U. S. Army Research Office

Report No. 75-2
June 1975

PROCEEDINGS OF THE TWENTIETH CONFERENCE ON THE DESIGN OF EXPERIMENTS

Part 1

Sponsored by the Army Mathematics Steering Committee

HOSTS

U. S. Army Operational Test and Evaluation Agency

and

U. S. Army Engineer Center at Fort Belvoir

23-25 October 1974

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

FOREWORD

The Army Mathematics Steering Committee (AMSC) sponsors, on behalf of the Chief of Research, Development and Acquisition, the series of conferences entitled, "Design of Experiments in Army Research, Development and Testing." It delegates the responsibility for the conduction of these meetings to its Subcommittee on Probability and Statistics. At the 30 November 1973 meeting of this subcommittee it was recommended that appropriate steps be taken to celebrate the twentieth anniversary of the design of experiments conferences. After some deliberation it was decided to ask members of the Program Committee to increase the usual number of invited speakers from five to eight, and to invite Army scientists to contribute many papers for both the technical and clinical sessions. In addition, some person should be asked to give a history of these conferences and their importance to the statistics used by the Army. This individual should point out the role played in these conferences by Professor Samuel S. Wilks, and also discuss his many contributions to the Army and to the other armed services. The Chairman of the Subcommittee, Dr. Walter Foster, reported that the coming conference would probably be held at Fort Belvoir and he hoped for confirmation of this in the near future.

In a letter under date of 20 February 1974, Lieutenant Colonel Harold P. Hoefekamp issued a formal invitation to hold the conference at Fort Belvoir on 23-25 October 1974. We quote the following paragraph from his letter: "The Operational Test and Evaluation Agency and the Engineer Center considers it an honor to host the Army's 20th Design Conference. Every effort will be made to insure that the best facilities and support are made available for this historic event. Both the Operational Test and Evaluational Agency and the Engineer Center are fully aware of the conference's significance, not only to the Army's scientific community, but to the Army as a whole." The sentiments expressed in this letter certainly guided the hosts in their handling of this meeting as it was one of the best conferences in this series. This was no doubt due largely to the expertise with which Fort Belvoir handled the arrangements and the visitors. Mr. Walter Hollis, Chairman on Local Arrangements, is to be commended on a very fine job. Unfortunately, he had to be out of the country on the dates of the conference. In his absence, Captain Stanley Dahlin took over his chores. He deserves special recognition for the manner in which he performed his assigned duties.

Each year the Program Committee is instructed 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 special interest to the host installation, is asked to present new developments in these fields. These selection criteria were certainly met by the gentlemen giving the talks in the General Sessions. The titles of their addresses are noted below:

Samuel S. Wilks and the Army Experiment Design Conferences
Dr. Churchill Eisenhart, National Bureau of Standards

Multidimensional Contingency Tables
Professor Solomon Kullback, The George Washington University

Multivariate Data Analysis
Professor Herbert Solomon, Stanford University

Order Statistics
Professor H. A. David, Iowa State University

Reliability
Professor Gerald Lieberman, Stanford University

Ranking and Selection Procedures
Professor Robert Bechhofer, Cornell University

Maximum Information from Experiments
Dr. Marion R. Bryson, U.S. Army Combat Development Experiment Command and
Dr. William Mallios, McDonald Service Company

The tenth Samuel S. Wilks Memorial Award of the American Statistical Association was presented to Mr. Cuthbert Daniel for his many outstanding contributions to the applications of statistics. The presentation of the medal, citation and honorarium was made by Professor Jerome Cornfield, President of the American Statistical Association. More details about this award appear in the body of these proceedings.

Probably the most valuable phases of these conferences are the technical and clinical sessions. In the technical sessions Army scientists announce their successes in handling a few of the many technical problems they face, while in the clinical sessions they have a chance to get help from nationally known scientists on ways to cope with some of their unsolved design problems. This year there were thirty-four (34) technical papers and eight (8) clinical papers on the agenda. We are pleased to be able to print many of these contributed papers in this technical manual.

Members of my Program Committee (Marion Bryson, Gerard Dobrindt, Walter Foster, Fred Frishman (Secretary), Walter Hollis, Badrig Kurkjian, Clifford Maloney, Herbert Solomon, Douglas Tang and Robert Thrall) are due my thanks for outlining the main events of this meeting and for selecting such an outstanding list of invited speakers. I would also like to express my appreciation to Francis Dressel for serving as secretary during the final phases of this conference.

FRANK E. GRUBBS
Conference Chairman

TABLE OF CONTENTS

Parts 1 and 2

(Part 2 starts on page 463)

TWENTIETH CONFERENCE ON THE DESIGN OF EXPERIMENTS

IN ARMY RESEARCH, DEVELOPMENT AND TESTING

23-25 October 1974

The U.S. Army Operational Test and Evaluation Agency
and the U.S. Army Engineer Center

* * * * * Wednesday, 23 October * * * * *

0800-0900 REGISTRATION - Main Lobby of Humphrey's Hall

0900-1130 GENERAL SESSION I - Auditorium of Humphrey's Hall

CALLING OF CONFERENCE TO ORDER

Walter Hollis, Chairman on Local Arrangements, U.S. Army Operational Test and Evaluation Agency

WELCOMING REMARKS

CHAIRMAN OF SESSION I

Dr. Ivan R. Hershner, Jr., Office of the Chief of Research and Development and Acquisition, The Pentagon, Washington, D.C.

SAMUEL S. WILKS AND THE ARMY EXPERIMENT DESIGN CONFERENCE SERIES

Dr. Churchill Eisenhart, National Bureau of Standards, Gaithersburg, Maryland

MULTIDIMENSIONAL CONTINGENCY TABLES

Professor Solomon Kullback, Department of Statistics, The George Washington University, Washington, D.C.

1130-1300 LUNCH

1300-1445 CLINICAL SESSION A - Auditorium of Humphrey's Hall

CHAIRMAN

Douglas B. Tang, Department of Biostatistics/Applied Mathematics, Walter Reed Army Institute of Research, Washington, D.c

1300-1445 CLINICAL SESSION A - (Cont'd)

PANELISTS

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

Churchill Eisenhart, National Bureau of Standards, Gaithersburg, Maryland

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

J. Richard Moore, U.S. Army Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland

NEEDED-MULTI-DIMENTIONAL, NON-GAUSSIAN, RANDOM PROCESSES WITH SPECIFIED COVARIANCE AND PROBABILITY DENSITY FUNCTIONS

James W. Wright, Advanced Sensors Directorate, U.S. Army Missile RD&E Lab, AMICOM, Redstone Arsenal, Alabama

DESIGN OF EXPERIMENTS FOR THE EVALUATION OF MATERIEL PERFORMANCE IN WORLDWIDE ENVIRONMENTS

B. O. Benn, Waterways Experiment Stations, Corps of Engineers, Vicksburg, Mississippi

STATISTICAL TESTING OF ELECTROEXPLOSIVE DEVICES SUBJECTED TO SHORT PULSE STIMULI

Burton V. Frank, Picatinny Arsenal, Dover, New Jersey
Ramie H. Thompson, Franklin Research Institute Laboratories, Philadelphia, Pennsylvania

1300-1445 TECHNICAL SESSION I

CHAIRMAN

William L. Shepherd, Instrumentation Directorate, U.S. Army White Sands Missile Range, White Sands Missile Range, New Mexico

TARGET VISIBILITY AND DECISION OPTIMIZATION

Timothy M. Small, U.S. Army Mobility Equipment Research and Development Center, Fort Belvoir, Virginia

OPTIMIZATION OF A PRODUCTION LINE

Eileen Weigand, Manufacturing Technology Directorate, Frankford Arsenal, Philadelphia, Pennsylvania

AN APPLICATION OF THE WEIBULL-GNEDENKO DISTRIBUTION FUNCTION FOR GENERALIZING FRAGMENT CONDITIONAL KILL PROBABILITIES

William P. Johnson, Modeling Branch, VL, Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland

1300-1445 TECHNICAL SESSION I - (Cont'd)

DECISION THEORY APPROACH TO GRADING BINOMIAL POPULATIONS

Paul E. Williams, U.S. Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland

1300-1445 TECHNICAL SESSION 2

CHAIRMAN

Fred K. McCoy, Metholology Branch, Test Design Division, U.S. Army Operational Test and Evaluation Agency, Fort Belvoir, Virginia

PSEUDO-BAYESIAN INTERVALS FOR RELIABILITY OF A SERIES SYSTEM GIVEN WEIBULL COMPONENT DATA

Ronald L. Racicot, Research Directorate, Benet Weapons Laboratory, Watervliet Arsenal, Watervliet, New York

THE UNIQUE APPLICATION OF BAYESIAN STATISTICS TO HIGH RELIABILITY TESTING

Charles Pleckaitis and Erwin Biser, Electronic Engineer, Avionics Laboratory, USAECOM, Fort Monmouth, New Jersey

ANALYTICAL APPROACH TO ROBUSTNESS FOR BAYESIAN DEVELOPMENTS IN RELIABILITY

Chris P. Tsokos and A.N.V. Rao, Department of Mathematics, University of South Florida, Tampa, Florida

A BAYESIAN APPROACH TO RELIABILITY GROWTH ANALYSIS

John G. Mardo, Product Assurance Directorate, Product Assurance Technology Division, Picatinny Arsenal, Dover, New Jersey

1445-1515 BREAK

1515-1700 CLINICAL SESSION B - Auditorium of Humphrey's Hall

CHAIRMAN

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

PANELISTS

O. P. Bruno, Reliability, Availability and Maintainability Division, U.S. Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland

Cuthbert Daniels, Rhinebeck, New York

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

1515-1700 CLINICAL SESSION B - (Cont'd)

J. Richard Moore, U.S. Army Ballistics Research Laboratories, Aberdeen Proving Ground, Maryland

THE LAUNCH TRANSIENT PROBLEM FOR OPTICAL CONTRAST (TV IMAGING) SEEKERS

Christopher E. Kulas and Joseph A. de Blaquiere, Jr., U.S. Army Missile RD&E Lab, Advanced Sensors Directorate, Optical Guidance Technology, AMICOM, Redstone Arsenal, Alabama

HYPOTHESES TESTING ON PRODUCT OF TWO BINOMIAL DISTRIBUTORS

Lang Withers, Operational Test Evaluation Agency, Fort Belvoir, Virginia

DETERMINATION OF "AVERAGE" NOISE OF U.S. ARMY CONSTRUCTION AND MATERIALS HANDLING EQUIPMENT

Samuel E. Wehr, U.S. Army Mobility Equipment Research and Development Center, Fort Belvoir, Virginia

1515-1700 TECHNICAL SESSION 3

CHAIRMAN

Badrig Kurkjian, U.S. Army Materiel Command, Alexandria, Virginia

EXPERIMENTAL COLLECTION OF STATISTICS BY COMPUTER SIMULATION: THE AUTOVON NETWORK

Egon Marx, Harry Diamond Laboratories, Washington, D.C.

AN ANALYSIS OF BUFFERS IN A PRODUCTION SYSTEM

Anton Hauschild, Manufacturing Technology Directorate, Frankford Arsenal, Philadelphia, Pennsylvania

RATE DEPENDENT FAILURE PROCESS SIMULATION

Martin Roffman and Robert Kuehn, Manufacturing Technology Directorate, Frankford Arsenal, Philadelphia, Pennsylvania

STATISTICAL MODEL FOR CONTROLLER PERFORMANCE MEASURES FOR AN AIR TRAFFIC AUTOMATED CENTER (ATMAC)

Erwin Biser, Avionics Laboratory, USAECOM, Fort Monmouth, New Jersey

1515-1700 TECHNICAL SESSION 4

CHAIRMAN

Edward W. Ross, Jr., U.S. Army Natick Laboratories, Natick, Massachusetts

1515-1700 TECHNICAL SESSION 4 - (Cont'd)

A FLEXIBLE, GENERAL PURPOSE COVARIANCE COMPUTER PROGRAM

Clifford J. Maloney and Lucille Carver, Bureau of Biologics, Bethesda, Maryland

OBSERVATIONS ON THE ALGEBRA OF NON-NORMAL FUNCTIONS

Donald M. Neal, Mechanics Research Laboratory, Army Materials and Mechanics Research Center, Watertown, Massachusetts

COMPUTATION OF MOMENTS OF A LOG RAYLEIGH DISTRIBUTED RANDOM VARIABLE

William L. Shepherd, Instrumentation Directorate, U.S. Army White Sands Missile Range, White Sands Missile Range, New Mexico

ON THE TYPE II ERROR OF THE 2 x 2 CONTINGENCY TABLE CHI-SQUARE STATISTIC

Robert L. Launer, Procurement Research Office, U.S. Army Logistics Management Center, Fort Lee, Virginia

1900-2000 SOCIAL HOUR - Mackenzie Hall (Officer's Club)

2000- BANQUET

PRESENTATION OF THE SAMUEL S. WILKS MEMORIAL AWARD

Dr. Frank E. Grubbs, Ballistic Research Laboratories, Master of Ceremonies

* * * * * THURSDAY * * * * *

0830-1000 CLINICAL SESSION C - Auditorium of Humphrey's Hall

CHAIRMAN

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

PANELISTS

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

Larry H. Crow, U.S. Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland

Gerald Lieberman, Department of Operations Research, Stanford University, Stanford, California

CLIMATIC CHANGES FOLLOWING VOLCANIC ERUPTIONS

John Bart Wilburn, Jr., Instrumentation and Methodology Branch, U.S. Army Electronic Proving Ground, Fort Huachuca, Arizona

VALIDATION OF ENGINEERING SIMULATION MODELS

Roland H. Rigdon, Rodman Laboratory, Rock Island Arsenal, Rock Island, Illinois

0830-1000 TECHNICAL SESSION 5

CHAIRMAN

Gerard T. Dobrindt, U.S. Army Test and Evaluation Command, Aberdeen Proving Ground, Maryland

PREDICTING METASTASIS BY DISCRIMINANT FUNCTION WHEN SMALL OPHTHALMIC MELANOMAS HAVE BEEN DIAGNOSED

Walter D. Foster and Ian McLean, Armed Forces Institute of Pathology, Washington, D.C.

FORECASTING MODELS FOR MOSQUITO POPULATION BEHAVIOR

Stephen Smeach and Chris P. Tsokos, Department of Mathematics, University of South Florida, Tampa, Florida

CURVE FITTING OF DISCRETE POINTS BY LEGENDRE POLYNOMIALS

O. M. Essenwanger, Physical Sciences Directorate, U.S. Army Missile Command, Redstone Arsenal, Alabama

0830-1000 TEHCNICAL SESSION 6

CHAIRMAN

Paul C. Cox, Quality Assurance Office, U.S. Army White Sands Missile Range, White Sands Missile Range, New Mexico

0830-1000 TECHNICAL SESSION 6 - (Cont'd)

AIRCRAFT ARMAMENT FIRE CONTROL SENSITIVITY ANALYSIS USING A PROGRAMMABLE CALCULATOR

Thomas O. McIntire, Methodology and Instrumentation Division, U.S. Army Yuma Proving Ground, Yuma Proving Ground, Arizona

APPLICATIONS OF A MORE EFFICIENT SEQUENTIAL SENSITIVITY TEST STRATEGY AND ESTIMATION METHODOLOGY TO RELIABILITY ASSESSMENTS

Gertrude Weintraub, Ammunition Development and Engineering Directorate, Concepts and Effectiveness Division, Picatinny Arsenal, Dover, New Jersey

STATISTICAL ANALYSIS AND MODELING OF SENSITIVITY AUGMENTATION IN CUTANEOUS COMMUNICATIONS

Richard J. D'Accardi and H. S. Bennett, U.S. Army Electronics Command, Fort Monmouth, New Jersey

1000-1030 BREAK

1030-1130 GENERAL SESSION II - Auditorium of Humphrey's Hall

CHAIRMAN

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

MULTIVARIATE DATA ANALYSIS

Professor Herbert Solomon, Department of Statistics, Stanford University, Stanford, California

1130-1300 LUNCH - Mackenzie Hall

1300-1415 TECHNICAL SESSION 7 - Auditorium of Humphrey's Hall

CHAIRMAN

Richard J. D'Accardi, U.S. Army Electronics Command, Fort Monmouth, New Jersey

SKIP-LOT PROCEDURE FORMULATION USING THE SIMPLIFIED MARKOV CHAIN METHOD

Richard M. Brugger, RAM Assessment Division, U.S. Army Armament Command, Rock Island, Illinois

SEMI MARKOV CHAINS APPLIED TO CONTINUOUS SAMPLING PLANS

David L. Arp, Naval Weapons Center, China Lake, California

1300-1415 TECHNICAL SESSION 8

CHAIRMAN

Gertrude Weintraub, Ammunition Development and Engineering Directorate, Concepts and Effectiveness Division, Picatinny Arsenal, Dover, New Jersey

PREDICTING RELIABILITY GROWTH

Larry H. Crow, U.S. Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland

MINIMUM VARIANCE SOLUTION OF A POLYNOMIAL FUNCTION OF TWO NOISY RANDOM VARIABLES

Oren N. Dalton, Mathematics Services Branch, U.S. Army White Sands Missile Range, White Sands Missile Range, New Mexico

1300-1415 TECHNICAL SESSION 9

CHAIRMAN

CPT Loit D. Proigler, Analysis Branch, Technical Support Division, U.S. Army Operational Test and Evaluation Agency, Ft. Belvoir, Virginia

THE PROBABILITY OF MOTOR CASE RUPTURE

Ronald S. Downs and Paul C. Cox, Quality Assurance Office, U.S. Army White Sands Missile Range, White Sands Missile Range, New Mexico

ON THE NONEXISTENCE OF SOME INCOMPLETE BLOCK DESIGNS

Alan W. Benton, U.S. Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland

1415-1445 BREAK

1445-1700 GENERAL SESSION III - Auditorium of Humphrey's Hall

CHAIRMAN

Professor Herbert Solomon, Department of Statistics, Stanford University, Stanford, California

ORDER STATISTICS

Professor H. A. David, Department of Statistics, Iowa State University, Ames, Iowa

RELIABILITY

Professor Gerald Lieberman, Department of Operations Research, Stanford University, Stanford, California

* * * * * FRIDAY * * * * *

0830-0915 TECHNICAL SESSION 10 - Auditorium of Humphrey's Hall

CHAIRMAN

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

SIMPLE STATISTICAL ALTERNATIVES TO THE METHOD OF LEAST SQUARES FOR THE DETERMINATION OF X-INTERCEPT, AND SLOPE

Joseph F. Hannigan, Research Institute, U.S. Army Engineering Test Center, Fort Belvoir, Virginia

A STATISTICAL APPROACH TO THE LOADING AND FAILURE OF STRUCTURES

Ronald Merritt, Construction Engineering Research Laboratory, Champaign, Illinois

0830-0915 TECHNICAL SESSION 11

CHAIRMAN

Lang P. Withers, Analysis Branch, Technical Support Division, U.S. Army Operational Test and Evaluation Agency, Fort Belvoir, Virginia

STRAIN GAGE INSTRUMENTATION FOR AMMUNITION TESTING

Paul D. Flynn, Pitman-Dunn Laboratory, Frankford Arsenal, Philadelphia, Pennsylvania

DATA ANALYSIS OF AUTOMATIC TRACKER TESTING

Nicholas P. Marasco, Robert T. Volz and Thomas G. Kelley, Photoelectric Branch, Fire Control Development and Engineering Directorate, Frankford Arsenal, Philadelphia, Pennsylvania

0830-0915 TECHNICAL SESSION 12

CHAIRMAN

Roy C. Schmidt, Test Design Division, U.S. Army Operational Test and Evaluation Agency, Fort Belvoir, Virginia

STATISTICAL INVESTIGATION INTO PULSE CHARGING OF NICKELCADMIUM BATTERIES

Walter Kasian and Erwin Biser, U.S. Army Electronics Command, Fort Monmouth, New Jersey

OPTICAL CHARACTERIZATION OF SURFACE ROUGHNESS

E. L. Church and J. M. Zavada, Frankford Arsenal, Philadelphia, Pennsylvania

* * * * * FRIDAY * * * * *

0915-1045 GENERAL SESSION IV - Auditorium of Humphrey's Hall

CHAIRMAN

Dr. Frank E. Grubbs, Chairman of the Conference, U.S. Army Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland

OPEN MEETING OF THE AMSC SUB-COMMITTEE ON PROBABILITY AND STATISTICS

Dr. Walter D. Foster, Computer Services Division, Armed Forces Institute of Pathology, Washington, D.C.

RANKING AND SELECTION PROCEDURES

Professor Robert Bechhofer, Department of Operations Research, Cornell University, Ithaca, New York

1045-1115 BREAK

1115-1215 GENERAL SESSION IV (continued)

MAXIMUM INFORMATION FROM EXPERIMENTS

Dr. Marion R. Bryson, U.S. Army Combat Developments Experiments Command, Fort Ord, California

Dr. William Mallios, Braddock, Dunn, McDonald Sérvice Company, Fort Ord, California

SAMUEL S. WILKS AND THE ARMY EXPERIMENT DESIGN CONFERENCE SERIES

Churchill Eisenhart
Senior Research Fellow
Institute for Basic Standards
National Bureau of Standards
Washington, D.C. 20234

ABSTRACT. A biography of Professor Samuel Stanley Wilks (1906-1964) of Princeton University, with particular attention to his early life, notes on the persons who shaped his professional development, review of his many facetted professional career and his role in initiating and launching the U.S. Army's annual series of Conferences on the Design of Experiments in Army Research, Development and Testing.

1. BIRTH, FAMILY, AND EARLY YEARS. Sam Wilks was born on the 17th of June 1906 in Little Elm, Denton County, Texas, the first of the three children of Chance C. and Bertha May Gammon Wilks. His father trained for a career in banking, but after a few years chose instead to make his livelihood by operating a 250-acre farm near Little Elm. His mother had a talent for music and art; and a lively curiosity, which she transmitted to her three sons. The predilection of their father, Chance C. Wilks, for alliteration is manifest in the given names of all three: Samuel Stanley, Syrrel Singleton, and William Weldon (Wilks).

Syrrel, less than two years younger than Sam, was his boyhood companion; studied biology (B.S., 1927) and physiology (Ph.D., 1936); became Associate Professor of Physiology at the Air Force School of Aviation Medicine; and passed away early this year (1974). In consequence of Sam's and Syrrel's initials being the same, their publications are sometimes lumped together under "S. S. Wilks" in bibliographic tools, e.g., in the successive volumes of the Science Citation Index.

Sam's "baby brother," William, was eight years younger. He also took a B.S. degree; became a research advisor to Bell Aircraft Company in Fort Worth, Texas; and is still living. The choice of "Weldon" for his middle name is merely a happenstance of his father's effort to achieve a triple alliteration, and has no genealogical significance: there is no known connection between the Chance Wilks family and that of the pioneer biometrician, W. F. R. Weldon (1860-1906), who died in London in April of the year in which our biographee was born; or with any other Weldons.

Sam began his early education in a typical one-room rural school house where, in the seventh grade, he had as his teacher William Marvin Whyburn, who became a distinguished mathematician, the president of Texas Technological College (1944-1948), and the chairman or head of two university departments of mathematics (UCLA, 1937-1944; University of North Carolina, 1948-1956, 1960-1965)--the first of an extraordinary number of prominent people who had a part in Sam's education. He attended high school in Denton, the county seat and the site of North Texas State Teachers College (now North Texas State University), and of a College of Industrial Arts for women (now Texas Woman's University). During the week he roomed in Denton, and went home on weekends, walking the 15 miles to his father's ranch when necessary. During his final year of high school, it was noticed that he was absent repeatedly from study hall. Inquiry revealed that he was skipping study hall in order to take a mathematics course at North Texas State Teachers College.

Following graduation from high school, Sam continued his studies at North Texas State Teachers College, where he followed an industrial arts program, with particular attention to mathematics. He received an A.B. in architecture in June 1926, a few days before his 20th birthday. A large drinking fountain, designed by Sam and a friend, on the campus of the College attests to his talent and serves as a reminder of his one-time interest in architecture. But believing his eyesight inadequate for the life of an architect, he turned to a career in mathematics.

2. TEACHING AND GRADUATE STUDY. During the school year 1926-1927, Wilks taught mathematics and manual training in a public school in Austin, Texas, and began graduate study of mathematics at the University of Texas there. He continued his studies at the University of Texas as a part-time instructor in mathematics 1927-1928; and received an M.A. in mathematics in 1928. His first course in advanced mathematics at the University of Texas was set theory, taught by R. L. Moore (1882-1974), renowned among mathematicians for his research in topology, his unusual methods of teaching, and the vigor and resoluteness of his opinions. Wilks was fascinated by the unfolding of this beautiful theory from a few simple definitions and axioms, but Moore's espousal of pure mathematics as a discipline wholly divorced from application, and Moore's scorn of applied mathematics as work on a level with washing dishes, were incomprehensible and unacceptable to him. Had Moore's attitude been otherwise, Sam might have become a topologist. But, as Alex M. Mood[1/] has said in his note on Sam's philosophy about his work, "Sam's character demanded that his work be immediately and obviously useful [and] Moore was the last man to persuade him that point set theory was useful." (MOOD 1965, p. 953) Much more in keeping with his "character" were probability and statistics, to which he was introduced by Edward L. Dodd (1875-1943), an inspiring teacher and distinguished scholar, noted for his researches on mathematical and statistical properties of various types of means.[2/]

An aside on Sam's views with respect to pure mathematics and pure mathematicians seems appropriate at this juncture, before taking up the next step in his education. To this end I can do no better than to quote further from Mood's note:

> "Wilks...saw little sense in pure mathematics unless it had some ultimate application. He generally believed that most pure mathematics would eventually justify itself in this way and was delighted when that did happen in his own work or that of others....The set theoretical foundation of probability theory developed by Kolmogorov gave Sam no end of pleasure partly because of that early course, perhaps, but more likely because it was a good piece of evidence that pure mathematicians were not, after all, wasting their time.
>
> "While Sam was generally optimistic about the eventual utility of pure mathematics he became less and less patient over the years with pure mathematicians themselves--especially those in the United States. For one thing he believed that their general refusal to apply their intellects even briefly to important practical problems was less than patriotic, to say the least. He rarely missed an opportunity to point out that almost all top-level Soviet mathematicians had at one time or another turned to an important field of application thus placing themselves, in his eyes, quite above many of America's leading mathematicians.
>
> "The thing that particularly annoyed Sam about pure mathematicians was their snobbishness about pure mathematics and, worse, their success in generating the same sort of snobbishness in every mathematically talented student that came along. Sam was a very even tempered man but this was a subject that could summon loud indignation from him. He believed that for reasonably even balance in the development of mathematics a substantial proportion of the most talented students should go into mathematical statistics, mathematical physics, applied mathematics, econometrics, etc. As it was, he believed that pure mathematics preempted over nine out of ten of the most talented students thus completely deforming mathematical progress in the United States. In his later years he maintained that it was impossible for him to persuade enough sufficiently promising college graduates to undertake work in statistics at Princeton and therefore he had to go to Britain and Canada to find good students whose attitudes had not been corrupted by pure mathematicians in the United States." (MOOD 1965, 953-954)

When Sam completed the requirements for his M.A. in mathematics at the University of Texas in 1928, Professor Dodd encouraged him to pursue further study of mathematical statistics at the University of

Iowa[3/] under Henry L. Rietz (1875-1943)[4/], the leader of his generation in American mathematical statistics.[5/] Wilks stayed on at the University of Texas as an instructor in mathematics during the summer of 1928, and the academic year 1928-29; applied for a fellowship at the University of Iowa; and to pick up some ready cash, served as a monitor for State bar exams given at the University.

In due course, Sam was offered, and accepted, a fellowship at the University of Iowa, in Iowa City. He arrived in Iowa City in the summer of 1929 to begin a two-year program of graduate study and research leading to a Ph.D. degree in mathematics, with a minor in education. During the second summer (1930), he was joined by two others whose names were later to become well-known in probability and mathematical statistics circles: Allen T. Craig and John H. Curtiss.

Curtiss had just received his A.B. in mathematics at Northwestern University, and had come to the University of Iowa to study actuarial mathematics preparatory to choosing actuarial work as a career. He was assigned to one of two desks arranged back-to-back in the Mathematics Department Library, the other occupied by Sam. He has a close-up picture of Sam taken from this vantage point.[6/]

Allen T. Craig, in contrast, had returned to the University of Iowa in the summer of 1930 for the express purpose of completing his doctoral thesis "On the Distribution of Certain Statistics Derived from Small Random Samples". I say "had returned to the University of Iowa" because Craig had been there during the academic year 1928-29, but had left Iowa City in the summer of 1929 to accept a position as an Instructor in mathematics at his alma mater, the University of Florida, in Gainesville, for the academic year 1929-30. Drawn together by common interests, Allen Craig and Sam Wilks immediately became close and lifelong friends. Craig, in his thesis (CRAIG 1932), gave a number of general results on the distributions of such statistics as the arithmetic mean, harmonic mean, geometric mean, median, quartile, decile and range of samples of small $\underline{n}$ items selected at random from a rather arbitrary (continuous) universe, together with a large number of explicit results for sampled universes of special types. Sam often said that his own work on the theory of nonparametric or distribution-free methods--an area in which Sam made a number of truly outstanding contributions[7/]--had its origins in the general formulas given by Craig for the distributions of the "median, quartile, decile, and range".

Sam's doctoral dissertation was, likewise, a contribution to "the theory of small samples". Entitled "On the distributions of statistics in samples from a normal population of two variables with matched sampling of one variable" (WILKS 1932a), it provided the small-sample distribution theory required to answer a number of questions drawn to

Sam's attention by Professor E. F. Lindquist, Professor of Education at the University of Iowa and Director of the Iowa Testing Programs, who had used the technique of "matched" groups in experimental work in educational psychology, and whose lectures Sam had attended.

Sam's thesis was preceded by a short note by Sam on "The standard error of the means of 'matched' samples" (WILKS 1931), published in the March 1931 issue of the Journal of Educational Psychology, where it was accompanied by an article by Lindquist (LINDQUIST 1931), describing the use and importance of "matched" groups as a statistical technique in experimental psychology and educational testing. Sam's predoctoral note and his doctoral dissertation were the first of a series of papers on multivariate analysis suggested by real-life problems in experimental psychology and educational testing, and mark the beginning of Sam's life-long association with the latter field.

Sam and Allen Craig both received their Ph.D.'s from the University of Iowa in June 1931--Sam in Mathematics, with a minor in Education; Allen, in Mathematics alone. "Father Rietz" was mighty proud of his "twins". Theirs were the first doctoral dissertations written at the University of Iowa on aspects of "the theory of small samples", the new area of mathematical research, initiated in 1908 by "Student" (William Sealy Gosset, 1876-1937) and developed to full flower by R. A. Fisher (1890-1962) between 1915 and 1928, to which an increasing number of American mathematicians were devoting attention at that time--notably C. C. Craig (at the University of Michigan in Ann Arbor), Harold Hotelling (at Stanford University, in California), Paul R. Rider (at Washington University, St. Louis), and Rietz (at the University of Iowa, in Iowa City).[8/] Rietz was doubly proud of their accomplishments; not only had each made a first-rate contribution to "the theory of small samples", but also the mathematics in their dissertations was intelligible to American mathematicians--which was a great deal more than one could say about the papers of R. A. Fisher.[9/] He therefore held out two "prizes" to his deserving "twins": (1) an appointment as an Associate (a rank between Instructor and Assistant Professor) in his department, and (2) his endorsement for a National Research Council Fellowship. Allen chose the appointment in the Department of Mathematics--stayed on to become a full Professor in 1945, and retired in 1970; Sam, the NRC Fellowship, and made plans to continue research in multivariate statistical analysis under Harold Hotelling (1895-1973), a pioneer in this field, and the individual in the United States most versed in the mathematics of the Student-Fisher theory of small samples.[10/]

After receiving his Ph.D., Wilks stayed on to attend the lectures given, and seminar conducted, during the first half of the summer session, 8 June - 16 July, by the British mathematical statistician, Egon S. Pearson; and gave a talk in the seminar series. Pearson's two papers with Jerzy Neyman on "The use and interpretation of certain test criteria for the purposes of statistical inference" (Part I, Biometrika, Vol. 20A (1928), pp. 175-240; Part II, ibid, pp. 263-294) had been well received by mathematicians interested in statistical theory. As you will recall, it was in these papers that they introduced and explored their likelihood-ratio technique for more or less automatically discovering "good" tests of various statistical hypotheses.[11/]

Wilks also met R. A. Fisher, who came over to Iowa City from Ames for a day during this period. By an extraordinary coincidence, R. A. Fisher was "in residence" that summer at Iowa State College, at Ames, 90 miles distant, giving a "competing" series of lectures on the material in his two books, Statistical Methods for Research Workers (3rd edition, 1930) and The Genetical Theory of Natural Selection (1930), during the first half of their summer session, 16 June-24 July. The overlap of the two programs, and the distance between the two institutions, made it physically impossible for faculty and students to take in both programs in their entirety.

3. MARRIAGE AND POSTDOCTORAL STUDY. Sam returned to Texas in midsummer 1931, and on September 1 married Gena Orr of Denton. The Wilks and Orr families had been friends for many years. Indeed, about one year before Chance Wilks finally won the hand of Bertha, she was being courted by Will Orr, while Chance was away from Little Elm, trying his hand at the banking business. But Chance returned in time to prevent my story from ending before it began--and in due course Gena was fathered by Will; and Sam, by Chance.

Sam and Gena had known each other from childhood. They attended the same high school in Denton; she was a student at the College of Industrial Arts, in Denton, at the same time that Sam was attending the North Texas State Teachers College there; and they both received their A.B. degrees in 1926; but they did not start "dating" until that summer. What brought them together was the wedding of Sam's cousin, James Hodge, and Jessie Hill, at which Sam was Best Man, and Gena a bridesmaid. Gena then taught school locally for a couple of years, while Sam was continuing his study of mathematics at the University of Texas, in Austin; and continued to date Sam from time to time when he was home on vacation. In due course she got herself over to the University of Texas, where she did graduate work in English, and received her Master's Degree in 1929.

As part of their honeymoon, Sam and Gena set off for New York City by boat, from Galveston, Texas. The trip took five days. They settled in an apartment on the 6th floor of the Columbia University-owned apartment building at 401 West 118th Street. During World War II the main offices of the Statistical Research Group--Columbia (SRG-C), of which Harold Hotelling was the Principal Investigator, were located in this building, in what had been Sam and Gena's apartment; and W. Allen Wallis, the Group's Director of Research, occupied what had been their bedroom.

Among those attending Hotelling's lectures on "Statistical Inference" that first year at Columbia in addition to Sam were Acheson J. Duncan, from whom I was later to receive my first course in this subject, and W. J. Youden (1900-1971), who was later to join me at the National Bureau of Standards (1948-1965) as practitioner, expositor and innovator of statistical methods par excellence. "Atch" Duncan was then an Instructor in Economics at Princeton University, and at my father's insistence had been sent at University expense to study modern statistical inference under Hotelling. I shall say more about this in a few moments. "Jack" Youden had received his Ph.D. in Chemistry from Columbia in 1924, was a Physical Chemist at the Boyce Thompson Institute for Plant Research in Yonkers, New York, and was commuting to New York to hear Hotelling's lectures on his own volition to gain a better grasp of Student-Fisher theory and methods.[12/]

In addition to auditing Hotelling's lectures, Sam joined Jack W. Dunlap and Warren G. Findley, then Ph.D. candidates at Columbia in Psychology and Educational Psychology, respectively, in attending the lectures, at Teachers College, of the English psychologist, Charles E. Spearman (1863-1945), revered by psychologists as the father of Factor Analysis (1904) and for development of a rational basis for determining general intelligence and for validating intelligence testing.[13/] I mention Jack Dunlap and Warren Findley explicitly because Sam's and their paths were to meet and join for a while at various times in later years, for example, when Dunlap was Director of Research of the National Research Council's Committee on Pilot Selection and Training (1941-42), and when Findley was Director of Test Development (1948-53), and later in charge of the Evaluation and Advisory Services (1953-56) of the Educational Testing Service in Princeton.

It was a year of exceptional productivity for Wilks: he wrote or completed four distinct papers in the area of multivariate analysis all of which saw almost immediate publication. In one (WILKS 1932b) he found the maximum likelihood estimates of the parameters of a bivariate normal distribution when some of the individuals in a sample yield observations on both variables, x and y, and some only on x, or on y, alone; in a second (WILKS 1932c), he showed that the distribution of the multiple correlation coefficient in samples from a normal population with a non-zero multiple correlation coefficient could be derived directly from Wishart's generalized product moment distribution (1928) without making use of the geometrical notions and an invariance property utilized by R. A. Fisher in his derivation (1928); in the third, his great paper on "Certain Generalizations in the Analysis of Variance" (WILKS 1932e), he defined the "generalized variance" of a sample of n individuals from a multivariate population, constructed multivariate generalizations of the correlation ratio and coefficient of multiple correlation; deduced the moments of the sampling distributions of these and other related functions in random samples from a normal multivariate population from Wishart's generalized product moment distribution (1928); constructed the likelihood ratio criterion for testing the null hypothesis that k multivariate samples of sizes n_1, n_2, --, n_k are random samples from a common multivariate normal population, now called "Wilks's Λ criterion", and derived its sampling distribution under the null hypothesis; and similarly explored various other multivariate likelihood ratio criteria; and in the fourth (WILKS 1932d), an outgrowth of attending Spearman's lectures, he obtained an exact expression for the standard error of an observed "tetrad difference"[14/] in samples of size n from a normal population (in the special case in which the intercorrelations of the four variables are all zero in the population).

I mention these details just to show to what a remarkable extent Sam was not only applying, but also extending the most advanced concepts and tools of Fisher, Hotelling, Neyman, E. S. Pearson and Wishart within one year of the receipt of his Ph.D.! I often heard my father, Luther Pfahler Eisenhart (1876-1965), remark when he was Chairman of the Department of Mathematics (1928-1945) and Dean of the Graduate School (1933-1945) of Princeton University, that what determined a man's stature in his chosen field was not the caliber of his doctoral dissertation, but rather the caliber of the papers that he wrote and published after receiving his Ph.D. Sam certainly passed that test in 1932 with a wide margin to spare! Furthermore, the high regard in which Sam's papers were held immediately following their publication is attested by the fact, already mentioned, that Irwin devoted 9 out of the 14 pages on "Exact sampling distributions" in his "Recent Advances..(1932)" (IRWIN 1934) to detailed consideration of Sam's thesis and the first three of these four postdoctoral papers. And E. S. Pearson more recently remarked that Sam's "stature as a statistician was I think early established by his Biometrika paper of 1932 on 'Certain generalizations in the analysis of variance' [which] must have been written during the winter after he gained his Ph.D. and as such was a remarkable performance." (PEARSON 1964, p. 597)

While at Columbia University, Sam went down to the Bell Telephone Laboratories at 463 West Street to visit Walter A. Shewhart (1891-1967), father of statistical quality control of manufacturing processes, with whose work he had become acquainted through Rietz and Hotelling.[15/] Sam became very interested in Shewhart's work, and shortly thereafter Sam and Gena paid a brief visit to Walter and Edna Shewhart at their home in Mountain Lakes, New Jersey. Several years ago, Mrs. Shewhart told me that she remembered well how, as soon as Sam and Gena had left, Walter had turned to her and said, "There is a young man who is going to be one of the top men in Statistics in this country", or words to that effect. This was the beginning of the friendship and collaboration of these two men that continued until Sam's death.

In the Spring of 1932, Sam obtained a renewal of his National Research Fellowship, as an International Research Fellowship. He and Gena set off in August 1932 for London, England, where Sam was to be in residence in Karl Pearson's Department of Applied Statistics at University College (of the University of London) during the "Michaelmas Term" (Sept.-Dec.).[16/] While there, Sam and Karl Pearson's son, Egon S. Pearson, wrote a joint paper (PEARSON and WILKS 1933) in which the likelihood ratio techniques of Sam's generalized analysis-of-variance paper are developed in greater detail for samples from a bivariate normal distribution, generalizing to this bivariate case the three tests developed by Neyman and Pearson (1931) for the univariate case. To illustrate the numerical application of the procedures they had developed, they included two worked examples, one based on data on the tensile strength and Rockwell hardness of aluminum dicastings, taken from Walter A. Shewhart's Economic Control of Quality of Manufactured Product (1931).[17/]

While in London, Sam met a great many of the leading British statisticians, and their disciples, either at University College or at the delightful teas that preceded the monthly meetings of the Royal Statistical Society. To add to the excitement--and to the strain of a married couple's attempting to live in London on the small stipend of an International Research Fellow--Sam and Gena's son Stanley Neal Wilks was born in London, in October 1932.[18/] Early in January 1933, the family of three moved to Cambridge so that Sam could work with John Wishart (1898-1956), whose work in multivariate analysis was close to Sam's main interest.[19/]

When Sam arrived at Cambridge, he found that Wishart and Bartlett had just completed an "independent" derivation of Wishart's generalized product-moment distribution "by purely algebraic methods", that is, by means of moment-generating functions in combination with the matrix algebra of quadratic forms (WISHART and BARTLETT 1933). Wilks found himself right at home in their company, and promptly wrote another major paper (WILKS 1934) in which he gave a method of deriving directly from the multivariate normal distribution (i.e., without using the Wishart distribution) the moments of the sampling distributions of functions of determinants of the types considered in his two Biometrika papers. Also, at the suggestion of G. Udny Yule (1864-1951), he wrote a paper, "On the Independence of Sums of Squares in the Analysis of Variance", in which by means of characteristic functions in combination with elementary matrix algebra, he demonstrated the independence of various row, column, etc., "sums of squares" involved an analysis-of-variance analysis of randomized blocks, Latin square, and certain other experimental arrangements, discussed previously by R. A. Fisher. Communicated to the Royal Society--not the Royal Statistical Society--by Yule, for publication in its Proceedings, the paper suffered rough treatment: it was apparently sent to Fisher to referee, who seems to have felt that by its very theme it implied that he had not already given adequate and intelligible proofs; then the manuscript was lost, and Sam had to provide a second copy; and then it was rejected. The publication shortly thereafter, in a publication of the Royal Statistical Society, of a similar, but somewhat more elementary, paper on the same subject, by one of Fisher's proteges, was a sore point with Sam for many years. (I have discussed this matter with the author of the "offending" paper. He assures me that he never saw Sam's manuscript; and, until our conversation, never knew of its existence.)

In May 1933 my father offered Sam an appointment as an Instructor in Mathematics in Princeton University. I first met Sam when he turned up in Princeton in time for the fall semester 1933, imported for the express purpose of teaching me--at least, that was what I thought at the time. My budding interest in probability and statistics may have helped a tiny weeny bit, but the true explanation was quite otherwise, and has an interesting background.

4. WILKS'S PRINCETON APPOINTMENT, AND STATISTICS AT PRINCETON BEFORE WILKS. The key figure in Wilks's appointment was my father, Luther Pfahler Eisenhart (1876-1965), who, in the spring of 1933, was not only willing, but, as Chairman of the Department of Mathematics (1928-1945), Dean of the Faculty (1925-1933), and Chairman of the University Committee on Scientific Research (1930-1945), was also able to effect Wilks's appointment to an Instructorship in Mathematics on a more or less emergency basis over the opposition of almost every member of his Department.[20]

An event that was to be instrumental in bringing both mathematical economics and modern statistical theory and methodology to the Princeton campus was the arrival of Charles F. Roos (1901-1958) as a National Research Fellow in Mathematics for the academic year 1927-28. Roos had received his Ph.D. in theoretical economics at the Rice Institute in 1926 under Professor G. C. Evans (1887-1973), who at that time was developing a new mathematical theory of economic phenomena termed "economic dynamics", and had spent 1926-27 at the University of Chicago working with Professor Henry Schultz (1893-1938) who at that time was deeply engaged in his epochal research on statistical laws of demand and supply as one facet of his life's work on the theory and measurement of demand. Roos came to Princeton primarily to broaden and sharpen his knowledge of mathematics as a basis for making further contributions to Professor Evans' new "economic dynamics". While there he succeeded in convincing some members of the Department of Economics and Social Institutions that the Department could not afford to continue to neglect much longer the advances in economic theory and methods pioneered by Evans and Schultz.

In 1928 my father became the Chairman of the Mathematics Department. One of his early acts in this capacity was to arrange for the loan by the Bell Telephone Laboratories, Inc. of a member of its Technical Staff, Dr. Thornton C. Fry, author of Probability and Its Engineering Uses (D. Van Nostrand, 1928), to give a course at Princeton on "Methods of Mathematical Physics" as a Visiting Lecturer in Mathematics during the first semester 1929-30. I remember going with my father to Bell Labs to visit Fry during either my spring or summer vacation of 1929--the necessary arrangements may have been broached, or perhaps firmed up on that occasion. Be that as it may, one result of Fry's visit to Princeton was that a course in probability, taught by H. P. Robertson (1903-1964), Associate Professor of Mathematical Physics, using Fry's book as the text, was offered by the Mathematics Department during the second semester of my sophomore year (1931-32). It was this course that first interested me in probability and mathematical statistics and started me on my career.

In 1931 steps were taken that led to a course in "modern statistical theory" being offered for the first time at Princeton by the Department of Economics and Social Institutions during the first semester of my senior year (1933-1934). What happened was this: Professor Frank D. Graham (1890-1949) of this department approached my father in his capacity as Dean of the Faculty, and suggested that one way to overcome lack of competence in his department with respect to the latest developments in mathematical and statistical methods in economics would be to send one of the young instructors in his department to study with Professor Henry Schultz at the University of Chicago. (The possibility of hiring a new staff member from the outside to this end had been considered earlier but put aside--the Depression was in full swing, and there was a freeze on new University appointments.) My father was favorable to this proposition, subject to an additional provision: that the individual concerned also study the modern theory of statistical inference with Harold Hotelling for the purpose of initiating a course in this subject on his return. The "victim" that Professor Graham had in mind was Acheson J. Duncan; and this is how it came to pass that Duncan, with financial assistance from the International Finance Section of Princeton University, spent the first half of the academic year 1931-32 studying with Professor Henry Schultz at the University of Chicago; and the second half with Professor Hotelling at Columbia.[21]

When Duncan arrived at Columbia University early in 1932, one of the first persons he met was Wilks. Another was W. R. Pabst, then a graduate student in Economics at Columbia, who years later, was to be instrumental in Duncan's becoming active as a teacher, author, and consultant on statistical methods in standardization and quality control. Duncan returned to Princeton in the fall of 1932, and began to ready himself to teach his projected new courses, unaware--as were also Wilks and my father--that before his course in "modern statistical theory" would get under way, Wilks would have joined the Princeton University faculty.

The program worked out for Duncan on his return to Princeton was this: He would participate as an assistant in the course, "Elementary Statistics", taught by Professor James G. Smith (1897-1946) in the Department of Economics and Social Institutions, scheduled for the Spring semester in 1933, serving as instructor in charge of the "laboratory" or "workshop" sessions in which the students gained practical experience in graphical and tabular presentation, and in the computation of descriptive statistics, index numbers, moving averages, link relatives, etc. Then, as a sequel to this course, Duncan's new course on "Modern statistical theory" would be offered by the same Department during the first semester of the academic year 1933-34.

I took these two courses in the Spring and Fall of 1933, respectively. In Smith's course we used as text <u>Principles and Methods of Statistics</u> by Robert E. Chaddock (1879-1940), published by the Houghton Mifflin Company in 1925, but the scope, nature, and mode of presentation is more accurately reflected by Professor Smith's <u>Elementary Statistics. An Introduction to the Principles of Scientific Methods</u>, published the following year (New York: Henry Holt and Company, 1934). Some of R. A. Fisher's contributions to statistical methodology were alluded to, but only very briefly, as tips on recent developments that would warrant looking into, not as integral parts of the course. In Duncan's course, on the other hand, built as it was around Hotelling's lectures, and the then available mimeographed chapters of Hotelling's never published book, <u>Statistical Inference</u>, the contributions of Student and R. A. Fisher occupied the center of the stage a large part of the time.

In the spring of 1933 a crisis developed of which I was totally unaware at the time, and the particulars of which I was not to learn until some years later. Wilks was at Cambridge University working with Wishart on the last lap of his two-year fellowship program and would be needing a permanent post, or at least a new source of income, by fall. He had sent résumés of his professional career to the universities in the United States known to have programs in probability and mathematical statistics, indicating that he was in need of an instructorship or other full-time position beginning with the academic year 1933-34. The replies that he received were all negative--the United States was in the depth of the Depression, colleges and universities were having to make do with dramatically reduced income from endowment and other sources, and all, it seemed, were tightening the belt, and none were planning to take on additional personnel. With an exceptional training in mathematical statistics, with four substantial research papers, and two research notes already published, one joint research paper accepted for publication, and two research papers nearly ready for publication, he was one of the most promising young men in mathematical statistics and applied mathematics generally, yet he had no prospect of a job. Wilks's situation seemed hopeless and was rapidly becoming desperate. Here he was in England with his wife and son; his fellowship funds, which were never really adequate for married people, or couples with children, were about to run out; and no prospect of employment.

Hotelling, knowing full well of my father's desire to build up a program in probability and mathematical statistics at Princeton and of the need of the College Entrance Examination Board for assistance from someone of Wilks's caliber on multivariate sampling distribution problems arising in educational testing, appealed directly to my father to take Wilks on at Princeton, stressing the long-term advantages to Princeton and the at-the-moment desperateness of Wilks's situation. Thus it came to pass late in the spring of 1933 that my father, as Chairman of the Mathematics Department, offered Wilks an instructorship in the Department of Mathematics for the academic year 1933-34, and advised him of a tentative arrangement that he had made with Professor Carl C. Brigham of the Department of Psychology and Associate Secretary of the College Entrance Examination Board (the central office of which had been at Princeton for some years) to work part-time also with the Board on problems arising in the scaling of achievement tests. It was not until many years later that I learned from my father that he had brought off this coup over the opposition of almost every member of his Department. I have often wondered whether he would have been able to bring it off a year or even six months later because, although he continued as Chairman of the Mathematics Department until 1945, in mid-1933 he gave up his post as Dean of the Faculty to become Dean of the Graduate School.

I also learned in later years, after I had returned from London and had become a close personal friend of Sam and Gena Wilks, that Sam had received only one other offer: at Rothamsted Experimental Station, in response to Wishart's repeated pressuring of R. A. Fisher on Wilks's plight and need. The offer itself, however, was humiliatingly niggardly and grossly inadequate to Wilks's needs, perhaps as a result of Wilks having already incurred Fisher's wrath over his analytical (in contrast to geometrical) exposition of the independence of sums of squares in the analysis of variance.

Wilks arrived in Princeton in September 1933. As a new instructor in the Department of Mathematics, he found himself teaching the usual undergraduate courses in analytic geometry, calculus, and so forth during the academic year 1933-34. In addition to such teaching that first year, Sam continued his research, primarily in multivariate analysis; gave me helpful guidance in the preparation on my senior thesis on "The Accuracy of Computations Involving Quantities Known Only to a Given Degree of Approximation"; and spent the remainder of his "spare time" on his "second job" with Professor Brigham and the College Entrance Examination Board. The following year, 1934-35, Sam's program was much the same, except that he now guided my post-graduate reading and study in probability and statistical theory and methodology in preparation for my becoming a doctoral candidate in Statistics under J. Neyman and E. S. Pearson at University College, London, 1935-37.

Wilks taught his first statistics course at the University of Pennsylvania, in Philadelphia, during 1935-36. (Dr. George Gailey Chambers, Professor of Mathematics, University of Pennsylvania, had died on 24 October 1935, shortly after his graduate course "Modern Theory of Statistical Analysis" had gotten under way. Sam was commissioned to complete the teaching of this course in his stead.) During the same period Sam gave an informal course--i.e., not listed in the official University course catalog--to three Princeton seniors, Walter W. Merrill, John O. Rohm, and William C. Shelton, on much the same material; and supervised Shelton's senior thesis on "Regression and Analysis of Variance". (Shelton continued in Statistics, rising to become Special Assistant to the Commissioner of Labor Statistics. Merrill and Rohm took up accounting and law, respectively.)

Wilks was promoted to an assistant professorship in 1936; and in 1936-37 taught his first statistics courses at Princeton: a graduate course during the Fall Term--see WILKS 1937--and an undergraduate course during the Spring Term. A Princeton senior that year who took the graduate course, Irving E. Segal (now a Prof. of Math at MIT), wrote a senior thesis under Sam's supervision that was subsequently published in the Proceedings of the Cambridge Philosophical Society (SEGAL 1938).

The publication, in the January 1973 issue of the IMS Bulletin, of Professor Harry C. Carver's letter of 14 April 1972 to Professor William Jackson Hall on the "beginnings of the Annals" prompts me to correct a mistaken conjecture contained therein on why Sam Wilks was not permitted to teach a course in mathematical statistics during his first few years as an instructor in the Mathematics Department there. Professor Carver wrote:

> "...one day I asked [Wilks] how it was that he was not teaching a course in mathematical statistics at Princeton. He replied that he had tried to start such a course there, but his superiors turned down his request each time,--probably because mathematical statistics and probability had not yet rung a bell in the staid Eastern Colleges."

The fact of the matter is that mathematical statistics and probability already had "rung a bell" at Princeton: two years before Wilks's arrival, Acheson J. Duncan had been sent off at University expense to study with Professors Henry Schultz and Harold Hotelling for the express purpose of readying himself to initiate courses in "mathematical economics" and "modern statistical theory" on his return. It was this prior arrangement and commitment, not lack of appreciation of the importance of mathematical statistics and probability--or of Wilks's exceptional qualifications--that constituted the primary obstacle to Wilks's offering an undergraduate course in mathematical statistics during his first three years as a member of the Mathematics Department of Princeton University. Duncan's course on "modern statistical theory" had been scheduled to be offered for the first time during the Fall Term of 1933 before the possibility of Wilks's coming to Princeton had even been considered. In view of the expense that the University had incurred in underwriting Duncan's year of training in preparation for the offering of this course, and the sacrifice that Duncan had made in postponing work on his doctoral dissertation in order to acquire the requisite training at the University's request, it would have been very improper and cruel to have shelved Duncan's course and let Wilks start one instead. I am sure that Wilks recognized this; and was also cognizant of the other factors that delayed his getting a course of his own in the Mathematics Department.

The three-year delay between Sam's arrival at Princeton and his first officially recognized course in statistics under the auspices of the Mathematics Department was the result of at least four factors.

First, there was the priority that circumstances had accorded to Duncan's course in the Department of Economics and Social Institutions. Furthermore, that Department had taken the initiative in the matter, and was desirous of modernizing its outlook and course offerings with respect to mathematical economics and statistics.[22]

Second, under the circumstances, any course on "mathematical statistics", "statistical analysis", "statistical inference", or whatever, to be offered by Wilks in the Mathematics Department would have to be an additional new course, and would require the approval of the all-powerful Course of Study Committee of the Faculty. A new course at Princeton had to be described in detail by the department proposing to offer it. Faculty approval gave the department the right to teach the described subject matter. I am not sure that this was an exclusive right, but I doubt that the Course of Study Committee would have approved teaching essentially the same material in two departments. Hence a major obstacle to Sam's teaching an undergraduate course in Statistics was the historical fact that Statistics had been the province of the Department of Economics and Social Institutions.

Third, until Sam was promoted to an assistant professorship in 1936, he was only an instructor; and in a department having the stature, nationally and internationally, of Princeton's Mathematics Department it was definitely not customary for an undergraduate, much less a graduate course, to be initiated by and be the sole responsibility of an individual with the rank of instructor.

A fourth, and very inhibiting factor was the unfavorable mathematical "climate" that prevailed in Fine Hall, which housed Princeton's Mathematics Department during Sam's early years at Princeton. Geometry had occupied the center of the stage in this Department, for over a quarter of a century, with Algebra and Analysis accorded much less exalted roles. Then, in 1932, the new Institute for Advanced Study, an institution completely distinct from Princeton University, had come into being, and the members of its School of Mathematics were granted office space in the Mathematics Department's Fine Hall until the completion of their first building, Fuld Hall, in 1939. Albert Einstein (1879-1955) arrived to take up his post in the Institute during the Winter of 1933, and Hermann Weyl (1885-1955) arrived a few months earlier. John Von Neumann (1903-1957) was already there (Lecturer, 1930-31, Princeton, then Professor of Mathematical Physics, 1931-33; Professor of Mathematics, Institute for Advanced Study, 1933-57); as were also E. U. Condon (1903-1961; Assistant Professor of Mathematical Physics, Princeton, 1928-31; Associate Professor, 1931-38, Professor, 1938-47), and E. P. Wigner (Lecturer in Mathematical Physics, Princeton, 1930; Professor, 1930-36; 1938-1971). With this galaxy of mathematical physicists all together ine one place for the first time, the mathematical theory of relativity and quantum mechanics were definitely the fashion of the day in Fine Hall--a difficult "climate" in which to initiate a program in mathematical statistics.

By 1936-37, the division of territory between the Department of Mathematics and the Department of Economics and Social Institutions had been resolved. The latter would be restricted to instruction in statistical theory and methods pertinent to the economic and social sciences; and the basic general undergraduate course(s) in statistical theory and methodology, and the graduate courses in advanced mathematical statistics would be the province of the Mathematics Department. As we have already said, Wilks taught his first statistics course at Princeton in the fall of 1936, the graduate course leading to his lithographed lecture notes on Statistical Inference--(1937); and in the spring of 1937, a sophmore course with calculus as prerequisite, quite possibly the first carefully formulated college underclass course in mathematical statistics at this level. It was offered thereafter for a number of years to students in all fields in the second half of the sophomore year. The material presented in this course, extended and polished, became generally available a decade later in his "blue book", Elementary Statistical Analysis (1948b). A third course, also one semester in length, was added in 1939-40. It was an upperclass course for students who wanted to specialize in statistics, and consisted of a rather thorough mathematical treatment of statistical theory in the classroom plus a laboratory section devoted to applications and computations. This course was taken also by beginning graduate students. Wilks's first doctoral student, Joseph F. Daly received his Ph.D. in 1939. George W. Brown and Alexander M. Mood followed in 1940. World War II demolished his plans for sabbatical leave to lecture in South America and accept an offered exchange professorship for one semester at the National University in Santiago, Chile. As World War II progressed, Sam became ever more deeply involved in war research--I shall return to this in a moment--and in due course was released from academic duties entirely. Helped by two of his graduate students, T. W. Anderson and D. F. Votaw, Jr., and Henry Scheffé, he succeeded in seeing through to lithoprinted publication the graduate level text, Mathematical Statistics (1943), before becoming totally involved in war work. This was the forerunner of his polished comprehensive treatment bearing the same title published as a type-set book in 1962.

In keeping with my father's policy of promotions as soon as merited without regard to leave of absence, Sam was promoted to a full Professor of Mathematics in 1944, effective on his return to academic duties; and plans were laid for a Section of Mathematical Statistics within the Department of Mathematics. Following the war there was a steady flow of able graduate students and postdoctoral research associates, some of whom, like Robert Hooke and Henry Scheffé, were changing from mathematics to statistics. By the time of Sam's death (1964), Princeton had granted Ph.D.'s to approximately 40 men in mathematical statistics and probability, all of whom had studied to some extent with Wilks, and the dissertations of about half had been supervised by him.

It would be a mistake to infer from the foregoing that Wilks's educational activities were limited to teaching and thesis guidance in mathematical statistics. He was deeply interested in the whole spectrum of mathematical education. In "Personnel and Training Problems in Statistics" (1947) he outlined the growing use of statistical methods, the demand for personnel, problems of training, and made recommendations that served as a guide in the rapid growth of university centers of training in statistics after World War II. Drawing on his experience at Princeton, he urged, in "Teaching Statistical Inference in Elementary Mathematics Courses" (1958), teaching the principles of statistical inference to freshman and sophmores, and further proposed revamping high school curricula in mathematics and the sciences to provide topics in probability, statistics, logic and other modern mathematical subjects. In furtherance of his ideas in this direction he co-authored, as a member of the Commission on Mathematics of the College Entrance Examination Board 1955-1958, the Introductory...Experimental Course (1957) that recommended major changes in the teaching of mathematics in the secondary schools and suggested inclusion of an option of Introductory Probability with statistical applications in the twelfth grade. During his last few years he worked with an experimental program in Miss Mason's School in Princeton which introduced new mathematics at the elementary level, down to kindergarden. During his final week of life, he was considering, as a member of the Advisory Board of the School Mathematics Study Group, how much time the following summer he would be able to devote to writing on probability and statistics for this group.

5. WILKS'S FURTHER CONTRIBUTIONS TO MATHEMATICAL STATISTICS. A few more words are in order on Wilks's further contributions to mathematical statistics before turning to his many services to the U.S. Government generally and to the Army in particular.

Wilks was definitely not an ivory tower researcher. A great many of his research papers in mathematical statistics were written to meet needs that he personally had encountered in his applied work; and, especially in his earlier papers, he usually included explicit worked examples of the application of the new theory concerned. Thus, his first important contribution to multivariate analysis after arriving in Princeton, "On the Independence of k Sets of Normally Distributed... Variables" (1935a), appears to have been written to meet a need Wilks encountered in his work with the College Entrance Examination Board in Princeton, N.J.; as do also many of his later contributions to multivariate analysis, e.g., "Weighting Systems for Linear Functions of Correlated Variables..." (1938b) and "Sample Criteria for Testing Equality of Means, Equality of Variance, and Equality of Covariances..." (1946);

and "Multivariate Statistical Outliers" (1963), the last of his total of fifteen research papers on topics in multivariate analysis, has a definitely applied flavor.

In addition to the extensive and penetrating studies of likelihood ratio tests for various hypotheses relating to multivariate normal distributions embodied in the aforementioned papers, Wilks investigated (1935b) likelihood ratio tests for various hypotheses relating to multinomial distributions and to independence in two- three- and higher-dimensional contingency tables, and provided (1938a) a compact proof of the basic theorem on the large-sample distribution of the likelihood ratio criterion for testing "composite" statistical hypotheses, i.e., when the "null hypothesis" tested specifies the values of, say, only m out of the h parameters of the probability distribution concerned. Jerzy Neyman's basic paper on the theory of confidence-interval estimation appeared in 1937. The following year Wilks showed (1938c) that, under fairly general conditions, confidence intervals for a parameter of a probability distribution based upon its maximum-likelihood estimator are on the average the shortest obtainable in large samples; and a year later, in a joint paper with J. F. Daly, generalized this result to the case of several parameters.

In response to a need expressed by Shewhart, Wilks, in "Determination of Sample Sizes for Setting Tolerance Limits" (1941), laid the foundations of the theory of statistical "tolerance limits", which are actually confidence limits, in the sense of Neyman's theory, not, however, for the value of some parameter of the distribution sampled as in Neyman's development, but rather for the location of a specified fraction of the distribution sampled. In this paper he showed that a suitably selected pair of ordered observations ("order statistics") in a sample of sufficient size from an arbitrary continuous distribution provide a pair of limits, statistical "tolerance limits", to which there corresponds a stated chance that at least a specified fraction of the underlying distribution is contained between these limits, thus providing the "distribution-free" solution needed when the assumption of an underlying normal distribution of industrial production is unwarranted. In the same paper he derived the corresponding parametric solution of maximum efficiency in the case of sampling from a normal distribution (based on the sample mean and standard deviation), and an expression for the relative efficiency of the distribution-free solution in this case. In "Statistical Prediction..." (1942), he found formulas for the probabilities that at least a fraction N_0/N of a second random sample of N observations from an arbitrary continuous distribution would (a) lie above the r^{th} "order statistic" (r^{th} observation in increasing order of size), $1 \leq r \leq n$, in a first random sample of size n from the same distribution; (b) be included between the r^{th} and s^{th} order statistics, $1 \leq r \leq s \leq n$, of the first sample; and illustrated the application of these results to the setting of one- and two-sided statistical tolerance limits. These papers embodied the earliest of a series of contributions made by Wilks to "nonparametric" or "distribution-free" methods of statistical inference, an area of research in which he persuaded a number of his students to write senior theses or doctoral dissertations; and of which he provided an extensive review in depth in "Order Statistics" (1948a), an expository paper that was in large part responsible for the ensuing blossoming of research activity in this area.

Wilks was one of the small group of mathematicians and statisticians who at Ann Arbor, Michigan, on September 12, 1935, founded the Institute of Mathematical Statistics, and thereafter was an active and leading member. At this meeting, Harry C. Carver, who had founded, edited, and personally financed and published the Annals of Mathematical Statistics (in affiliation with the American Statistical Association) from 1930, volunteered to turn over the editing and publication of Annals to the Institute as its official organ as soon as the Institute was able to assume these responsibilities. The Institute assumed full responsibility for the Annals, and Wilks took over as editor, with the June 1938 issue.[23/] He served as editor through the December 1949 issue, and guided the development of the Annals from a marginal journal with a small subscription list, to the foremost publication in its field, with a ten-fold increase in individual, and a five-fold increase in library subscriptions; and in the process, fostered the growth of the Institute, from a once marginal society to a mature international organization, large in both size and contribution. His editorship of the Annals was his greatest contribution to mathematical statistics.

In 1954 Wilks joined Walter Shewhart in editing the Wiley Publications in Statistics, a major U.S. publication effort that did much to change statistics from a subordinate branch of the social sciences in the 1930's, to a respected discipline in its own right with a large and solid literature in the 1960's.

6. HIS BROAD CAREER OF GOVERNMENT SERVICE, AND AS INITIATOR OF THESE EXPERIMENT DESIGN CONFERENCES. In 1936, when my father recommended Sam for promotion to Assistant Professor of Mathematics he noted in his recommendation that Sam had just received an appointment as a Collaborator in a United States Soil Conservation Program of the Department of Agriculture. A broad career of government service was underway that was to range widely and continue through the last twenty-eight years of his life. He served the United States Government as a member of the Applied Mathematics Panel, NDRC, OSRD, and director of its Princeton Statistical Research Group, 1942-1945; chairman, mathematics panel, Research and Development Board, DOD, 1948-1950; member, scientific advisory committee, Selective Service System, 1948-1953; "charter" member, ASA advisory committee to the Bureau of the Budget, 1951-1964;

member, divisional committee for the mathematical, physical and engineering sciences, NSF, 1952-1956; member, committee on battery additives, NAS, 1953; member, divisional committee for the social sciences, NSF, 1957-1962; member, scientific advisory board, NSA, 1953-1964 (chairman, 1958-1960); member, U.S. National Commission for UNESCO, 1960-1962; and academic member, Army Mathematics Advisory Panel (called "Army Mathematics Steering Committee", from 1956 on), 1954-1964. It was in this latter capacity that he initiated these Experiment Design Conferences.[24/]

General Leslie E. Simon, upon becoming Chief of the Research and Development Division in the Office, Chief of Ordnance, in 1951, entered into an agreement with Duke University to establish on that campus, an Office of Ordnance Research to sponsor external basic research initiated by non-government investigators with ordnance interests. Such research had always been carried out by all Army Technical Services, but previously under vague mandate and seldom on an appreciable scale. The level of effort had been wholly dependent on the sophistication of the administrators concerned. A Statistics Branch, and other units with statistical interests, were included in the setup.

In 1954 the Army Research Office--Durham (then the Office of Ordnance Research) upon the request of the Chief of Research and Development Division, Office, Assistant Chief of Staff G-4, Department of the Army, established the Army Mathematics Advisory Panel (AMAP) as an ad hoc committee to provide advice on the mathematical needs of the Army. (The Panel was reconstituted as a permanent body, the Army Mathematics Steering Committee, on 27 February 1956.)

Soon after its formation, the AMAP conducted a comprehensive inquiry into the Army's uses of mathematics; whether these uses could be advantageously extended; what future needs might be anticipated; and what measures might then be taken to insure a future capability adequate to these needs. As an academic member, Wilks surveyed thirty Army installations with the AMAP and reported that "the most frequently mentioned needs expressed by the scientific personnel were for greater knowledge of modern statistical theory of the design and analysis of experiments" (SIMON 1965, p. 958), clearly implying that a major deficiency of Army research, development and testing was insufficient use of modern statistical experiment design techniques. He proposed, therefore, that the Army establish a series of Army-wide conferences on design of experiments in Army research, development and testing. Dr. Frank E. Grubbs, who had chaired an Ordnance symposium on Statistical Methods in 1953,[25/] strongly indorsed Wilks' proposal for Army-wide conferences devoted primarily to design of experiments. General Simon gave the proposal a green light and his support. Upon making further inquiries it was found that a number of research workers at various facilities expressed an interest in contributing papers to such a conference. Others had unsolved or partially solved problems which they wished to present for discussion.

The AMAP decided to organize a three-day conference on the design of experiments with three kinds of sessions. The first group of sessions would consist of invited papers by well-known authorities on the philosophy and general principles of the design of experiments. The second group would consist of technical papers contributed by research workers from various Army research, development and testing facilities. The third group would be clinical sessions consisting of presentations and discussions of partially solved and unsolved problems which had arisen in these establishments.

Wilks agreed to serve as chairman of the first Conference, which was held on October 19-21, 1955 at the Diamond Ordnance Fuze Laboratories and the National Bureau of Standards in Washington, D.C. It was attended by over 230 registrants and participants representing some 50 organizations. Speakers and other participants in the conference came from the Bell Telephone Laboratories, Johns Hopkins University, Princeton University, Virginia Polytechnic Institute, Bureau of Ships, National Bureau of Standards, and 18 Army facilities.[26]

More specifically, the principal speakers, and their topics, were:

1. W. G. Cochran, The Philosophy Underlying the Design of Experiments.

2. Churchill Eisenhart, The Principle of Randomization in the Design of Experiments.

3. M. E. Terry, Finding Optimum Conditions by Experimentation.

4. Panel Discussion led by John W. Tukey on How and Where Do Statisticians Fit In. (The others on this Panel were: Besse B. Day, Cuthbert Daniel, Churchill Eisenhart, M. E. Terry, and S. S. Wilks).

5. W. J. Youden, Design of Experiments in Industrial Research and Development.

It was such a success that the Army has continued these conferences annually in October or November since 1955, following the same format. (See the Appendix for places and dates of the first nineteen Conferences, and names and topics of the invited speakers at these Conferences.) Wilks chaired the first nine of these Conferences (1955-1963), and wrote the Foreword to the Proceedings of the first eight. At the tenth

Conference, held in 1964 and dedicated to Wilks's memory, establishment of the Samuel S. Wilks Memorial Award and Medal was announced, to be administered by the American Statistical Association, and to be awarded annually "to a statistician...based primarily on his contributions...to the advancement of scientific or technical knowledge in Army statistics, ingenious application of such knowledge, or successful activity in the fostering of cooperative matters which coincidentally benefit the Army, the DOD and the Government, as did Samuel S. Wilks himself"; and the initial award presented to Dr. Frank E. Grubbs, Ballistic Research Laboratories, Aberdeen Proving Ground. In 1947, Wilks was awarded the Presidential Certificate of Merit for his contributions toward antisubmarine warfare and the solution of convoy problems; and the same year, the Centennial Alumni Award of the University of Iowa.

6. HIS DEATH, AND CONCLUDING REMARKS.

Sam became "my teacher" and guiding spirit at once in 1933; and in later years, he proved "a friend indeed", on a number of "difficult" occasions. He died most unexpectedly in his sleep, on March 7, 1964, at his home in Princeton, New Jersey. At that instant Statistics lost one of its greatest champions; government agencies, professional societies, and the field of education a devoted work mate, helping hand, and guide; and I, "my teacher" and "a friend indeed".

As W. G. Cochran has said: "He will be long remembered with affection and gratitude: no man of his generation did as much to ensure that the rapid growth of statistical theory, applications, and education in the United States took place along sound and healthy lines." (COCHRAN 1964, p. 191); and Egon S. Pearson: "...it is hard to think of any mathematical statistician of the past 30 years who combined to a greater extent an excellence in the field of theory with a power of inspiring confidence in government agencies, national research institutions, and educational authorities, as a wise counsellor in practical affairs." (PEARSON 1964, p. 597)

He is survived by his widow, Gena Orr Wilks, his son, Stanley N. Wilks; a brother, William Weldon Wilks, three granddaughters, one grandson; and a host of friends.

8. POSTSCRIPT AND ACKNOWLEDGMENTS. At the Tenth Conference (1964) dedicated to the memory of Professor Wilks, I spoke from notes on "Sam Wilks as I Remember Him". The material presented was for the most part subsequently written up and a typescript prepared, but unfortunately not in time for publication in the Proceedings of that Conference--nor in the Proceedings of the Eleventh Conference, as was suggested. Portions of the typescript were submitted to, and comments received in writing from

Alva E. Brandt, Acheson J. Duncan, the late Frederick F. Stephan (1903-1971) and the late George W. Snedecor (1881-1974). Large portions of that previous manuscript have been taken over bodily and incorporated in the present text, with revisions in the light of the comments received from the foregoing, for which I am very grateful. Use has also been made of comments received from Frederick Mosteller on the penultimate draft of a biography of Wilks prepared for publication in a forthcoming volume of the Dictionary of Scientific Biography (New York: Charles Scribner's Sons, Publishers, 1970-), likewise gratefully acknowledged. In addition, I have taken advantage of, and have very probably incorporated more than I realize, from the obituaries and other memorial articles on Wilks that have appeared during the past decade, especially: ANDERSON (1965), COCHRAN (1964), DIXON (1965), HANSEN (1965a, 1965b), MOOD (1965), MOSTELLER (1964, 1968), PEARSON (1964), SIMON (1965), STEPHAN AND TUKEY (1965), and TUKEY (1965). My thanks to these for what I have "borrowed", explicitly or otherwise. For whatever faults of commission or omission still afflict this memorial to Sam Wilks, I must assume full responsibility.

NOTES

1. Alexander Mc Farlane Mood was the second of Sam's graduate students to receive a Ph.D. (1940) in mathematical statistics from Princeton University. After teaching at the University of Texas, and serving as a statistician in the Bureau of Labor Statistics, Mood returned to Princeton during World War II as a research associate in the Statistical Research Group-Princeton, engaged in war research under Wilks's direction as an arm of the Applied Mathematics Panel (AMP) of the National Defense Research Committee (NDRC) of the Office of Scientific Research and Development (OSRD), under a contract between Princeton University and the OSRD. It was as a member of this group that he and Wilfrid J. Dixon wrote their famous memorandum, later published as an article in the Journal of the American Statistical Association (Vol. 43, No. (March 1948), 109-126), on the statistical theory of the "up-and-down" or "Bruceton" method of obtaining and analyzing sensitivity data, with which they had become acquainted in 1943 at the NDRC's Explosives Research Laboratory (now a unit of the Bureau of Mines, U.S. Department of the Interior), at Bruceton, Pennsylvania. Subsequently Mood became a professor of mathematical statistics at Iowa State College; deputy chief, mathematics division, RAND Corporation; president, General Analysis Corporation; a vice president of CEIR, Inc.; and at the time of writing his tribute to Wilks, was Assistant Commissioner of Education, U.S. Office of Education.

2. Dodd had joined the staff of the University of Texas in 1907 as Instructor in Pure Mathematics. He seems to have been silent publication-wise until 1912 when two papers by him appeared, one on plane and skew curves, and the other on the method of least squares and orthogonal transformations. These were followed immediately in 1913 by four papers on statistical properties of the arithmetic mean, the median and "other functions of measurements". One of these latter, entitled "The probability of the arithmetic mean compared with that of certain other functions of the measurements", was published in the Annals of Mathematics (Vol. 14, pp. 186-198, June 1913), of which my father (Luther Pfahler Eisenhart, 1876-1965) was then an editor. My father seems to have corresponded with Professor Dodd with regard to this paper. Thereafter Professor Dodd sent my father reprints of many of his subsequent papers on functional and statistical properties of various types of "means". These reprints proved to be very helpful to me when I became interested in such matters in the early '30's. I had the good fortune to meet Professor Dodd, when I went with Sam to the Joint Meeting of the American Mathematical Society and Institute of Mathematical Statistics in Indianapolis in December 1937. (For additional information on Dodd,

see footnote 4; J. C. Poggendorff, Biographisch-Literarisches Handwörterbuch für Mathematik..., Vol. 5 (1904-1922), Leipzig and Berlin, 1926, p. 299; and C. D. Simmons, "Edward Lewis Dodd, 1875-1943", Journal of the American Statistical Association Vol. 38, No. 222 (June 1943), 247-248.)

3. The University of Iowa in Iowa City (now known as the "State University of Iowa") was, in the 1920's, the leading center in the United States for research and training in mathematical statistics. It should not be confused with Iowa State College at Ames (renamed "Iowa State University" on the occasion of its centenary in 1958), which, during the same period, was the leading center for application of, and teaching the application of, modern statistical methods in the experimental sciences, especially in agricultural research and closely related fields.

4. Professor Dodd after receiving his Ph.D. in mathematics from Yale in 1904, had served as an Instructor in mathematics for two years (1904-06) at the University of Iowa, and one year (1906-07) at the University of Illinois, in Urbana. At the University of Illinois, Dodd had become acquainted with Rietz, who at that time was dividing his time about equally between his position of Assistant Professor of Mathematics in the Department of Mathematics, and his position of Statistician in the Experiment Station of the College of Agriculture. Rietz was teaching a course in the Mathematics Department entitled "Averages and Mathematics of Investment", which he had been induced to develop two years before, when a demand had arisen for a course in statistics which none of the members of the Mathematics Department were particularly prepared to give. Also, at that time Rietz was very busy working on his first publication in statistics, a 32 page appendix ("Statistical Methods. Appendix to Principles of Breeding") to A Treatise on Thremmatology by Eugene Davenport, Dean of the College of Agriculture and Director of the Agricultural Experiment Station (Boston: Ginn and Co., 1907, pp. 681-713); and also on his bulletin (with Dean Davenport) on Statistical Methods Applied to the Study of Type and Variability in Corn (Illinois Agriculture Experiment Station Bulletin No. 119, 1907). From then until he was called to the University of Iowa in 1918 as Head of the Department of Mathematics, Rietz published a long list of papers on statistical topics, some purely theoretical, some expositional, some arising out of his connection with the College of Agriculture. I mention these details to emphasize the fact that the development of statistical theory and methodology in the United States owes far more to the needs and support of workers in agriculture than many people realize today.

5. Under Rietz's leadership the University of Iowa rapidly became one of the leading centers of actuarial mathematics in the United States, and the leading center for research in mathematical statistics. (Other notable centers of actuarial mathematics and mathematical statistics were the University of Michigan, in Ann Arbor, under the leadership of James W. Glover (1868-1941) and Harry C. Carver, who in 1930 founded, and for five years personally financed the Annals of Mathematical Statistics; and Harvard University, under the leadership of Edward V. Huntington (1874-1952), Truman L. Kelley (1884-1961) and Warren M. Persons (1878-1937).) Two of Rietz's publications helped to firm up the University of Iowa's standing: (1) the Handbook of Mathematical Statistics (Boston: Houghton Mifflin Company, 1924) prepared by the "Members of the Committee on the Mathematical Analysis of Statistics of the Division of Physical Sciences of the National Research Council" (H. C. Carver, A. R. Crathorne, W. L. Crum, James W. Glover, E. V. Huntington, Truman L. Kelley, Warren M. Persons, H. L. Rietz, and Allyn A. Young) with Rietz serving as Editor-in-Chief; and (2) Rietz's own Carus Mathematical Monograph (No. 3) entitled Mathematical Statistics, published for the Mathematical Association of America by the Open Court Publishing Company in 1927, which served as the basis for courses in mathematical statistics given in Departments of Mathematics of many universities and colleges for years afterward. The jointly written Handbook was doomed, however, to become obsolete almost upon publication: the future of mathematical statistics was being shaped in the 1920's by the papers of R. A. Fisher; and the future of statistical methodology, by his Statistical Methods for Research Workers (1925), which rapidly became "the Bible" of statistical methodology Iowa State College, Ames, under the guidance of Professors George W. Snedecor (1881-1974) and A. E. Brandt. (For additional information on Rietz, see A. R. Crathorne, "Henry Lewis Rietz--In Memoriam", Annals of Mathematical Statistics, Vol. 15, No. 1 (March 1944), 102-108, which contains lists of selected publications of Rietz, of his books, and of doctorate dissertations written under his supervision; and Frank Mark Weida, "Henry Lewis Rietz, 1875-1943", Journal of the American Statistical Association, Vol. 39, No. 226 (June 1944), 249-250.)

6. After one year of graduate work in actuarial mathematics at Iowa, Curtiss decided against a career as an actuary, and went on to earn his Ph.D. in pure mathematics (analysis) at Harvard in 1935. However, five years later, as instructor in mathematics at Cornell University, and the most junior member of the Mathematics Department, he was assigned the responsibility of a course in mathematical statistics. To prepare for this course, to answer the teasing query of his senior colleagues, "What is there to statistics anyway?", he dug into the first ten volumes of the Annals of Mathematical Statistics, the first six volumes of the Supplement to the Journal of the Royal Statistical Society (borrowed from the late Frederick F. Stephan (1903-1971), and J. O. Irwin's series of reviews of "Recent Advances in Mathematical Statistics" in the Journal of the Royal Statistical Society,

and other sources. In the third of these reviews (for 1932), he no doubt noticed nine of the fourteen pages of the section on "Exact sampling distributions" were devoted to discussion of four papers of his friend Sam Wilks. During World War II, Curtiss, as a Lt. Commander, USNR, applied modern statistical theory and methodology to problems of naval engineering with considerable success in the Bureau of Ships of the U.S. Navy Department. (For discussion of some of these applications, see J. H. Curtiss "Statistical Inference Applied to Naval Engineering", Journal of the American Society of Naval Engineers, Vol. 58, No. 3 (August 1946), 335-398.) In April 1946, he was brought to the National Bureau of Standards by its new Director, Dr. E. U. Condon (1902-1974), and appointed statistical assistant to the Director for the express purpose of introducing modern statistical theory and methodology into the scientific and technical programs of the Bureau. However, before Curtiss could get such a program under way, Dr. Condon was obliged to turn over to him the day-to-day administration of the Bureau's new responsibilities in the development of large-scale automatic digital computers, and of an associated program of developing the mathematics of numerical analysis. John's original assignment at the Bureau was therefore placed on my shoulders, when I arrived at the Bureau to receive it on October 1, 1946--and the rest of that story you know.

