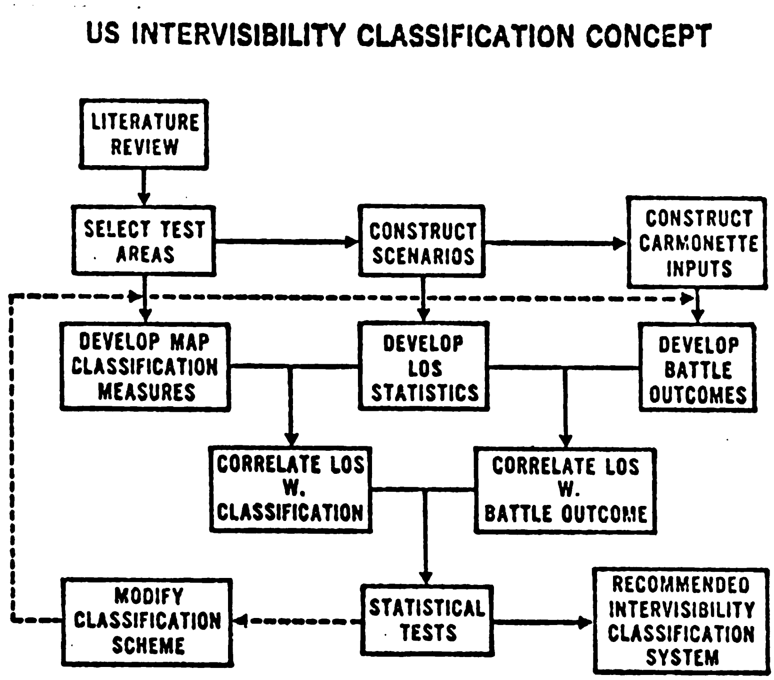


Figure 27 Intervisibility Classification Procedure

a. Test Areas.

Seventeen test map sheets have been selected based on the availability of appropriate digital data and the variability in surface roughness (See Figures 28-30). These regions can be grouped into ten test areas (Figure 31) using the 1974 Natick Landform Classification System, which together represent nearly 55 percent of the variability in surface roughness for West Germany. The test areas allow limited comparisons between two or more map sheets representing the same terrain "type". They also permit the comparison of statistics generated via computer models with those developed in certain field trials. Visualizations of the variability in surface roughness are shown schematically in Figure 32 and in a 3-D plot of the Neumarkt region in Figure 33. The codes for the Natick Landform Classification System (Ref. 11) are provided in Figure 34.

b. Intervisibility Classification System.

The Natick classification system depicts the typical surface roughness of a landform compartment (see Figure 30) by describing its maximum local relief, modal local relief, and number of positive features per mile along a random transect through the compartment. Together, the latter descriptors present a sinusoidal picture of the typical terrain profile in a compartment, modified periodically by larger hill masses ("outliers") as defined by the maximum local relief. For areas free of any surface clutter, there appears to be some positive correlation between landform type and intervisibility statistics. However, most intervisibility is further modified by vegetation and cultural features (urban areas) which appear on maps, but which are not described by the Natick system.

To remedy this difficulty, a tentative vegetation/urban classification system has been developed (see Figure 35) which is patterned after the structure of the Natick System. The TRASANA system for classifying this surface clutter uses three additional identifiers to describe vegetation and urban features in terms of their median height, median thickness, and median separation. The information required to classify a region in this manner is developed through computer processing of DMA digital topographic data. The computer programs allow the classification of vegetation and urban features either separately or jointly and provide statistics useful in the analysis process (see Figures 36-39). The map information summarized in Figure 40, plus LOS statistics will be used to test and modify the above classification system as necessary during the course of the study. A preliminary statistical design is provided in Appendix B.

c. Tactical Scenarios.

Since a major goal of the classification study is to develop a system which can predict military intervisibility conditions without large scale field trials or massive computer simulations, it is important that the LOS statistics to be used in the development of the classification system reflect

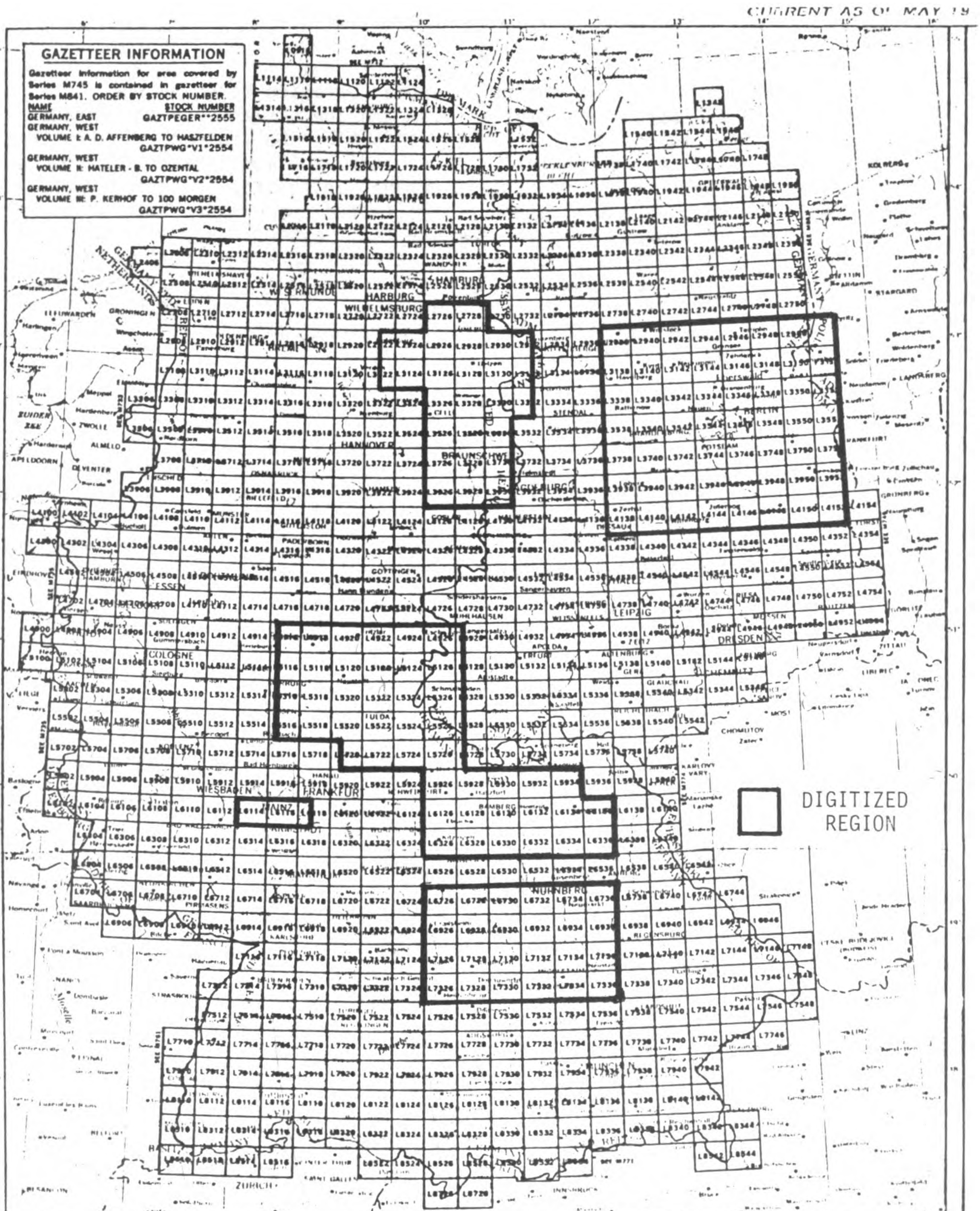


Figure 28 Special Digital Terrain Data - Germany

1.	L7128	NÖRDLINGEN
2.	L7130	TREUCHTLINGEN
3.	L3726	PEINE
4.	L3926	BAD SALZDETFURTH
5.	L2928	BAD BEVENSEN
6.	L2930	DAHLENBURG
7.	L2924	SCHNEVERDINGEN
8.	L6336	ESCHENBACH
9.	L5122	NEUKIRCHEN
10.	L6734	NEUMARKT
11.	L6736	VELBURG
12.	L6936	PARSBERG
13.	L5320	ALSFELD
14.	L5324	HÜNFELD
15.	L5928	HASSFURT
16.	L5524	FULDA
17.	L5526	MELLRICHSTADT

Figure 29 Terrain Classification Map Sheets

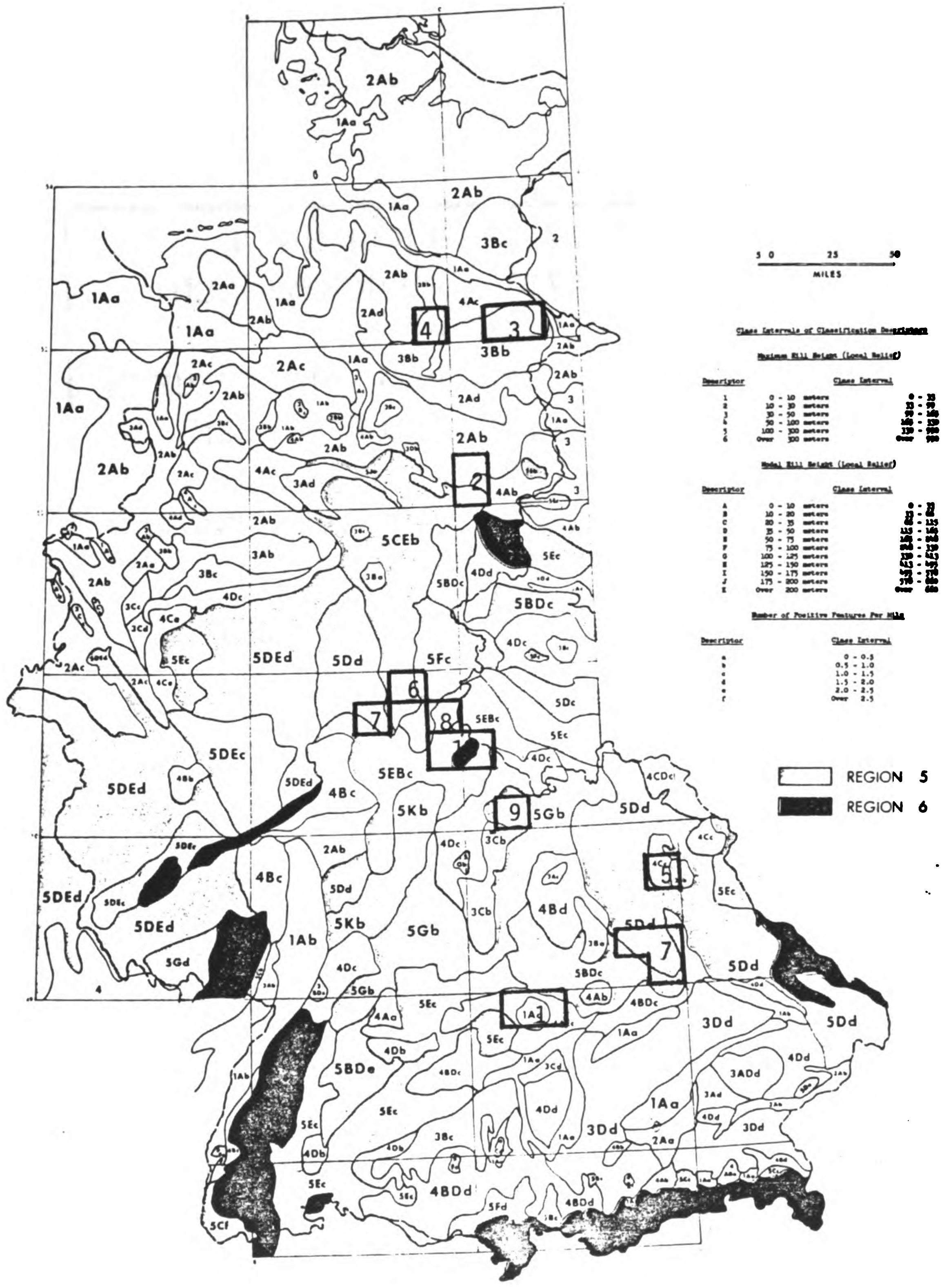


Figure 30 Natick Landform Test Areas

SCENARIO	MAP	TERRAIN TYPE	FRG %
1. NORDLINGEN	L7128, L7130	1Aa	8
2. BRAUNSCHWEIG (CHINESE EYE)	L3726, L3926	2Ab	18
3. UELZEN (KINGS RIDE)	L2928, L2930	3Bb	2
4. SCHNEVERDINGEN (TETAM; 3B, 4B)	L2924	3Bb	2
5. GRAFENWOHR (NATO)	L6336	4Cc	1
6. ALSFELD (TETAM 4F)	L5122	5CEb	5
7. HOHENFELS (TETAM, NATO)	L6734, L6736 L5320, L6936	5Dd	14
8. HUNFELD	L5324	5EBc	2
9. SCHWEINFURT	L5928	5Gb	3
10. GERSFELD (WILDFLECKEN)	L5524, L5526	6Ka	0.1
		TOTAL	55.1

Figure 31 Proposed Terrain Classification System Test Areas

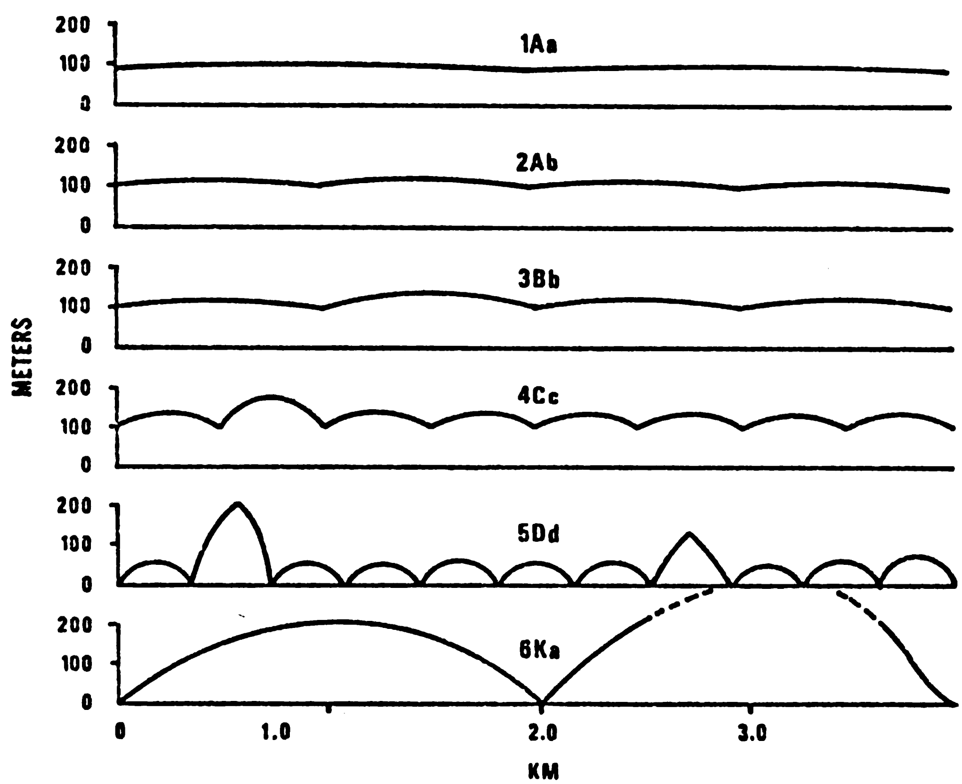


Figure 32 Sample Landforms (Natick Classification)

NEUMARKT
(L6734)

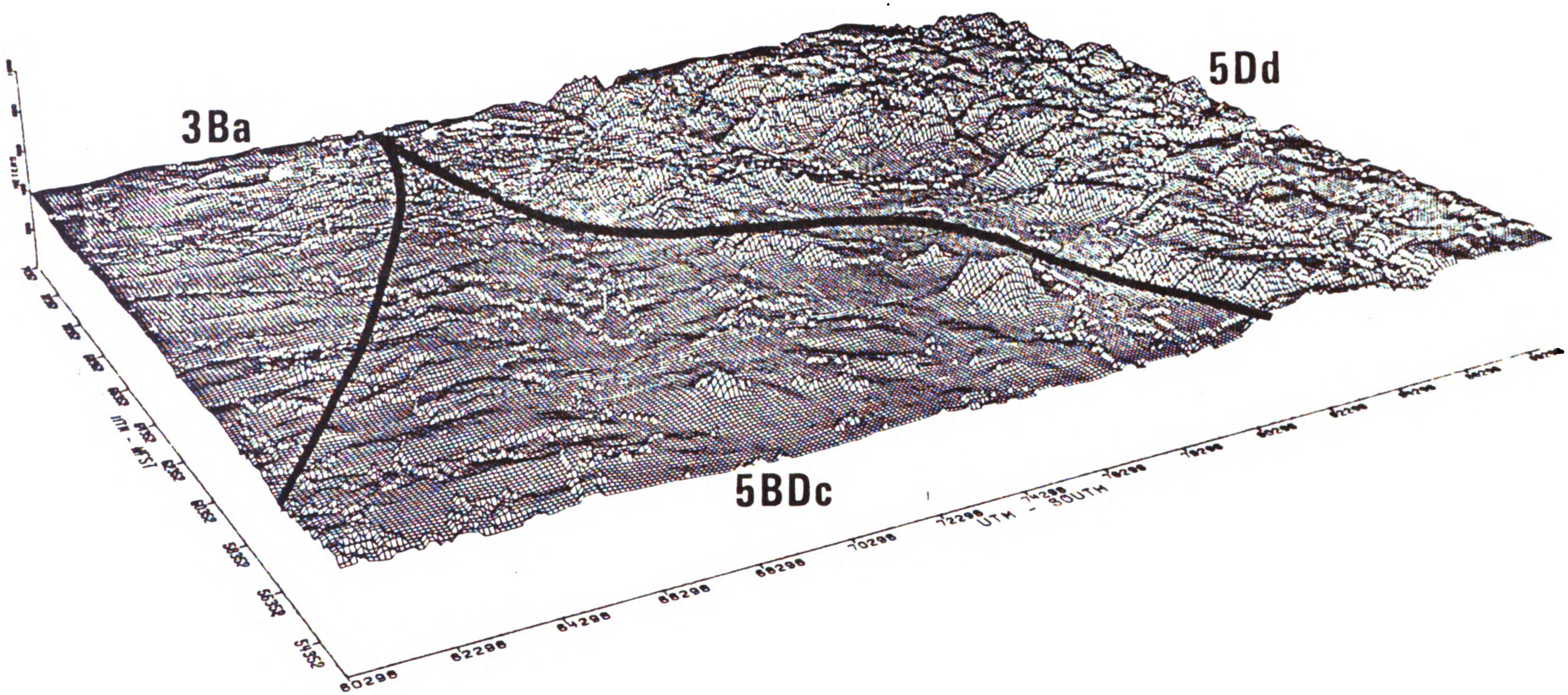


Figure 33 Neumarkt Region, showing Landform Compartments

MAXIMUM HILL HEIGHT (LOCAL RELIEF)

DESCRIPTOR	CLASS INTERVAL	
1	0 - 10 METERS	0 - 33 FEET
2	10 - 30	33 - 99
3	30 - 50	99 - 165
4	50 - 100	165 - 330
5	100 - 300	330 - 990
6	OVER 300	OVER 990

MODEL HILL HEIGHT (LOCAL RELIEF)

DESCRIPTOR	CLASS INTERVAL	
A	0 - 10 METERS	0 - 33 FEET
B	10 - 20	33 - 66
C	20 - 35	66 - 115
D	35 - 50	115 - 165
E	50 - 75	165 - 248
F	75 - 100	248 - 330
G	100 - 125	330 - 413
H	125 - 150	413 - 495
I	150 - 175	495 - 578
J	175 - 200	578 - 660
K	OVER 200	OVER 660

NUMBER OF POSITIVE FEATURES PER MILE

DESCRIPTOR	CLASS INTERVAL	
a	0 - 0.8 KILOMETERS	0 - 0.5 MILES
b	0.8 - 1.6	0.5 - 1.0
c	1.6 - 2.4	1.0 - 1.5
d	2.4 - 3.2	1.5 - 2.0
e	3.2 - 4.0	2.0 - 2.5
f	OVER 4.0	OVER 2.5