7. See WILKS 1941, 1942, 1948; pp. 18-19 of ANDERSON 1965; and items (40), (41), and (45) in the list of "The Publications of S. S. Wilks" appended thereto.

8. Rietz gave a paper, "Comments on Applications of Recently Developed Theory of Small Samples", at the 92nd Annual Meeting of the American Statistical Association, Cleveland, Ohio, 30 December 1930, which saw publication in the Journal of the American Statistical Association, Vol. 26, No. 175 (June 1931), 150-158.

9. Thus Paul Rider, in a valuable review article, A Survey of the Theory of Small Samples (Annals of Mathematics, 2nd Series, Vol. 31, No. 4, (October 1930), pp. 577-628), which was later to "save my neck" on a number of occasions, wrote (p. 578):

> "Undoubtedly the leading writer in the theory of small samples is R. A. Fisher, whose work in this field has revolutionized modern sampling theory. Much of it is to be found in his book, Statistical Methods for Research Workers, but this book is extremely unsatisfying to a mathematician, as it merely states results without proofs and usually without even indicating how a given result may be derived. It discusses such things as the distribution of t without telling what the distribution is. His

original papers are much more enlightening, but from the references as given in the book it is sometimes difficult to tell which paper treats of a given topic. Even these papers suffer in places from the same defects as those of the book, and they are often troublesome to follow."

I don't know whether Paul later retracted these remarks, or Fisher was forgiving, because, when I got to University College, London, in 1935, to study under J. Neyman and E. S. Pearson, there was Paul sitting at a desk up in "Fisher territory" (the Galton Laboratory and Department of Eugenics), working on moment functions for Fisher's k-statistics in samples from a finite population.

10. Hotelling's paper on "The distribution of correlation ratios calculated from random data", in Proceedings of the National Academy of Sciences, 11, no. 10 (October 1925), 657-662, made him the first person in the United States to respond in kind to R. A. Fisher's signal contributions to the theory of small samples--his derivation employed the same kind of geometrical reasoning in terms of Euclidean N-dimensional space that Fisher had used so effectively. This paper carries a footnote that I've always considered to be very significant. I believe it affords an explanation of why so many American mathematicians had difficulty following Fisher's geometrical proofs. Anyone who attempts to duplicate Fisher's geometrical reasoning soon discovers that a crucial step is the correct evaluation of the relevant element of volume. Hotelling, at this juncture in his paper, gives a general expression for the relevant element of volume, which he numbers "(17)", and then remarks in a footnote:

"This important expression for the volume element has been used in lectures by [at Princeton University] by Professors O. Veblen and L. P. Eisenhart. I do not find it in any of the treatises on Calculus, Analysis or Differential Geometry, save for the special case in which the manifold of integration is a surface. It may readily be proved by showing first that (17) is a relative invariant under arbitrary transformations of the parameters; and second, that if the parameters of the hypersurface are orthogonal at a point, (17) becomes at this point the simple expression for the volume element in cartesian coordinates."

Hotelling had gone to Princeton University as a J.S.K. Fellow in mathematics, 1921-1922, after receiving his A. B. (1919) and an M.S. (1921) from the University of Washington, in Seattle. His interests in statistics predated his going to Princeton in the Fall of 1921. He had hoped to find some work in probability theory and the mathematics of statistics going on there in the Mathematics Department. Finding none, he undertook instead a program of study and research in topology (then called "analysis situs") and differential geometry, under the direction of Professor Oswald Veblen (1880-1960) and my father, Luther Pfahler Eisenhart (1876-1965). He stayed on at Princeton, 1922-1924, as an Instructor in Mathematics and received his Ph.D. from Princeton University in June 1924, his doctoral dissertation being on "Three-dimensional manifolds of states in motion." In 1927 he published a paper "An application of analysis situs to statistics" (Bulletin of the American Mathematical Society, Vol. 33, (1927), pp. 467-476), which had to do with topological aspects of serial and multiple correlations.

Following receipt of his Ph.D., Hotelling returned to the West Coast, to Stanford University, where he was a Junior Research Associate (1924-25), and then Research Associate (1925-27), in the Food Research Institute; and finally, an Associate Professor of Mathematics (1927-31), in the Department of Mathematics. Hotelling visited Fisher in England, in 1929, hoping to persuade Fisher to join with him in the preparation of an up-to-date textbook on the mathematics of Statistical Inference. Fisher was not interested in the proposition. In 1931, Hotelling was called to Columbia University, in New York City, as Professor of Economics to develop further the existing work there in Mathematical Economics, and to initiate a program in Mathematical Statistics.

11. These papers had been followed by their more elegantly written "On the problem of two samples" (Bulletin de l'Académie Polonaise et des Lettres, Series A, 1930, 471-494), and "On the problem of k samples" (idem, 1931, 460-481), in which the likelihood-ratio technique had led directly to the now famous test for the homogeneity of variance involving the ratio of the weighted arithmetic mean of the sample variances (with weights subsequently modified by Bartlett). This great discovery was discussed by Pearson in one of his lectures, and no doubt contributed to Sam's enthusiasm for likelihood-ratio tests.

It was too early to claim that the tests thus found were "best" in some sense inasmuch as the Neyman-Pearson Lemma was yet to come in J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypothesis", communicated to the Royal Society of London in August 1932, "read" to the Society on November 10, 1932, and published on February 16, 1933 in the Society's Philosophical Transactions, Series A, Vol. 231, pp. 289-337; which, incidentally was refereed by Fisher who, at the time, considered it an important step forward.

12. Time and again during his years at the Bureau I would hear him tell a consultee, or an audience, that he was "a chemist", implying that he was not a statistician. Well, Jack may have been all chemist at one time, but by 1931 he was already on his way to becoming an exponent and practitioner of Fisherian methods too. He had come upon Student's t test "by accident...in 1925" (W. J. Youden, Risk, Choice and Prediction: An Introduction to Experimentation, Duxbury Press, North Scituate, Mass., 1974, p. 5). By the "summer of 1931 [he] had obtained one of the 1050 copies printed of the first edition" of Fisher's Statistical Methods for Research Workers (1925), and when Fisher "visited Cornell" to attend the 6th International Congress of Genetics, 24-31 August 1931, Youden "drove there...to show him an experimental arrangement". (Quotations are from p. 727 of W. J. Youden, "Memorial to Sir Ronald Aylmer Fisher," Journal of the American Statistical Association, Vol. 57, No. 300 (Dec. 1962), 727-728.) From Hotelling's lectures Youden "first got some hint that [Fisher's Statistical Methods...] also held a message for mathematicians...He told the young men listening to him not to be misled by the large print, the wide margins, and a text almost devoid of mathematical symbols, that in this book were concepts as new to the theorists as to the researchers". (Quoted from p. 47 of W. J. Youden, "The Fisherian Revolution in Methods of Experimentation," Journal of the American Statistical Association, Vol. 46, No. 253 (March 1951), 47-50.) During the next few years he published a variety of papers expounding and demonstrating the application of known statistical techniques to various problems arising in studies of apples, seeds, soils, leaves, tomatoes, trees and viruses. He had clearly "crossed the Rubicon"; was on his way to becoming an expert expositor and practitioner of statistical methods in experimentation; and from then on he became more and more of a statistician--or shall we say, "experimentrician"-- and less and less "chemist".

13. Spearman devoted over 40 years of his life to the development of a psychological theory of mental ability built around a General Factor, g, that characterizes an individual's "general mind power"--see his The Abilities of Man (New York: The Macmillan Company, 1972); but is most widely known among statisticians today for a comparatively minor contribution, his coefficient of rank-order correlation (1904).

14. Whether the population tetrad differences, $\tau_{1234} = \rho_{12}\rho_{34} - \rho_{13}\rho_{24}$ and $\tau_{1324} = \rho_{13}\rho_{24} - \rho_{14}\rho_{23}$, were both zero, both non-zero, or one zero and the other non-zero, where ρ_{ij} is the coefficient of correlation between the i-th and j-th traits, was of decisive importance in Spearman's theory of mental abilities of man.

15. Rietz had chaired the session on Statistical Methodology on the first day of the 92nd Annual Meeting of the American Statistical Association in Cleveland, Ohio, December 29-31, 1930, at which Shewhart had

presented his paper on "Statistical Method from an Engineering Viewpoint" (published in the Proceedings of the Meeting as "Applications of Statistical Method in Engineering", Journal of the American Statistical Association, Vol. 26, March 1931 Supplement, pp. 214-221); and the following day Shewhart had been the invited discussant of Hotelling's paper on "Recent Improvements in Statistical Inference" (same Supplement, pp. 79-87; discussion, pp. 87-89).

16. This was Karl Pearson's last year as the first Galton Professor of National Eugenics (1911-1933), as Editor of the Annals of Eugenics, which he had founded and edited since 1925, and as Head of the Department of Applied Statistics (1911-1933), which included the Biometric Laboratory (which Pearson had originated in 1895, as a center for postgraduate study in this new branch of applied mathematics when Goldsmid Professor of Applied Mathematics and Mechanics (1884-1911)) and the Francis Galton Laboratory of National Eugenics (which had been formed, and placed under Pearson's direction, in 1906 at Galton's request, as successor to Galton's own Eugenics Records Office established at University College in 1904 by a gift from Galton to the University of London for this purpose). He continued, however, to edit Biometrika, of which he was one of the three founders, always the principal editor (vols. 1-28, 1901-1936), and for many years the sole editor; and had almost seen the final proofs of the first half of volume 28 through the press when he died on 27 April 1936.

When I arrived at University College in October 1935 as a Ph.D. candidate in statistics, we were told that Karl Pearson's strength was rapidly failing, that he was still driving himself to shut out his grief over the thwarting of his ideal of an Applied Statistics Institute (with Readers in Genetics, Medicine, Psychology, Mathematical Statistics, etc.) by the break up of his Department into separate Departments of Eugenics and Applied Statistics; and that he was very reluctant to see visitors. The end came before Paul Rider and I and many of our fellow students were granted opportunities to meet him. I have never quite recovered from that lost opportunity.

17. E. S. Pearson had spent some time with Shewhart and his colleagues at the Bell Telephone Laboratories during his 1931 visit to the United States. He was one of the early exponents in England of Shewhart's control-chart techniques, and at the time of Sam's visit was engaged in the preparation of a paper on "Statistical Method in the Control and Standardization of the Quality of Manufactured Products", presented at the December 1932 meeting of the Royal Statistical Society, and later published in the Society's Journal, (Vol. 96 (1933), pp. 21-60). This paper was largely responsible for the formation of the Industrial and Agricultural Research

Section of the Royal Statistical Society on November 23, 1933, and the subsequent publication of the now-famous Supplement to the Journal of the Royal Statistical Society to provide a medium for publication of papers of this Section. (For further details, see E. S. Pearson, "Some Historical Reflections on the Introduction of Statistical Methods in Industry: The Statistician, Vol. 22, No. 3 (Sept. 1973), 165-179.)

18. Stanley, like his father, received an A.B.--but in mathematics, not architecture--from North Texas State College ("Teacher's" having been dropped from the name) in 1955. He studied at Cambridge University 1955-1956; married Jocelyn Wilkins, daughter of a classmate of Sam's at North Texas State, in 1958; received an M.S. in applied mathematics from Columbia University in 1961; has three daughters and a son; and works for the Department of Defense as a mathematician.

19. John Wishart had gained First Class Honors Degree in Mathematics and Natural Philosophy at the University of Edinburgh, in Scotland, in 1922. At Edinburgh he had attended the lectures of E. T. Whittaker (1873-1956), on "The Calculus of Observations" which were later to appear in book form (T. WHITTAKER and ROBINSON, The Calculus of Observations, London and Glasgow: Blackie and Son, Ltd., 1924), and had learned numerical mathematics "the hard way", i.e., without the benefit of a desk calculator, in Whittaker's Mathematical Laboratory. In the autumn of 1924, Wishart had joined Karl Pearson at University College, as a Research Assistant. One of Wishart's main tasks on arriving there was to get work on Pearson's Tables of the Incomplete Beta-Function underway.

Wishart stayed with Pearson for three years and then in the autumn of 1927 accepted a teaching position at the Imperial College of Science and Technology (of the University of London), inasmuch as he was a teacher by training and temperament. While still with Pearson he had collaborated with R. A. Fisher on a joint paper "On the distribution of the error of an interpolated value and on the construction of tables" (Proceedings of the Cambridge Philosophical Society, Vol. 23, Part 8 (October 1927), pp. 917-921). He was barely settled in his new post at Imperial College, when, at the beginning of 1928, he was offered and accepted an appointment as Statistical Assistant to R. A. Fisher at Rothamsted Experiment Station. With Fisher's encouragement, he derived "The generalized product moment distribution in samples from a normal multivariate population" (Biometrika, Vol. 22A, Parts 1&2 (July 1928), pp. 31-52), by a geometrical argument analogous to those used previously by Fisher, the simultaneous distribution of the sample estimates of the variances and covariances of a multivariate normal population corresponding to a sample of N items from such a population, and prepared an extensive tabulation of the moments and product moments of this distribution, which is now known as "Wishart's distribution". Wishart, during his three years at Rothamsted (1928-31) participated fully not only in the mathematical research on sampling distributions and their properties, but also in the advisory and service activities of Rothamsted Statistical Department during that period, as reflected by the twenty publications of which he was the single or joint author during this period.

In October 1931, a few months after G. U. Yule's retirement from full-time teaching as Reader in Statistics in the University of Cambridge, Wishart was appointed to a newly created post of Reader in Statistics in the Faculty of Agriculture, with responsibilities also for some teaching in the Faculty of Mathematics. This was an exceptionally fine appointment: at Cambridge, as at other English universities, a Readership is only one step below a Professorship, and until the late 1950's Professorships were very few and far between, there ordinarily being only one per established discipline (e.g., Mathematics), which Statistics certainly was not at that time. (Thus Yule himself had been merely a University lecturer in Statistics from 1912 until only a few months before his premature retirement owing to ill health). Wishart saw in his Cambridge appointment an opportunity to introduce statistics to mathematical undergraduates, and began at once to offer not only a general course on statistical methods in the Faculty of Agriculture, but also a course on mathematical statistics which undergraduate students in the Faculty of Mathematics could offer for Schedule B of the Mathematical Tripos. Among his early students in this program were M. S. Bartlett (B. A., Queens' College, 1932) and W. G. Cochran (B.A., St. John's College, 1933). (For additional information on Wishart, see E. S. Pearson, "John Wishart, 1898-1956", Biometrika, Vol. 44, Pts. 1&2 (June 1957), 1-8, which includes a bibliography of his published work; and M. S. Bartlett, "John Wishart, D.Sc., F.R.S.E.", Journal of the Royal Statistical Association, Series A, Vol. 119, Pt. 4 (1956), 492-493.)

20. A few words are in order on how my father became interested in, and partial to statistics.

My father's primary mathematical interest was differential geometry, and his research was exclusively in that area. Exactly when he began to take an "outside" interest in mathematical statistics I do not know. It may have been as early as 1913, when as noted earlier, he corresponded with Edward L. Dodd on various aspects of the latter's paper entitled "The probability of the arithmetic mean compared with that of certain other functions of measurements", which was published in the Annals of Mathematics (Vol. 14, pp. 186-198, June 1913), of which my father was then an editor. At any rate, thereafter Dodd sent my father reprints of many of his subsequent papers on functional and statistical properties of various types of "means", which my father kept and ultimately turned over to me when I became interested in such matters in the early '30's.

Early in 1924, "at the request of the Commission on New Types of Examination of the College Entrance Examination Board", my father "formed a committee of mathematicians to examine critically certain statistical methods used in the investigations of the Commission" (American Mathematical Monthly, Vol. 31, No. 4 (April 1924), p. 209). The "mathematicians" of the Committee included the economic statisticians W. Randolph Burgess and W. L. Crum (1894-1967) of the Federal Reserve System and Economics Department, Harvard, respectively; the mathematicians E. V. Huntington (1874-1952) and J. H. M. Wedderburn (1882-1948), of Harvard and Princeton, respectively; and the mathematical statistician, H. L. Rietz.

The findings of this Committee, my father's continued advisory relations with the higher-ups of the College Entrance Examination Board (CEEB), and Wilks's contributions at Iowa (and under Hotelling at Columbia) to the solution of statistical problems arising in educational testing, made it possible for my father to arrange a part-time appointment with the CEEB concurrent with his initial University appointment--a relationship with the Board, and its successor, the Educational Testing Service, that continued until Wilks's death.

As mentioned earlier, Hotelling, after receiving his Ph.D. in mathematics from Princeton in 1924, went to Stanford University, first to a position in the Stanford Food Research Institute, later in the Mathematics Department, Stanford University. During these years at Stanford (1924-1931) he wrote and published a stream of important original contributions to statistical theory and mathematical economics; reviews of American and English books on statistical methods, (e.g., of Statistical Analysis by Edmund E. Day (New York: The Macmillan Company, 1925), in Journal of the American Statistical Association, Vol. 21, No. 155 (Sept. 1926), 360-363), in which he deplored the obsoleteness of teaching and research in statistics in the United States and placed the blame squarely on the doorsteps of Departments of Mathematics; and expository articles on "British statistics and statisticians today" (Journal of the American Statistical Association, Vol. 25, No. 170 (June 1930), 186-190), "Recent improvements in statistical inference" (cited fully in footnote 15), etc., in which he did his very best to acquaint American readers with the "new look" in statistics. He regularly sent reprints of all of these to my father. When my father gave them to me in the Fall of 1932, as I was reading up on "Student-Fisher statistics", it was quite clear that my father had more than a superficial knowledge of the papers on statistical theory, and had "got the message" of Hotelling's book reviews and expository articles.

21. This assignment was very disruptive to Duncan at that time. When asked to undertake it he was already at work on his doctoral dissertation on "South African gold and international trade"; and his acceptance of it delayed until 1936 his completion of the requirements for his Ph.D. in Economics. He also lost out on one of the features that "sweetened" the proposition, an opportunity to visit the West Coast--when the plans were made, Hotelling was at Stanford University, but had moved on to Columbia University before the time arrived for Duncan to study under him. This assignment was to be instrumental in changing the direction of Duncan's subsequent career.

22. The aim of the Department of Economics and Social Institutions was to improve its own offerings in statistics for economics students by integrating and updating the Smith-Duncan sequence of courses within that department. The extent to which this aim was achieved is evidenced by the two volumes Fundamentals of the Theory of Statistics: Vol. 1, Elementary Statistics and Applications; Vol. 2, Sampling Statistics and Applications, authored jointly by Professors Smith and Duncan and published by the McGraw-Hill Book Company, Inc., in 1944, 1945, respectively.)

23. For further details on the founding and early years of the Annals of Mathematical Statistics see the letter from Harry C. Carver, dated 14 April 1972, to Professor [W. J.] Hall, reproduced in the Institute of Mathematical Statistics Bulletin, 2, No. 1 (Jan. 1973), 11-14; and Allen T. Craig, "Our Silver Anniversary", in Annals of Mathematical Statistics, 31, no. 4 (Dec. 1960), 835-837.

24. The material of the five following paragraphs is taken for the most part from MALONEY 1962 and SIMON 1965, where further details can be found on the history of statistical methodology in Army research, development and testing.

25. See Proceedings of the First Symposium on Statistical Methods: Sampling Techniques (4-5 November 1953), Ballistic Research Laboratories Report No. 897, Aberdeen Proving Ground, Maryland, January 1954.

26. Proceedings of the First Conference on the Design of Experiments in Army Research, Development and Testing, Office of Ordnance Research Report No. 57-1, Office of Ordnance Research, Durham, North Carolina, June 1957.

REFERENCES*

AMERICAN STATISTICAL ASSOCIATION (1965). Memorial to Samuel S. Wilks. Journal of the American Statistical Association, 60, no. 312 (Dec. 1965), 938-966. Eight parts:

1. Frederick F. Stephan and John W. Tukey, "Sam Wilks in Princeton" (939-944).

2. Frederick Mosteller, "His writings in applied statistics" (944-953).

3. Alexander M. Mood, "His philosophy about his work" (953-955).

4. Morris H. Hansen, "His contributions to government" (955-957).

5. Leslie E. Simon, "His stimulus to Army statistics" (957-962).

6. Morris H. Hansen, "His contributions to the American Statistical Association" (962-964).

7. W. J. Dixon, "His editorship of the Annals of Mathematical Statistics" (964-965).

8. (unsigned), "The Wilks Award" (965-966).

ANDERSON, T. W. (1965). Samuel Stanley Wilks, 1906-1964. Annals of Mathematical Statistics, 36, no. 1 (Feb. 1965), 1-27.

CARVER, HARRY C. (1973). Letter to Wm. J. Hall on "memories of events associated with the founding of the Annals and the IMS". Institute of Mathematical Statistics BULLETIN 2, no. 1 (Jan. 1973), 11-14.

COCHRAN, W. G. (1964). Samuel Stanley Wilks, 1906-1964. Review of the International Statistical Institute 32, pts. 1/2 (1964), 189-191.

COLLEGE ENTRANCE EXAMINATION BOARD (1957). Introductory Probability and Statistical Inference: An Experimental Course. Preliminary edition, College Entrance Examination Board, New York, 1957; revised edition, idem, Princeton, New Jersey, 1959; Spanish translation, Tosorio, Argentina, 1961.

CRAIG, A. T. (1932). On the distribution of certain statistics. American Journal of Mathematics 54, no. 2 (April 1932), 353-366.

---

*Exclusive of references incorporated in footnotes in support of statements there and not otherwise relevant to Wilks.

DIXON, W. J. (1965). See AMERICAN STATISTICAL ASSOCIATION (1965), pt. 7.

FISHER, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. Proceedings of the Royal Society, A, 121, no. A788 (1 Dec. 1928), 654-673.

HANSEN, MORRIS H. (1965a, 1965b). See AMERICAN STATISTICAL ASSOCIATION (1965), pts. 4 and 6.

IRWIN, J. O. (1934). Recent advances in mathematical statistics (1932). Journal of the Royal Statistical Society, 97, pt. 1 (Mar. 1934), 114-154.

LINDQUIST, E. F. (1931). The significance of a difference between "matched" groups. Journal of Educational Psychology, 22 (March 1931), 197-204.

MALONEY, CLIFFORD J. (1962). Statistics in Army research, development and testing. The American Statistician 16, no. 3 (June 1962), 13-17.

MOOD, A. M. (1965). See AMERICAN STATISTICAL ASSOCIATION (1965), pt. 3.

MOSTELLER, FREDERICK (1964). Samuel S. Wilks: statesman of statistics. The American Statistician 18, no. 2 (Apr. 1964), 11-17.

MOSTELLER, FREDERICK (1965). See AMERICAN STATISTICAL ASSOCIATION (1965), pt. 2.

MOSTELLER, FREDERICK (1968). S. S. Wilks. International Encyclopedia of the Social Sciences, 16 (New York, 1968), 550-553.

NEYMAN, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. Philosophical Transactions of the Royal Society, A, 236, no. 767 (30 Aug. 1937), 333-380 (repr. as paper no. 20 in A Selection of Early Statistical Papers of J. Neyman [Cambridge; Berkeley and Los Angeles, 1967].

NEYMAN, J. and PEARSON, E. S. (1928a). On the use and interpretation of certain test criteria. Part I. Biometrika 20A, pts. 1-2 (July 1928), 175-240 (repr. as paper no. 1 in Joint Statistical Papers of J. Neyman and E. S. Pearson [Cambridge; Berkeley and Los Angeles, 1967]).

NEYMAN, J. and PEARSON, E. S. (1928b). On the use and interpretation of certain test criteria. Part II. Biometrika 20A, pts. 3-4 (Dec. 1928), 263-294 (repr. as paper no. 2 in Joint...Papers).

NEYMAN, J. and PEARSON, E. S. (1931). On the problem of k samples. Bulletin International de l'Académie Polonaise des Sciences et des Lettres, Année 1931, no. 6A (June 1931), 460-481 (repr. as paper no. 4 in Joint...Papers).

PEARSON, E. S. (1964). Samuel Stanley Wilks, 1906-1964. Journal of the Royal Statistical Society, A, 127, pt. 4 (1964), 597-599.

PEARSON, E. S. and WILKS, S. S. (1933). Methods of statistical analysis appropriate for k samples of two variables. Biometrika 25, pts. 3-4 (Dec. 1933), 353-378.

SEGAL, IRVING E. (1938). Fiducial distribution of several parameters with application to a normal system. Proceedings of the Cambridge Philosophical Society 34, pt. 1 (Jan. 1938), 41-47.

SHEWHART, WALTER A. (1931). Economic Control of Quality of Manufactured Product (New York: D. Van Nostrand Company, Inc., 1931).

SIMON, L. E. (1965). See AMERICAN STATISTICAL ASSOCIATION (1965), pt. 5.

STEPHAN, FREDERICK F. and TUKEY, JOHN W. (1965). See AMERICAN STATISTICAL ASSOCIATION (1965), pt. 1.

TUKEY, JOHN W. (1965). Samuel Stanley Wilks (1906-1964). Yearbook of the American Philosophical Society, 1964 (Philadelphia, 1965), 147-154.

WILKS, S. S. (1931). The standard error of the means of 'matched' samples. Journal of Educational Psychology, 22, no. 3 (March 1931), 205-208.

WILKS, S. S. (1932a). On the distributions of statistics in samples from a normal population of two variables with matched sampling of one variable. Metron, 9, nos. 3-4 (March 1932), 87-126.

WILKS, S. S. (1932b). Moments and distributions of estimates of population parameters from fragmentary samples. Annals of Mathematical Statistics, 3, no. 3 (Aug. 1932), 163-195.

WILKS, S. S. (1932c). On the sampling distribution of the multiple correlation coefficient. Annals of Mathematical Statistics 3, no. 3 (Aug. 1932), 196-203.

WILKS, S. S. (1932d). The standard error of a tetrad in samples from a normal population of independent variables. Proceedings of the National Academy of Sciences 18, no. 8 (August 1932), 562-565.

WILKS, S. S. (1932e). Certain generalizations in the analysis of variance. Biometrika, 24, pts. 3-4 (Nov. 1932), 471-494.

WILKS, S. S. (1934). Moment-generating operators for determinants of product moments in samples from a normal system. Annals of Mathematics, 35, no. 2, (Apr. 1934), 312-340.

WILKS, S. S. (1935a). On the independence of k sets of normally distributed statistical variables. Econometrica, 3, no. 3 (July 1935), 309-326.

WILKS, S. S. (1935b). The likelihood test of independence in contingency tables. Annals of Mathematical Statistics, 6, no. 4 (Dec. 1935), 190-196.

WILKS, S. S. (1937). Lectures by S. S. Wilks on the Theory of Statistical Inference 1936-1937, Princeton University. Edwards Brothers, Ann Arbor, Michigan.

WILKS, S. S. (1938a). The large-sample distribution of the likelihood ratio for testing composite hypotheses. Annals of Mathematical Statistics, 9, no. 1 (Mar. 1938), 60-62.

WILKS, S. S. (1938b). Weighting systems for linear functions of correlated variables when there is no dependent variable. Psychometrika, 3, no. 1 (Mar. 1938), 23-40.

WILKS, S. S. (1938c). Shortest average confidence intervals from large samples. Annals of Mathematical Statistics, 9, no. 3 (Sept. 1938), 166-175.

WILKS, S. S. (1941). Determination of sample sizes for setting tolerance limits. Annals of Mathematical Statistics, 12, no. 1 (Mar. 1941), 91-96.

WILKS, S. S. (1942). Statistical prediction with special reference to the problem of tolerance limits. Annals of Mathematical Statistics, 13, no. 4 (Dec. 1942), 400-409.

WILKS, S. S. (1943). Mathematical Statistics. Princeton University Press, Princeton, New Jersey.

WILKS, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate population. Annals of Mathematical Statistics, 17, no. 3 (Sept. 1946), 257-281.

WILKS, S. S. (1947). Personnel and training problems in statistics. American Mathematical Monthly, 54, no. 9 (Nov. 1947), 525-528.

WILKS, S. S. (1948a). Order statistics. Bulletin of the American Mathematical Society, 54, no. 1 (Jan. 1948), 6-50.

WILKS, S. S. (1948b). Elementary Statistical Analysis. Princeton University Press, Princeton, N. J.

WILKS, S. S. (1958). Teaching statistical inference in elementary mathematics courses. American Mathematical Monthly, 65, no. 3 (Mar. 1958), 143-152.

WILKS, S. S. (1962). Mathematical Statistics. John Wiley and Sons, Inc., New York.

WILKS, S. S. (1963). Multivariate statistical outliers. Sankhya, 25A, pt. 4 (Dec. 1963), 407-426.

WILKS, S. S. and DALY, J. F. (1939). An optimum property of confidence regions associated with the likelihood function. Annals of Mathematical Statistics, 10, no. 4 (Dec. 1939), 225-235.

WISHART, J. and BARTLETT, M. S. (1933). The generalized product moment distribution in a normal system. Proceedings of the Cambridge Philosophical Society, 29, pt. 2 (10 May 1933), 260-270.

APPENDIX

Places, Dates and Hosts of Conferences on the Design of Experiments in Army Research, Development and Testing, with Names and Topics of Invited Speakers

1st: Washington, D.C., 19-21 October 1955.
Diamond Ordnance Fuze Laboratories and National Bureau of Standards.

W. G. COCHRAN, "The Philosophy Underlying the Design of Experiments."

CHURCHILL EISENHART, "The Principle of Randomization in the Design of Experiments."

M. E. TERRY, "Finding Optimum Conditions by Experimentation."

W. J. YOUDEN, "Design of Experiments in Industrial Research and Development."

Panel Discussion on "How and Where Do Statisticians Fit In."
The Panel: John W. Tukey, Chairman, Cuthbert Daniel, Besse Day, Churchill Eisenhart, M. E. Terry, and S. S. Wilks.

2nd: Washington, D.C., 17-19 October 1956.
Diamond Ordnance Fuze Laboratories and National Bureau of Standards.

C. A. BENNETT, "The Predesign Phase of Large Sample Experiments."

R. A. BRADLEY, "Recent Research in Statistical Problems in Subjective Testing."

B. G. GREENBERG, "Application of Order Statistics in Medical Experiments."

G. E. NICHOLSON, JR., "The Planning of Experiments in the Presence of Variation."

M. B. WILK, "Derived Linear Models in the Analysis of Variance."

JEROME CORNFIELD, "Choice of Error in the Design of Experiments."

3rd: Washington, D.C., 16-18 October 1957.
Diamond Ordnance Fuze Laboratories and National Bureau of Standards.

BENJAMIN EPSTEIN, "Life Testing."

R. A. FISHER, "Practical Problems in Experimental Design."

H. O. HARTLEY, "Changes in the Outlook of Statistics Brought About by Modern Computers."

A. W. MARSHALL, "Experimentation by Simulation and Monte Carlo."

4th: Natick, Massachusetts, 22-24 October 1958.
Quartermaster Research and Engineering Center.

C. I. BLISS, "Some Statistical Aspects of Preference Studies."

A. C. COHEN, "Simplified Computational Procedures for Estimating Parameters of a Normal Distribution from Restricted Samples."

A. W. KIMBALL, "Errors of the Third Kind in Statistical Consulting."

C. F. KOSSACK, "The AASHO Road Test as an Example of Large Scale Tests."

L. H. C. TIPPETT, "Statistical Methods Applied to the Textile Industry."

5th: Fort Detrick, Maryland, 4-6 November 1959.
U.S. Army Biological Warfare Laboratories.

JOSEPH BERKSON, "The Measure of Death."

H. A. DAVID, "The Method of Paired Comparisons."

D. B. DeLURY, "Sampling in Biological Populations."

W. J. DIXON, "Medical Health Statistics."

N. E. GOLOVIN, "Prediction of the Reliability of Complex Systems."

RICHARD WEISS, "The Army Research and Development Program as it Relates to the Civil Economy."

6th: Aberdeen Proving Ground, Maryland, 19-21 October 1960.
Ballistic Research Laboratories.

JAMES R. DUFFETT, "Reliability."

F. J. ANSCOMBE, "Examination of Residuals."

W. S. CONNOR, "Developments in the Design of Experiments."

J. E. JACKSON, "Multivariate Analysis Illustrated by Nike-Hercules:
I. Separation of Product and Measurement Variability.
II. Acceptance Sampling."

7th: Fort Monmouth, New Jersey, 18-20 October 1961.
U.S. Army Signal Research & Development Laboratory.

G. A. WATTERSON, "Time Series and Spectral Analysis."

J. M. HAMMERSLEY, "Monte Carlo Methods."

R. L. ANDERSON, "Designs for Estimating Variance Components."

G. S. WATSON, "Hazard Analysis."

8th: Washington, D.C., 24-26 October 1962.
Walter Reed Army Institute of Research, Walter Reed Army Medical Center.

EGON S. PEARSON, "A Statistician's Place in Assessing the Likely Operational Performance of Army Weapons and Equipment."

MARVIN A. SCHNEIDERMAN, "A General Survey of Screening Theory."

HERMAN CHERNOFF, "Optimal Design Experiments."

R. P. ABELSON, "An Experimental Design for Decisions Under Uncertainty."

H. C. BATSON, "Bio-Assay."

9th: Redstone Arsenal, Alabama, 23-25 October 1963.
Directorate of Research and Development, U.S. Army Missile Command.

SOLOMON KULLBACK, "Communication Theory."

FRANK PROSCHAN, "The Concept of Monotone Hazard Rate in Systems Reliability."

CHURCHILL EISENHART, "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems."

H. O. HARTLEY, "Nonlinear Estimation."

D. B. DUNCAN, "On the Simultaneous Estimation of a Missle Trajectory and the Error Variance Components Including the Error Power Spectra of Several Tracking Systems."

10th: Washington, D.C., 4-6 November 1964.
Army Research Office, Office Chief of Research and Development, Department of the Army.

MAJ. GEN. LESLIE E. SIMON (Ret'd), "The Stimulus of S. S. Wilks to Army Statistics."

OSCAR KEMPTHORNE, "Development of the Design of Experiments Over the Past Ten Years."

H. O. HARTLEY and A. W. WORTHAM, "Assessment and Correction of Deficiences in Pert Analysis."

CHURCHILL EISENHART, "Sam Wilks as I Remember Him."

W. J. YOUDEN, "An Operations Research Yarn and Other Comments."

JOHN W. TUKEY, "The Future of Processes of Data Analysis."

M. G. KENDALL, "Statistics and Management."

11th: Dover, New Jersey, 20-22 October 1965.
Headquarters, U.S. Army Munitions Command.

JOAN R. ROSENBLATT, "Confidence Limits for the Reliability of Complex Systems."

J. STUART HUNTER, "Non-Linear Models: Estimation and Design."

WILLIAM C. GUENTHER, "Target Coverage Problems."

H. O. HARTLEY, "Maximum Likelihood Estimates for the General Mixed Analysis of a Variance Model."

12th: Washington, D.C., 19-21 October 1966.
Harry Diamond Laboratories and National Bureau of Standards.

BRIAN W. CONOLLY, "Operations Research."

JOHN MANDEL, "Statistics as a Diagnostic Tool in Data Analysis."

W. G. COCHRAN, "Planning and Analysis of Observational Studies."

NORMAN L. JOHNSON, "Sample Censoring."

13th: Fort Belvoir, Virginia, 1-3 November 1967.
U.S. Army Mobility Equipment Center and U.S. Army Engineer Topographic Laboratories.

FRANCIS J. ANSCOMBE, "Regression Analysis."

K. A. BROWNLEE, "Some Comments on Matching."

I. J. GOOD, "Some Statistical Methods in Machine Intelligence Research."

FRANK PROSCHAN, "Maximum Likelihood Estimation of Reliability."

M. B. WILK, "Data Analysis."

14th: Edgewood Arsenal, Maryland 23-25 October 1968.
U.S. Army Edgewood Arsenal.

LT. GEN. WILLIAM B. BUNKER, "Broadening the Horizons of Experimental Design."

ROLF E. BARGMANN, "Structure and Classification of Patterns."

ACHESON J. DUNCAN, "Bulk Sampling."

EMANUEL PARZEN, "Time Series."

15th: Redstone Arsenal, Alabama, 22-24 October 1969.
U.S. Army Missile Command.

JOHN E. CONDON, "Reliability and Quality Assurance."

NANCY R. MANN, "Systems Reliability."

CLIFFORD J. MALONEY, "A Probability Approach to Catastrophic Threat."

RICHARD G. KRUTCHKOFF, "The Empirical Bayes Approach to the Design and Analysis of Experiments."

S. C. SAUNDERS, "On Confidence Limits for the Performance of a System When Few Failures are Encountered."

16th: Fort Lee, Virginia, 21-23 October 1970.
U.S. Army Logistics Management Center.

SOLOMON KULLBACK, "Minimum Discrimination Information Estimation and Application."

RICHARD J. KAPLAN, "Field Testing."

GARY G. KOCH, "The Analysis of Complex Contingency Table Data from General Experimental Designs and Sample Surveys."

A. CLIFFORD COHEN, "Estimation in Truncated Poisson Distributions with Concomitant Intervals and Truncation Points."

DANA QUADE, "Nonparametric Analysis of Covariance."

17th: Washington, D.C., 27-29 October 1971.
Walter Reed Army Institute of Research, Walter Reed Army Medical Center.

MARVIN ZELEN, "The Role of Mathematical Sciences in Biomedical Research."

BERNARD G. GREENBERG, "Randomized Response: A New Survey Tool to Collect Data of a Personal Nature."

GEOFFREY H. BALL, "Classification and Clustering Techniques in Data Analysis."

K. S. BANERJEE, "Hotelling's Weighing Designs."

JOHN J. GART, "The Comparison of Proportions: A Review of Significance Tests, Confidence Intervals and Adjustments for Stratification."

18th: Aberdeen Proving Ground, Maryland, 25-27 October 1972.
U.S. Army Test and Evaluation Command.

JOHN TUKEY, "Exploratory Data Analysis."

G. S. WATSON, "Orientation Analysis."

J. STUART HUNTER, "Sequential Factorial Estimation."

G. E. P. BOX,"Forecasting and Control."

RAYMOND H. MYERS, "Dual Response Surface Analysis."

19th: Rock Island, Illinois, 24-26 October 1973.
Headquarters U.S. Army Armament Command and U.S. Army Management Engineer Training Agency.

JEROME CORNFIELD, "Bayesian Statistics."

SHANTI GUPTA, "Ranking Based on Multiple Criteria."

J. SETHURAMAN, "A Strong Justification for the Use of Rank Tests in the Case of Non-Normality."

H. L. GRAY, "Generalized Jackknife Techniques."

FRANK PROSCHAN, "Reliability Growth."

SAM C. SAUNDERS, "Accelerated Life Testing."

WILLIAM A. THOMPSON, JR., "Reliability of Multiple Component Systems."

Solomon Kullback
The Florida State University
Department of Statistics

ABSTRACT. Through the use of the principle of minimum discrimination information estimation, leading to exponential families or multiplicative models or log-linear models it has been shown, using illustrative examples exhibiting different aspects of contingency table analysis, that:

(1) Estimates of the cell entries under various hypotheses or models can be obtained;
(2) The adequacy or fit of the model, or the null hypothesis, can be tested;
(3) Main effect and interaction parameters can be estimated;
(4) The structure of the table can be studied in detail in terms of the various interrelationships among the classificatory variables;
(5) The procedures can be applied to test hypotheses about particular parameters and linear combinations of parameters that are of special interest;
(6) The procedures provide indication of outlier cells;
(7) Since the procedures and concepts are based on a general principle a unified treatment of multi-dimensional contingency tables is possible;
(8) The procedure provides estimates based on an observed or sample table, which satisfy certain external hypotheses as to underlying probability relations in the population table. These estimates also preserve the inherent properties of the observed data not affected by the hypothesis;
(9) In general, the m.d.i. estimate is best asymptotically normal;
(10) The minimum discrimination information test statistics are asymptotically distributed as chi-squared with appropriate degrees of freedom;
(11) Convergent iterative computer algorithms are available for the analyses.

CONTINGENCY TABLES. There are two ways in which Statistical data are collected. In one form, actual measurements are recorded for each individual in the sample; in the other, the individuals are classified as belonging to different categories. On many occasions classifications are used to reduce original data on direct measurements. A well known example is that of "frequency-distributions". Data collected in the form of measurements may later be grouped and presented as a frequency distribution. An important advantage of grouping is that it results in a considerable reduction of data. On the other hand, it is not usually possible to convert grouped or classified data back into the original form.

Data which results from experiments in the physical sciences and engineering are usually outcomes of controlled experiments, and expressible in quantitative terms. In many other fields however, the data are seldom results of controlled experiments. In addition, the observations usually can be expressed only in qualitative or categorical terms, a yes - no, alive - dead, agree - disagree, class A - class B - class C, etc. type of response.

A contingency table is a form of presentation of grouped data. In the simplest case, a group of N items may be classified into just two groups, according to, say, presence or absence of a certain characteristic. For a fixed (given) characteristic the different groups of classification are called categories. For example, a group of N individuals may be classified according to hair color (characteristic), the categories being black, brown, blonde and "other". The categories may be qualitative as above, or may be quantitative, as for example in the classification by weight in pounds consisting of five categories: 40-80, 80-120, 120-160, 160-200, 200-240. When there is only one characteristic according to which data are classified we get a one-way-table. If there are two ways of classification, say according to Rows and Columns, the Row-classification having r categories and the Column-classification having c categories, the table is called a two-way table or a r x c table. The latter notation gives the number of categories in each classification. Carrying this notation further, a r x c x d table will have three characteristics of classification, the first having r categories, the second having c and the third d.

For example, an individual may be classified by sex, by race, by profession, by smoking habit, by age, by incidence of coronary heart disease. If we take observations over a sample of many such individuals, the result will be a multidimensional contingency table with as many dimensions as there are classifications. Contingency tables are cross-classifications of vectors of discrete random variables showing the number of subjects belonging to distinct categories of each of several qualitative or categorical classifications. The number of counts of individuals in a cell of this table represents that portion of the sample having the specific attributes within each of the classifications. A problem of interest, for example, might be to determine the factors that are associated with the presence or absence of coronary heart disease.

Data from many fields are often presented in this manner, that is, in a cross-tabulated form. Statistical analyses of these types of data has had a long history, but were mainly concerned with the simple kind, the two-way table. Analyses of multidimensional contingency tables have been investigated intensively only during the last decade or so.

Conclusions drawn from contingency tables may be only exploratory in nature. One of the difficulties can be the availability of meaningful and reliable data. The first problem one faces in the analysis of cross-classified data is the decision on the number of classifications to be included and the categories within each classification. Typical among the problems in the analysis is how to segregate the effect on the response of some of the background variables, individually or jointly, from that of the others that are of particular interest. The data analytic attitude is empirical rather than theoretical. A more empirical attitude is natural when detailed theoretical understanding is unavailable.

Estimation of parameters in models should be considered less as attempts to discover underlying truths and more as data calibrating devices which make it easier to conceive of noisy data in terms of smooth distributions and relations. With a given data set, a variety of models may be tried on, and one selected on the ground of looks and fit.

In the analysis of contingency tables we are usually interested in the relationship between one classification and one or more of the other classifications. As an example, consider a three-way r x c x d contingency table in which the row-classification represents the response of an experiment on animals, the column classification types of treatment and the depth classification sex. The following hypotheses may be of interest.

1. Response is independent of treatment irrespective of sex.
2. Response is independent of the different combinations of treatment and sex (as against the possibility that a particular treatment is more "effective" in terms of the response, for a particular sex).
3. Given sex, response is independent of treatment.

Of course, not all contingency tables can be interpreted in such a straightforward manner. In some instances, all three classifications can be considered as responses; then we may be interested in the independence or association among these responses. In other cases, a classification may be controlled, experimentally or naturally, like three specified levels of fertilizer applied or sex, and then the classification is termed a factor. For convenience, we shall group all the concepts of association, dependence, etc. under the general term of <u>interaction</u>. No interaction between treatment and sex appears to be a more acceptable phrase than independence between treatment and sex, since the term independence is usually reserved to express the relationship between random variables. We may also say that the interaction between response and treatment does not interact with sex, meaning the degree of association between response and treatment is the same for both sexes. The concept gives rise to the idea of second-order interaction. There are a number of different approaches to the mathematical formulation and interpretation of the concept of "no interaction". One such approach, through the concept of "generalized independence" is powerful and general enough to include all hypotheses of "no interaction" (formulated in a specific manner) <u>and</u> many other hypotheses about homogeneity, symmetry, etc. that we come across in analyzing contingency tables.

Consider, for example, an experiment to compare the effectiveness of safety release devices for regrigerators in relation to children's safety. Children between two to five years of age are induced to crawl into refrigerators equipped with six different types of release devices. If a child can open the door of the refrigerator, from inside, within a certain time period, the response is classified as a success, otherwise a failure. The background variables studied included age, sex, weight, socio-economic status of parents. The experimental variable was one of six devices. (A partial analysis of this data may be found in page 581 of Kullback, S., Kupperman, M., and Ku, H.H.(1962), Tests for contingency tables and Markov chains, <u>Technometrics</u>, 4, 573-608). Some balancing of the background variables was achieved.

In other instances none of the factors are subject to experimental control, and whatever available data could be collected is reported. The analysis of this type of data, though it may only be seeking preliminary information can be important in fields of health and safety. The uncontrolled experimental data are sometimes the only realistic data available when these data deal with life, death, health, and safety, and some of these factors and responses are only expressible in qualitative terms, in the present state of art.

It is expected that the number of problems calling for the techniques of the analysis of multidimensional contingency tables will increase. Experience at the George Washington University with such a growing demand confirms this. The examination and interpretation of data from social phenomena, housing, psychology, education, environmental problems, health, safety, manpower, business, experimental testing of devices, military research and development, etc., are potential source areas.

Classical problems in the historical development of the analysis of contingency tables concerned themselves primarily with such questions as the independence or conditional independence of the classificatory variables, or homogeneity or conditional homogeneity of the classificatory variables over time or space, for example, similar to such tests in multivariate analysis as independance, multiple correlation, partial correlation, canonical correlation, etc. Such classical problems turn out to be special cases of the technique we discuss.

These techniques result in analyses which are essentially regression type analyses. As such they enable us to determine the relationship of one or more "dependent" qualitative or categorical variables of interest on a set of "independent" classificatory variables, as well as the relative effects of changes in the "independent" variables on the "dependent variables". The object of the analyses is the study of the interaction between and among the classifications. The term interaction is used here in a general sense to cover both dependence and association.

Critics of methods for contingency table analysis have maintained that most of the procedures used, at least in the past, were only of a global chi-squared test nature. However, for a recent example of this see Patil, K.D. (1974), Interaction test for three-dimensional contingency tables, Journal Am. Statist. Assn, 69 164-168. Through the use of the principle of minimum discrimination information (m.d.i.) estimation, leading to exponential families or multiplicative models (generalized independence) or log-linear models we show that:

(1) Estimates of the cell entries under various hypotheses or models can be obtained;
(2) The adequacy or fit of the model, or the null hypothesis, can be tested;
(3) Main effect and interaction parameters can be estimated;
(4) The structure of the table can be studied in detail in terms of the various interrelationships among the classificatory variables;
(5) The procedures can be applied to test hypotheses about particular parameters and linear combinations of parameters that are of special interest;

(6) The procedures provide indication of outlier cells. These may cause a model not to fit overall, yet fit the other cells excluding the outliers;

(7) Since the procedures and concepts are based on a general principle a unified treatment of multidimensional contingency tables is possible. Sequences of generalizations step by step to higher order dimensional contingency tables are not necessary as has been the case with other ad hoc procedures (see for example, Patil (1974), Sugiura, N and Otake, H. (1974), An extension of Mantel-Haenszel procedure to k 2 x c contingency tables and the relation to the logit model, Communications in Statistics, to appear);

(8) The procedure provides estimates based on an observed or sample table, which satisfy certain external hypotheses as to underlying probability relations in the population table. These estimates also preserve the inherent properties of the observed data not affected by the hypothesis;

(9) In general, the m.d.i. estimates are best asymptotically normal (BAN) and in the many applications of fitting models to a table based on observed sets of marginal values or linear restraints of observed values, the m.d.i. estimates in particular are maximum-likelihood estimates;

(10) The test statistics are minimum discrimination information (m.d.i.) statistics which are asymptotically distributed as chi-squared with appropriate degrees of freedom. In the case of fitting models to a table based on observed sets of marginal values or linear restraints of observed values, the m.d.i. statistics are log-likelihood ratio statistics. The m.d.i. statistics are additive, as are the associated degrees of freedom, so that the total under an hypothesis can be analyzed into components each under sub-hypotheses. The analysis is analogous to analysis of variance and regression analysis techniques. It uses a design matrix, a set of regression parameters, and explanatory variables, and analysis of information tables.

(11) In models fitting estimates to an observed table based on sets of observed marginal values as explanatory variables, some estimates can be expressed explicitly as products of marginal values. However, this is not generally true, and expected cell frequencies (functions of marginal values), can be computed by an iterative proportional fitting procedure, and the use of a computer to perform the iterations becomes necessary. For the foregoing cases which we term internal, and problems involving tests of external hypotheses on underlying populations a number of iterative computer programs are available. They provide as output, design matrices, the observed cell entries and the cell estimates as well as their logarithms, parameter estimates, outlier values, m.d.i. statistics and their corresponding significance levels, and covariance matrices of parameter estimates, to assist in and simplify the numerical aspects of the inference. In this respect it is of interest to cite the following quotation from a book review by D.J. Finney in Journal Royal Statistical Society, Series A(General) Vol. 136(1973), part 3, p. 461, "No mention is made of the extent to which computers have destroyed the need to assess statistical methods in terms of arithmetical simplicity: indeed the emphasis on avoiding lengthy, but easily programmed, iterative calculations is remarkable".

MULTI-DIMENSIONAL, NON-GAUSSIAN, RANDOM PROCESSES WITH SPECIFIED COVARIANCE AND PROBABILITY DENSITY FUNCTIONS

James W. Wright
Advanced Sensors Directorate
US Army Missile Research, Development
and Engineering Laboratory
US Army Missile Command
Redstone Arsenal, Alabama

ABSTRACT. The simulation of radar scattering signatures, including radar cross section and glint, of complex targets for use in air defense system simulations is a difficult and time consuming task. Although it is possible to develop deterministic models of the radar signature, as a function of the target aspect angles, it is generally not possible to use these models in a realtime simulation because of the computational requirements involved in using such a model. Statistical and stochastical models of radar signatures are generally limited to the classical radar cross section models, although some models do include crudely correlated glint models as well. It is possible to describe statistically the radar scattering signature in terms of the probability density and covariance functions, but the processes generally are non-stationary, non-Gaussian, non-Markovian processes. Even reduction of the process to a stationary, Markovian, non-Gaussian process does not presently reduce the problem to an analytically solvable problem. The development of techniques to generate these multidimensional random processes is needed to make more realistic simulations practical.

I. INTRODUCTION. The simulation of realistic radar signatures of aircraft for use in air defense (AD) system simulations is a difficult task. Stochastic models capable of representing the multidimensional radar signature with any realism do not exist, so complex deterministic models are used when realism is required. These deterministic models require significant computer resources in terms of both computation time and memory, but they can represent the nonstationary multidimensional signature with sufficient accuracy to make them invaluable in all digital simulations. The computation time requirements generally exclude deterministic models from realtime hybrid simulations, i.e., those simulations with actual system hardware in the loop. It is for these realtime, hybrid simulations that realistic stochastic models are needed.

The purpose of this paper is to present the problem with a description of the process and the underlying phenomena from which it is derived.

II. DESCRIPTION OF THE PROBLEM

1. General. The first step is to establish the definition and description of the radar signature. The elements, or parameters, of the signature can be addressed one at a time or in pairs to arrive at a statistical description of the overall process.

The radar signature is defined to be the set of target induced parameters which are measured by the observing radar(s) or which directly influence the radar measurements or tracking systems. These include the radar cross section (RCS), azimuth glint (e_θ), elevation glint (e_φ), and intrinsic phase (α). Other parameters, such as range glint, which are dependent on system mechanization will not be considered. Each of these four parameters is range independent for far field conditions, which is assumed for simplicity. Since most of the Army AD systems are semi-active systems, each of these parameters must be considered twice: once for the ground radar, and once for the missile seeker. The signature, therefore, is eight dimensional.

The problem is further complicated by the fact that these parameters are functions of the target aspect angles which are nonstationary functions of time. Figure 1 depicts the plan view of an arbitrarily selected flight path and Figure 2 depicts the aspect angles of a perfectly controlled target in still air as seen from a ground radar located at the origin. (The + sysbols are taken at equal time intervals.) A realistic target will experience random rotational (and resulting translational) perturbations from wind gusts and autopilot noise and response characteristics, so that the actual aspect angles will be a two-dimensional random process with averages approximately as shown in Figure 2. The nature of the random perturbations is a function of the assumed environment and the target aerodynamic and control response characteristics.

The problem, then, is to develop a stochastic model of a nonstationary eight-dimensional random process. The correlations among the various parameters must be maintained because they significantly affect the response of the AD system. It is clear that some simplifications to the problem are required before any attack can be made on the problem. An investigation of the various parts of the problem will reveal some significant simplifications.

2. Target Signature Characteristics. The most realistic deterministic radar signature models are based on the N-body approach. This approach assumes that the signature can be assumed to be generated by N scatterers or scattering centers. The components of the radar signature can be given by [1]:

$$RCS = \sum_{i=1}^{N} \sum_{j=1}^{N} \sqrt{S_i S_j} \cos(\alpha_i - \alpha_j) \quad ,$$

$$e_\theta = \sum_{i=1}^{N} \sum_{j=1}^{N} \sqrt{S_i S_j} \; f_{1i} \cos (\alpha_i - \alpha_j) \quad ,$$

$$e_\varphi = \sum_{i=1}^{N} \sum_{j=1}^{N} \sqrt{S_i S_j} \; f_{2i} \cos (\alpha_i - \alpha_j) \quad ,$$

and

$$\alpha = \arctan \left(\frac{\sum_{i=1}^{N} \sqrt{S_i} \sin \alpha_i}{\sum_{j=1}^{N} \sqrt{S_j} \cos \alpha_j} \right) \quad ,$$

where

$$f_{1i} = -x_i \cos\theta \cos\varphi - y_i \cos\theta \sin\varphi - z_i \sin\theta \quad ,$$

$$f_{2i} = -x_i \sin\varphi + y_i \cos\varphi \quad ,$$

(x_i, y_i, z_i) are the coordinates of the i-th scatterer,

S_i is the RCS of the i-th scatterer,

and

α_i is the phase angle associated with the i-th scatterer.

The RCS is given in square meters, and the e_θ and e_φ are given as errors in the apparent target position, in meters, in a plane orthogonal to the line of sight at the target. It has been demonstrated that the glint errors, e_θ and e_φ, can be expressed as the gradients of the phase with respect to the appropriate angles [2,3]. Conversely, it is possible to compute the random component of the intrinsic phase as the integral of the phase gradient.

Figures 3, 4, and 5 depict the RCS, RCS* e_θ (theta component of non-radial power) and RCS* e_φ (phi component of non-radial power) for a selected region of aspect angles for a simple mathematical model of the MQM-34D (BQM-34A) target drone [1] for a frequency of 1 GHz. The non-linear nature of these functions is clear, but the correlation is not.

The correlation of interest is the statistical correlation. If it is assumed that the aspect angles have some statistical relationships, it is possible to compute the statistical relationships of the scattering components.

The nonstationary nature of the process can be handled in a relatively straightforward manner. Since the average aspect angles generally change slowly with respect to the random components, it is reasonable to consider the random component separately. It is assumed that the random components of the aspect angles are stationary Markov processes. This is not strictly true but the lack of definitive measurements on aircraft motion and the realism of the radar signature models make the assumption acceptable. Since the statistical characteristics of the signature components are functions of the aspect angles, the statistical characteristics of the signature are also assumed to be stationary processes when considered from the short term viewpoint.

3. Statistical Characteristics of the Radar Signature. The statistical characteristics of the radar signature of a target are functions of many variables, two especially important ones being radar frequency and aspect angle statistics. Figures 3, 4, and 5 were computed for 1 GHz. The lobing structure increases approximately linearly with frequency, but the averages and variances do not. A Monte Carlo type simulation of the target at a point corresponding to $\theta = 103.8$ degrees, $\varphi = 38.67$ degrees with θ and φ jointly normal and $\sigma_\theta = 3.07$ degrees $\sigma_\varphi = 1.46$ degrees, and $\rho_{\theta\varphi} = 0.543$ resulted in the results shown in Figures 6 through 11. These figures present plots of the radar signature parameters or related parameters. These plots indicate the type of random processes that are to be modeled. Figures 12 through 16 present probability density functions of the data in Figures 6 through 11 and Figures 17 through 20 present typical covariance functions for part of these data.

Three classical statistical RCS models are of interest. They are the Swerling 1, Swerling 3, and log-normal. The equations for these models are given by:

Swerling 1

$$f_s(s) = \frac{1}{\bar{s}} \exp\left(-\frac{s}{\bar{s}}\right), \quad s \geq 0$$

where

$$\bar{s} = E\{s\} \quad ,$$

Swerling 3

$$f_s(s) = \frac{4s}{\bar{s}^2} \exp\left(-\frac{2^s}{\bar{s}}\right), \quad s \geq 0 \quad ,$$

and log-normal

$$f_s(s) = \frac{1}{\sigma s \sqrt{2\pi}} \exp\left\{ - \frac{(\ell n\ s - \mu)^2}{2\sigma^2} \right\} \quad , \quad s \geq 0$$

where

$$\mu = E\ \{\ell n\ s\}$$

and

$$\sigma^2 = E\ \{(\ell n\ s - \mu)^2\} \quad .$$

Models of the nonradial components of power have not been developed at this time.

The problem of bistatic angles must also be addressed. It can be shown that the bistatic signature is best approximated by the monostatic signature for the aspect angles corresponding to those of the bisector of the ground radar aspect angles and the missile seeker aspect angles reduced by a scale factor which is a function of the bistatic angle. The covariance of the RCS as a function of one half of the bistatic angle is approximately the same shape as the covariance of the RCS as a function of the aspect angle.

The roll-off of RCS as a function of bistatic angle for the MQM-34D drone near nose-on appears to be exponential in shape with a reduction of approximately 4.5 decibels (multiplicative factor of approximately 3) in 30 degrees in average and standard deviation.

A review of the known data is appropriate at this point to determine what data are missing. The first order probability density functions and covariance functions can be assumed to be known for the monostatic RCS and nonradial components of power and for the bistatic RCS and nonradial components of power. The covariance function of the monostatic and bistatic RCS is also known. The covariance functions of the monostatic and bistatic nonradial components of power are not known, but theory and experiments indicate that for the frequencies of interest and bistatic angles exceeding 5 degrees, these covariance functions can be assumed to be zero. The intrinsic phase is most readily computed as the integral of the glint (phase gradient) since absolute phase is not important. Thus the stochastic process is actually reduced basically to a pair of three-dimensional random processes, with a correlation between the monostatic and bistatic RCS for small bistatic angles.

The first order probability density functions for the RCS will generally be selected from one of the three classical models previously given. The probability density functions for the nonradial components of power remain to be determined, at least in analytical form. The covariance functions appear to be approximately exponential for the cases studied to date. Analysis of other deterministic models, the MQM-34D and other targets, may indicate other shapes for the covariance functions, however.

The nonstationary aspect of the problem can probably be handled by using averages and variances which are functions of the aspect angles and, hence, of time.

That leaves one major problem area, the generation of a three-dimensional random process with non-normal probability density functions and specified covariance functions. Techniques are available for handling limited classes of one-dimensional random processes [4, 5, 6], but it appears that generalization to multidimensional processes, except for very special cases [5], has not been accomplished [7].

III. CONCLUSION. This paper has outlined one area where the multi-dimensional random processes are needed today. No attempt has been made to present all of the data necessary to completely define the problem. In fact, the author is not sure what data are needed to completely define the nature of the stochastic process that is to be modeled. The data presented are generally accepted as necessary but are probably not sufficient to permit complete characterization.

IV. REFERENCES

1. Wright, J. W. and Haddad, A. H., On the Statistical Modeling of Radar Targets, US Army Missile Command, Redstone Arsenal, Alabama, Report No. RE-72-19, November 1972.

2. Howard, D. H., "Radar Target Angular Scintillation in Tracking and Guidance Systems Based on Echo Signal Phase-Front Distortions," Proceeding of NEC, Vol 15, 1959.

3. Lindsay, J. E., "Angular Glint and the Moving, Rotating, Complex, Radar Target," IEEE Transactions on Aerospace and Electronic Systems, AES-4, No. 2, March 1968.

4. Haddad, A. H. and Valisalo, P. E., "Generation of Random Time Series Through Hybrid Computation," Sixth International Hybrid Computation Meetings.

5. Haddad, A. H., "Dynamical Representation of Markov Processes of the Separable Class," IEEE Transactions on Information Theory, IT-16, No. 5, September 1970.

6. Broste, N. A., Digital Generation of Random Sequences with Specified Autocorrelation and Probability Density Functions, US Army Missile Command, Redstone Arsenal, Alabama, Report No. RE-TR-70-5, March 1970.

7. Wong, E., "Recent Progress in Stochastic Processes - A Survey," IEEE Transactions on Information Theory, IT-19, No. 3, May 1973.

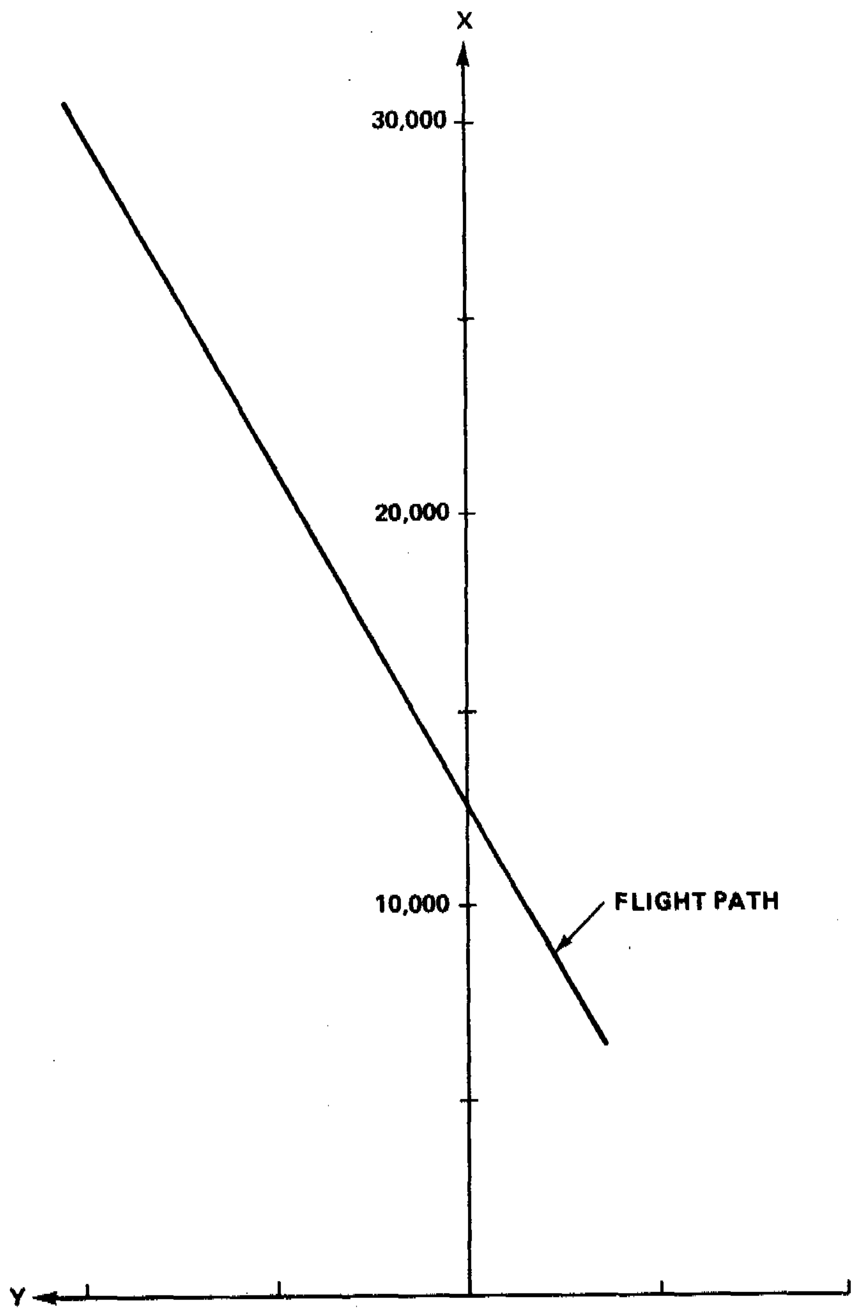


Figure 1. Plan view of target flight path.

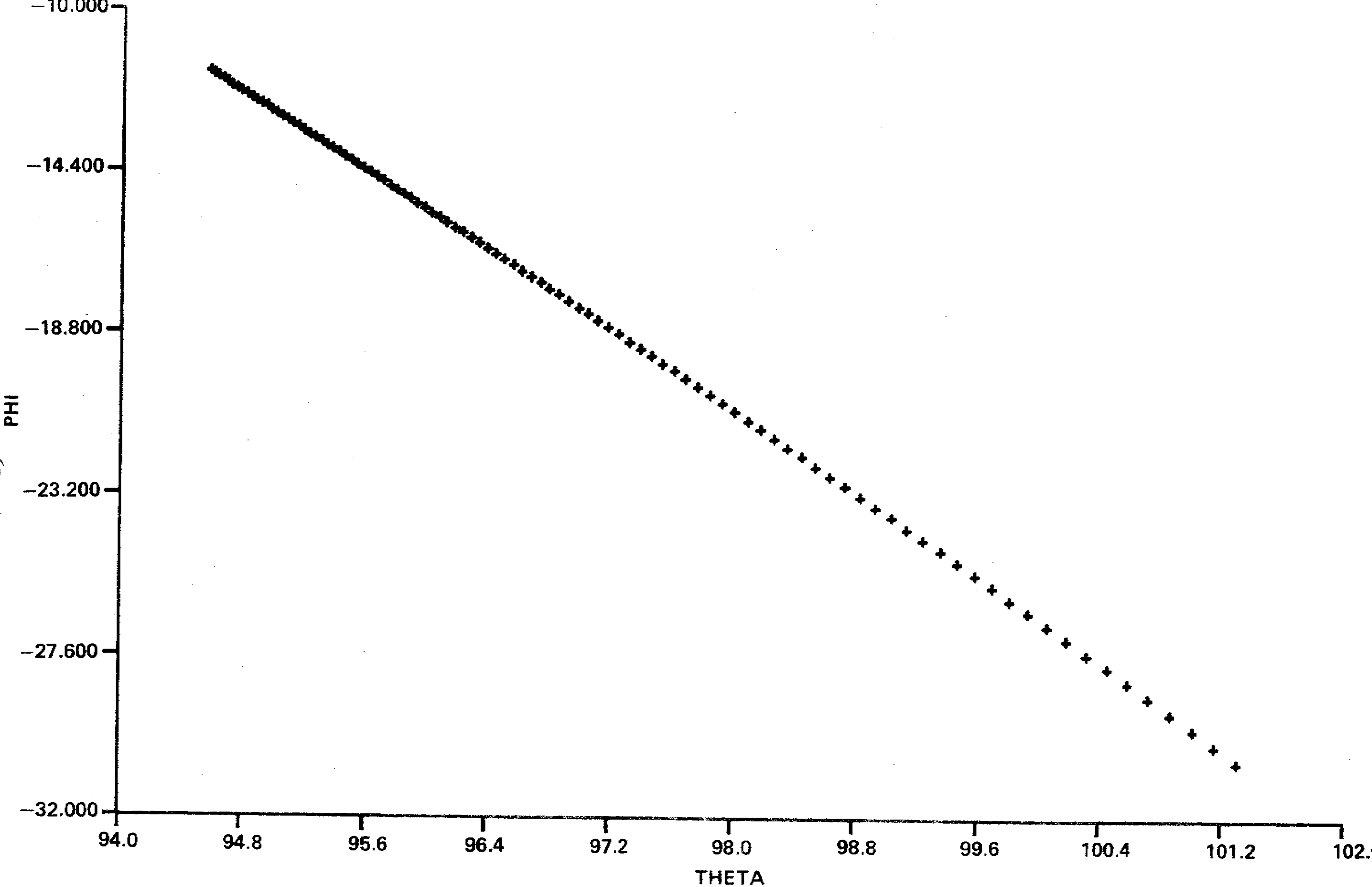


Figure 2. Target aspect angles to ground radar.

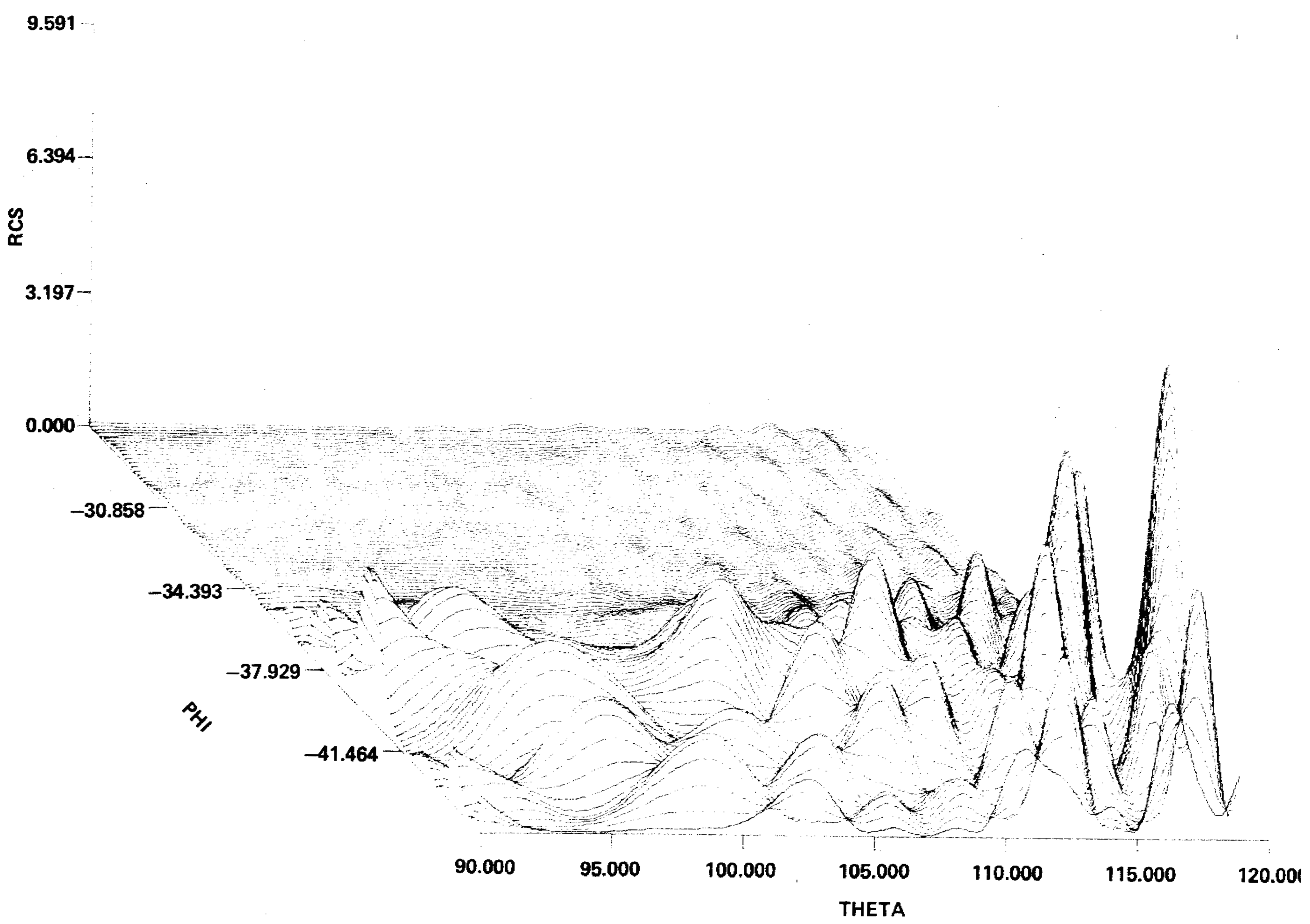


Figure 3. Radar cross section as a function of the aspect angles.

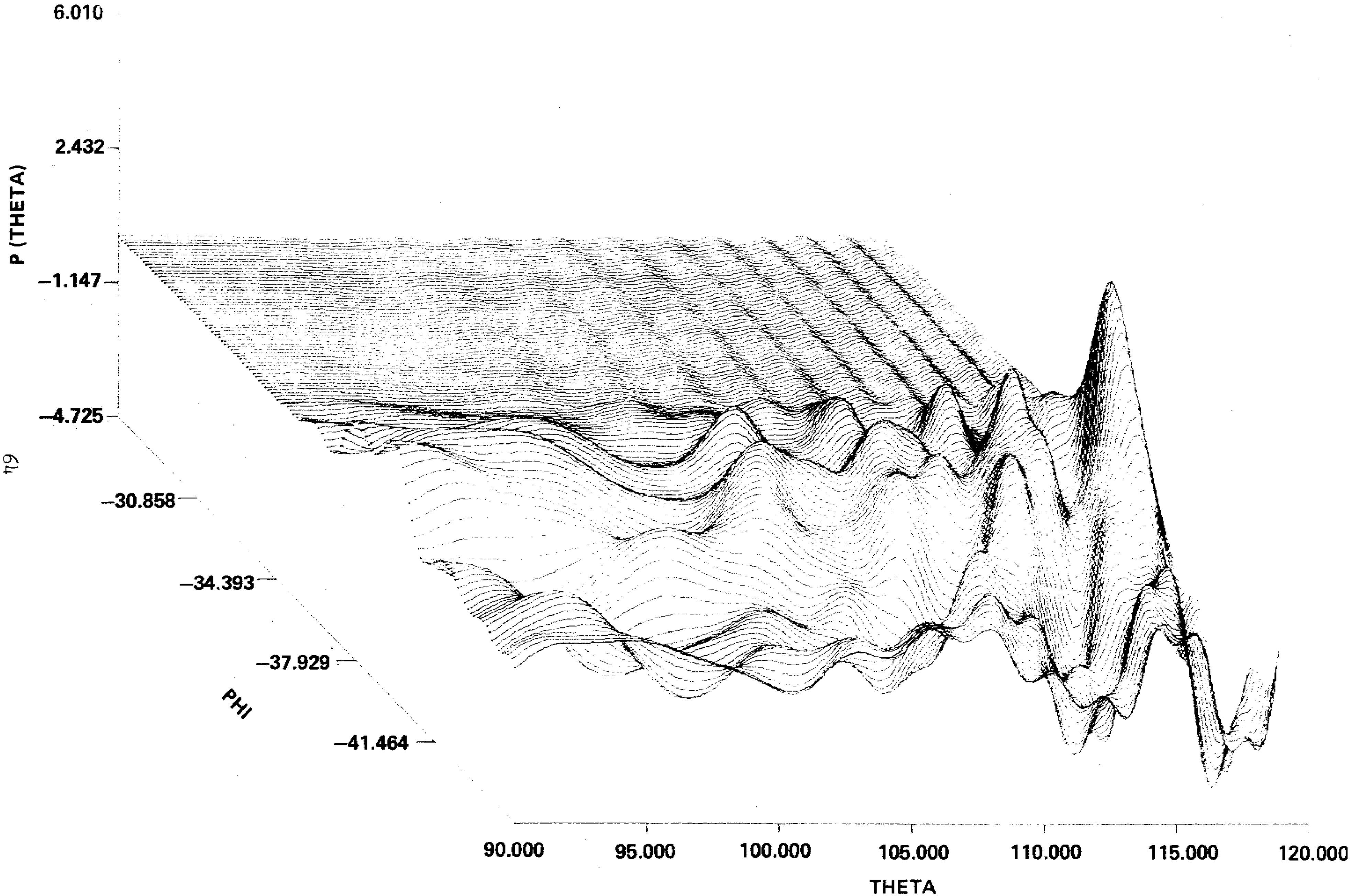


Figure 4. Theta component of non-radial power as a function of the aspect angles.

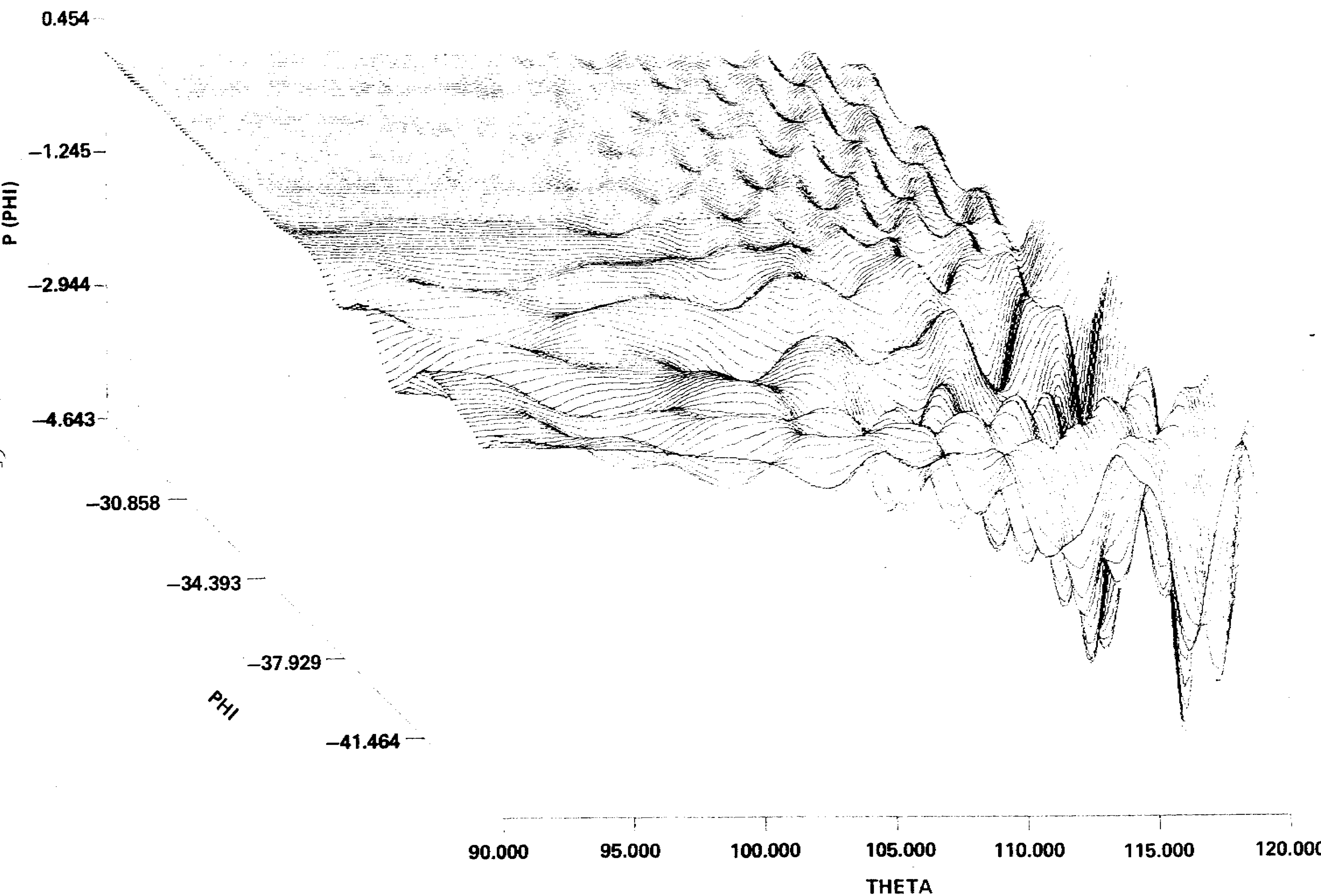


Figure 5. Phi component of non-radial power as a function

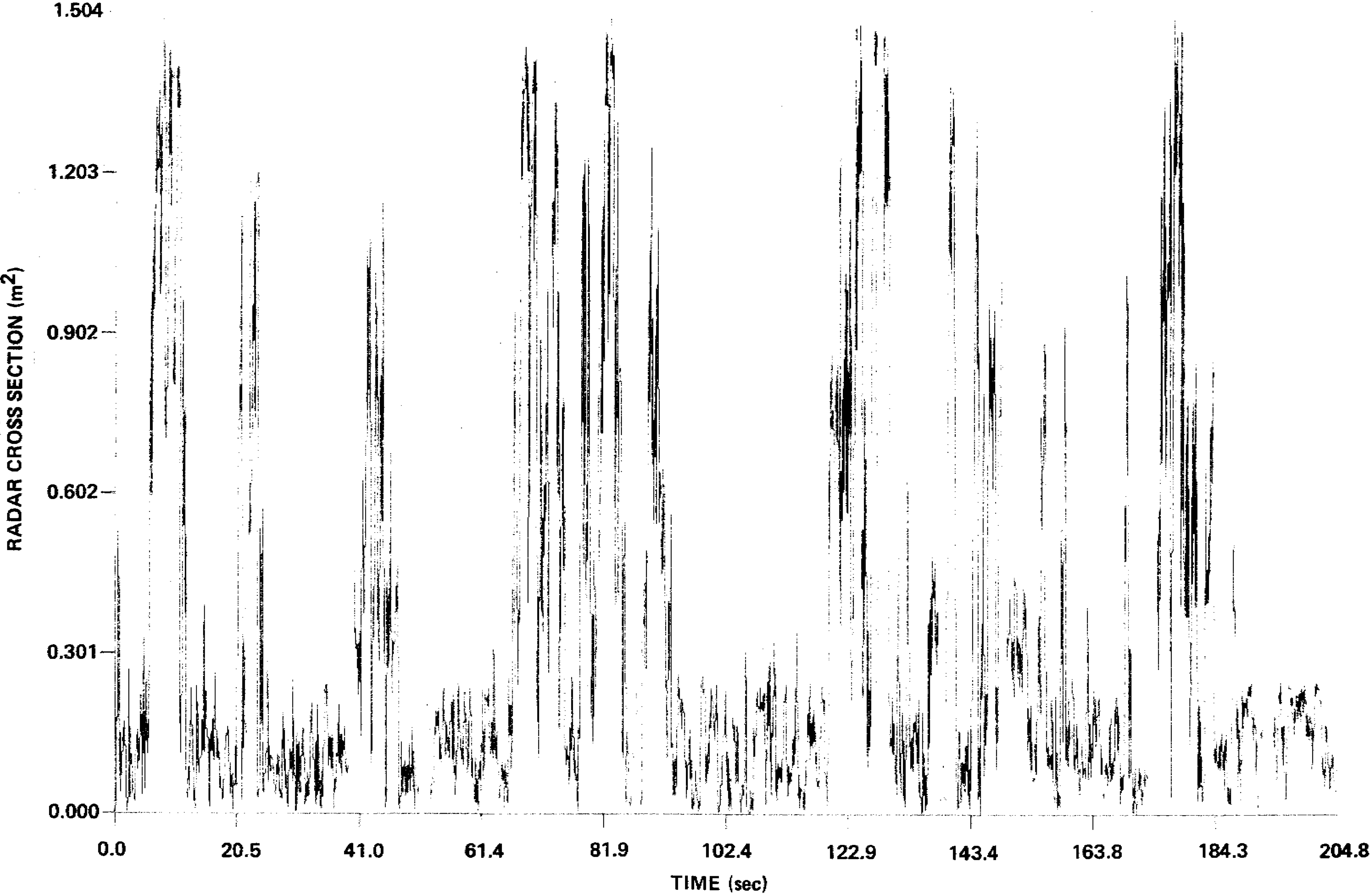


Figure 6. Simulated radar cross section history.

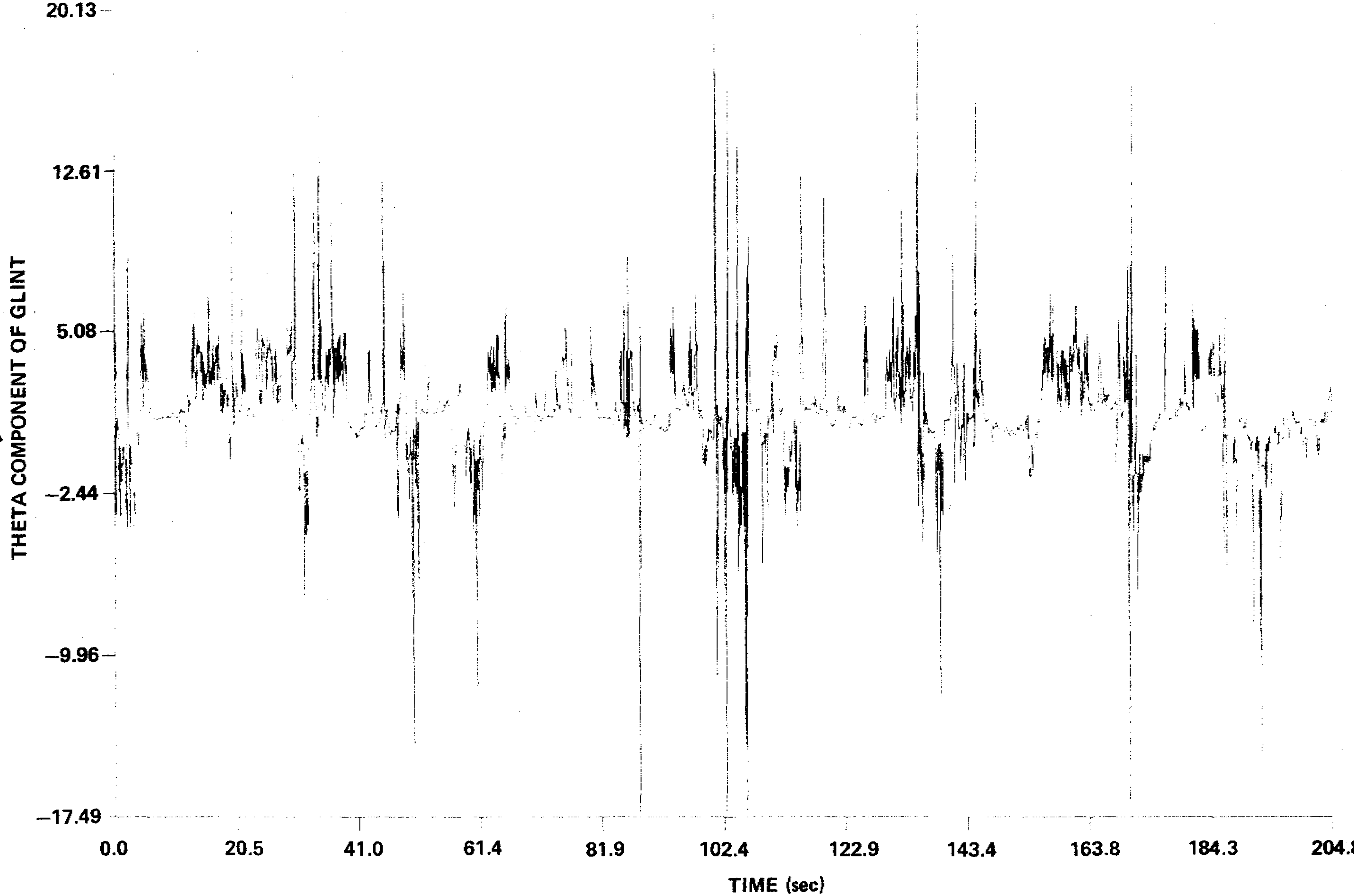


Figure 7. Simulated theta component of glint history.

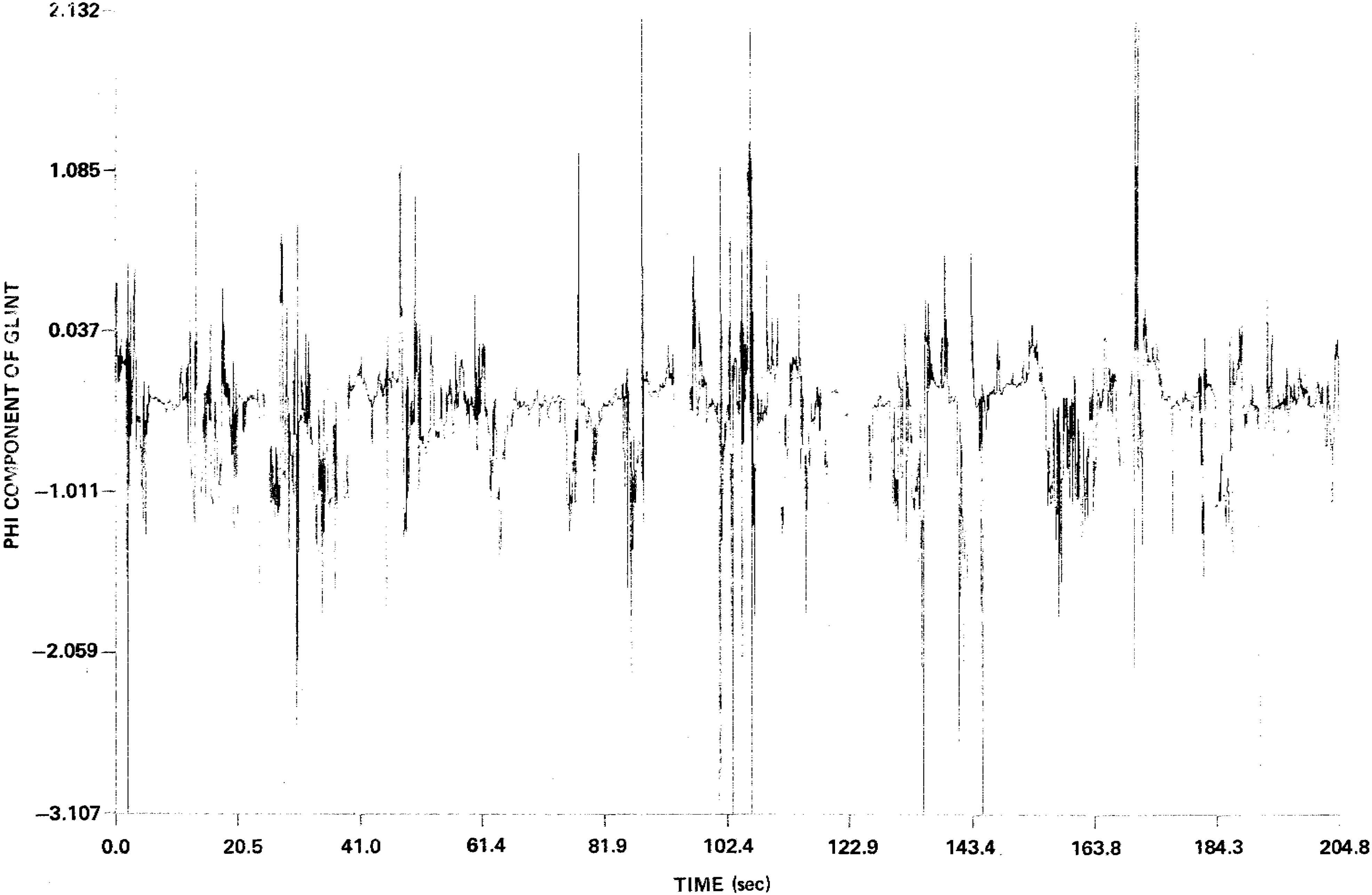


Figure 8. Simulated phi component of glint history.

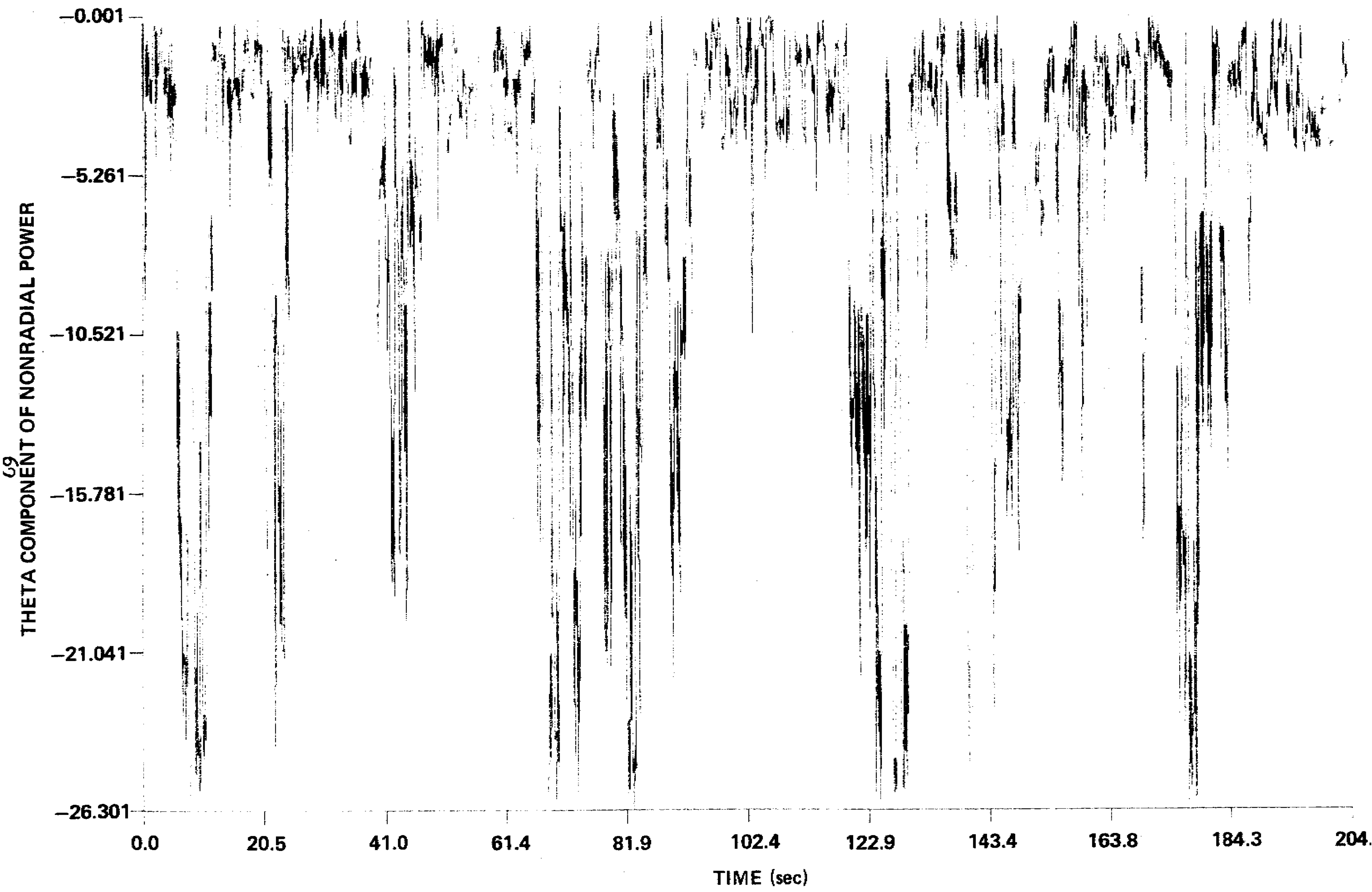


Figure 9. Simulated theta component of non-radial power history.

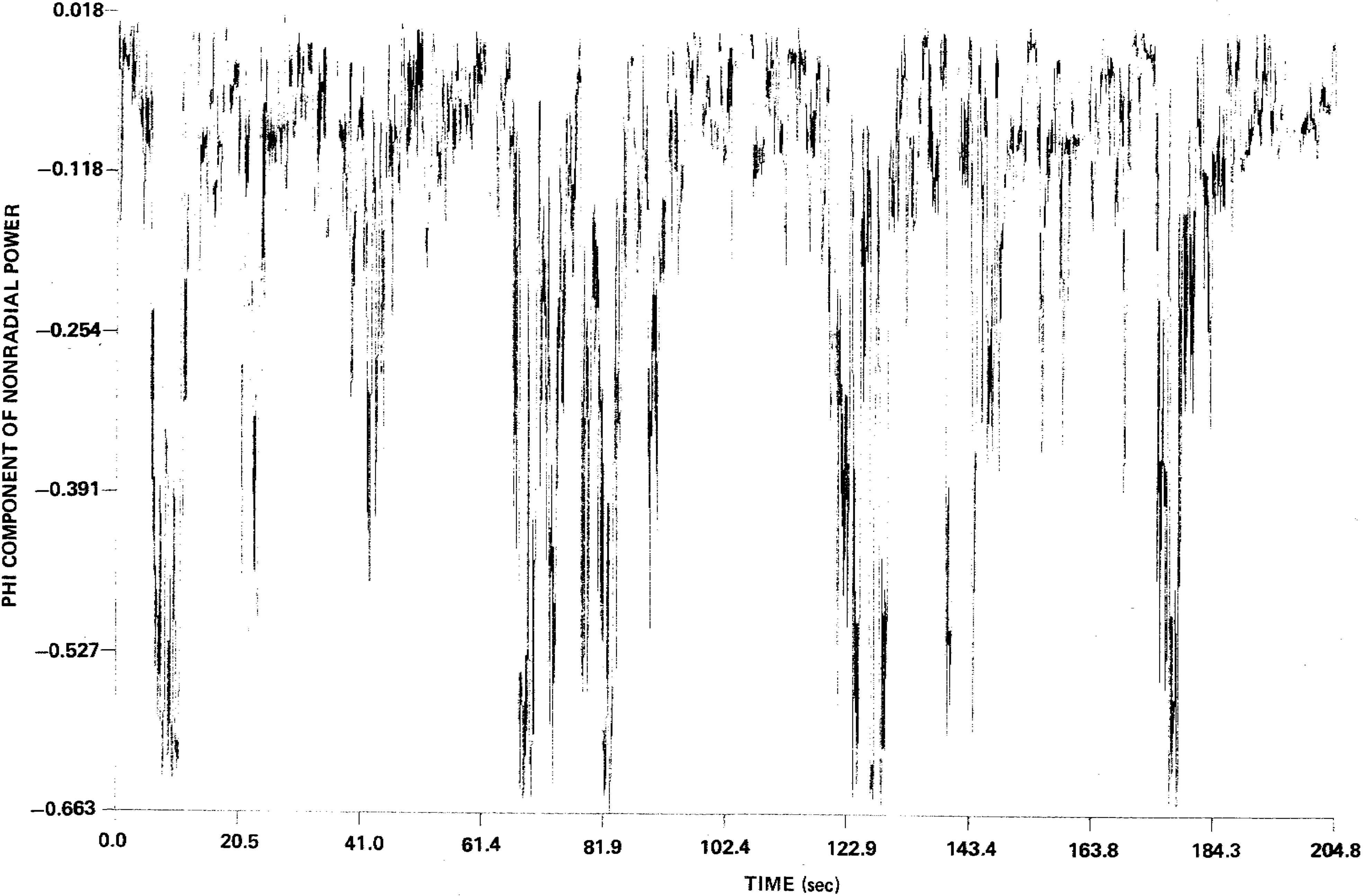


Figure 10. Simulated phi component of non-radial power history.

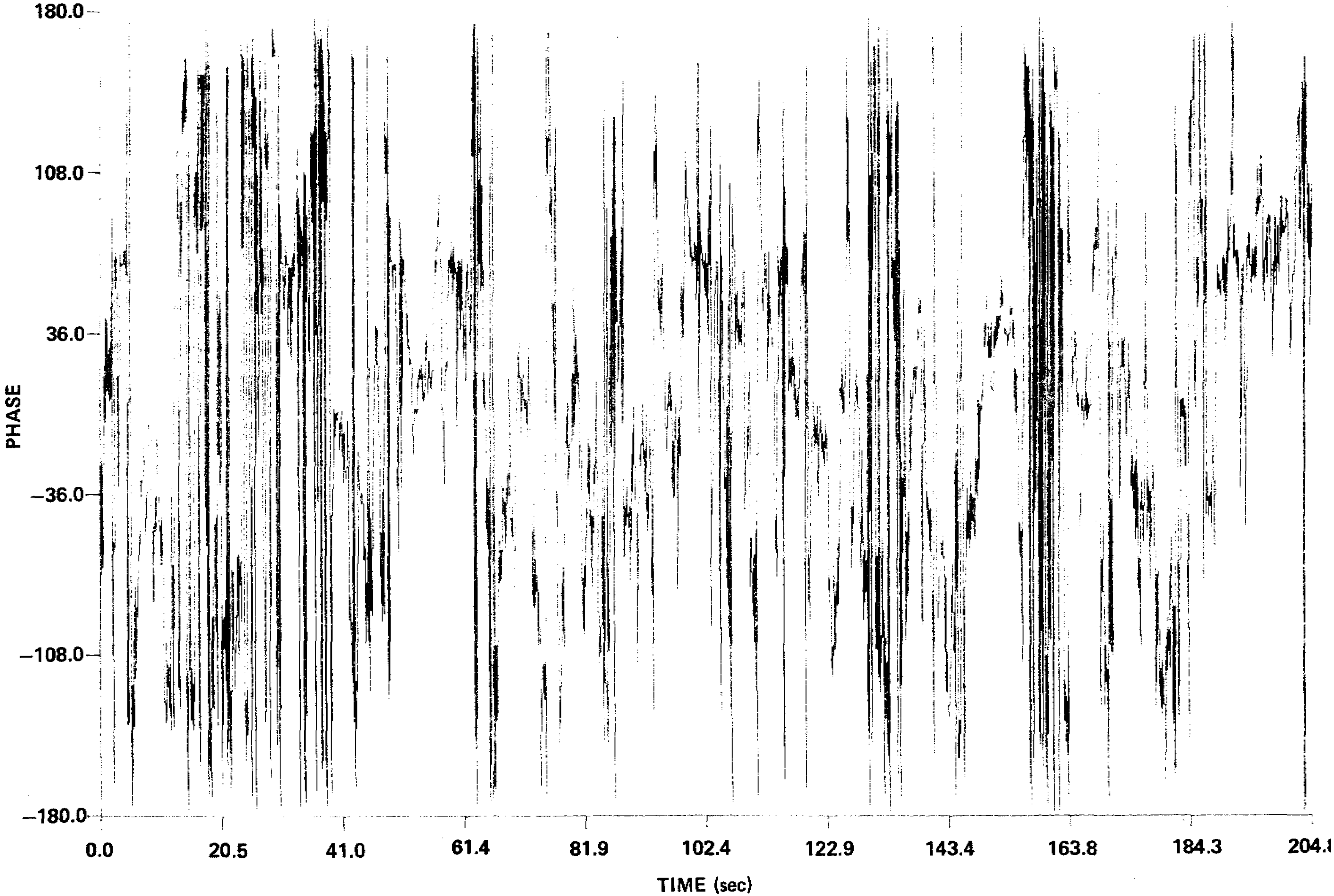


Figure 11. Simulated phase history.

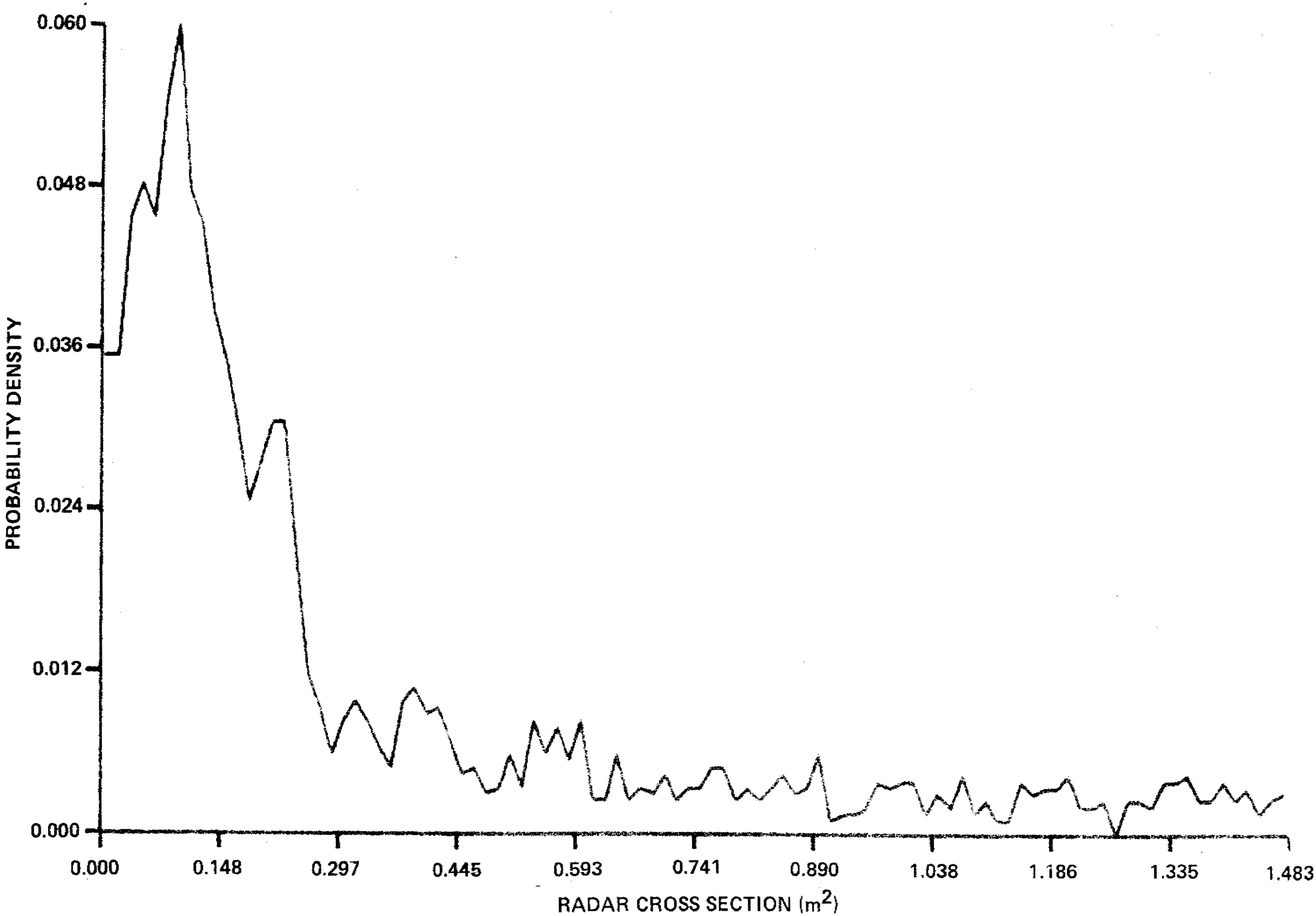


Figure 12. Probability density of radar cross section.

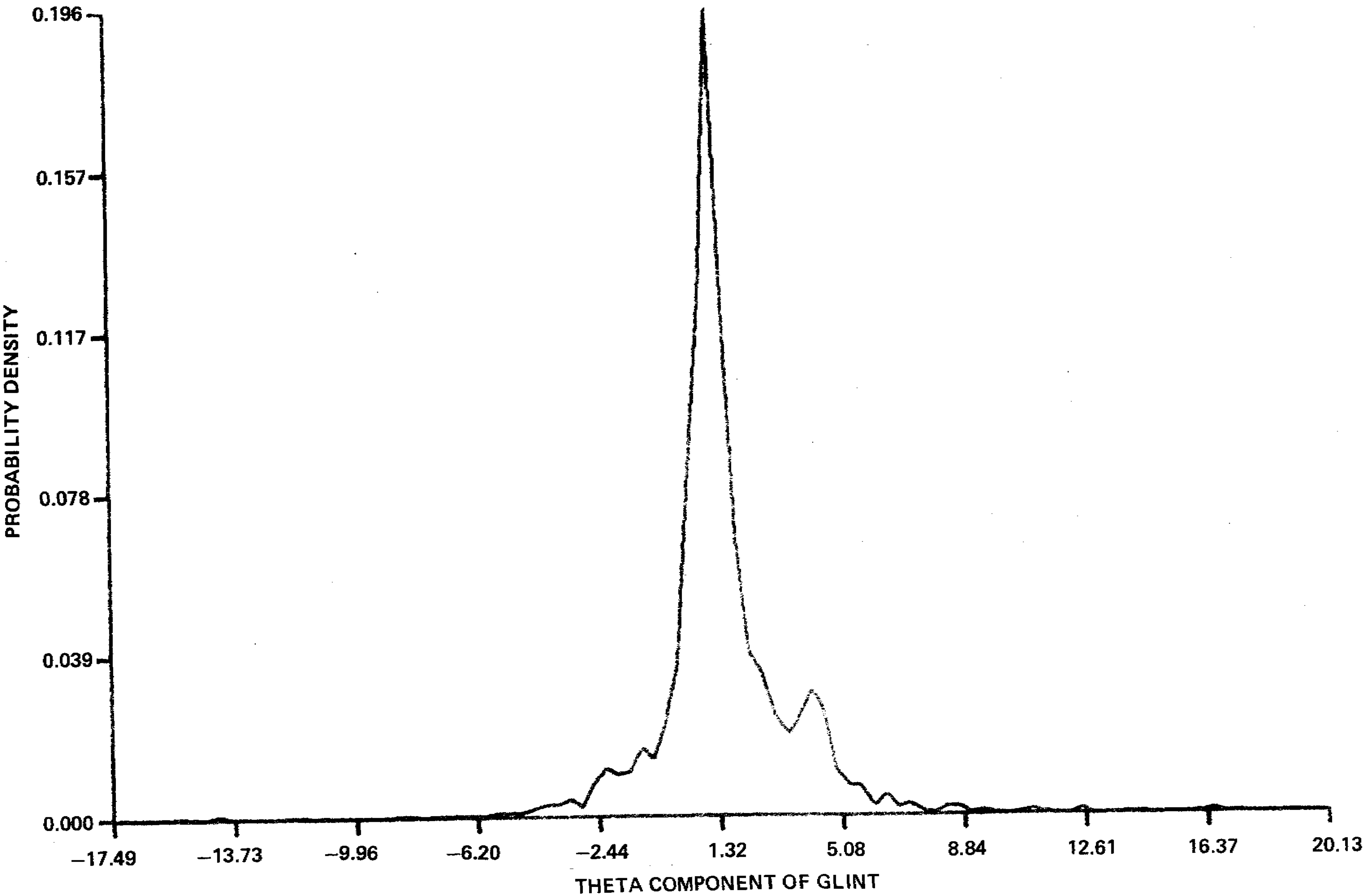


Figure 13. Probability density of theta component of glint.

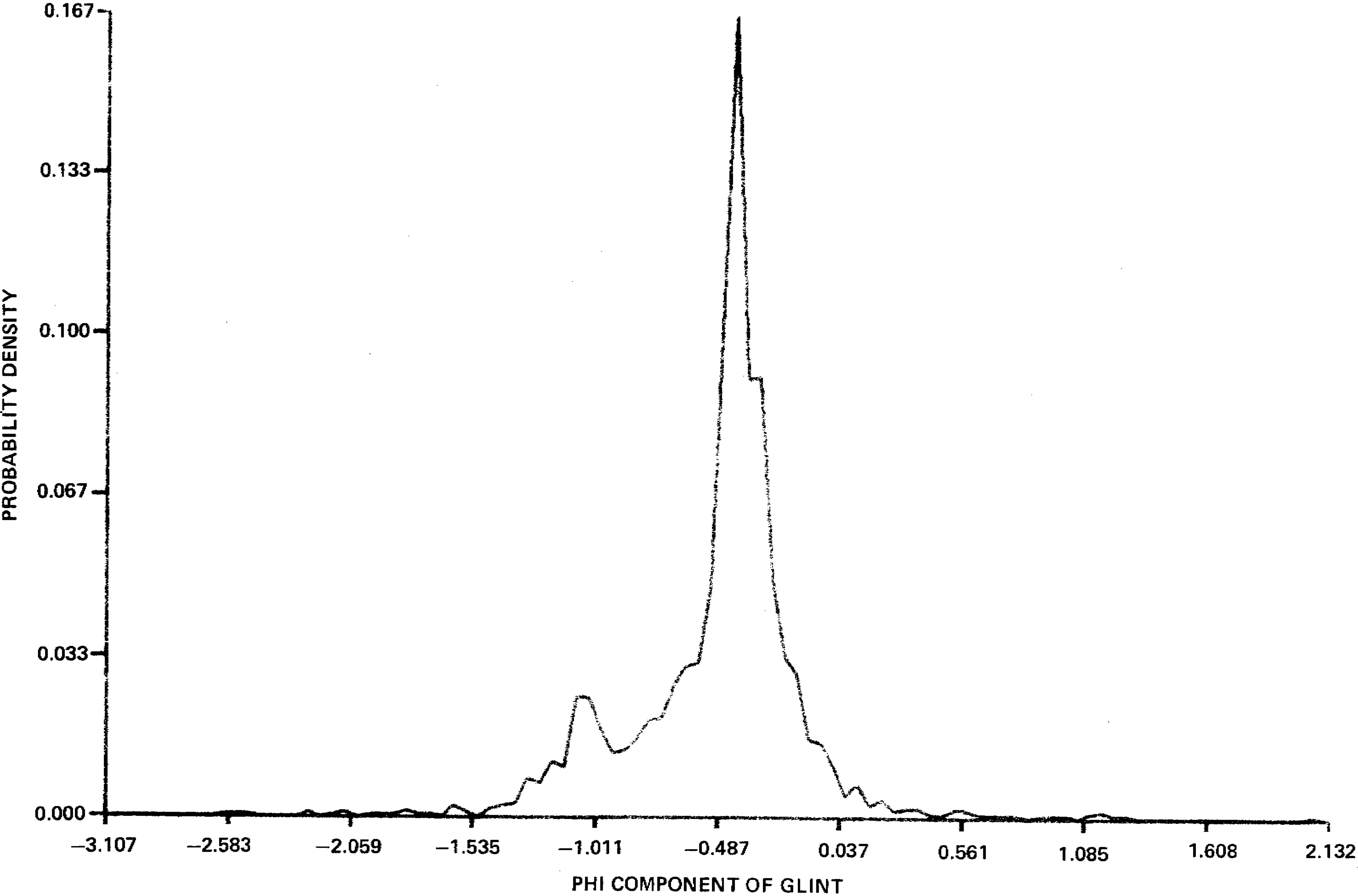


Figure 14. Probability density function of phi component of glint.

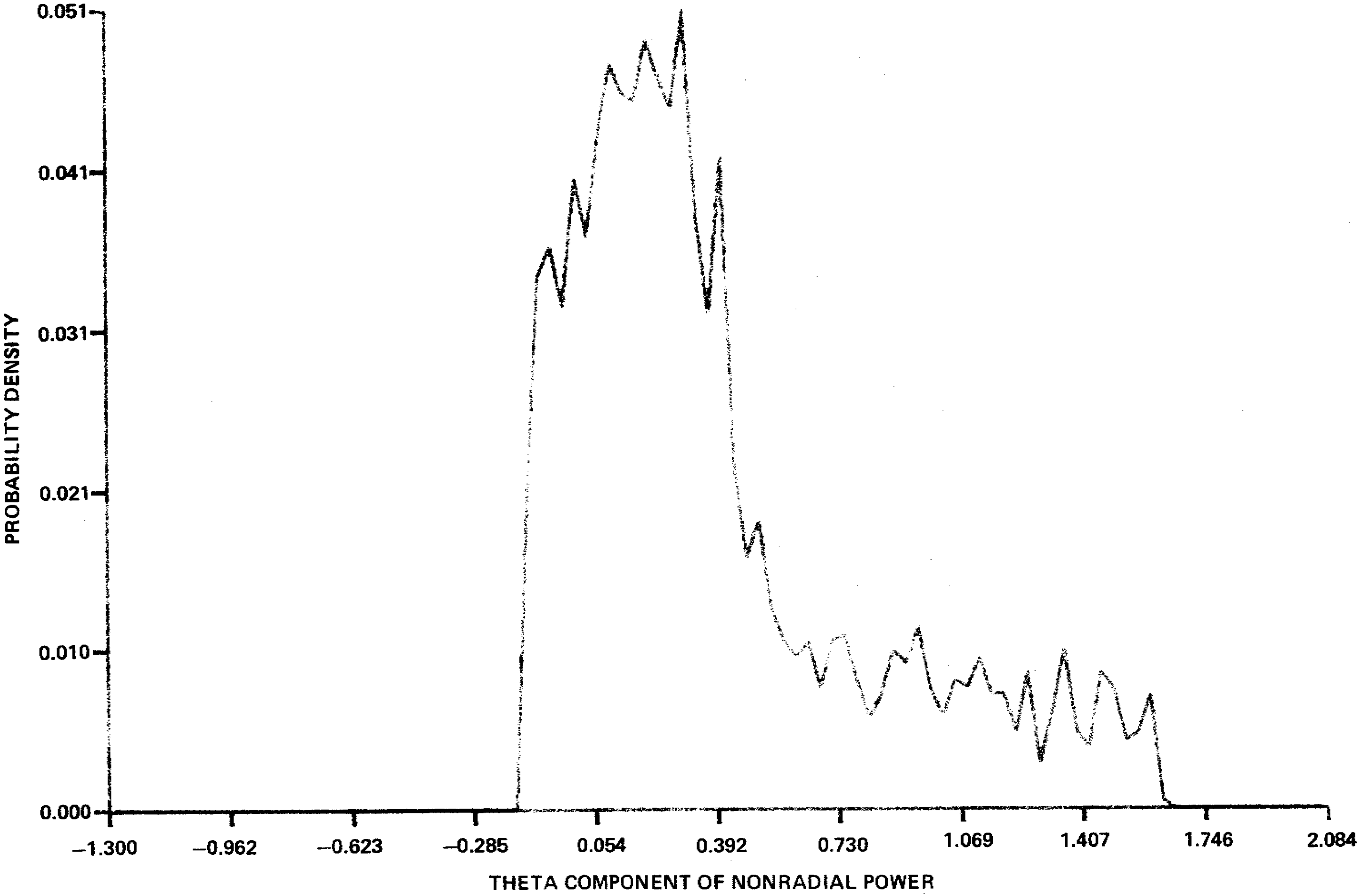


Figure 15. Probability density function of theta component of non-radial power.

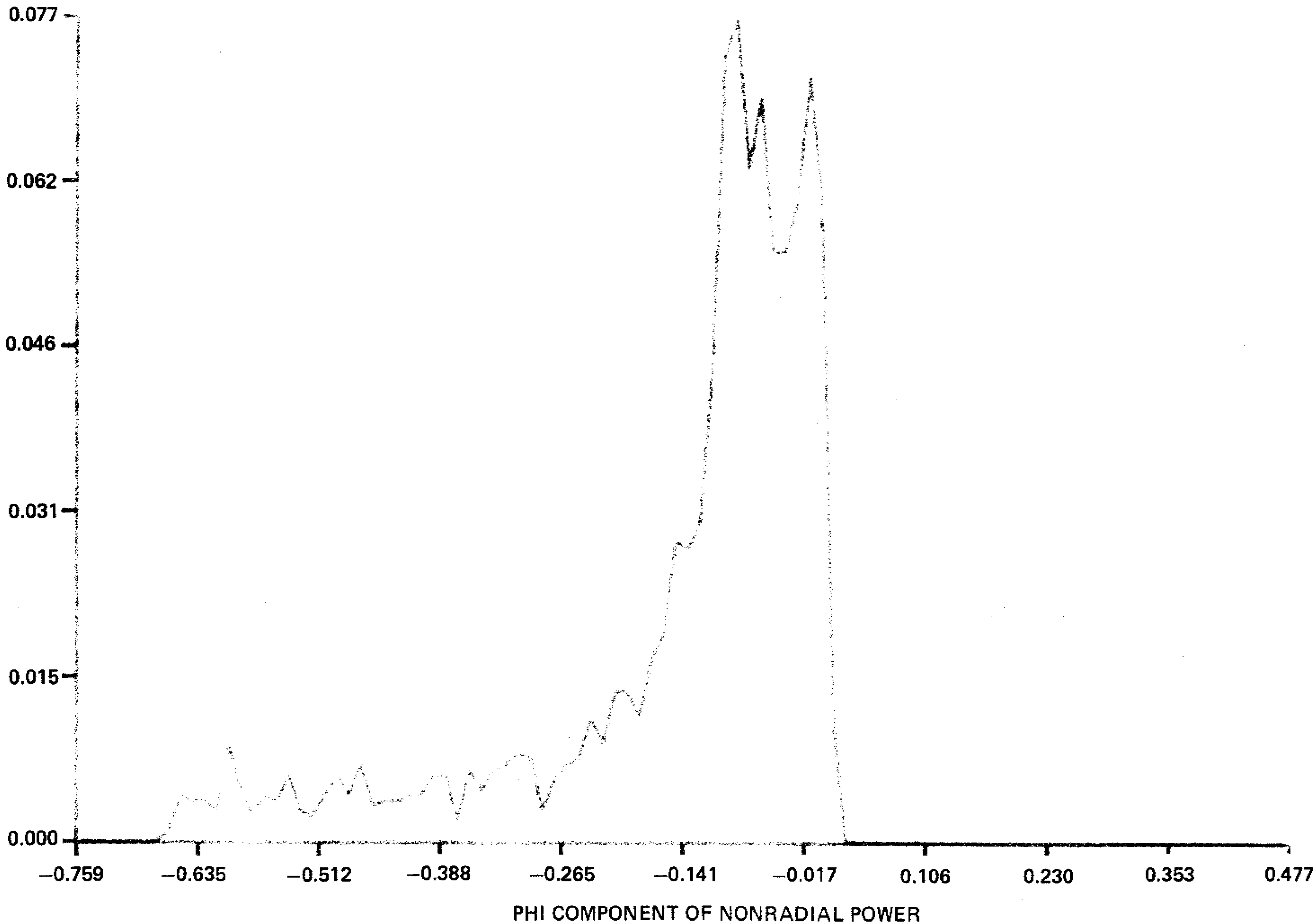


Figure 16. Probability density function of phi component of non-radial power.

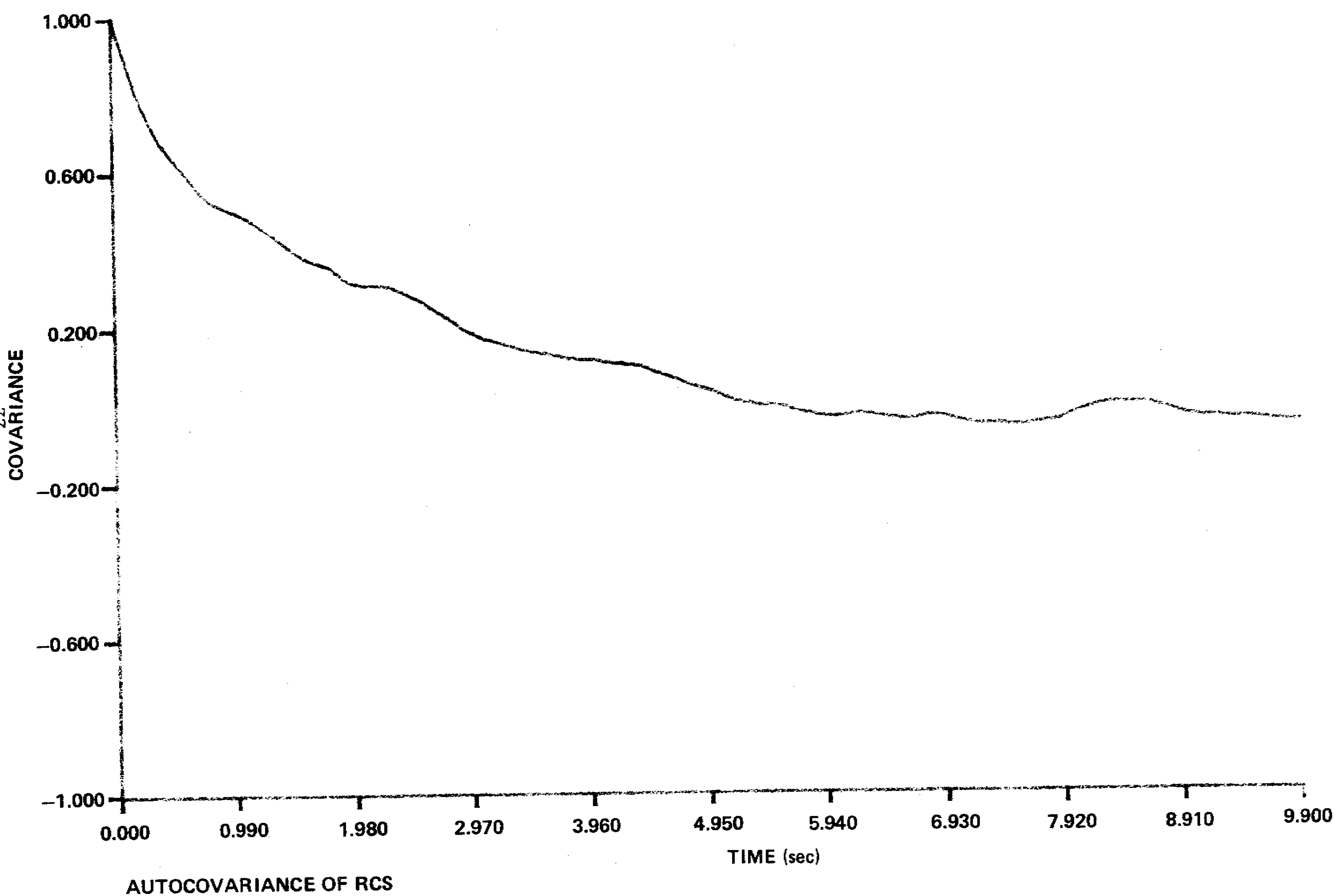


Figure 17. Normalized autocovariance of the radar cross section.

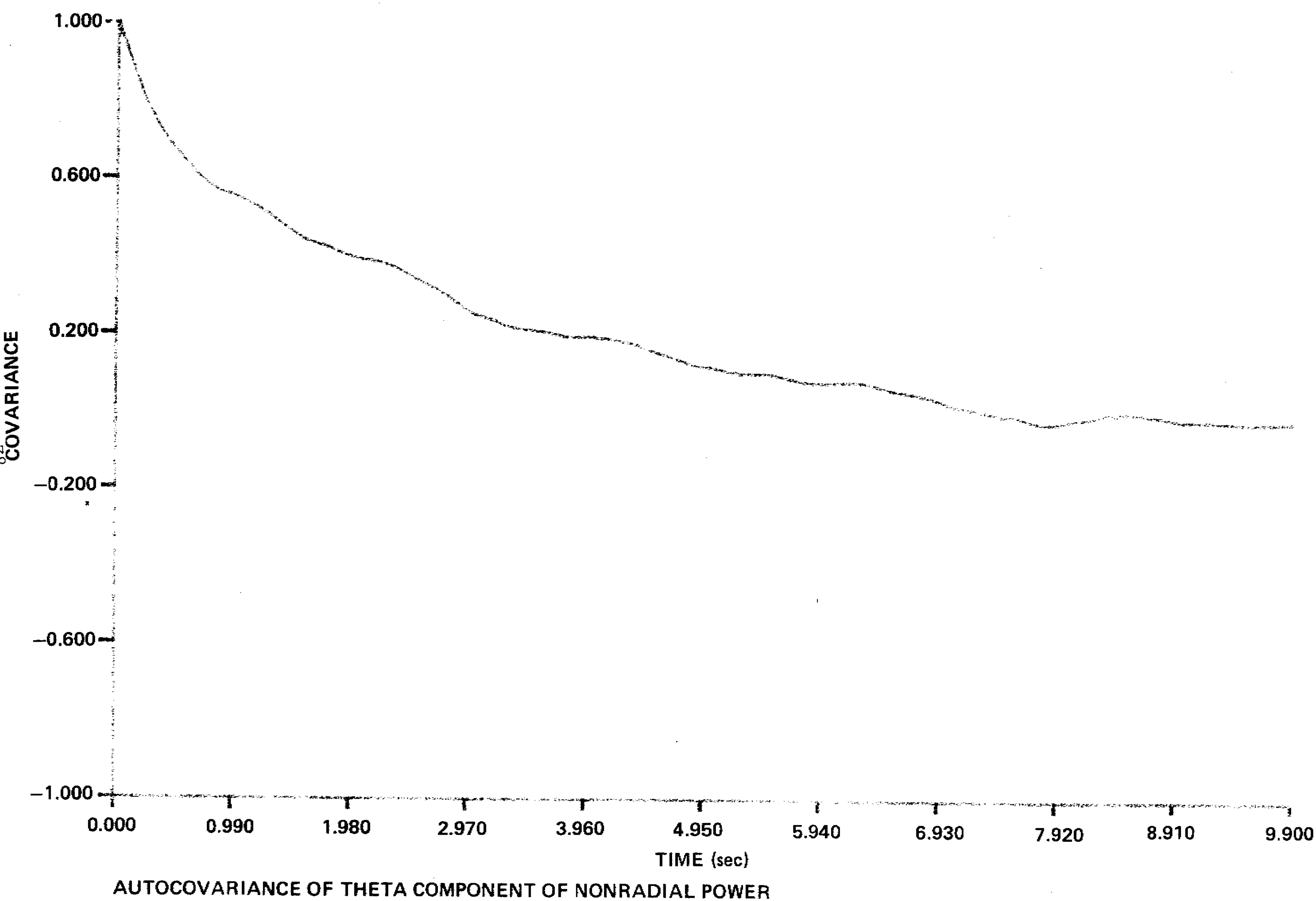


Figure 18. Normalized autocovariance of theta component of non-radial power.

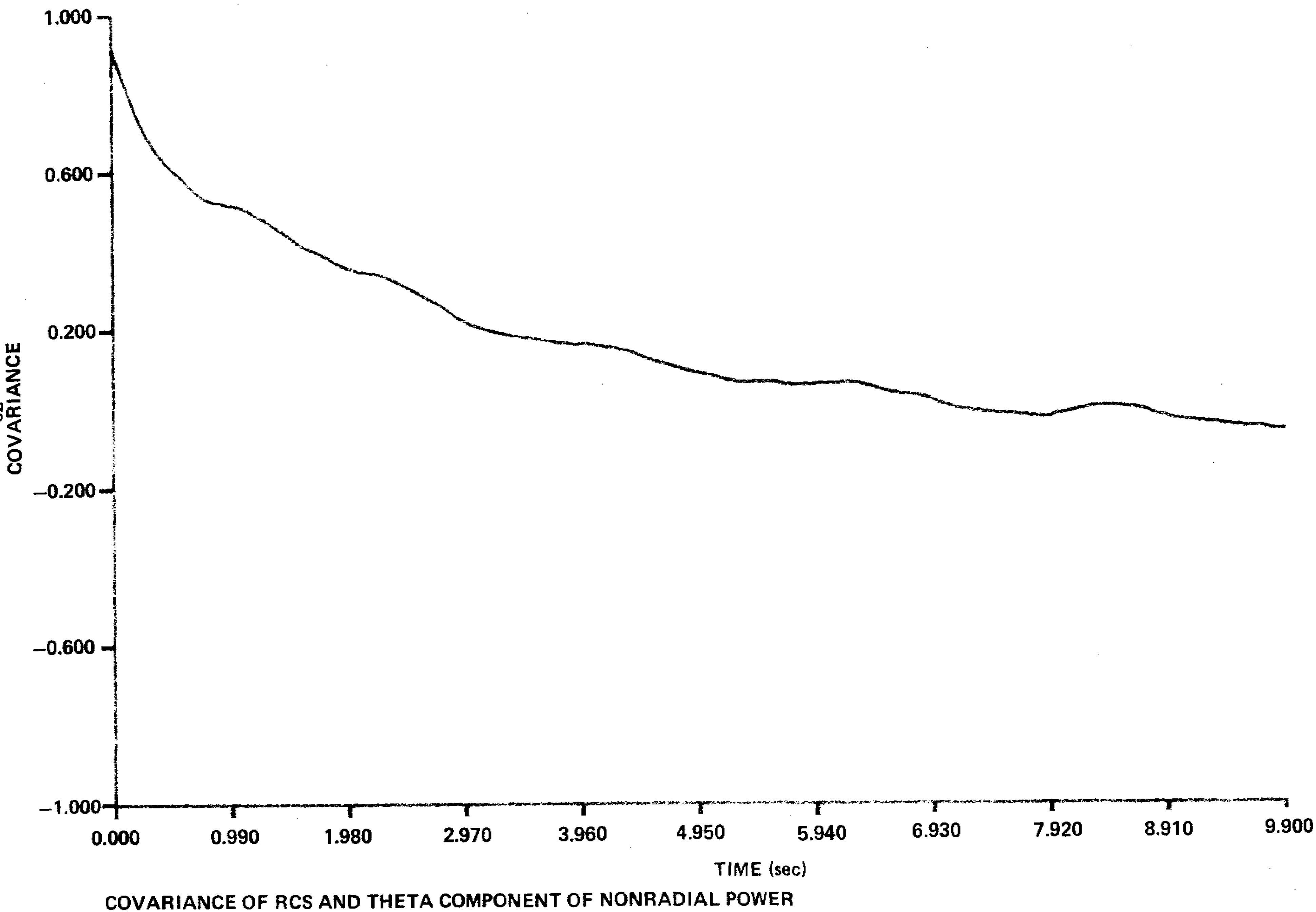
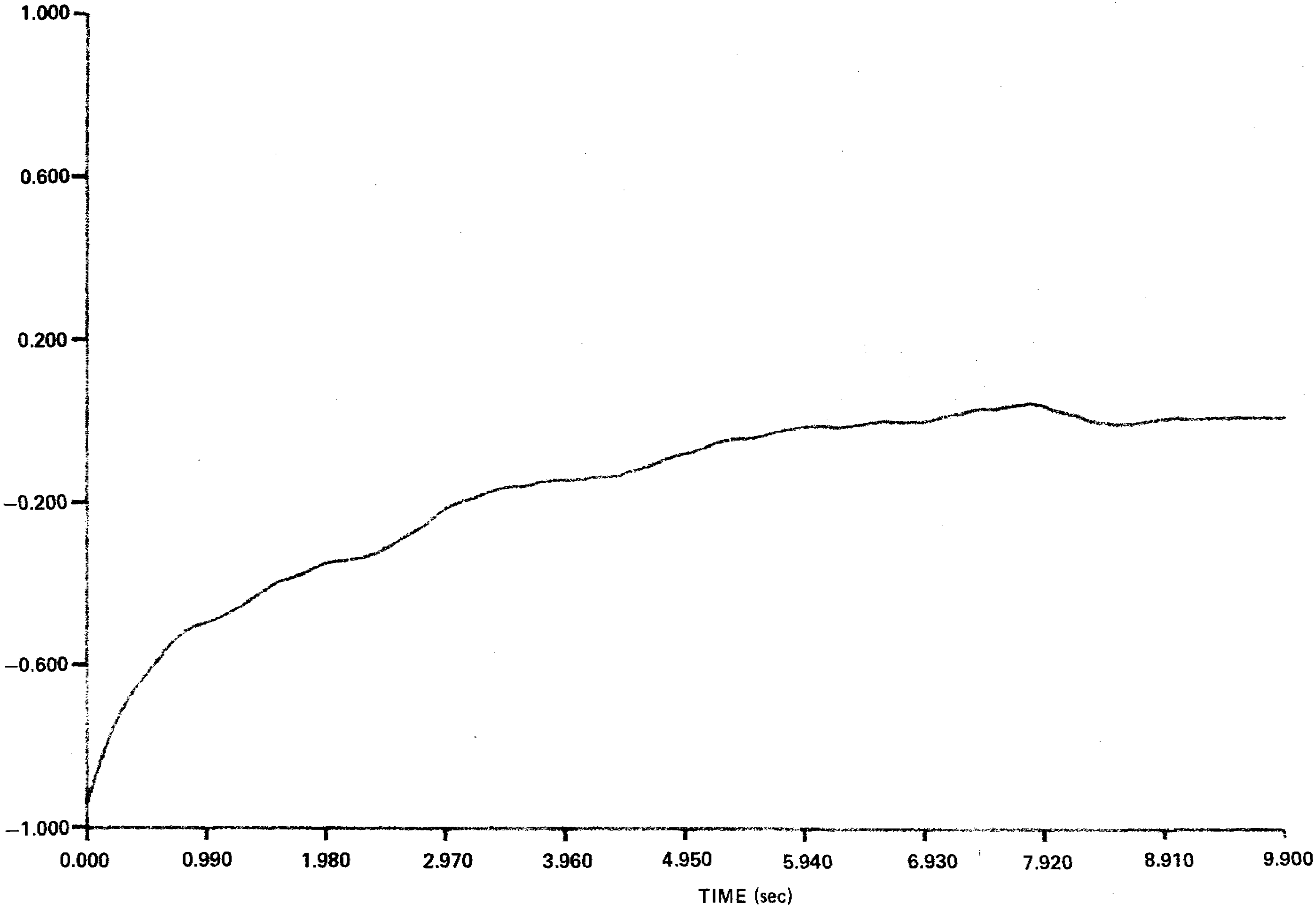


Figure 19. Normalized covariance of radar cross section and theta component of non-radial power.



COVARIANCE OF THETA AND PHI COMPONENTS OF NONRADIAL POWER

Figure 20. Normalized covariance of the theta and phi components of non-radial power.

DESIGN OF EXPERIMENTS FOR THE EVALUATION OF MATERIEL PERFORMANCE IN WORLDWIDE ENVIRONMENTS

Bob O. Benn
Waterways Experiment Station
Corps of Engineers
Vicksburg, Mississippi

Background

A review of recently approved Required Operational Capability (ROC) documents reveals that the user in the Army is requesting a host of materiel items with truly exceptional performance capabilities. A good example of this trend toward materiel sophistication can be found in the ROC's dealing with intrusion detection, target-position location, and target discrimination. It is axiomatic that the more complex the system, the more sensitive it can be to its operational environment. Nevertheless, the systems are intended to function adequately in the majority of worldwide terrain or environmental conditions.

The Test Methodology Directorate, U. S. Army Test and Evaluation Command (TECOM), has recognized for some time that adequate testing procedures are not available for comprehensive evaluation of materiel of the type discussed in the preceding paragraph. To a large extent, the inadequacy results because empirical tests to evaluate materiel items are conducted in a specific (and only a limited number of) terrain or environmental conditions; yet those test results must be extrapolated to worldwide conditions if the evaluations are to be conclusive. As part of TECOM's endeavor to improve its test and evaluation capability, the U. S. Army Engineer Waterways Experiment Station (WES) was asked to develop test guidance and analytical procedures that could be used to extrapolate to worldwide environments the results obtained in tests with unattended sensors and mines in specific terrains.

In general, the problem to be addressed was the development of guidelines for designing an experiment in which the items could be evaluated to determine if they will function above the minimum operating criteria stated in the ROC. Conventionally, the chief objectives of experimental design are to:

- a. Arrange the experiment so that the effects of changing each relevant condition or factor can be readily measured independent of the effects of changing the other conditions and of experimental error.
- b. Obtain a valid estimate of error appropriate for assessing independently or synergistically the statistical significance of the effects of the factors considered.
- c. Enable the effects to be measured with the required accuracy.

Normally, the experiment is arranged so that one (or at most two or three) of the factors or conditions which are known to be significant are varied incrementally while all others are held constant. This permits the effects of those factors to be determined, but only as independent variables. Such a procedure does not permit the evaluation of the effects of all of the relevant factors acting in concert, yet that is invariably the way they act when the device is in operational use. Nor does this procedure make any provision for the recognition of the factors which are not known to be significant at the beginning of the experiment or test. In general, classically-designed experiments respond to this situation only by increasing the experimental error.

The economic aspect of experimentation cannot be emphasized too strongly. Inductive inference from experimental data is subject to error that can be quantified with mathematical statistics; therefore, a measure of accuracy is obtainable. In practice, it is necessary to consider the cost of obtaining a particular accuracy and at what stage the cost of obtaining increased accuracy is too great.

Energy Exploitation

Materiel items that are used for intrusion detection, target-position location, and target discrimination must contain sensors that function by exploiting a wide range of energy propagation forms, such as seismic, acoustic, magnetic, electromagnetic, etc. In general, an item is designed to exploit the energy generated by a target of military interest. To be exploited, the generated energy must be propagated from the target to the sensor. The informational content of the propagated energy, i.e. the information at the sensor, is acted on by the signal processor (i.e. the logic) of the materiel item. The generation of energy in all the various forms is affected by interaction of the generator (target) and the environment, and the energy wave form is further affected by the medium through which it propagates. Every large region of the world exhibits a wide variety of terrain conditions, each of which may change the character of the wave in one or more ways. If the wave is changed to a form which the sensor logic has not been designed to "recognize," the sensor will not respond.

The implication of the above is that to design an adequate experiment to evaluate the item's performance, its interactions with the operational environment must be understood quantitatively. Prerequisite to this understanding is an identification (and quantification) of the factors of the environment that control or significantly influence the interactions. Thus, two major experimental design steps emerge: (a) modeling the item-terrain interaction and (b) quantification of the test environment.

Materiel Item Design

The design of advanced hardware often requires use of specialized technology, and most designs emerge from defense contractors, who may have proprietary rights to software and procedures they have developed. Because of idiosyncrasies in the design procedures, the contractors often optimize the device for certain environments. Unfortunately, testers may not be aware of this because they may not have access to critical information on the rationale and procedures used to develop the item's design. Furthermore, the as-built specifications are not always provided to the test agency. It is the tester's duty to be skeptical, and to try to stress the item in a manner analogous to actual use conditions.

The difficulty in implementing this duty can best be explained by designing an experiment for the evaluation of a specific item. For this discussion a seismic sensor that is required to discriminate among classes of targets is chosen. In this case it is instructional to consider the problems of designing sensor logic that can discriminate among targets.

The difficulty in designing such a seismic sensor lies in the fact that sensor logic must be capable of identifying seismic signal features that are consistently associated with a particular target, regardless of terrain or target conditions. This generalized problem breaks down into three components: (a) the inadequacy of the techniques available for extracting design criteria, (b) the inherent variability of seismic signals, and (c) the inadequacy of available seismic data.

The most popular method of developing design criteria consists of extracting candidate signal characteristics from an empirically generated data base. Briefly, the steps are as follows:

a. Establish signature design data base. This step involves measuring signature data from targets of interest operating at various speeds and in various terrain conditions.

b. Digitize the data and separate them into two batches.

c. Select candidate discriminating features measurable from the time- or frequency-domain signals. Examples include ratios of energy in selected frequency bands, Fourier coefficients, number of zero crossings, peak-to-peak ratios, root mean square values of selected frequency ranges, mean values of selected frequency ranges, etc.

d. Measure candidate features from one batch of the signature data.

e. Correlate features with target classes. Multiple correlation techniques are used in this step to identify the most persistent features and relate them to target classes.

f. Test the correlation derived in step *e*. This step involves testing the model against the second batch of signature data not used in the development of the model. The results are often shown as a probability of classification by target class.

Because these techniques must rely on a finite number of samples, they do not ensure that the logic will function for all situations in the total signature population, i.e. the data base may not have statistical representativeness. As a point of fact, a statistically valid data base may be exceedingly difficult to define. If the design data base is inadequate, the adequacy of resulting sensor design is subject to question.

Evaluation Test Design

Variables to be considered

Because the design technique is subject to statistical probability, the problem facing the test designer becomes one of developing a scheme to establish the performance envelope of the sensor, i.e. demonstrate under what operational and environmental conditions the sensor will and will not meet the design specifications as stated in the ROC. Conventional empirical test procedures in themselves are wholly inadequate to define the performance envelope of a sensor capable of discriminating among targets. For example, consider a sensor designed to discriminate among:

a. Tracked vehicles

b. Wheeled vehicles

c. Man or men

d. Rotary aircraft

e. Fixed-wing aircraft

f. Noise

To decide the number of test interactions required to positively define the sensor's performance envelope, the sources of variance in the signal characteristics (which are correlated with target class) must be examined. Within a given target class, such as tracked vehicles, a number of types exist, e.g. the M113 personnel carrier, the M60 tank, etc. Since the various types within a class do not have identical engines, drive trains, suspensions, etc., it is very reasonable to assume that values for specific signal characteristics (i.e. within a given terrain and at a given distance from the source) will not necessarily be the same for each type of tracked vehicle. Thus, some variance in the values of signal characteristics occurs because all tracked vehicles are not the same.

A similar, but more subtle, variation in the values of specific signal characteristics may occur because all members of a particular type of target within a class may not have exactly identical characteristics. For example, all M113 tracked personnel carriers certainly have similar power supplies, drive trains, suspensions, etc.; however, their physical characteristics (such as spring constants, effective horse power, weight, etc.) may not be identical because of inherent variability in the manufacturing of the vehicle and the different degrees of wear and histories of usage.

A third factor, and perhaps one of the more important, is the influence of terrain conditions on the values for the signal characteristics for a given member of a specific target type and within a specific target class. Three general phenomena must be considered:

- a. Generation of the signal.
- b. Influence of surface and subsurface conditions on signal characteristics.
- c. Influence of surface features (topography) on signal characteristics.

It must be emphasized that the influence of surface and subsurface conditions on signal characteristics is a function of distance from the source in many cases. Thus, the influence of the terrain can be complex indeed.

The generation of seismic signals is a complex process and depends primarily on the target and the terrain conditions on which the target superimposes an input stress. When the target is moving, it is reacting to the irregularities in the surface of the terrain, and as a function of time is passing onto and over a variety of the surface irregularities. The target, such as a vehicle, will react in various ways to different sizes and shapes of surface irregularities and therefore will produce variations in the generated signals as a function of time. In addition to this time/geometry problem, the signal generation process is affected by the subsurface terrain conditions; the coupling of energy into the terrain materials by the target is not the same for all subsurface terrain conditions. Thus, the characteristics of the generated signals can vary because of the variations in the energy coupling phenomena in different subsurface materials (both configurations and properties). The effects of both the surface irregularities and the subsurface characteristics are complicated by another variable, the speed of travel of the target. Thus, the variation in the generated signal due to surface and subsurface effects may have an additional component of variation with changes in target speed.

In addition to a consideration of signal generation, it is necessary to examine signal propagation. Once energy is coupled to the medium, it propagates away from the source in various modes. As the signals propagate, the terrain materials through which they travel alter the frequency and amplitude characteristics of the signals by acting as a filter. The filtering effect of the terrain materials is a function of distance, thus variance in measured signal characteristics can occur (for a given target) because of different terrain conditions and as a function of the distance from the source at which the signals are measured. Furthermore, surface irregularities come into play again, i.e. the signals propagating near the surface may be altered by reflection, refraction, and conversion from one propagation mode to another as a result of the interaction of the signal and the surface irregularities. An additional source of variation occurs if terrain conditions change between the source and the point of measurement. Thus, variations in terrain conditions in general are the cause of many sources of variation in measured signal features. Another complication can be added by noting that many terrain parameters, such as soil moisture content and soil strength, may vary considerably (i.e. at one position) because of changes in seasonal or climatic phenomena (e.g. rainfall, freezing, etc.).

Other sources of variance in signal characteristics exist, such as testing or measurement errors; however, elaboration on these topics is beyond the scope of this discussion. The cogent question to be answered is: How can these sources of variance in signal characteristics (upon which the sensor design is predicated) be isolated and their effects be accounted for in a test program to evaluate the performance of the sensor?

Empirical evaluation

If the assumption is made that the evaluation can be made by empirical testing, it is pertinent to examine the influence of the many sources of variance on the number of tests required. In any empirical study it is necessary to collect sufficient data to define the variation of a given variable under a specified set of conditions. For example, the seismic response (i.e. the amplitudes and frequencies of the seismic signal) of a specific M113 tracked vehicle has some distribution for a given set of terrain and test conditions. Since the initial estimates of variance values are not readily available, statistical theory cannot be used to calculate the number of tests necessary to achieve an adequate evaluation. For this reason, a somewhat cursory analysis must be made by listing the relevant variables and estimating the number of combinations of variables that must be tested.

As stated earlier, both surface and subsurface terrain conditions affect the generation and propagation of seismic energy. Further, these conditions (i.e. surface and subsurface) are dynamic phenomena that are closely related to soil moisture content. Also, the seismic signals are affected by the propagating medium as a function of range. Let it be assumed that the entire spectrum of terrain surface conditions of interest can be represented by 10 specific situations, and the spectrum of subsurface terrain conditions of interest can be limited to 100 specific conditions. Since both surface and subsurface conditions are dynamic phenomena that are closely related to moisture content, let us assume that five different situations (e.g. five moisture conditions, etc.) can occur. Also, let us assume we need measurements at 10 distances from the target to the source. In the extreme, but considering only these factors, the number of possible combinations that must be tested is a striking 50,000.

If this were not bad enough, consider the fact that it is necessary to define the variability in the signal characteristics that may occur for an individual target of a specific type (e.g. a specific M113). To do this let us assume the need for five repeat trials. Also, since all individuals may not react the same, let us test five individuals of each target type within each target class. In addition to this, it must be remembered that there are numerous types within each class, say five. Finally, we are dealing with six classes of targets. When all of these combinations are considered, the resultant number of combinations (or required field tests) could total 37,500,000. Clearly, this is not a viable solution, either technically or economically, and an alternative approach must be sought.

Empirical-theoretical evaluation

Perhaps the most viable solution consists of a balanced experimental and theoretical program. In this approach, well-controlled empirical tests are conducted to ensure that the hardware functions, i.e. it meets ruggedness and longevity specifications and the electrical circuits work. Equally important, the empirical tests demonstrate how the device works in a specific set of test conditions.

In the theoretical portion of the program, realistic simulation models are used to estimate how the device would function if the various terrain and target factors were varied throughout the range of interest. The deficiency in applying the balanced approach centers around the fact that simulation models adequately describing sensor performance as a function of target and environmental conditions are not readily available. Further, for practical applications, they must be formulated such that they accept unique and measurable target and environmental factors. Although difficult, formulation of adequate simulation models is both possible and practical. To illustrate, the following paragraphs briefly describe a seismic sensor performance model developed at the WES. Also presented are examples of how well the signals predicted with the model compare with measured signals from man-walking and vehicle targets. Also, presented are examples of how the signals change as a function of terrain conditions.

The simulation process

The seismic prediction chain for the simulation process developed at the WES is shown in fig. 1. Stress signals are predicted by the various intruder models (e.g. footstep and wheeled and tracked vehicles) for the forces applied to the ground media as the intruder travels over it. The stress signals are used by the microseismic signal model to compute microseismic signals. The microseismic signals are applied to the seismic sensor model, which is used to compute sensor response as a function of site and target properties. The simulation models are described in detail in the WES report entitled "Effects of Environment on Microseismic Wave Propagation Characteristics in Support of SID Testing at Fort Bragg, N. C.; Report 2, Comparison of Summer- and Winter-Season Conditions," by T. L. Engdahl and H. W. West, soon to be published.

It should be noted that the model simulating the sensor can be in the form of a mathematical transfer function, or the predicted signal can be input via magnetic tape directly into the sensor itself. For this reason the sensor does not necessarily have to be modeled, and the critical link in the simulation process is the prediction of the seismic signal. Fig. 2 demonstrates how well this can be accomplished for a signal resulting from a footstep. The two sets of measured curves (figs. 2a and 2b) are for the same walking man at the same Fort Bragg, N. C., site, but the second set (fig. 2b) was recorded after a heavy rain some 10 days later than the first (fig. 2a). In the soils at Fort Bragg (predominantly sand), the footstep signals have approximately the same amplitude and frequency content in both the wet and the dry conditions. This is not necessarily the case for all soils, but, under this situation an intermediate condition ought to show roughly the same signal characteristics. The predicted signals (fig. 2c) are for such an intermediate condition; the surface rigidity data for the footstep model were collected a short time after both sets of footstep signal measurements were taken, when the soil moisture content was intermediate to before-rain and after-rain conditions. It can be seen that the predicted signals have wave forms with characteristics (amplitude, frequency content, and signal duration) similar to those in both sets of measured signals.

Measured and predicted signals from Fort Bragg for an M151 jeep at 32 km/hr are compared in figs. 3 and 4. The measured and predicted time-domain signals in fig. 3 show very good agreement at ranges of 50, 100, 150, and 200 m. For this particular set of predictions, the signals are primarily generated by the suspension as the jeep travels cross-country. If the vehicle had been traveling much slower or traveling over a smooth surface, the suspension component would be reduced and the seismic signal would reveal the engine signal components. Both the predominant frequency and the amplitudes are reduced as the range increases. This is shown more clearly in the frequency-domain signals for the same test (fig. 4). As the range increases, the high-frequency signal components are reduced in amplitude at a much greater rate than the low-frequency components. This causes the peak in the spectrum to reduce in amplitude and shift to lower frequencies.

The terrain inputs required for the man-walking target predictions are: compression spring constants ($k_{s,c}$) and deflection at maximum bearing capacity (Z_{max}) obtained from plate-load tests; and thickness (T), compression wave velocity (V_P), shear wave velocity (V_S), of the various soil layers as defined by seismic refraction surveys, and soil wet density (ρ). Normally good comparison of predicted and measured signals can be obtained if only the first and second layers are considered. In addition to the surface parameters (i.e. $k_{s,c}$ and Z_{max}) discussed above for a man-walking target, a surface geometry profile is required as an input to both the wheeled and tracked vehicle models. Subsurface data (i.e. T, V_P, V_S), and ρ) requirements are identical to that required for making a prediction for the signal from footsteps.

<u>Extrapolation of test results</u>

Extrapolation of test results to environmental conditions outside the test area is accomplished by varying the environmental factors discussed in the preceding paragraph. For example, various combinations of the factors representing soft soil, firm soil, and frozen ground are shown in figs. 5, 6, and 7. Predicted time- and frequency-domain signals for a man-walking target at a range of 5 m is shown in fig. 8. These examples (which cover a wide range of soil conditions) show that the particle velocity amplitude for the frozen ground is about two orders of magnitude (from $\approx$ 20 to .2 cm/sec x 10^{-3}) less than that for the soft soil. Also, the energy is propagated at higher frequencies as the soil rigidity increases.

Fig. 9 shows predicted results for a Soviet light tank (PT76) on the same terrain conditions as in the man-walking predictions. The speed of the vehicle is 5 mph, and it is at a range of 75 m from the sensor. These plots show dramatically how the shape of wave forms from a vehicle depends on terrain conditions. As with footstep signals, the amplitude decreases with soil rigidity. Also, the dominant frequency increases with soil rigidity.

Conclusions

It must be recognized that the various terrain and target parameters can combine so as to have a synergistic effect on the resultant wave forms; therefore, many combinations of terrain factors must be evaluated. The WES sensor performance models use algorithms that can be solved efficiently; therefore, they provide a means for generating a relatively large number of predictions at a low unit cost. Work is now being directed toward devising a listing or matrix of terrain factors to provide a data base for the comprehensive evaluation of any seismic sensor. It is felt that this balanced approach, i.e. balance between empirical testing and theoretical extrapolation, can be directly applied to the evaluation of many items of advanced materiel.