Figure 34 Natick Landform Classification Descriptors

VEGETATION/URBAN CLASSIFICATION DESCRIPTORS

MEDIAN HEIGHT

DESCRIPTOR	CLASS INTERVAL	
1	0 – 2.5 METERS	0 – 8.3 FEET
2	2.5 – 5	8.3 – 16.5
3	5 – 10	16.5 – 33.0
4	10 – 15	33.0 – 49.5
5	15 – 20	49.5 – 66.0
6	OVER 20	OVER 66.0

MEDIAN THICKNESS

DESCRIPTOR	CLASS INTERVAL	
A	0 – 0.2 KM	0 – 0.1 MI
B	0.2 – 0.4	0.1 – 0.3
C	0.4 – 0.6	0.3 – 0.4
D	0.6 – 0.8	0.4 – 0.5
E	0.8 – 1.0	0.5 – 0.6
F	1.0 – 2.0	0.6 – 1.3
G	2.0 – 3.0	1.3 – 1.9
H	3.0 – 5.0	1.9 – 3.1
I	OVER 5.0	OVER 3.1

MEDIAN SEPARATION

DESCRIPTOR	CLASS INTERVAL	
a	0 – 0.2 KM	0 – 0.1 MI
b	0.2 – 0.5	0.1 – 0.3
c	0.5 – 1.0	0.3 – 0.6
d	1.0 – 1.5	0.6 – 0.9
e	1.5 – 2.0	0.9 – 1.3
f	2.0 – 3.0	1.3 – 1.9
g	3.0 – 4.0	1.9 – 2.5
h	4.0 – 5.0	2.5 – 3.1
i	OVER 5.0	OVER 3.1

Figure 35 Class Intervals — Vegetation/Urban Descriptors

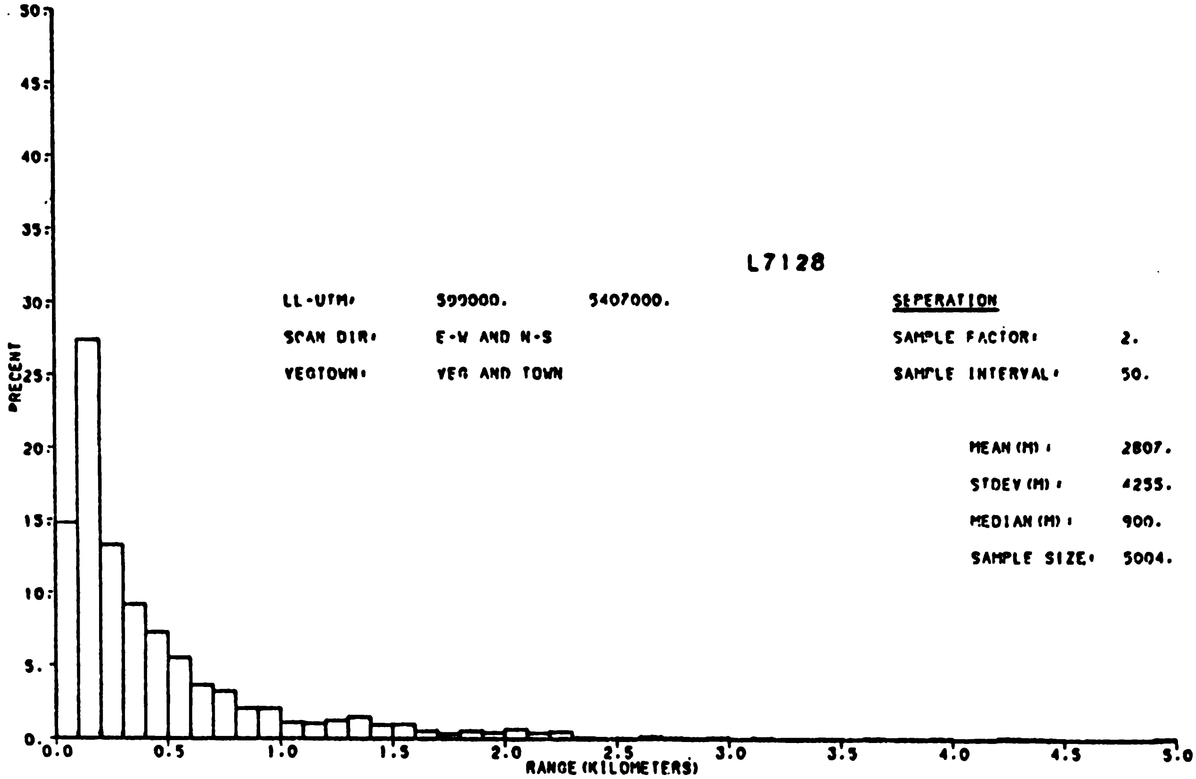


Figure 36 Distribution of Vegetation/Urban Cell Thicknesses

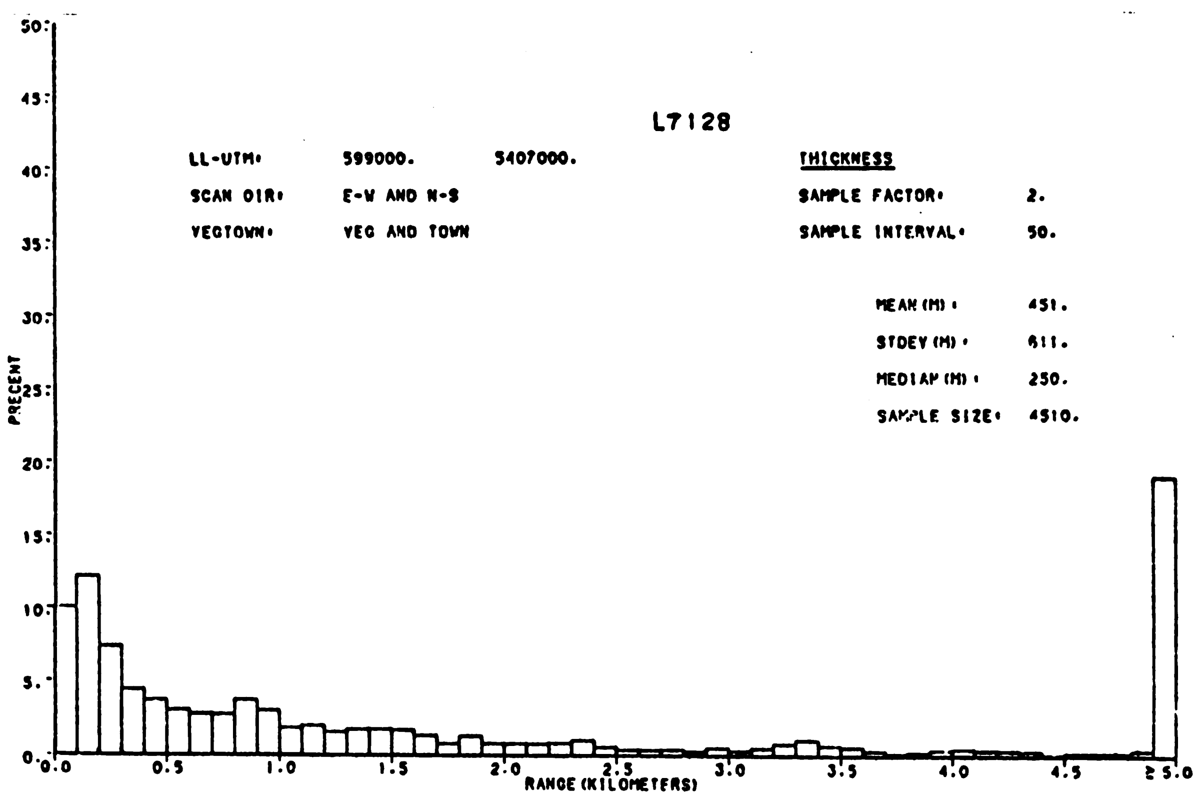


Figure 37 Distribution of Vegetation/Urban Cell Separation Distances

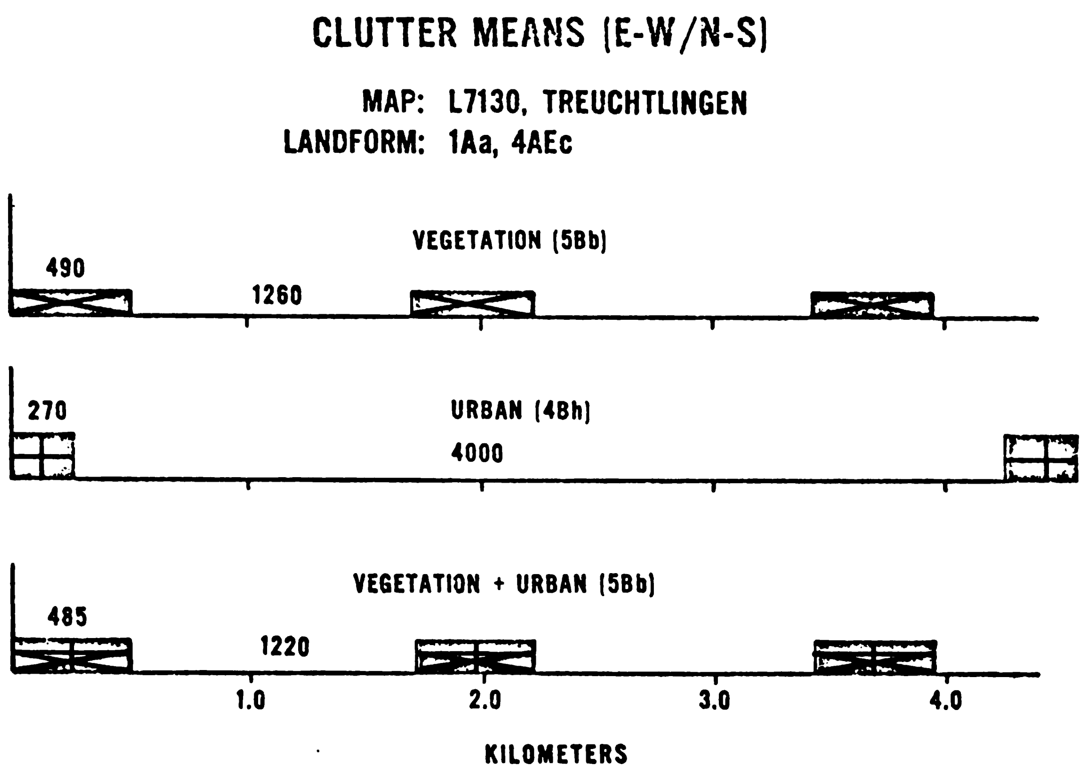


Figure 38 Sample Vegetation/Urban Distribution (Treuchtlingen)

	ELEVATION			LAND FORM CLASS		VEGETATION		URBAN		VEGETATION/URBAN		
MAPSHEET	MIN	MAX	DIF	OVERLAY (NATICK)	COMPUTER	CLASS	%	CLASS	%	CLASS	%	PATCHES
L7128	406	663	257	1 Aa	5CAa	2 Bc	11.3	1 B1	1.3	2 Bc	12.6	449/56
L7130				1 Aa								
L3726	54	110	56	2 Ab	4 Aa	1 Bg	6.7	5 Bf	8.1	2 Bd	14.6	214/389
L3926	80	315	235	2 Ab	5 CBb	3 De	25.0	3 Bf	4.3	3 Bd	29.3	152/518
L2928	13	108	95	3 Bb	4 Bb	4 Bc	33.0	2 Bg	3.9	4 Bb	36.9	1048/432
L2930	9	140	131	3 Bb	5 Bc	6 Bb	54.2	2 Bg	2.7	6 Bb	56.9	599/392
L2924	41	168	127	3 Bb	5 Bc	4 Bb	34.5	2 Ag	2.7	4 Bb	37.2	1042/708
L6336	394	560	146	4 Cc	5 Cb	5 Bb	42.5	2 Af	2.3	5 Bb	44.8	1545/644
L5122	208	635	427	5 CEb	6 EFc	6 Cb	51.2	2 A1	2.0	6 Cb	53.2	214/389
L6734	370	635	265	5 Dd	5 Ec	4 Bb	35.4	1 C1	1.6	4 Bb	37.0	1969/22
L6736	346	630	284	5 Dd	5 Cb	6 Bb	52.9	1 C1	0.2	6 Bb	53.1	1181/10
L5320	218	570	352	5 Dd	6 Eb	4 Cb	34.1	2 Bh	2.9	4 Cc	37.0	534/380
L6936	334	587	253	5 Dd	5 Cb	4 Bb	38.7	1 B1	0.4	4 Bb	39.1	2114/37
L5324	218	702	484	5 ECc	6 Eb	3 Bc	26.5	2 Bh	3.3	4 Bc	39.8	655/230
L5926	128	510	382	5 Gb	6 Ec	4 Bb	39.2	2 Ag	3.5	5 Bb	42.7	1308/551
L5524	269	947	678	6 Ka	6 EFGKb	4 Bc	32.5	2 Af	3.9	4 Bc	36.4	783/753
L5526	251	900	649	6 Ka	6 Eb	4 Bb	37.6	2 B1	2.4	5 Bb	40.0	924/312

Figure 39 Summary Map Sheet Analysis

MAP STATISTICS

1. TOPOGRAPHY
   - HIGH POINT (ELEVATION, COORDINATES)
   - LOW POINT (ELEVATION, COORDINATES)
   - MAP RELIEF (HIGH – LOW)

2. LANDFORMS
   - MAXIMUM HILL HEIGHT
   - MODAL HILL HEIGHT
   - POSITIVE FEATURES

3. VEGETATION
   - PERCENT VEGETATION
   - NUMBER OF PATCHES
   - MINIMUM AREA, DIMENSIONS
   - MAXIMUM AREA, DIMENSIONS
   - DIMENSION DISTRIBUTION
   - HEIGHT DISTRIBUTION

4. URBAN AREAS
   - PERCENT URBAN
   - NUMBER OF PATCHES
   - MINIMUM AREA, DIMENSIONS
   - MAXIMUM AREA, DIMENSIONS
   - DIMENSION DISTRIBUTION
   - SPACING DISTRIBUTION
   - HEIGHT DISTRIBUTION

Figure 40 Map Statistics

a valid military setting. To insure this, detailed tactical scenarios and overlays are being prepared for each test area. An example is provided in Figures 41-45 for the Peine Map Sheet (L3726). This scenario portrays a BLUE covering force area (CFA) and main battle area (MBA) battle (Figures 42 and 43). Only the last positions of the units in the CFA are included. The scenario consists of one cavalry troop in the CFA and one tank battalion task force in the MBA. In the MBA, the battalion is deployed with two company teams forward with one company positioned on the commanding terrain in the center. A total of 6 tank platoons, 3 mech platoons and 3 cavalry platoons are available to the force. After completing the covering force mission, the cavalry troop is attached to the tank battalion and is attached by platoon to the companies in the MBA.

The RED attack (Figures 44 and 45) consists of a reinforced motorized rifle regiment attacking on two axes. The attack uses three battalions abreast with the fourth battalion being committed in the center after seizure of the initial objectives. The RED plan depicts the route of advance of each battalion and the formation of these units as they proceed in the attack.

The overlay originals are color-coded and keyed with the respective legends to allow the considerations of all likely vehicle attack paths and occupied defensive positions during the intervisibility analysis.

In this analysis, it is desired that all statistics gathered reflect information from tactically realistic scenarios. In order to do this, the movement of an attacking threat force must be simulated. That is to say, once a given attacker route is specified, the movement and location of individual vehicles within a given formation is needed. This is referred to as the "multiple route" problem. Digitization of individual routes was tried for one scenario. This is not a feasible solution to the problem for several reasons. First, the volume of data is enormous. Second, movement and phasing of individual vehicles within a formation is still a problem. Therefore, a computer algorithm was generated which gives the time/position history for each vehicle along a given attacker route. Basically, the routine generates the position for a lead vehicle in a formation and maintains the integrity of that formation by having all other vehicles follow a course parallel to the attacker route at constant trailing distance.

d. CARMONETTE Analysis.

A subset of the available tactical scenarios will be prepared for insertion in the CARMONETTE battalion-level Monte Carlo combat simulation. The intent of this effort is to develop time phased engagement results which can be compared with LOS statistics to examine the effects of intervisibility on battle outcome. Although this is an interesting research topic in its own right (see Ref. 12), the main objective of this effort will be to gain insight concerning multiple lines-of-sight in a realistic m-on-n target engagement environment. These results will be useful in the further examination of LOS "coherence" phenomena (occasions in which LOS correlations exceed 0.7).

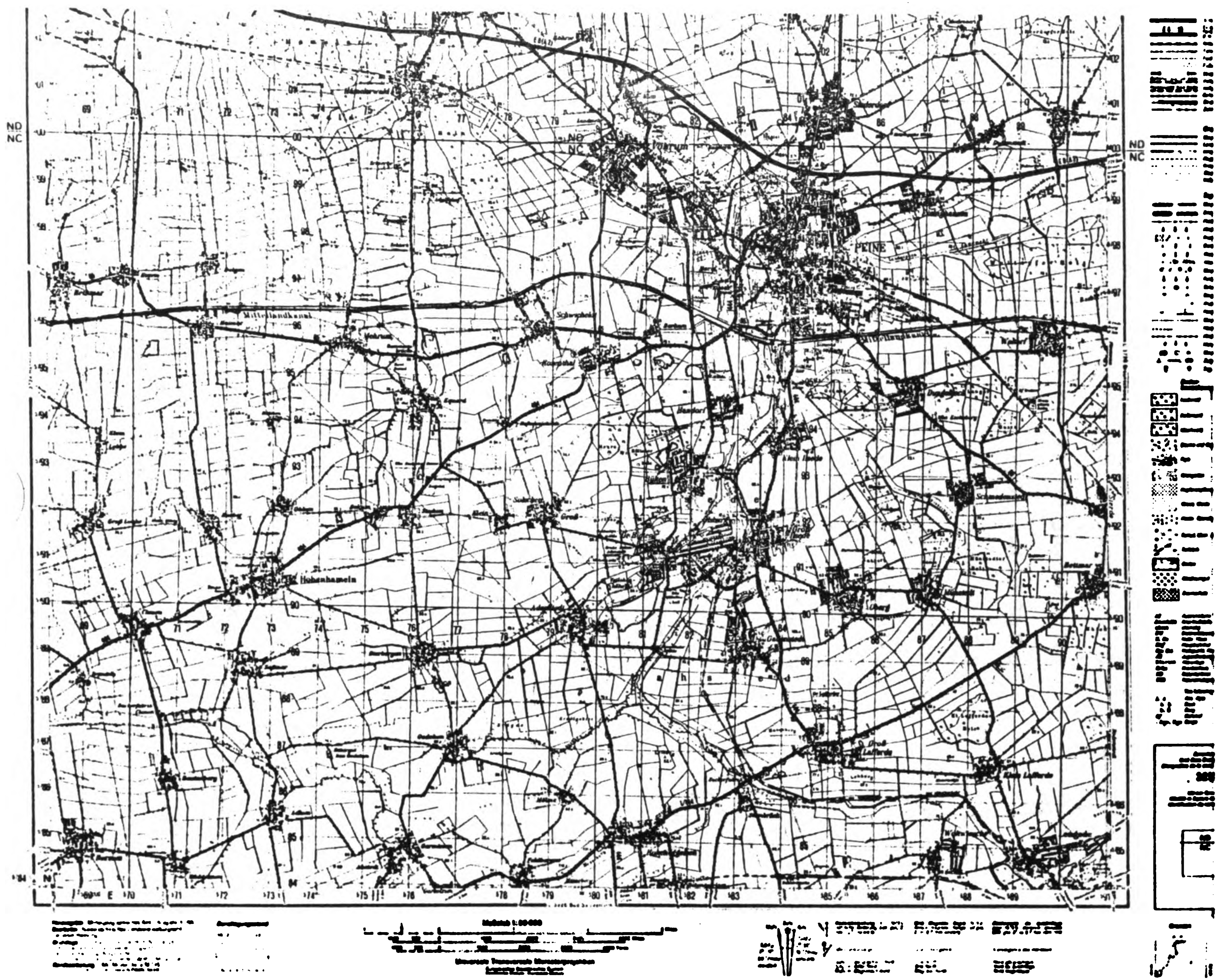


Figure 41 Peine Map Sheet (L3726)