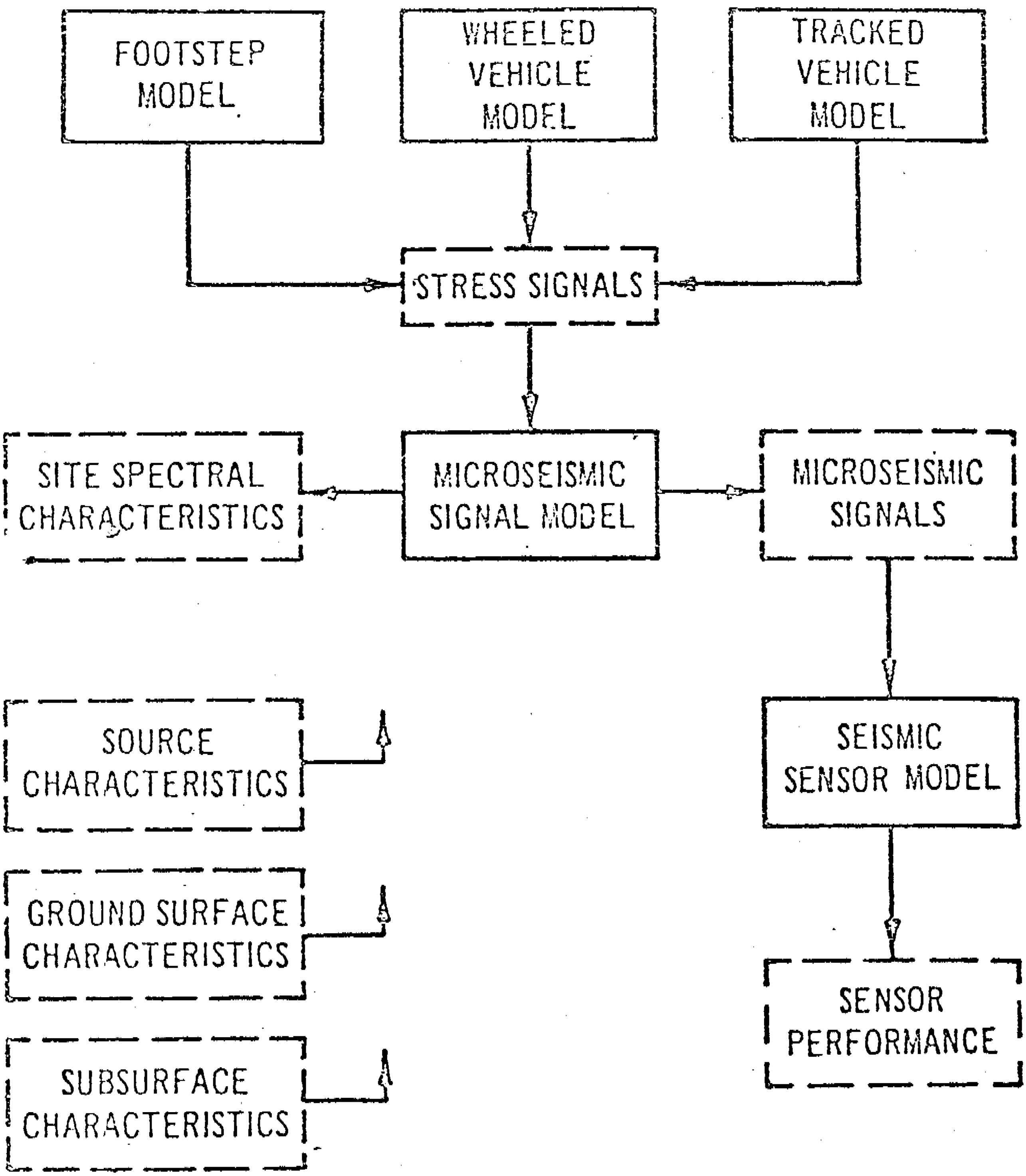
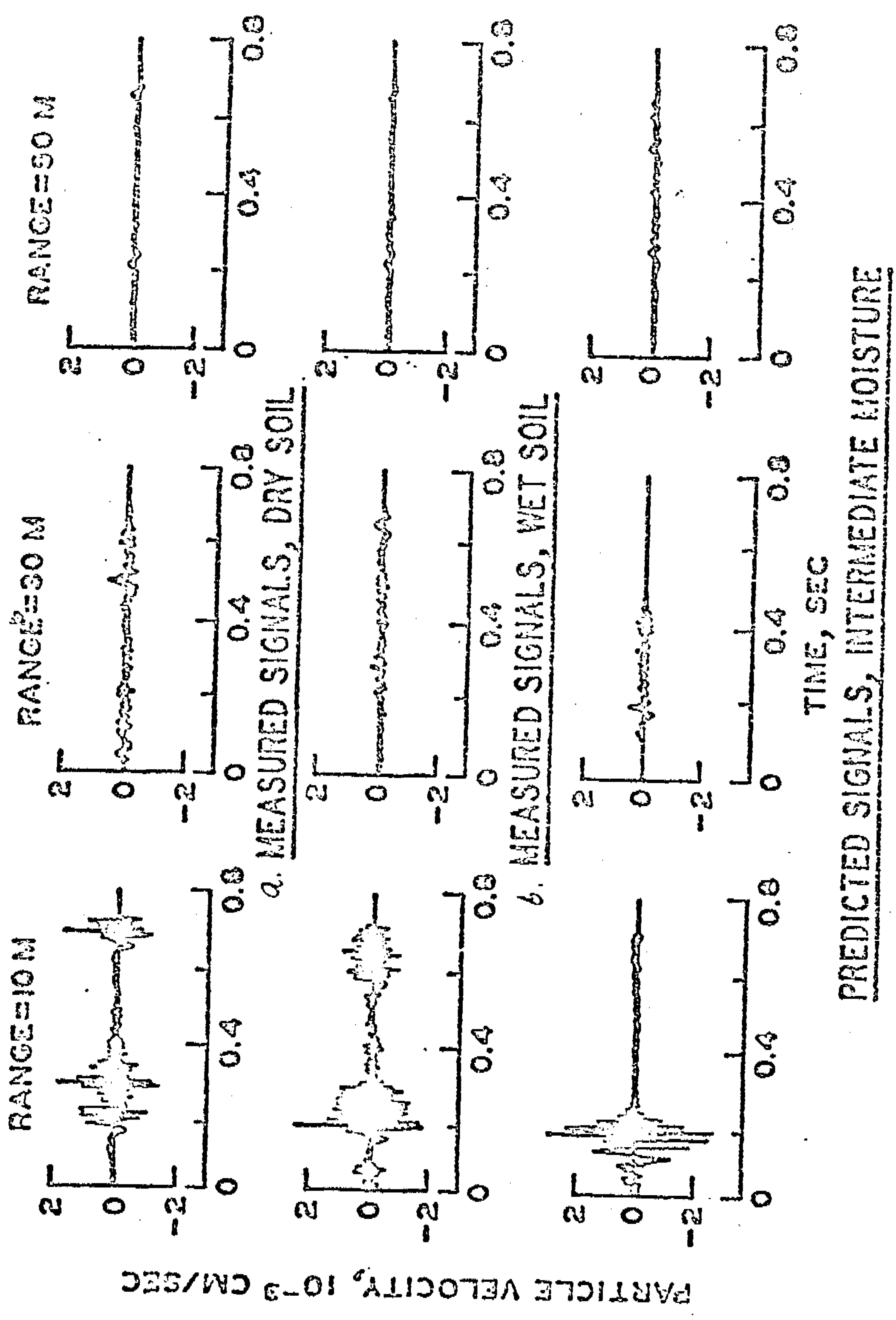
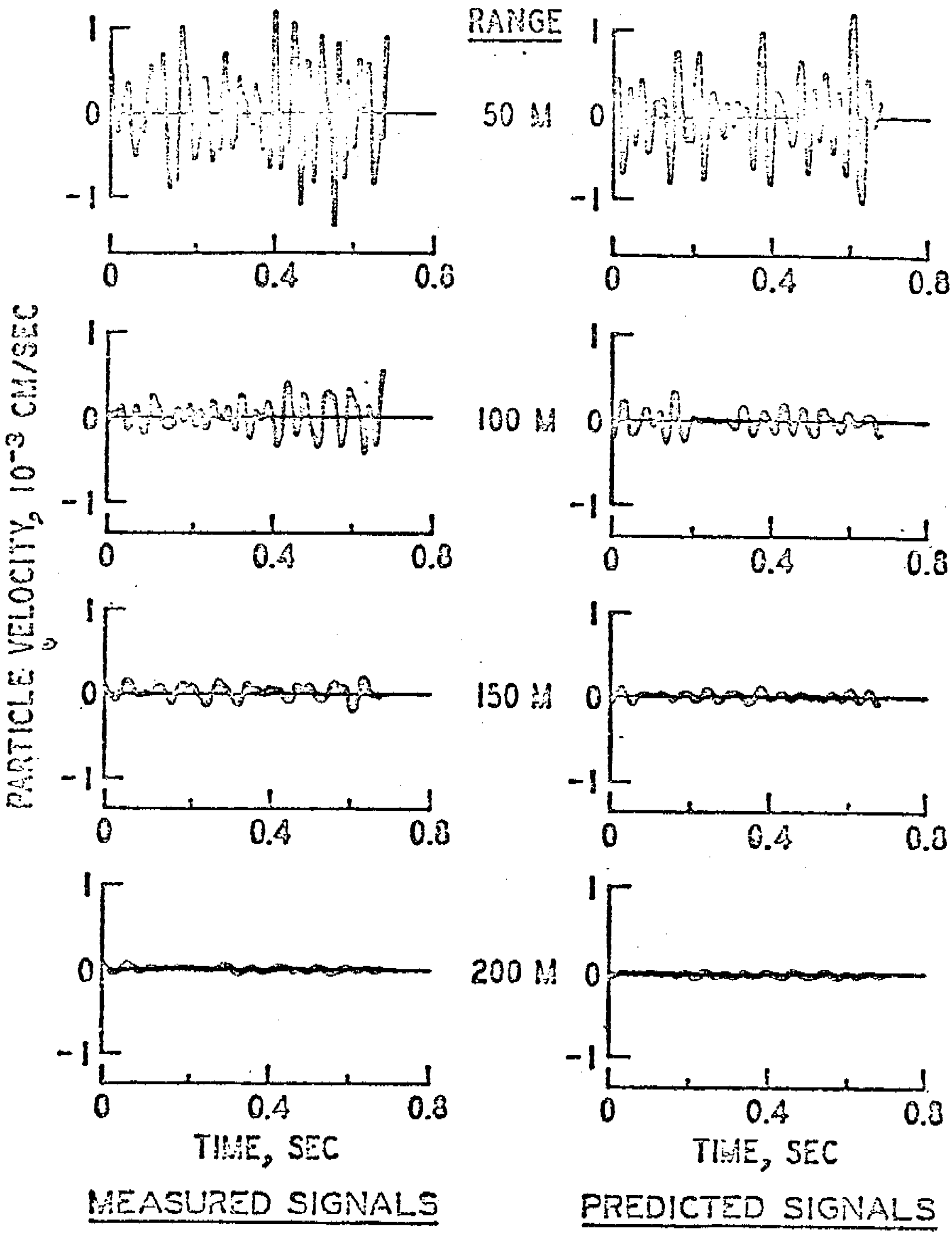

FOOTSTEP MODEL
WHEELED VEHICLE MODEL
TRACKED VEHICLE MODEL
STRESS SIGNALS
SITE SPECTRAL CHARACTERISTICS
MICROSEISMIC SIGNAL MODEL
MICROSEISMIC SIGNALS
SOURCE CHARACTERISTICS
SEISMIC SENSOR MODEL
GROUND SURFACE CHARACTERISTICS
SENSOR PERFORMANCE
SUBSURFACE CHARACTERISTICS

FIGURE 1



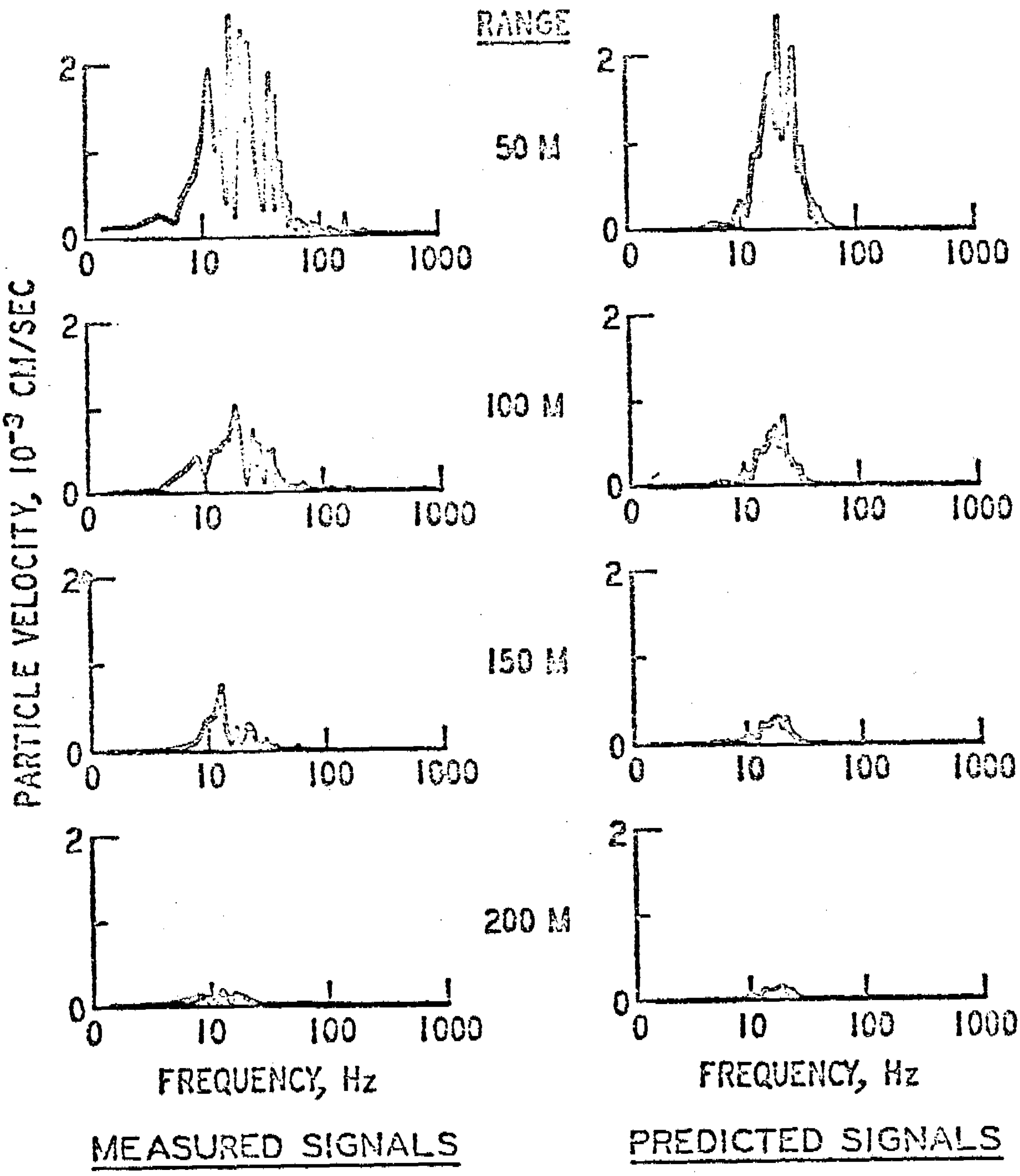
COMPARISON OF FOOTSTEP SIGNALS, FORT BRAGG, N.C.

FIGURE 2



M151 (JEEP) SIGNALS
32 KM/HR, TIME DOMAIN

FIGURE 3



M151 (JEEP) SIGNALS
32 KM/HR, FREQUENCY DOMAIN

FIGURE 4

SOFT SOIL

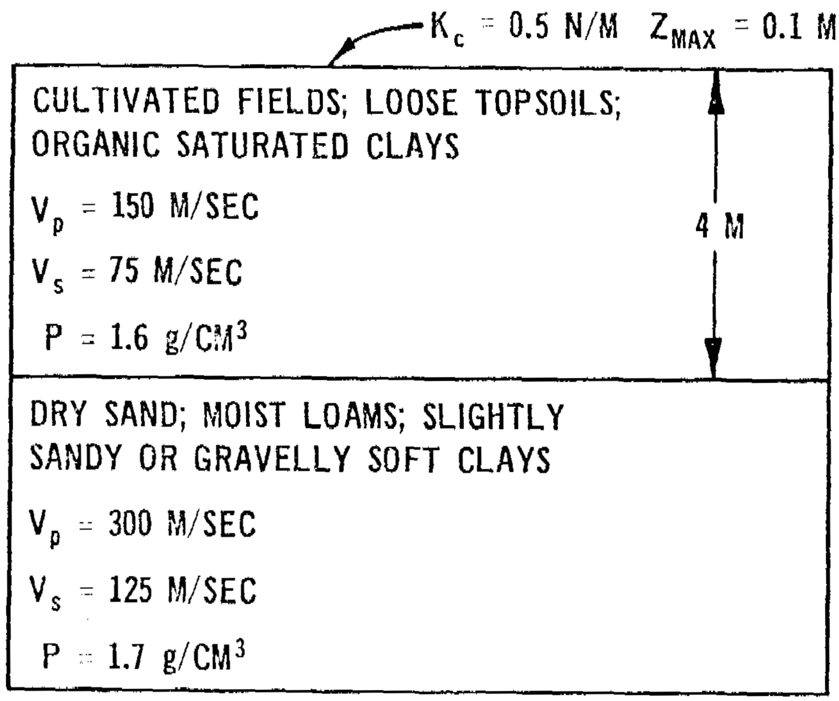


FIGURE 5

FIRM SOIL

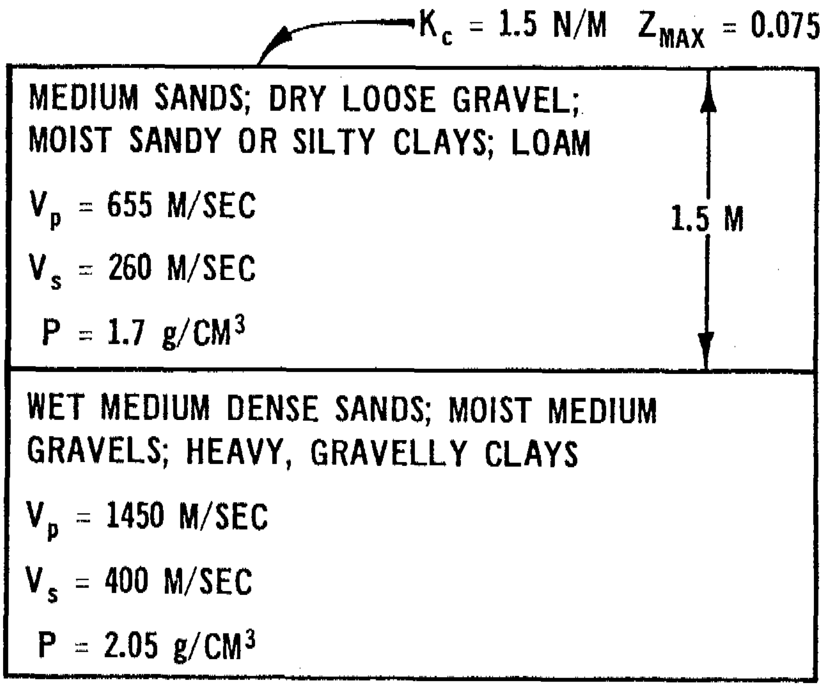


FIGURE 6

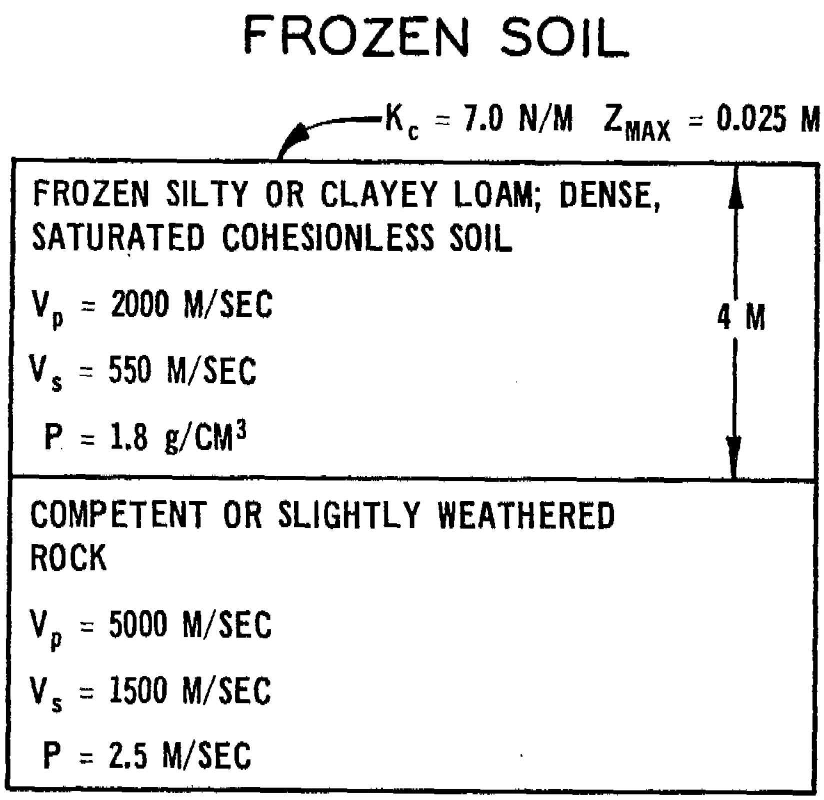
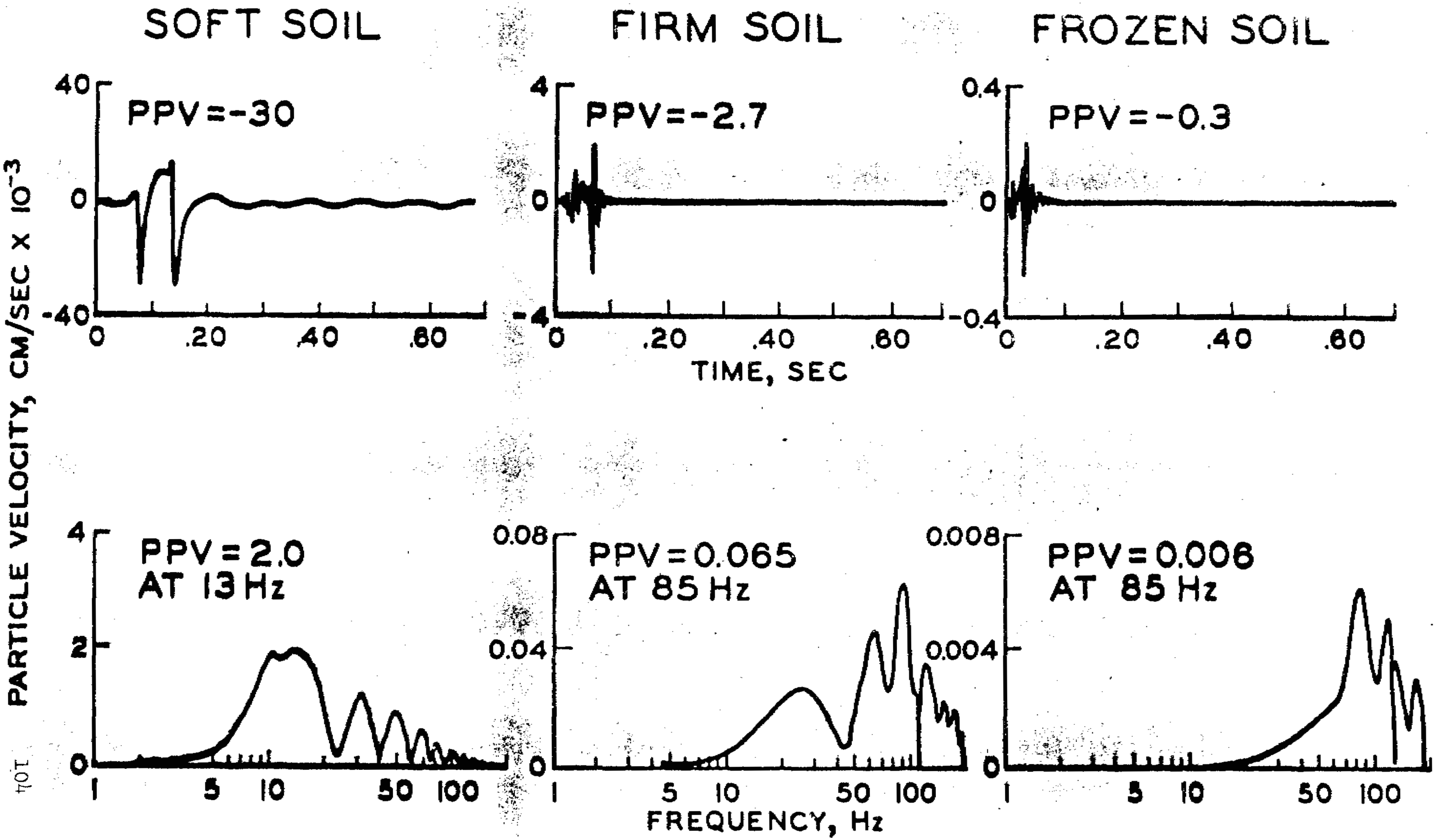
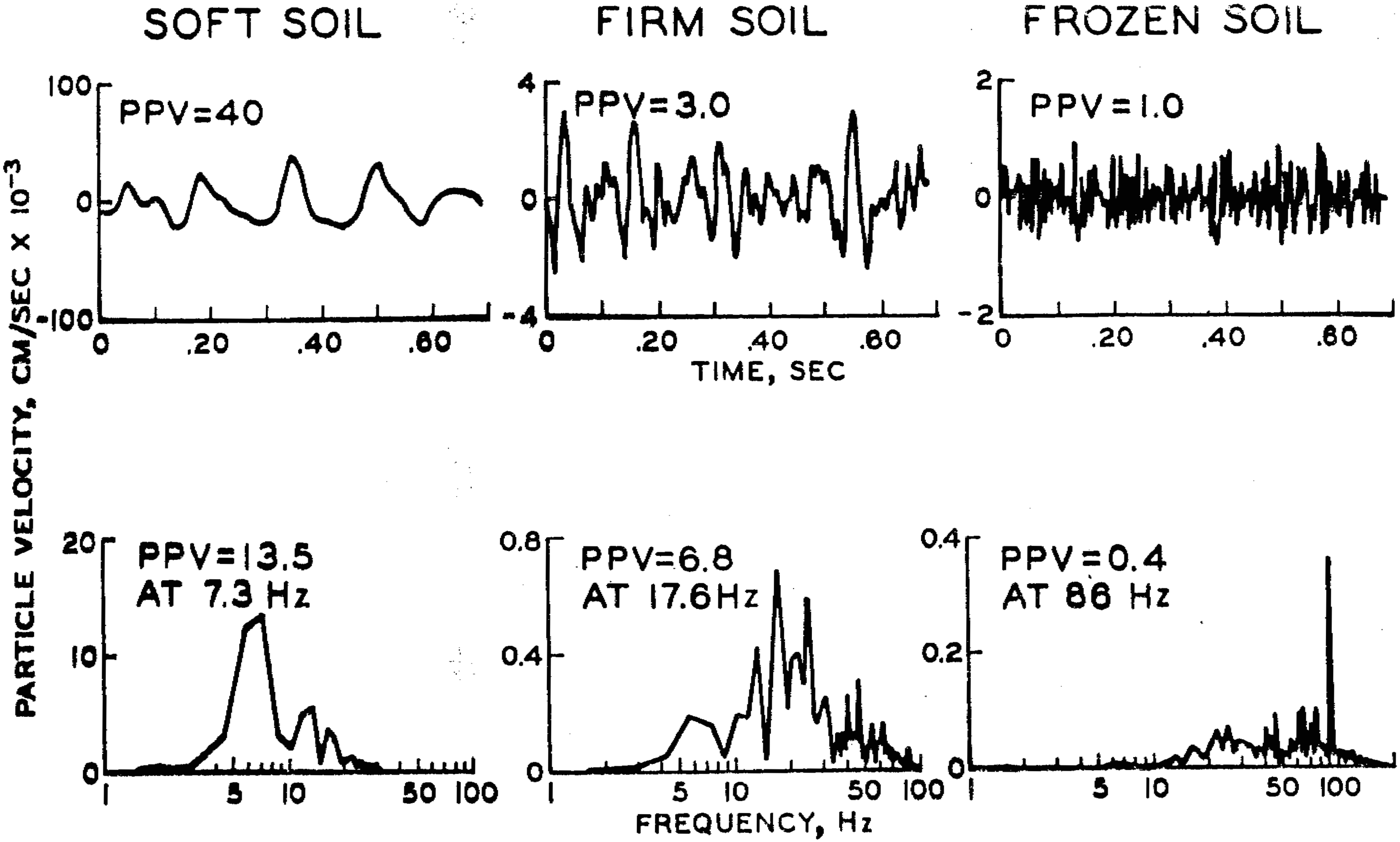

FROZEN SOIL
K_c = 7.0 N/M Z_{MAX} = 0.025 M
FROZEN SILTY OR CLAYEY LOAM; DENSE, SATURATED COHESIONLESS SOIL
V_p = 2000 M/SEC
V_s = 550 M/SEC
P = 1.8 g/CM^3
4 M
COMPETENT OR SLIGHTLY WEATHERED ROCK
V_p = 5000 M/SEC
V_s = 1500 M/SEC
P = 2.5 M/SEC

FIGURE 7



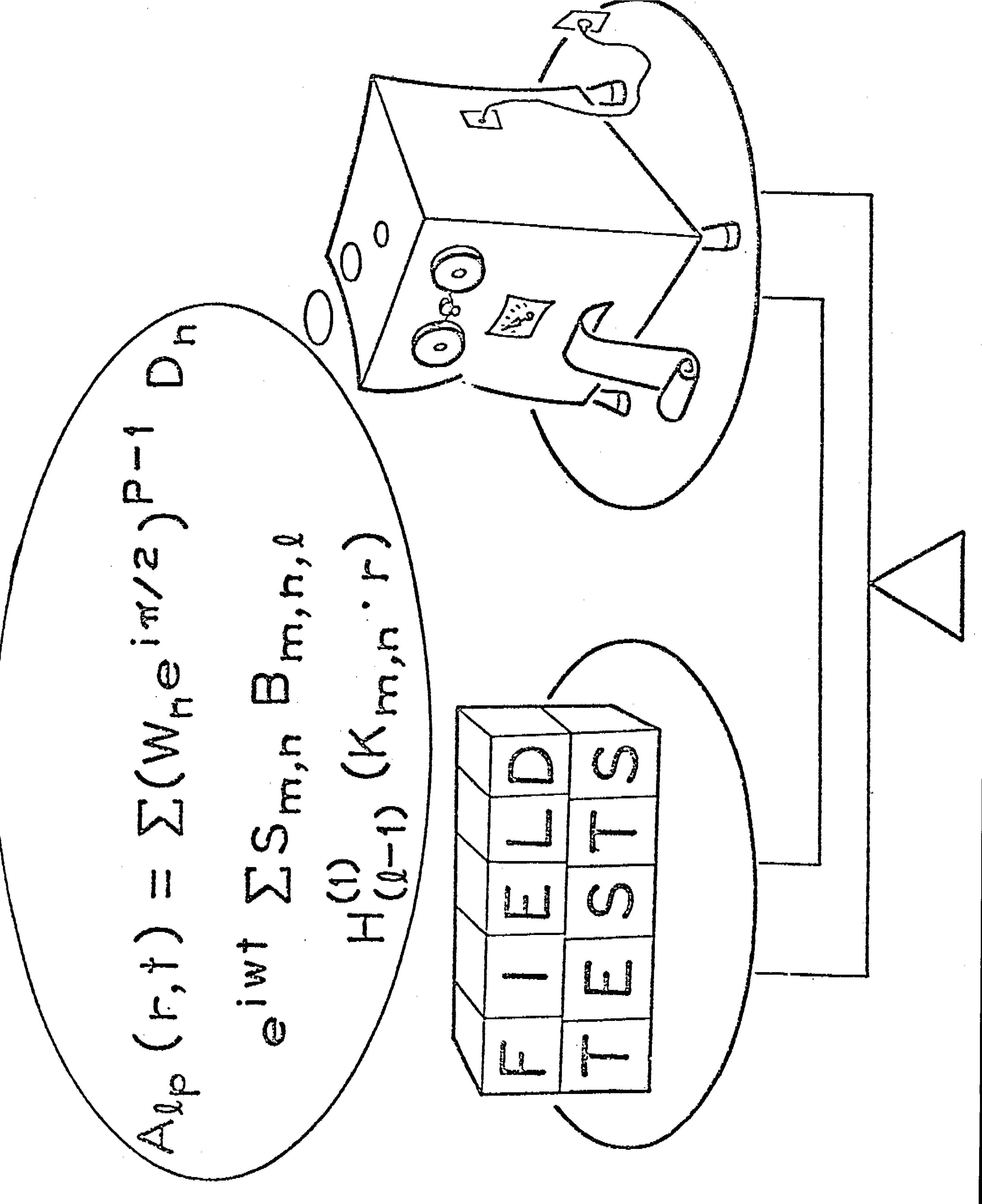
PREDICTED SEISMIC RESPONSE AT 5 M
(MAN WALKING)

FIGURE 8



PREDICTED SEISMIC RESPONSE AT 75M
(PT 76, 5 MPH CROSS-COUNTRY)

FIGURE 9



$A_{\ell p}(r,t) = \sum(W_n e^{i\pi/2})^{P-1} D_n$
$e^{iwt} \sum S_{m,n} B_{m,n,\ell}$
$H^{(1)}_{(\ell-1)}(K_{m,n} \cdot r)$
FIELD
TESTS

FIGURE 10

SHORT PULSE TESTING OF EEDs
AND
THE BRUCETON PROBLEM

Ramie H. Thompson
The Franklin Institute Research Laboratories

and

Burton V. Frank
Nuclear Engineering Directorate
Picatinny Arsenal

1. INTRODUCTION

The authors have recently completed tests of the short pulse response of some electroexplosive devices. The primary results were evaluated using the classical electroexplosive device statistical techniques. This paper describes the test equipment and technique in some detail. Hopefully, enough detail to allow the reader to get a feeling for the accuracy of the techniques and yet not bore him completely. In any case the section describing the stimuli and their generation can be skipped without any great loss of knowledge about the central question we wish to raise. That question is straightforward. Are the commonly used statistical techniques of electroexplosive evaluation adequate for the use to which we put the information generated by these techniques?

The authors suspect that the presently used techniques are the best of a bad lot-the result of economic and theoretical compromise-and frankly seek suggestions for improvement or alternate techniques.

2. BACKGROUND

The Applied Physics Laboratory of The Franklin Institute Research Laboratories has been involved with the evaluation of the DC and RF (both pulsed and continuous wave) responses of Electroexplosive Devices (EED's) for about twenty years. Recently an interest in the response of EED's to very short duration/high amplitude electrical stimuli has been prompted by concern about possible Electromagnetic Pulse (EMP) interactions with EED's. The original work we performed in this field used damped sinusoidal stimuli but interest has shifted to the more easily produced and controlled rectangular pulse shape. All of the work discussed here uses the short rectangular pulse as the basic EED stimuli.

How short is a short pulse? We have conducted extensive tests using 25 ns, 50 ns and 100 ns pulses but pulse lengths can be increased without trouble to about 10 microseconds. The primary advantage of our specialized pulse supplying equipment is the ability to monitor our high amplitude stimuli and responses without interference.[†]

3. THE STIMULI AND THEIR GENERATION[‡]

Conventional type twin lead EED's can be initiated by two ways:

(1) Passing current through the bridgewire (the conventional firing mode) and

(2) Application of a high voltage between the pins of the EED and the metal case.

Our pulse generating equipment can supply short pulses to the EED in either of these firing modes. Figure 1 shows a typical high current short duration pulse applied to the bridgewire of a conventional type hot wire EED. The oscilloscope traces shown here have been traced from the actual oscilloscope photographs. Note that the waveform on the left shows the current thourgh the bridgewire of the EED and the waveform on the right shows the voltage across the bridgewire. The time scale is the same on both photographs and the individual sweeps start at the same time. Figure 2 shows a high voltage short pulse applied to the pin-to-case firing mode of an EED. Note that the current is very low and that this is associated with a "no-fire" response of the EED. This current is that which actually flows between the pins and the case of the EED during the voltage pulse application.

Figures 3 and 4 show simplified equivalent circuit schematics for the pin-to-pin and pins-to-case short pulse testing equipment configurations and Figure 5 is an overall schematic for the actual system.

[†]The short pulse generating equipment is described in great detail in "Pins-to-Case Short Pulse Sensitivity Studies for the Atlas PC Switch," FIRL report I-C3410 produced for Picatinny Arsenal.

[‡]Most of the material in this section is taken directly from "Short Pulse Testing of EED's," a paper by R. H. Thompson given at the 8th Symposium on Explosives & Pyrotechnics at The Franklin Institute, Phila., Pa. in Feb. 1974.

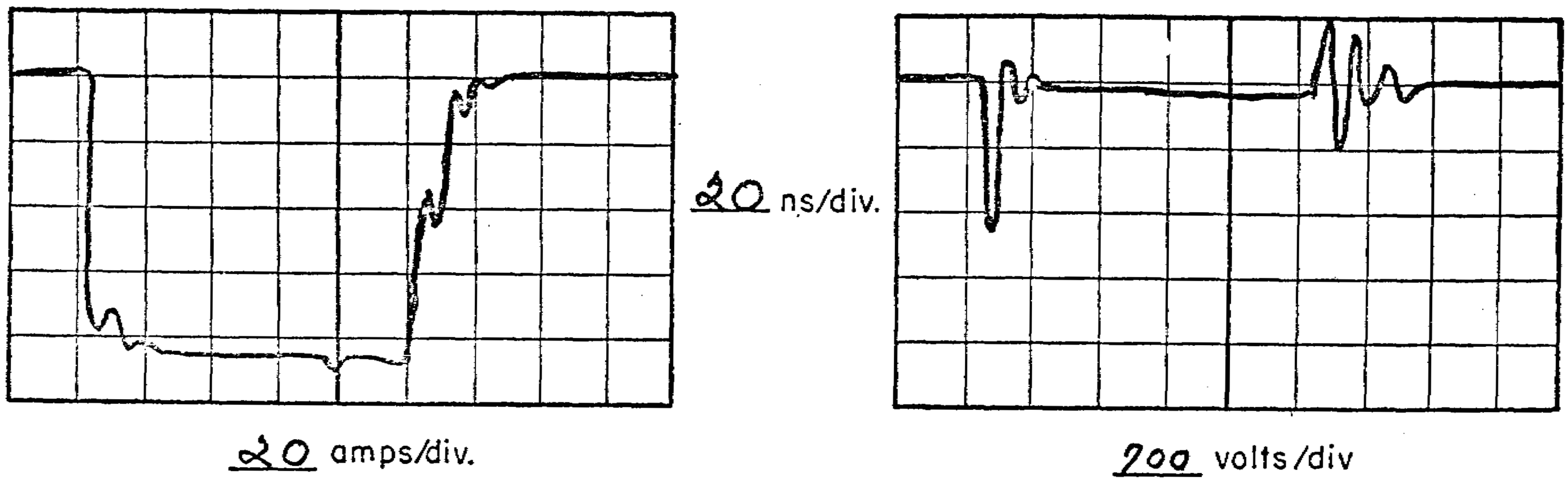


Figure 1. High Current Square Wave Applied to a Bridgewire

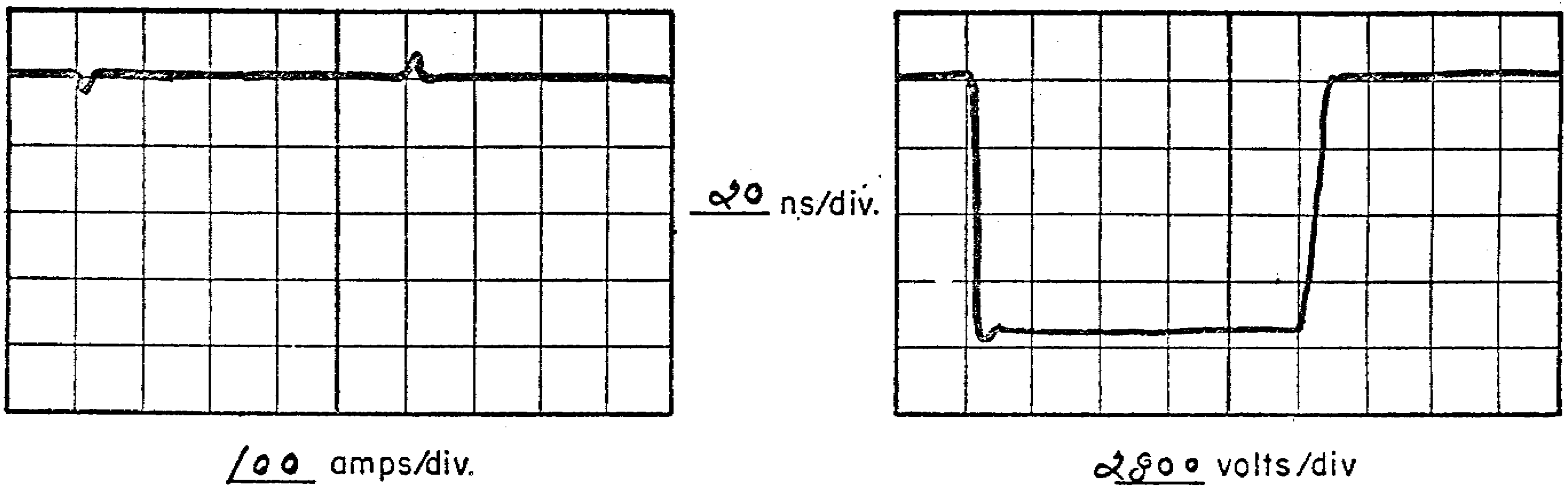


Figure 2. High Voltage Square Wave Applied to an EED Pin-to-Case Impedance

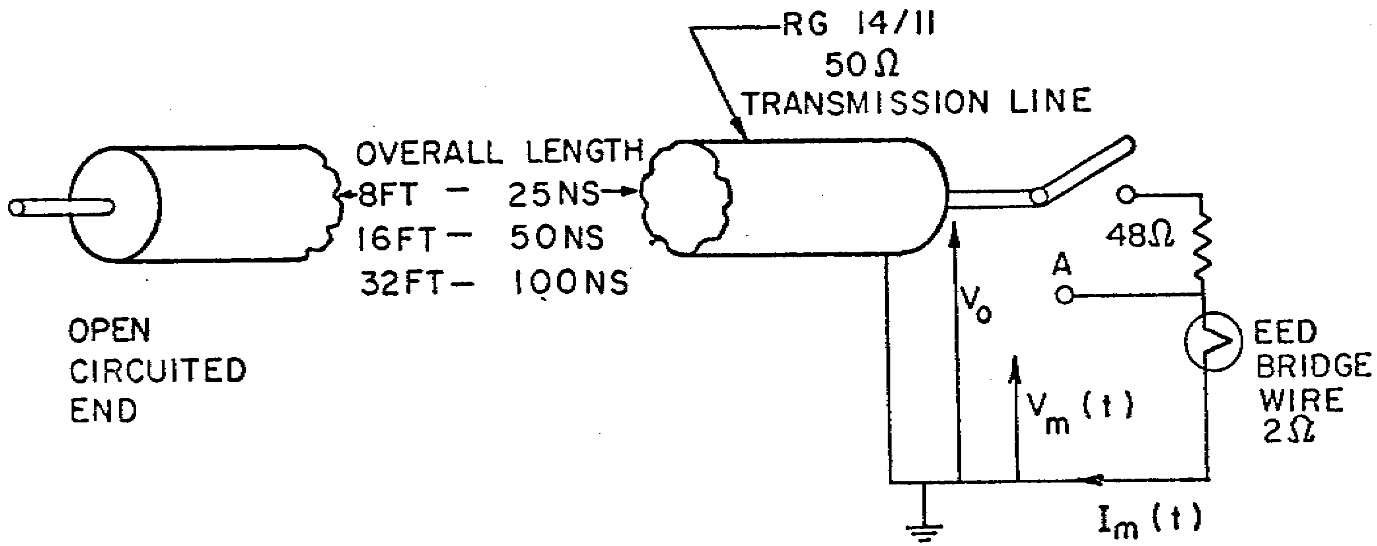


Figure 3. Equivalent Circuit for Pin-to-Pin Exposures

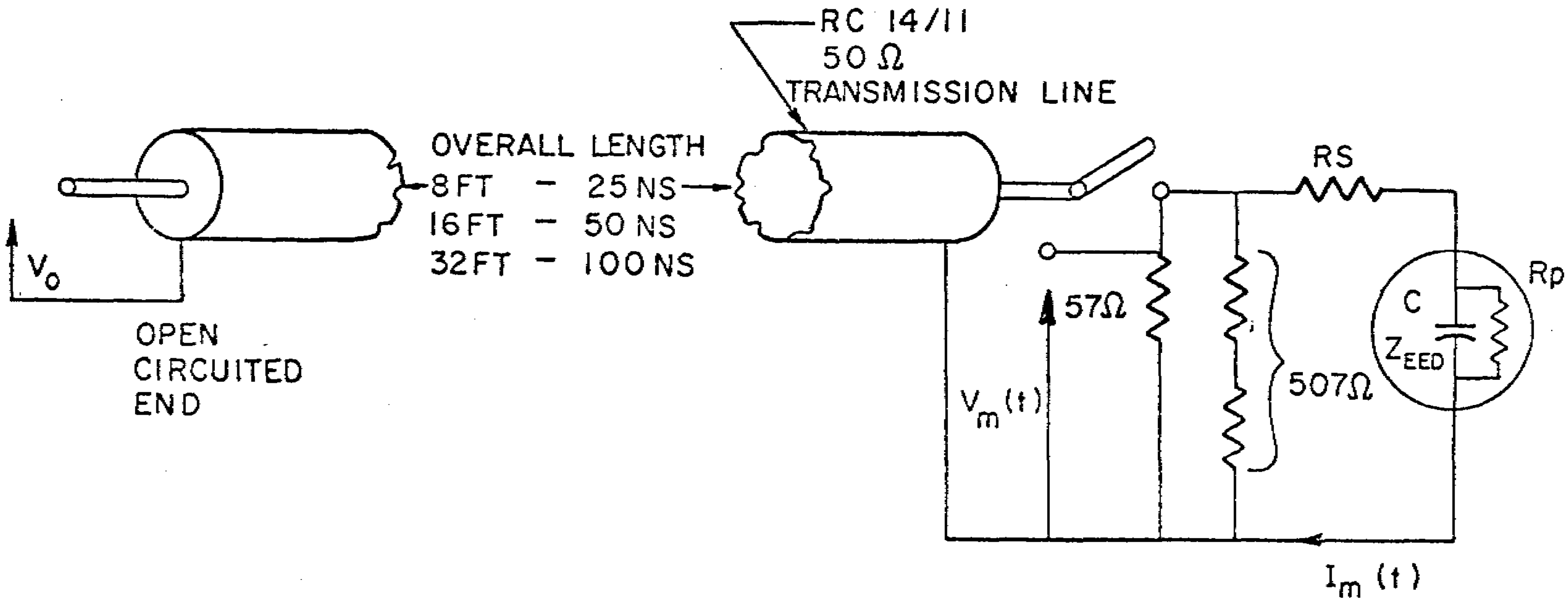


Figure 4. Equivalent Circuit for Pin-to-Case Exposures

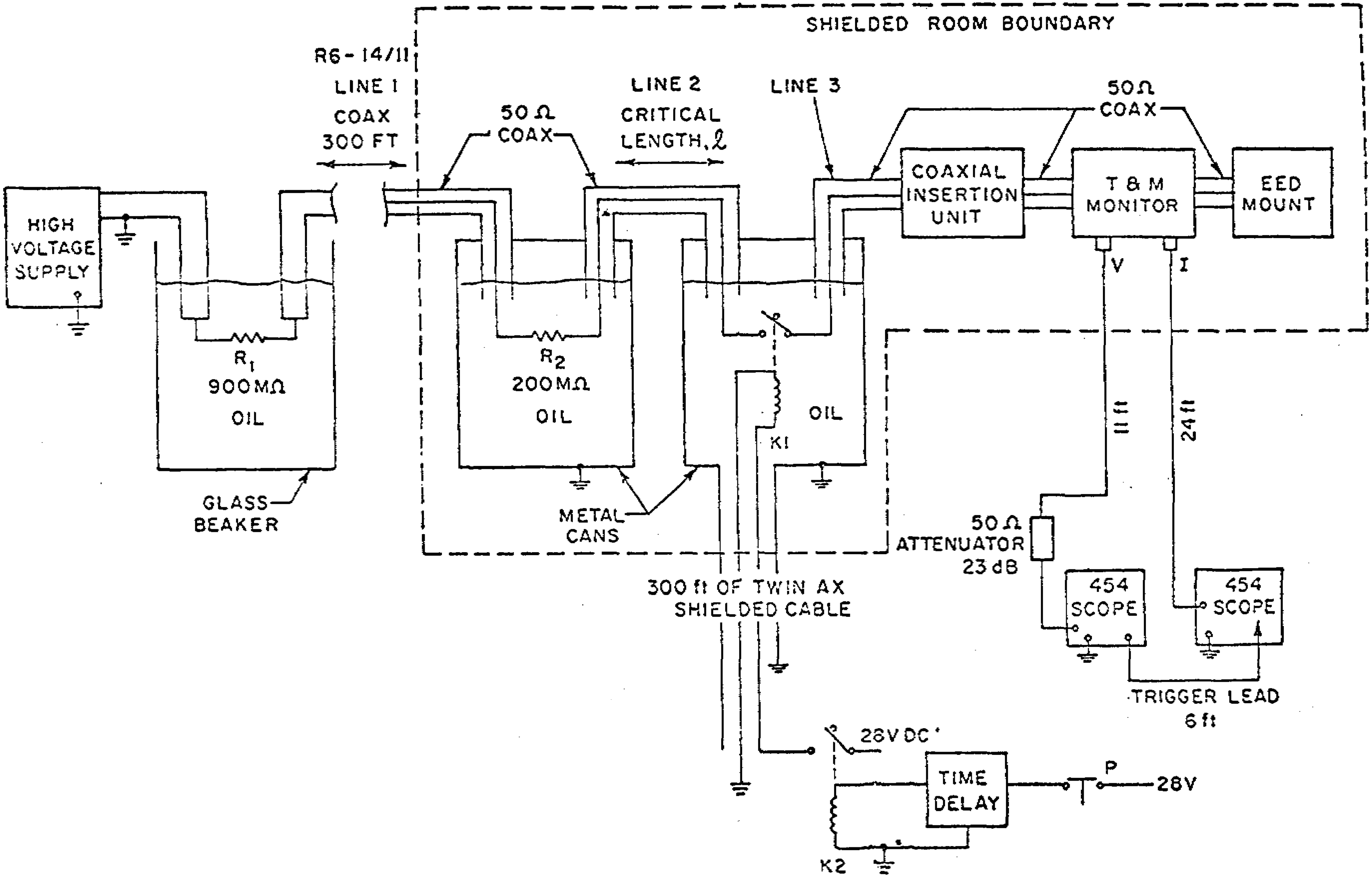


Figure 5. The Overall System Schematic

In operation the equipment is used in this order: (Refer to Figure 5).

1) Adjust the high voltage supply for the desired voltage and polarity. This charges transmission lines 1 and 2 through the large resistors R_1 and R_2.

2) The closure of pushbutton P energizes time delay relay K2 for approximately 2 seconds.

3) Relay K2 energizes the 40 KV vacuum relay K1 for 2 seconds.

4) As the contacts of K_1 begin to close they eventually reach a point where the charged transmission line (line 2) arcs to the movable contact of K1. This arc effectively connects transmission lines 2 and 3 together. This switching action is roughly equivalent to that of a perfect switch so we approximate the theoretical situation of a charged transmission line connected instantaneously to a load.

5) The electromagnetic disturbance caused by the switching propagates from the relay contacts in two directions: both through line 3 toward the insertion unit and back through line 2 toward R_2. R_2 is chosen to be very large in relation to the 50 ohm characteristic impedance of line 2 so that it approximates an open circuit. The insertion unit contains a series or parallel resistor such that the impedance looking toward the EED is about 50 ohms. For pin-to-pin tests the insertion unit contains a 50 ohm resistor in series. For pin-to-case tests a 55 ohm parallel resistor is usually used. In general the choice of impedance insures that the system behaves as if line 2 is open circuited at the R_2 end and loaded with 50 ohms at the relay end. This choice results in a theoretically rectangular voltage across the input impedance to the insertion unit. The aptitude of the pulse if theoretically one-half the DC charging voltage and its duration is twice the one way delay time of line 2.

6) The damped rectangular stimulus is coupled through a teflon filled coaxial cable, through the T&M monitor, through the teflon filled coaxial firing mount and thus to the EED.

7) The voltage and current for the EED are monitored by the T&M monitor and coupled to the oscilloscope by fairly short coaxial cables. The traces of the scopes are photographed on high speed film.

Note that the block diagram shows several components in an oil bath. The oil used is standard transformer oil. It eliminates corona discharge, and its resulting interference, from the system. The metal cans shown in the block diagram are constructed to completely confine the electromagnetic noise due to the arc at the switch. Note that the only connections between the shielded volume containing the arc and the outside volume containing the oscilloscopes is through the 300 feet of RG 14/11, thourgh the 300 feet of shielded Twinax or though the voltage and current monitor cables. Noise on the monitor cables is a real perturbation to our measured signals and it will be displayed on the scope as it should be. The other type of noise that would normally affect our oscilloscopes, due to direct coupling between the arc and the scopes, does not show on our photographs because it must, in our system, propagate at least the length of the RG 14/11 and/or the Twinax before it can interact with the scope. This takes enough time that our photography is finished and the scopes have finished their sweep long before the noise arrives. We thus get clean pictures of the actual voltage and current across the load.

The two metal shielding cans shown in Figure 5 differ in construction. The can that contains R_2 is a modified one gallon paint can. The shield line 1 is soldered 360 degrees to the bottom of the can which it penetrates. The lid of the can is penetrated by line 2 and this shield is also soldered 360 degrees. In use R_2 is soldered in place with the lid almost closed, the can is filled with oil, the lid closed and the can submerged in an oil bath. The can containing relay K1 is a modified 50 caliber ammunition can. The can has both input and output connectors. They are Teflon insulated, General Radio, 50 ohm connectors. RF gasketing material has been applied to the lid/body mating line. The overcentered closure device thus makes a good RF joint when the can is closed. The shield of the twin lead relay control cable is 360 degree soldered where it penetrates the can. Line 2 is dressed with a Teflon insulated connector on the end opposite the paint can lid. This connector mates with the input connector on the ammo can. In use the can with its input connector is submerged in an oil bath. The output connector is above the oil level.

Note that line 2 is constructed with a paint can lid soldered at one end and a connector at the other end. The length of this line controls the pulse length of the overall system. We have constructed lines giving 25, 50 and 100 nanosecond pulses and with the arrangement described above we can substitute one for another in about five minutes.

The T&M monitor contains a shunt 506 ohm voltage divider and a 0.051 ohm series resistor for current monitoring. The voltage division ratio is 98.6:1 and the output impedance is 50 ohms. The voltage output of the T&M monitor is coupled through six feet of 50 ohm line to 23 dB of General Radio 50 ohm pads. These pads are in turn connected through five feet of 50 ohm cable to a shunt 50 ohm load at the input to the voltage monitoring 454 Tektronix oscilloscope. The current monitoring output of the T&M monitor (which is across the 0.051 series resistor) is connected by 24 feet of 50 ohm cable to the input of the current monitoring scope. This input is also shunted by a 50 ohm load. The current monitoring scope is externally triggered by the voltage monitoring scope's sweep gate signal through a six foot 50 ohm cable. The voltage monitoring scope is internally triggered by the voltage input. The cable lengths are critical (within a few feet) for proper time relation of the voltage and current.

The 98.6:1 voltage divider and the 23 dB pad result in an equivalent deflection factor for the voltage monitoring scope of:

Stimulus Volts/Division	Scope Volts/Division
1393	1
2786	2
6964	5

The 0.051 current monitoring series resistor results in an equivalent deflection factor for the current monitoring scope of

Stimulus Amps/Division	Scope Volts/Division
0.98	0.05
1.96	0.1
3.92	0.2
9.8	0.5
19.6	1.0
39.2	2.0
98.0	5.0
196.0	10.0

Since the reading of the oscilloscope pictures can seldom be done to better than 5% accuracy the calibration factors above are rounded off in the quotation of scope deflection factors. Thus we give 40 amperes per division for an actual 39.2 ampere per division factor and we quote 2800 volts per division for an actual factor of 2786 volts per division, etc.

The scopes are calibrated against their internal voltage standards and the system monitoring equipment is calibrated (or checked) by connecting a 50 ohm load on the end of the T&M monitor and observing the magnitude of voltage and current indicated. During this check the horizontal positioning controls of the scopes are adjusted so the pulses begin at the same place on the scope faces. This facilitates comparison of the voltage and current photographs.

Considerable short rectangular pulse testing has been done on the Atlas Squib Switch. Figure 6 shows a cutaway view of the overall switch and Figure 7 shows a disassembled switch. Most of our testing was done on the "plug" alone. This subassembly is indicated in Figure 6.

4. SOME RESULTS

Figure 8 is a representative set of traces of a 25 nanosecond pin-to-pin Bruceton test for the Atlas Squib Switch plug. If we assume that the bridgewire resistance is constant during the test (and it seems likely from study of the voltage trace in Figure 8) we can calculate firing energies. Figure 9 compares the mean firing energies for various other Brucetons with that determined by our 25 ns test. Note that the means all roughly compare. This result points to the fact that the pin-to-pin response to short pulses is in keeping with the theory applicable to longer pulses and that no new phenomena are evidenced in this shorter time regime.

Figures 10 and 11 show double exposure results of high voltage pin-to-case short pulse tests on the Atlas Squib Switch. Both photos show a normal "no-fire" response with very small currents and a "fire" response that clearly shows breakdown and a large pin-to-case current. A considerable number of such photos shows that pin-to-case initiation is a phenomena that takes

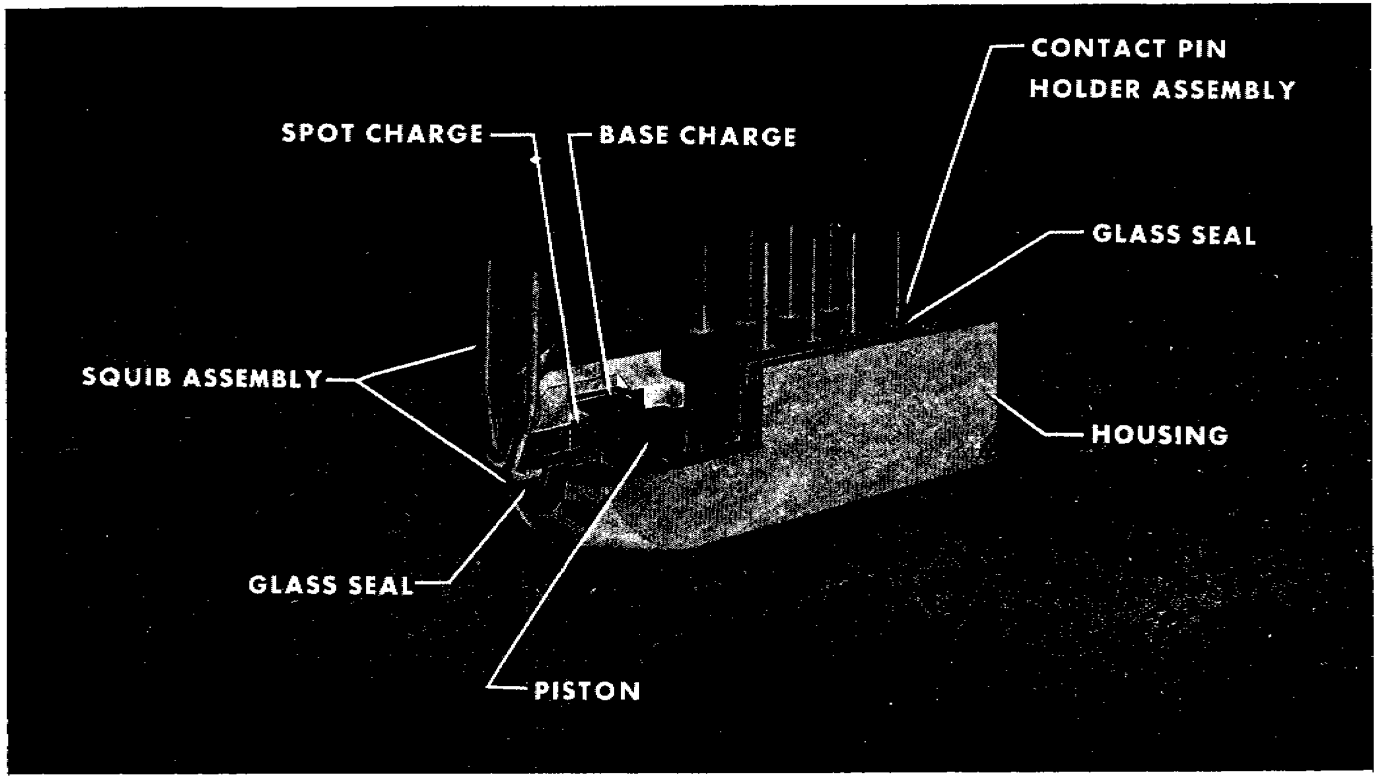


Figure 6. Switch, Squib Actuated Non Delay

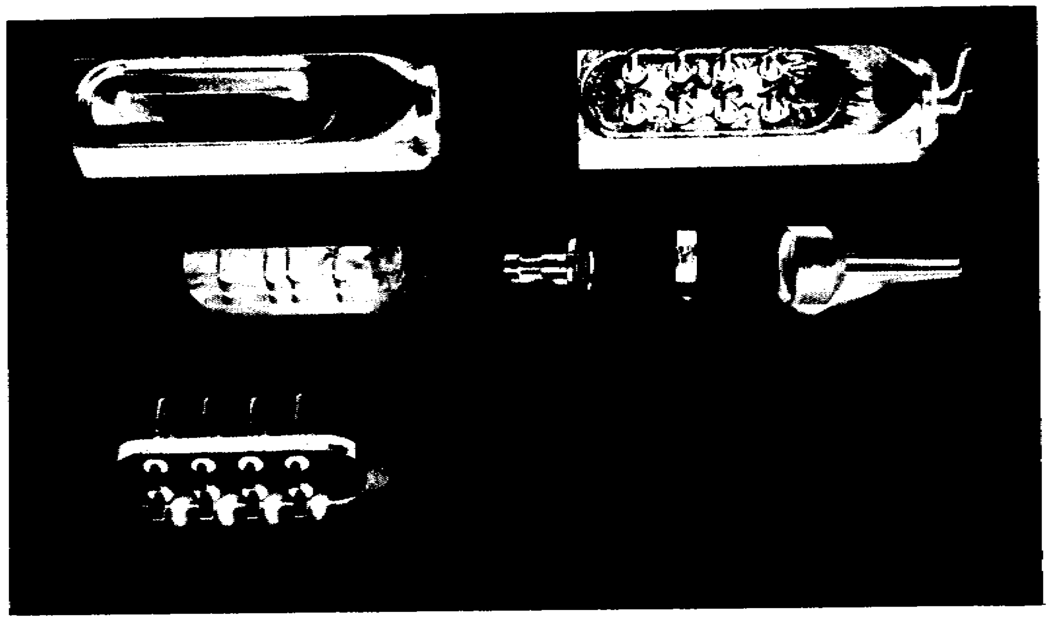


Figure 7. Disassembled Switch

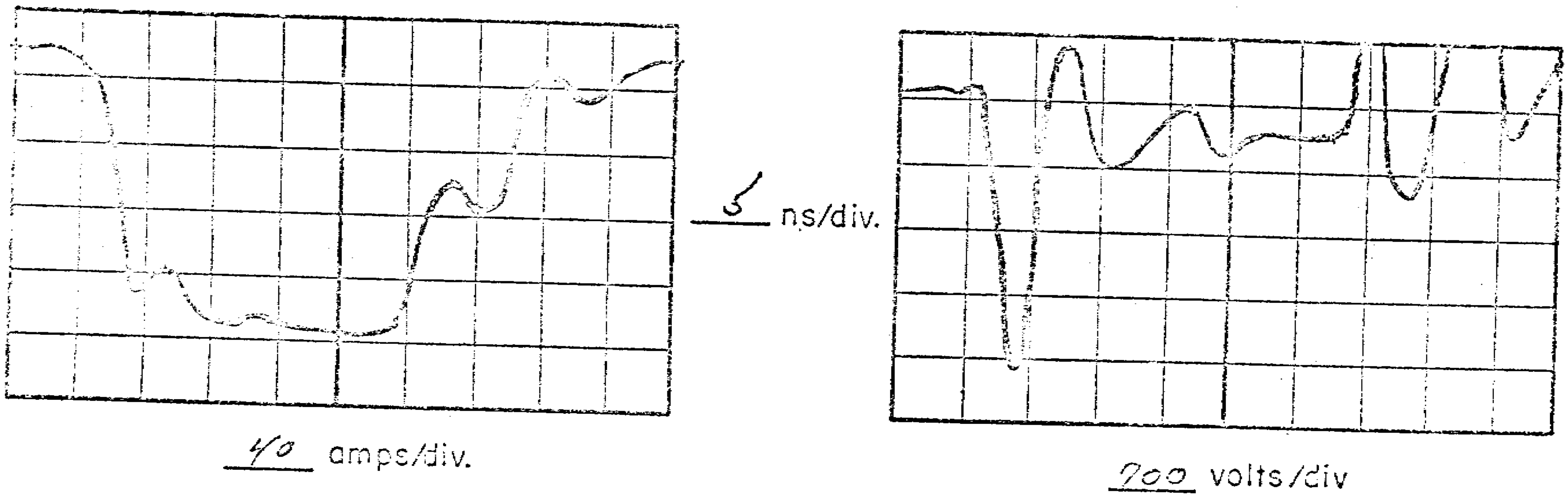


Figure 8. Typical Traces From the Pin-to-Pin Bruceton

SUMMARY OF SENSITIVITY DATA
-PIN-TO-PIN MODE-

Probability Level (95% Confidence) (%)	Pulse Time 25 n sec Energy (erg)	Pulse Time 50 n sec Energy (erg)	Pulse Time 100 n sec Energy (erg)	Pulse Time 270 n sec Energy (erg)	Pulse Time 1 micro sec Energy (erg)	Pulse Time 10 micro sec Energy (erg)
99.9	76,000	30000	22000	18300	17000	20100
50	13,000	16400	13500	13500	13300	15000
0.1	2,300	8900	8300	9900	10400	11100

NOTE: ALL ENERGIES COMPUTED USING A RESISTANCE OF 1.83 OHMS

Figure 9. Energies Computed Using a Resistance of 1.83 Ohms

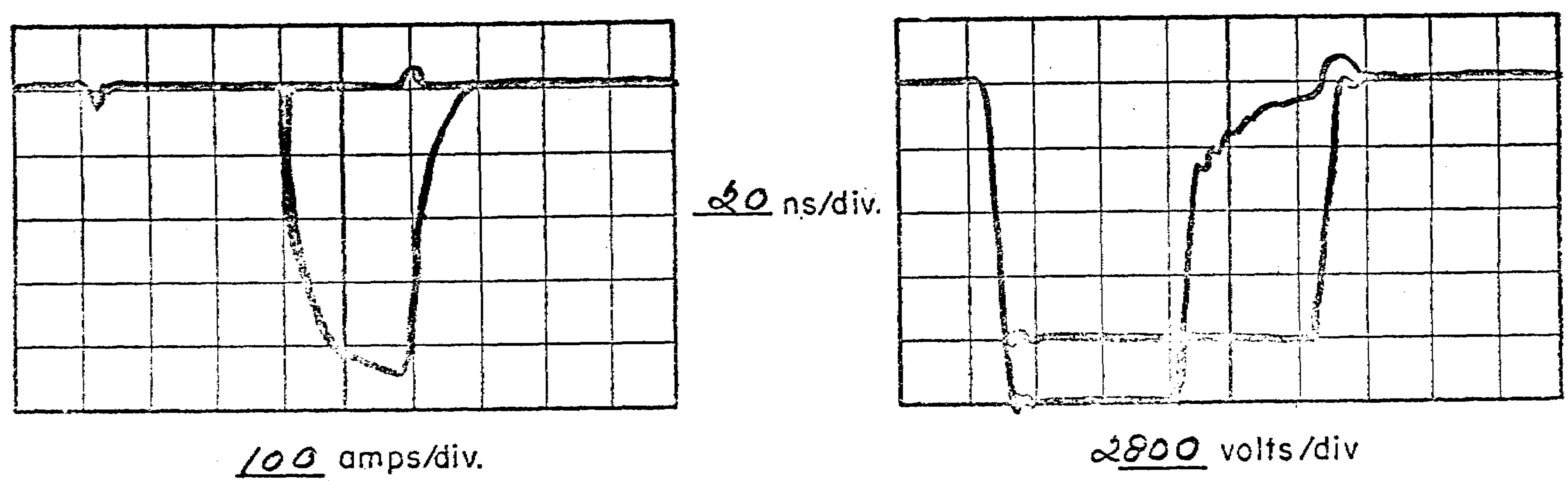


Figure 10. Typical Pin-to-Case Traces Showing Both Breakdown and No-Breakdown Responses

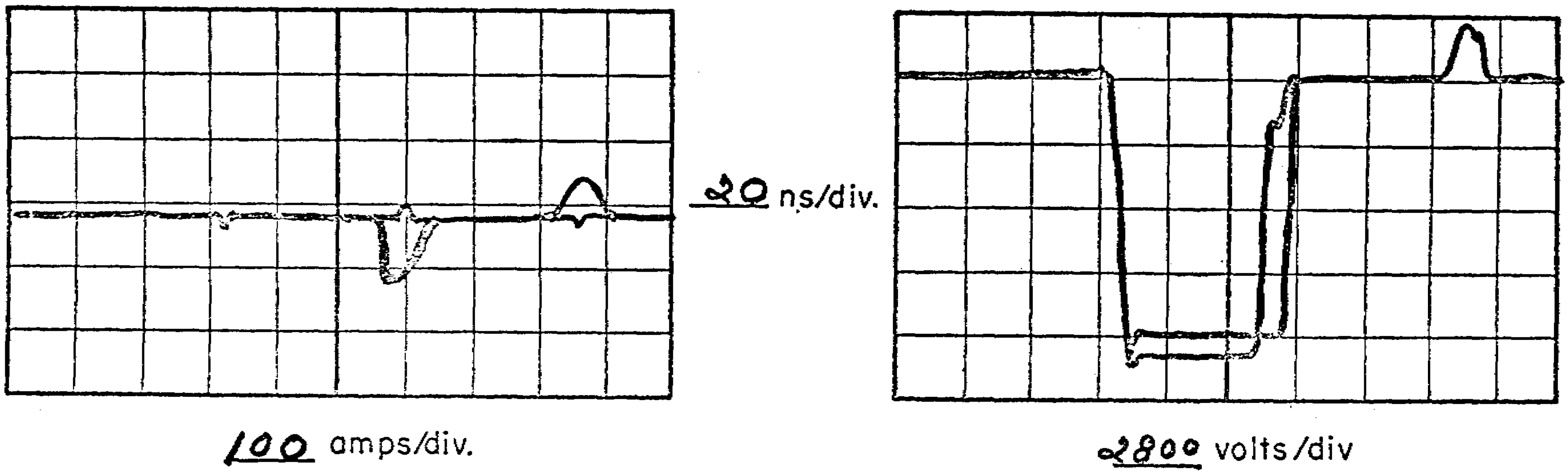


Figure 11. Pin-to-Case Traces Showing Low Energy Fire

place on a nanosecond time scale. Several determinations of the energy delivered to the breakdown impedance have been made. The minimum energy determined to date is about 25,000 ergs but more work is obviously necessary to determine the minimum energy necessary for initiation. We feel that continued work in this area could lead to a much better understanding of overall pin-to-case firing characteristics of electroexplosives.

5. THE BRUCETON AND THE PROBLEM

Figure 12 shows the computer output from a not atypical Bruceton* calculation using the short pulse data. Note that we compute the Bruceton statistics using both the "fires" and the "no fires" and then average the results. The confidence levels are computed using the average data. The computer program is written in Fortran IV and can produce 95% or 90% confidence levels at choice. It can also produce the same sort of output as shown in Figure 12 for the 1%, 99% levels and the 10%, 90% levels. Other levels can of course be obtained but we would need to change the program cards slightly.

The most common use of the no-fire (i.e. 0.1%) probability levels as determined by the Bruceton test procedure is as an absolute safety level. For example many radio frequency/EED safety analyses use the 0.1% power level (with 95% conf.) of a radio frequency Bruceton test as the absolute maximum of power that can be coupled to the EED and have the overall system considered safe. Other analyses add a safety factor by dividing the no fire level by two or ten. In any event a sensitivity test that predicts a no-fire level that is consistently lower than the actual tends to err on the conservative or safe side.

Our end uses of the Bruceton results mentioned in this report were for the estimation of "no-fire" levels. We have used the 0.1% level (with

*The "Bruceton" test we refer to is a test of the type described in "Statistical Analysis for a New Procedure in Sensitivity Experiments," a report submitted by the Statistical Research Group, Princeton University, to the National Defense Research Committee July 1944.

95% confidence) as this no-five level.

Our primary questions about the application of the Bruceton in this manner are two:

1) How does the error in the determination of the test levels influence the Bruceton results?

2) Is the Bruceton test procedure a useful tool in this application or are there other more "optimum" techniques?

```
***B R U C E T O N   A N A L Y S I S***

DATA,RF2410 P-P25NANASEC,PULSE DC+C,PLUG C3410-01 5-16-73

SER. RES.  FUNCT.     L E V E L S          LEVEL  STIMULUS                                  V A L I D I T Y  T E S T S
NO. (OHMS) TIMES.
           ( SEC)   1 2 3 4 5 6 7 8 9 10   NO.    (AMPS )    I   I*I    NO    NX
            1                         X
            2                       0       1     .1580+03   0    0     1     0          EQUALITY OF
            3                         X
            4                       X       2     .1617+03   1    1     1     1          OCCURRENCE -OX
            5                     X
            6                 0             3     .1654+03   2    4     2     1
            7                     X
            8                   X           4     .1693+03   3    9     5     2          NO. OF RUNS= 16
            9                   X
           10                 X             5     .1732+03   4   16     2     5
           11               X
           12             X                 6     .1773+03   5   25     1     3          LENGTH OF
           13           0
           14             0                 7     .1814+03   6   36     0     2                 RUNS= 6
           15               0
           16                 0             8     .1856+03   7   49     1     1
           17                   X
           18                 X             9     .1900+03   8   64     0     2
           19               0
           20                 0            10     .0000      9   81     0     0
           21                   X
           22                 0                                         NO=13  NX=17
           23                   0
           24                     X        LOG OF FIRST LEVEL= 2.19866  D=  .010
           25                   0
           26                     X        AO=   40          AX=   79
           27                   X
           28                 0            BO=  160          BX=  427
           29                   X
           30                 0            MO= 2.84024       MX= 3.52249

                                           MEANO= 2.23443  MEANX= 2.24013

                                           SIGMO=  .04838  SIGMX=  .05930  SIGMA=  .05484

                                           S= 5.48385        G=  .900      G*G=   .810000

                                           H= 2.735          H*H=  7.481929

                                           CONFIDENCE   INTERVAL=  .21131

                                           LOG OF 99.9%(95%CONF)= 2.61842    99.9%(95%CONF)= 415.356 AMPS

                                           LOG OF MEAN(50%)LEVEL= 2.23766          MEAN(50%)= 172.846 AMPS

                                           LOG OF 0.1%(95%CONF) = 1.85690    0.1%(95%CONF)=  71.928 AMPS
```

Figure 12. Pin-To-Pin Bruceton Results

TARGET VISIBILITY AND DECISION OPTIMIZATION

Timothy M. Small
Countermine/Counter Intrusion Department
U.S. Army Mobility Equipment Research and Development Center
Fort Belvoir, Virginia

ABSTRACT. The detection optimization problem can be reduced to simple terms, even though the specific technology used may be sophisticated. Basically, the target's response signal and its contrast to background is to be maximized and it should provide as unique a signature as possible. The first two experimental conditions optimize target visibility; the latter minimizes false signals. In other words, the objective is to optimize the measure of a detection system's performance - the detection efficiency and target specificity.

This paper provides a generalized analysis of detection efficiency optimization in a system which measures a spectral response of a target in a consistent background. The spectral response is assumed to be a Gaussian shaped enhancement mixed within a uniform background. The analysis related signal intensity, signal to background ratio, background determination error, efficiency and specificity. Optimization of the signal window width and decision threshold are constraining conditions. Calculated results and the relative sensitivity of each of these parameters will be presented. This analysis provides the detection system designer with the information needed to specify critical parameters or to predict the performance of a given system.

1. INTRODUCTION. Another title for this paper could be "A Pedestrian's Approach to Detection Theory". Being neither a statistician nor familiar with the state-of-the-art in detection theory, I have developed an approach to detection optimization in a way which makes sense to me and would like to share with you.

The approach to target detection which we are going to consider is intended to be practical in nature and is in some respects model dependent. For clarity's sake only a simple but somewhat generalized case will be described. Many of the constraining assumptions may be easily changed to better approximate a particular technique with relatively simple modifications to the analysis.

The essence of target detection is the accurate determination of whether or not a target is within the field of view of a detection system - often under conditions in which the signal intensity is quite limited. This entails tailoring the detection system so as to maximize the visibility of the target and providing decision logic which maximizes the probability of correctly cueing on a target while minimizing the introduction of false cues. This in turn assumes that the detection system provides a binary response. A cue is given if a target-like signature is detected, otherwise no cue is given. We assume that this decision must be rendered in a single pass over the target and based entirely on the response of a single detector.

In actual detection scenarios an object which simulates a target may be encountered, thereby inducing a false signal. Because the description of this class of false signals is closely linked to the specific technique used, it will not be further considered here, but rather assumed to be an isolated problem to be considered by the detector designer. Another source of false signals is spurious cueing due to statistical fluctuations in the detector response. We assume this to be the predominant consideration in optimizing detection specificity.

The model we will consider is that of a target which emits either spontaneously or by stimulation a characteristic spectral signature, which in turn is measured in one pass by the detector. The reconstructed spectrum may look something like Figure 1. It would be composed of a background and a signal whose centroid is at a known position in the spectrum. The target emission or detector response is assumed to be quantized or quantizable, thus the spectral intensity on this graph is defined as the number of counts, N, corresponding to the detected quanta, per interval in the spectral parameter, x, such as counts per energy, wavelength, or time of arrival. Quantized data is particularly convenient for computer processing, which this analysis lends itself to. We will assume that both signal and background count fluctuations within an interval obey Gaussian statistics with one standard deviation corresponding to the square root of the number of counts within the interval. Such Poisson fluctuations are adequately approximated by Gaussian statistics if the number of counts is large. In addition, the background shape is assumed predictable so that the background under the signal can be inferred by normalizing the counts in a pure background portion of the spectrum. For simplicity the background is assumed flat in the region of the signal so that the signal shape is not distorted. The signal is assumed to have a Gaussian shape. This is a good approximation in many cases, particularly if the original signal width is much less than the detection system resolution, so that it is smeared into the Gaussian shape.

2. <u>MODEL SPECIFICATION</u>. Figure 2 illustrates some of the basic parameters we will need. The resolution of many systems is defined in terms of the full width at half maximum. For a Gaussian, one standard deviation is slightly narrower than the half width at half maximum and a half-bin width parameter, x_B, remains to be calculated. This parameter defines the upper and lower bounds of the optimal signal window. The signal which falls within this window is called S. This window correspondingly defines the amount of background which is collected, B. S_o and B_o are

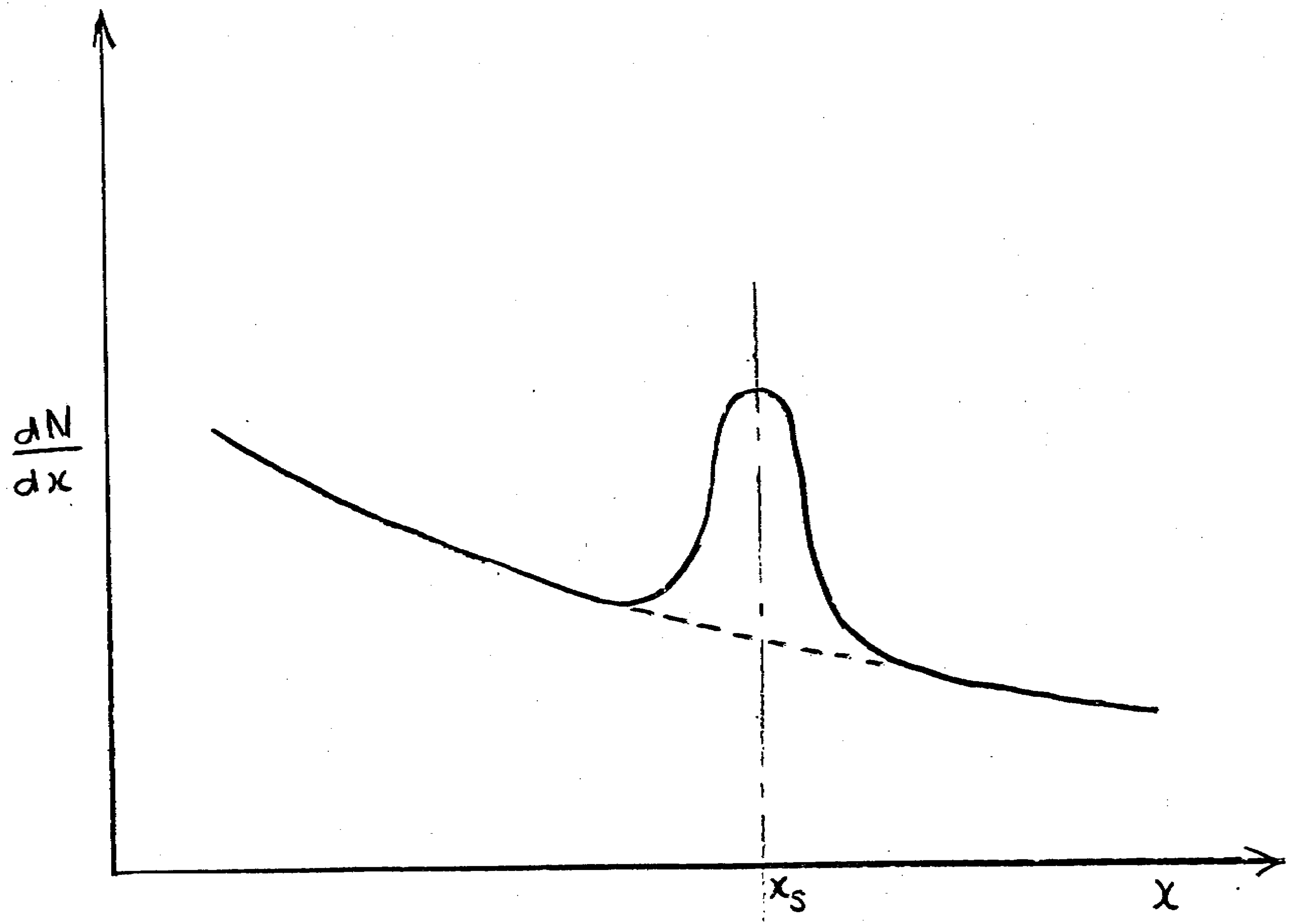


Figure 1

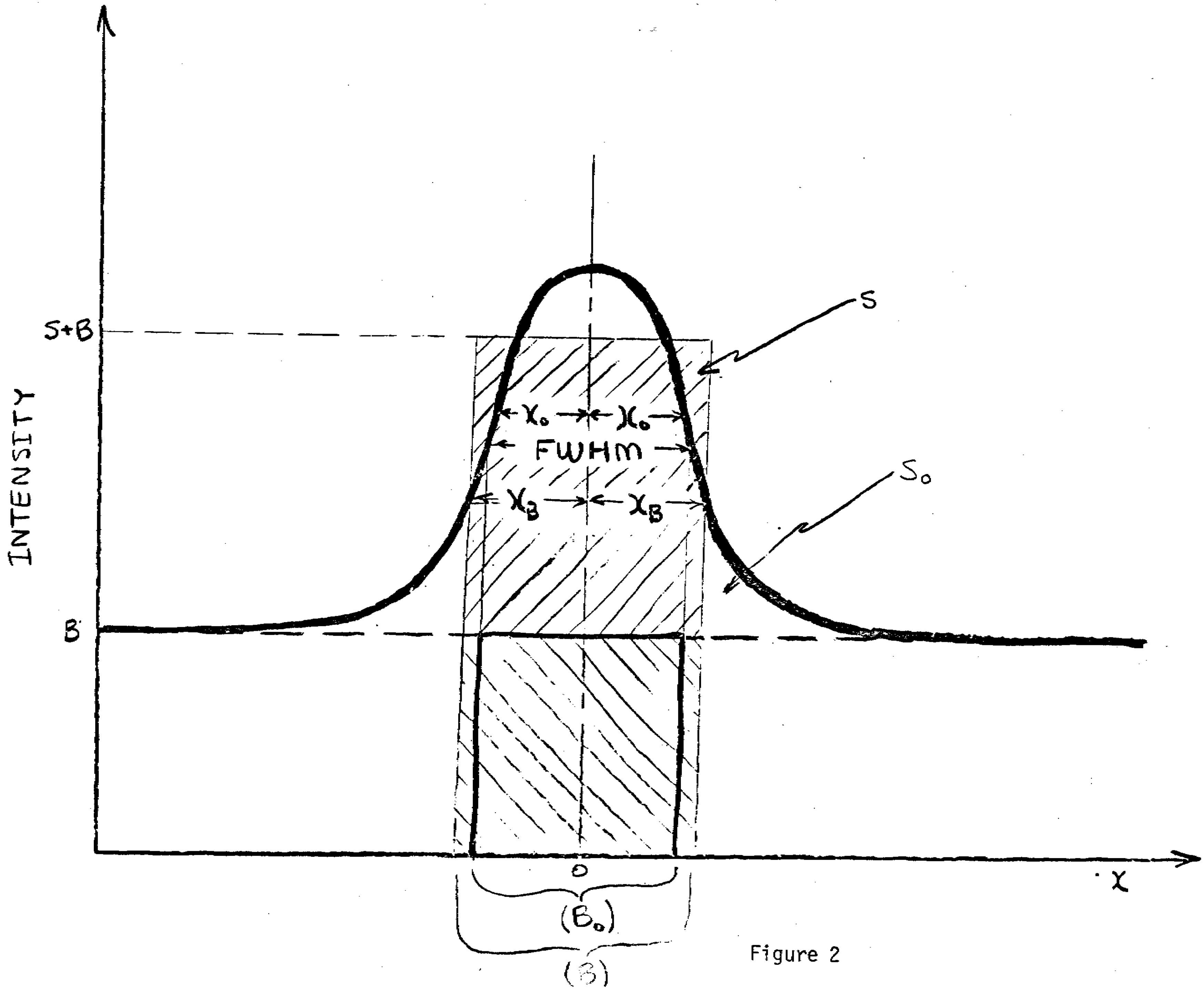


Figure 2

internally calculated numbers, since they are a function of x_B, a calculated parameter. In order to externally specify the signal and background we should select experimentally meaningful definitions. S_o accordingly is defined as the total signal counts in the entire Gaussian, and B_o is the background counts within the FWHM interval of the signal.

The detection system is designed to cue when the total counts within the window exceeds a certain threshold. The threshold should be as much above the background as possible for good specificity - to avoid cueing on statistical fluctuations - and below the total signal plus background for good detection efficiency. We can write this

$$T_1 = B + \sigma_s \sqrt{B}$$

$$T_2 = S + B - \sigma_e \sqrt{S + B}$$

where σ_s is the number of standard deviations above the background and σ_e is the corresponding parameter below the signal plus background. Note that the σ's are the number of standard deviations and not the standard deviation itself.

The error on the background is more than just the statistical error. There is also a determination error, since B must be inferred. Thus the complete expressions for a standard deviation are

$$T_1 = B + \sigma_s \sqrt{B + \Delta B^2}$$

$$T_2 = S + B - \sigma_e \sqrt{S + B + \Delta B^2}$$

where ΔB is the determination error. In actuality $T_1 = T_2$ which is the detection threshold, T, and it is optimized when both σ's are maximized.