LEGEND

	WITHDRAWAL ROUTE, 3rd PL, CAV TROOP
	WITHDRAWAL ROUTE, 2nd PL, CAV TROOP
	WITHDRAWAL ROUTE, MECH PLT, A CO
	WITHDRAWAL ROUTE, 1st PLT, CAV TROOP
	WITHDRAWAL ROUTE, 1st TK PLT, A CO
	WITHDRAWAL ROUTE, 2nd TK PLT, A CO
	WITHDRAWAL ROUTE, 2nd TK PLT, B CO
	WITHDRAWAL ROUTE, 1st TK PLT, B CO
	WITHDRAWAL ROUTE, MECH PLT, B CO
	WITHDRAWAL ROUTE, 1st TK PLT, C CO
	WITHDRAWAL ROUTE, 2nd TK PLT, C CO
	WITHDRAWAL ROUTE, MECH PLT, C CO
	PLT BATTLE POSITION
	SUBSEQUENT PLT BP (UNOCCUPIED AND PREPARED)
H	HELICOPTER UNIT
GEMSS .004	MINEFIELD (GEMSS LAYED-DENSITY .004 MINES/M^2
PP 1	PASSAGE POINT
●	A GUN POSITION WITHIN PLATOON BATTLE POSITION

NOTES

1. ONE DS BATTALION IN SUPPORT (155M)
2. ONE BATTALION 8"/MLRS GSR

Figure 42 BLUE Force Legend (Peine)

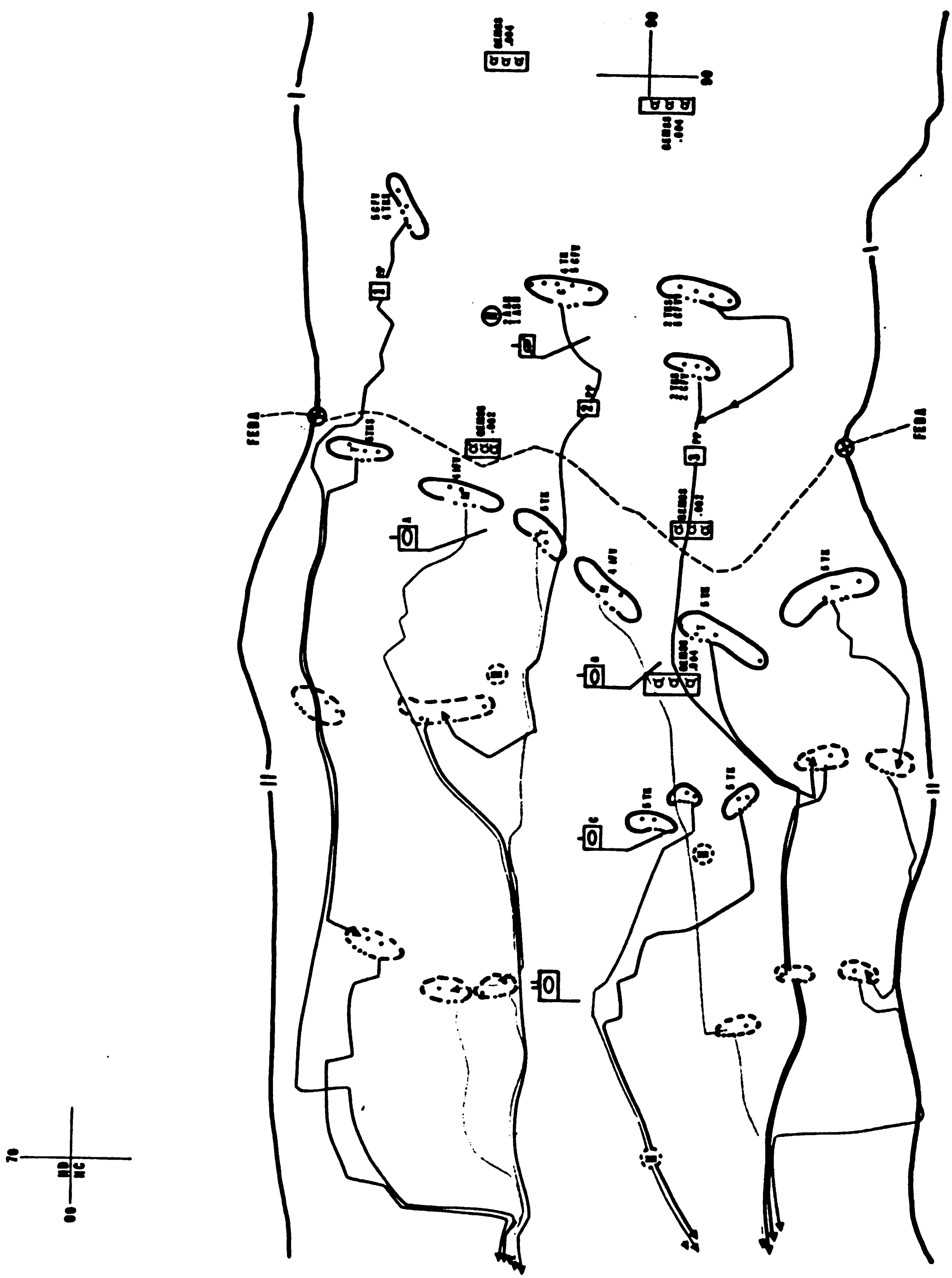


Figure 43 BLUE Overlay (Peine)

LEGEND

COLUMN FORMATION MOTORIZED RIFLE BN
4 TKS, 10 BMPS, 4 TKS, 10 BMPS, 2 ZU 23-4,
4 TKS, 10 BMPS, 2 BTR-50, 2 SA-8
(50 M BETWEEN VEHICLES)

COMPANIES ON LINE - PLTS IN COLUMN

```
TTTT      TTTT
 BBB       BBB
 BBB       BBB
 BBB       BBB
                  (50 M BETWEEN VEHICLES;
TTTT               100 M BETWEEN COMPANIES)
 BBB
 BBB
 BBB
```

COMPANIES IN ASSAULT LINE - PLTS IN COLUMN

```
TTTT      TTTT      TTTT
 BBB       BBB      .BBB     (50 M BETWEEN VEHICLES;
 BBB       BBB       BBB      100 M BETWEEN COMPANIES)
 BBB       BBB       BBB
```

ASSAULT LINE - TANKS LEADING INFANTRY

```
  T   T   T   T       (25 M BETWEEN BMPS; 100 M BETWEEN TANKS;
B B B B B B B B B               50 M BETWEEN LINES)
        B
```

COMPANY COLUMN
(4 TKS, 10 BMPS) - 50 M BETWEEN VEHICLES

2d ECHELON TANK BATTALION IN COLUMN
40 TANKS (100 M BETWEEN VEHICLES)
MOVE 5 KM BEHIND LEADING BATTALIONS
COMMUTED PAST OBJ URAL AFTER SEIZURE

TANK COMPANIES IN ASSAULT LINE - PLTS IN COLUMN

```
T T T   T T T   T T T
T T T   T T T   T T T     (50 M BETWEEN VEHICLES;
T T T   T T T   T T T      100 M BETWEEN COMPANIES)
T T T   T T T   T T T
  T       T       T
```

H HELICOPTER UNIT

RECON COMPANY (REGT) IN COLUMN
4 PT 76, 3 BMP, 2 BTR

OVERWATCH VEHICLES

Figure 44 RED Force Legend (Peine)

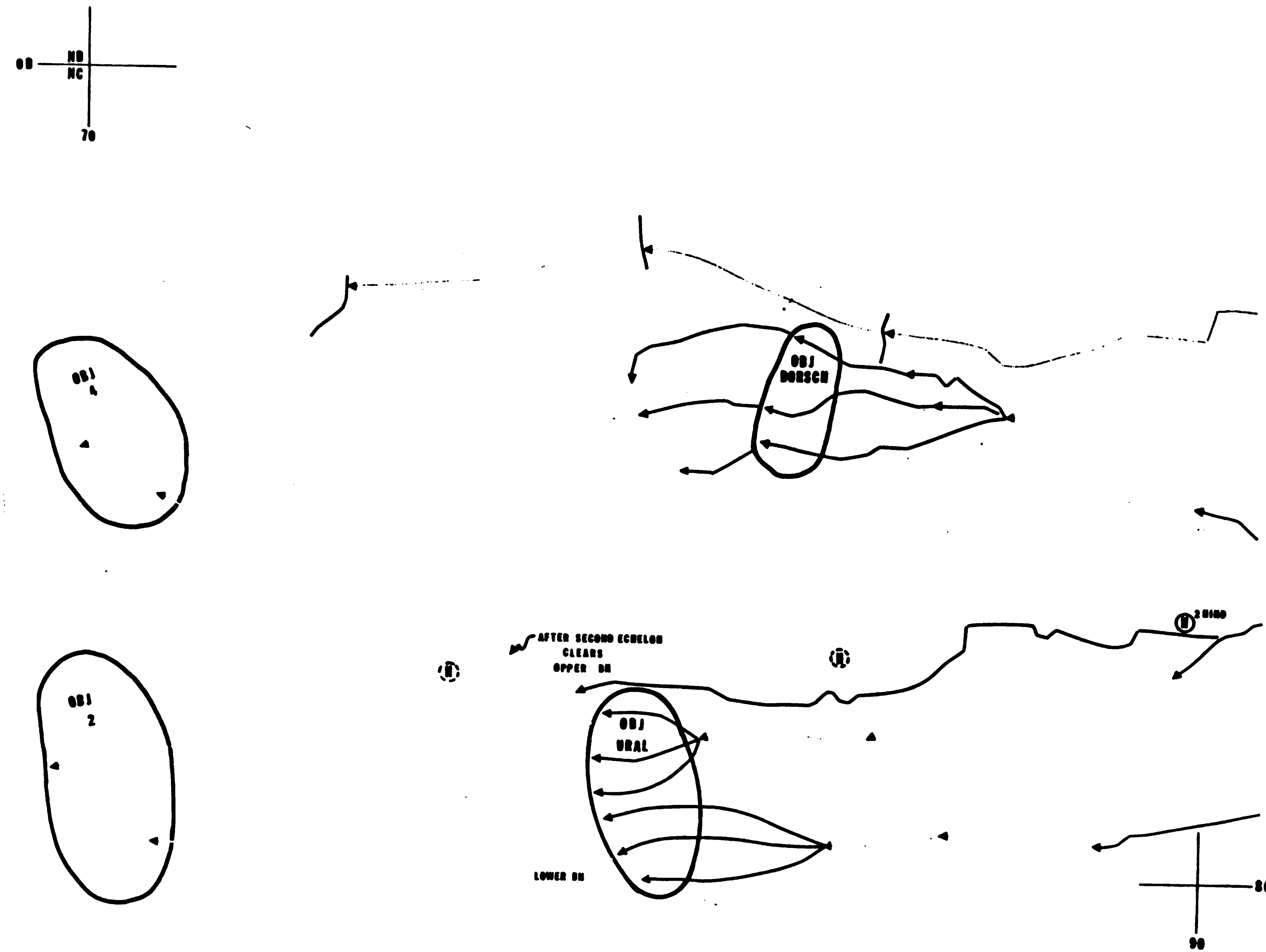


Figure 45 RED Overlay (Peine)

e. Tactical LOS Measures.

During the study several tactical intervisibility measures will be developed for use in the classification effort, as well as in extending the knowledge of battlefield characteristics (See Figure 46). Some measures (Figure 47) are classical in nature and have been used in a variety of past trials and studies. Others, especially ALOS/DLOS and coherence, are relatively recent in usage and give promise of being highly correlated with predicted battle outcome (see Figures 48-52). Additional statistics suggested by the CHINESE EYE and KINGS RIDE field trials or this analysis will also be examined.

f. Classification Work Outline.

As of 1 October 1981, the intervisibility classification work outlined is approximately 50 percent completed. Figure 53 indicates the major tasks and shows partially (/) or fully completed (X) work for each of the ten test areas. It is anticipated that the remaining effort will be completed and a recommended interim intervisibility classification system will be available in Mid-1982.

g. Mobility Classification.

A parallel international effort concerning the classification of terrain with regard to tactical mobility is also in progress. This effort will use a screening process to arrive at a group of map sheets for detailed analysis (perhaps utilizing some of the same scenarios generated for the intervisibility work). From this work, a library of digitized mobility data will be developed which represents specific map sheets in great detail, and which will be related through factor overlays to broader regions, for inferential purposes. The results of the mobility classification research are expected in 1983.

h. Summary.

Although the effort outlined above involves a difficult problem whose solution has evaded researchers for several decades, TRASANA believes that the scope of the current work program is broad enough and the tools are now available to permit a major advancement of the understanding of intervisibility in a tactical environment. If the work remains uninterrupted, a breakthrough in the ability to classify terrain and use that classification system for predictive purposes (see Figure 54) is anticipated. Such a system would have multiple uses within the military operations research community.

TACTICAL LOS MEASURES

- PROBABILITY VS RANGE
- IN-VIEW DISTRIBUTION
- OUT-OF-VIEW DISTRIBUTION
- OPENING RANGE DISTRIBUTION
- ALOS/DLOS
- CORRELATION
- COHERENCE

Figure 46 Tactical Line-of-Sight Measures

L 3726 SCENARIO NO. 7

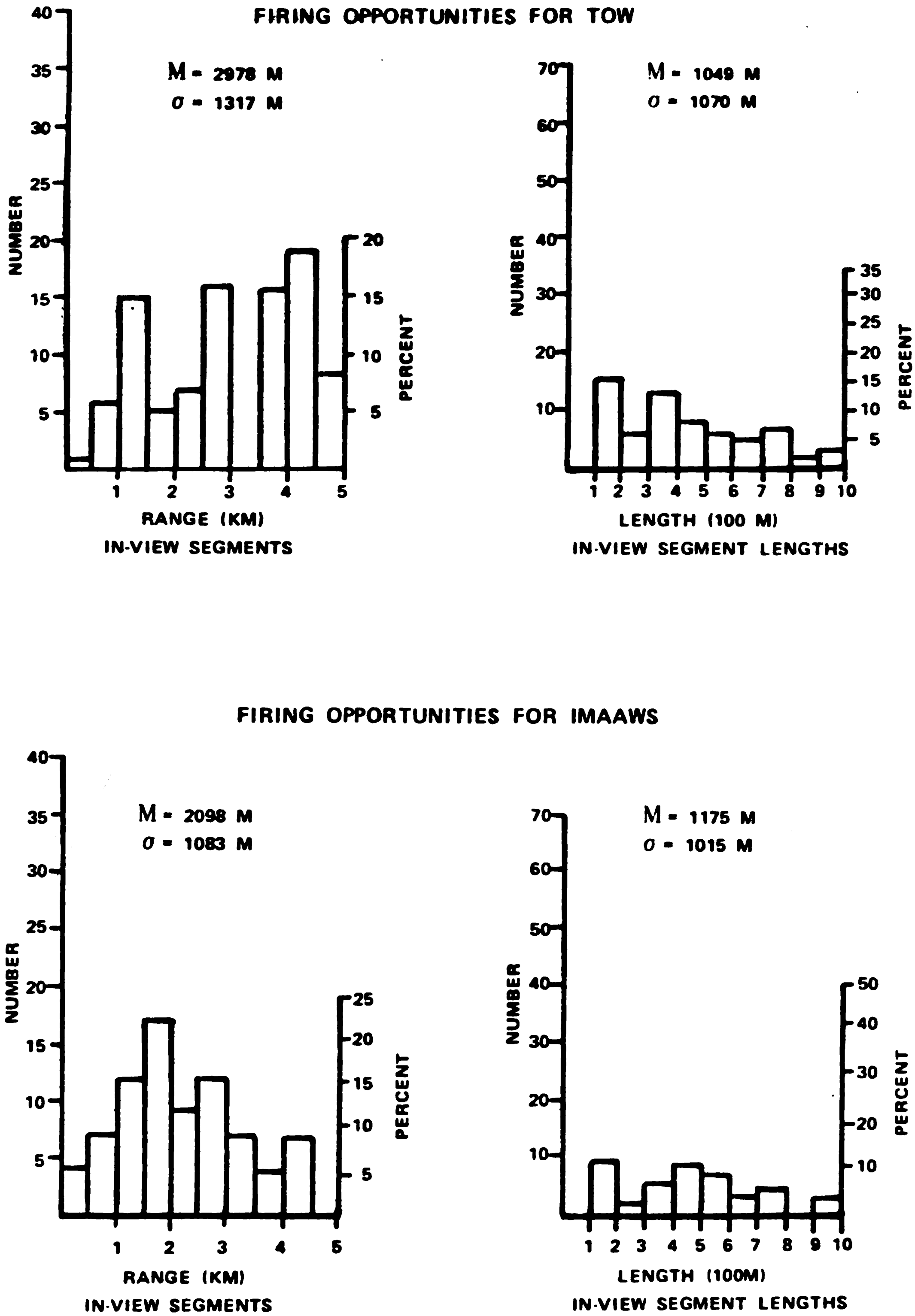


Figure 49 ATGM Firing Opportunities

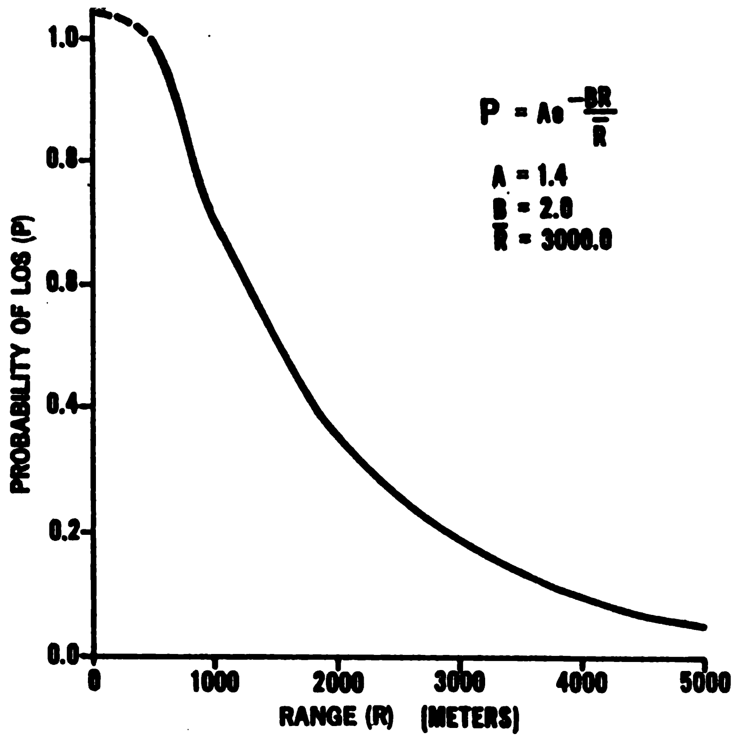


Figure 47 Functional Curve Fit — LOS Probability

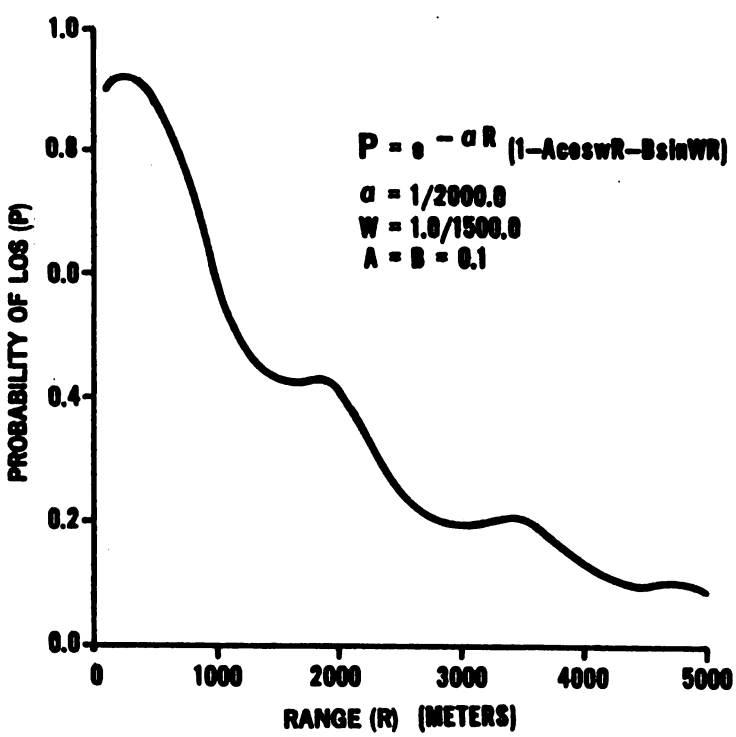


Figure 48 Functional Curve Fit — LOS Probability

FARREL TERRAIN/TACTICS RESEARCH
(VRI DUSA-1 FR77-1)

PREMISE: ANY CLASSIFICATION OF TERRAIN BY EXPOSURE CHARACTERISTICS IS USEFUL ONLY TO THE EXTENT THAT IT PROVIDES INFORMATION REGARDING THE POTENTIAL OUTCOMES (PDFWIN) OF COMBAT ON ALL TERRAINS IN A CLASS.