The specific detection problem determines the relative importance of efficiency and specificity. This ratio is designated, κ, the signal to background ratio is ρ, and β is the ratio of background measurement error to statistical error.

$$\kappa \equiv \frac{\sigma_e}{\sigma_s}$$

$$\rho \equiv \frac{S}{B}$$

$$\beta \equiv \frac{\Delta B}{\sqrt{B}}$$

Making these substitutions and setting $T_1 = T_2$ we obtain the following expression.

$$\frac{1}{\sigma_S} \equiv \phi(\rho, S) = \frac{\sqrt{1+\beta^2} + \kappa\sqrt{1+\rho+\beta^2}}{\sqrt{\rho S}}$$

Since the σ's are to be maximized the expression ϕ should be minimized.

The signal is written as a function of the spectral parameter x whose origin is assumed to be the signal centroid.

$$\mathcal{S}(x) = \frac{S_o}{x_o\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{x_o}\right)^2}$$

S_o is the total signal counts and as before x is the Gaussian's width expressed as one standard deviation and is approximately 42 percent of the full width at half maximum.

$$x_o = \gamma\,(\text{FWHM}),\ \gamma = (8\,\ell n\, 2)^{-\frac{1}{2}} \approx 0.42$$

To provide generalized dimensionless variables we define $z \equiv x/x_o$ and define A as the definite integral of the Gaussian.

$$A(z) \equiv \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{z^2}{2}}\, dz$$

This expression can be numerically approximated by an inverted polynomial series. The signal can then be written

$$S = \int_{-z_B}^{z_B} \mathcal{S}(z)\, dz = 2S_o\, A\,(z_B),$$

where z_B is the half signal window width. An experimentally convenient way to express the background is as counts per FWHM interval. We have designated this B_o, so the background within the signal window, B, is

$$B = 2\gamma B_o z_B.$$

This reflects the assumption that the background is flat. Thus, the signal to background ratio

$$\rho \equiv \frac{S}{B} = \frac{S_o}{\gamma B_o} \frac{A(z_B)}{z_B}$$

Now with these expressions for S and ρ, ϕ becomes a constant times ϕ', a reduced minimization function.

$$\phi(z) = \frac{1}{\sqrt{2 S_o \rho_o/\gamma}} \phi'(z),$$

where

$$\rho_o \equiv \frac{S_o}{B_o}, \quad \alpha \equiv 1 + \beta^2,$$

$$\phi'(z) = \frac{\sqrt{\alpha z} + \kappa \sqrt{\alpha z + \rho_o A(z)/\gamma}}{A(z)}$$

Note that all variables are dimensionless and that z_B is a function only of the three input parameters, κ, ρ_o, and β.

3. MODEL SOLUTION. The complexity of the expression for ϕ' doesn't lend itself to an analytic solution of its derivitives to determine the z corresponding to its minima. Thus iterative solution by computer is used. A program, GREØP, has been written to run on the MERDC CDC 6600 to calculate z_B. Some of these results are summarized in Figure 3. Here z_B is not displayed - rather z_B/γ, where $\gamma \approx 0.42$, is contoured on a ρ_o versus κ plot in which the background is assumed well known. If $\beta > 0$ the optimal width decreases slightly. Note that a window width approximately 20 to 30 percent above the signal FWHM is in many cases optimal. The next question is how sensitive is the choice of z_B? Figure 4 shows that it is relatively insensitive. Changes of 20 percent induce a decrease in the σ's of less than 2 percent. Two extreme conditions among those I have calculated with $\beta = 0$ are shown and are quite similar. If $\beta > 0$ then the trough narrows slightly. Once z_B is determined several other parameters can be calculated.

Fraction of S_o within window = $2 A(z_B)$

Optimal window width as fraction of FWHM = $2z_B/\gamma$

Relative Gaussian amplitude at window boundary = $e^{-\frac{z^2}{2}}$

The actual signal to background ratio within the window is ρ,

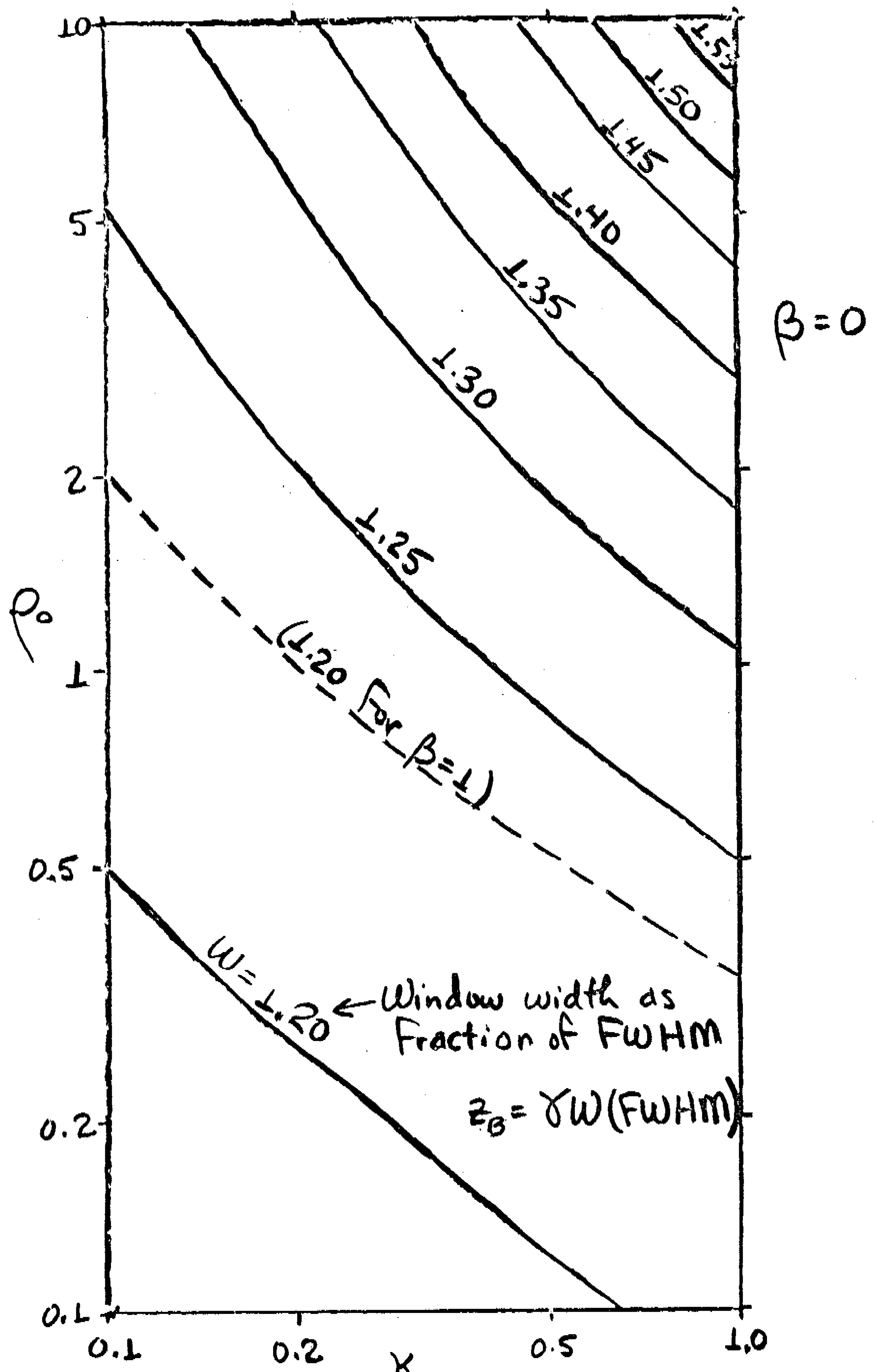

10
5
2
1
0.5
0.2
0.1
ρ_0
0.1
0.2
0.5
1.0
K
1.55
1.50
1.45
1.40
1.35
1.30
1.25
(1.20 for β=1)
W=1.20
← Window width as Fraction of FWHM
$z_B = \gamma W$(FWHM)
β=0

Figure 3

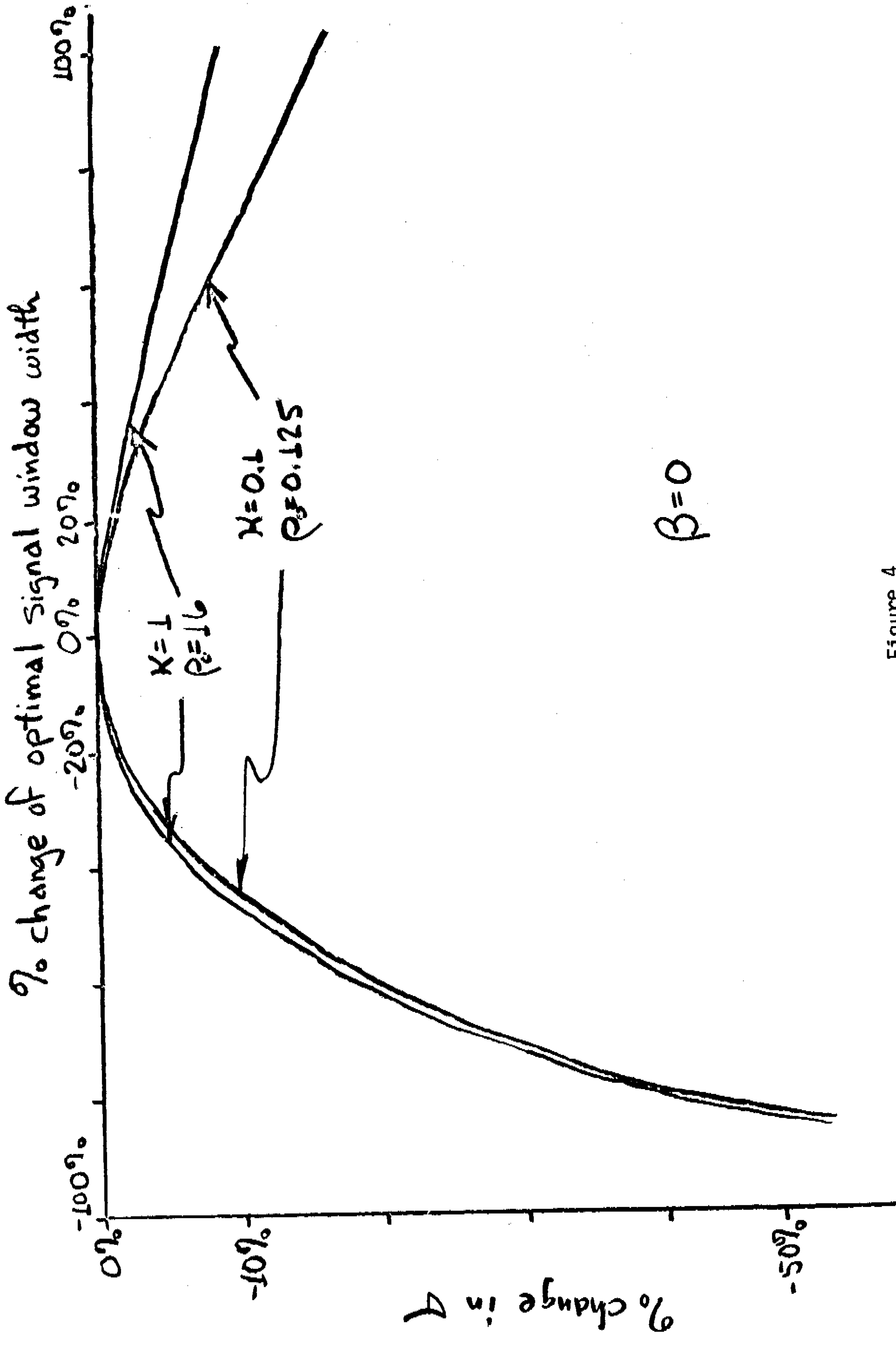

% change of optimal signal window width
-100%
-20%
0%
20%
100%
K=1
K=0.1
ρ0=0.125
β=0
0%
-10%
-50%
% change in

Figure 4

$$\rho = \frac{\rho_o}{\gamma} \frac{A(z_B)}{z_B}$$

The optimal threshold to background ratio, τ, also follows after a little artthmetic.

$$\tau \equiv \frac{T}{B} = \frac{\rho + 1 + \kappa \sqrt{1 + \rho/\alpha}}{1 + \kappa \sqrt{1 + \rho/\alpha}}$$

These dimensionless ratios are shown plotted in Figure 5 for the case in which $\beta = 0$. The threshold/background ratio, τ, is plotted versus the measured signal/background ratio, ρ. A family of curves can be drawn for each efficiency/specificity ratio parameter, κ. The input ρ_o is also shown. As β increases the optimal threshold also increases, as expected in order to move it further from the less well known background.

Up to this point all parameters have been dimensionless ratios. If we now specify S_o, B_o, σ_s, or σ_e - also dimensionless but not independent numbers since ρ_o and κ have been specified - then the others can be calculated. For example, given various signal levels, S_o counts, corresponding to expected target emission intensities, the background per FWHM, the σ's and the signal and background within the window follow.

$$B_o = S_o/\rho_o$$

$$\sigma_s = 1/\phi(z_B)$$

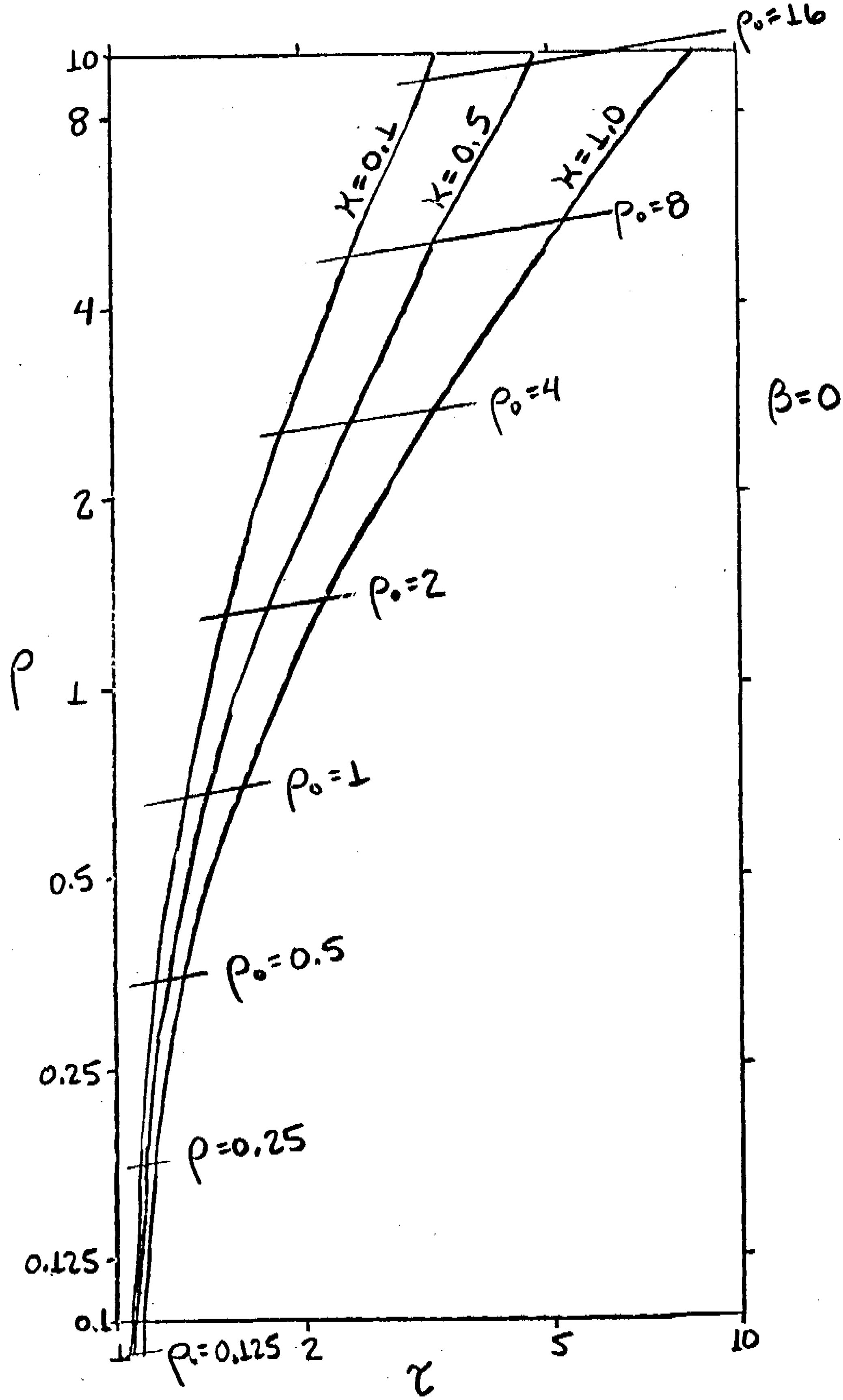
$$\sigma_e = \kappa\sigma_s = \kappa/\phi(z_B)$$

$$S = 2S_o A(z_B)$$

$$B = S/\rho = 2S_o z_B \gamma/\rho_o$$

The optimal threshold, false signal probability due to statistical fluctuations, and the target detection efficiency can now be calculated.

$$T = B + \sigma_s \sqrt{\alpha B} = \tau B$$



$\rho_0=16$
$\kappa=0.1$
$\kappa=0.5$
$\kappa=1.0$
$\rho_0=8$
$\rho_0=4$
$\beta=0$
$\rho_0=2$
$\rho_0=1$
$\rho_0=0.5$
$\rho=0.25$
$\rho_0=0.125$
ρ
10
8
4
2
1
0.5
0.25
0.125
0.1
1
2
5
10
τ

Figure 5

$$\text{False signal probability: 1 per } \frac{1}{1/2 + A(\sigma_s)}$$

$$\text{Target detection efficiency} = 1/2 + A(\sigma_e)$$

4. RESULTS. Figure 6 is a portion of the printout from GREØP for the case in which $\beta = 1/2$, $\kappa = 1/2$, and $\rho_o = 1$. Those variables which were previously assigned a symbol, have that symbol written in here. Note particularly the false signal frequency due to statistical fluctuations and the detection efficiency as they vary with total signal. A signal of 200 counts produces almost 100 percent efficiency and essentially a complete lack of spurious false signals. Such rejection is surely much better than the false signal rate due to target simulation in most scenarios. GREØP generates a four dimensional matrix of conditions spanning some of the most likely values of the input parameters, so that cross comparisons can be performed.

Several related problems have not been considered in this paper, such as the extension of this analysis to non-quantized input and the problem of collecting data continually rather than processing one sample at a time. The price paid to achieve continuous converage is that of not knowing the time domain in which the target is viewed, thus somewhat higher signal count rates are required to yield the same performance as with a static configuration. Other modifications such as multiple targets each with a distinct signature, multiple signals from a single target and non-flat backgrounds are relatively simple extensions of this analysis.

5. GAUSSIAN RESPONSE FUNCTION. To digress a bit, one may think that a weighting function which matched the response function would provide a better window to filter out the signal. That is to say that the σ's would be larger if the window were Gaussian weighted than if weighted by a box function as we have done.

$$S_G = \int_{-\infty}^{\infty} \frac{S_o}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \left(e^{-\frac{z^2}{2}}\right) dz = \frac{S_o}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2} dz$$

$$= \frac{S_o}{\sqrt{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{S_o}{\sqrt{2}}$$

TYPICAL GREØP ØUTPUT

(INPUT) $\beta = 0.5$, $X = 0.5$, $\rho_0 = 1.0$

FRACTION OF SIGNAL WITHIN BOUNDS= 85.38 PERCENT.
WIDTH= 1.249 OF FWHM.
HEIGHT= .335 OF PEAK.
ρ SIGNAL/BACKGROUND WITHIN BOUNDS= .688
τ OPTIMAL THRESHOLD/BACKGROUND= 1.424

S_0 (INPUT) SIGNAL	σ_s SIGMA, SPECIFICITY	σ_e SIGMA, EFFICIENCY	FALSE SIGNALS 1 PER X SAMPLES	DETECTION EFFICIENCY
1.0	.424	.212	2.977	.5639
2.0	.599	.299	3.542	.6177
5.0	.947	.474	5.820	.6621
10.0	1.339	.670	11.082	.7435
20.0	1.894	.947	74.352	.8232
50.0	2.995	1.497	720.148	.9329
100.0	4.235	2.118	869-1.579	.9829
200.0	5.989	2.995	75346397.666	.9986
500.0	9.470	4.735	***************	1.0000
1000.0	13.393	6.695	R	1.0000

S SIGNAL WITHIN BOUNDS	B BACKGROUND WITHIN BOUNDS	ΔB MEASURE ERROR	$\sqrt{B}$ STATISTICAL ERROR	T OPTIMAL THRESHOLD
.859	1.249	.559	1.118	1.778
1.718	2.498	.790	1.580	3.556
4.294	6.244	1.249	2.499	8.891
8.588	12.488	1.767	3.534	17.781
17.176	24.977	2.498	4.998	35.562
42.938	62.442	3.951	7.902	88.906
85.876	124.884	5.588	11.175	177.812
171.753	249.769	7.902	15.804	355.624
429.382	624.421	12.494	24.988	889.060
858.763	1248.843	17.669	35.339	1778.121

Figure 6

$$B_G = \int_{-\infty}^{\infty} \gamma B_o \left(e^{-\frac{z^2}{2}}\right) dz = \sqrt{2\pi}\,\gamma B_o$$

$$\rho_G = \frac{S_G}{B_G} = \frac{1}{2\sqrt{\pi}\,\gamma}\rho_o$$

Notice the Gaussian weight is not normalized to unit area since this would unnecessarily attenuate the signal. A Gaussian weighted Gaussian signal can be seen here to yield a detected signal of 71 percent of the total signal and a signal to background of about 2/3, see Figure 7. Both values are less than that of the previously presented GREØP calculation, 86 percent and .69, respectively. Only when ρ_o and κ are large and β is small, does ρ_G exceed ρ, which drops to .55 under the worst conditions; but under these conditions over 95 percent of the signal is processed. Under all situations calculated the processed signal is larger. Thus if there is a better filtering function than that used, it is not Gaussian in shape.

6. SUMMARY. Much of what I have discussed is related to communication theory. But not being familiar with the intricies and subtilies of this field I have started from scratch, in a detection format, assuming the signal is digital and have kept the approach as convenient for application as possible. The input parameters follow naturally from the detection conditions. For example, β (the relative background determination error) follows from the detector characteristics, κ (the relative importance of efficiency to specificity) from the nature of the detection problem and ρ_o (the signal to background ratio) from the target characteristics. The optimal threshold can then be determined directly since it is proportional to the background with a proportionality constant τ, which can be calculated once β, κ and ρ_o are known.

Another practical example is the determination of the minimum signal intensity,which yields the desired detection characteristics - that is the false signal rate and detection efficiency. Or the inverse in which a given signal intensity is available and it is desired to determine how to best optimize performance. Given the range or parameter values which constrain his system, the designer can trade off one for the other to obtain the best performance with the weakest signal. Thus if nothing else this approach to target visibility and detection decision optimization provides the desigher with an organized approach to improvement of his detection system.

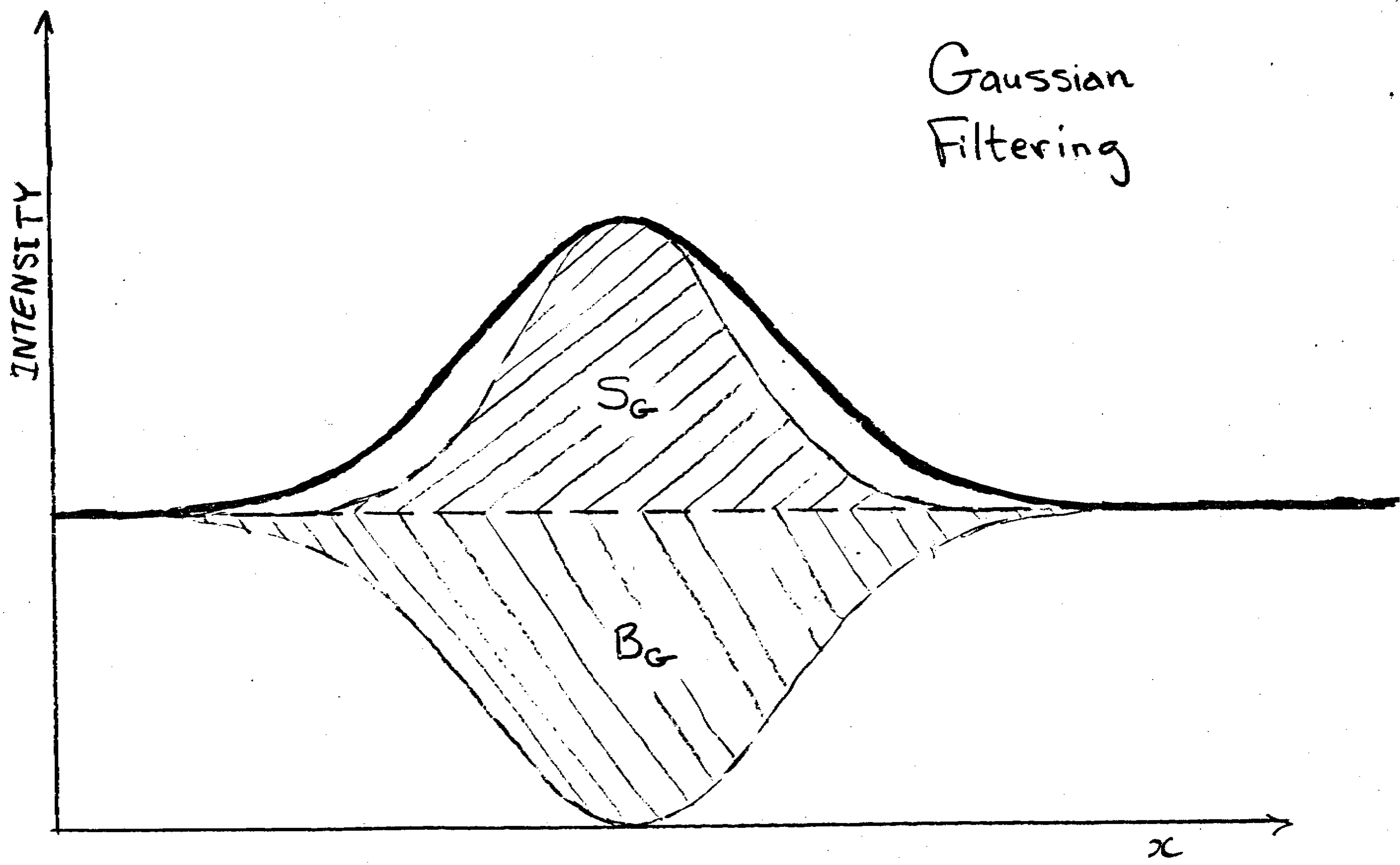


Figure 7

OPTIMIZING A PRODUCTION LINE
FOR COST AND QUANTITY

Eileen M. R. Weigand
Manufacturing Technology Directorate
Small Caliber Ammunition
Modernization and Engineering Division
Frankford Arsenal, Philadelphia, Pennsylvania

ABSTRACT. At various times in the production life of any manufacturing equipment, it is desired to optimize one or more outputs of the production machinery. Such outputs might be cost per round, number of pieces produced per minute or system reliability. Variable, but controllable, factors can influence the measured outputs. In the SCAMP program, the operating RPM of the equipment and the length of the running time between repairs and preventive maintenance may seriously affect the outputs stated above.

A statistically designed experiment will be used to generate mathematical equations. These equations will then be used to predict the effects which operating rate and time between equipment shut-down will have on operating cost and pieces produced.

The results will be used to establish the operating and maintenance policies for the SCAMP production lines. Since each output may be optimized by different values of rates and time, the machinery may well be operated and maintained differently in peacetime than in a state of full mobilization.

Response surface techniques will be utilized to generate the mathematical equations. The equations will then be used to establish the operating and maintenance policies based on the factor to be optimized. Moreover, the response surface equations should then be used throughout the life cycle of the production equipment to monitor the machinery for deviations from the original conditions which may indicate a need for major repair and/or replacement.

1. INTRODUCTION. At the present time, new generation equipment designed to automate the 5.56mm ammunition production line is now entering its final stages of development. The equipment consists of a group of submodules connected in series which, when integrated, shall be known as the "SCAMP" production module.

The Small Caliber Ammunition Modernization Program consists of a Case Submodule which manufactures brass cartridge cases. These cases, in turn, are fed into the Primer Insert Submodule which inserts the primer charge into the base of the brass case. At the same time, the Bullet

Submodule manufactures the copper bullet which is mated to the primed and charged brass case by the Load and Assemble Submodule. The finished rounds of ammunition are then prepared for shipping by the Packaging Submodule. The entire operation is completely automated and computer controlled.

The submodules are composed of high speed rotary turrets linked by a continuous transfer mechanism. The twenty-four station turrets are designed for rapid tool changes and offline repairs. These turrets are also designed to operate at a maximum speed of fifty revolutions per minute. The cartridges can then be manufactured and assembled at rates approaching twelve hundred rounds per minute.

Up to now the submodules have been tested individually to debug the equipment and to fine tune the various steps of the production process. During this time, each of the prototype submodules has displayed its own distinctive operating characteristics, and a data bank has been developed for each one. We know the range of speeds over which the individual submodules can operate, approximately how and when the individual tool stations on each turret will fail, and how frequently the entire submodule must be stopped for repairs and maintenance. Turrets and transfer mechanisms were redesigned to increase the mean time between failures and to reduce the repair and maintenance downtime.

The individual submodules have now reached the stage where they are ready to be linked into a continuous production line. At this point, the thrust of the SCAMP statistical research shifts to the module as an integrated system. In particular, the emphasis of this research shall be to determine how well the equipment meets its design criteria, and then to construct and formalize operating and maintenance routines for the entire production module. Thus, this paper will present one technique which is under consideration to optimize both the cost and the yield of this highly complex, computer controlled automated production line.

2. DEVELOPMENT OF THE STATISTICAL MODEL. It is desired that the production system produce, at least, a daily average of 384,000 acceptable quality pieces at the lowest possible unit cost during an eight hour production day. Several factors of the production module affect both the number of good pieces produced and the unit cost. Among these factors are the modular operating speed, the amount of time the module operates before it is shut down for maintenance, the length of time it takes to perform maintenance, the quality of the input raw materials, lubrication, and the initial settings and adjustments of the tooling. Of these factors, some, such as lubrication, are not easily controlled. Other factors, however, appear to exert the greatest effect on submodule productivity and cost. From previous tests the following appear most likely. They are the operating speed of the equipment and the duration of time the production module runs before it is shut down for maintenance. Fortunately, they are easily controlled. Thus, we shall concentrate on this latter set.

To simplify the problem, the submodules shall be considered from an integrated system perspective only. We are primarily interested in the number of pieces produced and their unit cost when the module is run for a specified period of time at a predetermined operating rate. Therefore, one is interested in determining the predictive responses of (1) cost and (2) quantity for a certain range of values of speed and time. This suggests a multiple polynomial regression approach for the problem resolution since the yield, or response, may be perceived as a function of the controlled variables. (Figure 4) This function is called the response surface. If the function is not known, it can sometimes be satisfactorily approximated within the experimental region by a polynomial in X. Again it is fortunate that we already know the boundary conditions. For the operating speed, thirty and fifty revolutions per minute are the constraints. Moreover, since the production schedule is eight hours, we can establish reasonable boundaries for the operating time of the equipment. Let's say three to seven hours.

Once again prior experience with the machinery has limited our experimental region to the areas over which the module can and will be operated. Thus, we are looking for the optimum response, or yield, within the above constraints. Moreover, the previous data lends one to believe that the response surface will be adequately approximated by a quadratic function. (Figure 5) Therefore, the experimental points chosen for the test are speeds 30, 40 and 50 revolutions per minute (i.e., 720, 960 and 1200 pieces per minute) and times of 3, 5 and 7 hours. For these values of speed and time, we have a 3^2 factorial design or 9 treatment combinations. Coding the minimum values as -1, the maximum as +1, and the mid value as 0, we test for the responses,

Y_i, i = 1,..., 9, at the nine design points, (X_{1j}, X_{2j}), j = 1, 2, 3 (Figure 6). The X matrix for this model is shown in terms of the coded variables in Figure 7. However, the variance-covariance matrix for this model is not diagonal.

Since a diagonal variance-covariance matrix is desired to eliminate any interaction between the coefficients of the equation, standard techniques suggest we rewrite the response surface equation as shown in Figure 8 to obtain a new X matrix. In this model $X^2{}_{mi}$ is the mean of the squared coded variables for i = 1,2. Now the variance-covariance matrix, shown in Figure 9, is diagonal and we can determine the least squares estimates for the values of B by the matrix equation given at the bottom of Figure 6. These values of B will give us the response surface equation. Knowing the estimates for the regression coefficients, a regression analysis of variance is performed to determine the accuracy of the response surface equations for (1) pieces produced and (2) cost.

3. <u>ANALYSIS OF THE FITTED SURFACE</u>. Now, after we have determined the proper second order response functions, we are prepared to analyze the fitted surfaces. The maximum point, if it exists, will be the set of conditions such that the first partial derivatives are simultaneously equal to zero. This set of conditions is called the stationary point.

This point, however, is not necessarily that which maximized the response. In fact this point can be either a maximum, a minimum, or a saddle point as illustrated in Figure 10. In addition, there may not be a point at all, but some type of a ridge which may be classified as rising, falling or stationary. The determination of the nature of the stationary point, and the entire response surface, is the ultimate goal of the experiment.

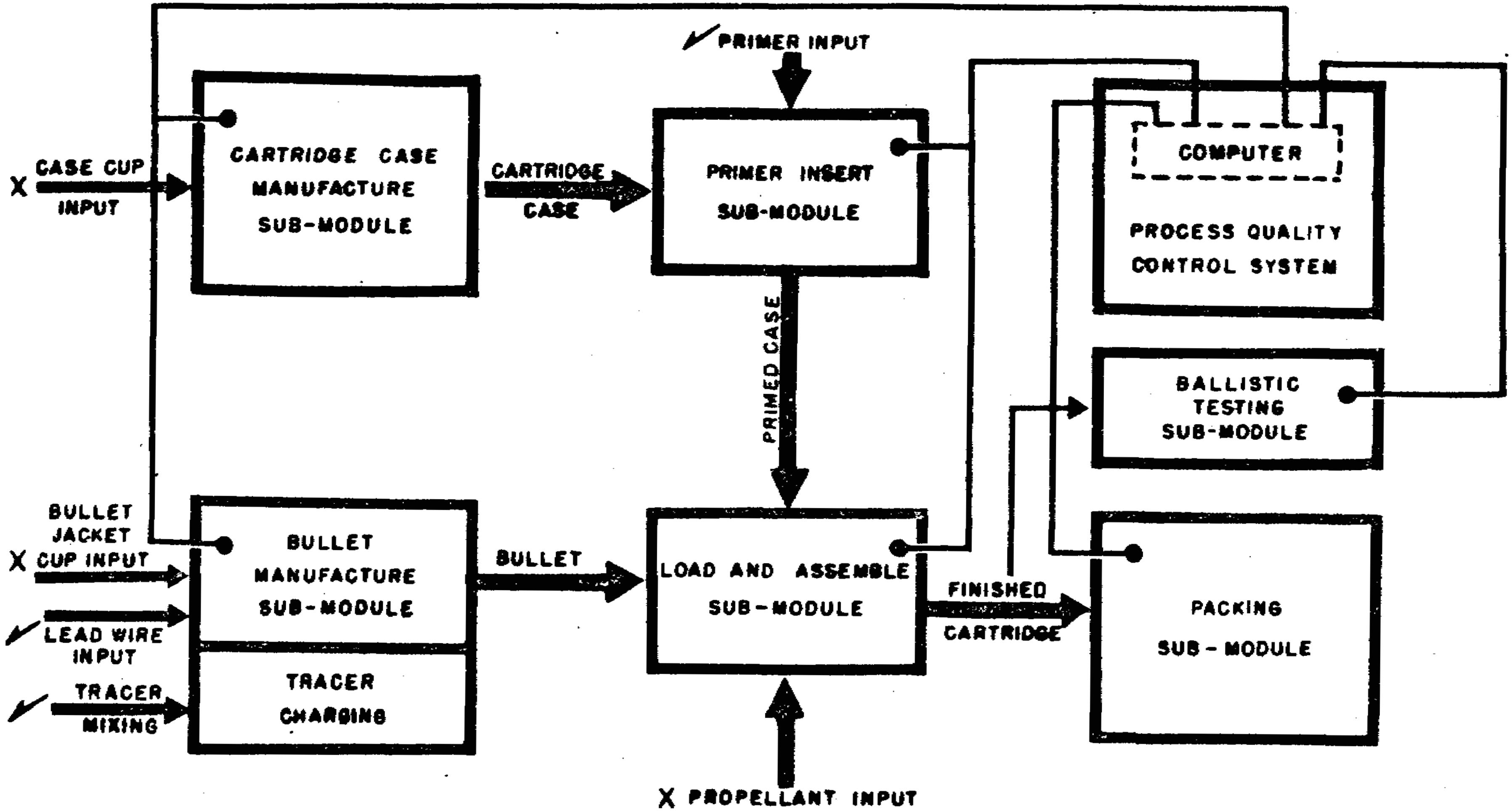
The analysis begins with a translation of the response surface to the stationary point, X. Then the response function is expressed in terms of new variables Z_1 and Z_2 . This corresponds to a rotation of the axes to correspond to the principle axes of the contour system. The form of the function in terms of the Z variables is called its cononical form. Now, by moving along the new axes one can see the quickest direction to travel to find the maximum or the minimum responses.

4. OUTLOOK. Slippages in schedules, due to equipment debugging have delayed the integration tests needed to obtain data for this experiment. But looking at the preliminary data generated by the submodules we can theorize that the response surface for the number of pieces produced will be a rising ridge. For the cost surface, we expect some type of a basin. Thus, the conditions under which we will operate the equipment will then be somewhere in that space enclosed by the two response surfaces. If the two surfaces do not intersect, or in times of national emergency, then we must choose whether we want to operate for maximum yield or minimum cost.

Positive results from the operating equipment will be obtained in approximately six months. But we are very confident that these techniques will give us our desired results, in the least amount of time and for minimum experimental costs. The procedures are well-defined and, as soon as the equipment becomes available, we can generate our two equations to predict the number of acceptable pieces produced and the unit cost. These equations will then be used to establish the optimum operating and maintenance policies for the SCAMP production line.

FRANKFORD ARSENAL • PHILADELPHIA, PA.

SMALL CALIBER AMMUNITION PRODUCTION MODULE



● INSPECT AND CONTROL INFORMATION

✓ PRODUCED IN PLANT

X PURCHASED MATERIAL

DEFINITIONS

SUBMODULE: The equipment necessary in the production or processing of major components such as the cartridge case.

MODULE: One integrated series of submodules with a common production rate and capacity, coupled together by an automated component transfer and process quality

CRITERIA

- MAXIMUM NUMBER OF PIECES
- MINIMUM COST
- EIGHT HOURS

Figure 2

FACTORS AFFECTING RESPONSES

- OPERATING SPEED
- FREQUENCY OF REPAIR

Figure 3

$$y_r = f(x_{1r}, x_{2r}, \ldots, x_{kr}) + e_r$$

k = NUMBER OF INDEPENDENT VARIABLES

r = NUMBER OF RESPONSES

Figure 4

$$y_r = b_0 x_0 + b_1 x_{1r} + b_2 x_{2r} + b_{11} x_{1r}^2 + b_{22} x_{2r}^2 + b_{12} x_{1r} x_{2r}$$

x_1 = SPEED VARIABLE

x_2 = TIME VARIABLE

x_0 = DUMMY VARIABLE FOR COMPUTATION

Figure 5

$$\eta = \gamma_0 + \gamma_1 \theta_1 + \gamma_2 \theta_2 + \gamma_{11} \theta_1^2 + \gamma_{22} \theta_2^2 + \gamma_{12} \theta_1 \theta_2$$

$$D = \begin{bmatrix} x_1 & x_2 \\ -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix}$$

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

Figure 6

THE DESIGN POINTS ARE:

x_1			
	(1, -1)	(1, 0)	(1, 1)
	(0, -1)	(0, 0)	(0, 1)
	(-1, -1)	(-1, 0)	(-1, 1)
		x_2	

IN TERMS OF THE CODED VARIABLES,

$$
X = \begin{bmatrix}
x_0 & x_1 & x_2 & x_1^2 & x_2^2 & x_1 x_2 \\
1 & -1 & -1 & 1 & 1 & 1 \\
1 & -1 & 0 & 1 & 0 & 0 \\
1 & -1 & +1 & 1 & 1 & -1 \\
1 & 0 & -1 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & -1 & 1 & 1 & -1 \\
1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
$$

Figure 7

$$n = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} (X_1^2 - X_{m1}^2) + \beta_{22} (X_2^2 - X_{m2}^2) + \beta_{12} X_1 X_2$$

$$X = \begin{bmatrix} x_0 & x_1 & x_2 & x_1^2 - x_{m1}^2 & x_2^2 - x_{m2}^2 & x_1 x_2 \\ 1 & -1 & -1 & 1/3 & 1/3 & 1 \\ 1 & -1 & 0 & 1/3 & -2/3 & 0 \\ 1 & -1 & 1 & 1/3 & 1/3 & -1 \\ 1 & 0 & -1 & -2/3 & 1/3 & 0 \\ 1 & 0 & 0 & -2/3 & -2/3 & 0 \\ 1 & 0 & 1 & -2/3 & 1/3 & 0 \\ 1 & 1 & -1 & 1/3 & 1/3 & -1 \\ 1 & 1 & 0 & 1/3 & -2/3 & 0 \\ 1 & 1 & 1 & 1/3 & 1/3 & 1 \end{bmatrix}$$

Figure 8

$$X'X = \begin{bmatrix} 9 & 0 & 0 & 0 & 0 & 0 \\ & 6 & 0 & 0 & 0 & 0 \\ & & 6 & 0 & 0 & 0 \\ & & & 2 & 0 & 0 \\ & & & & 2 & 0 \\ & & & & & 4 \end{bmatrix}$$

Figure 9

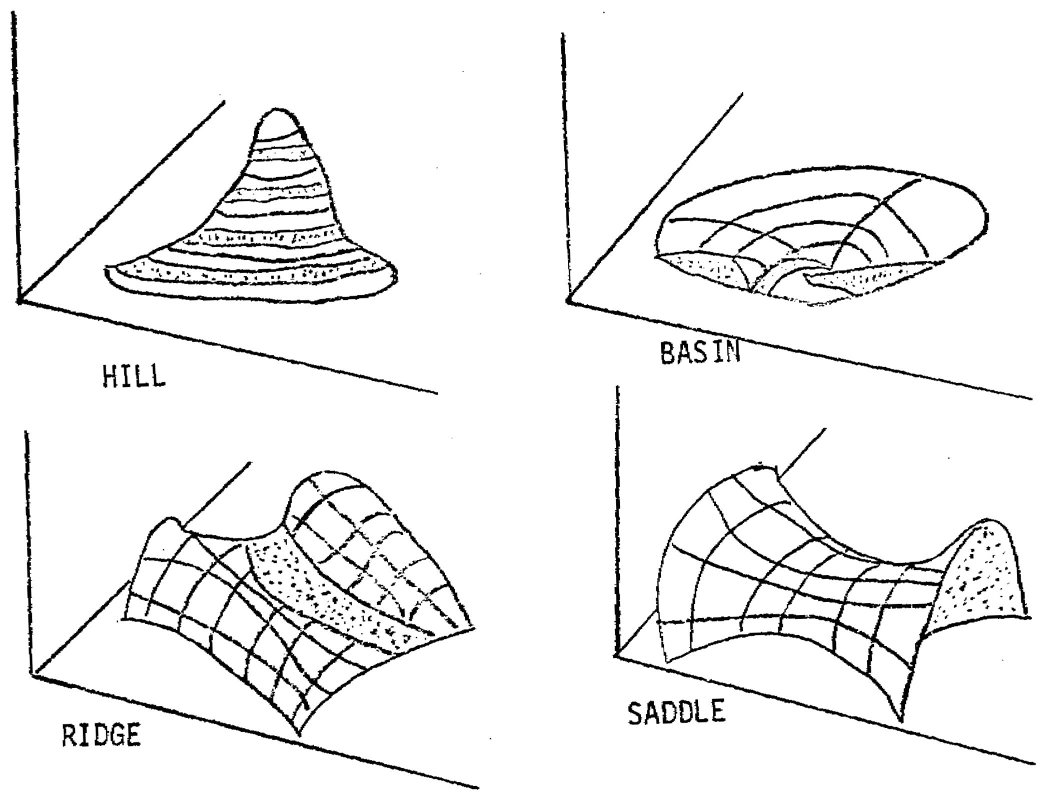
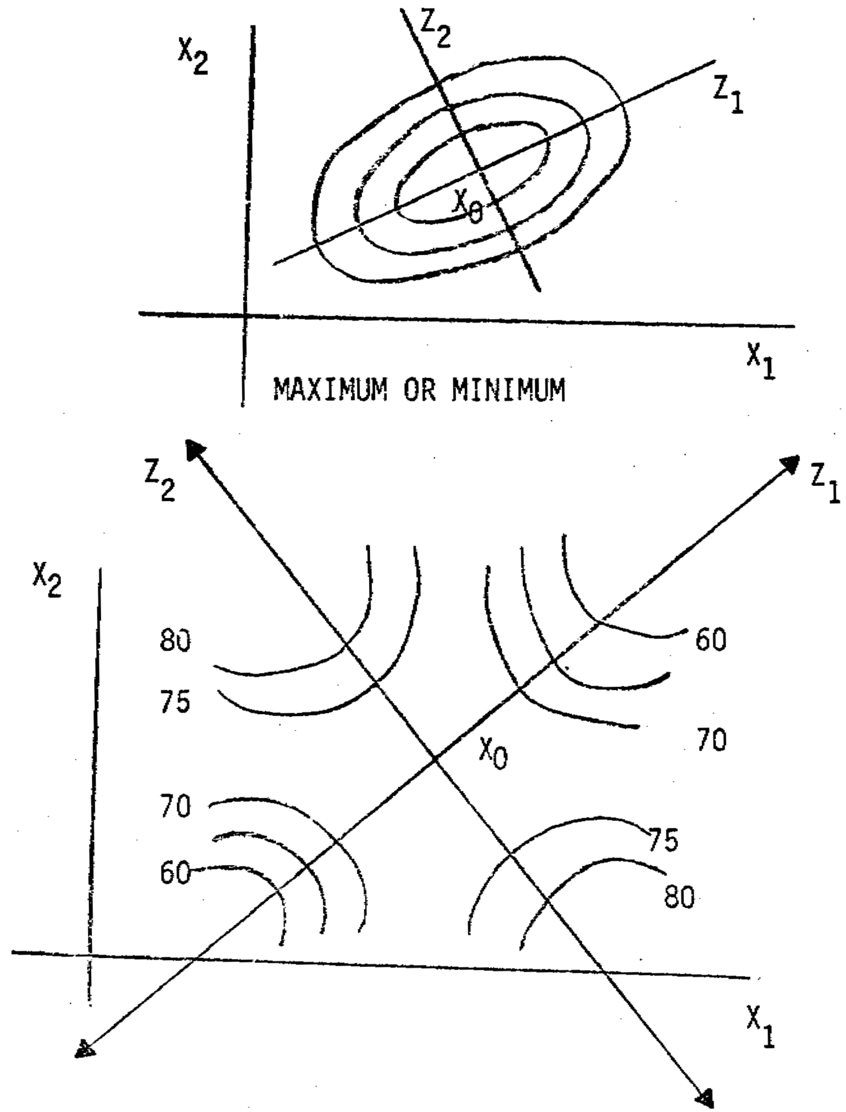


Figure 10

CANONICAL FORM FOR A RESPONSE SURFACE IN TWO VARIABLES



MAXIMUM OR MINIMUM

SADDLE POINT IN A FITTED SURFACE

Figure 11

AN APPLICATION OF THE WEIBULL-GNEDENKO DISTRIBUTION FUNCTION FOR GENERALIZING CONDITIONAL KILL PROBABILITIES OF SINGLE FRAGMENT IMPACTS ON TARGET COMPONENTS

William P. Johnson
Vulnerability Modeling Branch
Concepts Analysis Laboratory
US Army Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland 21005

ABSTRACT. It has previously been shown in a number of works on reliability that the Weibull-Gnedenko distribution gives a good description of the lifetime of numerous elements in electronic equipment when the failure of these elements is regarded as the exceeding of established limits and the part of any of the parameters.

In this study a modified version of the Weibull-Gnedenko distribution has been used to establish a relationship between the conditional probability of kill of target components ($P_{K/H}$) and the momentum per unit area of the impacting fragment (MV/A). A threshold of sensitivity is assumed for each component resulting in a three parameter distribution function. Techniques have been developed and are presented which allows calculation of approximate values for each of the three parameters.

1. INTRODUCTION. When a high explosive munition detonates, the casing that surrounds the explosive charge is fragmented and projected outward. The resultant fragments are usually irregular in shape, vary in weight and lose velocity through air at a rate that is proportional to several physical parameters. Equipment and/or weapon systems struck by these fragments may be unharmed, incapacitated, or killed. For this reason, a portion of the Army's defense effort is devoted to determining the relationship between parameters of the impacting fragments and the resultant system damage. An important part of this effort has been the establishment of conditional kill probability curves.

Conditional kill probabilities, for critical target components, have been developed as a function of the striking mass and selected velocities of the striking fragment.[1] The impracticality of using these same techniques to make measurements over the entire spectrum of fragment mass-velocity combinations, makes a generalized model a necessity. Models of this type facilitate the vulnerability analyses designed to rate the capabilities of existing and prospective systems on the basis of their abilities to withstand impacts from fragments or shaped charges.

[1] R. E. Kinsler, "Conditional Kill Probabilities for Single Fragment Impacts on Components of the Soviet KRAZ-214 Truck (U)," Ballistic Research Laboratories Memorandum Report No. 1995, July 1969, DDC AD504240L, CONFIDENTIAL

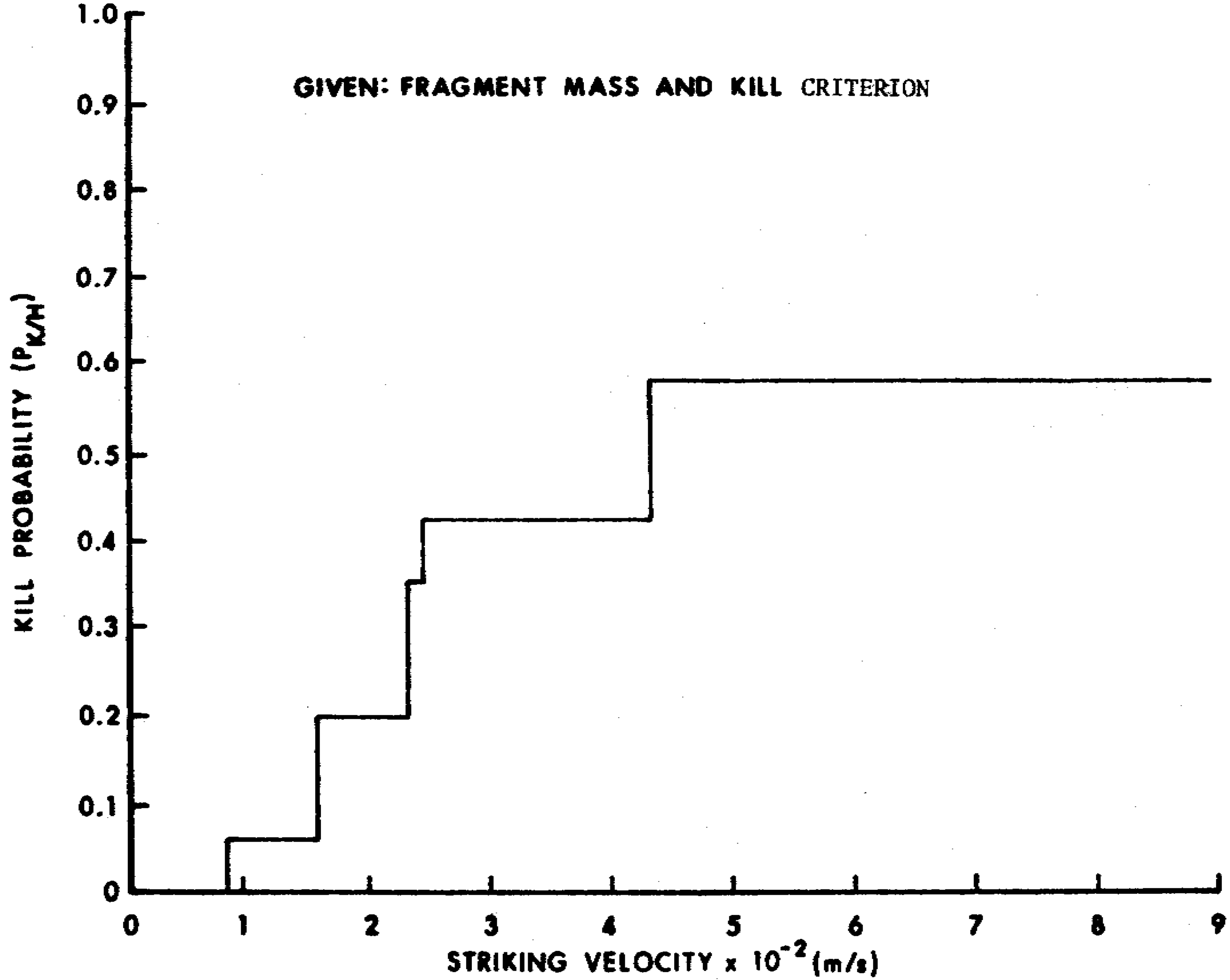


Figure 1 - Typical Probability Function for Fragments Striking a Hypothetical Component.

Previous attempts to relate fragment size to conditional kill probabilities used step functions. The use of these functions required the user to interpolate between curves for masses which were not included in the original analysis and between points of a given curve for velocities not included. Computer programs using these curves required an extensive amount of memory in order to accommodate all of the input data normally required for a detailed analysis. Computer run time was usually extensive, due in part to the numerous conditional statements required to determine whether the right combination of curves was being used in the interpolation process. Figure 1 illustrates a typical step function representation of the experimental data. It should be kept in mind that this approach required a curve of this type for each mass included in the sample population.

The model outlined in the subsequent pages of this report was developed to overcome the limitations in the step function approach. The primary aim was to develop a mathematical relationship which would satisfactorily represent the data points and would expedite subsequent computer analysis of the vulnerability of the target. The accuracy of the data points themselves were not questioned, but merely accepted as the best available representation of the true data points. A description of the technique used to develop the experimental data may be found in references 2 and 3.

2. PROCEDURE. The experimental data inlcuded conditional kill probabilities as a function of kill criterion, fragment mass, fragment velocity and attack orientation. Of these four variables, kill type and fragment attack orientation were held constant for each analysis. More specifically, only the random attack condition was used to develop the model. The random condition assumes that attack of the target is equally likely from all directions. The same procedures used to establish the model could also be used to determine regression constants that would provide conditional kill probabilities as a function of any other attack orientation or kill criterion.

The weights of the fragments included in the experimental data were the following:

Weight (grains)	Weight (kilograms)
1	6.48×10^{-5}
2	1.30×10^{-4}
5	3.24×10^{-4}
10	6.48×10^{-4}
15	9.72×10^{-4}

Weight (grains)	Weight (kilograms)
30	1.94×10^{-3}
60	3.89×10^{-3}
120	7.78×10^{-3}
240	1.56×10^{-2}
500	3.24×10^{-2}
1000	6.48×10^{-2}
2000	1.30×10^{-1}

Striking velocities ranged from 91.44 m/s to 2133.6 m/s (300 to 7000 fps.).

Independent plots were made for several of the target components. Several conditional probability of kill ($P_{K/H}$) - abscissa combinations were considered before it appeared that a relationship between the conditional probabilities and the ratio of the momentum of the fragment to its average presented area existed.

The root mean square deviation (erms) of the experimental data points from the regression curve was used as the criterion for selecting the analytical model which generally best described the data population. Erms was calculated from the following equation:

$$\text{erms} = \left[\frac{\sum_{i=1}^{SS} \left(Y_{p_i} - Y_{o_i} \right)^2}{SS-3} \right]^{1/2} \qquad (1)$$

where: Y_{p_i} = predicted conditional kill probability

Y_{o_i} = observed conditional kill probability

SS = sample size.

3. THE WEIBULL - GNEDENKO DISTRIBUTION.[4] Let us consider a system consisting of a group of elements and possessing the following properties: (1) The

[4]Gertsbakh, I. B., Kordonsky, Kh. B., "Models of Failure," Springer-Verlag, New York, 1969

failures of the elements are mutually independent. (2) Failure of any element is treated as a failure of the entire system. We call such systems chain systems. Let α denote the lifetime of the chain system and let α^i denote the lifetime of the ith element of the system for i = 1, 2, 3,...Z. In such a case:

$$\alpha = \min (\alpha^1, \alpha^2, \alpha^3, \ldots \alpha^Z)$$

We point out an important special case suppose that all elements of the chain system have an exponential distribution of the lifetime. In accordance with formulas derived in reference 4, the distribution function F(t) is equal to:

$$F(t) = 1 - e^{-\sum_{i=1}^{Z} \alpha^i t}$$

where the α^i are parameters of the distribution of the elements of the chain system.

Of special interest is the situation generalizing the case just described but having the following features. The number, Z, of elements of the chain system is great and all the distribution functions $F_i(t)$ are such that:

$$F_i(t) = gt^\lambda + O(t)^\lambda \qquad (2)$$

Where g and λ are positive, as $t \to 0$.

This relationship determines the order of the infinitesimal F(t) for small t. One can show that, for large Z, the distribution function F(t) is well approximated by an expression of the form:

$$F(t) = \begin{cases} 1 - \exp[-\beta(t)^\lambda] & , t \geq 0 \\ 0 & , t \leq 0 \end{cases} \qquad (3)$$

This distribution was proposed by W. Weibull in 1939 without mathematical foundation. A rigorous mathematical treatment of related problems was done by B. V. Gnedenko in 1941.[5]

Equation 3 shall be called the Weibull-Gnedenko Distribution throughout this report.

It has been shown in a number of works on reliability that a Weibull-Gnedenko distribution gives a good description of the distribution of the lifetime of numerous elements in radio-electric equipment

[5] Bolotin, V. V., "Statistical Methods in Construction Mechanics," Stroyizdat, 1965

when the failure of those elements is regarded as the exceeding of established limits on the part of any of the parameters.

Often there are situations in which there is a threshold of sensitivity, t_o, that leads to a displacement of the distribution. For these situations the Weibull-Gnedenko Distribution becomes:

$$F(t) = \begin{cases} 1 - \exp[-\beta \cdot (t - t_o)^{\lambda}] & , t \geq t_o \\ 0 & , t < t_o \end{cases} \qquad (4)$$

Where t_o is defined as the threshold of sensitivity. The definition originated in metrology. Its significance, is that until the parameter t exceeds or equals the threshold of sensitivity the device under investigation does not "feel" the effect of the load and it is only when $t \geq t_o$ that this influence becomes perceptible and causes a probability of failure.

4. APPLICATION OF WEIBULL-GNEDENKO DISTRIBUTION AND CALCULATION OF PARAMETERS. For these studies of a modified version of the Weibull-Gnedenko distribution was found to represent the experimental data population extremely well.

This variation assumes the following form:

$$P_{K/H} = \begin{cases} P_{max}\ [1 - e^{-B(MV/A - K)^N}] & , MV/A > K \\ 0 & , MV/A \leq K \end{cases} \qquad (5)$$

Where the combination MV/A is substituted for the parameter t in the original distribution, and K becomes the threshold of sensitivity. In addition the distribution is multiplied by a constant P_{max}. This multiplication factor was included so as to provide constraints on the regression curve such that $0 \leq P_{K/H} \leq P_{max}$ in contrast to the constraints provided by the original distribution $0 < F(t) < 1.0$. These new constraints were prompted by practical consideration which indicate that the maximum probability of kill that can be obtained on certain components is less than 1.

An explanation of the terms included in the above equation follows:

$P_{K/H}$ - represents the conditional kill probability

M - represents the weight of the fragment (kg)

V - represents the striking velocity of the fragment upon the component (m/s)

A - represents the average presented area of the fragment (cm^2)

P_{max} - Maximum value of $P_{K/H}$ in experimental data set

e - represents the base of the natural logarithm

B & N are regression constants.

In the above equation, K dictates the MV/A value at which the predicted kill probability diverges from zero. In order to determine this constant a computer program was developed in which cutoff points were selected at small intervals between zero and the smallest ratio of MV/A available for the component - kill type combinations under consideration. A combination of the constants B & N were calculated with each selection of K, from the solution of normal equations for a straight line, after a double logarithmic transformation had been made on equation 5. The root mean square error was calculated from each combination of B, N & K by using the calculated values of these constants in Equation 5. It should be noted that the erms was determined thru the use of Equation 5 and not thru its logarithmically transformed version. The logarithmic transformation served solely as a means of obtaining values of the constants.

Equation 1 was used to determine the average deviation of the regression line from the data population. The combination B, N & K selected to be used was that group which minimized erms.

A second equation was used to establish 95 percent confidence limits on the individual values of $P_{K/H}$ for a second value of MV/A. This equation assumed the following form:

$$P_{K/H} - t_{.025, (SS-3)} S_P \leq P_{K/H} \leq P_{K/H} + t_{.025 (SS-3)} S_P \qquad (6)$$

where: $P_{K/H}$ - represents the conditional kill probability

SS - represents the sample size

S_P - represents the standard deviation of the $P_{K/H}$ population

$t_{.025}$ - represents the value taken from the students t table with (SS-3) degrees of freedom.

Confidence intervals of this form have a minimum at the mean of the independent variable values included in the experimental data set.

Therefore, it is expected that predictions made on either extreme of the data set would have larger confidence intervals. In these analyses these intervals were restricted to the interval $0 \leq P_{K/H} \leq 1$, although mathematically, they would have fallen outside of these intervals. This restriction forces the confidence intervals to reach a plateau at each of these limits in some instances. Figure 2 illustrates the goodness of fit of the above model to data from a typical component.

5. MODEL VALIDATION. Upon completion of the program development, constants were calculated for several data sets, and given to vulnerability analysts to test in established vulnerability programs. The purpose was to determine whether the use of the equation rather than the step-functions would in fact decrease the running time of the programs as had been speculated and also to determine whether significant statistical differences existed between the results of the two methods.

It was observed that the vulnerable area calculations based on mass-velocity combinations contained within the original data set were reasonably close to values obtained thru the use of step-functions, however, vulnerable area values determined from extrapolated values of the data set were generally larger.

A plot of the $P_{K/H}$ data from a typical data set revealed that the step-functions and the generalized curve diverged at the maximum $P_{K/H}$ value for each of the fragment masses. As can be observed in Figure 3, the step function approach assumes that a maximum $P_{K/H}$ can be achieved and the $P_{K/H}$ curve will remain constant at this value. But because the Weibull-Gnedenko distribution provides a curve with $P_{K/H}$ values directly proportional to the independent variable MV/A, $P_{K/H}$ values greater than those obtained in the experimental data set were obtained for each mass smaller than the largest.

To circumvent this problem it was decided that a second equation was needed which would enable the analyst to determine the maximum $P_{K/H}$ value obtainable for a given component as a function of a fragments mass. The logic was that during the actual calculation of vulnerable areas in the analyst's computer programs, a comparison could be made between the $P_{K/H}$ value determined from the generalized curve and the maximum $P_{K/H}$ value which could be obtained from the fragments mass. The smaller of these two values would then be used in the vulnerable area calculations.

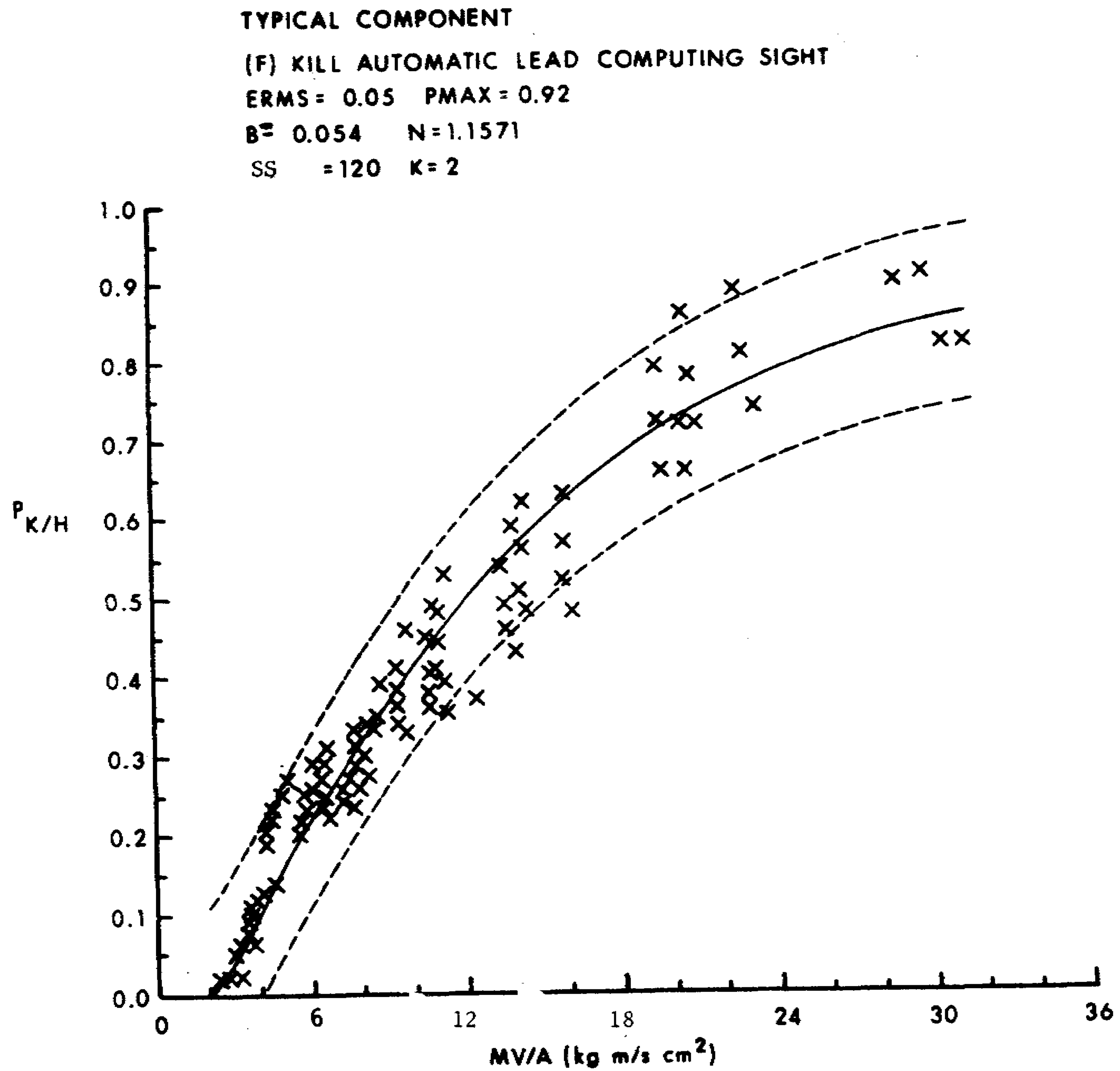


Figure 2- Illustration of Typical Data Fit.

To carry out this logic a relationship between the maximum $P_{K/H}$ values for each mass contained in the experimental data set and the mass itself was needed. Interpolation of the maximum value of $P_{K/H}$ for masses not included in the data set was the main reason for requiring a generalized relationship.

As had been the case earlier, it appeared that the Weibull-Gnedenko Distribution provided the best relationship between the variable of interest. The relationship:

$$P_{max} = 1 - e^{-B1 \left[(\log_{10}M + 1) - K1\right]^{N1}} \qquad (7)$$

gave excellent regression fits of the experimental data with small erms values as determined by Equation 1 (Figure 4).

In the above equations the independent variable can be chosen as ($\log_{10}M$) if the masses of interest are greater than 1 unit. The above form will provide $P_{K/H_{max}}$ values for fragments with mass ≥ 0.1 units. The actual unit of mass chosen to use in the equation is immaterial since the values of B1, N1 and K1 will adjust to the unit used.

Upon completion of the generalization of the maximum $P_{K/H}$ as a function of each fragment's mass, the idea to use the P_{max} value predicted from Equation (7) as a variable in Equation (5) evolved.

The objective was to provide an upper limit on the $P_{K/H}$ predictions for each mass under consideration and to avoid the comparisons required in the earlier method. To determine whether this was a feasible technique, both sides of Equation (5) were divided by P_{max} giving:

$$P_{K/H} / P_{max} = 1 - e^{-B (MV/A - K)^{N}} \qquad (8)$$

Equation (8) indicates that the distribution function for the ratio of the conditional probability of kill to the maximum probability of kill for a given fragment is given by the Weibull-Gnedenko equation. All variables in Equation (8) are identical to those described for Equation (5) with the exception of P_{max}. P_{max} in the above equation is the maximum

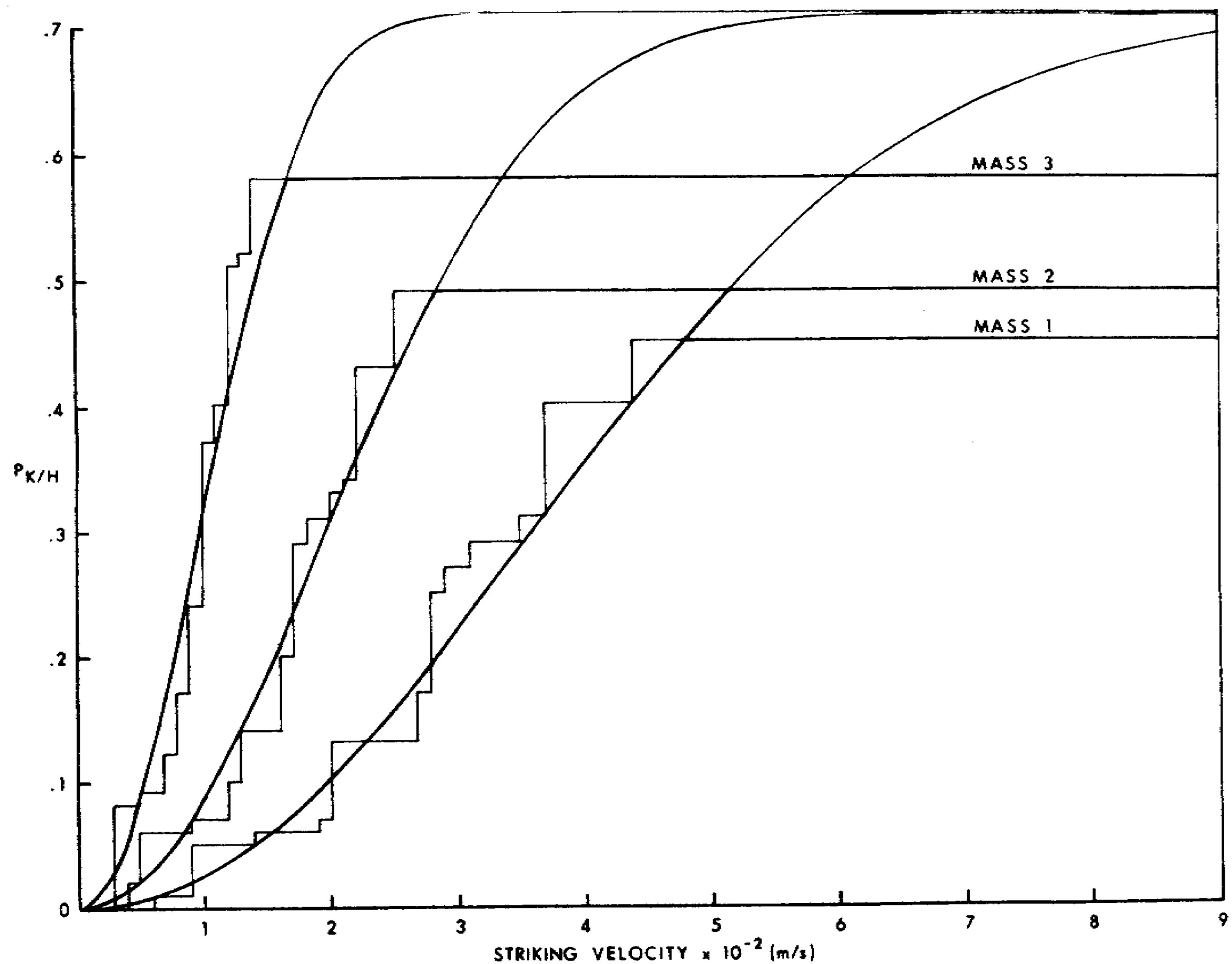


Figure 3 - Divergence of Model and Step Functions

probability of kill obtainable against a selected component for the desired fragment weight. Since P_{max} is determined from Equation (7), Equation (7) must be developed prior to using Equation (8).

In studies designed to document the prediction error of $P_{K/H}$ values obtained from Equation (8) and those obtained from Equations (5) and (7), both methods appeared to work equally well. To illustrate this point nine components are listed below with associated root mean square errors (erms) obtained from both methods:

COMPONENT NUMBER	ERMS EQUATION 5, 7	ERMS EQUATION 7,8
1	.072	.100
2	.003	.004
3	.013	.008
4	.000	.001
5	.052	.027
6	.069	.042
7	.166	.192
8	.137	.120
9	.110	.110

Although there appears to be little difference in the erms values calculated from each of the two methods, Equation (8) may be easier to use during actual vulnerability calculations and is therefore recommended. Both methods prevent the divergence of the step-function and the generalized curves illustrated previously, and both appear to reduce computer running time over that which was previously required.

Though it appears that the objectives of the generalization have been achieved, more formal tests are needed to determine whether the resulting vulnerable area values produced by (1) the step-function curves and (2) the generalized equation are statistically equivalent. Suggested techniques are discussed in the next section.

6. STATISTICAL METHODS.[6,7] In order to determine whether there is a significant statistical difference between the vulnerable area values obtained with the step-functions and the values obtained with the generalized equations, the following hypothesis, level of significance and test statistics are suggested.

[6] Dixon, Wilfrid J. and Massey, Frank J. Jr., "Introduction to Statistical Analysis," 2nd Edition, McGraw-Hill Book Company, Inc., 1957, Chapters 9 & 14.

[7] D. V. Huntsberger, "Elements of Statistical Inference," Second Edition, Allyn & Bacon, 1967, Chapter 2.

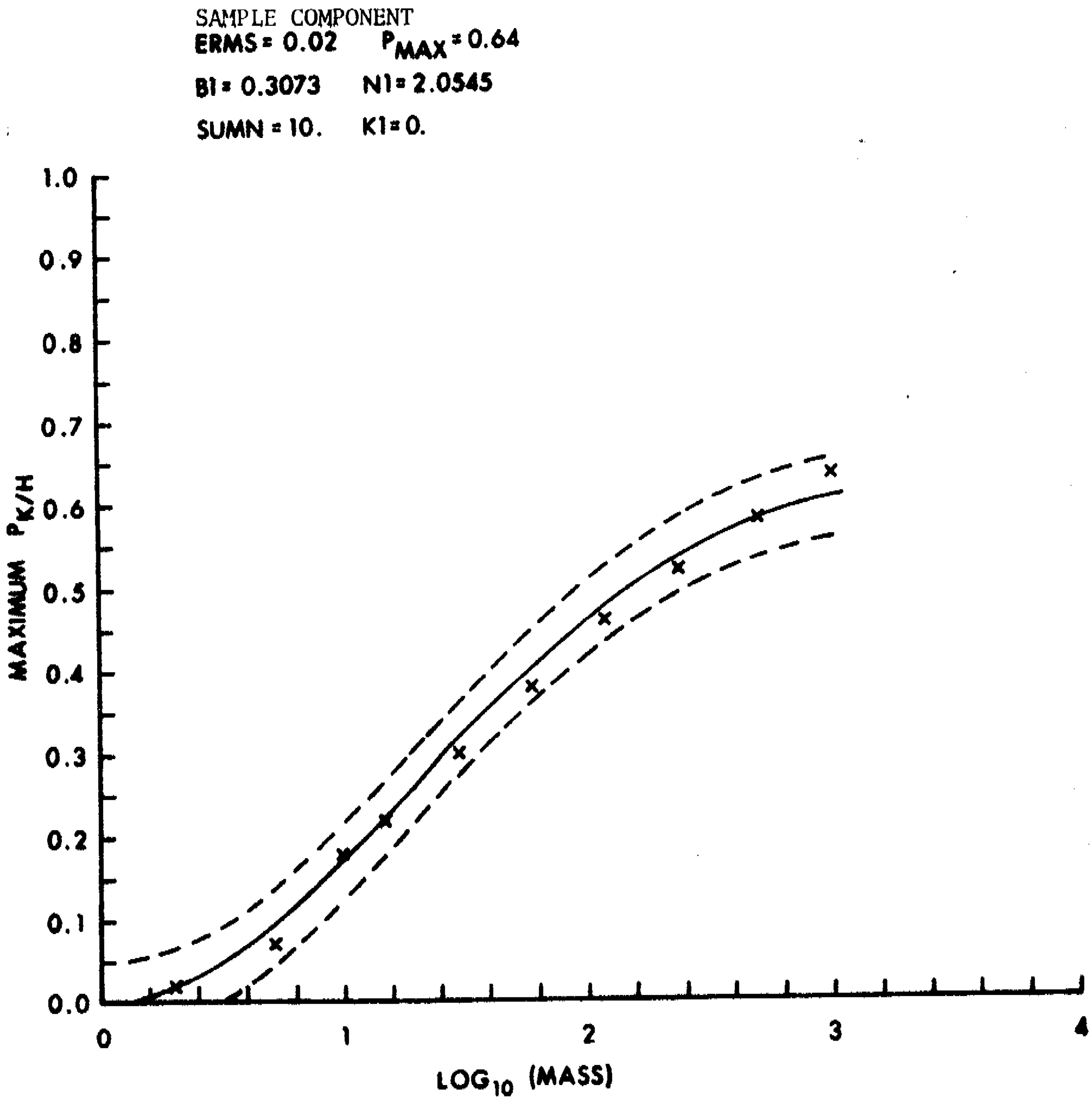


Figure 4 - Maximum $P_{K/H}$ Versus Fragment Weight

1. Let: X_s be the vulnerable area value obtained from mass (X) and velocity (Y) using the step-function approach.

X_g be the vulnerable area value obtained from mass (X) and velocity (Y) using the generalized equation approach.

Let the difference between the two vulnerable area values previously defined above equal X where $X = X_s - X_g$. Clearly, X will only be equal to 0, that is no difference between the individual values exists when $X_s = X_g$ otherwise X will be either a positive or a negative number ($X > 0$ or $X < 0$).

Let SS represent the number of differences to be considered.

2. Test the hypothesis that the mean difference between the vulnerable area values generated by step-functions (μ_s) and the vulnerable area values generated by the generalized $P_{K/H}$ equations (μ_g) is statistically equivalent to zero, (i.e. $\mu_d = \mu_s - \mu_g \stackrel{\sim}{=} \Sigma X/SS = 0$).

3. A test of significance is in general terms a calculation by which the sample results are used to throw light on the truth or falsity of a null hypothesis. A quantity called a test statistic is computed, which measures the extent to which the sample departs from the null hypothesis in some relavant aspect. If the value of the test criterion falls beyond certain limits into a region of rejection the departure is said to be statistically significant or more concisely significant. Tests of significance have the property that if the null hypothesis is true (not difference between means) the probability of obtaining a significant result has a known value most commonly referred to as α and chosen as 0.05 or 0.01. This probability is the significance level of the test.

For the purpose of the analysis one should choose:

4. $\alpha = .05$ or $\alpha = .01$. The test criterion should be:

5. $$\hat{t} = \frac{\mu_d}{S/\sqrt{SS}} = \frac{\dfrac{\Sigma\left(X_s - X_g\right)}{SS}}{S/\sqrt{SS}} = \frac{\bar{X}}{S/\sqrt{SS}}$$

where: X_s = Vulnerable area estimate from step-functions

X_g = Vulnerable area estimate from generalized curve

SS = Sample size

S = estimate of population standard deviation and is computed from the following:

$$S = \left[\frac{\Sigma (X_s - X_g)^2 - \frac{(\Sigma(X_s - X_g))^2}{SS}}{SS-1} \right]^{.5}$$

6. Our population should be large enough so that the central limit theorem is applicable. Therefore, our test statistic will have a t distribution with SS-1 degrees-of-freedom.

7. Our rejection region defined in paragraph 5 will be obtained from standard tabular values of the t distribution with SS-1 degrees-of-freedom.

In other words we will only be able to statistically state that vulnerable area values computed from the two equations are different if the computed value of the test criterion t is outside of the region defined for the combination, that is:

$$|\hat{t}| \geq |t_{1/2\alpha,\ SS - 1.}|$$

Otherwise, we will have to say that there is not enough evidence to reject the hypothesis.

An alternate approach to the above would be the use of a two way analysis of variance which would offer the advantage of determining whether the differences, if any exist, are between the row effects (masses), column effects (velocities) or both.

7. SUMMARY. The Weibull-Gnedenko Distribution function has been used to establish a relationship between the conditional probability of kill ($P_{K/H}$) of a fragment and the fragments momentum per unit area (MV/A). Statistical tests are suggested which determine whether a significant statistical difference exists between vulnerable area values calculated using this distribution function and vulnerable area values calculated from currently used step-functions.

If it is determined that no significant difference exists, then it is recommended that this distribution function be used in subsequent vulnerability analyses because of its convenience and the reduction obtained in computer running time.