CONCLUSIONS:

1. COMBAT RESULTS NOT CORRELATED WITH AVERAGE:

   A. EXPOSURE LENGTH
   B. PERCENT OF PATH EXPOSED
   C. NUMBER OF SIMULTANEOUS EXPOSURES
   D. LAST COVERED RANGE (90%)
   E. OPENING RANGE
   F. PLOS

2. PDFWIN IS STRONGLY CORRELATED WITH:

   A. ALOS(N)/DLOS(N)
   B. ATTACKER EXPOSURE COORDINATION

3. WEAPON SYSTEM STUDIES DON'T NEED MULTIPLE SCENARIO TEAMS WORKING ON SAME TERRAIN; DO NEED MULTIPLE SCENARIOS/TERRAINS.

Figure 50 Farrell Terrain/Tactics Study Conclusions

ALOS(N_0) = MEAN NUMBER OF SECONDS (IN RANGE BAND) IN WHICH AN ATTACKER WEAPON SYSTEM WOULD HAVE N_0 OR MORE EXPOSED TARGETS AND WOULD BE TACTICALLY PERMITTED TO FIRE.

DLOS(N_0) = MEAN NUMBER OF SECONDS (IN RANGE BAND) IN WHICH A DEFENDER WEAPON WOULD HAVE N_0 OR MORE EXPOSED TARGETS AND WOULD BE TACTICALLY PERMITTED TO FIRE.

$$\frac{ALOS}{DLOS} = \frac{(1-H)\ MP_1 T + F}{(1-H)\ (1-(1-P_2)^C)T + F}$$

WHERE

H = DEGREE TO WHICH ATTACKER USES COVERED ROUTES

P_1 = FRACTION OF OPENING RANGE BAND IN WHICH A RANDOM POINT IN THE AREA (1KM DEEP AND AS WIDE AS THE DEFENDER FRONT) IS WITHIN TACTICAL RANGE OF AND HAS LINE-OF-SIGHT TO AT LEAST 10% OF THE DEFENDERS

M = FRACTION OF TIME ATTACKER IS ABLE TO SPEND FIRING WHILE ADVANCING

T = TIME REQUIRED FOR ATTACKER TO ADVANCE 1KM

F = AVERAGE AMOUNT OF TIME ATTACKER SPENDS STOPPED IN FIRING POSITION

P_2 = AVERAGE FRACTION OF THE OPENING RANGE BAND AREA VISIBLE TO RANDOM DEFENDER

C = ATTACKER'S DEGREE OF EXPOSURE COORDINATION, A NUMBER BETWEEN 1 AND THE NUMBER OF ATTACKERS REPRESENTING THE AVERAGE NUMBER OF <u>INDEPENDENT</u> TARGETS PROVIDED BY THE ATTACKER

Figure 51 ALOS/DLOS Definition

COHERENCE/CORRELATION

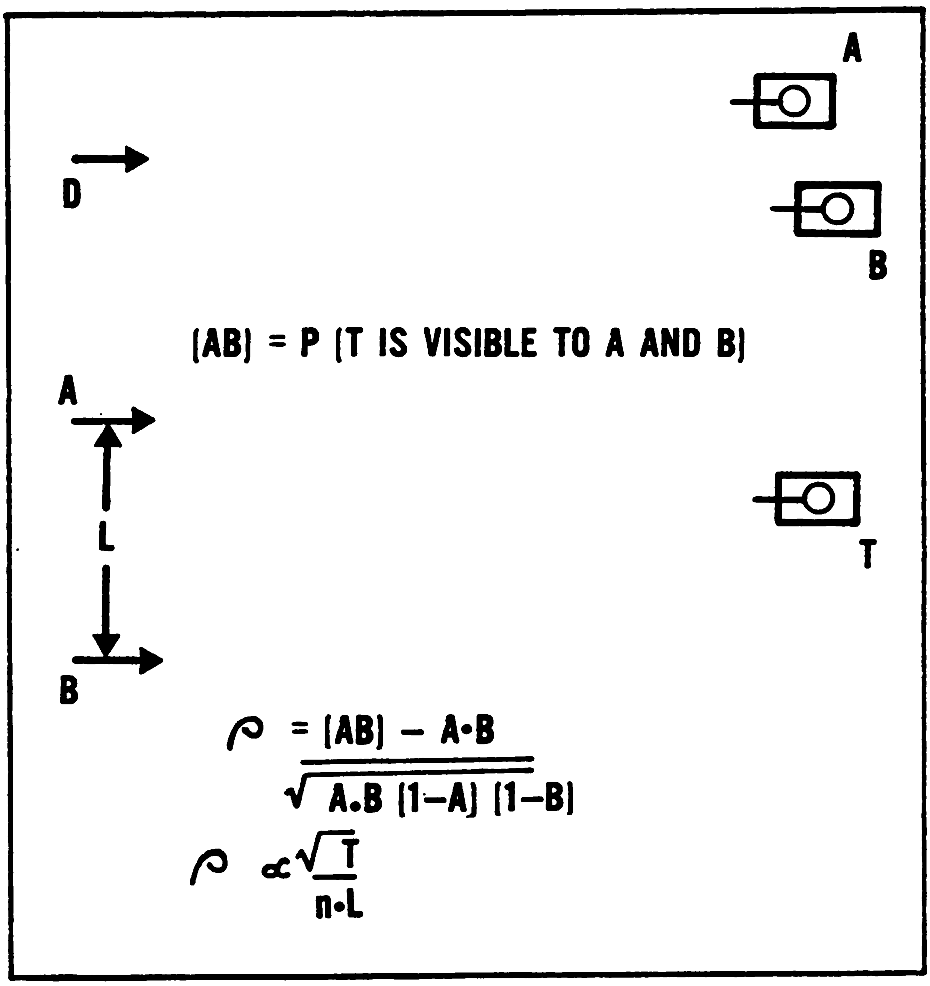


Figure 52 LOS Correlation Calculation

	1	2	3	4	5	6	7	8	9	10
DIGITIZE SCENARIOS	X	X	X	X	X	X	X	X	X	X
CHECK LOS	X	X	X	X	X	X	X	X	X	
DEVELOP 50-METER GRID FILES	X	X	X	X	X	X	X	X	X	X
GENERATE MULTIPLE PATHS	X	X	X	X		/	/			
MODIFY SCENARIOS	X	X	X	X	X	X	X	X	X	
RECHECK NATICK CLASS	/	/	/	/	/	/	/	/	/	/
FINISH MAP STATISTICS	/	/	/		/	/	/	/		
DEFINE VEG/URBAN PATTERNS	X	X	X	X	X	X	X	X	X	X
PHOTOGRAPH MAPS	X	X	X	/	/	X	/	/	/	/
PHOTOGRAPH OVERLAYS	X	X	X	/	/	X	/	/	/	/
DOCUMENT SCENARIOS		X	X		X	X		X		
SET UP LAND DECKS	/	X								
CARMONETTE ORDERS		/								
3-D PLOTS										
DEVELOP LOS STATISTICS	/	/	/	/						
DEVELOP BATTLE OUTCOMES		/								

Figure 53 Terrain Classification Work Outline (1 Oct 81)

1. EXTEND LOS INFORMATION DERIVED IN PREVIOUS STUDIES/TESTS TO NEW REGIONS OF SIMILAR SURFACE GEOMETRY.
2. GUIDE NEW WEAPONS ACQUISITION AND FORCE TAILORING.
3. INCREASE SURVIVABILITY ON THE BATTLEFIELD.
4. PREDICT LOS CHARACTERISTICS AS FN OF CLUTTER AND SURFACE ROUGHNESS:

$$P_1 = Ae^{\frac{-BR}{R}}$$

$$P_2 = e^{\cdot CR}\,(A\ \mathrm{Cos}\,\omega R + B\ \mathrm{Sin}\,\omega R)$$

Figure 54 Intervisibility Classification Uses

8. Conclusions. Many advances in terrain analysis software have been made since the late 1970s. Visualization of the battlefield on computerized color graphics is now commonplace. It is now also possible to develop detailed LOS masking information complete with statistical analysis in under six minutes of computer processing time for regions up to 40 x 70km in size (Refs. 9 and 15). This represents a ten-fold quantum jump in processing capability which did not exist two years ago. The accuracy of computer predictions has also been quantified and found reasonable for most applications. As processing capability has improved, so has data base coverage, although to a lesser degree. Possible near-term development of a new Defense Mapping Agency land combat data base would do much to improve analysis capabilities. Longer term, DMA production would improve the capability to do land combat analysis in a wide variety of locations around the world. Once this feasibility is demonstrated, transfer of this intelligence/operational planning capability from the laboratory to tactical units in the field is sure to follow. Finally, if the terrain classification research program is successful, it will be possible to generalize and extend analytic efforts to other regions when time or lack of data prevent detailed analysis --and do it with confidence. The decade of the 1980s promises more improvements in store for this important "combat multiplier".

REFERENCES

1. US Army Field Manuel (FM) 100-5, HQDA, Washington, DC, 1 July 1976 (UNCLASSIFIED).

2. ABCA Terrain Models Catalog, Quadripartite Advisory Publication 2, 8 May 1981 (UNCLASSIFIED).

3. ABCA Terrain Data Catalog, Quadripartite Advisory Publication, 8 May 1981 (UNCLASSIFIED).

4. McCoy, D. Hue, and Champion, Danny C., Telemetry Positioning at Ft. Irwin, Letter Report LR-30-80, US Army TRADOC Systems Analysis Activity, White Sands Missile Range, NM, 4 Dec 80 (UNCLASSIFIED).

5. Champion, Danny C., Telemetry Positioning at Ft. Sill, Draft Letter Report LR-35-81, US Army TRADOC Systems Analysis Activity, White Sands Missile Range, NM, 15 Sep 81 (UNCLASSIFIED).

6. Infantry Close Combat Advanced Antiarmor Requirements Study (ICCAARS), TRASANA COEA 27-81, US Army TRADOC Systems Analysis Activity, White Sands Missile Range, NM, August 1981 (SECRET).

7. Deason, Paul J., Determination of the Average Expected Movement Segment Lengths for Tanks in Combat, Unpublished TRASANA Memorandum, 10 Aug 81 (UNCLASSIFIED).

8. Report of the Special Study Group on the Infantry Fighting Vehicle and the Cavalry Fighting Vehicle, Prepared by US Army Combined Arms Combat Development Activity, Ft. Leavenworth, KS, 28 Aug 78 (SECRET).

9. Olson, Warren K., McCoy, D. Hue, and Deason, Paul J., GE/US SOTAS/PAVE MOVER Intervisibility Analysis, Unpublished Paper, US Army TRADOC Systems Analysis Activity, White Sands Missile Range, NM, 3 Sep 81 (UNCLASSIFIED).

10. Olson, Warren, K., A Terrain Analysis of Four Tactical Situations, Technical Memorandum No. 158, US Army Materiel Systems Analysis Activity, APG, MD, December 1972 (UNCLASSIFIED).

11. Anstey, Robert, J., The Natick Landform Classification System, Technical Report No. 100, US Army Materiel Systems Analysis Activity, APG, MD, October 1974 (UNCLASSIFIED).

12. Farrell, Robert L., Investigation of the Tactical Control of Cover and Exposure and its Relation ot Predicted Combat Results, Report No. VRI DUSA-1 FR 77-1, Vector Research, Inc., Ann Arbor, MI, February, 1977 (UNCLASSIFIED).

13. Reed, George W., An Evaluation of the Natick Landform Classification Systems, Interim Note No. T-13, US Army Materiel Systems Analysis Activity, APG, MD, December 1978 (UNCLASSIFIED).

14. West, H. W. and Krivitzky, D., Digital Terrain and Mobility Data Bases for E-FOSS, Misc. Paper EL-80-1, US Army Engineer Waterways Experiment Station, Vicksburg, MS, January 1980 (UNCLASSIFIED).

15. Broome, Barbara D., SEEFAR: An Improved Model for Processing Line-of-Sight Maps, Technical Report No. 225, US Army Materiel Systems Analysis Activity, Aberdeen Proving Ground, MD, September, 1980 (UNCLASSIFIED).

APPENDIX A 26 August 1981

PROTOTYPE FORT LEWIS-YAKIMA TRAINING AREA*
TACTICAL TERRAIN ANALYSIS DATA BASE (TTADB) FORMAT
FOR 1:50,000 SCALE PRODUCTS

Data Element	Bit Designation	# of Bits	Code	Value Represented
I. Surface Configuration Overlay				
Elevation (m)	1 - 16	(16)		
Slope (%)	17 - 20	(4)	0	No Data
			1	0 - 3
			2	3 - 10
			3	10 - 20
			4	20 - 30
			5	30 - 45
			6	>45
			7	Naturally and/or culturally dissected land (0->45) (Numerous small hillocks, sand dunes, glacial debris, landfills, dumps, etc.)
			8-14	
			15	Open Water
II. Vegetation Overlay				
Type	21 - 26	(6)	0	No Data
			1	Agriculture (dry crops)
			2	Agriculture (wetland rice)
			3	Agriculture (terraced crops, both wet and dry)
			4	Agriculture (shifting cultivation)
			5	Brushland/Scrub (<5m high, nearly open to medium spacing)
			6	Brushland/Scrub (<5m high, medium to dense spacing)
			7	Coniferous/Evergreen Forest
			8	Deciduous Forest
			9	Mixed Forest
			10	Orchard/Plantation (rubber, palm, fruit, etc.)
			11	Grassland, Meadows, Pasture
			12	Grassland with Scattered Trees, Some Scrub Growth
			13	Forest Clearings (cutover areas, burns, etc.)
			14	Swamp (mangrove, cypress, etc.)
			15	Marsh/Bog (peat, muskeg, etc.)
			16	Wetlands (L.S.I., low-lying wet areas)

*NOTE: On account of programming time limitations, this Fort Lewis-Yakima prototype is a condensation of the full proposed Tactical Terrain Analysis Data Base. Numerous data fields have had to be compressed, omitted, or specifically tailored to the Fort Lewis-Yakima terrain conditions to meet these limitations.

Data Element	Bit Designation	# of Bits	Code	Value Represented
Type (Con't)			17	Vineyard/Hop-garden
			18	Bamboo
			19	Bare Ground
			20-22	
			23	Open Water
			24	Built-up Areas
			25-29	
			30-63	Not Used
Canopy Closure (%)	27 - 29	(3)	0	No Data
			1	0 - 25
			2	25 - 50
			3	50 - 75
			4	75 - 100
			5-7	
Tree Height (m)	30 - 33	(4)	0	No Data
			1	0 - 2
			2	2 - 5
			3	5 - 10
			4	10 - 15
			5	15 - 20
			6	20 - 25
			7	25 - 30
			8	30 - 35
			9	> 35
			10-15	Not Used
Stem Diameter (m)	34 - 37	(4)	0	No Data
			1	0
(Note: CCM formula uses meters and these ranges were selected because they best correspond to the push-over limits of the vehicles for which we compute CCM)			2	.00 - .02
			3	.02 - .04
			4	.04 - .06
			5	.06 - .08
			6	.08 - .10
			7	.10 - .15
			8	.15 - .20
			9	.20 - .25
			10	.25 - .50
			11	.50 - 1.00
			12	1 - 3
			13	3 - 5
			14	5 - 10
			15	> 10
Stem Spacing (m)	38 - 41	(4)	0	No Data
			1-9	0 - 4.0 (by 0.5)
			10	4 - 5
			11	5 - 6
			12	6 - 8
			13	8 - 10
			14	10 - 15
			15	> 15

Data Element	Bit Designation	# of Bits	Code	Value Represented
Vegetation Roughness Factor	42 - 46	(5)	0	0
			1	.1
			2	.2
			3	.3
			4	.4
			5	.5
			6	.6
			7	.7
			8	.8
			9	.9
			10-31	Not Used
Undergrowth	47 - 48	(2)	0	No Data
			1	None
			2	Sparse
			3	Dense
Tree Crown Diameter (meters)	49 - 51	(3)	0	No Data
			1-7	Not Used

(Note: Need if we intend to do inter-visibility or line-of-sight products; otherwise will remain zeroed out)

Data Element	Bit Designation	# of Bits	Code	Value Represented
Height of Lowest Branch (meters)	52 - 54	(3)	0	No Data
			1-7	Not Used

(Note: Same as above)

III. Surface Material Overlay

Data Element	Bit Designation	# of Bits	Code	Value Represented
Type	55 - 60	(6)	0	No Data
			1	GW - Gravel, well graded
			2	GP - Gravel, poorly graded
			3	GM - Gravel, silty
			4	GC - Gravel, clayey
			5	SW - Sand, well graded
			6	SP - Sand, poorly graded
			7	SM - Sand, silty
			8	SC - Sand, clayey
			9	ML - Silt
			10	CL - Clays
			11	OL - Organic silts
			12	MH - Inorganic silts
			13	CH - Fat clays
			14	OH - Fat organic clays
			15	PT - Organic, peat
			16	Snowfield/Glacier
			17	Rock outcrops
			18	Evaporite
			19-61	
			62	Open water
			63	Not evaluated (built-up areas, etc)

Data Element	Bit Designation	# of Bits	Code	Value Represented
Qualifier	61 - 65	(5)	0	No Data
			1	None
			2	Boulder field
			3	Quarry, mine, diggings
			4	Bare rock, smooth
			5	Lava fow
			6	Dunes
			7	Loose
			8	Karst
			9	Lateritic
			10	Permafrost
			11	Frequent stone or rock outcrops
			12	Dissected
			13	Metal/ore slag dump
			14	Tailings, waste pile
			15	Strip mine
			16	Rugged bedrock
			17-31	
State of the Ground	66 - 69	(4)	0	No Data
			1	Dry
			2	Approximately 50% saturated
			3	Wet (saturated)
			4-14	0 - 1.0 (by 0.1--fraction of soil moisture in top one half meter at time of measurement or CCM synthesization)
			15	
Depth of Surface Material (meters)	70 - 71	(2)	0	No Data
			1	0 - 0.5
			2	>0.5
			3	
Surface Roughness Factor - Medium and Heavy Tanks (M-1 Abrams, M60, and T-72 Tanks, etc)	72 - 75	(4)	0	No Data
			1	0
			2	0.05
			3	0.1
	0-5.5in		4	0.2
			5	0.3
			6	0.4
			7	0.5
			8	0.6
			9	0.7
(Note: One SR table will be needed for each vehicle type for which a CCM map is to be prepared)			10	0.75
			11	0.80
			12	0.85
			13	0.90
			14	0.95
			15	1.00

Data Element	Bit Designation	# of Bits	Code	Value Represented
Surface Roughness Factor - Large Wheeled Vehicles (M35 Truck, etc)	76 - 79	(4)	0 1-15	No Data Not Used
Surface Roughness Factor Small Wheeled Vehicles (M151 Jeep, etc)	80 - 83	(4)	0 1-15	No Data Not Used
Surface Roughness Factor - Light Tracked Vehicles (M113 APC, etc)	84 - 87	(4)	0 1-15	No Data Not Used
Surface Roughness Factor Foot Troops	88 - 91	(4)	0 1-15	No Data Not Used
Depth to Bedrock (meters)	92 - 95	(4)	0 1-15	No Data Not Used

(Note: Need if we intent to support penetration or engineering studies; otherwise will remain zeroed out)

IV. Surface Drainage Overlay

Data Element	Bit Designation	# of Bits	Code	Value Represented
Type	96 - 98	(3)	0	No Data
			1	Stream Channel (Dry Wash or Intermittent, e.g., Arroyo)
			2	Lakes, Ponds, Reservoirs
			3	Stream Channel (Perennial)
			4	Stream Channel (Subject to Tidal fluctuations)
			5	Channelized Stream/Canal/ Irrigation Canal/Drainage Ditc
			6	Off-Route Ford (Entrance and Exit points connected)
			7	Dam/Lock
Gap Width (Bank to Bank (m))	99 - 101	(3)	0	No Data
			1	< 4.5
			2	4.5 - 18
			3	18 - 50
			4	50 - 100
			5	100 - 142
			6	> 142
			7	
Bottom Material	102 - 104	(3)	0	No Data
			1	Clay and Silt
			2	Silty Sand
			3	Sand and Gravel
			4	Gravel and Cobble
			5	Rocks and Boulders
			6	Bedrock
			7	Paved

Data Element	Bit Designation	# of Bits	Code	Value Represented
Height, right bank (m)	105 - 107	(3)	0	No Data
			1	<0.5
			2	0.5 - 1.0
			3	1.0 - 5.0
			4	>5.0
			5-7	
Height, left bank (m)	108 - 110	(3)	0	No Data
			1	<0.5
			2	0.5 - 1.0
			3	1.0 - 5.0
			4	>5.0
			5-7	
Slope, right bank (%)	111 - 113	(3)	0	No Data
			1	<30
			2	30 - 45
			3	45 - 60
			4	>60
			5-7	
Slope, left bank (%)	114 - 116	(3)	0	No Data
			1	<30
			2	30 - 45
			3	45 - 60
			4	>60
			5-7	
Water velocity, average (meters/seconds)	117 - 118	(2)	0	No Data
			1	≤2.5
			2	>2.5
			3	Not Used
Water depth, average (m)	119 - 121	(3)	0	No Data
			1	<0.8
			2	0.8 - 1.6
			3	1.6 - 2.4
			4	>2.4
			5-7	
Dense Vegetation Along Stream Banks	122	(1)	0	No Data
			1	>50% Segment Length

(Note: Normally closely spaced row of trees, which could possibly hinder stream crossing operations)

Data Element	Bit Designation	# of Bits	Code	Value Represented
V. Transportation Overlay				
Type	123 - 126	(4)	0	No Data
			1	Bridge - Road
			2	Bridge - Railroad
			3	Tunnel - Road
			4	Tunnel - Railroad
			5	Dual Lane/Divided Highway/ Expressway
			6	Highway/Road
			7	Railroad
			8	Airfield
			9	Inland Waterway
			10	Lock
			11-15	
Condition	127 - 129	(3)	0	No Data
			1	Good — Operational
(Note: Need part of this field now, will need the rest if we intend to support air operations or on-route trafficability studies)			2	Fair — "
			3	Poor/Deteriorated — "
			4	Damaged
			5	Destroyed
			6	Abandoned/Dismantled
			7	Under construction
Qualifier	130 - 132	(3)	0	No Data
			1	Road Constriction, <4 meters
(Note: Now do all except the grade in excess of 3% for railroads)			2	Grade in excess of: 7% for roads or 3% for railroads
			3	Sharp curve with radius <30 meters
			4	Ferry Site
			5	On Route Ford Site
			6	Electrified Line
			7	
Length (meters)	133 - 138	(6)	0	No Data
			1	Unknown
(Note: For this Fort Lewis-Yakima prototype, length only refers to Bridges, Tunnels, Airfields, and other transportation types less than 100 meters long. All lengths greater than 100 meters are determined by the the number of points used to digitize the length of the feature - 0.25 mm (approx. 0.01 inches) equals 12.5 meters on the ground at 1:50,000 scale).			2	0 - 10
			3	10 - 20
			4	20 - 30
			5	30 - 40
			6	40 - 60
			7	60 - 80
			8	80 - 100
			9	>100
			10-31	Not Used

Data Element	Bit Designation	# of Bits	Code	Value Represented
Average Width (meters)	139 - 142	(4)	0	No Data
			1	Unknown
			2	0 - 3
(Note: For this Fort Lewis-Yakima prototype, width refers to any transportation type, except Railroads, less than 50 meters wide. All widths greater than 50 meters are determined the same as lengths greater than 100 meters - see above).			3	3 - 4
			4	4 - 5
			5	5 - 7
			6	7 - 10
			7	10 - 20
			8	20 - 50
			9	>50
			10	Not Used
Surface	143 - 146	(4)	0	No Data
			1	Paved
(Note: Need part of this field now, will need the rest if we intend to support on-route trafficability studies)			2	Hard
			3	Loose/Gravel
			4	Unpaved
			5	Natural Earth
			6	Grass
			7	Macadam/Asphalt/Bituminous
			8	Concrete
			9	Stone/masonry/brick
			10-15	Not Used
Highways and/or Roads:				
Type	147 - 149	(3)	0	No Data
			1	All Weather
			2	Fair/Dry Weather
			3	Cart Tracks
			4	Trails
			5-7	
Railroads:				
Type	150 - 152	(3)	0	No Data
			1	Normal Gauge, single track
			2	Normal Gauge, dual track
			3	Normal Gauge, multiple (2 or more) tracks
			4	Narrow Gauge, single track
			5	Narrow Gauge, multiple track
			6	Broad Gauge, single track
			7	Broad Gauge, multiple track
Passing tracks, sidings & yards (meters)	153 - 154	(4)	0	No Data
			1	Passing track ≥ 280
			2	Siding ≥ 280
			3	Yard ≥ 280

Data Element	Bit Designation	# of Bits	Code	Value Represented
Tunnels:				
Height (meters)	155 - 157	(3)	0	No Data
			1	3 - 6
			2	6 - 8
			3	8 - 12
			4	>12
			5-7	
Bridges:				
Type	158 - 161	(4)	0	No Data
			1	Truss
(Note: Need if we intend to support engineer or on-route trafficability studies; otherwise will remain zeroed out)			2	Girder
			3	Beam
			4	Slab
			5	Arch
			6	Suspension
			7	Floating
			8	Cable stayed
			9	Cantilever
			10-15	Not Used
Movement	162 - 164	(3)	0	No Data
			1	Fixed
(Note: Same as above)			2-7	Not Used
Overhead Clearance	165 - 166	(2)	0	No Data
			1	Unknown
			2	Unlimited clearance
			3	Possible obstruction to military traffic
Horizontal Clearance	167 - 168	(2)	0	No Data
			1	Unknown
			2	Unlimited Clearance
			3	Possible obstruction to military traffic
Underbridge Clearance	169 - 171	(3)	0	No Data
			1	Unknown
			2	0 - 5
			3	5 - 10
			4	10 - 50
			5	>50
			6-7	Not Used

Data Element	Bit Designation	# of Bits	Code	Value Represented
Bypass Conditions (Potential within 2km)	172 - 173	(2)	0	No Data
			1	Easy
			2	Difficult
			3	Impossible
Construction Material	174 - 176	(3)	0	No Data
			1	Other
			2	Wood
			3	Stone/masonry/brick
			4	Steel
			5	Concrete
			6	Reinforced concrete
			7	Prestressed concrete
Classification (one-way wheeled) (metric tons)	177 - 180	(4)	0	No Data
			1	50
			2-9	
			10-15	Not Used
Classification (one-way tracked) (metric tons)	181 - 184	(4)	0	No Data
			1-7	0 - 60 (by 10)
			8	61 - 100
			9	>100
			10-15	Not Used
Reliability of Bridge Classification	185 - 186	(2)	0	No Data
			1	Unknown
			2	Known
			3	Estimated
Spans (number)	187 - 190	(4)	0	No Data
			1	Unknown
			2	1
			3	2
			4	3
			5-8	4 - 11 (by 2's)
			9	≥12
			10-15	
Span Length (meters)	191 - 193	(3)	0	No Data
			1	Unknown
			2	<25
			3	25 - 50
			4	50 - 100
			5	>100
			6-7	

Data Element	Bit Designation	# of Bits	Code	Value Represented
VI. Obstacles Overlay				
Type	194 - 196	(3)	0	No Data
			1	Road and RR cuts and fills
(Note: Linear obstacles are defined as any hinderance to movement which is greater than 1.5 meters high, has a 45% or greater slope, and is at least 250 meters long. Areal obstacles are defined as any area which is so characterized as to severely restrict, stop, or otherwise make movement impractical. This overlay depicts land obstacles only--user is referred to Surface Drainage Overlay for hydrologic obstacles.)			2	Natural linear obstacles (escarpments, dikes, cliffs, etc.)
			3	Walls and/or fences (hedgerows, rock and wire fences and walls, retaining walls, etc.)
			4	Other man-made linear obstacles (dikes, moats, embankments, etc)
			5	Military obstacles (antitank ditches, airfield and/or road craters, blown bridges, debris choked valleys and/or towns, impact areas, minefields, roadblocks, trenches, wire entanglements, etc.)
			6	Man-made areal obstacles (mining operations--pits, quarries, strip mines, etc.--terraced hills and/or paddies (wet and dry))
			7	Natural areal obstacles (craters, dissected land, talus piles, depressions, sink holes, open water, etc.)
Height (m)	197 - 199	(3)	0	No Data
			1	>1.5
			2	1.5 - 5
			3	5 - 10
			4	10 - 20
			5	20 - 35
			6	>35
			7	

The Prototype Fort Lewis-Yakima Training Area TTADB matrixed format ends with bit 199. The full proposed TTADB includes 224 bits and includes additional overlays for:

(1) Aerial Obstructions.

(2) Special Features/Product Synthesis (customer related overlays/standardized and future CCM, Concealment, etc. overlays).

(3) Text Data (two) - for such information as bridge tables, climatic data, hydrologic flow graphs, names, generalized descriptors, etc. On account of the limitations mentioned on page 1, rather than carry 25 bits packed with zeroes at the end of each data point, it was decided to omit these fields from the Fort Lewis-Yakima Training Area TTADB.

APPENDIX B 18 Sep 81

SUBJECT: Simple Analysis Methods by Which to Discriminate Among Terrain Compartments[1]

The purpose of this paper is to simply describe what data may be used and, due to whether this data is expressed as discrete or continuous variables, what analytic tools might be brought to bear in the determination of whether the terrain compartments are equal or whether there are differences between them. These analyses may lend themselves to answer such questions as: whether the Natick plus Veg-Urb classifications adequately and/or accurately describe the various compartments; what variables show the greatest discriminative ability, acting alone, between compartments; are there interaction effects between the levels of two variables; etc. These procedures are labeled simple analyses in that they are precursors to more involved techniques which seek to uncover the actions and interactions of several variables as they are expressed in the compartments. The techniques which shall be used on "continuous" variables (those for which at least the mean, standard deviation, and n are available) consist of one way analyses of variance (ANOVA), and following a significant F, the determination of which means are equal, which are unequal through the use of such a posteriori methods as the least significant difference (lsd) test or the Student-Newman-Keuls multirange procedure.[2]

One Way ANOVA for 13 Compartments

Source	SS	df	MS	F	MS expected value
Compartments	SS_c	12	SS_c/df	MS_c/MS_e	$\frac{N^2-\sum nj^2}{N(12)}\sigma_t^2 + \sigma_\varepsilon^2$
Error	SS_e	N-13	SS_e/df		$E(MS_{error})=\sigma_\varepsilon^2$
Total	SS_t	N-1			

In several occasions, where there are two levels of a variable (such as in-view/out-of-view) a two way ANOVA should be appropriate, with the determination of the significance of the interaction between the two variable levels and the compartments.

---

[1] *This paper was done not only to outline the ways to analyze the terrain data, but also to serve as a refresher in the mechanic's of the various techniques to those who may be a little "rusty".*

[2] *References for the ANOVA and a posteriori contrasts are Steel, R.G.D. & J.H. Torrie; Principles and Procedures of Statistics; New York, McGraw-Hill; 1960 and Winer, B.J.; Statistical Principles in Experimental Design, 2nd Edition, New York, McGraw-Hill; 1971.*

Two Way ANOVA (equal cell size) for 13 Compartments					
Source	ss	df	ms	f	ms expected value
Compartments	SS_c	12	ss_c/df_c	ms_c/ms_e	$2n\sigma^2_c + \sigma^2_e$
IV/OOV	SS_i	1	ss_i/df_i	ms_i/ms_e	$13 \quad n^2_I + \sigma^2_\varepsilon$
Comp x IV/OOV	SS_x	12	ss_x/df_x	ms_x/ms_e	$n\sigma^2_{ci} + \sigma^2_\varepsilon$
Error	SS_e	13 * (n-1)	ss_e/df_e		σ^2_ε
Total		13*2*(n)-1			

An aposteriori contrast test is a systematic procedure for comparing all possible pairs of group means. The groups are divided into homogeneous subsets where the difference in the means of any two groups in a subset is not significant at some prescribed significance level (alpha). The procedure is based on the test:

$$[\bar{X}_i - \bar{X}_j] < R(alpha, g,f)*S\bar{X}$$

Where R(alpha,g,f) is a range based on a significance level (alpha), the number of groups n the subset (g), and the degrees of freedom (f) in the between-groups sum of squares (error degrees of freedom). $S_{\bar{X}}$ is the standard error in the combined subset, and is equal to $(MSerror)^{1/2}$.

The least significant difference test (lsd) is essentially a Student's t test between group means. It is usually not recommended since as the number of groups increases, so does the experiment wise error rate. However, with the prior determination of a significant F, which puts an upper bound on the experiment wise error rate, the lsd procedure is considered are appropriate liberal test. lsd is also exact for unequal group sizes. The lsd is computed as follows: lsd=t sd, where t is the Student's t for the chosen significance level and error degrees of freedom. $Sd=(MS_{error}(\frac{1}{ri} + \frac{1}{rj}))^{1/2}$ where ri and rj are the sample sizes of the two means being compared. If $|xi-xj|>lsd$, then the two means are considered significantly different. The alpha level of significance may be modified for the lsd procedure, and take care of the expanding experiment wise error rate by choosing a lower alpha.

The Student-Newman-Keuls (SNK) test attempts to avoid the expanding error rate problem in the lsd by using a different range value for subsets of different sizes (the larger the number of groups in a subset the larger the difference in the means must be in order to be declared significant). The SNK test also uses the concept of a special protection level rather than a significance level: the probability of finding a significant difference, given that two groups are in fact equal, is less than or equal the specified significance level. SNK holds the experiment wise error rate to alpha for each stage of the testing procedure. This procedure is more conservative than the lsd procedure (a greater tendency to find means not significant than the lsd), however the validity of this procedure for handling unequal group sizes has not been verified. The following is offered as an example of the one-way ANOVA followed by a posteriori comparisons. This is a comparison of four terrain compartments for first observation ranges:

	1	2	3	4
N	86	282	108	142
$\bar{X}$	3240	3298	2404	2788.3
Sd	988.62	11878.2	1384.18	1506.84

One Way ANOVA	df	ms	f	
Compartments	3	24854429.42	15.2	sig<.001
Error	614	1635639.407		
Total	617			

A posteriori comparisons -

Compartment	$\bar{X}$	difference 4	1	2
3	2404	383.7	836	894
4	2788.3		451.7	509.7
1	3240			58
2	3298			

lsd $\quad t_{.99,617}=2.33$ one tail, $lsd=t(MS_e(\frac{1}{ri}+\frac{1}{ri}))^{1/2}$

mean 3&4 lsd-380.5 ∴ Sig, 4>3 mean 4&1 lsd=407.168∴Sig 1>4
mean 3&1 lsd=430.66∴ Sig, 1>3 mean 4&2 lsd=306.63∴Sig 2>4
mean 3&2 lsd=337.2∴ Sig, 2>3 mean 1&2 lsd=367.07∴NS 1=2

The results of the lsd test can be summarized by saying:

The mean of compartment 3 is less than the mean of compartment 4, which is less than the means of compartments 1 and 2, which are equal; or 3<4<1=2.

The SNK test is performed by first looking up in a table the multipliers associated with the number of means across which the test is to be made. For instance, in the above table of ordered means a test of 3 and 4 would be across two means, 3 and 1 across three means, 3 and 2 across four means. If one looks in a table of the Studentized range statistic[3] for alpha of .01, df of ∞ the multiplier q for 2 is 3.64, for 3 is 4.12, for 4 is 4.40. The critical value for the difference between two means is

$$q \left(\frac{MS_{error}}{\tilde{N}}\right)^{1/2}$$

where $\tilde{N}$ is the harmonic mean, k is number of compartments

$$\tilde{n} = \frac{k}{(1/n_1) + 1/n_2 + \cdots + (1/n_k)} \Rightarrow \frac{4}{(1/86)+(1/282)+(1/108)+(1/142)} = 127.08$$

so $\left(\frac{MS_{error}}{\tilde{n}}\right)^{1/2} = 113.45$, and the critical value between 2 means is 412.95, between 3 is 467.41, between 4 is 499.18. By examining the mean difference table above, one may see that:

$\bar{X}_4-\bar{X}_3=383.7 < 412.95 \therefore$ 3&4 not different

$\bar{X}_1-\bar{X}_3=836 > 467.41 \therefore$ significant 1>3

$\bar{X}_2-\bar{X}_3=894 > 499.18 \therefore$ significant 2>3

$\bar{X}_1-\bar{X}_4=451.7 > 412.95 \therefore$ significant 1>4

$\bar{X}_2-\bar{X}_4=509.7 > 467.41 \therefore$ significant 2>4

$\bar{X}_2-\bar{X}_1=58 < 412.95 \therefore$ 1&2 not different

The results of the SNK may be summarized as follows: 3=4<1=2.

This also serves as an example of the difference between the "liberal" lsd and the "conservative" SNK. Take the difference between mean 3 and mean 4; lsd found these means as significantly different, even though close to the critical value, whereas SNK did not reject the null hypothesis of no difference between them. It depends on the audience, but I would tend to go with the lsd results, due to the prior significant F and the associated increase in power the lsd allows.

---

[3] *Winer, opcit pp 870-871.*

For the two way analysis of variance, the same procedures apply; to test between two means, use a t-test or lsd test, using the MS within cell and its associated df to determine the criterion value of t. In making all possible tests between ordered means, the SNK procedure is to be used, using the harmonic mean of the cell sample sizes for n. It is also possible to test the significance of the simple effects of one factor over one level of a second factor, especially useful if there is interaction present between the two factors. For example, the sample effects of the four means of a factor B over the second level of A, A2, would be

$$SS_{b\,for\,a_2} = \bar{N}h\left[(\bar{b}_{1,2})^2+(\bar{b}_{2,2})^2+(\bar{b}_{3,2})^2+(\bar{b}_{4,2})^2-\frac{(\sum_1^4 b_{i,2})^2}{4}\right]$$

$$MS_{b\,for\,a_2} = \frac{SS_{b\,for\,a_2}}{3}$$

$$F = \frac{MS\ b\ for\ a_2}{MS\ within\ cell}$$ (the MS within cell is from the 2-way ANOVA)

The df for this F are 3 and the df within cell.

At this time, the following continuous variables are available for the terrain compartments:

a. Thickness and separation measures for vegetation, urban features, and the combination of vegetation and urban features.

b. In-view and out-of-view segment lengths.

c. First opening range.

d. Number of in-view and out-of-view segments per route.

It may be necessary to transform some of these variables so that they more nearly represent a normal distribution. This will be determined as the data becomes available which will allow a determination of the underlying distribution. Battle outcome may also be considered a continous variable when sufficient replications are performed; Battle outcome will then be handled in a repeated measures ANOVA design.