DECISION THEORY APPROACH TO GRADING BINOMIAL POPULATIONS

Paul Williams
U S Army Materiel Systems Analysis Activity
Aberdeen Proving Ground, Maryland

ABSTRACT. A method of classifying lots of mass produced articles into one of k categories on the basis of the number of defectives allowable in a single sample size n (n<k-1) is presented. A beta distribution is assumed as the prior distribution of the true lot fraction defective where the parameters are to be estimated using knowledge obtained from lots inspected in the past. An optimum decision rule for obtaining the allowable number of defectives in a lot is developed where the allowable lot fraction defective and sample size are determined prior to testing.

1. INTRODUCTION. The production of large quantities of mass produced artilces often necessitates dividing the articles, for the purpose of homogeneity, into groups called lots. It is then desirable for the producer or consumer to place the lot in a category based on the quality or reliability determined from some characteristic of the individual articles in the lot. If the articles in the lot are to be labeled effective or defective after inspection, then the inspection is to be by attributes since inspection by variables is based upon quantitative measurements.

Where large lots are concerned, the cost of testing of each item may be excessive or where the item is destroyed by testing, a sampling plan to estimate the number of defectives in the lot must be devised. Such a plan is the single sampling plan where a sample of n items is selected from a large lot of size N. After each item has been inspected or tested, the number of defective items (r) is determined and the lot is placed into one of k categories based on the number of defectives in the sample.

A basic solution to the single sampling problem is to decide on an acceptable quality level p_1 such that the consumer desires to accept almost all lots of fraction defective p_1 or less and also to specify an objectionable quality level p_2 which represents lots of quality so inferior that the consumer cannot accept more than a few lots of this quality. By specifying quality levels p_1 and p_2, the risk α of rejecting a lot of lot fraction defective p_1 and the risk β of accepting a lot of objectionable quality can be determined given the distribution of the process sampled. The problem then is to find the smallest sample size and acceptance number which will give the desired protection. In the case of fixed sample size, the problem is to determine a minimum acceptance number.

The sampling plan considered in this paper is the single sampling plan and is solved using the decision theory approach, that is, a priori information will be combined with data from the sampling program. However, economic considerations are assumed to be unimportant or small when compared to the risk of making a wrong decision.

2. DEFINITIONS OF THE ACTION SPACE. The sampling plan then calls for obtaining a fixed sample of n items from the lot and testing to determine the number of items (r) which are defective. If the lot contains c or less defectives, the lot is accepted, but if the lot contains c+1 defectives the lot is rejected. With any sampling plan there is the possibility of making a wrong decision, that is, taking an action which would not be taken if the true quality of the lot were known before making a decision. The set of all possible actions which may be taken to solve a problem is known as the action space. As an example, the space of possible actions could for a decision about the disposition of a lot contain two points; (1) accept the lot or (2) reject the lot. The correct action to take would depend on the true state of nature; that is the actual proportion of defective items in the lot. The difficulty of course is that the true state of nature is not known unless the entire lot is tested.

The action space can then be defined as placing a lot into one of k categories. A lot is placed into one of k categories by using the following rule:

If	$o \leq r \leq c_1$	Lot is Grade A
	$c_1 + 1 \leq r \leq c_2$	Lot is Grade B
	. .	. .
	. .	. .
	. .	. .
	. .	. .
	$c_{k-1} + 1 \leq r \leq c_k$	Lot is Grade k

Each of the above intervals should be determined from the allowable lot fraction defective for each category.

3. PRIOR DISTRIBUTION. In considering a production lot in which the items can be classified into two groups, effective and defective, a natural probability model is the Bernolli process which has probability mass function:

$$(1) \qquad f(x) = p^x (1-p)^{1-x} \qquad \begin{array}{l} x = o \text{ Success} \\ x = 1 \text{ Failure} \end{array}$$

where p is the probability that a randomly chosen item in the lot will be defective. As the prior distribution of the parameter p in the Bernolli process, we choose a Beta Distribution of the form

$$(2) \qquad f(p;AB) = \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} p^{A-1} (1-p)^{B-1}$$

where: $o \leq p \leq 1$

$A > o$

$B > o$

where A and B are to be estimated from prior knowledge, that is previous testing. The Beta Distribution has been chosen as a prior distribution for the true defect rate in the lot for two reasons: (1) The random variable p in the Beta Distribution is defined in the interval $o \leq p \leq 1$ as is the parameter in the Bernolli process. (2) The Beta Distribution represents a rich family of possible densities to express our knowledge of prior information.

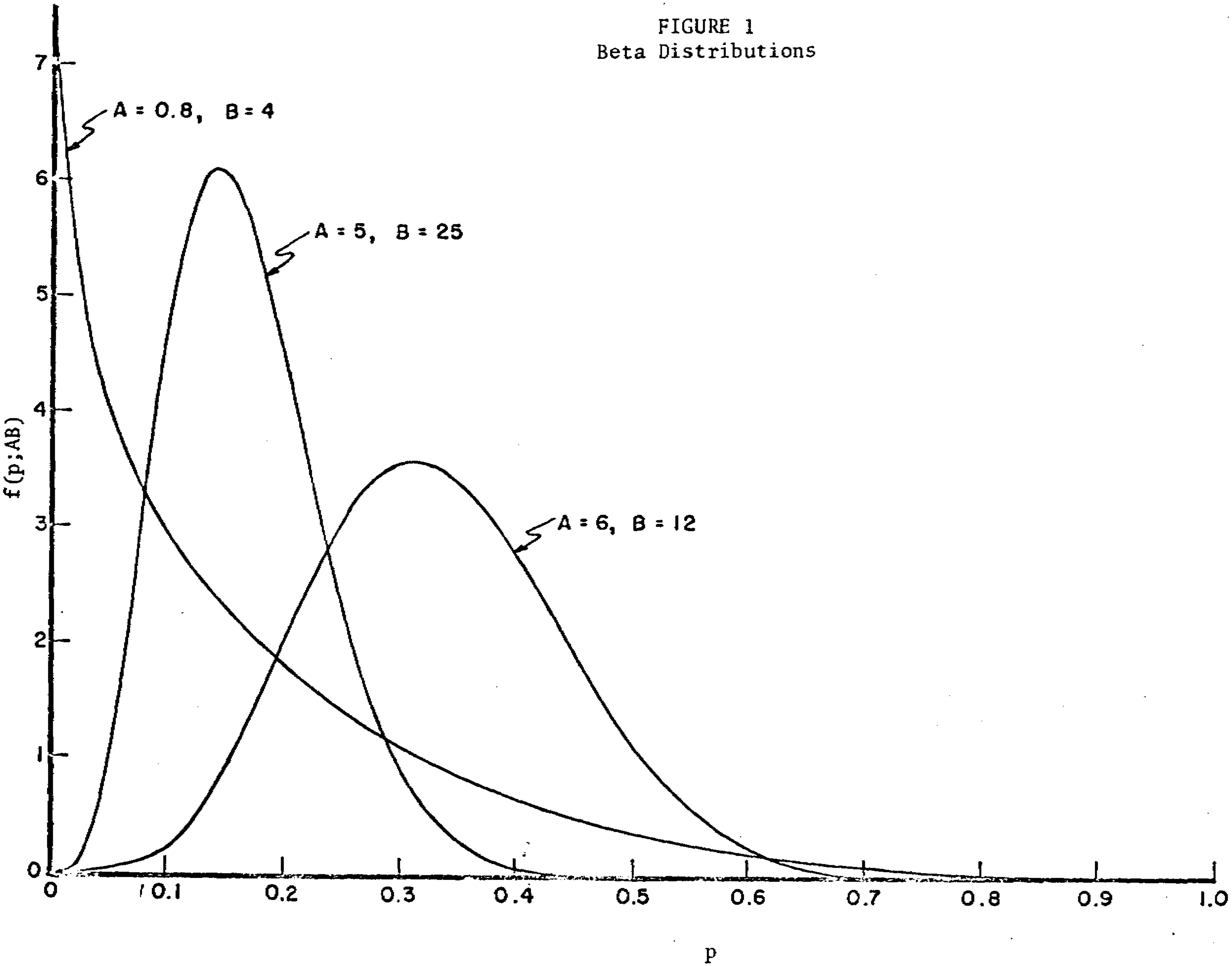
For various values of A and B we can generate the family of Beta Distributions, some examples of which are given in Figure 1. To select a member of the family of Beta Distributions we can choose its mean and variance as an expression of our prior beliefs about the unknown parameter p. The first and second moments about the origin are:

$$(3) \qquad E(p) = \mu_1 = \frac{A}{A+B}$$

$$(4) \qquad E(p^2) = \mu_2{}^1 = \frac{A(A+1)}{(A+B)(A+B+1)}$$

Using the number of defects as observations on p, parameter estimates for the Beta can be estimated by:

FIGURE 1
Beta Distributions



(5) $$\hat{A} = \frac{\hat{\mu}_1 \, (\hat{\mu}_1 - \hat{\mu}^1{}_2)}{\hat{\mu}_1 \, (\hat{\mu}_2{}^1 - \hat{\mu}_1) + \hat{\mu}_2{}^1 \, (1 - \hat{\mu}_1)}$$

(6) $$B = \frac{(\hat{\mu}_1 - \hat{\mu}^1{}_2 \, (1 - \hat{\mu}_1)}{\hat{\mu}_1 \, (\hat{\mu}^1{}_2 - \hat{\mu}_1) + \hat{\mu}^1{}_2 \, (1 - \hat{\mu}_1)}$$

In some actual problems the choice of a specific beta prior distribution may be largely subjective at the start of production. However, after a few lots have been produced the selection of the prior distribution becomes more objective by incorporating previous lot results.

4. PROBABILITY OF CORRECTLY GRADING A LOT. Under the plan stated in paragraph 3 and using the formula of total probability*, the probability of correctly grading the lot for the two action state case is the sum of the probabilities of accepting lots of true quality p_1 and rejecting lots of true quality p_2. This probability can be expressed by an equation of the form:

Probability of Correctly Grading Lot = P

$$= \Pr\,(o \leq p < p_1) \;\; \Pr\,(o \leq r \leq c_1 | o \leq p < p_1)$$

$$+ \;\; \Pr\,(p_1 \leq p < p_2) \;\; \Pr\,(c_1 + 1 \leq c \leq c_2 | p_1 \leq p < p_2)$$

Using the distribution assumptions, the probability of correctly grading k categories in the above expression can be written as:

(7) $$P = \int_o^{p_1} \frac{(A+B-1)!}{(A-1)!\;(B-1)!} \; p^{A-1} \; (1-p)^{B-1} \sum_{r=o}^{c_1} \binom{n}{r} p^r \; (1-p)^{n-r} \, dp$$

$$+ \int_{p_1}^{p_2} \frac{(A+B-1)!}{(A-1)!\;(B-1)!} \; p^{A-1} \; (1-p)^{B-1} \sum_{r=c_1+1}^{c_2} \binom{n}{r} p^r \, (1-p)^{n-r} \, dp$$

*B.V. Gredenko "Theory of Probability"

$$+\int_{P_2}^{P_3} \frac{(A+B-1)!}{(A-1)!\ (B-1)!}\ p^{A-1}\ (1-p)^{B-1} \sum_{r=c_2+1}^{c_3} \binom{n}{r} p^r\ (1-p)^{n-r}\ dp$$

$$\ldots + \int_{P_{k-1}}^{P_k} \frac{(A+B-1)!}{(A-1)!\ (B-1)!}\ p^{A-1}\ (1-p)^{B-1} \sum_{r=c_k+1}^{c_1} \binom{n}{r} p^r\ (1-p)^{n-r}\ dp$$

The problem then is to determine the number of defectives $c_1 c_2 c_3 \ldots c_k$ which will make placing a lot into one of k categories an optimum act under uncertainty.

5. TERMINAL ANALYSIS. The first step in terminal analysis is to determine the posterior distribution of the process. If we let $X_1\ X_2\ X_3 \ldots X_n$ represent a random sample from the lot, the distribution of any x given p can be written:

$$f\ (x/p) = p^x\ (1-p)^{1-x} \qquad x = 0,\ 1 \tag{8}$$

Since the observations are independent the conditional density of all the X's in the sample is:

$$f\ (X_1, X_2, X_3 \ldots Xn|p) = p^{\sum x}\ (1-p)^{n-\sum x} \tag{9}$$

To simplify the problem, the n - tuple $(X_1\ X_2\ X_3 \ldots Xn)$ is replaced by a 2-tuple. For the Bernolli process with binomial sampling, $Z = (r, n)$, where $r = \sum x_i$ is a sufficient statistic and will therefore replace the n - tuple in the analysis. Thus equation (9) may be written as:

(10) $$f\ (r|p) = \binom{n}{r} p^r\ (1-p)^{n-r}$$

$$r = 0,\ 1,\ 2\ \ldots.n$$

If we put a beta prior distribution on p, the posterier distribution is:

(11) $$h\ (p|r) = \frac{(n+A+B-1)!}{(A+r-1)!\ (n+B-r-1)!}\ p^{A+r-1}\ \ (1-P)^{n+B-r-1}$$

The problem now is to find a function $\emptyset$ defined by $\hat{P} = \emptyset\ (r)$ which will minimize the posterier expected risk in estimating p. The definition of expected risk is:

(12) $$E\left[R\ (\emptyset,\ p)\right] = \int_0^1 R\ (\emptyset,\ p)\ f\ (p)\ dp$$

The concept of correct action leads to the definition of another term, the loss function, denoted by

$$L\left[\emptyset\ (r);\ p\right]$$

The loss function is intended to give the loss which is incurred when a certain action is taken when a certain state of nature prevails. Since the interest is in the estimate of $\hat{p}$ being close to p, there difference should be small. For this purpose we use the squared error loss function which is:

$$L\ \ [\emptyset\ \ (r);\ p]\ = (\hat{p}-p)^2$$

The posterier risk then can be written as:

$$(13) \qquad \nu(\hat{p};r) = \frac{(n+A+B-1)!}{(A+r-1)!\ (n+B-r-1)!} \int_0^1 (\hat{p}-p)^2\, p^{A+r-1}\, (1-p)^{n+B-r-1}\, dp$$

$$= \hat{p}^2 - 2\hat{p}\frac{(A+r)}{(n+A+B)} + \frac{(A+r)\ (A+r+1)}{(n+A+B)\ (n+A+B+1)}$$

The value of $\hat{p}$ as a function of the sample which will minimize the posterier risk is then obtained by taking the derivative of $\nu(\hat{p};r)$ with respect to $\hat{p}$ and setting the result equal to zero; that is:

$$\frac{\partial\, \nu(\hat{p};r)}{\partial\, \hat{p}} = 2\hat{p} - 2\frac{A+r}{n+A+B} = 0$$

or

$$(14) \qquad r = \hat{p}\ (n+A+B) - A$$

Thus an optimum decision rule has been obtained for finding the value of c in terms of $\hat{p}$. The criteria for finding the number of allowable defectives in a sample of n can be given as

Reject if: $c_1 > p_1\ (n+A+B) - A$

Accept if: $c_1 \leq p_1\ (n+A+B) - A$

This criteria can readily be extended to more than two categories by specifying the allowable lot fraction defective $\hat{P}_k$ for each of k categories.

If another lot is to be tested, A+r and n+B-r can be used as parameters for a new prior distribution. It may be noted that after a number of lots have been tested, the terminal analysis tends to become less sensitive to the parameters of the initial prior distribution and more dependent on the accumulated test experience.

6. CONCLUDING REMARKS. The decision theory approach applied to the sampling situation in this paper is productive since a small amount of experimental data is combined in a rationale manner to make decisions among alternative courses of action. It is assumed that prior test experience is available so that the prior distribution may be determined since the entire structure of prior convictions is expressed by the beta distribution. Also, the method considered does not depend on the specification of consumers on producers risk but only requires the experimenter to answer a non-statistical question about good or bad quality. That is, what percent defective is acceptable for the lots under test? In most practical situations, this question can easily be answered.

BIBLIOGRAPHY

(1) Mood, A. M. and Graybill, F.A., Introduction to the Theory of Statistics, McGraw Hill Book Co., Inc., Second Edition, New York 1963.

(2) Raiffa, H. and Schlaifer, R., Applied Statistical Decision Theory, Division of Research, Graduate School of Business Administration, Harvard University, Boston 1961.

(3) Gredenko, B. V., The Theory of Probability, Chelsea Publishing Co., New York 1962.

(4) Golub, Abraham, Designing Single - Sampling Inspection Plans When The Sample Size is Fixed, Journal of the American Statistical Association, Volume 48 (1953), pp. 278-288.

(5) The American Statistical Association, Acceptance Sampling, presented at the 105th meeting, Cleveland, Ohio, January 27, 1946.

PSEUDO-BAYESIAN INTERVALS FOR RELIABILITY OF A SERIES SYSTEM GIVEN WEIBULL COMPONENT DATA

Ronald L. Racicot
Research Directorate
Benet Weapons Laboratory
Watervliet Arsenal
Watervliet, New York 12189

ABSTRACT. A pseudo-Bayesian solution for confidence intervals on average reliability of a series-system composed of Weibull components has been formulated. The term pseudo-Bayesian is used since the goal is to choose priors that lead to classical limits and not the usual Bayesian limits. Uniform priors are assumed for population parameters to approximate complete prior ignorance. The distribution assumed for component interarrival failure times is the 2-parameter Weibull with both parameters unknown.

In the solution derived, an approximation is used to compute average system reliability from average component reliabilities. A normal distribution is then assumed for the log of system reliability with mean and variance being computed from the posterior means and variances of the individual component log-reliabilities. The bias in the mean log-reliability was also investigated and an unbiasing factor can be used to reduce the potentially large errors in system reliability resulting from the accumulation of biases in the component means.

Monte Carlo trials were conducted to determine frequency exactness of the derived intervals for particular cases. Near exactness was observed for a number of cases depending on Weibull shape parameters, true component reliabilities and sample sizes.

NOTATION.

$f(t)$	pdf of interarrival times of failures;
$F(t)$	cdf corresponding to f(t);
$\bar{F}(t)$	1 - F(t)
$F_{Ra}(\cdot)$	cdf of average reliability R_a;
$h(t)$	renewal rate; the unconditional pdf of component failure and subsequent renewal;
n_c	number of components in system;
n_f	number of component failures;
n_m	number of missions over system life;

$R(t,\tau)$	reliability at time t for an interval τ;
$R_j(t,\tau)$	reliability of the jth component;
$R_a(\tau)$	average reliability over system life for mission interval τ;
$R_{ja}(\tau)$	average reliability of the jth component;
$R_s(t,\tau)$	system reliability at time t for an interval τ;
$R_{sa}(\tau)$	average system reliability;
t	system or component age;
t_i	starting time for the ith mission;
X	sample outcome including both failure and censoring times;
x_f	component failure time;
α	Weibull scale parameter;
β	Weibull shape parameter; and
τ	mission length for which reliability is required.

1. INTRODUCTION. The general problem is to determine the reliability of a series system from component test results where the term "series" implies that a failure of any component in the system results in system failure. The interarrival failure times of each component are assumed to follow the 2-parameter Weibull distribution with both parameters unknown. In addition, the assumption of ideal repair is made wherein a component is instantaneously renewed with a like new component whenever it fails during system operation. Finally, a fixed number of failures with associated failure times are assumed given for each component.

Consider first a single component in the system. The failure times for a component subject to ideal repair form a renewal process. The theory for renewal processes and interval-reliability are well covered in the literature; so only the final results are summarized here [1-5]. The failure times of components within a system are not known in advance and are treated probabilistically by introducing the renewal rate (unconditional failure rate) h(t) over the population of all systems. The renewal rate in this case is distinguished from the hazard or conditional failure rate which describes failure of a non-repairable item. The renewal rate is a function of the underlying failure distribution [2, 3]:

$$h(t) = f(t) + \int_0^t f(t-x)h(x)dx. \qquad (1)$$

Interval or mission reliability at system time t for mission length τ can be determined from the renewal rate [4]:

$$R(t,\tau) = \bar{F}(t+\tau) + \int_0^t \bar{F}(t+\tau-x)h(x)dx \qquad (2)$$

in which

$$F(t) = 1 - \exp(-\alpha t^{\beta}) \qquad (3)$$

for the Weibull distribution.

Since interval-reliability given by (2) is transient, there is some motivation to define a single reliability index that would characterize a component in a system throughout system life. For this, one can define the worst mission reliability, the asymptotic reliability or, as is done in this paper, the arithmetic average reliability for some fixed system life given in number of missions:

$$R_a(\tau) = \frac{1}{n_m} \sum_{i=1}^{n_m} R(t_i,\tau) \qquad (4)$$

in which t_i = starting time for the ith mission.

Average component reliability, as defined by (4), does not readily lend itself to classical confidencing approaches since a statistic for reliability could not be found which depends only on the true reliability. The statistics considered depended on both of the unknown Weibull parameters. This is in contrast to reliability of a non-repairable component in which, for the Weibull example, the distribution of the maximum likelihood estimate of reliability depends only on the true reliability [6, 7].

A Bayesian approach was consequently used to at least render this problem numerically tractable. The goal was not to determine the usual Bayesian limits, in which prior information is to be used, but rather to choose priors which gave near classical frequency limits; hence the term pseudo-Bayesian.

The solution for the single Weibull component in which uniform priors were assumed for the population parameters is presented in another paper [8]. The final result for the posterior cdf for average component reliability from this paper is given by the following expression:

$$\bar{F}_{Ra}(z|X;\tau,n_m)=K^{-1} \int_0^{\infty} a(\beta)(b(\beta))^{-n_f-1} P(n_f+1,w(\tau,n_m,z,\beta))d\beta \qquad (5)$$

in which $K \equiv \int_0^\infty a(\beta)(b(\beta))^{-n_f-1} d\beta$

$$a(\beta) \equiv \beta^{n_f} \prod_{i=1}^{n_f} (x_{fi})^{\beta-1}$$

$$b(\beta) \equiv \sum_{j=1}^{N} x_j^{\beta}$$

N = total sample size including both failures and censoring times

$$w(\tau,n_m,z,\beta) \equiv R_a^{-1}(\tau,n_m,z,\beta)b(\beta)$$

P(n;x) = incomplete gamma function

$$= 1-e^{-x} \sum_{i=0}^{n-1} x^i/i! \text{ for integer } n \text{ [9].}$$

Solution of (5) for given z, τ, n_m and sample outcome X is accomplished by numerical quadrature. Confidence limits on component reliability can be determined from (5) by computing the probability limit z for a given probability level.

In this paper, results are presented for a few of the problems encountered in determining system reliability from the component results using a similar pseudo-Bayesian approach and using the derived posterior distribution (5) for average component reliability.

2. NUMERICAL COMPUTATION OF SYSTEM RELIABILITY. The first problem encountered is the computation of average system reliability from the average component reliabilities. The reliability of a series system is given as

$$R_s(t,\tau) = \prod_{j=1}^{n_c} R_j(t,\tau). \qquad (6)$$

Average system reliability can be defined in a similar manner as average component reliability:

$$R_{sa}(\tau) = \frac{1}{n_m} \sum_{i=1}^{n_m} \prod_{j=1}^{n_c} R_j(t_i,\tau) \qquad (7)$$

This is a difficult equation to work with since the time dependent component reliabilities are required. Ideally, one would like to express average system reliability in terms of average component reliabilities so that use can be made of the posterior distribution given by (5).

Three approximations to average system reliability were investigated:

$$R_{sa}(\tau) \simeq 1 - \sum_{j=1}^{n_c} (1-R_{ja}(\tau)) \tag{8a}$$

$$\simeq \exp\left[- \sum_{j=1}^{n_c} (1-R_{ja}(\tau))\right] \tag{8b}$$

$$\simeq \prod_{j=1}^{n_c} R_{ja}(\tau) \tag{8c}$$

In the first two, component reliabilities were approximated by exponential forms which are accurate for high reliability components. In the third, the product and summation signs for system reliability (7) have been interchanged and the definition of component reliability (4) was used. Equality would exist in this third case if geometric averages had been considered instead of arithmetic averages. The geometric average is close to the arithmetic average for either high reliability components or if there is a relatively small variation in reliability as a function of time. Table 1 lists some of the computations performed using (8a), (8b) and (8c). Based on these and other computations, it was concluded that (8c) represents an adequate approximation to system reliability over a wide range of reliability levels. It also gives a somewhat conservative result in that a lower than true reliability is generally computed.

3. BAYESIAN SOLUTION FOR SYSTEM RELIABILITY. The next step is to formulate a Bayesian solution for average system reliability. Taking logs of (8c) gives

$$\ell n\, R_{sa}(\tau) = \sum_{j=1}^{n_c} \ell n\, R_{ja}(\tau). \tag{9}$$

This equation, in a Bayesian sense, represents $\ell n\, R_{sa}$ as a random variable which is equal to the sum of the random variables $\ell n\, R_{ja}$.

TABLE 1

APPROXIMATIONS TO AVERAGE SYSTEM RELIABILITY*

RUN NO.	AVERAGE COMPONENT RELIABILITIES	WEIBULL SHAPE PARAMETER	R_{sa}, TRUE	R_{sa}, (8a)	(Approx.) (8b)	(Approx.) (8c)
1	.99, .97, .94	3, 3, 3	.9026	.8997	.9046	.9024
2	.951, .905, .861, .819	1, 1, 1, 1	.6062	.5350	.6282	.6062
3	.951, .899, .850, .519	3, 3, 3, 3	.3804	.2191	.4580	.3773
4	.951, .899, .850, .800	3, 3, 3, 3	.5848	.4998	.6064	.5812
5	.99, .95, .90, .90, .85, .803	1, 2, 3, 4, 5, 6	.5218	.3927	.5448	.5198
6	.99, .99, .99, .99, .99, .99	1, 2, 3, 4, 5, 6	.9413	.9394	.9412	.9409

*Series System - Weibull Components

Number of Missions = 150

Mission Time = 1.0

Using the Central Limit Theorem, the posterior distribution of $\ln R_{sa}$ can be assumed to asymptotically approach the normal distribution. The mean and variance of $\ln R_{sa}$ required in the normal distribution can be derived from the means and variances of $\ln R_{ja}$:

$$E(\ln R_{sa}) = \sum_{j=1}^{n_c} E(\ln R_{ja}) \tag{10a}$$

and

$$Var\ (\ln R_{sa}) = \sum_{j=1}^{n_c} Var\ (\ln R_{ja}) \tag{10b}$$

in which statistical independence is assumed for the random variables $\ln R_{ja}$. It remains then to determine $E(\ln R_{ja})$ and $Var\ (\ln R_{ja})$ from the posterior distribution of R_{ja}. These are obtained from the following relations:

$$E(\ln R_{ja}) = -\int_0^1 (F_{R_{ja}}(z)/z)dz \tag{11a}$$

$$E(\ln^2 R_{ja}) = -2\int_0^1 (F_{R_{ja}}(z)\ln z/z)dz \tag{11b}$$

$$Var(\ln R_{ja}) = E(\ln^2 R_{ja}) - E^2(\ln R_{ja}) \tag{11c}$$

in which $F_{R_{ja}}(z)$ is given by (5).

Using the normal distribution assumption for $\ln R_{sa}$, probability limits on R_{sa} can then be determined using a standard normal table.

4. BIAS OF THE BAYESIAN ESTIMATES. From previous work and from work on similar approaches presented in the literature, the final solution for system confidence limits can be sensitive to the bias of the estimators used for component reliability. That is, if the component reliability estimates are biased, then a function of these such as system reliability can become highly biased. The statistical bias of component log reliabilities was consequently studied to

determine potential bias of the system reliability estimates.

Table 2 summarizes some of a number of results obtained for the true mean and for Bayesian mean of $\ell n\ R_{ja}$ derived from Monte Carlo simulation. Generally, the results indicate that the mean of $\ell n\ R_{ja}$ is biased on the conservative side of true reliability. Unbiasing factors were studied which are analogous to the exponential case where one failure is subtracted from the total number of failures to yield an unbiased estimate of mean failure rate [10]. Unbiasing factors obtained by subtracting 0.5 to 1.0 failures gave the best overall results for Weibull shape parameter in the range of 2 to 6 as indicated, for example, in Table 2.

5. FREQUENCY INTERPRETATION OF THE BAYESIAN INTERVALS. The final question regarding the Bayesian limits is whether or not there is a frequency interpretation of the resulting intervals with and without an unbiasing factor. In order to check this, a number of Monte Carlo simulations were conducted using the previously described confidencing procedure. Monte Carlo is not used here to derive the confidence intervals but rather to check for exactness.

Various systems of 3 and 6 components were assumed. The shape parameter, true reliability and number of failures for each component were fixed at different assumed values. Test samples were then artificially generated from random numbers using the assumed parameters. Exactness was then checked for the generated samples. For these trials the system life was 150 missions with mission time τ being equal to 1.0.

Table 3 lists some of the results of the Monte Carlo trials. In this table the Kolmogorov-Smirnov rejection error is presented [11]. This error represents significance level or risk in rejecting the hypothesis that the confidence intervals are exact at all confidence levels when in actuality the hypothesis is true. From the results given in Table 3 it can be seen that when no unbiasing factor is used, the resulting confidence intervals are not exact at all confidence levels. Using an unbiasing factor of $(n_f-0.5)/n_f$ gave the best overall results except for the case of $\beta=1.0$.

In addition to the K-S statistic given in Table 3, the relative distribution of the confidence limit was also generated to determine exactness for the lower confidenced reliability in the range of 90 to 95% confidence. It was found that in all cases considered the lower confidenced reliability is conservative when no unbiasing factor is used. That is, the proportion of the time that the true reliability

TABLE 2

SUMMARY OF COMPUTATIONS TO CHECK BIAS OF BAYESIAN MEAN OF LOG OF RELIABILITY

n_f	β	e^{μ}	$e^{\mu\left(\frac{n_f-1}{n_f}\right)}$	$e^{\mu\left(\frac{n_f-0.5}{n_f}\right)}$	True R_a
10	1.0	.879	.890	.885	.900
	2.0	.897	.907	.902	.900
	3.0	.896	.905	.901	.900
10	1.0	.940	.945	.943	.950
	2.0	.947	.953	.950	.950
	3.0	.949	.954	.951	.950
	6.0	.949	.954	.951	.950
10	1.0	.989	.9901	.9895	.990
	2.0	.9891	.9902	.9896	.990

μ = Mean of $\ell n\ R_a$ from Monte Carlo Simulation (1000 trials)

n_f = Number of Failures

β = Weibull Shape Parameter

TABLE 3

K-S TEST FOR MONTE CARLO TRIALS TO CHECK EXACTNESS OF PSEUDO-BAYESIAN INTERVALS (100 TRIALS PER CASE)

Run No.	n_c	R_j	β_j	n_{fj}	K-S, Rejection Error* No Unbiasing Factor	K-S, Rejection Error* Unbiasing Factor $= (n_f-0.5)/n_f$
1	3	.90, .90, .90	1, 2, 3	20, 20, 20	0.05	>0.20
2	3	.90, .90, .90	1, 2, 3	10, 10, 10	0.01	0.10
3	3	.90, .90, .90	1, 1, 1	20, 20, 20	0.01	0.01
4	3	.90, .90, .90	2, 2, 2	10, 10, 10	0.01	0.15
5	3	.90, .90, .90	3, 3, 3	10, 10, 10	0.01	>0.20
6	6	0.97, 0.97, 0.97, 0.97, 0.97, 0.97	1, 2, 3, 4, 5, 6	10, 10, 10, 10, 10, 10	0.01	>0.20
7	6	0.90, 0.90, 0.90, 0.90, 0.90, 0.90	1, 2, 3, 4, 5, 6	10, 10, 10, 10, 10, 10	0.01	>0.20
8	6	0.90, 0.90, 0.90, 0.90, 0.90, 0.90	1, 2, 3, 4, 5, 6	5, 5, 5, 5, 5, 5,	0.01	>0.20
9	3	0.95, 0.95, 0.95	3, 3, 3	10, 10, 10	0.01	0.15
10	3	0.99, 0.99, 0.99	2, 2, 2	10, 10, 10	0.01	>0.20

*Significance or risk in rejecting hypothesis that confidence intervals are exact at all levels when hypothesis is true.

was greater than the lower confidence limit was at least equal to the confidence level. These results hold for Weibull components with shape parameter greater than 1.0.

It is also of interest to compare the pseudo-Bayesian limits derived for Weibull components to the usual method used for determining confidenced reliability from component results. Generally, the exponential assumption is made for failure time distribution regardless what the true underlying distribution may be. The exponential assumption yields a constant failure rate (constant reliability) for components and hence for the entire system. This assumption is usually made since confidence intervals can often be derived classically [10]. Table 4 lists some results of the average lower 90% confidence limit on reliability for the pseudo-Bayesian Weibull and the classical exponential methods. From these results it can be seen that the Bayesian limits, although previously shown to be conservative, are not as conservative as the exponential limits. The degree of difference depends on true reliability and shape parameter.

6. CONCLUSIONS. Although numerically tedious, the pseudo-Bayesian method of confidencing system reliability for Weibull components described in this paper appears to be a sound approach to a problem which generally has no other solution. Two conclusions that can be made based on the results of this study are:

a. The approximation $\ell n\, R_{sa} = \sum \ell n\, R_{ja}$ for system reliability with no unbiasing factor gives a conservative lower confidence limit for all cases considered but not as conservative as assuming the exponential distribution.

b. The unbiasing factor $(n_f-0.5)/n_f$ to $(n_f-1.0)/n_f$ gives the best overall results for exact confidence limits at all confidence levels for $\beta>1.0$.

TABLE 4

AVERAGE LOWER 90% CONFIDENCE LIMIT FOR BAYESIAN-WEIBULL AND CLASSICAL EXPONENTIAL METHODS

Trial No.	No. of Comp.	True Comp. Rel.	Weibull Shape Parameter	No. of Failures Per Comp	True System Rel.	Average Lower 90% Confidence Limit on System Reliability(1)	
						Weibull(2)	Exponential(3)
1	3	.995	6.0	10	.985	.982	.965
2	3	.995	3.0	10	.985	.979	.967
3	3	.995	2.0	10	.985	.978	.968
4	3	.995	6.0	5	.985	.979	.965
5	3	.995	3.0	5	.985	.981	.968
6	3	.995	2.0	5	.985	.970	.965
7	6	.995	2.0	10	.970	.959	.941
8	10	.995	2.0	10	.951	.940	.910
9	3	.990	2.0	10	.970	.963	.956
10	3	.950	2.0	10	.857	.831	.829

(1) Average of 10 Monte Carlo trials.
(2) Pseudo-Bayesian limits described in this paper.
(3) Mann-Grubbs frequency limits [10].

REFERENCES.

[1] R. E. Barlow and F. Proschan, Mathematical Theory of Reliability, New York: Wiley, 1965, Chap. 3.

[2] D. R. Cox, Renewal Theory, New York: Wiley, 1963.

[3] E. Parzen, Stochastic Processes, San Francisco: Holden-Day, 1962, Chap. 5.

[4] B. V. Gnedenko, Yu. K. Belyayev and A. D. Solovyev, Mathematical Methods of Reliability Theory, New York: Academic Press, 1969, Chap. 2.

[5] E. Pieruschka, Principles of Reliability, Englewood Cliffs, New Jersey: Prentice-Hall, 1963, Chap. 7.

[6] D. R. Thoman, L. J. Bain and C. E. Angle, "Maximum Likelihood Estimation, Exact Confidence Intervals for Reliability, and Tolerance Limits in the Weibull Distribution", Technometrics, Vol. 12, May 1970, pp. 363-371.

[7] M. V. Johns, Jr. and G. J. Lieberman, "An Exact Asymptotically Efficient Confidence Bound for Reliability in the Case of the Weibull Distribution," Technometrics, Vol. 8, February 1966, pp. 135-175.

[8] R. L. Racicot, "Numerical Solution and Inference for Interval-Reliability of Repairable Components," to appear in IEEE Transaction on Reliability in April 1975.

[9] M. Abramowitz and I. A. Stegun, Editors, Handbook of Mathematical Functions, U. S. Dept. of Commerce, Nat. Bur. of Stds., Appl. Math. Series No. 55, Ninth Printing Nov. 1970, Formula No. 6.5.13.

[10] N. R. Mann and F. E. Grubbs, "Approximately Optimum Confidence Bounds on Series System Reliability for Exponential Time to Failure Data," Biometrika, Vol. 59, 1972, pp. 191-204.

[11] S. Siegel, Nonparametric Statistics for the Behavioral Sciences, New York: McGraw-Hill, 1956, Chap. 4.

THE UNIQUE APPLICATION OF BAYESIAN STATISTICS TO HIGH RELIABILITY TESTING

Charles A. Pleckaitis
Electronic Engineer
Erwin Biser
Operations Research Analyst
Avionics Laboratory
U.S. Army Electronics Command
Fort Monmouth, New Jersey

ABSTRACT. The paper deals with simulating the reliability predictions of Heavy Lift Helicopter (HLH) Fly-by-Wire system. Acceptance decisions are dependent on the prior distributions (based on available subjective judgments and experimental information) of the equipment Reliabilities which constitute the "states of nature" (A_i).

The prior probabilities $P(A_i)$ associated with the states of nature (reliabilities) are determined from test data, historical system reliability data predictions, and inferences from experiences with similar equipments.

The A_i's are arbitrary discrete values near the midpoints of various reliability bands. These are chosen to cover the reliability scale from 0 to 1. A_2, for example, is the event within cell No. 2 with its associated reliability of 0.999 999 99: $P(A_2) = 0.9(8)$. The cell values **(midpoints)** are chosen to be closer to one another at the higher values of reliabilities and spread apart as the reliabilities decrease.

This is done with the objective of obtaining a sharp descrimination of reliability values at the required high spectrum. This procedure will enable one to investigate the associated probabilities at the higher reliability regions. See Table I and Figure I.

Reliability tests were performed, and Bayesian algorithms were used to update the priors based on the test data. This results in a posterior distribution which is used as a new prior for the next phase of testing.

Various initial priors were applied to different sets of test data in order to observe how sensitive the priors are, after Bayesian updating, in affecting the resultant posterior distributions. Charts and graphs showing these results are presented.

Table I. Discrete Reliability Cell Formulation

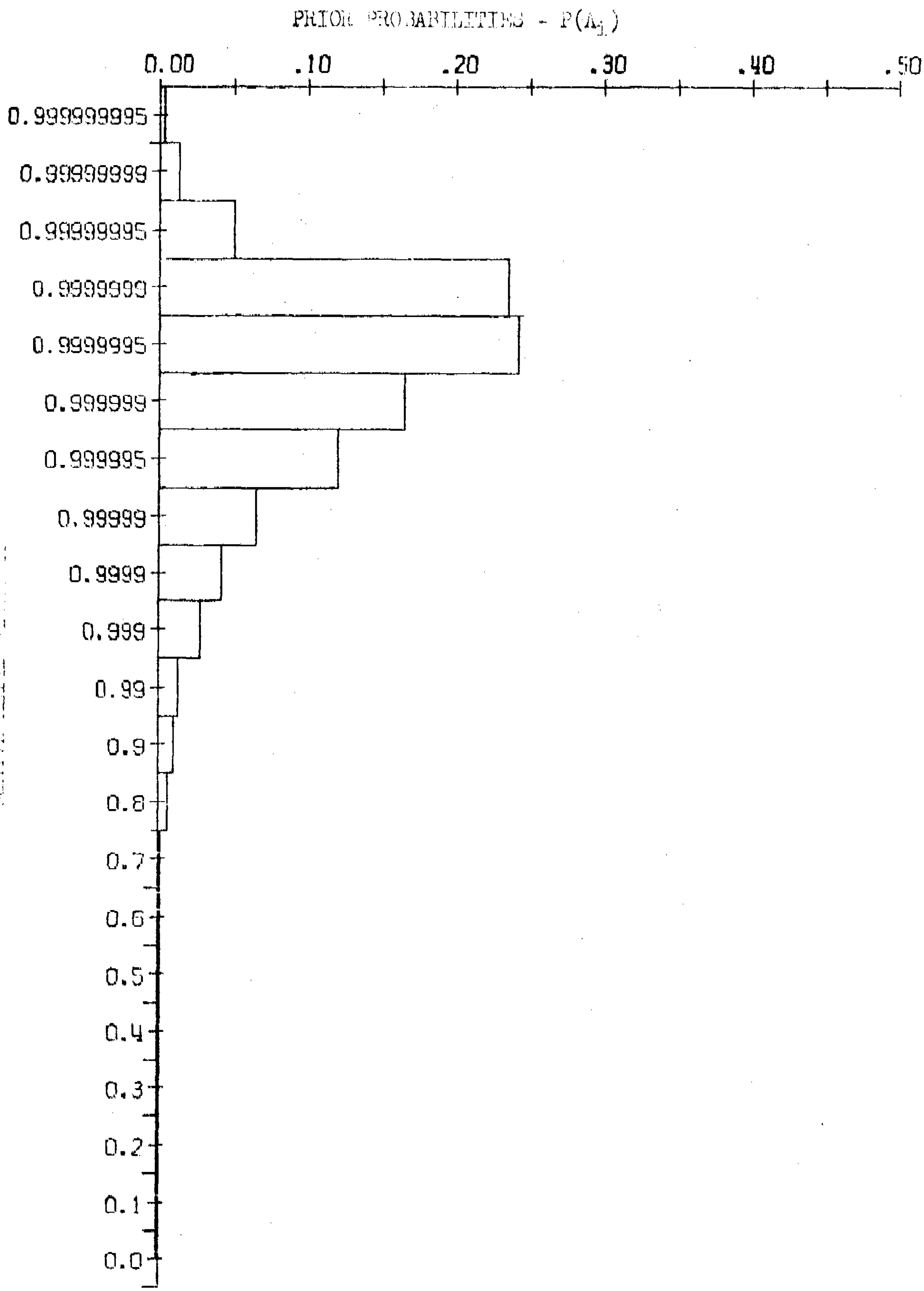
Cell No.	Cell Reliability Values	Reliability Bands
1	0.999 999 995	0.999 999 992 5 - 1.0
2	0.999 999 99	0.999 999 975 - 0.999 999 992 5
3	0.999 999 95	0.999 999 925 - 0.999 999 975
*4	0.999 999 9	0.999 999 75 - 0.999 999 925
5	0.999 999 5	0.999 999 25 - 0.999 999 75
6	0.999 999	0.999 997 5 - 0.999 999 25
7	0.999 995	0.999 992 5 - 0.999 997 5
8	0.999 99	0.999 975 - 0.999 992 5
9	0.999 9	0.999 75 - 0.999 975
10	0.999	0.997 5 - 0.999 75
11	0.99	0.975 - 0.997 5
12	0.9	0.85 - 0.975
13	0.8	0.75 - 0.85
14	0.7	0.65 - 0.75
15	0.6	0.55 - 0.65
16	0.5	0.45 - 0.55
17	0.4	0.35 - 0.45
18	0.3	0.25 - 0.35
19	0.2	0.15 - 0.25
20	0.1	0.05 - 0.15
21	0	0 - 0.05

*This is the reliability we would like to obtain

1.0 BACKGROUND. Early in the investigations of the Heavy Lift Helicopter's (HLH) development, it was recognized that certain areas of its development would be a technological challenge. The area of flight controls stimulates avionics interest due to the needs of container cargo off shore loading, Fly-By-Wire (FBW) primary control, precision hovering, and sling load stabilization.

A major technological breakthrough in the HLH is the FBW primary control. In this approach the conventional mechanical primary flight control mechanical linkage is eliminated and the pilot's stick commands the control surface actuator or swashplate through electrical signal wires. Some areas of primary concern in the use of a FBW system are: flight safety reliability demonstration, mission reliability test demonstration, and nuclear vulnerability. The FBW system is triple redundant with a flight safety reliability goal of 0.999 999 900 for two hours of flight (or a MTBF of approximately 10,000,000 hours).

2.0 INTRODUCTION. The objective of this paper is to find a technique to test demonstrate the Heavy Lift Helicopter (HLH) triple redundant Fly-By-Wire (FBW) system's flight safety reliability goal of 0.999 999 900 for two hours. It investigates by simulating the reliability predictions of the HLH FBW system.



PRIOR PROBABILITIES - P(A_i)
0.00
.10
.20
.30
.40
.50
0.999999995
0.99999999
0.99999995
0.9999999
0.9999995
0.999999
0.999995
0.99999
0.9999
0.999
0.99
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

Ideally, the purpose of a reliability demonstration technique is to establish, in the shortest possible test time and at the minimum cost, whether this high reliability goal can be met. The FBW reliability demonstration testing environment consists of:

- High Reliability Requirements
- Limited Funds
- Short Test Times Available
- Low Test Risks (which requires long test times)

In other words, it is desirable to have low producers' and consumers' risks which results in long test times due to the high reliability requirements. Generally, the accuracy of the demonstration reliability tests and the measure of test confidence increases as the number of observed failures increases. However, high reliability requirements mean long times to (observed) failures. This environment creates a cost/time problem which is apparently unsolvable by traditional Classical methods. This is exactly the dilemma involved in testing the highly reliable and expensive HLH FBW system.

When one applies the traditional Classical method to reliability demonstration tests the following characteristics emerge:

- The technique is easy to use and to understand
- It is assumed that the desired reliability R is fixed (R is actually a random variable)
- Previous (or prior) failure information is ignored (such as laboratory design and development tests, initial buyer acceptance tests, etc.)
- It requires long test times for equipments of high reliability

After investigations, the conclusion was reached that it is not practical to use traditional Classical techniques to test demonstrate systems of 0.999 999 or higher with a reasonable degree of confidence. As an alternative, Bayesian statistical techniques are considered. In this methodology it is essential that a prior distribution be used and the reliability parameter R be considered as a random variable.

3.0 BAYES THEOREM. Bayes Theorem is essentially a simple relation between probabilities of the occurrences of two different events. It will be applied in this paper to a discrete reliability cell formulation.[1]

The basic expression which describes the Bayes relationship between two events is:

$$P(A_i/B) = \frac{P(A_i)\ P(B/A_i)}{P(B)}$$

where $P(B) = \sum_{i=1}^{n} P(A_i)\ P(B/A_i)$

and the "/" is to be read "given that". This relationship in words states:

$$\begin{bmatrix}\text{New Probability}\\ \text{concerning specified}\\ \text{reliability hypothesis}\\ \text{given test data}\end{bmatrix} = \begin{bmatrix}\text{Prior Probability}\\ \text{concerning specified}\\ \text{reliability hypothesis}\end{bmatrix} \times \frac{\begin{bmatrix}\text{Probability of}\\ \text{Observed Test}\\ \text{outcome under the}\\ \text{Reliability Hypothesis}\end{bmatrix}}{\begin{bmatrix}\text{Total Probability of}\\ \text{Observed test outcomes}\\ \text{under all specified}\\ \text{Reliability Hypotheses}\end{bmatrix}}$$

The terms of the relationship are defined as follows:

A_i: States of nature; or the event of having cell number "i" with its associated reliability.

B: test event; or evidence bearing upon either the success or failure of an observed reliability test (that is, test result which has a bearing upon the credibility of the A_i events).

$P(A_i)$: Reliability associated with cell number "i", given by prior distribution assignments, before test evidence B becomes available.

---

[1] "Use of Bayes' Theorem In Its Discrete Formulation For Reliability Estimation Purposes", W.J. Mac Farland, from Transactions of the 7th Reliability and Maintainability Conference, July 1965, pages 352-365, 1968 Annual Assurance Sciences.

$P(B/A_i)$: The probability of the observed outcome B assuming the occurrence of event A_i. ["The probability that the test resulted in a success (the equipment worked satisfactorily at the end of a fixed time interval)-given that the associated cell A_i reliability is, say, 0.90-is 0.90. Also, the probability of an observed test failure B- given that the associated cell A_i reliability is, say, 0.90-is obviously 0.10].

$P(B)$: The probability of the observed event B evaluated across the entire weighted ensemble of events A_i.

$P(A_i/B)$: The posterior or modified (new) probability, or the probability assigned to events A_i as a result of the new test evidence B by the use of Bayes' technique. Thus, Bayes' Technique provides the basis to recompute $P(A_i)$ based on additional new test evidence B.

4.0 BAYESIAN DISCRETE UPDATING. Table 1 shows a discrete reliability cell formulation consisting of 21 cells and 21 reliability ranges (the number of chosen cells is arbitrary). The A_i's are arbitrary discrete values near the midpoints of the various reliability bands or ranges. These bands are chosen to cover the reliability scale from 0 to 1. The cell reliability values (midpoints), the A_i's, are selected close to one another at the desired reliability of interest; and spread apart as they diverge from the desired reliability. This is done with the objective of obtaining a sharp discrimination of reliability values for the required high spectrum. This procedure enables one to investigate the associated probabilities in the higher reliability regions.

Bayesian acceptance decisions are dependent on the prior distributions (based on available subjective judgments and experimental information) of the system reliabilities which constitute the "states of nature" (A_i).

The next step is to establish from historical and empirical data a prior reliability distribution. One must determine from existing historical and empirical failure data the probability values, $P(A_i)$, for each cell reliability value A_i.

5.0 DETERMINATION OF PRIOR PROBABILITIES. For the equipment in question, knowledgeable individuals such as the component manufacturers, the design engineers, reliability experts, and other responsible individuals should be gathered to determine a suitable prior distribution. As a hypothetical example, a component manufacturer is asked to estimate the percent of time, a thousand components (in the equipment in question) can be expected to function successfully for two hours within the cell number 1 reliability band (0.999 999 9925 to 1.0). The component manufacturer may indicate one percent, i.e.) $P(A_1) = 0.01$ based upon his available historical

TABLE (I). DISCRETE RELIABILITY CELL FORMULATION.

Cell No.	(STATES OF NATURE) A = Cell Reliability Value	Reliability Bands
1	0.999 999 995	0.999 999 992 5 - 1.0
2	0.999 999 99	0.999 999 975 - 0.999 999 992 5
3	0.999 999 95	0.999 999 925 - 0.999 999 975
*4	0.999 999 9	0.999 999 75 - 0.999 999 925
5	0.999 999 5	0.999 999 25 - 0.999 999 75
6	0.999 999	0.999 997 5 - 0.999 999 25
7	0.999 995	0.999 992 5 - 0.999 997 5
8	0.999 99	0.999 975 - 0.999 992 5
9	0.999 9	0.999 75 - 0.999 975
10	0.999	0.997 5 - 0.999 75
11	0.99	0.975 - 0.997 5
12	0.9	0.85 - 0.975
13	0.8	0.75 - 0.85
14	0.7	0.65 - 0.75
15	0.6	0.55 - 0.65
16	0.5	0.45 - 0.55
17	0.4	0.35 - 0.45
18	0.3	0.25 - 0.35
19	0.2	0.15 - 0.25
20	0.1	0.05 - 0.15
21	0	0 - 0.05

* THE DESIRED RELIABILITY

or empirical data. He is given this test for each of the 21 reliability bands with the condition that the sum of the $P(A_i)$ must equal to one (see table 2). In table (2), it should be noted, that the low $P(A_i)$'s are given values of 0.001 because an initial assignment of zero probability could never be updated by Bayes' technique regardless of what the new test data might indicate. Here each $P(A_i)$ value is designated as a probability of the corresponding cell value A_i. Estimates of $P(A_i)$ are obtained, based upon historical and empirical failure data, for each cell reliability value A_i from the design engineers, reliability group, inferences from similar equipments and other sources. These estimates, tabulated in Table (2), are then averaged and listed as Prior distribution values or Average $P(A_i)$. By examining this table, A_{10} is the event within cell reliability value number 10 (A_{10} = 0.990) with its associated reliability of 0.0275, that is, $P(A_{10})$ = 0.0275.

One can also observe from the Average $P(A_i)$ column that the A_5 cell value of 0.999 999 5 is the most probable, that is, $P(A_5)$ = 0.2425. Each value of the Average $P(A_i)$ column now forms an estimated probability distribution, called the prior distribution $P_0(A_i)$, and is shown in histogram form in figure (1). This histogram shows a prior distribution of averages. Referring to Table (2), each value of this distribution of averages is incorporated as a probability of the corresponding cell (reliability) value (A_i). For example, such a probability associated with cell A_6 (0.999 999) is 0.165 in Table (2) which plotted as the abscissa in figure 1.

The cell reliability values A_i are drawn on the histogram such that they are not proportional in width to the amount of probability or cell reliability value contained. **However, in this paper, each A_i is drawn to a uniform scale width.**

Using the histogram of figure 1 as our prior, let us consider some examples showing the results of the Bayesian updating calculations upon this prior distribution. Let us call this prior distribution the binomial type average prior.

6.0 SPECIFIC APPLICATION OF THE DISCRETE BAYESIAN TECHNIQUE TO PARTICULAR RELIABILITY TESTS. A test is performed under simulated environmental conditions which are expected to be encountered in nature. For example, a piece of Avionics equipment can be tested during actual helicopter test flights (and ground tests). The test measurements are to record the failure times which should indicate whether or not the equipment fails at the end of, say, every two hours.

TABLE 2. DETERMINATION OF PRIOR PROBABILITIES, $P(A_i)$

(STATES OF NATURE) A_i = Cell Value	P(A) PRIOR PROBABILITY ESTIMATES				
	(1) Component Manufacturer	(2) Design Engineers	(3) Reliability Group	(4) Inference From Similar Avionics Equipments	Prior Distribution. Values of Average of (1) (2) (3) & (4)
A_1 = 0.999 999 995	0.01	0.001	0.001	0.001	0.00325
A_2 = 0.999 999 99	0.03	0.001	0.02	0.001	0.0130
A_3 = 0.999 999 95	0.10	0.03	0.05	0.02	0.050
A_4 = 0.999 999 9	0.36	0.15	0.25	0.18	0.235
A_5 = 0.999 999 5	0.28	0.25	0.21	0.23	0.2425
A_6 = 0.999 999	0.10	0.20	0.16	0.20	0.165
A_7 = 0.999 995	0.04	0.16	0.12	0.16	0.120
A_8 = 0.999 99	0.03	0.08	0.07	0.08	0.065
A_9 = 0.999 9	0.02	0.05	0.04	0.06	0.0425
A_{10} = 0.999	0.01	0.04	0.03	0.03	0.0275
A_{11} = 0.99	0.01	0.01	0.02	0.01	0.0125

TABLE 2. DETERMINATION OF PRIOR PROBABILITIES, $P(A_i)$ (CONT'D)

A_i = Cell Value	P(A) PRIOR PROBABILITY ESTIMATES				
	(1) Component Manufacturer	(2) Design Engineers	(3) Reliability Group	(4) Inference From Similar Avionics Equipments	Prior Distribution Values Or Average of (1), (2), (3) & (4)
A_{12} = 0.9	0.001	0.01	0.02	0.01	0.01025
A_{13} = 0.8	0.001	0.01	0.001	0.01	0.0055
A_{14} = 0.7	0.001	0.001	0.001	0.001	0.001
A_{15} = 0.6	0.001	0.001	0.001	0.001	0.001
A_{16} = 0.5	0.001	0.001	0.001	0.001	0.001
A_{17} = 0.4	0.001	0.001	0.001	0.001	0.001
A_{18} = 0.3	0.001	0.001	0.001	0.001	0.001
A_{19} = 0.2	0.001	0.001	0.001	0.001	0.001
A_{20} = 0.1	0.001	0.001	0.001	0.001	0.001
A_{21} = 0	0.001	0.001	0.001	0.001	0.001
	$\Sigma P(A_i) = 1.0$	$\Sigma P(A_i) = 1.0$	$\Sigma P(A_i) = 1.0$	$\Sigma P(A_i) = 1.0$	$\Sigma P(A_i) = 1.0$

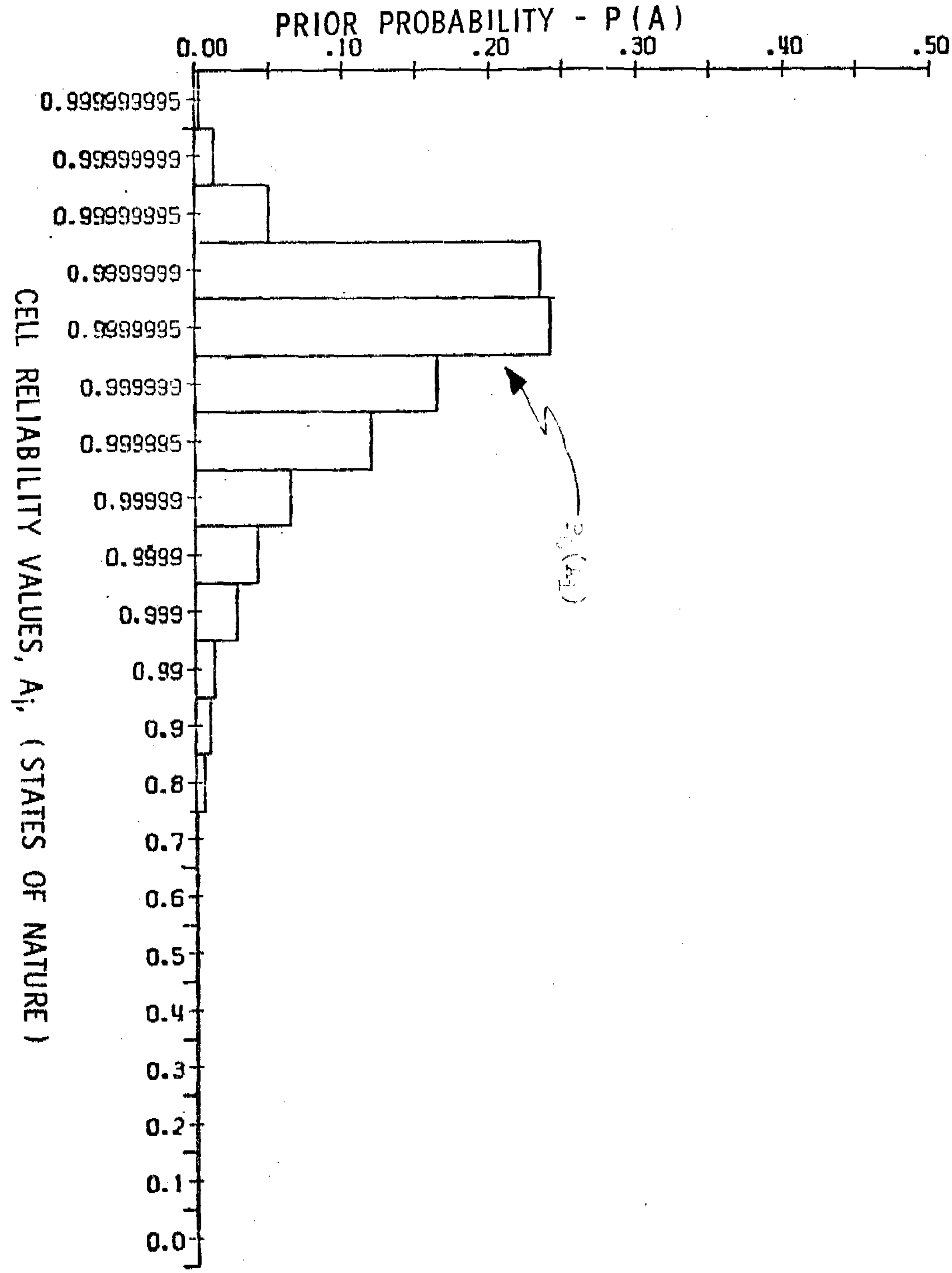


FIGURE 1. PRIOR DISTRIBUTION OF AVERAGES, A

EXAMPLE 1. The first test, at the end of two hours, resulted in a FAILURE (Test Event B).

BAYESIAN UPDATING COMPUTATIONS. We desire to update each $P(A_i/B)$ based upon the one empirical observation that the first test resulted in failure. The $P(A_i/B)$, for each i (in this case i = 1 to 21), is the set of posterior probabilities of A_i, given B. These can be used as the new priors for the next phase of testing. As an illustration, the calculations for cell values A_1, and A_{10} are given below. Table (3) shows the results of all the Bayesian cell value calculations for this example.

FOR CELL VALUE A_1

From Table(1) A_1 = 0.999 999 995, From Table(2) $P(A_1)$ = 0.00325, we want to compute $P(A_1/B)$.

Here A_1 is the event of having cell number 1 with its associated reliability of 0.999 999 995. $P(A_1)$ is the prior probability assigned to cell number 1 before test evidence B is available. In other words, it is the prior probability that the equipment reliability falls within a band of reliability (in this case 0.999 999 992 5 to 1.0) centered approximately about 0.999 999 995.

Applying Bayes' Theorem we get,

$$P(A_1/B) = \frac{P(A_1)\ P(B/A_1)}{P(B)}$$

$P(B/A_1)$ is the likelihood (the probability) of the observed test outcome B, given that cell A_1 is the case, that is, A_1 is a band of reliabilities centered about 0.999 999 995. If this is in fact true, (that is, if test B is a success), then $P(B/A_1)$ = 0.999 999 995. In other words, the best estimate (based on the success of the first test B) of the equipment reliability falling within the A_1 Reliability Band (i.e., 1.0 to 0.999 999 995) is 0.000 000 005 (1-0.999 999 995).

The Bayesian Calculations are:

$$P(A_1/B) = \frac{P(A_1)\ P(B/A_1)}{P(B)}$$

$$P(B) = \sum_{i=1}^{n=21} P(B/A_i)\ P(A_i) \quad \text{(from Table 3)}$$

$$P(B) = 0.0074833$$

TABLE (3). BAYESIAN UPDATING CALCULATIONS IF FIRST TEST RESULTS IN FAILURE.

Cell Values, A_i	P(Ai)	P(B/Ai)*	P(B/Ai) P(Ai)	$P(Ai/B) = \frac{P(A_i)P(B/Ai)}{P(B)}$
A_1 = 0.999 999 995	0.00325	0.000 000 005	1.625×10^{-11}	0.000 00
A_2 = 0.999 999 99	0.0130	0.000 000 01	1.30×10^{-11}	0.000 00
A_3 = 0.999 999 95	0.050	0.000 000 05	25×10^{-10}	0.000 00
A_4 = 0.999 999 9	0.235	0.000 000 1	0.000 000 023 5	0.000 00
A_5 = 0.999 999 5	0.2425	0.000 000 5	0.000 000 121 25	0.000 02
A_6 = 0.999 999	0.165	0.000 001	0.000 000 165	0.000 02
A_7 = 0.999 995	0.120	0.000 005	0.000 000 600	0.000 08
A_8 = 0.999 99	0.065	0.000 01	0.000 000 650	0.000 09
A_9 = 0.999 9	0.0425	0.000 1	0.000 004 25	0.000 57
A_{10} = 0.999	0.0275	0.001	0.000 027 5	0.003 67
A_{11} = 0.99	0.0125	0.01	0.000 125	0.016 70
A_{12} = 0.9	0.01025	0.1	0.001 025	0.136 97

*P(B/Ai) for test failure is given by 1-Ai where Ai is the cell value. Therefore, $P(B/A_1) = 1-A_1 = 1-0.999\ 999\ 995 = 0.000\ 000\ 005 = 5 \times 10^{-9}$.
If the test had been a success P(B/Ai) is given by Ai.

TABLE (3). Bayesian updating calculations if first test results in a failure (contd).

Cell Values, A_i	P(Ai)	P(B/Ai)*	P(B/Ai) P(Ai)	$P(Ai/B) = \frac{P(A_i)P(B/A_i)}{P(B)}$
$A_{13} = 0.8$	0.0055	0.2	0.001 1	0.146 99
$A_{14} = 0.7$	0.001	0.3	0.000 3	0.040 09
$A_{15} = 0.6$	0.001	0.4	0.000 4	0.053 45
$A_{16} = 0.5$	0.001	0.5	0.000 5	0.066 82
$A_{17} = 0.4$	0.001	0.6	0.000 6	0.080 18
$A_{18} = 0.3$	0.001	0.7	0.000 7	0.093 54
$A_{19} = 0.2$	0.001	0.8	0.000 8	0.106 90
$A_{20} = 0.1$	0.001	0.9	0.000 9	0.120 27
$A_{21} = 0.0$	0.001	1.0	0.001 0	0.133 63
	$P(B) = \sum_{i=1}^{21} P(B/A_i)P(A_i) = 0.0074833$			

*P(B/Ai) for test failure is given by 1-Ai where Ai is the cell value. Therefore, $P(B/A_1) = 1 - A_1 = 1 - 0.999\,999\,995 = 0.000\,000\,005 = 5 \times 10^{-9}$.

IF THE TEST HAD BEEN A SUCCESS P(B/Ai) IS GIVEN BY A_i.

$$P(A_i/B) = \frac{(0.00325)\ (0.000\ 000\ 005)}{(0.0074833)}$$

$$P(A_i/B) = \frac{1.625 \times 10^{-11}}{0.0074833} = 1.216 \times 10^{-13}$$

<u>For cell value A_{10}</u>

From Table 1, $A_{10} = 0.999$

From Table 2, $P(A_{10}) = 0.0275$

We want to compute $P(A_{10}/B)$.

The Bayesian Calculations are:

$$P(A_{10}/B) = \frac{P(A_{10})\ P\ (B/A_{10})}{P(B)}$$

$$P(B) = \sum_{i=1}^{n=21} P(B/A_i)\ P(A_i) = 0.0074833$$

$$P(A_{10}/B) = \frac{(0.0275)\ (0.001)}{(0.0074833)} \quad \frac{0.000275}{0.0074833}$$

$$P(A_{10}/B) = 0.00367$$

Table 3 shows the results of the remaining $P(A_i/B)$ Bayesian calculations. These $P(A_i) = P_0$, and $P(A_i/B) = P_1$ are plotted in figure 2. The P_1 posterior distribution is used as the new prior for the next test (see Table 4).

Figure 2 shows a histogram of the prior distribution updated by the Bayesian calculations <u>after a single test resulted in a failure.</u> A shift in the $P(A_i)$ probabilities is observed with the higher reliability values, 0.999 999 9 and 0.999 999 5, being shifted to 0.9 and 0.8, respectively. Also significant values of reliability occur between 0.7 down to 0.0, and the updated distribution shows a trend clustered near zero. This section indicates that even one initial test failure may radically alter a prior distribution.*

The results of Bayesian updating calculations for a second test resulting in a success after an initial failure is shown in example 2, histogram, figure 3 and table 5.

---

*It should be mentioned that the case of more than one failure in a two hour test is still to be treated as one failure in this approach. In other words, either the system functions properly or fails at the end of the two hour test.

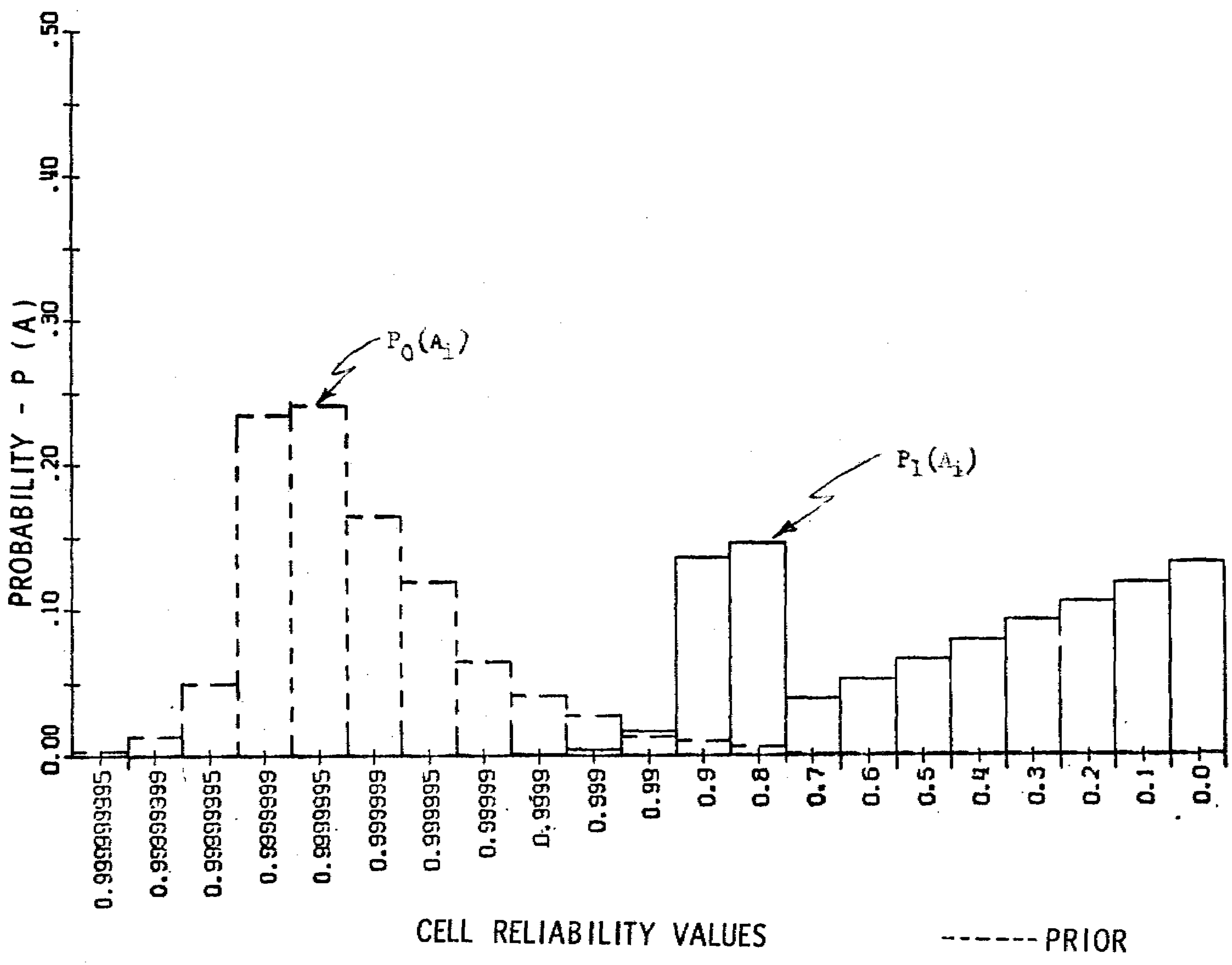


FIGURE 2. BAYESIAN UPDATED RESULTS OF ONE FAILURE (1F) ON PRIOR A + 1 F

TABLE 4. Prior and Posterior (Or Final) Distributions After First Test Results in Failure

$P(A_i)$: Prior : $P_o(A_i)$	$P(A_i/B)$: Posterior (or Final) or $P_1(A_i)$ new Prior for Next Phase of Testing
0.00325	0.000 00
0.01300	0.000 00
0.05000	0.000 00
0.23500	0.000 00
0.24250	0.000 02
0.16500	0.000 02
0.12000	0.000 08
0.06500	0.000 09
0.04250	0.000 57
0.02750	0.003 67
0.01250	0.016 70
0.01025	0.130 27
0.00550	0.140 99
0.00100	0.040 09
0.00100	0.053 45
0.00100	0.066 82
0.00100	0.080 16
0.00100	0.093 54
0.00100	0.106 90
0.00100	0.120 27
0.00100	0.133 63

NOTE: The accuracy of these Bayesian calculations were carried out to thirteen decimal places by computer. It is is used in subsequent calculations in thirteen decimals although the results are only listed as five decimals in the tables.

TABLE 5. Bayesian updating calculations if second test results in a success after an initial failure.

Cell Values, A_i	$P(A_i)$ is $P(A_i/B)$ of Table (3)	$P(B/A_i)$*	$P(B/A_i)\ P(A_i)$	$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(B)}$
0.999 999 995	0.000 00	0.999 999 995	0.000 00	0.000 00
0.999 999 99	0.000 00	0.999 999 99	0.000 00	0.000 00
0.999 999 95	0.000 00	0.999 999 95	0.000 00	0.000 00
0.999 999 9	0.000 00	0.999 999 9	0.000 00	0.000 00
0.999 999 5	0.000 02	0.999 999 5	0.000 02	0.000 04
0.999 999	0.000 02	0.999 999	0.000 02	0.000 04
0.999 995	0.000 08	0.999 995	0.000 08	0.000 18
0.999 99	0.000 09	0.999 99	0.000 09	0.000 20
0.999 9	0.000 57	0.999 9	0.000 57	0.001 27
0.999	0.003 67	0.999	0.003 67	0.008 18
0.99	0.016 70	0.99	0.016 53	0.036 84

*$P(B/A_i)$ for a test success if given by the A_i cell values. If this second test had been a failure then $P(B/A_i)$ is given by $(1-A_i)$.

TABLE 5. **Bayesian updating calculations if second test results in a success after an initial failure (contd).**

Cell Values, A_i	$P(A_i)$ is $P(A_i/B)$ of Table (3)	$P(B/A_i)$*	$P(B/A_i)\ P(A_i)$	$P(A_i)/B) = \frac{P(A_i)P(B/A_i)}{P(B)}$
0.9	0.136 97	0.9	0.123 27	0.274 59
0.8	0.146 99	0.8	0.117 60	0.261 94
0.7	0.040 09	0.7	0.028 06	0.062 51
0.6	0.053 45	0.6	0.032 07	0.071 44
0.5	0.066 82	0.5	0.033 41	0.074 42
0.4	0.080 18	0.4	0.032 07	0.071 44
0.3	0.093 54	0.3	0.028 06	0.062 51
0.2	0.106 90	0.2	0.021 38	0.047 63
0.1	0.120 27	0.1	0.012 03	0.026 79
0	0.133 63	0.0	0.000 00	0.000 00
	$P(B) = \sum_{i=1}^{21} P(B/A_i)P(A_i)$		= 0.448 93	

*$P(B/A_i)$ for a test success if given by the A_i cell values. If this second test had been a failure then $P(B/A_i)$ is given by $(1-A_i)$.

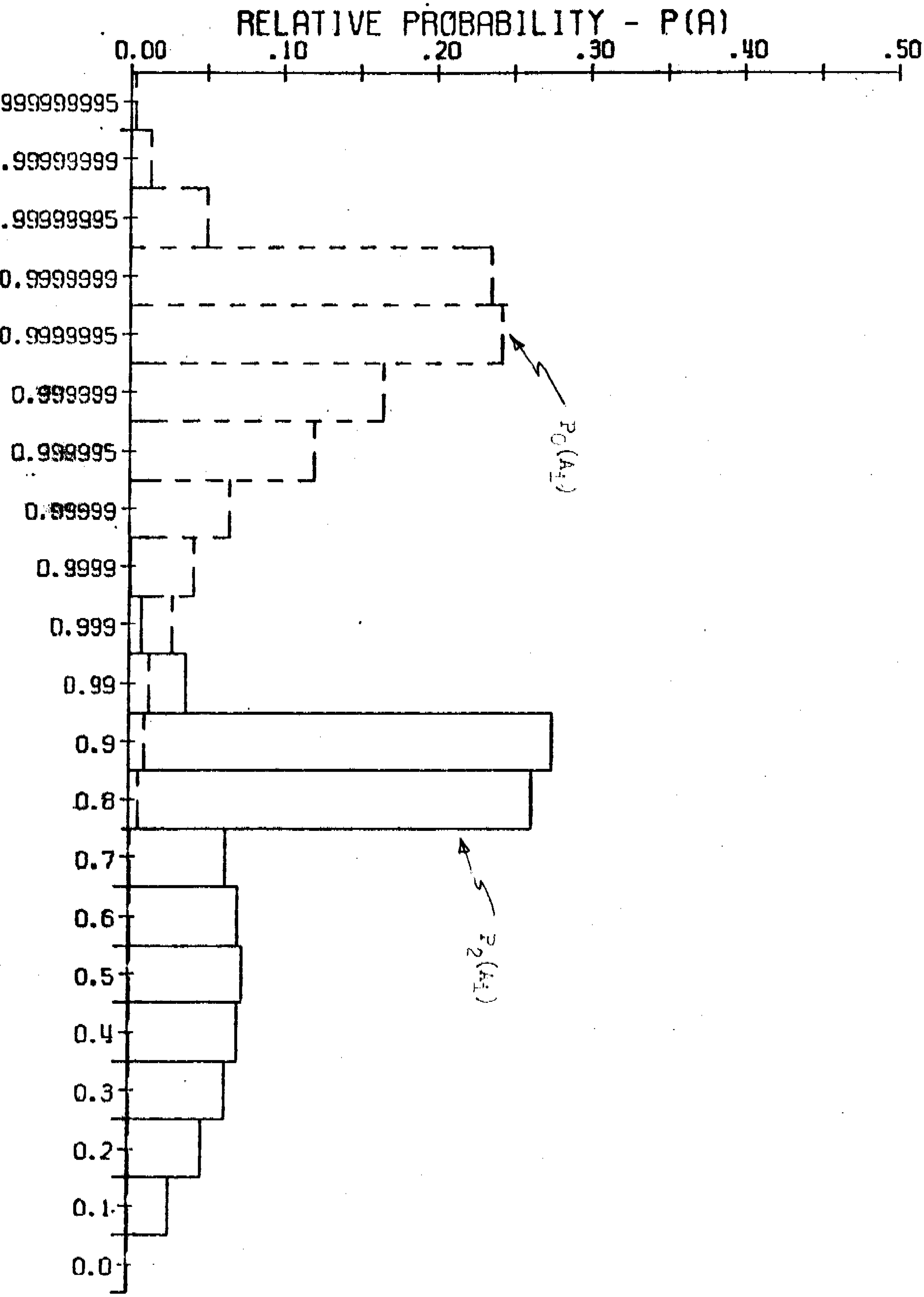


FIGURE 3. BAYESIAN RESULTS OF A SUCCESS AFTER A FIRST FAILURE, A+1F+1S

EXAMPLE 2. The second test results in a success after an initial failure. The Bayesian sample calculations follow for A_1 and A_{10} cell values using $P_1(A_i)$ as the new prior probability.

For Cell Value A_1. From Table 4, $P_1(A_1) = 0.000\ 00$. Since we cannot use zero, we must use the number calculated from example 1 which is $P(A_1/B) = 1.216 \times 10^{-13} = P_1(A_1)$. The new posterior distribution is now calculated:

$$P(A_1/B) = \frac{P(A_1)\ P(B/A_1)}{P(B)}$$

$$P(B/A_1) = 0.999\ 999\ 995$$

From Table 5, $P(B) = 0.44893$

$$P(A_1/B) = \frac{P_1(A_1)\ P(B/A_1)}{P(B)} = \frac{(1.216 \times 10^{-13})(0.999\ 999\ 995)}{(0.44893)}$$

$P(A_1/B) = 2.709 \times 10^{-13}$ (shown as 0.000 00 in Table 5)

For Cell Value A_{10}. From Table 4, $P_1(A10) = 0.00367$. The new posterior (or final) value for $P(A_{10}/B)$ is now calculated for cell $A_{10} = 0.999$ and a second test resulting in a success.

$$P(A_{10}/B) = \frac{P_1(A_{10})\ P(B/A_{10})}{P(B)}$$

From Table 5, $P(B) = 0.44893$ and $P(B/A_{10})$ is 0.999 because second test was a success.

$$P(A_{10}/B) = \frac{(0.00367)\ (0.999)}{(0.44893)} = 0.00818$$

Table 5 shows the results of the remaining $P(A_i/B)$ Bayesian calculations which are plotted in histogram form in figure 3 along with its prior. By observing this histogram, one can see that the reliability cells are now shifted away from 0.0 and a signficant increase in relative probability values is observed at 0.9 and 0.8.

Table 6 shows the flow of data for each cell from the Prior Distribution to second test a success following a first test failure.

TABLE 6. Prior and Final Distributions After Second Test is a Success and First Test was a Failure.

PRIOR DISTRIBUTION $P_{\theta}(A_i)$, i=1 to 21 Figure 1	FIRST TEST A FAILURE $P_1(A_i)$ Figure 2	SECOND TEST A SUCCESS $P_2(A_i)$ Figure 3
0.00325	0.000 00	0.000 00
0.01300	0.000 00	0.000 00
0.05000	0.000 00	0.000 00
0.23500	0.000 00	0.000 00
0.24250	0.000 02	0.000 04
0.16500	0.000 02	0.000 04
0.12000	0.000 08	0.000 18
0.06500	0.000 07	0.000 20
0.04250	0.000 57	0.001 27
0.02750	0.003 67	0.008 18
0.01250	0.016 70	0.036 04
0.01025	0.136 97	0.274 59
0.00550	0.146 99	0.261 94
0.00100	0.040 09	0.062 51
0.00100	0.053 45	0.071 44
0.00100	0.066 82	0.074 42
0.00100	0.080 18	0.071 44
0.00100	0.093 54	0.062 51
0.00100	0.106 90	0.047 63
0.00100	0.210 27	0.026 79
0.00100	0.133 63	0.000 00

7.0 VARIOUS PRIOR DISTRIBUTIONS AS APPLIED TO DIFFERENT SETS OF TEST DATA. Other initial prior distributions, termed "A", are chosen in order to indicate how sensitive the selection of prior failure distributions are to the final resultant distributions after Bayesian updating calculations. In other words, the following histograms are used to illustrate the degree change of prior distribution that occurs after Bayesian updating as new reliability test data becomes available. Table 7 gives the reliability values at each reliability cell, A_i, for the following prior distributions: Binomial type average distribution of table 2; Uniform distribution; Peaked (at 4th cell mid value) distribution; Peaked (at 7th cell mid value) distribution; and Skewed Binomial type distribution. Figures 1, 4, 5, 6, and 7 show relative frequency histogram graphs of each of these prior distributions.

The tables and histogram graphs of Appendix A will be used to illustrate the change that occurs in the Bayesian estimates as new reliability test data becomes available. The Bayesian calculations were performed with an EAI 8400 digital computer with a double precision calculation accuracy to thirteen decimal places.* This accuracy is necessary because it is important to use extremely small numbers for each Bayesian calculation without rounding them to zero.

The results of Bayesian updating calculation of ten tests with various success/failure combinations for various initial prior distributions are shown in tables A.1 to A.5 and in histogram figures A.1 to A.25 inclusive. In the cases where no failures occurred in ten tests (figures A.1, A.6, A.11, A.16, A.21 it can be observed that the final histogram data (solid lines) is generally shifted to the left or higher reliability regions with a slight increase in the relative probability values at the higher reliability cell values. It can be seen that the selection of prior distributions (dotted lines) biases the results of the Bayesian calculations if all ten tests are successes, and the final histogram looks very similar to the prior.

The Bayesian calculations which updated the Binomial Type Prior Distribution for ten two-hour tests with various success/failure combinations are in table A.1 and are plotted in figures A.1 to A.5 inclusive. Figure (A.2) shows the updated Bayesian results (solid lines) of one failure in ten tests upon this binomial prior (dotted lines). One would intuitively

---

*See Appendix B for Computer Program

	PRIOR DISTRIBUTIONS				
Cell Reliability Values, A_i	Binomial Type Avg. Prior of Table 2	Uniform Prior Distribution	Peaked Prior (at 4th cell)	Peaked Prior (at 7th cell)	Skewed Binomial Prior
0.999 999 995	0.00325	0.1400	0.002	0.002	0.0010
0.999 999 99	0.01300	0.1400	0.002	0.002	0.0010
0.999 999 95	0.0500	0.1400	0.002	0.002	0.4800
0.999 999 9	0.2350	0.1400	0.960	0.002	0.1900
0.999 999 5	0.2425	0.1400	0.002	0.002	0.1000
0.999 999	0.1650	0.1400	0.002	0.002	0.0660
0.999 995	0.1200	0.1400	0.002	0.960	0.0470
0.999 99	0.0650	0.0018	0.002	0.002	0.0300
0.999 9	0.0425	0.0014	0.002	0.002	0.0200
0.999	0.0275	0.0014	0.002	0.002	0.0150
0.99	0.0125	0.0014	0.002	0.002	0.0110
0.9	0.01025	0.0014	0.002	0.002	0.0100
0.8	0.00550	0.0014	0.002	0.002	0.0090
0.7	0.0010	0.0014	0.002	0.002	0.0050
0.6	0.0010	0.0014	0.002	0.002	0.0040
0.5	0.0010	0.0014	0.002	0.002	0.0030
0.4	0.0010	0.0014	0.002	0.002	0.0020
0.3	0.0010	0.0014	0.002	0.002	0.0020
0.2	0.0010	0.0014	0.002	0.002	0.0020
0.1	0.0010	0.0014	0.002	0.002	0.0010
0.0	0.0010	0.0014	0.002	0.002	0.0010

Table 7 Initial Prior Failure Distributions, A.

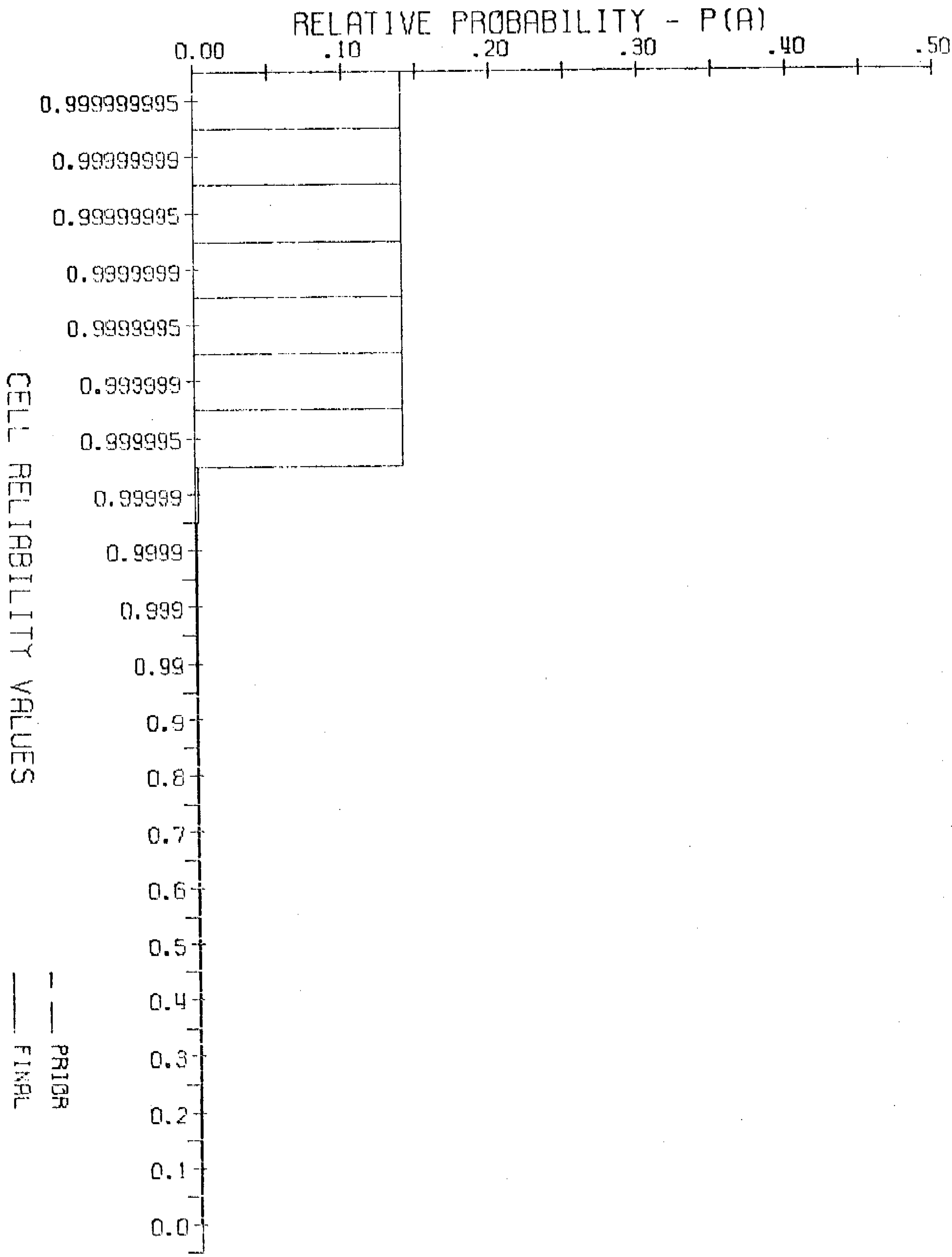


FIGURE (4) UNIFORM PRIOR DISTRIBUTION

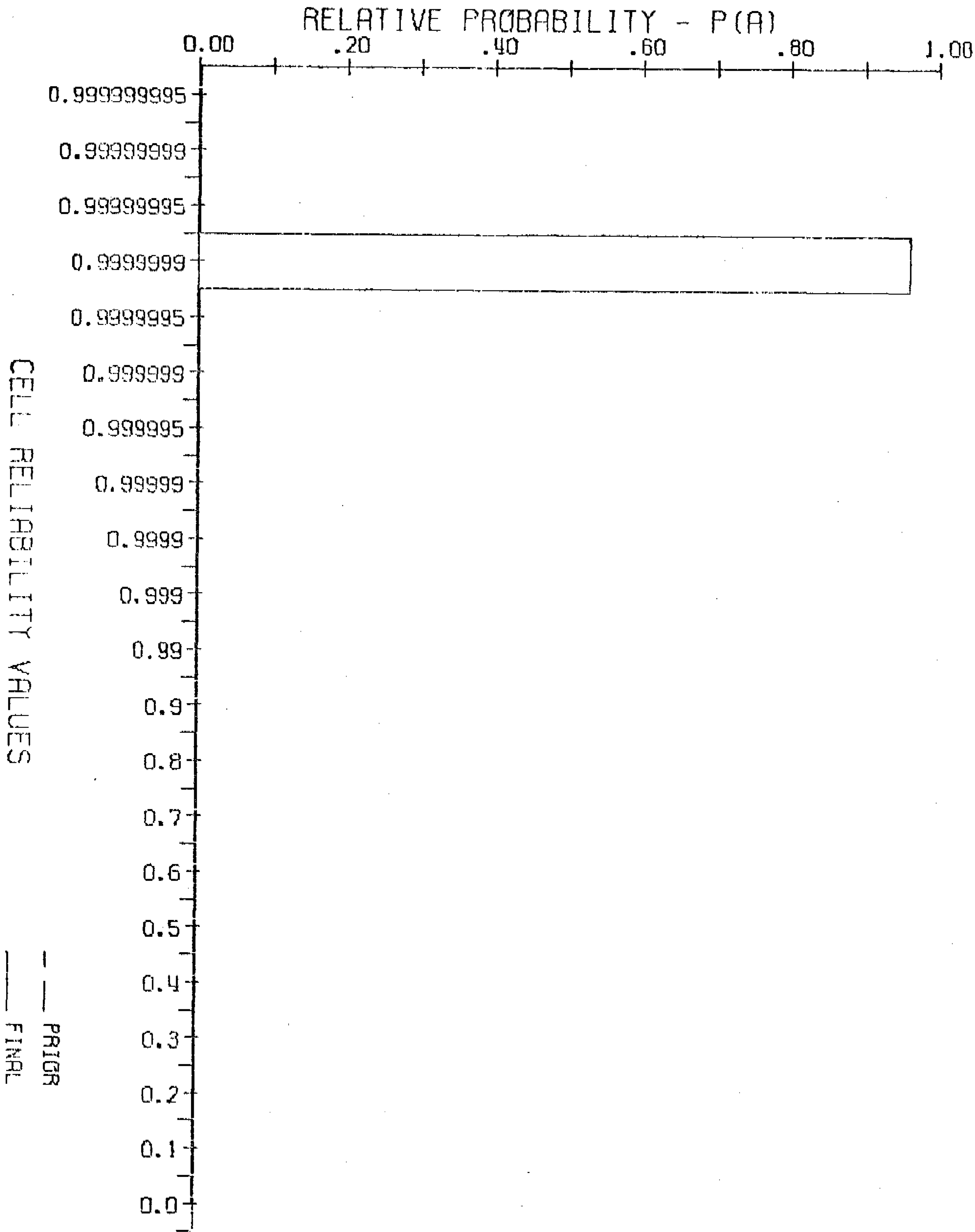


FIGURE (5) PEAKED (AT FOURTH CELL) PRIOR DISTRIBUTION

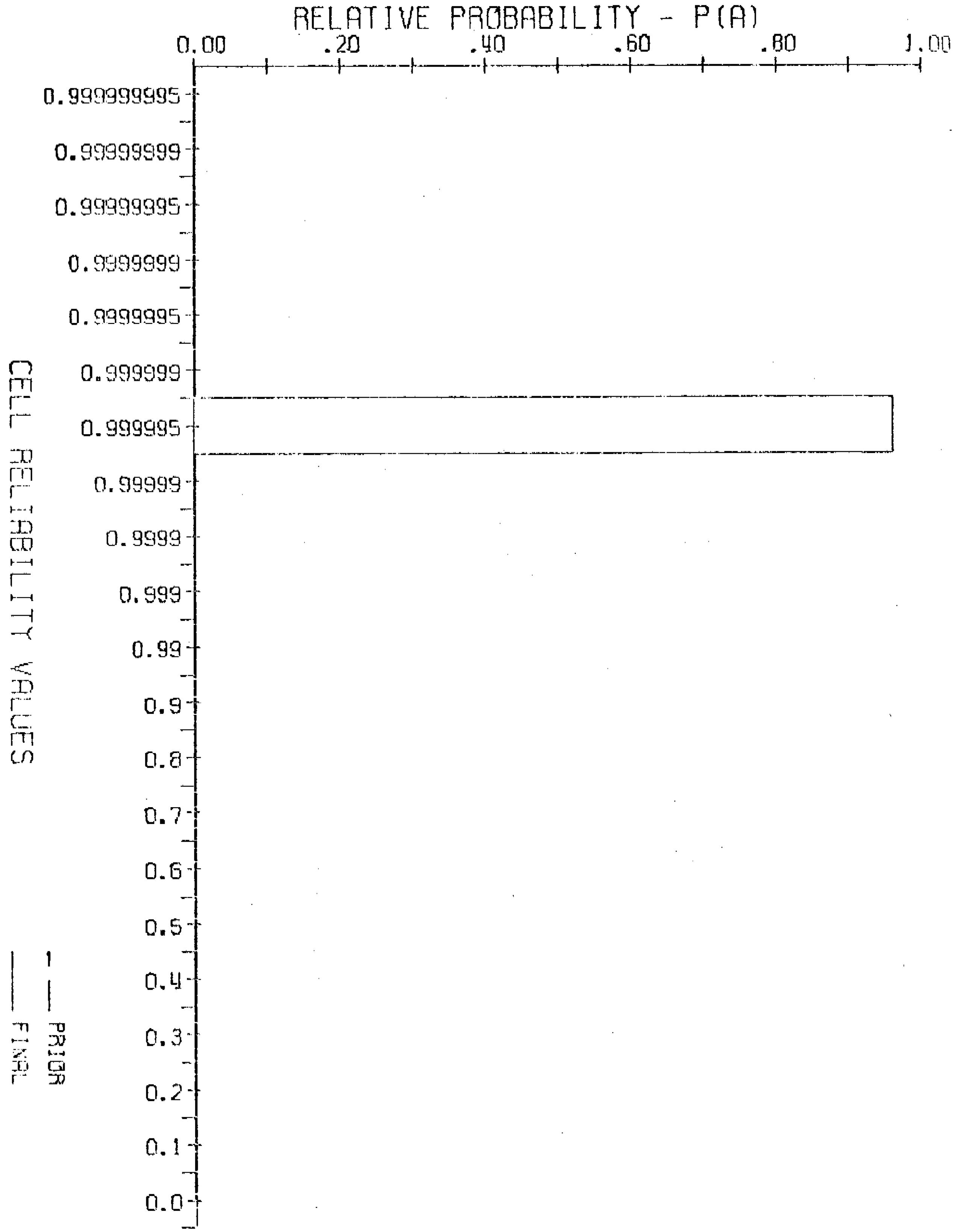


FIGURE (6) PEAKED (AT SEVENTH CELL) PRIOR DISTRIBUTION

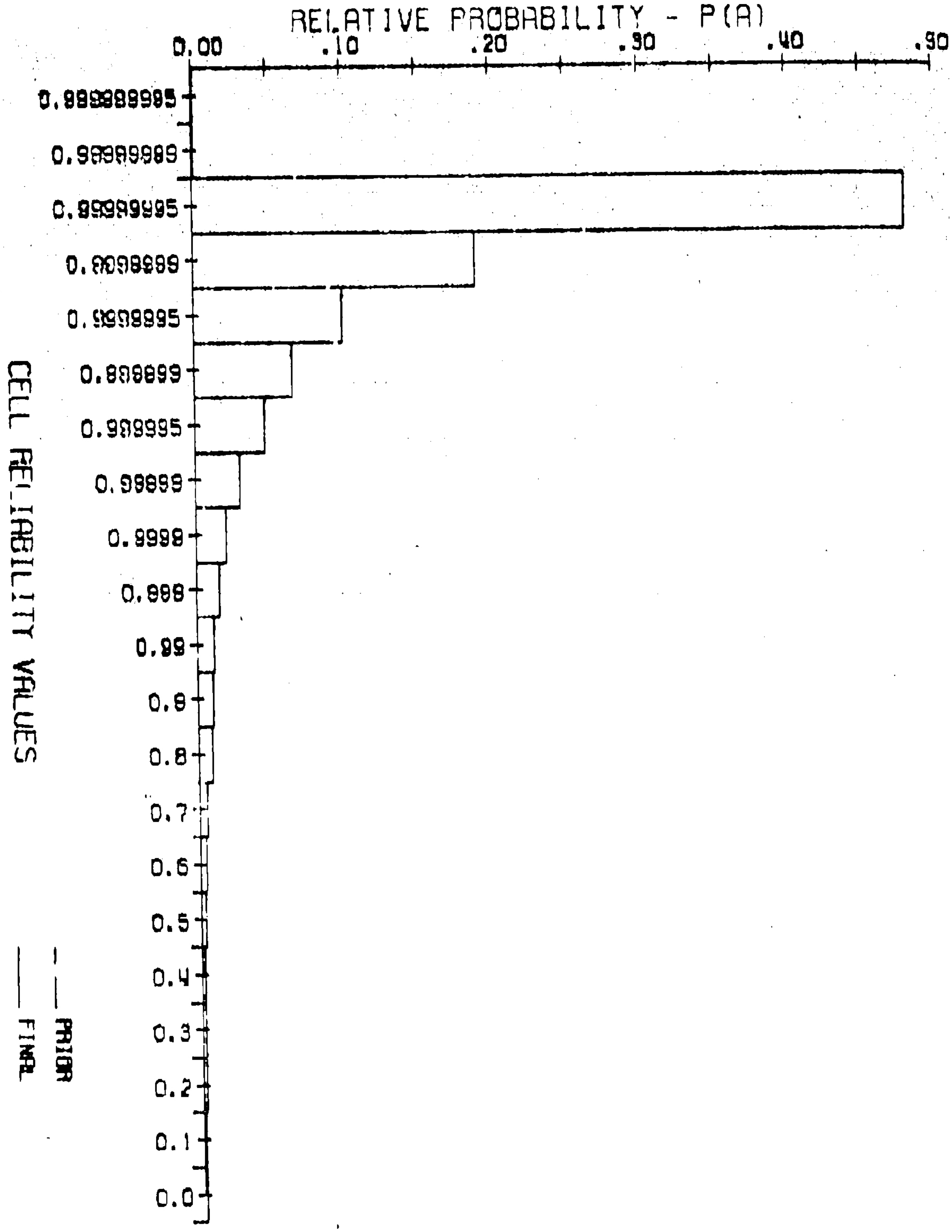


FIGURE (7) SKEWED BINOMIAL TYPE PRIOR DISTRIBUTION expect the reliability distribution to be centered around 0.9 which is clearly indicated in the figure. Similarly, for the Bayesian results of two failures in ten tests one would expect the reliability point estimate to be near 0.8. For two failures in ten tests, figure (A.3) shows most of the data to be distributed around 0.9 because the prior distribution biases the final distribution towa~'s itself. Since the prior distribution is : the high reliability region, in this case the final distribution did not completely eliminate the prior with only ten tests. For five failures in ten tests, the inference is that most of the final distribution should be near the reliability point estimate of 0.5. Figure (A.4) shows the final distribution for five failures in ten tests. Most of the data is distributed about the 0.5 cell reliability value with a secondary relative probability peak at 0.8 due to the biasing or weighting caused by the prior distribution. The histogram in figure (A.5) shows the results of ten failures in ten tests. Here the inference is obvious since most of the final distribution should be about the 0.0 cell reliability value. Thus, ten failures in ten tests have radically altered the prior distribution, as should be expected.

Table A.1 shows the results of Bayesian calculations upon the following tests 1F + 9S (one failure followed by nine successes), 4S + 1F + 5S (four successes followed by one failure and then five successes), and 9S + 1F (nine successes followed by one failure). In general, the results of the Bayesian calculations, given any one prior, are the same with any equal number of total tests regardless of the order in which the test data is received.

The Bayesian results which update the Uniform Prior Distribution for ten two-hour tests with various success/failure combinations are given in table A.2 and are plotted in histogram form in figures A.6 to A.10 inclusive. The final histogram data for this prior is distributed around the reliability point estimates which one would expect to obtain for the particular test results indicated.

The Bayesian results for the Peaked (at fourth cell) Prior Distribution are given in table A.3 and plotted in figures A.11 to A.15 inclusive for ten tests with various success/failure combinations. Again the final histogram data for this prior is distributed around the reliability point estimates or class intervals which one would expect for the test results given.

Table A.4 and figures A.16 to A.20 inclusive show the Bayesian results for the Peaked (at seventh cell) Prior Distribution for ten tests with various success/failure combinations. The final histogram data is again distributed around the cell reliability values expected except for figure A.17. This final distribution for one failure in ten tests shows approximately 2.5 percent of the data distributed in cell value 0.999 995 due to prior distribution biasing.

For the Skewed Binomial Prior Distribution with various success/failure combinations of ten tests, the Bayesian results are given in table A.5 and are plotted in figures A.21 to A.25 inclusive. The final histogram data is distributed as would be expected except for the case of five failures in figure A.24. Here the data is slightly biased by the prior to the left towards the prior.

To illustrate the influence of the prior distribution upon the final results of the updating Bayesian calculations and estimates from new reliability test data, consider next the results of one failure in 100, 500, 1000, 5000, and 10,000 tests for each prior distribution. The Bayesian results of these larger number of tests are given in tables A.6 to A.10 and histogram figures A.26 to A.50 inclusive. It can be observed that the prior distribution in most cases in this paper does not bias the final histograms to any large degree except in figures A.32, A.33, A.34, A.35, and A.39 to A.45 inclusive. Let us briefly summarize the results of this test data upon our various prior distributions.

Given a binomial prior distribution (figure 1) with the most probable cell at 0.999 999 5, it can be observed that, as the number of tests increase and the percent defective decreases (figures A.26, A.27, A.28, A.29, A.30) the final or posterior distribution will be gradually shifted towards the ordinate or in the direction of increasing reliability. Concurrently, the most probable cell value of the posterior distribution will change from 0.99 to 0.9999. However, if our test results indicate poor reliability (far to the right of the binomial prior) the binomial prior will be almost entirely washed out.

Given a uniform prior distribution (figure 4) it can be observed that as the number of tests increase and the percent defective decreases (figures A.31, A.32, A.33, A.34,A.35) the posterior distribution will be gradually shifted towards the ordinate or in the direction of increasing reliability. Concurrently, the most probable cell of the posterior distribution will change from .99 to .999 995. However, if the test results indicate poor reliability (far to the right of the uniform prior), the uniform prior will be reduced in amplitude or almost refuted. There are instances where two peaks appear in the posterior or final distribution. This is caused by the biasing of the uniform prior, particularly in the range of high reliability. In these cases, the test data results are starting to reinforce the last uniform prior cells.

Given a peaked (at 4th cell) prior distribution (figure 5) it can be observed that as the number of tests increase and the percent defective decreases (figures A.36, A.37, A.38, A.39, A.40) the posterior distribution will be gradually shifted towards the ordinate or in the direction of increasing reliability. Assuming that the reliability increases during testing and approaches the fourth cell (0.999 999 9), the peaked prior distribution is reinforced. Concurrently, the most probable cell of the posterior distribution (cell with highest reliability) will change from 0.99 to 0.999 999 9. For the 5000 and 10,000 test cases, the large jump of one posterior or final peak to 0.999 999 9 (or the fourth cell) is the result of the selection of the prior with most of the relative probability or amplitude in the fourth cell. Here the test data inferences or results are starting to reinforce the prior distribution. Finally, if the test results indicate poor reliability (that is, far to the right of the fourth cell), the peaked prior will be totally washed out.

Given a peaked (at 7th cell) prior distribution (figure 6) it can be observed as the number of tests increase and the percent defective decreases (figures A.41, A.42, A.43, A.44, A.45) the posterior distribution will be gradually shifted towards the ordinate or in the direction of increasing reliability. Assuming the reliability increases during testing and approaches the seventh cell (0.999 995) the peaked prior distribution is reinforced. Concurrently, the most probable cell of the posterior distribution will change from 0.99 to 0.999 995. The centering of posterior information or data in the seventh cell (.999 995) is the result of the prior being placed almost entirely in the seventh cell, and the fact that the test data inferences are starting to reinforce the prior. Also, if the test data results indicate poor reliability (that is, far to the right of the seventh cell), this peaked prior will be totally washed out.

Given a skewed binomial prior distribution (figure 7) with the most probable cell at 0.999 999 95, it can be observed that as the number of tests increases and the percent defective decreases (figures A.46, A.47, A.48, A.49, A.50) the posterior distribution will be gradually shifted towards the ordinate or in direction of increasing reliability. Concurrently, the most probable cell of the posterior distribution will change from 0.99 to 0.999 9. However, if our test results indicate poor reliability (that is, far to the right of the prior cells), then the skewed binomial prior will be almost entirely washed out.

In order to indicate the Bayesian results of no failures for various test data upon several prior distributions, tables A.11, A.12, and A.13, and figures A.51, A.52, and A.53 have been generated. The conclusion is that the final histograms look very similar to the prior distribution if all tests are successes and the number of tests are not greater than the indicated cell reliability bands order of magnitude. In other words, if the reliability of the sample tested is greater than the most probable cell or class interval of the prior distribution, then the final distribution will be distributed around the reliability of the sample tested. For example, if the most probable cell of the prior is (0.99 and 10,000 tests are completed with no failures), one can expect the final distribution's most probable cell value to be 0.999 9.

The following conclusions can be reached concerning the Bayesian Discrete Reliability inference procedure. In the discrete reliability cell formulation, the choice of the number of cells, the location of the cell reliability values, and cell boundaries are all arbitrary when covering the complete reliability scale of numbers from zero to unity and can be chosen to suit the user. The time for each test is picked to be equal to the defined or specified equipment reliability time interval. For example, if we choose or require our equipment or mission reliability to be 0.99 per hour, then each test or data point is to be one hour. If we require a reliability of 0.98 for three hours of operation, then each test is to be three hours long. If more than one failure is recorded in our test time interval, the failures are still to be treated as one failure. In this approach either the system functions properly or fails during our test time interval. The final or posterior distribution will not be affected by the order of failure occurrence. In other words, the Bayesian results, given any one prior, are the same with the same total number of tests regardless of the order in which the test data is received.

An important issue that must be resolved in this application of the Bayes' Technique is the choice of an appropriate prior distribution. From this report, it is clear that the prior distribution can bias the final distribution in certain cases. From the histograms it is evident that the least biasing type of prior distribution is the binomial type. There are some questions that must be answered in order to postulate appropriate and/or usable prior distributions. What does the prior distribution look like? Is it to be a continuous or discrete one? For a particular choice of prior distribution, can a bound be placed on the error (in the probability sense) one can expect in the final distribution? In the implementation of the Bayesian statistics in reliability testing, can agreement be reached between the contractor and government on the selection of an approriate prior distribution? These are just a few of the questions that should be answered before a prior distribution is to be selected.

In this Bayesian discrete approach the final distribution is available for examination after each test. In other words, the final distribution can be observed after each laboratory and experimental tests (each prototype and production flight tests) and updated at each of the program phases on a continuing basis. Finally, flight test operational data can be used by the Bayesian method to update the previous final (posterior) reliability distribution.

8.0 CONCLUSIONS.

a. Based on analyses of traditional Classical reliability test demonstration methods, it is not feasible to demonstrate with a reasonable confidence the reliability of equipments with reliabilities of 0.999 999 or higher. This is due to the excessive amount of test time required.

b. The application of Bayesian Statistics to high reliability demonstration testing, though promising, has not been conclusively demonstrated.

c. The method of assigning and choosing the prior distribution based upon existing historical failure data and existing test data requires additional investigations.

d. The application of the Bayesian methodology to high reliability testing is contingent on an appropriate prior distribution.

e. Present Classical reliability test demonstration techniques can only be reasonably applied to each nonredundant subsystem of a triple redundant system of high reliability. This triple redundant system requires only one out-of-three subsystems to function properly for successful operation.

9.0 RECOMMENDATIONS.

a. Research efforts are necessary to provide a technique of determining optimal prior distribution by combining pertinent historical and test failure data such as analytical predictions, engineering judgments, and inferences from similar equipments.

b. Bayesian methodologies should be applied to actual equipments (existing and/or new). A comparison of the actual data and that projected by the Bayesian technique should be performed.

c. Apply the Bayesian methodology to each phase of equipment analyses and testing on an **iterative** basis. For example, establish an equipment's prior distribution say P_0 from (a) above. Next apply the Bayesian updating calculations to available laboratory design failure data to result in a final distribution P_1 as shown below:

PRIOR DISTRIBUTION		FINAL DISTRIBUTION
P_0	BAYESIAN UPDATING CALCULATIONS →	P_1 (Lab Design Failure Data)
$P_{1,0}$	→	P_2 (Prototype Lab Test Data)

PRIOR DISTRIBUTION		FINAL DISTRIBUTION
$P_{2,0}$	⟶	P_3 (Acceptance Test Data)
⋮		⋮
$P_{n-1,0}$	⟶	P_n (Final Flight Operational Test Data)

The final distribution P_1 is now used as the prior distribution called $P_{1,0}$ to be updated by prototype laboratory test failure data resulting in a new final distribution P_2. This procedure is continued on an iterative basis using acceptance and production test data to update each prior distribution. In other words, the final or posterior reliability distribution can be obtained after each laboratory and experimental test, each prototype and production flight test, and updated at each of the program phases on a continuing basis.

ACKNOWLEDGEMENT

Many thanks to Marietta Bowlin for preparing and correcting this manuscript.

APPENDIX A

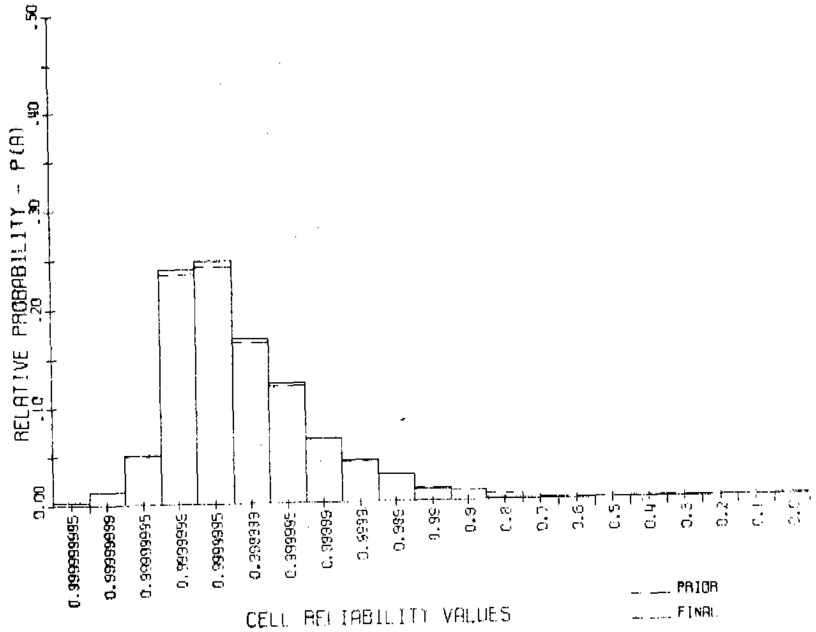
POSTERIOR DISTRIBUTIONS DERIVED FROM VARIOUS HYPOTHETICAL PRIOR DISTRIBUTIONS AND TEST DATA

Cell Reliability Values A_i	10 Two Hour Tests with Binomial Type Average Prior Distribution of Table 2				
	10S Tests	1F +9S Tests / 4S +1F+5S Tests / 9S +1F Tests	2F +8S Tests / 3S +2F+5S Tests / 8S +2F Tests	5F +5S Tests / 5S +5F Tests	10F Tests
0.999 999 995	0.0033	0.0000	0.0000	0.0000	0.0000
0.999 999 99	0.0133	0.0000	0.0000	0.0000	0.0000
0.999 999 95	0.0511	0.0000	0.0000	0.0000	0.0000
0.999 999 9	0.2401	0.0000	0.0000	0.0000	0.0000
0.999 999 5	0.2477	0.0002	0.0000	0.0000	0.0000
0.999 999	0.1686	0.0002	0.0000	0.0000	0.0000
0.999 995	0.1226	0.0008	0.0000	0.0000	0.0000
0.999 99	0.0664	0.0009	0.0000	0.0000	0.0000
0.999 9	0.0434	0.0060	0.0000	0.0000	0.0000
0.999	0.0278	0.0384	0.0003	0.0000	0.0000
0.99	0.0115	0.1610	0.0126	0.0000	0.0000
0.9	0.0037	0.5599	0.4831	0.0146	0.0000
0.8	0.0006	0.2082	0.4041	0.1395	0.0000
0.7	0.0000	0.0171	0.0568	0.0988	0.0000
0.6	0.0000	0.0057	0.0294	0.1926	0.0001
0.5	0.0000	0.0014	0.0107	0.2362	0.0007
0.4	0.0000	0.0002	0.0026	0.1926	0.0041
0.3	0.0000	0.0000	0.0004	0.0988	0.0189
0.2	0.0000	0.0000	0.0000	0.0254	0.0720
0.1	0.0000	0.0000	0.0000	0.0014	0.2338
0.0	0.0000	0.0000	0.0000	0.0000	0.6705

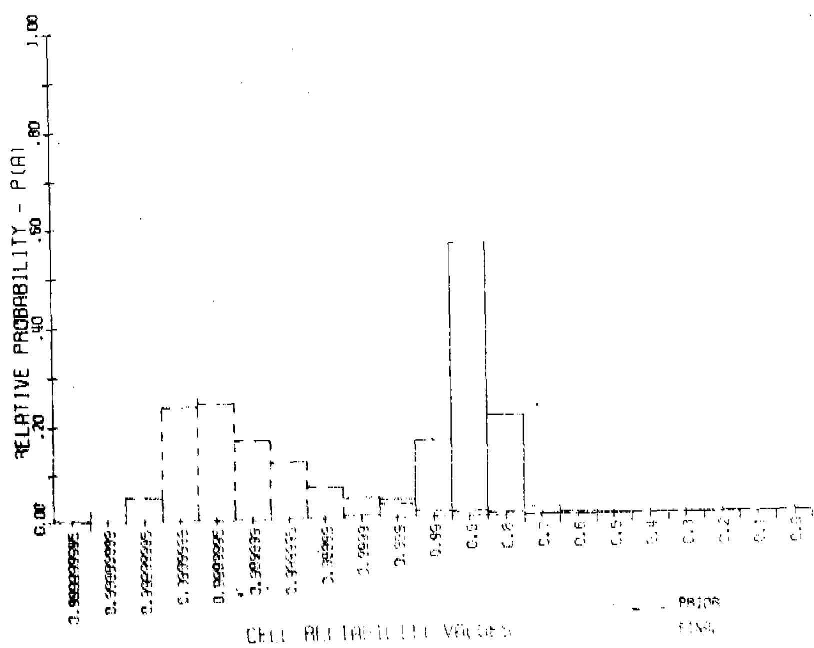
S ≡ Success

F ≡ Failure

Table A.1 Bayesian Calculation Results of Ten Two Hour Tests with Various Success/Failure Combinations, for Binomial Type Prior.



FIGURE(A.1).BINOMIAL PRIOR.A.RESULTS OF TEN TESTS.NO FAILURES A+10S



FIGURE(A.2).BINOMIAL PRIOR.A.RESULTS OF TEN TESTS.ONE FAILURE

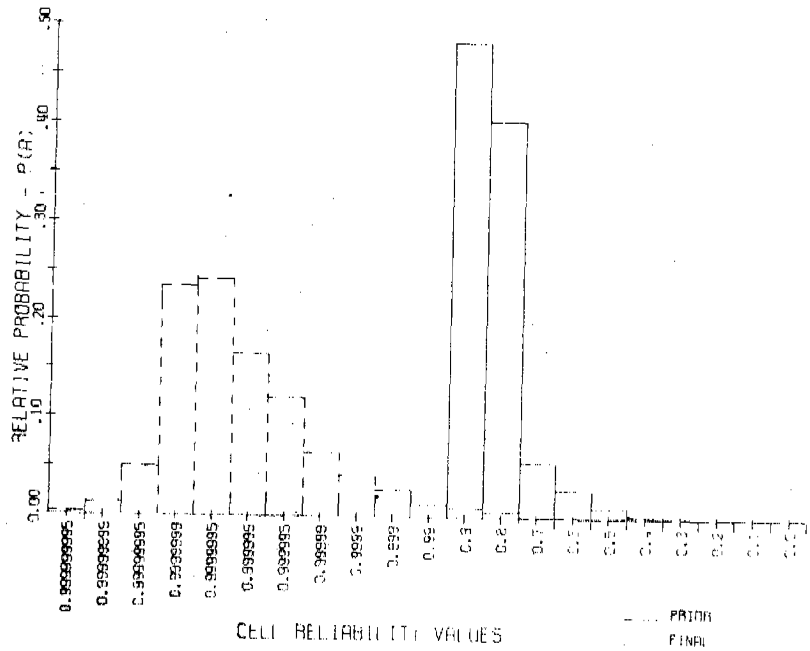


FIGURE (A.3).BINOMIAL PRIOR.A.RESULTS OF TEN TESTS.TWO FAILURES (r=2)(.85

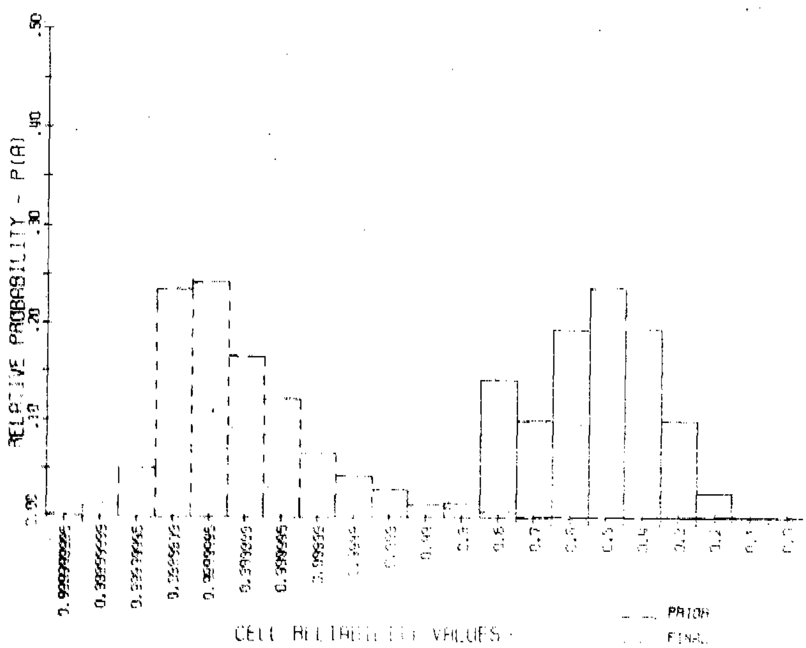
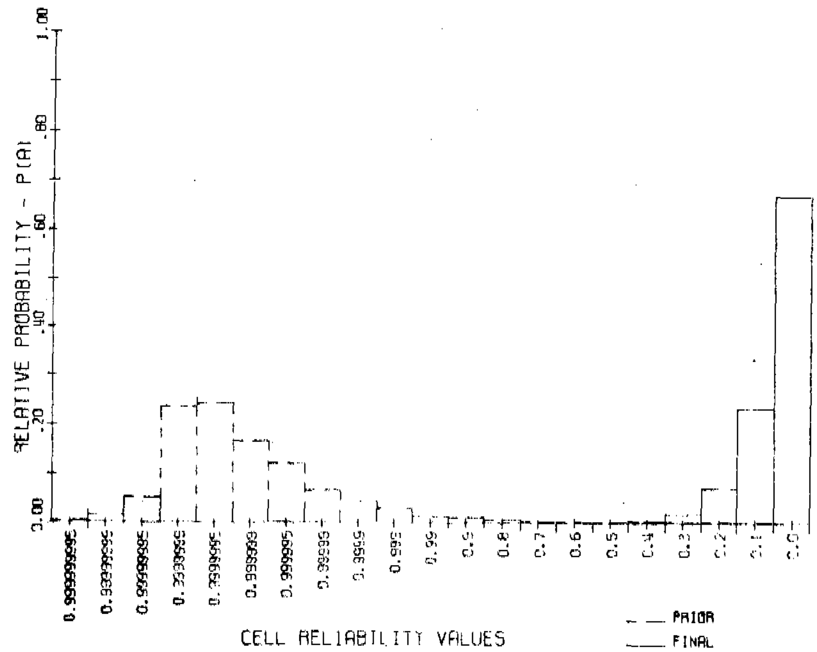
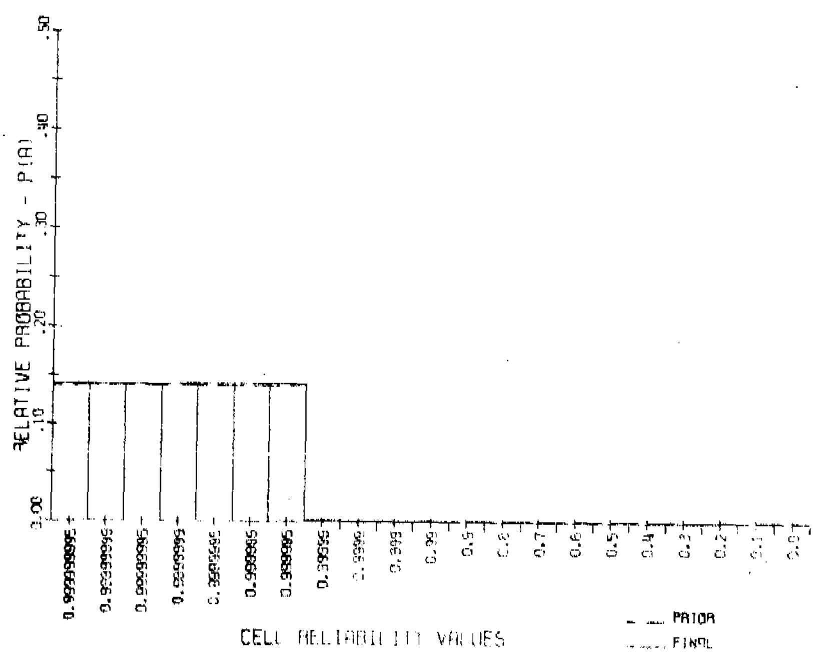


FIGURE (A.4).BINOMIAL PRIOR.A.RESULTS OF TEN TESTS.FIVE FAILURES (r=5)(.55



FIGURE(A.5).BINOMIAL PRIOR.A.RESULTS OF TEN TESTS.TEN FAILURES A+10F



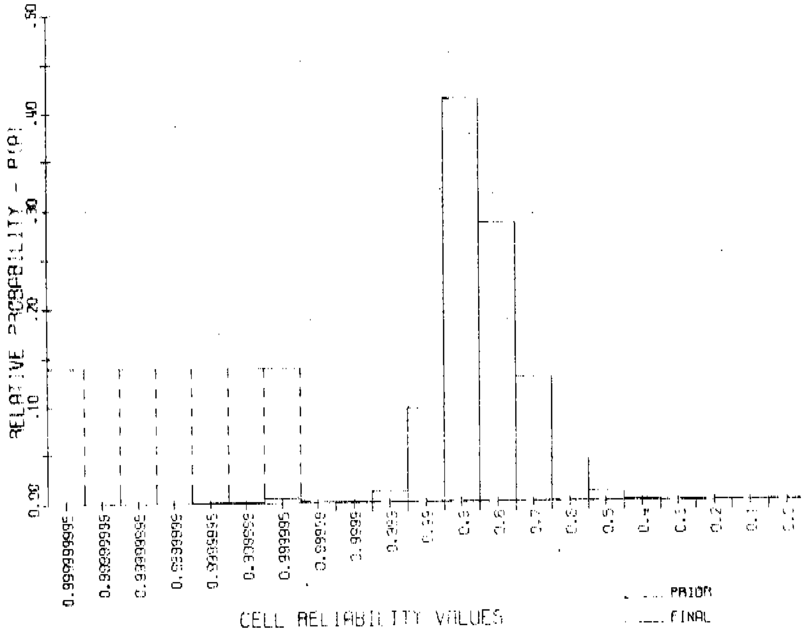
FIGURE(A.6).UNIFORM PRIOR.A.RESULTS OF TEN TESTS.NO FAILURES A+10S

Cell Reliability Values A_i	10S Tests	1F +9S Tests / 4S +1F+5S Tests / 9S +1F Tests	2F +8S Tests / 3S +2F+5S Tests / 8S +2F Tests	5F +5S Tests / 5S +5F Tests	10F Tests
	10 Two Hour Tests with Uniform Prior Distribution				
0.999 999 995	0.1419	0.0000	0.0000	0.0000	0.0000
0.999 999 99	0.1419	0.0000	0.0000	0.0000	0.0000
0.999 999 95	0.1419	0.0001	0.0000	0.0000	0.0000
0.999 999 9	0.1419	0.0001	0.0000	0.0000	0.0000
0.999 999 5	0.1419	0.0005	0.0000	0.0000	0.0000
0.999 999	0.1419	0.0011	0.0000	0.0000	0.0000
0.999 995	0.1419	0.0053	0.0000	0.0000	0.0000
0.999 99	0.0018	0.0001	0.0000	0.0000	0.0000
0.999 9	0.0014	0.0011	0.0000	0.0000	0.0000
0.999	0.0014	0.0106	0.0000	0.0000	0.0000
0.99	0.0013	0.0974	0.0046	0.0000	0.0000
0.9	0.0005	0.4131	0.2128	0.0016	0.0000
0.8	0.0002	0.2863	0.3317	0.0291	0.0000
0.7	0.0000	0.1291	0.2565	0.1132	0.0000
0.6	0.0000	0.0430	0.1328	0.2207	0.0001
0.5	0.0000	0.0104	0.0483	0.2707	0.0007
0.4	0.0000	0.0017	0.0117	0.2207	0.0041
0.3	0.0000	0.0001	0.0016	0.1132	0.0189
0.2	0.0000	0.0000	0.0001	0.0291	0.0720
0.1	0.0000	0.0000	0.0000	0.0016	0.2338
0.0	0.0000	0.0000	0.0000	0.0000	0.6705

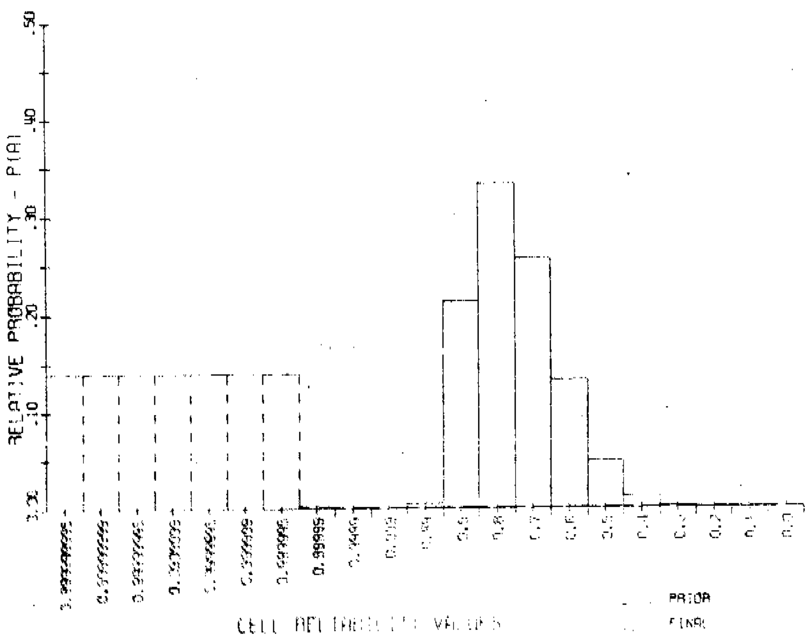
S ≡ Success

F ≡ Failure

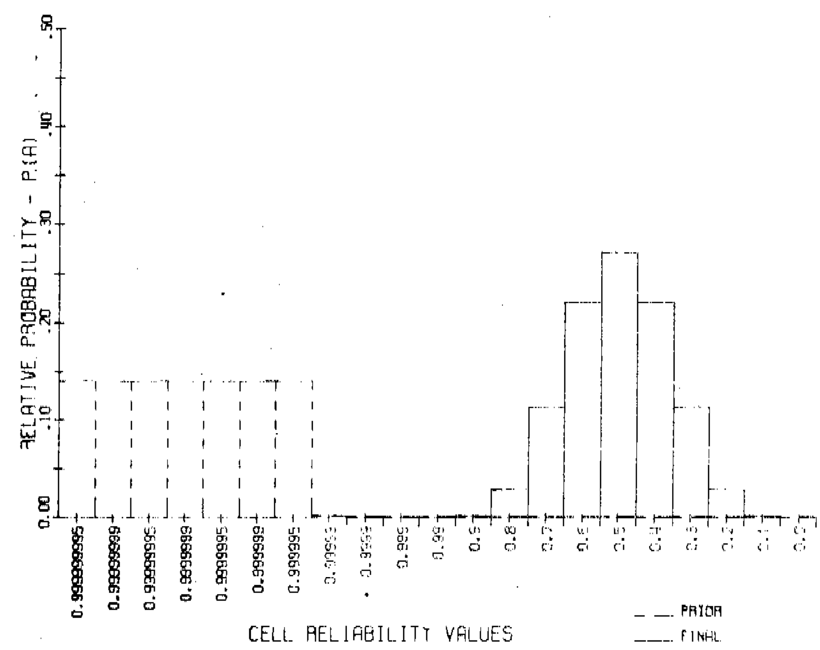
Table A.2 Bayesian Calculation Results of Ten Tests with Various Success/Failure Combinations, for Uniform Prior.



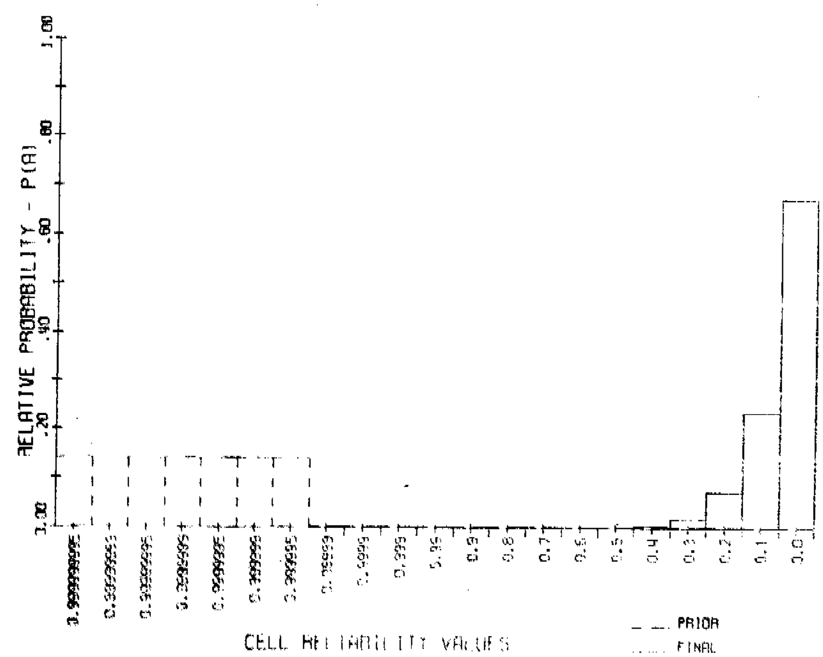
FIGURE(A.7).UNIFORM PRIOR.A.RESULTS OF TEN TESTS.ONE FAILURE A+1F+9S



FIGURE(A.8).UNIFORM PRIOR.A.RESULTS OF TEN TESTS.TWO FAILURES A+2F+8S



FIGURE(A.9).UNIFORM PRIOR.A.RESULTS OF TEN TESTS.5 FAILURES A+5F+5S



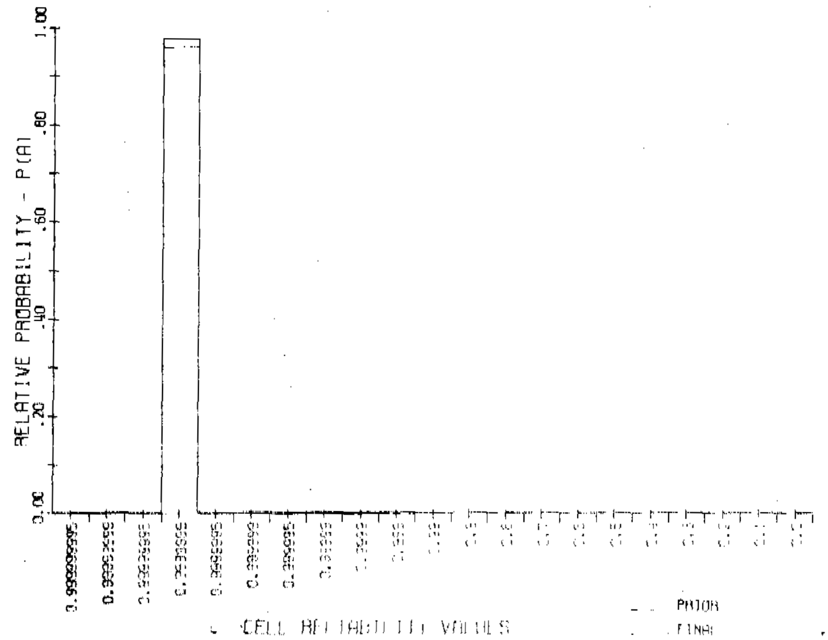
FIGURE(A.10).UNIFORM PRIOR.A.RESULTS OF TEN TESTS.TEN FAILURES A+10F

Cell Reliability Values, A_i	10 Two Hour Tests With Peaked (at 4th Cell) Prior Distribution									
	10S Tests	1F +9S Tests	4S +1F+5S Tests	9S +1F Tests	2F +8S Tests	3S +2F+5S Tests	8S +2F Tests	5F +5S Tests	5S +5F Tests	10F Tests
0.999 999 995	0.0020		0.0000			0.0000			0.0000	0.0000
0.999 999 99	0.0020		0.0000			0.0000			0.0000	0.0000
0.999 999 95	0.0020		0.0000			0.0000			0.0000	0.0000
0.999 999 9	0.9788		0.0005			0.0000			0.0000	0.0000
0.999 999 5	0.0020		0.0000			0.0000			0.0000	0.0000
0.999 999	0.0020		0.0000			0.0000			0.0000	0.0000
0.999 995	0.0020		0.0001			0.0000			0.0000	0.0000
0.999 99	0.0020		0.0001			0.0000			0.0000	0.0000
0.999 9	0.0020		0.0011			0.0000			0.0000	0.0000
0.999	0.0020		0.0106			0.0000			0.0000	0.0000
0.99	0.0018		0.0981			0.0046			0.0000	0.0000
0.9	0.0007		0.4159			0.2128			0.0016	0.0000
0.8	0.0002		0.2881			0.3317			0.0291	0.0000
0.7	0.0001		0.1299			0.2565			0.1132	0.0000
0.6	0.0000		0.0433			0.1328			0.2207	0.0001
0.5	0.0000		0.0105			0.0483			0.2707	0.0007
0.4	0.0000		0.0017			0.0117			0.2207	0.0041
0.3	0.0000		0.0001			0.0016			0.1132	0.0189
0.2	0.0000		0.0000			0.0001			0.0291	0.0720
0.1	0.0000		0.0000			0.0000			0.0016	0.2338
0.0	0.0000		0.0000			0.0000			0.0000	0.6705

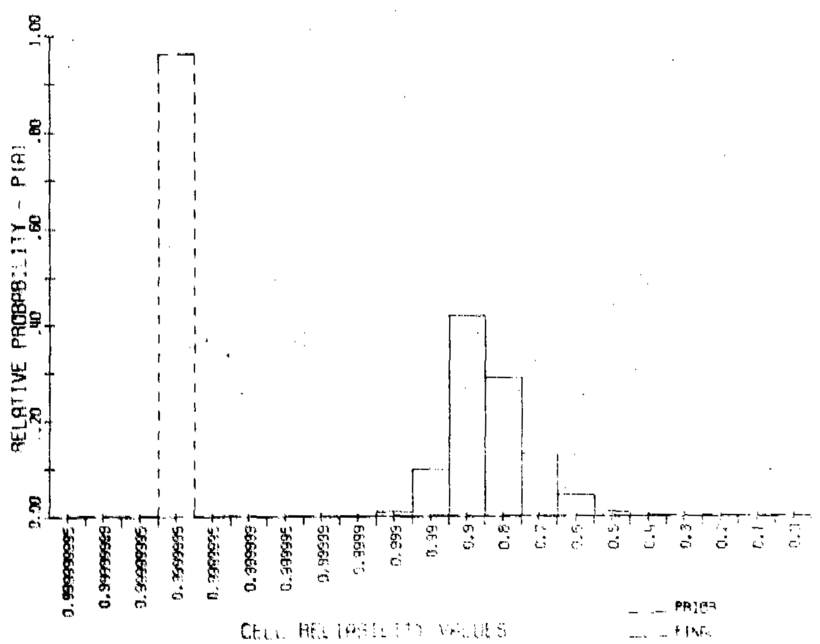
S ≡ Success

F ≡ Failure

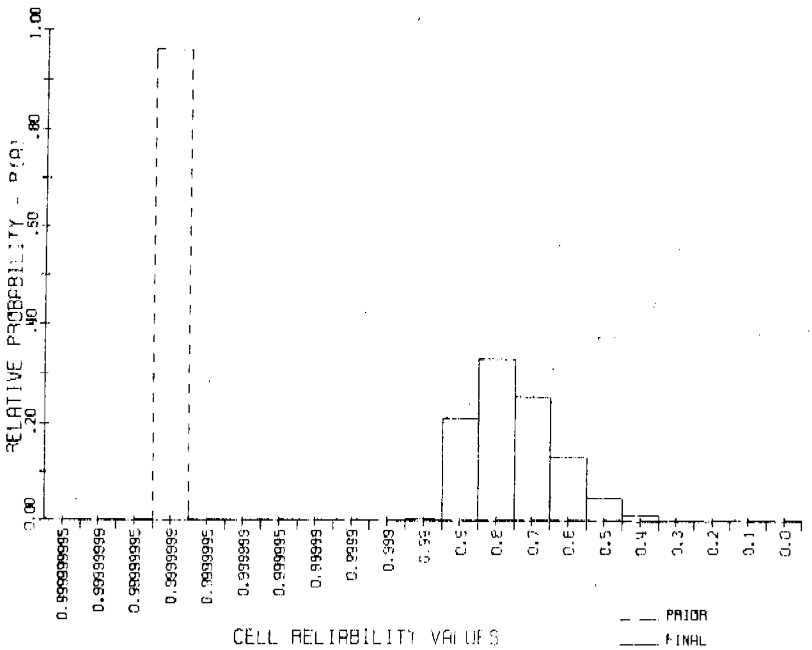
Table A.3 Bayesian Calculation Results of Ten Tests with Various Success/Failure Combinations, for Peaked (at 4th Cell) Prior.



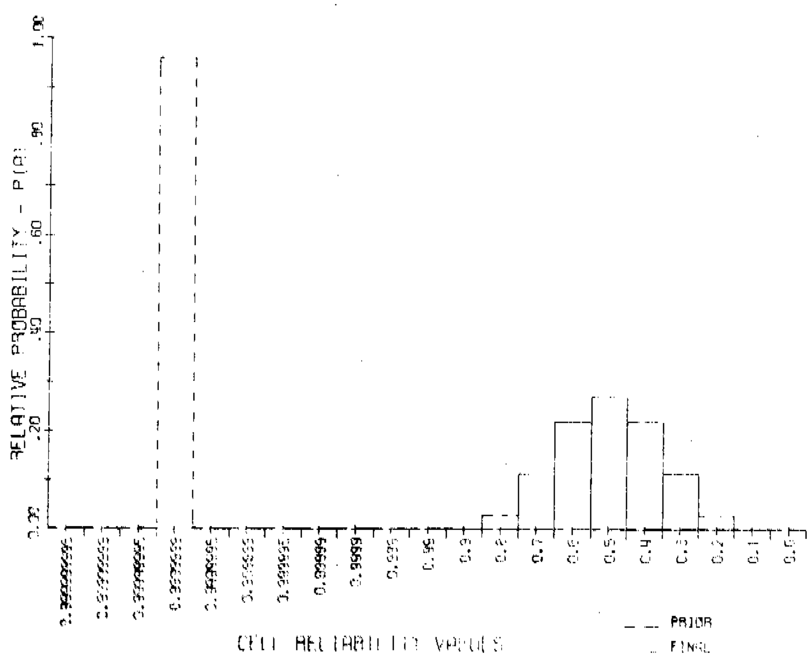
FIGURE(A.11).PEAKED(4TH CELL)PRIOR.A.RESULTS OF 10 TESTS.NO FAILURES [illegible]



FIGURE(A.12).PEAKED(4TH CELL)PRIOR.A.RESULTS OF 10 TESTS.1 FAILURE [illegible]



FIGURE(A.13).PEAKED(4TH CELL)PRIOR.A.RESULTS OF 10 TESTS.2 FAILURES.A+2F+8S



FIGURE(A.14).PEAKED(4TH CELL)PRIOR.A.RESULTS OF 10 TESTS.5 FAILURES.A+5F+5S

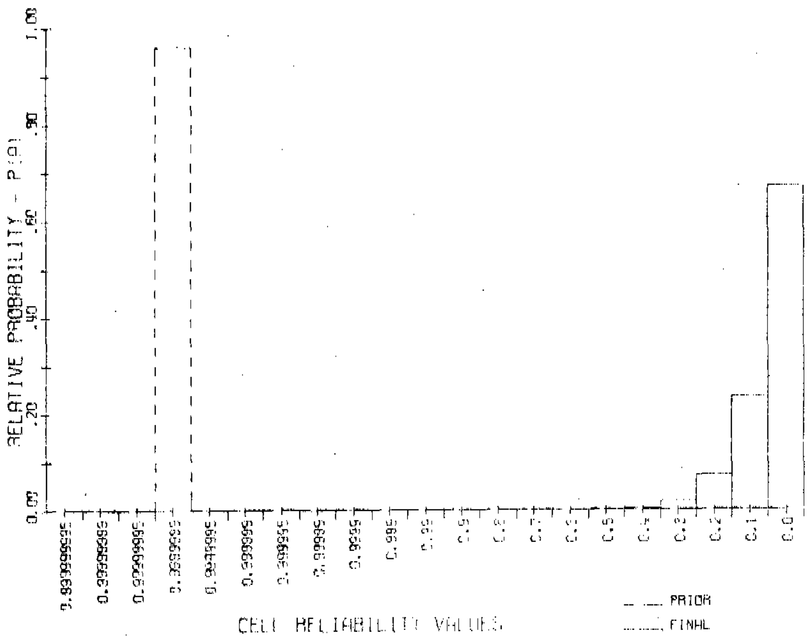


FIGURE (A.15). PEAKED (4TH CELL) PRIOR, A. RESULTS OF 10 TESTS, 10 FAILURES A+10F

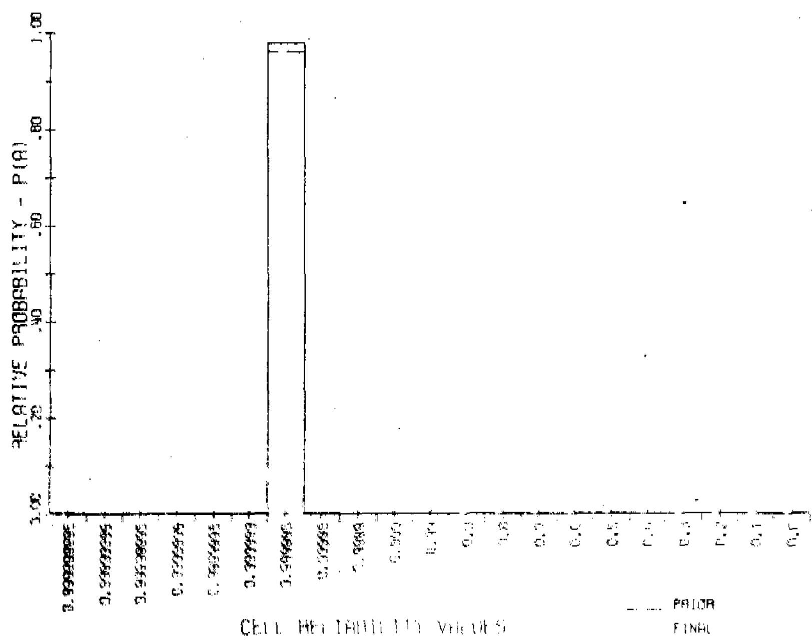


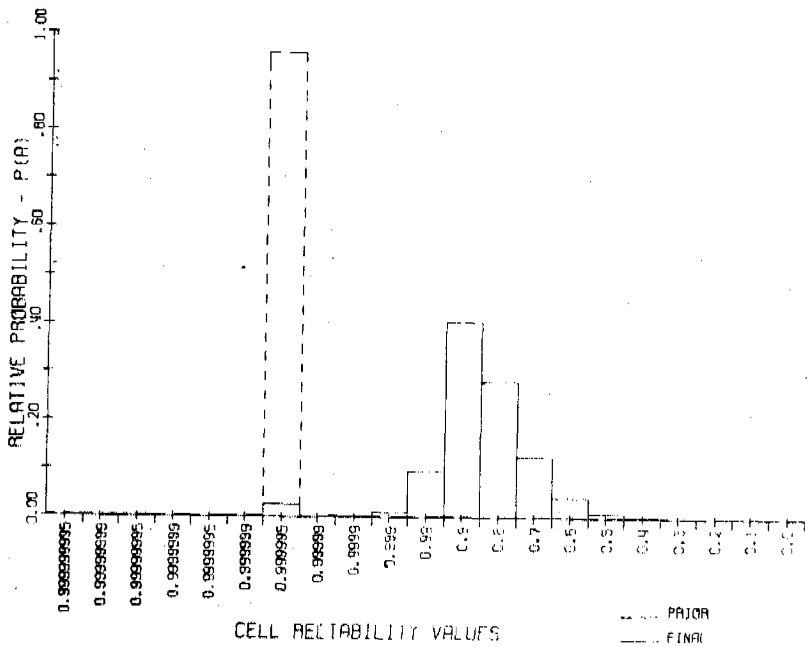
FIGURE (A.16). PEAKED (7TH CELL) PRIOR, A. RESULTS OF 10 TESTS, NO FAILURES A+10S

Cell Reliability Values A_i	10 Two Hour Tests With Peaked (at 7th Cell) Prior Distribution									
	10S Tests	1F +9S Tests	4S +1F+5S Tests	9S +1F Tests	2F +8S Tests	3S +2F+5S Tests	8S +2F Tests	5F +5S Tests	5S +5F Tests	10F Tests
0.999 999 995	0.0020		0.0000			0.0000		0.0000		0.0000
0.999 999 99	0.0020		0.0000			0.0000		0.0000		0.0000
0.999 999 95	0.0020		0.0000			0.0000		0.0000		0.0000
0.999 999 9	0.0020		0.0000			0.0000		0.0000		0.0000
0.999 999 5	0.0020		0.0000			0.0000		0.0000		0.0000
0.999 999	0.0020		0.0000			0.0000		0.0000		0.0000
0.999 995	0.9788		0.0251			0.0000		0.0000		0.0000
0.999 99	0.0020		0.0001			0.0000		0.0000		0.0000
0.999 9	0.0020		0.0010			0.0000		0.0000		0.0000
0.999	0.0020		0.0104			0.0000		0.0000		0.0000
0.99	0.0018		0.0956			0.0046		0.0000		0.0000
0.9	0.0007		0.4056			0.2128		0.0016		0.0000
0.8	0.0002		0.2811			0.3317		0.0291		0.0000
0.7	0.0001		0.1268			0.2565		0.1132		0.0000
0.6	0.0000		0.0422			0.1328		0.2207		0.0001
0.5	0.0000		0.0102			0.0483		0.2707		0.0007
0.4	0.0000		0.0016			0.0117		0.2207		0.0041
0.3	0.0000		0.0001			0.0016		0.1132		0.0189
0.2	0.0000		0.0000			0.0001		0.0291		0.0720
0.1	0.0000		0.0000			0.0000		0.0016		0.2338
0.0	0.0000		0.0000			0.0000		0.0000		0.6705

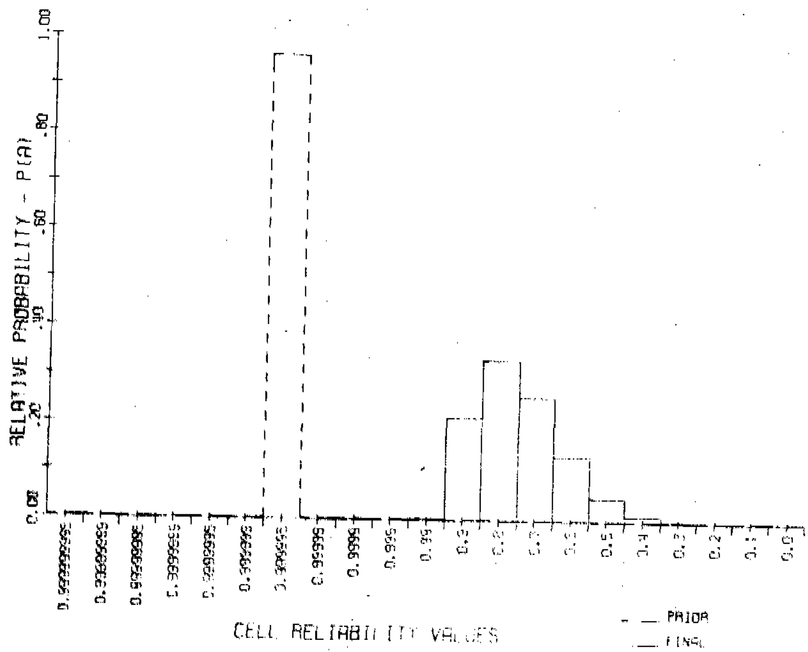
S ≡ Success

F ≡ Failure

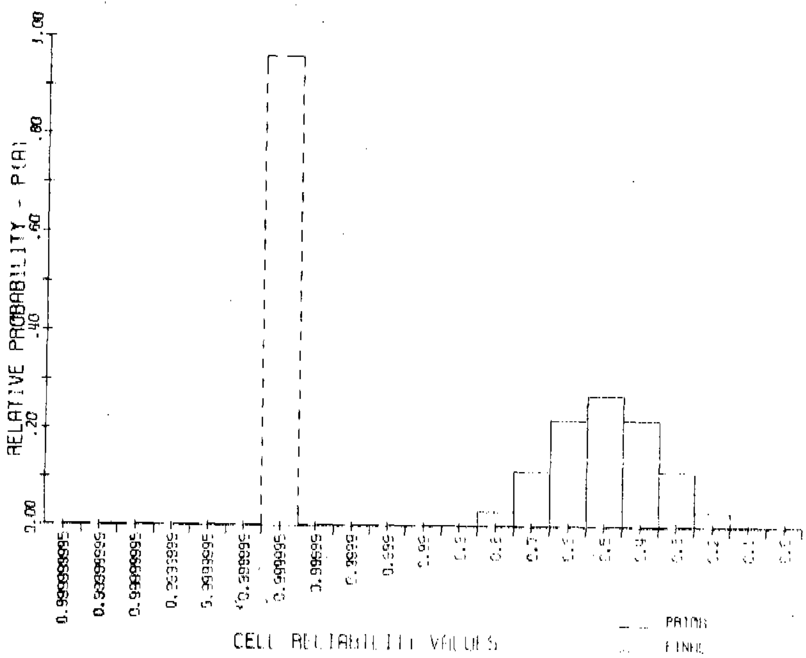
Table A.4 Bayesian Calculation Results of Ten Tests with Various Success/Failure Combinations, for Peaked (at 7th Cell) Prior



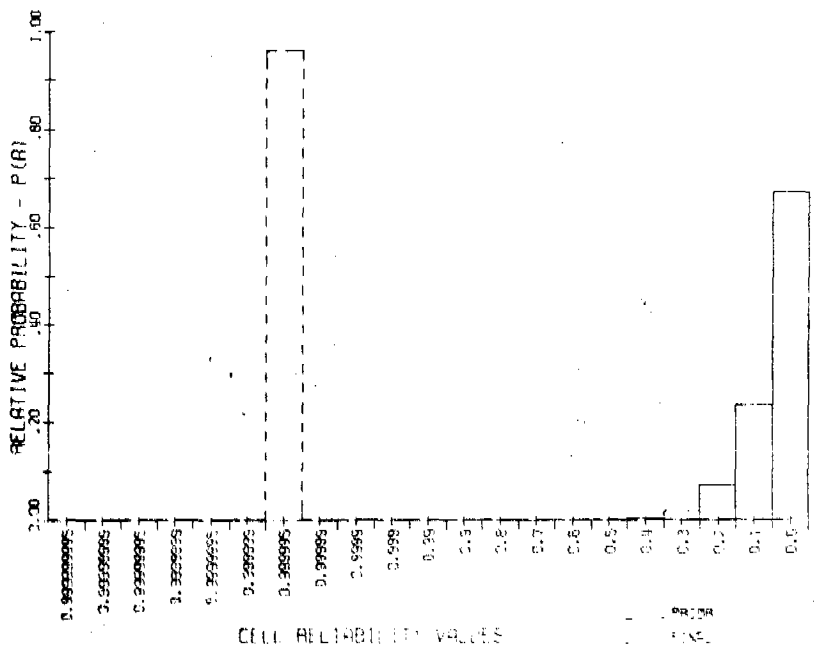
FIGURE(A.17).PEAKED(7TH CELL)PRIOR.A.RESULTS OF 10 TESTS.1 FAILURE A+1F+9S



FIGURE(A.18).PEAKED(7TH CELL)PRIOR.A.RESULTS OF 10 TESTS.2 FAILURES.A+2F+8S



FIGURE(A.19).PEAKED(7TH CELL)PRIOR.A.RESULTS OF 10 TESTS.5 FAILURES.A(5F+5S



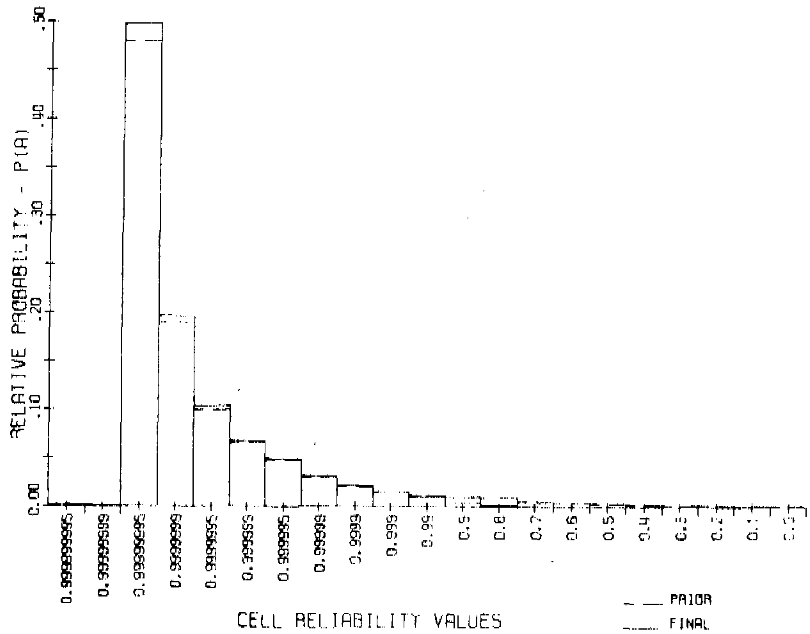
FIGURE(A.20).PEAKED(7TH CELL)PRIOR.A.RESULTS OF 10 TESTS.10 FAILURES A(10F

Cell Reliability Values A_i	10S Tests	1F +9S Tests / 4S +1F+5S Tests / 9S +1F Tests	2F +8S Tests / 3S +2F+5S Tests / 8S +2F Tests	5F +5S Tests / 5S +5F Tests	10F Tests
	10 Two Hour Tests With Approximate Skewed Binomial Prior Distribution				
0.999 999 995	0.0010	0.0000	0.0000	0.0000	0.0000
0.999 999 99	0.0010	0.0000	0.0000	0.0000	0.0000
0.999 999 95	0.4977	0.0000	0.0000	0.0000	0.0000
0.999 999 9	0.1970	0.0000	0.0000	0.0000	0.0000
0.999 999 5	0.1037	0.0001	0.0000	0.0000	0.0000
0.999 999	0.0684	0.0001	0.0000	0.0000	0.0000
0.999 995	0.0487	0.0003	0.0000	0.0000	0.0000
0.999 99	0.0311	0.0004	0.0000	0.0000	0.0000
0.999 9	0.0207	0.0024	0.0000	0.0000	0.0000
0.999	0.0154	0.0180	0.0001	0.0000	0.0000
0.99	0.0103	0.1215	0.0070	0.0000	0.0000
0.9	0.0036	0.4685	0.2976	0.0050	0.0000
0.8	0.0010	0.2921	0.4176	0.0801	0.0000
0.7	0.0001	0.0732	0.1794	0.1733	0.0000
0.6	0.0000	0.0195	0.0743	0.2703	0.0003
0.5	0.0000	0.0035	0.0203	0.2486	0.0018
0.4	0.0000	0.0004	0.0033	0.1351	0.0074
0.3	0.0000	0.0000	0.0004	0.0693	0.0345
0.2	0.0000	0.0000	0.0000	0.0178	0.1313
0.1	0.0000	0.0000	0.0000	0.0005	0.2132
0.0	0.0000	0.0000	0.0000	0.0000	0.6115

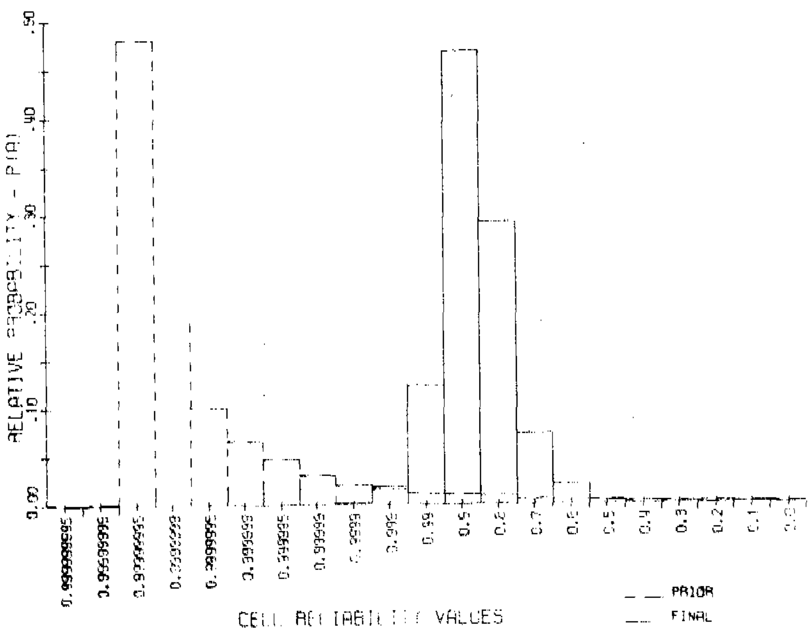
S ≡ Success

F ≡ Failure

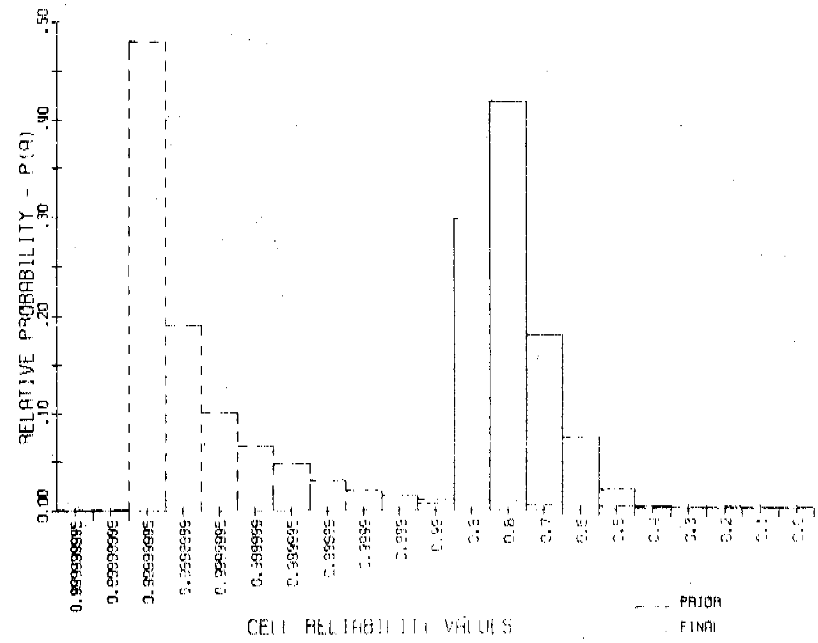
Table A.5 Bayesian Calculation Results of Ten Tests with Various Success/Failure Combinations, for Skewed Binomial Prior.