Discrete variables are those in which the data is broken into categories, which have a frequency of occurrance attached (bean counting)[4]. When the categories are considered nominal (just names-no relation between them)

---

[4]*References for nonparanictric analysis procedures include Bradley, J.V.; Distribution-Free Statistical Tests; Englewood Cliffs, N.J., Prentice-Hall, 1968 and Siegel, S; Nonparametric Statistics: For the Behavioral Sciences; New York, McGraw-Hill; 1956.*

measures of association such as Chi-square, or Mueller's lambda are appropriate. However, by ordering the terrain compartments by some means (such as the Natick system) then it is possible to use measures of association which depend on ordinal variables (ordered, but not assuming a constant separation between variables) such as tau, Gamma, or Somer's D.

The Chi-square is a test which measures whether a systematic relationship exists between two variables (that is, whether the variables are dependent or independent). This is done by computing the cell frequencies which would be expected if no relationship is present between the variables given the existing row & column totals (marginals). The expected cell frequencies are then compared to the actual values found in the contingency table according to the following formula:

$$\chi^2 = \sum_i \frac{(f_o^i - f_e^i)^2}{f_e^i}$$

where f_o^i equals the observed frequency in each cell, and f_e^i equals the expected frequency calculated as

$$f_e^i = \left(\frac{c_i r_i}{N}\right)$$

where c_i is the frequency in a respective column marginal, r_i is the frequency in a respective row marginal, and N stands for the total number of valid cases. Small values of χ^2 indicate the absence of a relationship, or statistical independence. Conversely, a large χ^2 implies that a systematic relationship of some sort exists between the variables.

Mueller's asymmetric lambda measures the percentage of improvement in the ability to predict the value of one variable once the value of the other variable is known. This is based on the assumption that the best strategy for prediction is to select the category with the most cases (modal category), since this will minimize the number of wrong guesses. This concept is called the proportional reduction in error. The formula for lambda is

$$\lambda_b = \frac{\sum_k \max f_{jk} - \operatorname{Max} f_{.k}}{N - \max f_{.k}}$$

where $\max f_{jk}$ represents the sum of the maximum values of the cell frequencies in each column, and $\max f_{.k}$ represents the maximum value of the row totals.

Tau b, Gamma, and Somer's D are measures of association between two ordinal-level variables, and are built upon a common basis. They use the information about the ordering of categories of variables by considering every possible pair of cases in the table. Each pair is checked to see if their relative ordering on the first variable is the same (concordant) as their relative ordering on the second variable, or if the ordering is reversed (discordant).

The first step is to compute the number of concordant pairs (P) and discordant pairs (Q). If P is larger than Q, this means that there is a preponderance of pairs ordered in the same direction on both variables, and the statistic will be positive. Conversely, a larger Q will result in a negative statistic. This positive or negative association is also referred to as correlation.

$$\text{Tau b} = \frac{P-Q}{[{}^1/2(n(n-1)-\Sigma T_1\ (T_1-1))^1/2(N(N-1)-\Sigma T_2(T_2-1))]^1/2}$$

where T_1 is the number of ties on row variables, T_2 is the number of ties on the column variables.

Gamma makes no adjustment for ties or table size, and is simply

$$\text{Gamma} = \frac{P-Q}{P+Q}$$

A variation on Gamma, which accounts for ties but not table size is Somer's

$$\text{Somer's D (symmetyric)} = \frac{P-Q}{P+Q+{}^1/2(T_1+T_2)}$$

These measures differ basically on the manner in which ties are handled, and are basically used to give an indication of the type and relative strength of the association between two variables.

A Friedman 2 way ANOVA by ranks, may be appropriate depending on the appearance of the data and the number of tied values. The comparison of two distributions for equality may be performed by using the Kolmogorov-Smirnov two-sample test. These procedures are appropriate for comparing between terrain compartments for modal hill height, and PLOS vs. range when range is expressed as range bands.

The Friedman two-way ANOVA would be used to test, for example, that each terrain compartment was ranked the same in each range band. Given k compartments in columns and N rangebands in rows, rank the scores in each row from 1 to k. Determine the sum of ranks in each column: Rj, and compute

$$\chi^2_r = \frac{12}{Nk(k+1)} \sum_{j=1}^{k} (Rj)^2-3N(k+1)$$

wnere N=number of rows
k=number of columns
Rj=sum of ranks in column j

χ^2_r is approximated by χ^2 with df=k-1. If the value of χ^2_r is equal to or larger than a tabled value of χ^2_{k-1}, the implication is that the sum of the ranks for the various columns differ significantly. The Kolmogorov-Smirnov two-sample test can be used as a suitable a posteriori comparison of distributions after the Friedman test. The procedure for this test would be to compute the cumulative distribution of each of the two compartments under test, then locate the range band where the difference between the cumulative percentage in one compartment is maximized with relation to the second compartment. This difference D is computed as

$$D = \mathrm{Max}\left[\frac{cf_1}{n_1} - \frac{cf_2}{n_2}\right] \text{ for a one-tail test.}$$

where n_1 and n_2 are the sample sizes of compartment 1 and compartment 2, respectively. The statistic is computed, for large n, as

$$\chi^2 = 4D^2 \left[\frac{n_1 n_2}{n_1 + n_2}\right]$$

which is approximated by the χ^2 distribution with df=2. If this statistic is found significant, then the two distributions are different. The rationale behind this test is that if the two samples were distributed equally, their cumulative distributions would not be very different.

Some of the variables encountered have a single value per terrain compartment. These may be used, given the terrain compartments are ordered, by using a technique such as the Spearman or Kendall rank-order correlation. A significant positive correlation would indicate that the compartments were properly ordered with respect to the observed variables. Single value data that would be analyzed with this technique would be maximum hill height, the number of positive features per kilometer, the total number of positive features, and battle outcome results if not replicated.

The Kendall[5] and Spearman rank correlations are correlations based upon the ordinal ranks of the observations of two variables, and not on their observed value. let $r_1, r_2, \ldots r_n$ represent the ranks of the values of one variable, and $s_1, s_2, \ldots s_n$ the representative ranks of the second variable. Correlation can be tested by arranging the n units in increasing order on the r variable, and testing the resulting order of the s variable for randomness.

---

[5] *refer also to Kendall, M.G., and A. Stuart; The Advanced Theory of Statistics, Vol 2, 3rd Ed; New York, Hafner, 1973, pp 494-505.*

If the two variates are independent (so that there is no correlation), the resulting sequence of s observations is equally likely to be any of the n! possible permutations of the n S's. However, if the two variables are linearly (or even "monotonically") correlated, the S-observations should tend to form an increasing or decreasig sequence, and any statistic that reflects this increase or decrease can be used to test for correlation. (The advantage of the Kendall over the Spearman is that the Kendall can be generalized to a partial correltion coefficient.) The Kendall rank correlation τb is defined as

$$\tau b = \frac{P-Q}{N(N-1)} \quad \text{where}$$

P=twice the number of pairs of rankings such that both $rj>r\ell$ (the increasing order) and $Sj>s\ell$ (also increasing order, thereby agreement in rank order direction).

Q=twice the number of pairs of rankings such that $rj>r\ell$ and $Sj<S\ell$ (disagreement in rank order due to the "inversion" of s from the "natural" ascending order). When rankings are tied in either r or s the formula for t_b is

$$t_b = \frac{P-Q}{([N(N-1)-T_1][N(N-1)-T_2])^{1/2}}$$

Where $T=\Sigma^* T_i(T_{i-1})$ and T_i is the number of observations tied with a single value. The sum Σ^* is over all distinct values for which a tie exists. T_1 is the total for the first variable r, T_2 for s.

The Spearman rank correlation is defined as

$$r_s = 1 - \left[\frac{6\Sigma(r_j - s_j)^2}{N^3 - N}\right]$$

when rankings are tied, r_s is modified to

$$r_s = \frac{A+B-D}{2\,(AB)^{1/2}}$$

where $A = (N^3-N-T_1)/12$

$B = (N^3-N-T_2)/12$

$D = \Sigma(r_j-s_j)^2$

and $T = \Sigma^* (T_i^3-T_i)$, defined as above.

As an example of these procedures, consider the following set of data[6]

Group	Attribute A	Attribute B	$r_j - s_j$	$(r_j - s_j)^2$
A	2	3	-1	1
B	6	4	2	4
C	5	2	3	9
D	1	1	0	0
E	10	8	2	4
F	9	11	-2	4
G	8	10	-2	4
H	3	6	-3	9
I	4	7	-3	9
J	12	12	0	0
K	7	5	2	4
L	11	9	2	4
				$\Sigma = 52$

for the Spearman rank correlation, this would turn out to be

$$r_s = 1 - \frac{6 \sum_{j-i}^{n} (r_i - s_j)^2}{N^3 - N} = 1 - \frac{6(52)}{1716} = .82$$

when N is 10 or larger, the significance of an obtained r_s under the null hypothesis may be tested by $t = r_s \left(\frac{n-2}{1-r_s^2}\right)^{1/2}$ and compare with student's t for df = N-2.

For this example

$$t = .82\left(\frac{12-2}{1-(.82)^2}\right)^{1/2} = 4.53 > t_{10,.005} \text{ for a one tail test.}$$

[6]*from Siegel pp 205*

The corresponding Kendall correlation is computed as follows:

Group	D	C	A	B	K	H	I	E	L	G	F	J
Att A(r_j)	1	2	3	4	5	6	7	8	9	10	11	12
Att B(s_j)	1	5	2	6	7	3	4	10	11	8	9	12

As an easy way of computing P and Q, $1/2$P and $1/2$Q is measured as the number of ranks of s; to the right of each individual s_j that are either greater than or less than the subject rank (or $1/2\ P_j$ and $1/2\ Q_j$, respectively). Therefore

$$1/2(P-Q)=(11-0)+(7-3)+(9-0)+(6-2)+(5-2)+$$
$$(6-0)+(5-0)+(2-2)+(1-2)+(2-0)+(1-0)=44.$$

(The attribute rank farthest to the left is 1. This rank has 11 ranks to the right which are larger, 0 which are smaller so $1/2\ (P_1-Q_1)=(11-0)$, and so on).

$$\tau_b=\frac{2(44)}{12(11)}=.67$$

When N is larger than 10, τ may be considered normally distributed with Mean=M_τ=0 and standard deviation=$\sigma\tau\ \left(\frac{2(2N+5)}{9N(N-1)}\right)^{1/2}$

$$\text{so } z=\frac{\tau-Mt}{\sigma T}=\frac{\tau}{\left(\frac{2(2N+5}{9N(N-1}\right)^{1/2}}$$

is approximately normally distributed with zero mean and unit variance. Thus the significance of z may be determines by reference to an appropriate z table. For the example, the test of whether the two variables are associated can be computed as:

$$Z=\frac{.67}{\left(\frac{2[(2)(12)+5]}{(q)(12)(11)}\right)^{1/2}}=3.03$$

by reference to a z table it is found that z $>$3.03 has the probability of occurance under Ho: no association, of p=.0012, this Ho is rejected, and it is concluded that the two variables are associated. You will note that the Spearman and Kendall procedures produce different coefficients of correlation when both were computed from the same pair of rankings. These examples illustrate the fact that τ and r_s have different underlying scales, and so numerically are not directly comparable to each other. However, both coefficients utilize the same amount of information, and thus both have the same power to detect the existance of association in the population. When used on data to which the Pearson r is properly applicable, both τ and r_s have efficiency of 91 percent.

As was initially stated, these methods are initial ones, meant to see how similar and/or dissimilar the terrain compartments are. Higher order analyses planned are: factor analyses to see how the several variables relate; multiple linear and nonlinear regression, to determine the effects of variables on dependent variables of interest, such as segment length, or observer-target distance; time series analyses to determine the terrain effects on variables of interest as time progresses; and analyzing differences in distributions as to their possible underlying causes. These higher level analyses will be undertaken as the data becomes available and programs are operational.

Also to be undertaken in the analysis of terrain is whether items of interest (such as line of sight, or battle outcome) vary over time, or distance from observer to target, and whether this variation can be expressed by an equation which is distinct for the separate types of terrain as identified by the (modified) Natick system. Variables which may be analyzable in this determination may be:

a. Percentage of Line of Sight (P_{LOS}) - for each unit of time t, for observer array O, target array T, the proportion of targets to which intervisibility exists, denoted P (t).

b. Correlation (or Coherence)[7] - for each unit of time t, for O and T, the correlation of several observers seeing several targets, denoted $\rho(t)$.

c. Percentage of Line-of-Sight and correlation, as above, varying as the distance L from O and T (in the case of correlation this distance may need to be to the centroid of the array T), denoted $P(\ell)$ and $\rho(\ell)$.

d. Battle outcome - in terms of targets killed by observers killed per unit time t, denoted k(t).

---

[7] *In the sample case of two observers seeing a group of targets, correlation is defined as* $p = \dfrac{P(AB)-P(A)P(B)}{(P(A)(1-P(A))P(B)(1-P(B)))^{\frac{1}{2}}}$.

Dr. Wilbur Payne defines coherence as the state when $p>0.7$.

In several recent articles[8] there has been described the potential for describing terrain with an equation, and that the spectral density of fluctuations in terrain (as well as many other physical quantities such as music and speech) vary as 1/f, where f is the frequency. This 1/f behavior implies some correlation in fluctuating quantities, such as PLOS, overall times corresponding to the frequency range for which the spectral density is 1/f, such as may be found in a homogeneous terrain compartment.

The spectral density Sp(f) of a quantity P(t) fluctuating with time t is a measure of the mean squared variation $\langle P^2 \rangle$ over the terrain compartment. A second characterization of the average behavior of V(t) is the auto correlation function, $\langle P(t)P(t+\tau)\rangle$, which is a measure of how the fluctuating quantities at times t and $t+\tau$ are related. For a stationary process $\langle P(t)P(t+\tau)\rangle$ is independent of t and depends only on the time difference τ. Sp(f) and $\langle P(t)Pt+\tau)\rangle$ are not independent, but are related by the Wiener-Khintchine relations[9].

$$\langle P(t)P(t+\tau)\rangle = \int Sp(f)\cos(2\pi f\tau)df$$

and $$Sp(f)=4\int\langle V(t)V(t+\tau)\rangle\cos(2\pi f\tau)d\tau$$

The methods for determining the appropriate f, spectral densities, and series analysis have not been fully worked out at this time, but will probably involve the use of the Spectral Anaysis and the Box-Jenkins Analysis routines contained in the 1981 release of the BMDP stastistical package[10].

PAUL DEASON

[8] *Cunningham, E.P.; "Single-Parameter Terrain Classification for Terrain Following"; J. Aircraft, 1980, pp 909-974.*
Voss, RF, and John Clarke; "1/f noise" in music: Music from 1/f noise" Acoust.SocAm 63(1), Jan 78, 258-263. Gardner, M.;"White and brown music, fractal curves, and one-over-f fluctuations" Mathematical Games Section of Scientific American, 1978.

[9] *Reif, F. Fundamentals of Statistical and Thermal Physics; New York, McGraw-Hill, 1965, pp 585-587.*

[10] *Biomedical Computing Programs, Health Sciences Computing Facility, University of California at Los Angeles, 1981.*

ERROR PROPAGATION IN PHYSICAL MODELS

Jerry Thomas and J. Richard Moore
US Army Ballistic Research Laboratory

ABSTRACT. Let $Y = g(X)$ be an equation which describes a physical process. If the argument of this function is subject to error, then Y will reflect this error. The transfer of error in X to error in Y through g(X) is commonly called error propagation. This paper gives a collection of procedures for studying commonly used error propagation equations and improving them when necessary.

1. INTRODUCTION. A physical model is defined for the purpose of this discussion to be either a mathematical equation or a set of simultaneous mathematical equations which describe the behavior of a real or conceptual physical system. Some examples of such models are:

$$Y = a + bX + cX^2 \tag{1}$$

and

$$Y = a'e^{b'X} \tag{2}$$

where Y is the muzzle velocity of a projectile launched from a gun with a propellant charge weight of X and a, b, c, a' and b' are parameters depending on other ballistic conditions.

Equations (1) and (2) form an example of two different models which may be used to describe the same phenomenon. If you were to fire a number of rounds, systematically varying propellant charge weight and recording charge weight and muzzle velocity for each round fired, you could fit either equation (1) or equation (2) to your data, i.e., you could find values for the unknown parameters a, b, c, a' and b' which best predict the observed Y's using the observed X's. The next natural step is to use one of the fitted equations to predict the muzzle velocity of a projectile fired with a propellant charge weight of X. This prediction would contain three sources of error.

One source of prediction error is the systematic error resulting from the inexact structure of the model used. Clearly predictions given by equation (1) will differ from those given by equation (2) and one can never be certain which of the two is the most reasonable to use or even whether some other functional form of the model would be appropriate. This type of error seems to be best controlled by careful consideration of the physical principles underlying the process to be modeled and recourse to sound engineering judgements. This type of error propagation will not be considered further in our discussion.

The other two errors which manifest themselves are random variables. They are the errors in estimating the parameters in the model and the inexact value of the independent variable(s) from which predictions are to be made. These are the errors which will be discussed below.

For a good treatment of the problem of propagation of errors in the model input (independent) variables the reader is referred to references 1 and 2.

We do not propose to give a collection of recipes for error analysis for physical models. Rather, we will discuss several techniques and some of their implications under usual assumptions.

2. ONE FUNCTION OF ONE RANDOM VARIABLE. Consider the physical model

$$Y = g(X) \tag{3}$$

where X is a random variable with distribution function $F(x;\theta)$ with θ being a vector of one or more parameters. Since X is a random variable it follows that Y is also a random variable and we are interested in the properties of Y. The question most often asked is, "What are the mean and variance of Y?"

The approximate answer to this question is usually obtained by expanding g(X) in a Taylor series about E(X), truncating the series appropriately and taking the expected value of the truncated series with respect to the random variable X.[1,2]

In the case of one independent variable, if derivatives of all order exist, it follows that

$$Y = \sum_{r=0}^{\infty} \frac{1}{r!} g^{(r)}(\mu)(X-\mu)^r \tag{4}$$

where $E(X) = \mu$ and $g^{(r)}(\mu)$ is the r^{th} derivative of g(X) evaluated at the point $X = \mu$.

If we define $\mu_r = E(X-\mu)^r$; $r = 0, 1, 2, \ldots$; we can write

$$E(Y) = \sum_{r=0}^{\infty} \frac{1}{r!} g^{(r)}(\mu)\mu_r. \tag{5}$$

It then follows that

$$Y - E(Y) = \sum_{r=0}^{\infty} \frac{1}{r!} g^{(r)}(\mu)[(X-\mu)^r-\mu_r]$$

and

$$\mathrm{Var}(Y) = E[Y-E(Y)]^2 = \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{1}{r!} \frac{1}{s!} g^{(r)}(\mu) g^{(s)}(\mu) \; E[(x-\mu)^r-\mu_r][(x-\mu)^s-\mu_s]$$

$$= \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \frac{1}{r!} \frac{1}{s!} g^{(r)}(\mu) g^{(s)}(\mu) [\mu_{r+s} - \mu_r \mu_s] \qquad (6)$$

since $\mu_0 = 1$ and $\mu_1 = 0$.

Equations (5) and (6) give the mean and variance of $Y = g(X)$ in terms of the higher moments of X. Similar relationships can be easily derived for the higher moments of Y.

For purpose of illustration, let us apply equations (5) and (6) to find the mean and variance of the quadratic model of equation (1).

$$Y = g(X) = a + bX + cX^2$$

$$g'(X) = b + 2cX$$

$$g''(X) = 2c$$

$$g^{(r)}(X) = 0 \quad \text{for } r \geq 3$$

It follows from equation (5) that

$$E(Y) = a + b\mu + c\mu^2 + c\sigma^2 \qquad (7)$$

where $\sigma^2 = \mu_2$ is the variance of X. The usual estimate of the expected value of Y is given by $E(Y) \approx g(\mu) = a + b\mu + c\mu^2$. This estimate is in error by $c\sigma^2$, which may or may not be negligible.

It follows from equation (6) that

$$\mathrm{Var}(Y) = (b+2c\mu)^2\sigma^2 + 2c(b+2c\mu)\mu_3 + c^2(\mu_4-\sigma^4) \qquad (8)$$

so that the usual estimate of the variance of Y, $\mathrm{Var}(Y) \approx [g'(\mu)]^2\sigma^2 = (b + 2c\mu)^2\sigma^2$ is in error by $2c(b + 2c\mu)\mu_3 + c^2(\mu_4 - \sigma^4)$. If the distribution of X is symmetric about μ, then $\mu_3 = 0$ and the error reduces to $c^2(\mu_4 - \sigma^4)$. Taking our analysis one step further, we note that if X has a normal distribution, or approximately so via the central limit theorem, then

$$\mu_r = E(X-\mu)^r = 0 \qquad ; \ r \text{ odd}$$

$$= \frac{\sigma^r r!}{2^{r/2}(r/2)!} \qquad ; \ r \text{ even}$$

and $\mathrm{Var}(Y) = (b + 2c\mu)^2\sigma^2 + 2c^2\sigma^4$ so the error becomes $2c^2\sigma^4$.

We next consider the usual case in which the parameters a, b and c are random variables. This occurs when observations are made on X and Y and these observations are used to estimate the parameters.

Consider the random vector (a,b,c)' which is independent of the random variable X (the input to our predictive model). Let

$$E[(a,b,c)'] = (A,B,C)'$$

and

$$\mathrm{Cov}[(a,b,c)'] = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} & \sigma_{ac} \\ \sigma_{ab} & \sigma_b^2 & \sigma_{bc} \\ \sigma_{ac} & \sigma_{bc} & \sigma_c^2 \end{bmatrix}$$

If we use least squares to fit the parameters to the data and assume independent normally distributed residuals, then we can calculate good estimates of the vector of means and the covariance matrix.

In this case

$$E(Y) = E(a + bX + cX^2) = A + B\mu + C(\sigma^2 + \mu^2) = A + B\mu + C\sigma^2 + C\mu^2, \qquad (9)$$

which is precisely the same as when the parameters were not random variables provided that they are unbiased estimates of the true coefficients.

Similarly

$$\begin{aligned} \mathrm{Var}(Y) &= \sigma_a^2 + \sigma_b^2\,\mathrm{Var}(X) + \sigma_c^2\,\mathrm{Var}(X^2) \\ &\quad + 2\,\mathrm{Cov}(a, bX) + 2\,\mathrm{Cov}(a, cX^2) \\ &\quad + 2\,\mathrm{Cov}(bX, cX^2) \\ &= \sigma_a^2 + \sigma_b^2\,\sigma^2 + \sigma_c^2\,\mathrm{Var}(X^2) + \text{covariance terms.} \end{aligned}$$

The covariance terms in the above relationship are calculated as follows:

$$\mathrm{Cov}(a, bX) = E(abX) - E(a)E(bX) = \mu\sigma_{ab}$$

$$\mathrm{Cov}(a, cX^2) = E(acX^2) - E(a)E(cX^2) = \sigma_{ac}(\mu^2 + \sigma^2)$$

$$\mathrm{Cov}(bX, cX^2) = E(bcX^3) - E(bX)E(cX^2)$$

$$= \sigma_{bc}(\mu_3 + 3\mu\sigma^2 + \mu^3) + BC\mu_3 + 2BC\mu\sigma^2.$$

Substituting these expressions in the above relationship leads to

$$\mathrm{Var}(Y) = \sigma_a^2 + \sigma_b{}^2\sigma^2 + \sigma_c^2(\mu_4 + 4\mu\mu_3 + 4\mu^2\sigma^2 - \sigma^4) + 2\mu\sigma_{ab}$$
$$+ 2\,\sigma_{ac}(\mu^2 + \sigma^2) + 2\,\sigma_{bc}(\mu_3 + 3\mu\sigma^2 + \mu^3)$$
$$+ 2\,BC(\mu_3 + 2\,\mu\sigma^2), \tag{10}$$

which is quite different from the usual estimate.

Equations of the form of equation (10) should prove useful for determining sample sizes required for experiments to collect data for estimation of the parameters of a mathematical model which will be used for predictive purposes.

3. TWO OR MORE FUNCTIONS OF ONE RANDOM VARIABLE.

Let $Y_1 = g_1(X)$ and

$Y_2 = g_2(X)$

$\vdots$

$Y_k = g_k(X)$,

that is, we have $k > 2$ functions of one random variable.

In this case we use equations (5) and (6) to find $E(Y_i)$ and $\mathrm{Var}(Y_i)$ for all $i = 1, 2, \ldots, k$. The only undetermined moments are $\mathrm{Cov}(Y_i, Y_j)$, $i \neq j$.

We note that

$$Y_i - E(Y_i) = \sum_{r=0}^{\infty} \frac{1}{r!} g_i^{(r)}(\mu)[(X-\mu)^r - \mu_r] \quad \text{and}$$

$$Y_j - E(y_j) = \sum_{s=0}^{\infty} \frac{1}{s!} g_j^{(s)}(\mu)[(X-\mu)^s - \mu_s].$$

Thus

$$\mathrm{Cov}(Y_i, Y_j) = E[\,Y_i - E(Y_i)][Y_j - E(Y_j)]$$

$$= \sum_{r,s=0}^{\infty} \frac{1}{r!s!} g_i^{(r)}(\mu) g_j^{(s)}(\mu) E[(X-\mu)^{r+s} - \mu_r(X-\mu)^s - \mu_s(X-\mu)^r + \mu_r\mu_s]$$

$$= \sum_{r,s=0}^{\infty} \frac{1}{r!s!} g_i^{(r)}(\mu) g_j^{(s)}(\mu)[\mu_{r+s} - \mu_r\mu_s].$$

If we recall that $\mu_o = E(X-\mu)^o = 1$ and $\mu_1 = E(X-\mu) = 0$, we see that for $r = 0$ or $s = 0$ $\mu_{r+s} - \mu_r\mu_s = 0$ so we can write

$$\mathrm{Cov}(Y_i, Y_j) = \sum_{r,s=1}^{\infty} \frac{1}{r!s!} g_i^{(r)}(\mu) g_j^{(s)}(\mu)[\mu_{r+s} - \mu_r\mu_s]. \tag{11}$$

4. ONE FUNCTION OF TWO OR MORE RANDOM VARIABLES. The problem of one function of two or more random variables is treated completely analogously to that of one function of one random variable. That is, we expand the function in a Taylor series about the vector of means of the random input vector and truncate appropriately. Before proceeding with this, we state the following theorem which can be found in reference 3:

Theorem 6.2 (Apostle)

Let g have continuous partial derivatives of order m at each point of an open set S of E_n. If a and b are both elements of S, $a \neq b$ and the line segment joining a and b lies in S, then there exists a point z on the line segment L(a,b) such that

$$g(a) - g(b) = \sum_{k=0}^{m-1} \frac{1}{k!} d^k g(a;b-a) + \frac{1}{m!} d^m g(z;b-a)$$

where:

$$a = (a_1, a_2 \ldots, a_n)' ,$$

$$b = (b_1, b_2, \ldots, b_n)'$$

$$X = (X_1, X_2, \ldots, X_n)',$$

$$d^k g(a;b-a) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_k=1}^{n} D_{i_1 i_2 \ldots i_k} g(a)\; t_{i_1} t_{i_2} \ldots t_{i_k} ,$$

$$t_{i_j} = b_{i_j} - a_{i_j}$$

and $$D_{i_1 i_2 \ldots i_k} = \frac{\partial^k g(X)}{\partial X_{i_1} \partial X_{i_2} \ldots \partial X_{i_k}}$$

We will use the following theorem from reference 3 to simplify expressions in our series expansions.

<u>Theorem</u>: If $D_i g(X)$, $D_j g(X)$ and $D_{ij} g(X)$ are continuous in a neighborhood of the point (X_i, X_j) in E_2, then $D_{ji} g(X_1,X_2)$ exists and $D_{ij}\, g(X_1,X_2) = D_{ji}\, g(X_1,X_2)$.

Before using the relationships given in this section one should verify that the function g(X) satisfies the conditions of the above two theorems.

We consider the case where n = 2, that is, Y = g(X) is a function of two random variables.

Let $X = (X_1, X_2)'$, $E(X) = (\mu_1, \mu_2)'$

and $$Cov(X) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$ and define

$$\mu_{r,s} = E[(X_1-\mu_1)^r (X_2-\mu_2)^s].$$

Expanding g(X) about the point $X = \mu$ in a Taylor series leads to

$$Y = g(X) = g(X_1, X_2) \doteq g(\mu) + \sum_{i=0}^{m-1} \frac{1}{k!} d^k g\ (\mu; X-\mu)$$

where we neglect the remainder term, $R = \frac{1}{m!} d^m g(z; X-\mu)$, $z \in L(X,\mu)$. Introducing the notation

$$g_{r,s}(\mu) = \left. \frac{\partial^{r+s} g(X)}{\partial X_1^r\ \partial X_2^s} \right|_{X=\mu}$$

and using the fact that $g_{r,s}(\mu) = g_{s,r}(\mu)$, our expansion becomes

$$Y = g(X) \doteq \sum_{r=0}^{m-1} \frac{1}{r!} \sum_{s=0}^{r} \binom{r}{s} g_{s,r-s}(\mu)(X_1-\mu_1)^s (X_2-\mu_2)^{r-s} \tag{12}$$

The expected value of Y is approximately

$$E(Y) = E[g(X)] \doteq \sum_{r=0}^{m-1} \frac{1}{r!} \sum_{s=0}^{r} \binom{r}{s} g_{s,r-s}(\mu)[\mu_{s,r-s.}] \tag{13}$$

Squaring equation (12) gives

$$Y^2 = \sum_{r=0}^{m-1} \sum_{u=0}^{m-1} \frac{1}{r!} \frac{1}{u!} \sum_{s=0}^{r} \sum_{v=0}^{u} \binom{r}{s}\binom{u}{v} g_{s,r-s}(\mu) g_{v,u-v}(\mu)(X_1 - \mu_1)^{s+v}(X_2 - \mu_2)^{r+u-s-v},$$

and it follows that

$$E(Y^2) \doteq \sum_{r=0}^{m-1} \sum_{u=0}^{m-1} \frac{1}{r!u!} \sum_{s=0}^{r} \sum_{v=0}^{u} \binom{r}{s} \binom{u}{v} g_{s,r-s}(\mu) g_{v,u-v}(\mu) \mu_{s+v,r+u-s-v} \tag{14}$$

Squaring equation (13) and subtracting the result from equation (14) leads to

$$\mathrm{Var}(Y) = E(Y^2) - [E(Y)]^2$$

$$\doteq \sum_{r=1}^{m-1} \sum_{u=1}^{m-1} \frac{1}{r!u!} \sum_{s=0}^{r} \sum_{v=0}^{u} \binom{r}{s}\binom{u}{v} g_{s,r-s}(\mu) g_{v,u-v}(\mu)$$

$$[\mu_{s+v,r+u-s-v} - \mu_{s,r-s}\mu_{v,u-v}] \tag{15}$$

To illustrate the use of the above, we consider the case in which X has a bivariate normal distribution. Here, $\mu_{r+s} = 0$ for r+s odd, and all other values of μ_{r+s} depend only on μ_1, μ_2, σ_1^2, σ_2^2 and σ_{12}. In this case

$$E(Y) \doteq g(\mu) + \frac{1}{2}g_{20}(\mu)\sigma_1^2 + 2g_{11}(\mu)\sigma_{12} + \frac{1}{2}g_{02}(\mu)\sigma_2^2$$
$$+ \frac{1}{8}g_{40}(\mu)\sigma_1^4 + \frac{1}{2}g_{31}(\mu)\sigma_1^2\sigma_{12} + \frac{3}{4}g_{22}(\mu)[\sigma_1^2\sigma_2^2 + 2\sigma_{12}^2]$$
$$+ \frac{1}{2}g_{13}(\mu)\sigma_2^2\sigma_{12} + \frac{1}{8}g_{04}(\mu)\sigma_2^4 \,. \tag{16}$$

and

$$\mathrm{Var}(Y) \doteq \sigma_1^2[g_{10}(\mu)^2] + \sigma_{12}[2g_{10}(\mu)g_{01}(\mu)] + \sigma_2^2[g_{01}(\mu)^2]$$
$$+ \sigma_1^4[g_{10}(\mu)g_{30}(\mu) + \frac{1}{2}g_{20}(\mu)^2]$$
$$+ \sigma_2^4[g_{01}(\mu)g_{03}(\mu) + \frac{1}{2}g_{02}(\mu)^2]$$
$$+ \sigma_1^2\sigma_2^2[g_{01}(\mu)g_{21}(\mu) + g_{10}(\mu)g_{12}(\mu) + g_{11}(\mu)^2]$$
$$+ \sigma_1^2\sigma_{12}[g_{01}(\mu)g_{30}(\mu) + 3g_{10}(\mu)g_{21}(\mu) + 2g_{11}(\mu)g_{20}(\mu)]$$
$$+ \sigma_2^2\sigma_{12}[g_{10}(\mu)g_{03}(\mu) + 3g_{01}(\mu)g_{12}(\mu) + 2g_{11}(\mu)g_{02}(\mu)]$$
$$+ \sigma_{12}^2[2g_{01}(\mu)g_{21}(\mu) + 2g_{10}(\mu)g_{12}(\mu) + g_{02}(\mu)g_{20}(\mu) + g_{11}(\mu)^2] \tag{17}$$

Though equations (16) and (17) may be lengthy, they are straightforward and can be easily evaluated on a computer.

Equations (13) and (15) readily generalize to the case of a function of k > 2 variables. Since the extended equations are messy and notationally complex, this effort will be delegated as a chore for the reader.

5. SEVERAL FUNCTIONS OF MORE THAN ONE VARIABLE.

Let $Y = [Y_1, Y_2, \ldots, Y_p]'$

$$= [g^{(1)}(X), g^{(2)}(X), \ldots, g^{(p)}(X)]'$$

where X is a random vector with properties given in the previous section and $Y_r = g^{(r)}(X)$. Equations (13) and (15) give approximations for $E(Y_i)$ and $Var(Y_i)$. It remains for us to find $Cov(Y_r, Y_u)$.

The argument that was used to derive equation (15) leads immediately to

$$Cov(Y_r,Y_u) \doteq \sum_{r=1}^{m-1} \sum_{u=1}^{m-1} \frac{1}{r!u!} \sum_{s=0}^{r} \sum_{v=0}^{u} \binom{r}{s}\binom{u}{v} g^{(r)}_{s,r-s}(\mu) g^{(u)}_{v,u-v}(\mu) \qquad (18)$$

$$[\mu_{s+v,r+u-s-v} - \mu_{s,r-s}\mu_{v,u-v}].$$

6. SUMMARY AND CONCLUSIONS.

Error propagation in mathematical models is simply another name for the study of one or more functions of one or more random variables. The ideal treatment of this problem is to derive the actual joint distribution of the new set of random variables which are transformations of the original set. Unfortunately this problem is unsolved for all but a relatively few commonly used functions.

Common practice for these unsolved problems is to approximate the means, variances and covariances of the new variables and use these approximations to construct error bars.

We have given procedures for refining the most commonly used approximations. Our equations become a bit awkward for random vectors of high dimension but they remain straightforward and manageable on a computer.

The procedures given herein should be useful for evaluating approximations currently in use and they promise to furnish a vehicle for developing improved approximations for special classes of functions, e.g., convex functions.

7. ACKNOWLEDGEMENT.

The authors express their gratitude to Ms. Ann McKaig for evaluating the higher moments of the bivariate normal distribution which were used to develop equations (16) and (17).

REFERENCES

1. Ku, H.H., "Notes on the Use of Propagation of Error Formulas," Journal of Research of the National Bureau of Standards - C. Engineering and Instrumentation, Vol. 70C, No. 4, (1966).

2. Meyer, Stuart L., Data Analysis for Scientists and Engineers, John Wiley and Sons (1975).

3. Apostol, Tom M., Mathematical Analysis, A Modern Approach to Advanced Calculus, Addison-Wesley Publishing Company, Inc., (1964).

On Regenerative Processes In Discrete Time

Walter L. Smith

University of North Carolina, Chapel Hill

In this paper we shall present a somewhat heuristic account of results we have been obtaining which we hope will prove to be of value in applying probabilistic ideas to many stochastic problems which may arise in practice.

We begin by explaining the idea of a renegerative stochastic process. First we must understand what is meant by a tour, a typical one of which we shall denote $\mathcal{T}$. For the present discussion, in which we shall be confining ourselves to events occurring on a discrete time scale, $\mathcal{T}$ will consist of

(a) An integer X called the duration of the tour.

(b) A real-valued function T(t) defined for $t=0,1,2,\ldots,X$; this is called the graph of the tour.

(c) A final value or residue, denoted Y, which can be regarded as T(X).

We imagine a population of tours, from which repeated sampling will produce an iid sequence $\mathcal{T}_1$, $\mathcal{T}_2,\ldots,$ and so on. The successive durations are X_1, $X_2,\ldots,$ and it is clear that they form a recurrent event process on the discrete time scale. We call this recurrent event process the bed for the cumulative process. Let us denote the probability distribution of the $\{X_n\}$ by f_1, f_2, ..., etc., noting that we shall always assume that $P\{X_n = 0\} = 0$. We can then call into play the pgf $F(z) = \sum_1^\infty z^n f_n$, which will converge for small enough $|z|$. For the present exposition we shall assume there is a $\zeta > 1$ such that $F(\zeta) < \infty$. This will ensure that $F(z)$ is analytic in the open disk $|z| < \zeta$.

We then build up a process on the infinite discrete time scale by concatenating the tours $\mathcal{T}_1, \mathcal{T}_2, \ldots$. Let us write $S_n = X_1 + X_2 + \cdots + X_n$ for the time at which $\mathcal{T}_{n+1}$ begins (so that $S_0 = 0$). Then if $S_n \leq t < S_{n+1}$ we can define the value of the regenerative process at t by $T_n(t - S_n)$, where T_n stands for the graph of $\mathcal{T}_n$. In this conection it will be convenient to write $N(t)$ for the largest integer k for which $S_k \leq t$. We shall then write $T^*_t(\cdot)$ for the graph of $\mathcal{T}_{N(t)+1}$ (which is the tour operating at time t). Similarly we shall write $S^*(t)$ for $S_{N(t)}$.

It will also help if we write Y_j for the residue of $\mathcal{T}_j$. We can then define a cumulative process $W(t)$ as follows.

$$W(t) = \sum_{j \leq N(t)} + T^*_t(t - S^*(t)).$$

In words, $W(t)$ is the sum of all residues of tours which ended prior to time t, together with the current value of the regenerative process (as described above).

In all really interesting applications, X_j and Y_j are not independent; of course, the very nature of the underlying sampling procedure ensures that the successive vectors (X_n, Y_n) are independent.

The main question which we try to answer in a study of such cumulative processes concerns the asymptotic behavior of $W(t)$, as t grows without bound. It is possible to define multivariate versions of all that we have described, and it is possible to obtain results in the more general case, but we shal eschew this generality in the present exposition.

A very useful tool in tackling $W(t)$ is the characteristic function, defined for real values of the dummy θ, by the equation

$$\Phi(\theta,t) = \mathbb{E} \exp[i\theta\, W(t)]$$

It then turns out that one needs the following functions

$$\psi(\theta) = \mathbb{E} \exp\, [i\theta Y_1)$$

$$G(z,\theta) = \mathbb{E}\, z^{X_1} \exp\, [i\theta Y_1]$$

$$R(z,\theta) = \mathbb{E} \textstyle\sum^{*} z^{m} \exp\, [i\theta T_1(m)],$$

where the sum $\sum^{*}$ extends over the range $m = 0$ to $m = X_1 - 1$.

The underlying model should, in many cases, allow the calculation of these various functions we have just defined. However, the determination of $\Phi(\theta,t)$ is difficult. What is possible is to determine the generating function

$$\Phi^{o}(\theta,z) = \textstyle\sum_{t=0}^{\infty} z^{t}\, \Phi(\theta,t).$$

A certain amount of fairly straightforward reasoning will then lead to the neat result

$$\Phi^{o}(\theta,z) = R(z,\theta)/[\,1 - G(z,\theta)]$$

There are two special cases of particular interest.

CLASS A

In this class it is supposed that $T_1(t) = t$ for all $0 \le t \le X_1$. This corresponds to the situation where Y_1 is a cost incurred by the tour, and this cost has to be paid at the instant the tour begins.