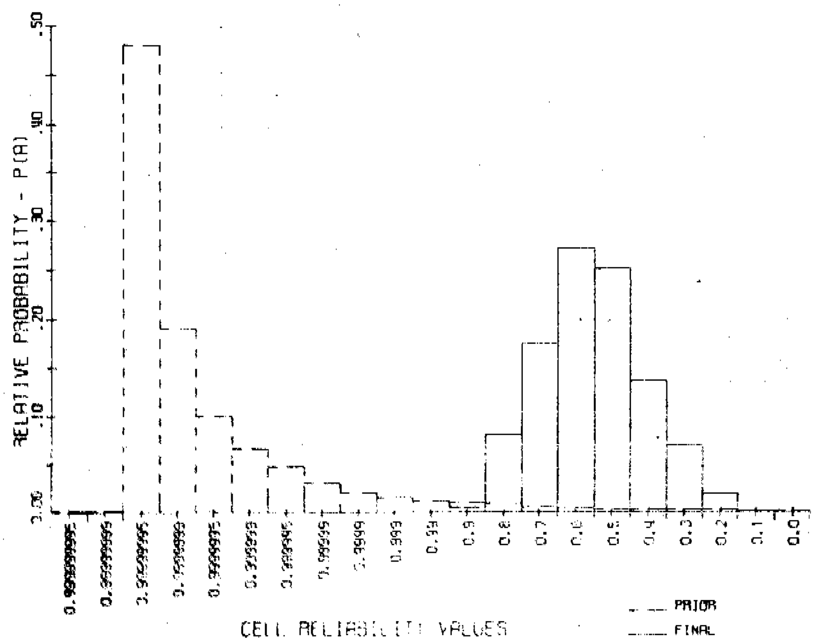
FIGURE(A.21).SKEWED BINOMIAL PRIOR.A.RESULTS OF 10 TESTS.NO FAILURES A+10S



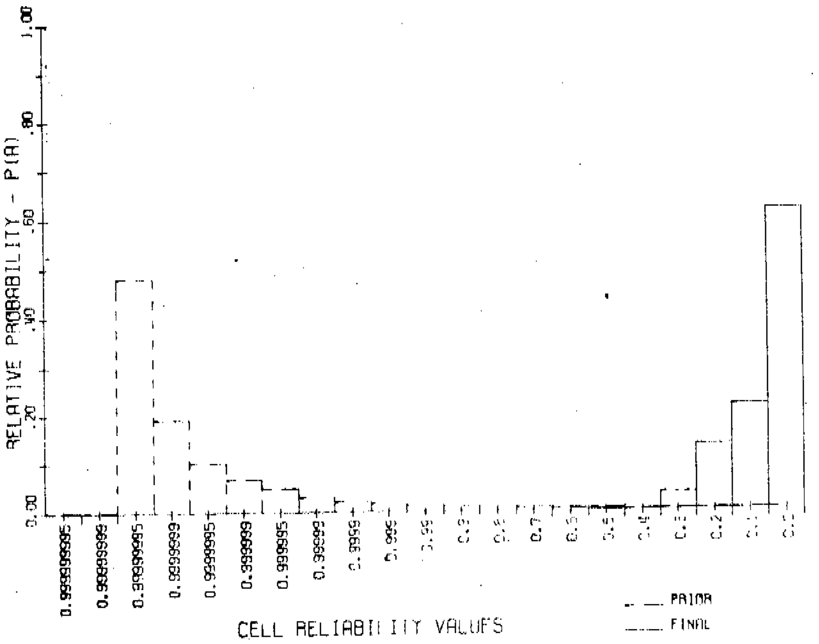
FIGURE(A.22).SKEWED BINOMIAL PRIOR.A.RESULTS OF 10 TESTS.1 FAILURE A+1F+9S



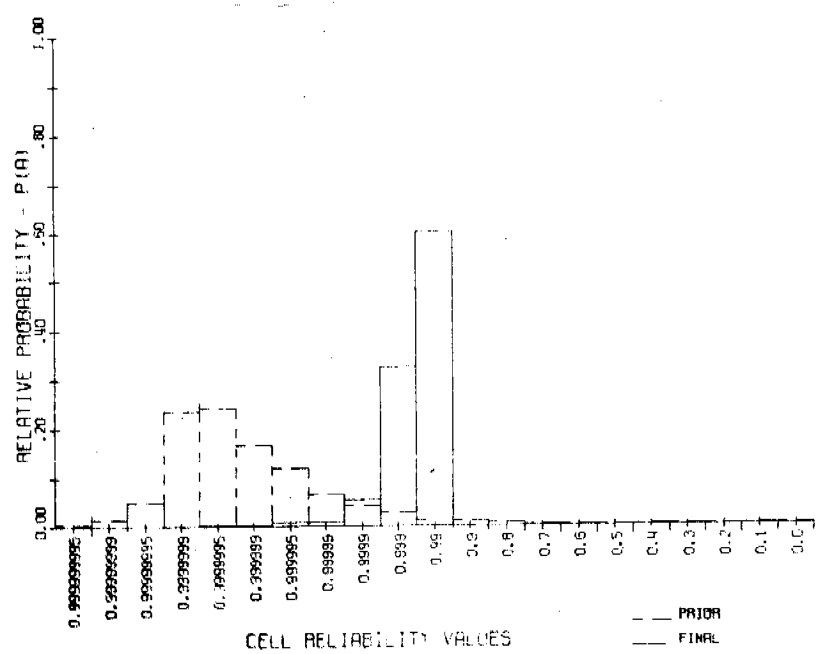
FIGURE(A.23).SKEWED BINOMIAL PRIOR.A.RESULTS OF 10 TESTS.2 FAILURES A+2F+8S



FIGURE(A.24).SKEWED BINOMIAL PRIOR.A.RESULTS OF 10 TESTS.5 FAILURES A+5F+5S



FIGURE(A.25).SKEWED BINOMIAL PRIOR.A.RESULTS OF 10 TESTS.10 FAILURES A+10F



FIGURE(A.26).BINOMIAL PRIOR.A.RESULTS OF 100 TESTS.1 FAILURE A+99S+1F

Cell Reliability Values A_i	Tests With Binomial Type Average Prior Distribution of Table 2				
	99S + 1F or 1F + 99S	499S + 1F or 1F + 499S	999S + 1F or 1F + 999S	4999S + 1F or 1F + 4999S	9999S + 1F or 1F + 9999S
0.999 999 995	0.0000	0.0000	0.0000	0.0000	0.0000
0.999 999 99	0.0000	0.0000	0.0000	0.0000	0.0000
0.999 999 95	0.0000	0.0001	0.0002	0.0006	0.0008
0.999 999 9	0.0003	0.0010	0.0015	0.0055	0.0077
0.999 999 5	0.0016	0.0052	0.0078	0.0283	0.0398
0.999 999	0.0021	0.0071	0.0106	0.0384	0.0538
0.999 995	0.0078	0.0259	0.0385	0.1368	0.1881
0.999 99	0.0084	0.0280	0.0415	0.1445	0.1939
0.999 9	0.0547	0.1749	0.2477	0.6026	0.5154
0.999	0.3238	0.7219	0.6519	0.0433	0.0004
0.99	0.6008	0.0359	0.0004	0.0000	0.0000
0.9	0.0004	0.0000	0.0000	0.0000	0.0000
0.8	0.0000	0.0000	0.0000	0.0000	0.0000
0.7	0.0000	0.0000	0.0000	0.0000	0.0000
0.6	0.0000	0.0000	0.0000	0.0000	0.0000
0.5	0.0000	0.0000	0.0000	0.0000	0.0000
0.4	0.0000	0.0000	0.0000	0.0000	0.0000
0.3	0.0000	0.0000	0.0000	0.0000	0.0000
0.2	0.0000	0.0000	0.0000	0.0000	0.0000
0.1	0.0000	0.0000	0.0000	0.0000	0.0000
0.0	0.0000	0.0000	0.0000	0.0000	0.0000

Table A.6 Bayesian Calculation Results for Various Numbers of Tests, Each with One Failure, for Binomial Type Prior.

Cell Reliability Values, A_i	Tests With Uniform Prior Distribution				
	99S +1F or 1F + 99S	499S + 1F or 1F + 499S	999S + 1F or 1F + 999S	4999S + 1F or 1F + 4999S	9999S + 1F or 1F + 9999S
0.999 999 995	0.0001	0.0003	0.0004	0.0007	0.0007
0.999 999 99	0.0002	0.0007	0.0009	0.0014	0.0015
0.999 999 95	0.0009	0.0035	0.0044	0.0068	0.0073
0.999 999 9	0.0019	0.0069	0.0088	0.0136	0.0145
0.999 999 5	0.0093	0.0346	0.0440	0.0681	0.0722
0.999 999	0.0186	0.0691	0.0880	0.1357	0.1436
0.999 995	0.0928	0.3448	0.4381	0.6652	0.6900
0.999 99	0.0024	0.0088	0.0112	0.0167	0.0169
0.999 9	0.0184	0.0658	0.0797	0.0827	0.0534
0.999	0.1682	0.4196	0.3241	0.0092	0.0001
0.99	0.6867	0.0459	0.0004	0.0000	0.0000
0.9	0.0005	0.0000	0.0000		
0.8	0.0000				
0.7	0.0000				
0.6	0.0000				
0.5	0.0000				
0.4	0.0000				
0.3	0.0000				
0.2	0.0000				
0.1	0.0000				
0.0	0.0000	0.0000	0.0000	0.0000	0.0000

Table A.7 Bayesian Calculation Results for Various Numbers of Tests, Each with One Failure, for Uniform Prior.

	Tests With Peaked (at 4th Cell) Prior Distribution				
Cell Reliability Values, A_i	99S + 1F or 1F + 99S	499S + 1F or 1F + 499S	999S + 1F or 1F + 999S	4999S + 1F or 1F + 4999S	9999S + 1F or 1F + 9999S
0.999 999 995	0.0000	0.0000	0.0000	0.0000	0.0001
0.999 999 99	0.0000	0.0000	0.0000	0.0001	0.0001
0.999 999 95	0.0000	0.0001	0.0001	0.0004	0.0005
0.999 999 9	0.0100	0.0576	0.0917	0.3654	0.4788
0.999 999 5	0.0001	0.0006	0.0010	0.0038	0.0050
0.999 999	0.0002	0.0012	0.0019	0.0076	0.0100
0.999 995	0.0010	0.0060	0.0100	0.0371	0.0475
0.999 99	0.0021	0.0119	0.0189	0.0724	0.0904
0.999 9	0.0208	0.1142	0.1729	0.4619	0.3674
0.999	0.1899	0.7287	0.7032	0.0512	0.0004
0.99	0.7752	0.0797	0.0008	0.0000	0.0000
0.9	0.0006	0.0000	0.0000	0.0000	0.0000
0.8	0.0000	0.0000	0.0000	0.0000	0.0000
0.7	0.0000	0.0000	0.0000	0.0000	0.0000
0.6	0.0000	0.0000	0.0000	0.0000	0.0000
0.5	0.0000	0.0000	0.0000	0.0000	0.0000
0.4	0.0000	0.0000	0.0000	0.0000	0.0000
0.3	0.0000	0.0000	0.0000	0.0000	0.0000
0.2	0.0000	0.0000	0.0000	0.0000	0.0000
0.1	0.0000	0.0000	0.0000	0.0000	0.0000
0.0	0.0000	0.0000	0.0000	0.0000	0.0000

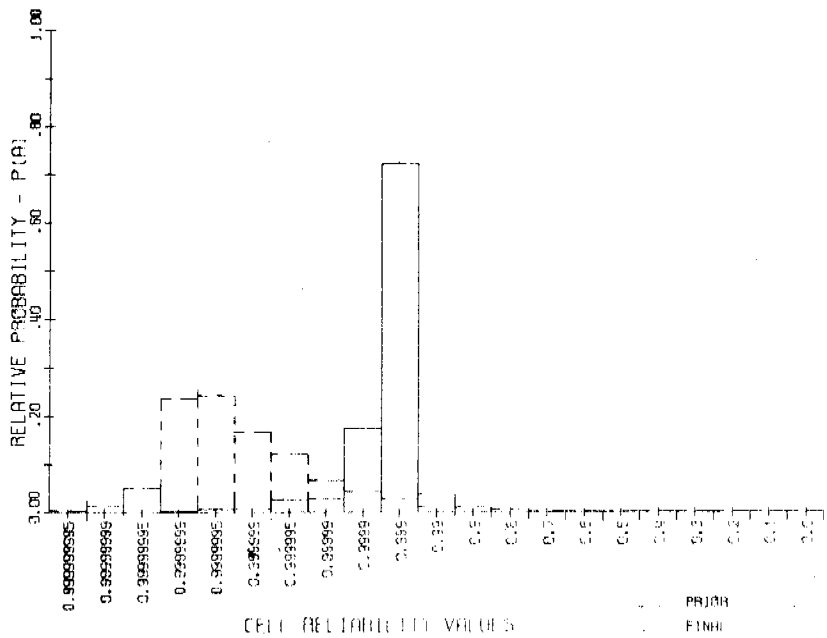
Table A.8 Bayesian Calculation Results for Various Numbers of Tests, Each with One Failure, for Peaked (at 4th Cell) Prior.

	Tests With Peaked (at 7th Cell) Prior Distribution					
Cell Reliability Values, A_i	99S + 1F or 1F + 99S	499S + 1F or 1F + 499S	999S + 1F or 1F + 999S	4999S + 1F or 1F + 4999S	9999S + 1F or 1F + 9999S	
0.999 999 995	0.0000	0.0000	0.0000	0.0000	0.0000	
0.999 999 99	0.0000	0.0000	0.0000	0.0000	0.0000	
0.999 999 95	0.0000	0.0000	0.0000	0.0000	0.0000	
0.999 999 9	0.0000	0.0000	0.0000	0.0000	0.0000	
0.999 999 5	0.0001	0.0002	0.0002	0.0002	0.0002	
0.999 999	0.0001	0.0003	0.0003	0.0004	0.0004	
0.999 995	0.3371	0.7542	0.8354	0.9675	0.9800	
0.999 99	0.0014	0.0031	0.0035	0.0039	0.0039	
0.999 9	0.0139	0.0300	0.0317	0.0251	0.0158	
0.999	0.1273	0.1912	0.1288	0.0028	0.0000	
0.99	0.5196	0.0209	0.0002	0.0000	↑	
0.9	0.0004	0.0000	0.0000	↑		
0.8	0.0000	↑	↑			
0.7	↑					
0.6						
0.5						
0.4						
0.3						
0.2						
0.1	↓	↓	↓	↓	↓	
0.0	0.0000	0.0000	0.0000	0.0000	0.0000	

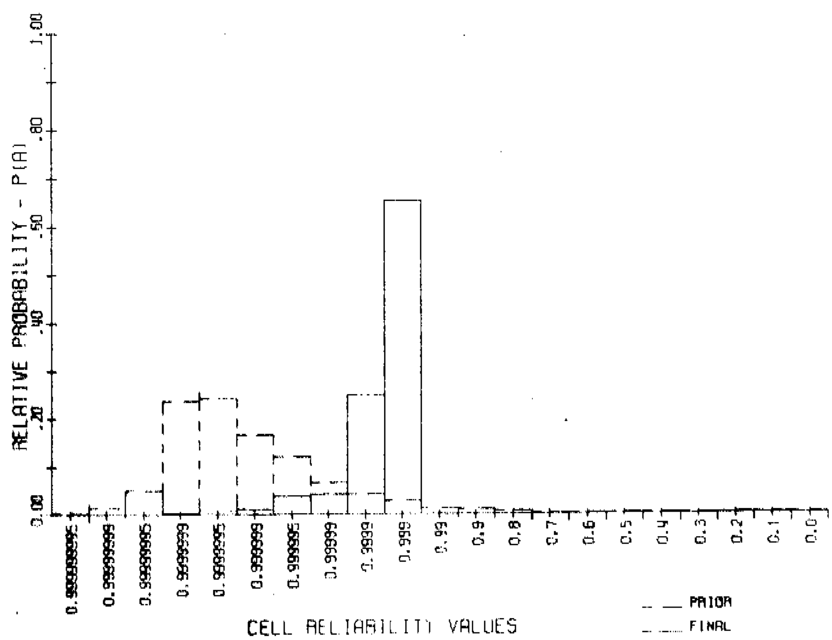
Table A.9 Bayesian Calculation Results for Various Numbers of Tests, each with One Failure, for Peaked (at 7th Cell) Prior.

Cell Reliability Values A_i	Tests With Approximate Skewed Binomial Prior Distribution				
	99S +1F or 1F +99S	499S + 1F or 1F +499S	999S+1F or 1F+999S	4999S+1F or 1F+4999S	9999S +1F or 1F+9999S
0.999 999 995	0.0000	0.0000	0.0000	0.0000	0.0000
0.999 999 99	0.0000	0.0000	0.0000	0.0000	0.0000
0.999 999 95	0.0004	0.0019	0.0030	0.0121	0.0173
0.999 999 9	0.0003	0.0015	0.0024	0.0096	0.0137
0.999 999 5	0.0009	0.0040	0.0062	0.0251	0.0358
0.999 999	0.0012	0.0053	0.0082	0.0330	0.0470
0.999 995	0.0041	0.0189	0.0291	0.1153	0.1609
0.999 99	0.0053	0.0240	0.0370	0.1436	0.1954
0.999 9	0.0348	0.1531	0.2255	0.6105	0.5300
0.999	0.2385	0.7325	0.6880	0.0508	0.0005
0.99	0.7140	0.0587	0.0006	0.0000	0.0000
0.9	0.0005	0.0000	0.0000	↑	↑
0.8	0.0000	↑	↑		
0.7	↑				
0.6					
0.5					
0.4					
0.3					
0.2					
0.1	↓	↓	↓	↓	↓
0.0	0.0000	0.0000	0.0000	0.0000	0.0000

Table A.10 Bayesian Calculation Results for Various Numbers of Tests, Each with One Failure, for Skewed Binomial Prior.



FIGURE(A.27).BINOMIAL PRIOR.A.RESULTS OF 500 TESTS.1 FAILURE A+1F+999S



FIGURE(A.28).BINOMIAL PRIOR.A.RESULTS OF 1000 TESTS.1 FAILURE A+999S+1F

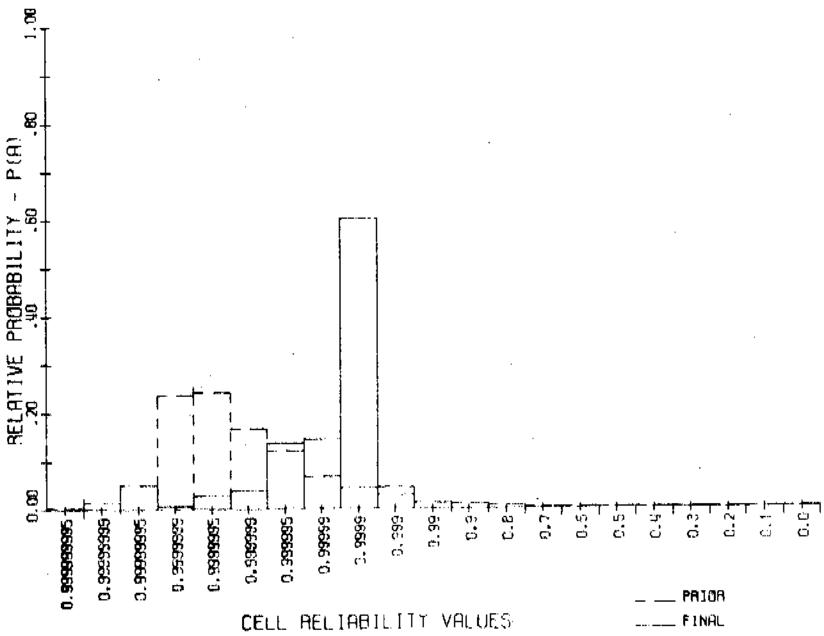


FIGURE (A.29). BINOMIAL PRIOR.A.RESULTS OF 5000 TESTS.1 FAILURE A+1F+4999S

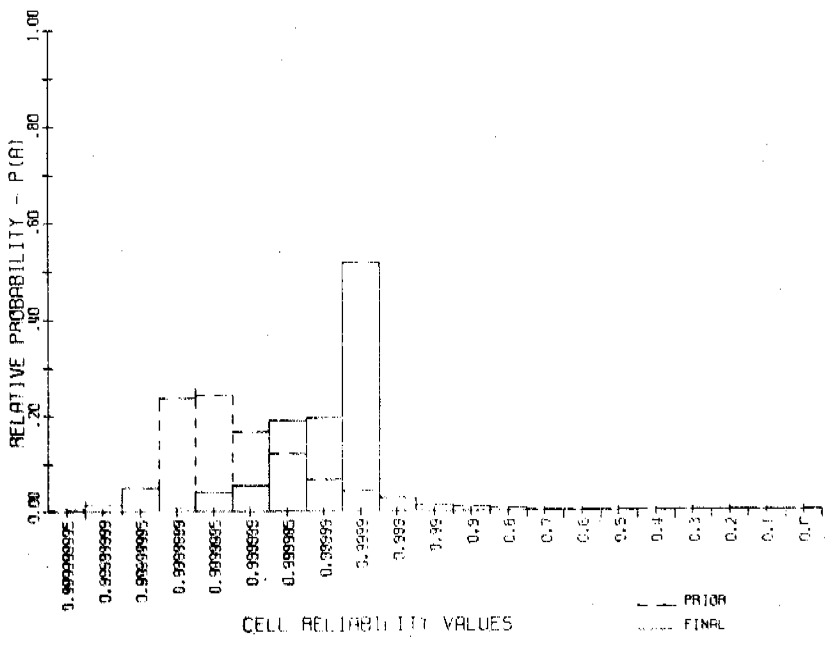
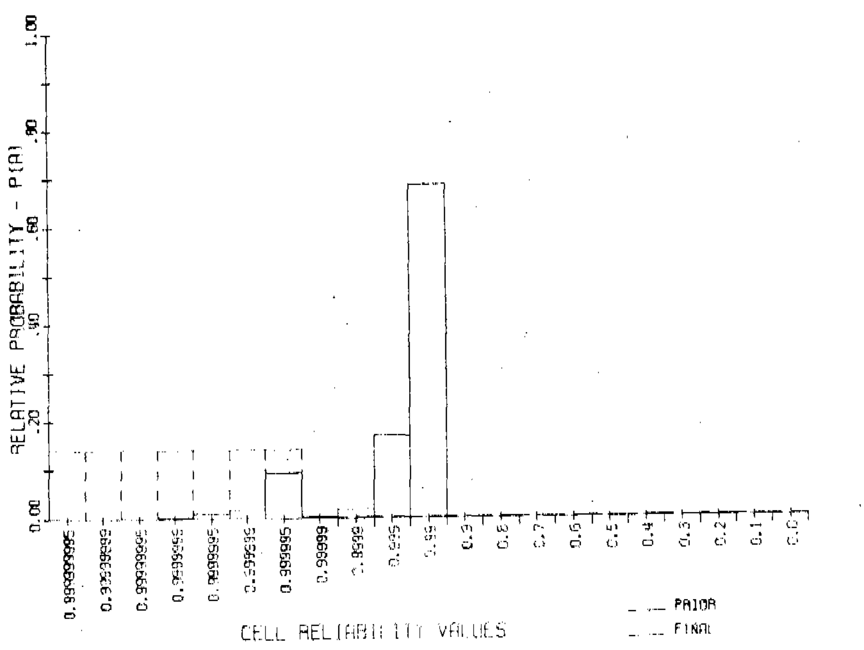
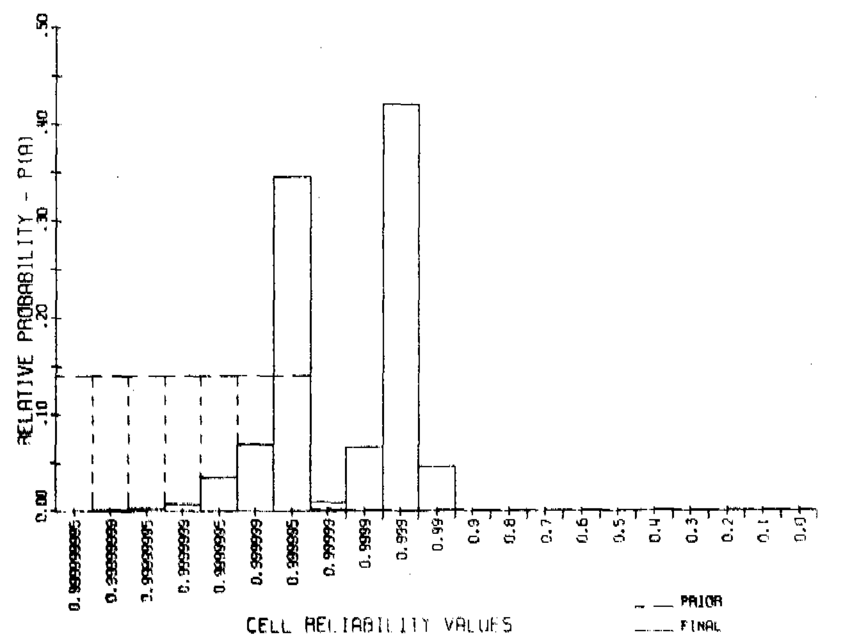


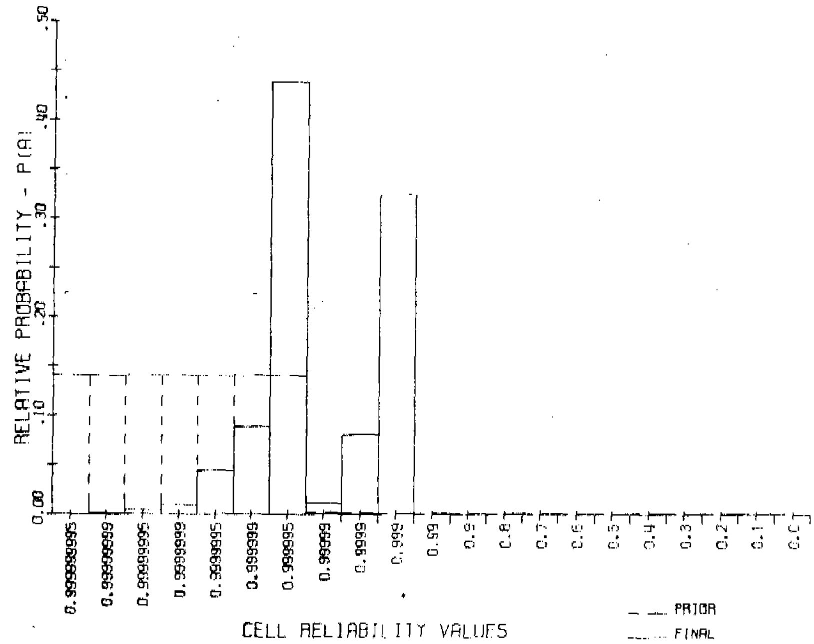
FIGURE (A.30). BINOMIAL PRIOR.A.RESULTS OF 10000 TESTS.1 FAILURE A.9999S+1F



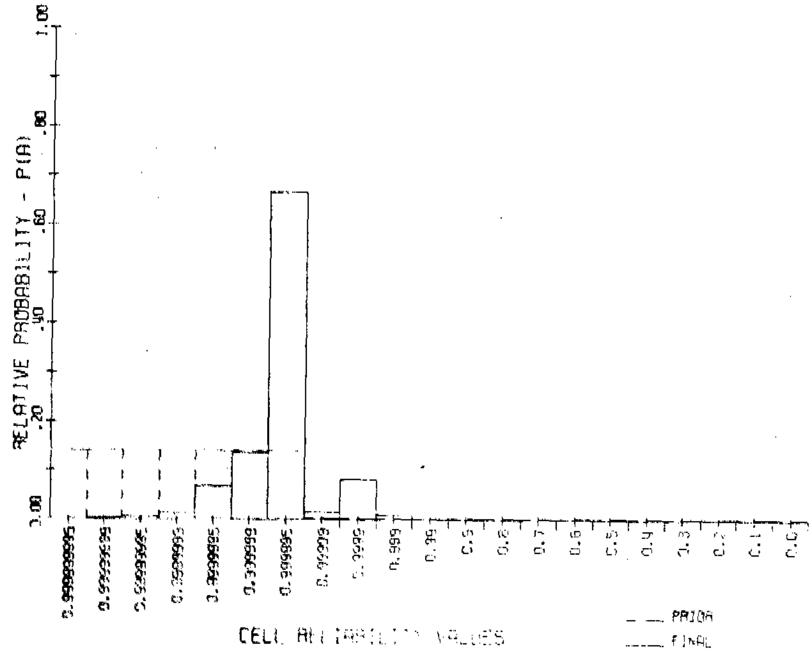
FIGURE(A.31). UNIFORM PRIOR,A,RESULTS OF 100 TESTS,1 FAILURE A+1F+99S



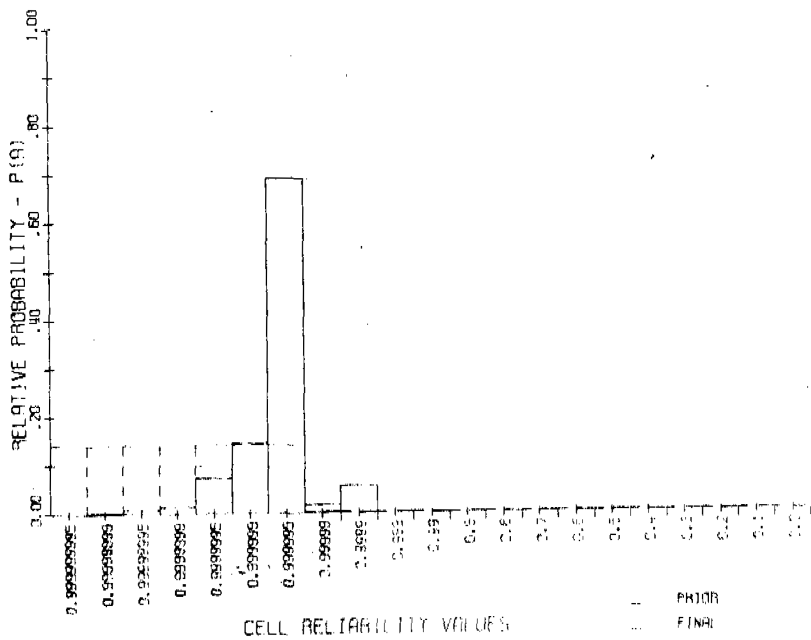
FIGURE(A.32). UNIFORM PRIOR,A,RESULTS OF 500 TESTS,1 FAILURE A+499S+1F



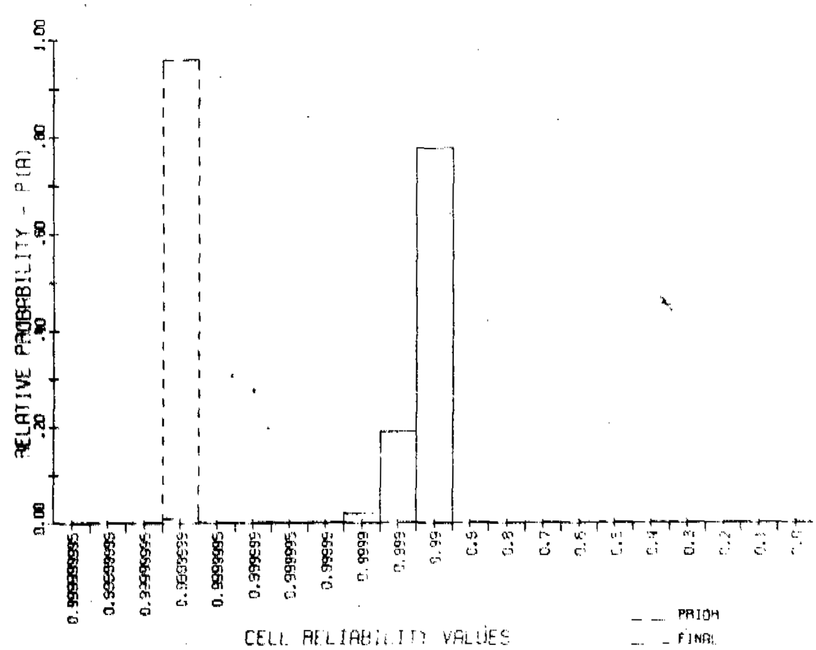
FIGURE(A.33). UNIFORM PRIOR,A,RESULTS OF 1000 TESTS,1 FAILURE A+1F+9995



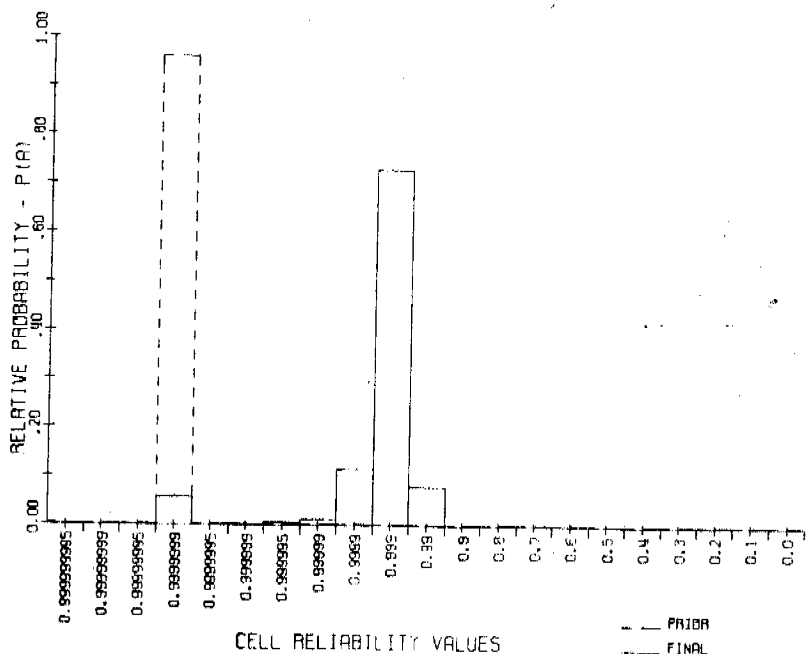
FIGURE(A.34). UNIFORM PRIOR,A,RESU[illegible] OF 5000 TESTS,1 FAILURE A+49995-[illegible]



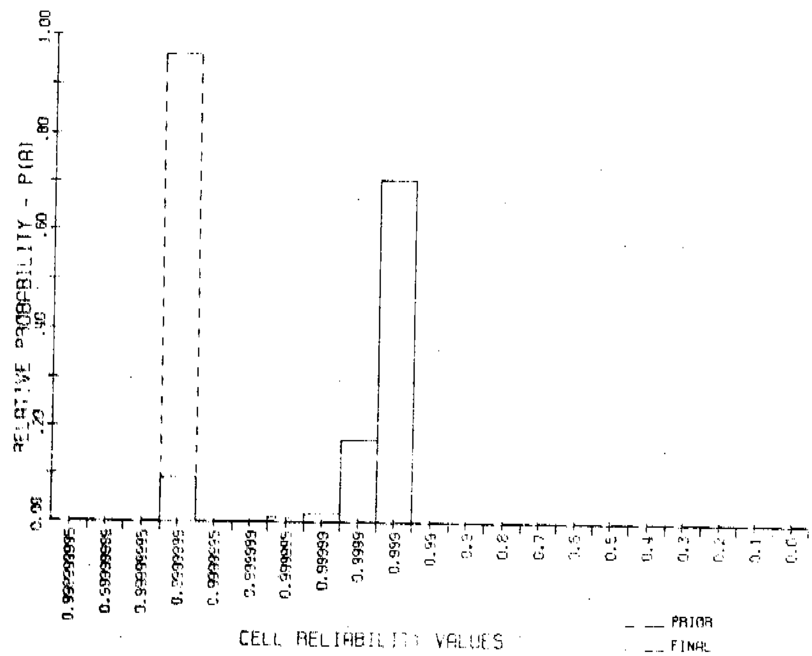
FIGURE(A.35). UNIFORM PRIOR,A,RESULTS OF 10000 TESTS,1 FAILURE A+1F .99995



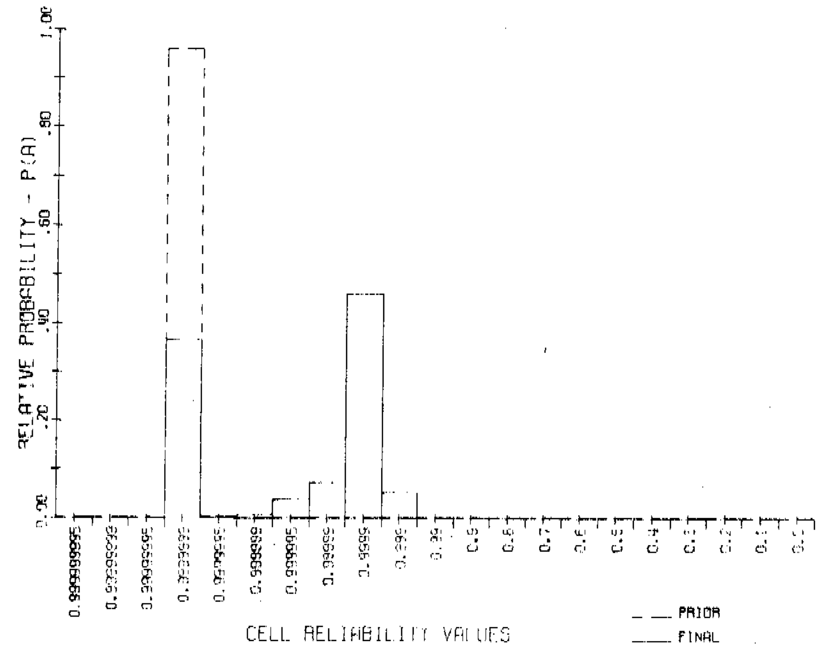
FIGURE(A.36). PEAKED(4TH CELL)PRIOR,A,RESULTS OF 100 TESTS,1 FAILURE A+995+1F



FIGURE(A.37). PEAKED(4TH CELL)PRIOR,A,RESULTS OF 500 TESTS,1 FAILURE A+1F+499S



FIGURE(A.38). PEAKED(4TH CELL)PRIOR,A,1000 TESTS,1 FAILURE A+999S+1F



FIGURE(A.39). PEAKED(4TH CELL)PRIOR,A.5000 TESTS.1 FAILURE A+1F+49995

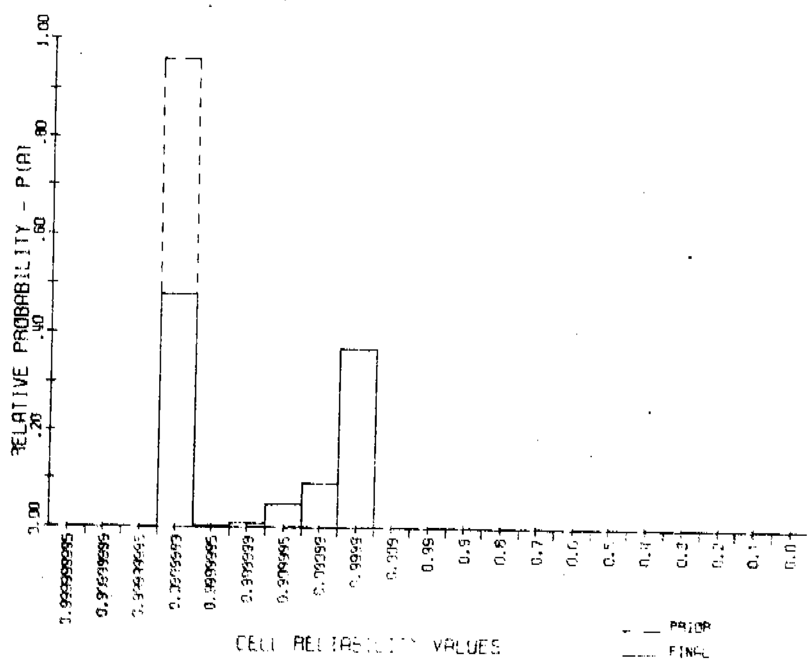
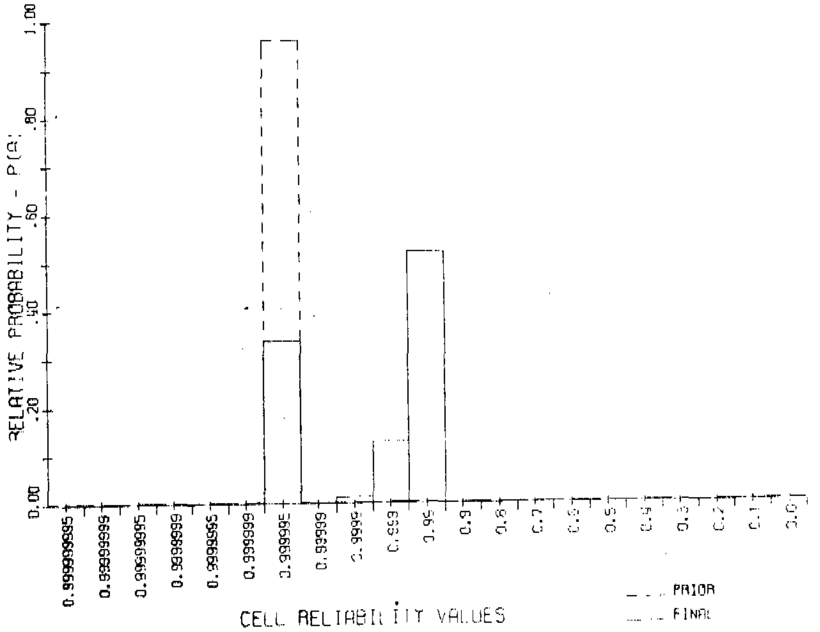
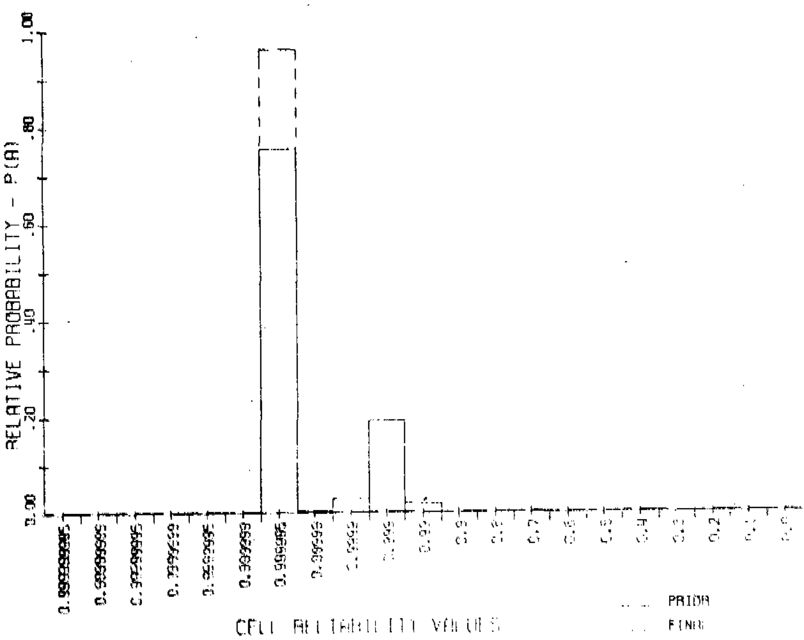


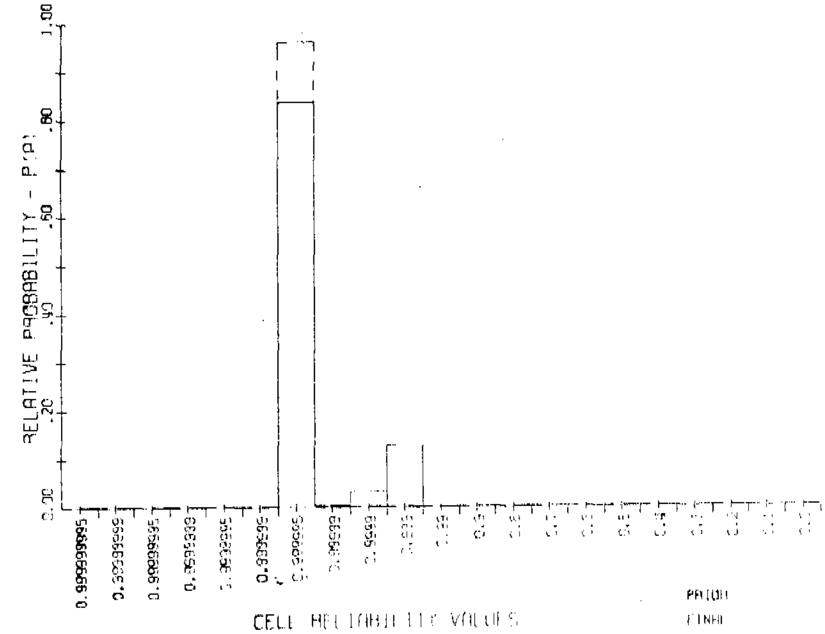
FIGURE (A.40). PEAKED(4TH CELL)PRIOR,A.10000 TESTS.1 FAILURE A+99995+1F



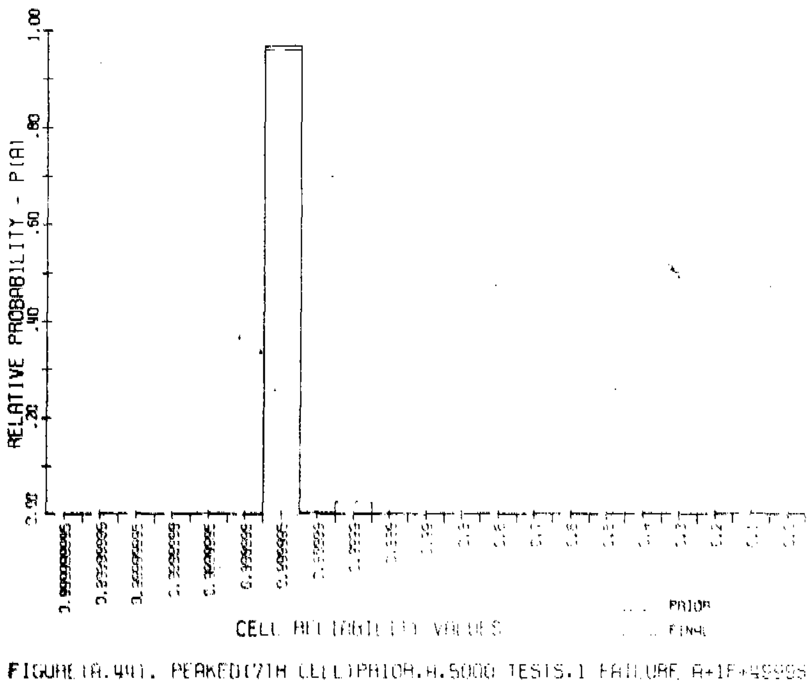
FIGURE(A.41). PEAKED(7TH CELL)PRIOR.A.RESULTS OF 100 TESTS.1 FAILURE A:995+1F



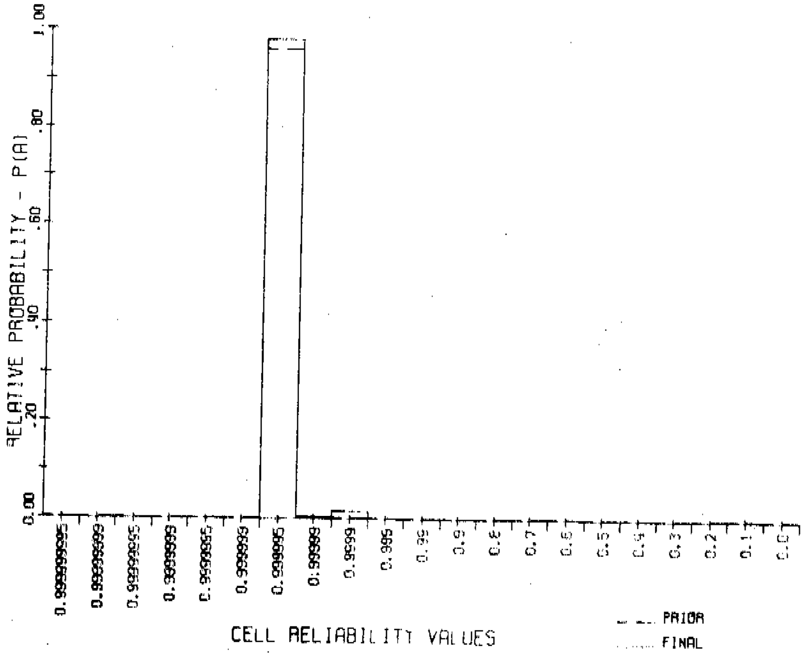
FIGURE(A.42). PEAKED(7TH CELL)PRIOR.A.500 TESTS.1 FAILURE A:4995+1F



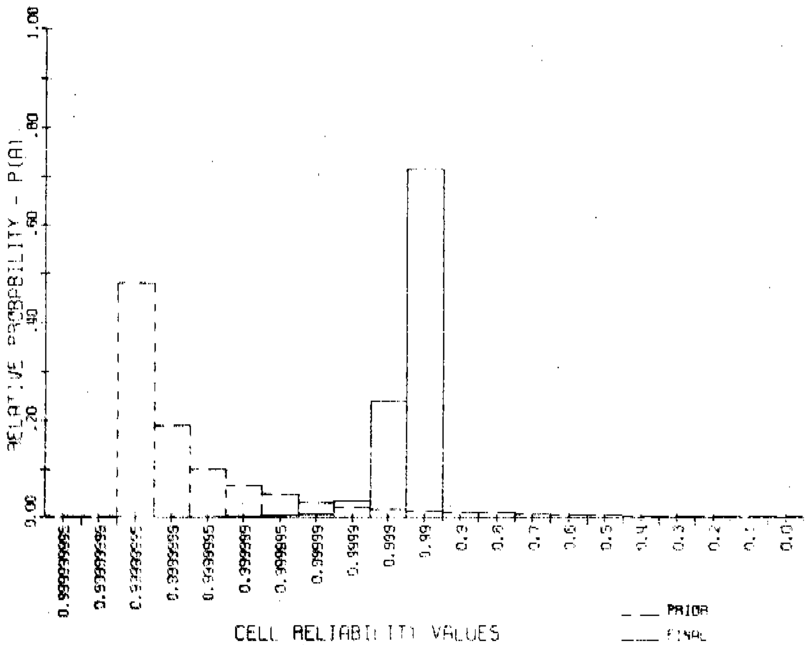
FIGURE(A.43). PEAKED(7TH CELL)PRIOR,A.1000 TESTS,1 FAILURE



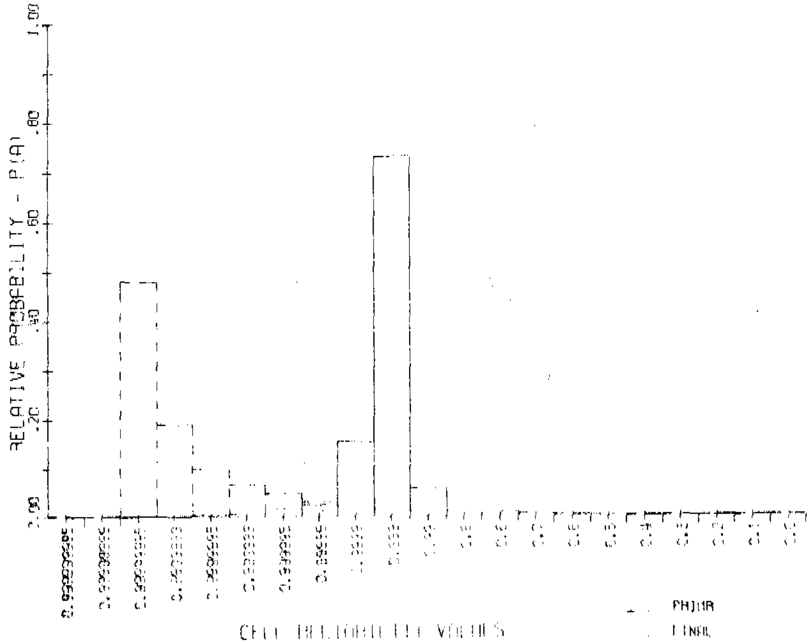
FIGURE(A.44). PEAKED(7TH CELL)PRIOR,A.5000 TESTS,1 FAILURE A+1F+4999S



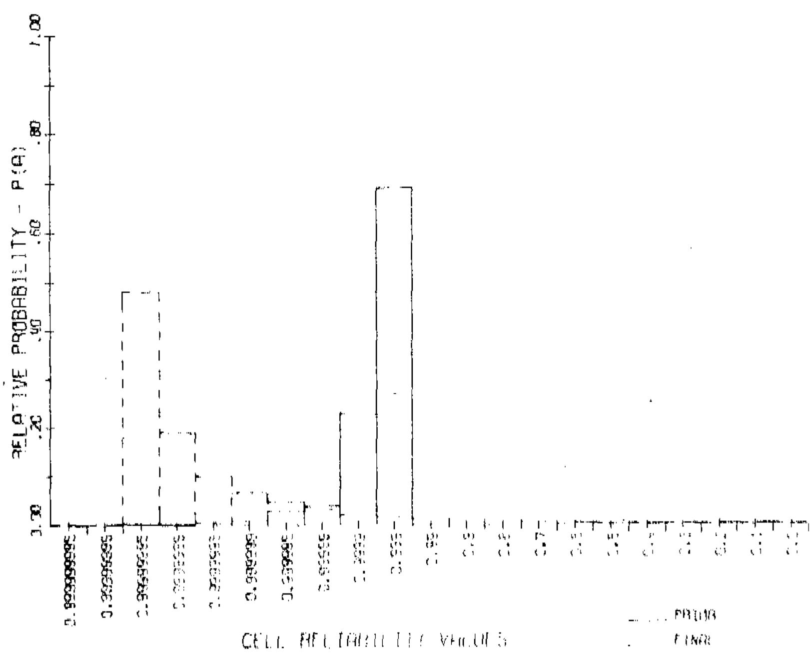
FIGURE(A.45). PEAKED(7TH CELL)PRIOR,A,10000 TESTS,1 FAILURE A+9999S+1F



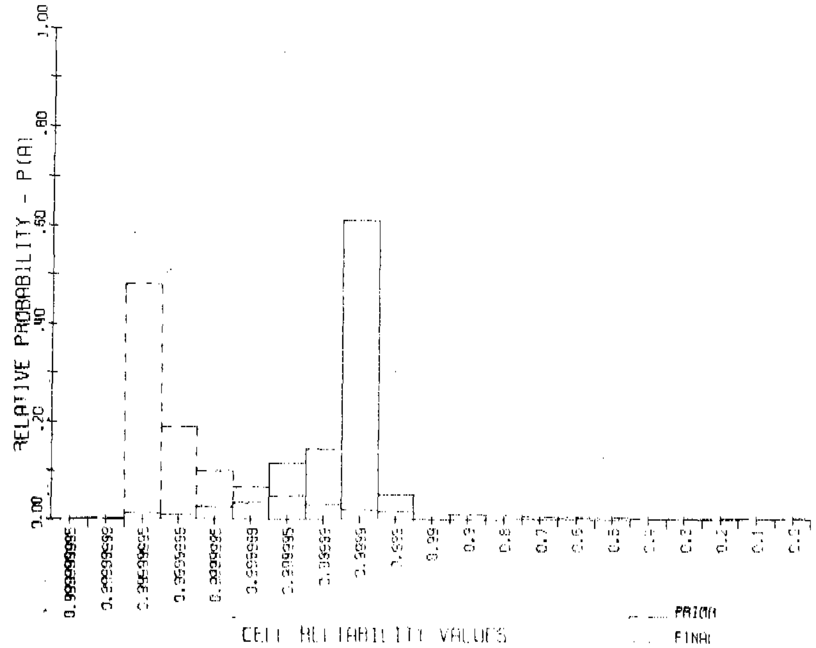
FIGURE(A.46). SKEWED BINOMIAL PRIOR,A,RESULTS OF 100 TESTS,1 FAILURE A+99S+1F



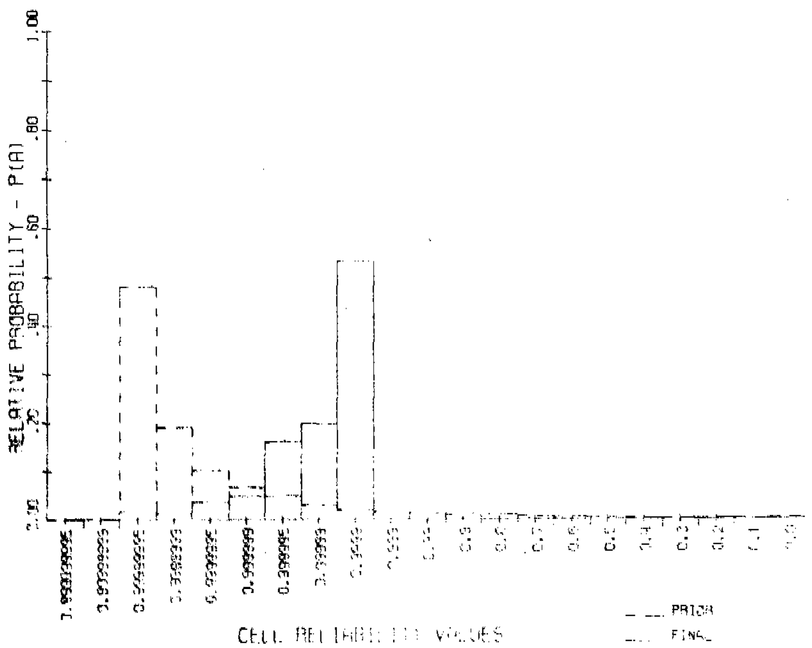
FIGURE(A.47). SKEWED BINOMIAL PRIOR,A,500 TESTS,1 FAILURE A+1F+.9995



FIGURE(A.48). SKEWED BINOMIAL PRIOR,A,1000 TESTS,1 FAILURE A+.9995+1F



FIGURE(A.49). SKEWED BINOMIAL PRIOR,A.5000 TESTS,1 FAILURE A+1F+4999S



FIGURE(A.50). SKEWED BINOMIAL PRIOR,A.10000 TESTS,1 FAILURE A+9999S+1F

TESTS WITH NO FAILURES AND BINOMIAL TYPE PRIOR DISTRIBUTION

CELL RELIABILITY VALUES, A_i	10000 S	
0.999 999 995	0.0036	
0.999 999 99	0.0145	
0.999 999 95	0.0559	
0.999 999 9	0.2625	
0.999 999 5	0.2698	
0.999 999	0.1827	
0.999 995	0.1276	
0.999 99	0.0658	
0.999 9	0.0175	
0.999	0.0000	
0.99	↑	
0.9		
0.8		
0.7		
0.6		
0.5		
0.4		
0.3		
0.2		
0.1	↓	
0.0	0.0000	

TABLE A.11 BAYESIAN CALCULATION RESULTS FOR 10000 TESTS WITH NO FAILURES, FOR BINOMIAL TYPE PRIOR DISTRIBUTION.

TESTS WITH NO FAILURES AND UNIFORM PRIOR DISTRIBUTION

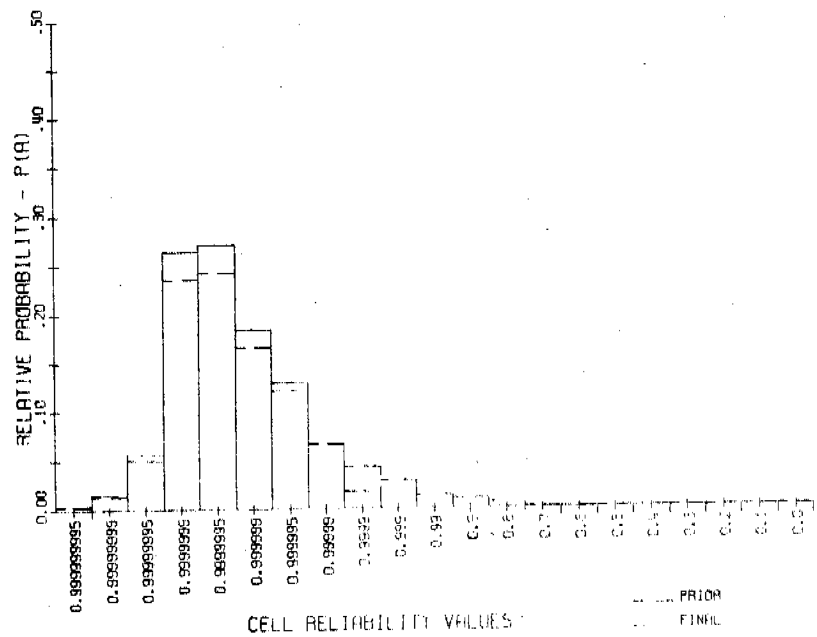
CELL RELIABILITY VALUES, A_i	1000 S	10,000 S
0.999 999 995	0.1425	0.1439
0.999 999 99	0.1425	0.1439
0.999 999 95	0.1425	0.1438
0.999 999 9	0.1425	0.1437
0.999 999 5	0.1424	0.1432
0.999 999	0.1423	0.1425
0.999 995	0.1418	0.1369
0.999 99	0.0018	0.0017
0.999 9	0.0013	0.0005
0.999	0.0005	0.0000
0.99	0.0000	
0.9		
0.8		
0.7		
0.6		
0.5		
0.4		
0.3		
0.2		
0.1		
0.0	0.0000	0.0000

TABLE A.12 BAYESIAN CALCULATION RESULTS FOR 1000 and 10,000 TESTS WITH NO FAILURES, FOR UNIFORM PRIOR DISTRIBUTION.

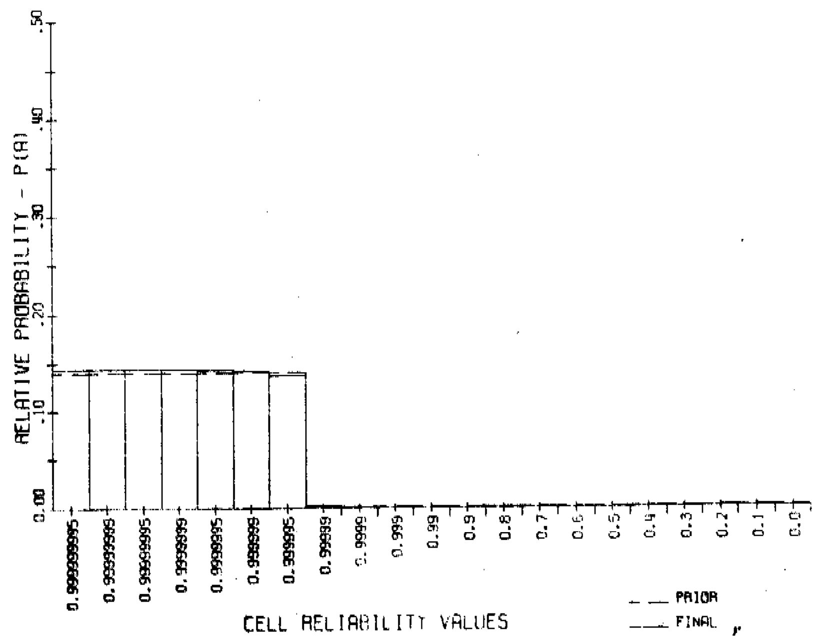
TESTS WITH NO FAILURES AND SKEWED BINOMIAL PRIOR DISTRIBUTION

CELL RELIABILITY VALUES, A_i	10000 S	
0.999 999 995	0.0011	
0.999 999 99	0.0011	
0.999 999 95	0.5240	
0.999 999 9	0.2073	
0.999 999 5	0.1057	
0.999 999	0.0714	
0.999 995	0.0488	
0.999 99	0.0296	
0.999 9	0.0090	
0.999	0.0000	
0.99	↑	
0.9		
0.8		
0.7		
0.6		
0.5		
0.4		
0.3		
0.2		
0.1	↓	
0.0	0.0000	

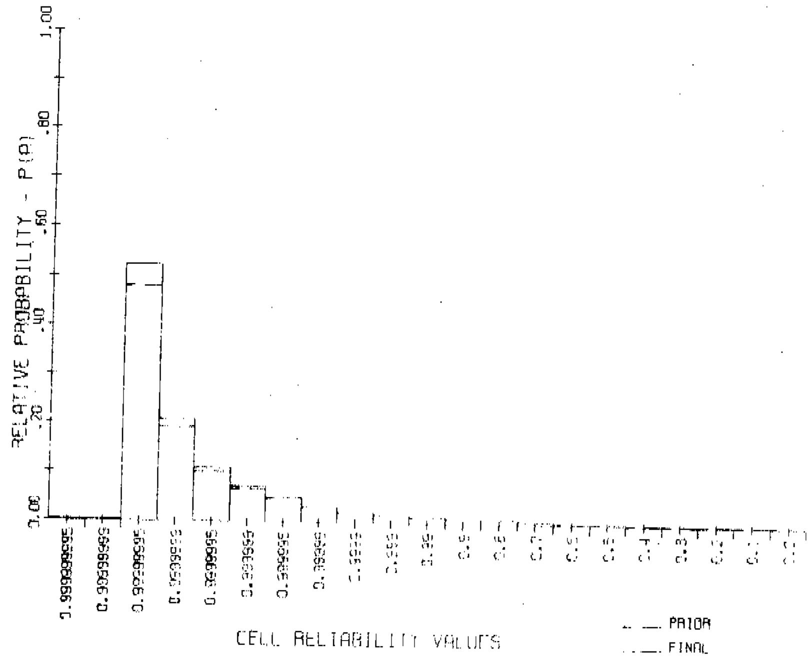
TABLE A.13 . BAYESIAN CALCULATION RESULTS FOR 10000 TESTS WITH NO FAILURES, FOR SKEWED BINOMIAL PRIOR DISTRIBUTION.



FIGURE(A.51). BINOMIAL PRIOR,A,10000 TESTS,NO FAILURES A+10000S



FIGURE(A.52). UNIFORM PRIOR,A,10000 TESTS,NO FAILURES A+10000S



FIGURE(A.53).SKEWED BINOMIAL PRIOR,A,10000 TESTS,NO FAILURES A+10000S

APPENDIX B

Sample of Bayesian Calculation Computer Program in FORTRAN IV

These calculations are for 10,000 tests in which the first nine tests and last test were failures with the remaining all successes, or 10 failures and 9990 successes.

```
FORTRAN 16K COMPILER PPN 07-825-0004-10-N
      DIMENSION A(24),B(24),P(24),C(24),PR(24),FI(24)
      DIMENSION IND (10000)
      READ(5,5)N1,N2
      READ(5,15) (A(I),I=1,N1)
      READ(5,15) (P(I),I=1,N1)
      READ(5,15) (B(I),I=1,N1)
    5 FORMAT(I3,I5)
   10 FORMAT(25I3)
   15 FORMAT(6D12.9)
      DO 60 I=1,N2
      IND(I)=1
      IF (I.EQ.1) IND (I) = 0
      IF (I.EQ.1) GO TO 19
      IF(I.EQ.2)IND(I)=0
      IF(I.EQ.3)IND(I)=0
      IF(I.EQ.4)IND(I)=0
      IF(I.EQ.5)IND(I)=0
      IF(I.EQ.6)IND(I)=0
      IF(I.EQ.7)IND(I)=0
      IF(I.EQ.8)IND(I)=0
      IF(I.EQ.9)IND(I)=0
      IF(I.EQ.N2)IND(I)=0
      IF(I.NE.N2) GO TO 25
   19 WRITE(6,20)
   20 FORMAT(1H1,10X,1HA,15X,4HP(A),16X,6HP(B/A),16X,10HP(B/A)P(A),
     116X,6HP(A/B)//)
   25 SUM=0.0
      DO 30 J=1,N1
      C(J)=B(J)*P(J)
      IF(IND(I).EQ.1) C(J)=A(J)*P(J)
```

```
30   SUM=SUM+C(J)
     DO 50 J=1,N1
     D=C(J)/SUM
     IF(I.EQ.1) GO TO 35
     IF(I.NE. N2) GO TO 47
     IF(IND(I) .EQ. 1) GO TO 45
 35  WRITE(6,40) A(J),P(J),B(J),C(J),D
40   FORMAT(1X,D16.9,F16.6,D26.9,F26.6,F24.6)
     IF(I.NE.1) GO TO 47
     PR(J)=P(J)
     IF(P(J).LT.0.0001) PR(J)=0.0001
     GO TO 47
45   WRITE(6,40) A(J),P(J),A(J),C(J),D
47   P(J)=D
     FI(J)=D
     IF(D.LT.0.0001) FI(J)=0.0001
50   CONTINUE
60   CONTINUE
     WRITE(2,70)
 70  FORMAT(47H TYPE A 1 FOR PUNCHED DATA, OTHERWISE TYPE A 0   )
     READ(2,71) IOP
 71  FORMAT(I1)

     IF(IOP.NE.1) GO TO 80
     WRITE(1,75) (PR(J),J=1,N1)
     WRITE(1,75) (FI(J),J=1,N1)
 75  FORMAT(10F8.5)
 80  CONTINUE
     STOP
     END
```

ROBUSTNESS STUDIES FOR BAYESIAN DEVELOPMENTS IN RELIABILITY

CHRIS P. TSOKOS AND A.N.V. RAO

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTH FLORIDA
TAMPA, FLORIDA 33620

ABSTRACT

The Weibull probability density function is considered as a failure model under the influence of a stochastic scale parameter. Bayesian estimates of the scale parameter, reliability function and hazard rate are given for the conjugate prior distribution. A method for evaluating the robustness of the conjugate prior distribution which characterizes the behavior of the parameter is presented. The proposed method is based on any acceptable level of the average mean square error of the Bayesian reliability estimate under the conjugate prior distribution. The analytical procedure utilizes the ratio of the average mean square error when the prior probability distribution is different from the conjugate prior (and its Bayes reliability estimate is used) to the average mean square error when he prior is assumed to be the conjugate distribution (and again the Bayes estimate of the reliability function is used).

A computer simulation to investigate the robustness of the conjugate prior distribution with respect to six other prior probability distributions, namely, beta, Poisson, inverted gamma, truncated normal, log-normal and extreme value, were employed. These results indicate that there is a significant variation in the average mean squared error even when the priors were chosen so that their first two moments approximately agreed with that of the conjugate.

This research was supported by the United States Air Force, Air Force Office of Scientific Research, under Grant No. AFOSR-74-2711.

1. *INTRODUCTION*

It has recently become quite evident that Bayesian reliability methods are very useful techniques in the reliability analysis and intepretation of various complex systems. A Bayesian analysis implies the expolitation of a suitable prior information in association with Bayes' Theorem. It rests on the notion that a parameter within a failure model is not simply an unknown fixed quantity but rather a stochastic variable which follows some probability distribution. In life testing problems, it is quite realistic to assume that a life parameter is stochastically dynamic. This assertation is strongly supported by the fact that the complexity of electronic or structural systems is likely to cause undetected component interactions resulting in unpredictable fluctuations of the life parameter.

Drake [4] and Evans, [5], have given an excellent account for the importance and implementation of Bayesian statistics in reliability problems. From a practical point of view, Feducia, [6], has given some constructive arguments concerning the Bayesian usefulness to reliability. Finally, Crelin, [3], among others has given an eloquent presentation of the basic philosophy and fruitfulness of the Bayesian approach to reliability.

Bhattacharya, [1], studied the exponential failure model under the assumption that the parameter behaved as a stochastic variable. Soland, [9], suggested a Bayesian analysis of the Weibull failure model. Canavos and Tsokos, [2], developed a Bayesian analysis of the scale and shape parameter in the Weibull failure model and its corresponding reliability function. For the scale parameter and reliability function, Bayesian estimates were obtained for various characterizations of the stochastic behavior of this parameter. Recently, Papadopoulos and Tsokos, [8], have developed the theory for constructing Bayesian confidence bounds for the random scale parameter and the reliability function of the Weibull failure model. The usefulness of the analytical findings were illustrated by employing a Monte Carlo simulation.

In the above studies among others, the most "popular" prior probability distribution to characterize the stochastic behavior of the scale parameter of the Weibull failure model is the conjugate distribution. Perhaps the most important reason for employing such a prior probability density is because the Bayesian estimates of the scale parameter and reliability function result in an analytical tractable form. Thus, this popularity encourages the investigation of what happens if the prior probability distribution is different from the assigned conjugate, but we proceed to employ the Bayes reliability estimator. Also, how "good" is the Bayes estimate under different probability distributions that characterize the random behavior of the scale parameter.

It is the aim of the present study to investigate the significance of the above questions. We shall propose to utilize the ratio of the average mean square errors as a measure of "goodness" of the Bayes reliability estimator. In Section 2 we introduce some of the preliminary concepts essential to our investigations. The Bayes estimates of the *scale parameters*, *reliability function* and the *hazard rate* are given in Section 3. In Section 4 we propose an analytical approach to investigate the robustness of the conjugate prior. A similar procedure to study the behavior of the Bayesian estimate of the hazard rate is given in Section 5. In Section 6 we present a brief computer simulation in an attempt to illustrate the usefulness of the basic analytical developments. The conclusions and a summary of our findings in the present study are given in Section 7.

2. PRELIMINARY CONCEPTS

In this section we shall summarize some known results of the *Weibull* failure model that we will be utilizing in fulfilling the aims of the present study.

Consider the Weibull failure model with the time to failure probability density function given by

$$f(t;\alpha,\theta) = \alpha\theta t^{\alpha-1} \exp\{-\theta t^{\alpha}\}, \qquad \begin{array}{l} 0 < t < \infty \\ 0 < \alpha,\ \theta \end{array} \tag{2.1}$$

where θ and α are the *scale* and *shape* parameters, respectively. It is well known that the *reliability function*, $R(t;\alpha,\theta)$ and the *hazard rate*, $\rho(t;\alpha,\theta)$ of the above failure model are given by

$$R(t;\alpha,\theta) = \exp\{-\theta t^{\alpha}\} \tag{2.2}$$

and

$$\rho(t;\alpha,\theta) = \alpha\theta t^{\alpha-1}, \tag{2.3}$$

respectively.

Suppose now that the scale parameter θ behaves as a *stochastic variable* and that the shape parameter α is known or can be estimated. To estimate the *Bayes* reliabilty function under the described situation we conduct a life test on n items and observe the ordered times to failure r ($\leq$ n) items to be $t_1, t_2, t_3, \ldots, t_r$. Let

$$\underline{t} = (t_1, t_2, \ldots, t_r). \tag{2.4}$$

Then the likelihood function, $L(\underline{t};\alpha,\theta)$, of the complete sample is given by

$$L(\underline{t};\alpha,\theta) = \frac{n!}{(n - r)!} (\alpha\theta)^r \prod_{i=1}^{r} t_i^{\alpha-1} \exp\{-\theta T_r\} \tag{2.5}$$

where

$$T_r = \sum_{i=1}^{r} t_i^{\alpha} + (n - r)t_r^{\alpha} . \tag{2.6}$$

It is known, [2], that the probability density function of T_r is of the form

$$f_r(y|\theta) = \frac{1}{(r - 1)!} \theta^r y^{r-1} \exp\{-\theta y\}, \quad 0 < y < \infty . \tag{2.7}$$

3. GAMMA PRIOR DISTRIBUTION FOR THE SCALE PARAMETER

Let us assume that the stochastic scale parameter, Θ, of the Weibull failure model, (2.1), is being characterized by the *natural conjugate* distribution, that is, the *gamma** probability density function given by

$$g(\theta) = \begin{cases} \dfrac{\mu^{\nu}}{\Gamma(\nu)} \theta^{\nu-1} \exp\{-\mu\theta\}, & 0 < \theta < \infty \\ 0, \text{ elsewhere} & 0 < \mu, \nu \end{cases}$$

Then the *posterior* probability density function of Θ given $\underline{t}$ is

$$g_1(\theta;\alpha|\underline{t}) = \frac{\theta^{-(r+\nu+1)} \exp\{-\frac{1}{\theta}(T_r+\mu)\}}{\int_0^{\infty} \frac{1}{\theta^{-(r+\nu+1)}} \exp\{-\frac{1}{\theta}(T_r+\mu)\} d\theta}$$

$$= \frac{1}{(T_r+\mu)\,\Gamma(r+\nu)} \left\{\frac{T_r+\mu}{\theta}\right\}^{r+\nu+1} \exp\left\{-\frac{1}{\theta}(T_r+\mu)\right\} \quad (3.2)$$

where T_r is defined by equation (2.6)

---

*Note if the failure model (2.1) had been given in the form of

$$f(t;\alpha,\theta) = \frac{\alpha}{\theta} t^{\alpha-1} \exp\left\{-\frac{t^{\alpha}}{\theta}\right\}, \qquad \begin{matrix} 0 < t < \infty \\ 0 < \alpha, \theta \end{matrix}$$

then the conjugate prior distribution will be the inverted gamma, that is,

$$g(\theta) = \frac{(\frac{\mu}{\theta})^{\nu+1}}{\mu\Gamma(\nu)} \exp\left\{-\frac{\mu}{\varepsilon}\right\} \qquad \begin{matrix} 0 < \theta < \infty \\ 0 < \mu, \nu \end{matrix}$$

For Θ to be characterized by the natural conjugate prior and for a *square error loss function*, it is known, [2], that the *Bayes estimator*, $\hat{\theta}_c(t)$, of the scale parameter θ is given by the conditional expectation of Θ given $\underline{t}$. That is,

$$\hat{\theta}_c(t) = E\{\Theta|\underline{t}\} = \frac{T_r + \mu}{r + \nu - 1} , \; 1 < r + \nu \; . \tag{3.3}$$

Also under the above conditions the *Bayes reliability estimator*, $\hat{R}_c(t)$, is given by the conditional expectation of the reliability function, (2.2), given $\underline{t}$, (2.4). It can be easily calculated to be

$$\hat{R}_c(t) = E\{R(t;\alpha|\theta)|\underline{t}\}$$

$$= \frac{1}{\left(1 + \frac{t^{\alpha}}{T_r + \mu}\right)^{r+\nu}} \; . \tag{3.4}$$

Similarly, the *Bayes hazard rate estimate*, $\hat{\rho}_c(t)$, is given by

$$\hat{\rho}_c(t) = E\{\rho(t;\alpha|\theta)|\underline{t}\}$$

$$= \frac{\alpha t^{\alpha-1}(r + \nu)}{(T_r + \mu)} \; . \tag{3.5}$$

In the following sections of the present study we shall employ the above Bayesian estimates to investigate the robustness of the Bayesian reliability estimate as we deviate from the assumption that the scale parameter is being characterized by the gamma distribution.

4. ANALYTICAL APPROACH TO ROBUSTNESS

In the preceding sections we have obtained estimates for the Bayesian reliability and hazard rate estimators for the Weibull failure model under the assumption that the stochastic behavior of the scale parameter is being characterized by the conjugate distribution. Perhaps one of the main reasons for choosing such a prior probability density is because it is analytically tractable. Thus, it becomes quite significant to ask:

(i) *What if the prior distribution is different from the assigned natural conjugate, but we employ the Bayes reliability estimator,* $\hat{R}_c(t)$, *developed for the inverted gamma?*

(ii) *How "good" is the estimate* $\hat{R}_c(t)$ *under a different prior probability distribution for the scale parameter?*

To answer this question, we shall propose to employ the ratio of the *average mean square errors* as a measure of "goodness" of the Bayes reliability estimator, $\hat{R}_c(t)$. That is, the ratio of the average mean square error when the prior is different from the conjugate prior (and $\hat{R}_c(t)$ is used as the reliability estimate) to the average mean square error when the prior is assumed to be the conjugate (and $\hat{R}_c(t)$ is again used as the Bayesian reliability estimate). Thus, it is clear that the closer this ratio is to one, the more robust is the Bayesian reliability estimate, $\hat{R}_c(t)$.

We proceed by developing an expression for the average mean square error in terms of the moments of the prior probability distribution which characterizes the behavior of the scale parameter.

The definition of the average mean square error, $\overline{M.S.E.}$, is given by

$$\overline{M.S.E.} = E_\Theta\{E_{\underline{t}}[R(t;\alpha,\theta) - \hat{R}_c(t)]^2|\theta\}.$$

More conveniently the $\overline{M.S.E.}$ can be written in the following form

$$\begin{aligned}\overline{M.S.E.} &= E_\Theta\{E_{\underline{t}}[R^2(t;\alpha,\theta)|\theta]\}\\ &\quad - 2\, E_\Theta\{R(t;\alpha,\theta)E_{\underline{t}}[\hat{R}_c(t)|\theta]\}\\ &\quad + E_\Theta\{E_{\underline{t}}[\hat{R}_c^2(t)]|\theta\}. \end{aligned} \tag{4.1}$$

We proceed by working with the term

$$E_\Theta\{R(t;\alpha,\theta)E_{\underline{t}}[\hat{R}_c(t)|\theta]\}$$

of equation (4.1). Utilizing equations (2.2) and (2.7) we can write the above expression as

$$\begin{aligned}&E_\Theta\{R(t;\alpha,\theta)E_{\underline{t}}[\hat{R}_c(t)|\theta]\}\\ &= E_\Theta\{R(t;\alpha,\theta) \int_0^\infty \frac{\theta^r}{\Gamma(r)} \frac{(T_r+\mu)^{r+\nu}}{(T_r+\mu+t^\alpha)^{r+\nu}} T_r^{r-1} \exp[-\frac{T_r}{\theta}]dT_r\}. \end{aligned} \tag{4.2}$$

Let us consider the integral

$$\int_0^\infty \frac{\theta^r}{\Gamma(r)} \frac{(y+\mu)^{r+\nu}}{(y+\mu+t^\alpha)^{r+\nu}} y^{r-1} \exp\{-\frac{y}{\theta}\}\, d\theta. \tag{4.3}$$

We can write the above integral, equation (4.3), as follows

$$\int_0^\infty \frac{\theta^r}{\Gamma(r)} \frac{(y+\mu)^{r+\nu}}{(y+\mu+t^\alpha)^{r+\nu}} y^{r-1} \exp\{-\frac{y}{\theta}\}\, d\theta$$

$$= \int_0^N \frac{\theta^r}{\Gamma(r)} \frac{(y+\mu)^{r+\nu}}{(y+\mu+t^\alpha)^{r+\nu}} y^{r-1} \exp\{-\frac{y}{\theta}\}\, d\theta$$

$$+ \int_N^\infty \frac{\theta^r}{\Gamma(r)} \frac{(y+\mu)^{r+\nu}}{(y+\mu+t^\alpha)^{r+\nu}} y^{r-1} \exp\{-\frac{y}{\theta}\}\, d\theta$$

$$\equiv I + II. \qquad (4.4)$$

We shall show that for any $\varepsilon > 0$, we can find an $N > 0$ such that

$$E_\Theta\{R(t;\alpha,\theta)\ II\} < \varepsilon$$

under the assumption that the probability density function of Θ is bounded. Thus