For a Class A process it is easy to show that

$$R(z,\theta) = [\psi(\theta) - G(z,\theta)]/[1 - z]$$

CLASS B

Here it is assumed that $T_1(t) = 0$ for all $t < X_1$, although it may well be that Y_1 differs from 0. This corresponds to the situation when there is a payment for a tour, but it is only made at the end of that tour.

This case gives a particularly simple result

$$r(z,\theta) = [1 - F(z)]/[1 - z]$$

A Class B process of great importance, and much studied in its own right, arises when every residue has the value unity. In this case the cumulative process $W(t) = N(t)$, and is the "renewal count."

For the renewal count process one can show that

$$\Phi^{o}(\theta,z) = [1 - F(z)]/[1-z][1 - e^{i\theta}F(z)].$$

It transpires that one can prove an identity for Class A processes, which parallels closely the famous Fundamental Identity of Abraham Wald in sequential analysis. Actually the argument here is much easier than it is in sequential analysis. The identity we can get is

$$E \frac{e^{i\theta W(t)}}{[\psi(\theta)]^{N(t)+1}} = 1$$

Unfortunately there is no correspondingly elegant result for the Class B process (or, indeed, the general cumulative process)

A trick that M.S.Bartlett exposed in his contribution to a Symposium on Stochastic Processes (Bartlett, 1949) can be used to extract information from this Fundamental Identity. It consists in looking for a $\zeta \neq 0$ for which $\psi(\zeta) = 1$. For example, if we suppose that $\psi(\theta) = (\lambda/\lambda - i\theta)^2$ we find that $\zeta = -2i\lambda$ and are led from the Fundamental Identity to the result

$$E\, e^{2\lambda W(t)} = 1,$$

from which it is easy to infer the inequality

$$P(W(t) \geq x) \leq e^{-2\lambda x}.$$

To see how easily the identity may be derived, let us choose any θ such that $\psi(\theta) \neq 0$, and set

$$\tilde{Y}_j = Y_j - \frac{\log \psi(\theta)}{i\theta}$$

If we use $\tilde{Y}_j$ in place of Y_j in the formula for the Class A process, then we find that

$$\tilde{\Phi}^0(\theta,z) = \frac{\tilde{\psi}(\theta) - \tilde{G}(z,\theta)}{(1-z)(1-\tilde{G}(z,\theta))} .$$

But it is a simple matter to show that $\tilde{\psi}(\theta) = 1$, from which it follows at once that $\tilde{\Phi}^0(\theta,z) = 1/(1-z)$. This obviously implies that $\tilde{\Phi}(\theta,t) = 1$ for all $t = 0,1,2,\ldots$. The derivation of the Fundamental Identity is now a simple matter.

Let us now consider a particular example of the general case which is neither a Class A nor B. Suppose that each residue Y is $N(0, 1+X)$. This represents a situation where there is a unit expected set-up cost for each tour and then a random extra cost whose expectation, conditional on the duration X of the tour, is X. For this case we can see that

$$G(z,\theta) = e^{-\theta^2/2} F(ze^{-\theta^2/2}).$$

To understand the behavior of $\Phi(\theta,t)$ we need to look at the singularities in the complex z-plane of $\Phi^0(\theta,z)$. These plainly occur where $G(z,\theta) = 1$, and, in the present special example, this is where

$$F(ze^{-\theta^2/2}) = e^{\theta^2} \tag{1}$$

Let us choose a small $\delta > 0$ and assume that $|\theta| \leq \delta$. Then (1) will have a solution $z = \zeta(\theta)$, say, which tends to zero as θ tends to zero, and there will be a $\epsilon > 1$ such that $\zeta(\theta)$ is the only zero of $1 - G(z,\theta)$ in the circle $|z| \leq \epsilon$.

To carry the example further, let us be specific about the distribution of the durations X_n. Let us suppose that X_n has a geometric distribution such with pgf

$$F(z) = pz/(1 - qz).$$

Then we shall find that

$$\zeta(\theta) = \frac{e^{\theta^2/2}}{q + pe^{-\theta^2/2}}$$

Notice that that this formula shows that $\zeta(\theta) \to 1$ as $\theta \to 0$ and $\to \infty$ as $\theta \to \infty$.

Our purpose in examining this very special example is to show the sort of behavior to be expected in the general situation. Indeed, careful complex analysis shows this sort of behavior to be a feature of the general case; but it is felt that, for a simple exposition such as the present, an illustration using a special case will be more acceptable to the reader.

Let us now resume the discussion of the most general situation and let C_2 be the circle $|z| = \epsilon > |\zeta(\theta)|$, and let C_1 be the circle $|z| = \rho < |\zeta(\theta)|$, interior to C_2. Then by the usual function-theory argument we are led to the result

$$\Phi(\theta,t) = \frac{1}{2\pi i}\int_{C_1}\left\{\frac{R(z,\theta)}{1 - G(z,\theta)}\right\}\frac{dz}{z^{t+1}}$$

Suppose we now allow ρ to grow so that the contour C_1 passes across the pole at $\zeta(\theta)$. In doing this we shall increase the value of the contour integral by the residue X, say, of the function

$$\left\{\frac{R(z,\theta)}{1 - G(z,\theta)}\right\}\frac{1}{z^{t+1}}$$

at the pole at $\zeta(\theta)$. It then apppears that we shall have

(2) $$\Phi(\theta,t) = -X + \frac{1}{2\pi i}\int_{C_2} \text{etc.}$$

But it is not too difficult to show that the integral in (2) is $O(\epsilon^{-t})$; this integral will thus go to zero <u>much</u> faster than any terms we shall retain, and so from now on we shall ignore it. The remaining problem is to determine the value of X, which, of course, will be a function of t. Let us write $G_z(\zeta(\theta),\theta)$ to denote

$$\left[\frac{\partial}{\partial z} G(z,\theta) \right]_{z = \zeta(\theta)}$$

Then we shall find that

$$\Phi(\theta,t) = \frac{R(\zeta(\theta),\theta)}{G_z(\zeta(\theta),\theta)} [\zeta(\theta)]^{-(t+1)} + \text{negligible terms}$$

At this stage it is convenient to introduce a function $A(\theta)$ with the property that $\zeta(\theta) = \exp A(\theta)$; thus $A(\theta) \rightarrow 0$ as $\theta \rightarrow 0$. This function $A(\theta)$ is very important, and so we shall give it a careful examination. It is possible to show that it will have a Taylor series expansion

$$[A(\theta)\mu_{10} + (i\theta)\mu_{01}] + (1/2)[(a(\theta))^2\mu_{20} + 2a(\theta)(i\theta)\mu_{11} + (i\theta)^2\mu_{02}] + \text{etc} = 0$$

It is possible, though with increasing complexity, to compute the $\langle A_n \rangle$ in terms of "known" moments. The calculation is based on the equation

$$G(e^{A(\theta)},\theta) = 1,$$

and this leads to the following relation,

$$[A(\theta)\mu_{10} + (i\theta)\mu_{01}] + (1/2)[(A(\theta))^2\mu_{20} + 2A(\theta)(i\theta)\mu_{11} + (i\theta)^2\mu_{02}] + \text{etc} = 0$$

There seems no way of escaping a considerable amount of tedious computation at this point. One must equate coefficients of powers of θ to zero and determine the $\langle A_n \rangle$ from the resulting equations. For $n > 5$ the formulae for these $\langle A_n \rangle$

become very large. Here we shall content ourselves with quoting the only the formulae for A_1 and A_2. These are as follows

$$A_1 = -\frac{\mu_{01}}{\mu_{10}}$$

$$A_1 = -\frac{\mu_{01}}{\mu_{10}} - \frac{\mu_{20}(\mu_{01})^2}{(\mu_{10})^3} - \frac{2\mu_{11}\mu_{01}}{(\mu_{10})^2}$$

Let us now set

$$\tau^2 = E\left(Y_j - \frac{\mu_{01}}{\mu_{10}} X_j\right)^2$$

Then it is possible to show that a much simpler formula exists for A_2,

$$A_2 = -\tau^2 / \mu_{10}.$$

In terms of the quantities we have now introduced we thus have the equation

$$E\, e^{i\theta W(t)} = \left\{ \frac{R(\zeta(\theta),\theta)}{G_z(\zeta(\theta),\theta)} \right\} e^{-(t+1)[A_1(i\theta) + (1/2)A_2(i\theta)^2 + \dots]}$$

and this will hold for all $|\theta| \leq \delta$, for some small $\delta > 0$.

But for <u>any</u> θ, if t is large enough, we shall find that

$$\left| \frac{\theta}{\sqrt{(t+1)}} \right| \leq \delta$$

Thus, if we replace θ by $\theta/\sqrt{(t+1)}$, we shall find that

$$(3) \qquad E \exp i\theta \left\{ \frac{W(t) + (t+1)A_1}{\sqrt{(t+1)}} \right\}$$

$$= \left\{ \frac{R(\zeta(\theta/\sqrt{(t+1)}),\theta/\sqrt{(t+1)})}{G_z(\zeta(\theta/\sqrt{(t+1)}),\theta/\sqrt{(t+1)})} \right\} e^{-(1/2)A_2(i\theta)^2 - A_3(i\theta)^3/(6\sqrt{[t+1]})}$$

+ negligible terms.

Various deductions can be made from (3). Let us first examine the dominant terms in this equation. It is not hard to show that

$$\left\{ \frac{R(\zeta(\theta/\sqrt{(t+1)}),\theta/\sqrt{(t+1)})}{G_z(\zeta(\theta/\sqrt{(t+1)}),\theta/\sqrt{(t+1)})} \right\} \to 1 \text{ as } t \to \infty.$$

This will therefore imply from (3) that

$$E \exp i\theta \left\{ \frac{W(t) + (t+1)A_1}{\sqrt{(t+1)}} \right\} = Z(t) \text{ ,say,}$$

is asymptotically $N(0, \tau^2/\mu_{10})$.

This Central Limit Theorem is useful. It can be, and has been, proved without the details in which we have become embroiled; but these details are necessary if we want to obtain the more delicate results.

With a great deal of tedious computation, one can obtain Edgeworth-type improvements, in the form of series expansions, in place of the single dominant normal density term. Evidently we can write

$$G_z(z,\theta) = Ez^X e^{i\theta Y},$$

so that

$$G_z(e^{A(\theta)},\theta) = \mu_{10} + (i\theta)[A_1(\mu_{20} - \mu_{10}) + \mu_{11}] + \text{etc...}$$

However, once one asks for more than the dominant term, the function $R(e^{A(\theta)},\theta)$ has to be taken into consideration. In this connection some curious moment-like numbers

$$\rho_{rs} = E \sum_{j=0}^{X_1-1} j^r (T_1(j))^s$$

emerge. They should be determinable from the known specification of the underlying model. If one makes use of these numbers one is led to an expansion like the following (which we quote to only one term beyond the dominant one)

$$\left\{ \frac{R(\zeta(\theta/\sqrt{(t+1)}),\theta/\sqrt{(t+1)})}{G_Z(\zeta(\theta/\sqrt{(t+1)}),\theta/\sqrt{(t+1)})} \right\} = 1 + \frac{(i\theta)C_0}{\sqrt{(t+1)}} .$$

Of course, nothing but hard work prevent ones from obtaining as many extra terms in this expansion as one wishes. In the above abbreviated expansion

$$C_0 = -\frac{\mu_{11}}{\mu_{10}} + \frac{\mu_{20}\mu_{01}}{2(\mu_{10})^2} - \frac{\mu_{01}}{2\mu_{10}} + \frac{\rho_{01}}{\mu_{10}} .$$

If we now write v in place of τ^2/μ_{10}, we are thus led to the result

$$E\, e^{i\theta Z(t)} = e^{-v\theta^2/2} \left\{ 1 + \frac{C_0(i\theta)}{\sqrt{(t+1)}} - \frac{A_3(i\theta)^2}{6\sqrt{(t+1)}} + O(1/t) \right\} .$$

From this last result one can thus infer that asymptotic probability statements about Z(t) will be better if, instead of a simple normal density, we use the "density"

$$\frac{e^{-x^2/2v}}{\sqrt{(2\pi v)}}\left\{1+\frac{C_1x^2-C_3x^3}{\sqrt{(t+1)}}+\ldots\right\},$$

where

$$C_1 = vC_0 + (1/2)A_3v^3,$$
$$C_3 = (1/6)A_3v^3$$

In deriving this result we have used the fact that multiplying a Fourier Transform by $(i\theta)$ corresponds to differentiating the original function of x, with repsect to x.

It is possible to extend all the ideas we have discussed, and derive roughly similar results, when we replace the simple renewal-type "bed" of the cumulative process with a "Semi-Markov" one. At the conference, some brief discussion was given of this latter more general model. However, we shall not discuss it here, preferring to delay an exposition until somewhat more comprehensive results have been obtained.

REFERENCES

Bartlett, M.S. (1949), "Some evolutionary stochastic processes," J. R. Statist. Soc., B, 11, 211-229.

Gupta, Y.P. (1976), "Joint distributions and asymptotic values of the moments for Markov renewal processes in space," J. Indian Statist. Assn., 14, 23-38.

Khirasaga, A.M. & Gupta, Y.P. (1967), "Asymptotic values of the first two moments in Markov renewal processes," Biometrika, 54, 597-603.

------- & -------- (1969), "Some results in Markov renewal processes," Calcutta Stat. Assoc. Bull., 18, 61-72.

Smith, W.L. (1955), "Regenerative stochastic processes," Proc. Roy. Soc., A, 232, 6-31.

------- (1958), "Renewal theory and its ramifications," J. R. Statist. Soc., B, 20, 243-302.

ERRATA

The following statments,supplied by the author of two articles appearing in earlier issues of these "Proceedings", corrects certain typographical errors.

ARO Report 77-2 (Proc. 22nd conf on DOE)

page 263. Footnote should read "*Quartiles, $r = 0.697$, 0.799, and 0.879....."

page 290. The expressions under the two radical signs are identical. Only the sign which precedes the radical is different.

page 294. Formula (D-5) should read

$$T_{i+1} = \frac{m-(k+i)}{k+i+1} \cdot \frac{(n+1) - (k+i)}{(N-m-n) + (k+i+1)} \cdot T_i$$

ARO Report 81-2 (Proc. 26th conf on DOE)

page 227. Second line from bottom should read

$$= 1 - \frac{2}{3}\,px - \frac{1}{9}\,p^2x^2 - \frac{4}{81}\,p^3x^3$$

27th CONFERENCE ON THE DESIGN OF EXPERIMENTS

Attendance Sheet

NAME	AFFILIATION	TELEPHONE #
1. Patrick D. Allen	CAA	(301)295-5297
2. William E. Baker	BRL	AV 283-4980
3. Carl B. Bates	CAA	AV 295-1542
4. Virginia Bates	AVRADCOM	AV 693-1575
5. Robert Bechhofer	Cornell Univ.	
6. Ray Bell	AMSAA	(301)278-5882
7. Henry Betz	AMSAA	AV 283-2135
8. COL Andrew P. Blasco	Dept/Math/USMA	(914)938-2407
9. Barry Bodt	BRL	AV 283-4482
10. Kurt F. Buckley	ARMTE, WSMR	AV 258-1242
11. Donald Burdick	Duke Univ.	
12. Howard Burnette	USALOGC	AV 687-2360
13. Oscar J. Castro	WSMR	AV 358-2181
14. Jagdish Chandra	ARO	AV 935-3331
15. John Chapman	NVEOL	AV 354-3765
16. John R. Crigler	NSWC	(703)663-7892
17. Larry Crow	AMSAA	(301)278-5882
18. Paul J. Deason	TRASANA	AV 258-1551
19. Francis Dressel	ARO	AV 935-3331
20. Bill H. Dunn	EPG	(602)538-6418
21. Rubright Earl	OTEA	AV 289-1088
22. Walter D. Foster	AFIP	AV 281-3181
23. Clarke J. Fox	AMSAA	AV 283-4291
24. Bret Graham	CAA	AV 295-5237
25. George M. Hanna	AMSAA	AV 283-3277

NAME		AFFILIATION	TELEPHONE #
26.	Bernard Harris	Univ. of Wisconsin	(608)262-2614
27.	Larry Hatch	NSA	(301)796-6461
28.	Jerry Highfill	EPA-HERL	(919)541-4068
29.	David C. Hoaglin	Harvard Univ.	(617)495-2583
30.	Robert S. Jackson	DARCOM ITC	AV 829-2501
31.	Norman L. Johnson	Univ. of North Carolina	(919)962-2187
32.	Oran Johnston	USASC&FG	AV 780-3709
33.	James R. Knaub, Jr.	TRADOC (Ft. Lee)	AV 687-3063
34.	Eugene O. Kessler	TRADOC	AV 737-9017
35.	Pierre Ladouceur	Nat't Defense	AV 672-5878
36.	Robert Launer	ARO	AV 935-3331
37.	Freeman G. Lee	Wright-Pat AFB	AV 787-2229
38.	Suzanne Leicht	NSA	(301)796-6461
39.	Thomas Leonard	Univ. of Wisconsin	
40.	Mary Anne Maher	WSMR	AV 278-1882
41.	COL Duff G. Manges	ARO	AV 935-3331
42.	John Mardo	ARRADCOM	AV 880-3008
43.	Michael P. McMiller	AVRADCOM	AV 693-1525
44.	Grady Miller	AMSAA	(301)278-5882
45.	Walter B. Miller	WSMR	(505)678-1528
46.	Robert Mioduski	AMSAA	AV 283-2135
47.	J. Richard Moore	BRL	AV 283-4980
48.	Mike Moore	JPG	AV 480-7484
49.	Richard L. Moore	ARRADCOM	AV 880-4639
50.	Bernard Most	Northrop Services, Inc.	(919)541-2390
51.	Donald M. Neal	AMMRC	AV 955-3165
52.	William S. Phoebus	APG, Maryland	AV 283-3943
53.	Everett C. Reich	AMSAA	(301)278-3151

NAME		AFFILIATION	TELEPHONE #
54.	Edward W. Ross, Jr.	NLABS	(617) 633-5209
55.	Donald B. Rubin	Univ. of Wisconsin	(608)263-5115
56.	Carl T. Russell	CRTC	(907)873-1181
57.	Edward Saibel	ARO	AV 935-3331
58.	Nozer Singpurwalla	George Washington Univ.	
59.	Walter L. Smith	Univ. of North Carolina	(919)962-1357
60.	Devinder S. Sodhi	CRREL	AV 684-3267
61.	Dan Solomon	NCSU	
62.	Leonard A. Stefanski	ARO & UNC	(919)933-5863
63.	Donald M. Swingle	Consultant	(505)522-5197
64.	Douglas Tang	WRAIR	AV 576-2011
65.	Audrey E. Taub	NSWC	(703)663-7892
66.	Malcolm S. Taylor	BRL	AV 283-5533
67.	Jerry Thomas	BRL	
68.	Paul Thrasher	WSMR	AV 678-2173
69.	Langhorne Withers	OTEA	(202)756-2160
70.	Thomas M. Webb	MERADCOM	AV 664-5771
71.	Douglas Wolfe	Ohio State Univ.	(614)422-0293
72.	C. F. Jeff Wu	Univ. of Wisconsin	(608)262-7938
73.	Norman Wykoff	JPG	AV 480-7208

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ARO Report 82-2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PROCEEDINGS OF THE TWENTY-SEVENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING		5. TYPE OF REPORT & PERIOD COVERED Interim Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s)		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Army Mathematics Steering Committee on Behalf of the Chief of Research, Development and Acquisition		12. REPORT DATE June 1981 13. NUMBER OF PAGES 511
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) U. S. Army Research Office P. O. Box 12211 Research Triangle Park, NC 27709		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

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

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This is a technical report from the Twenty-Seventh Conference on the Design of Experiments in Army Research, Development and Testing. It contains most of the papers presented at that meeting. These treat various Army statistical and design problems.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

discrete distributions
kinetic energy penetrators
supermartingales
ARMA analysis of time series
multimodal systems
reliability
two-phase sampling
Westenberg type tests
error propagation
terrain classification
random number generator
Weibull process models
3-factor interactions
software reliability
statistical hypotheses
complex computer simulation models
experimental designs
long-term storage problems
ballistic testing
two-way layouts
tracking maneuvering aircraft
error analysis
failure prediction

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE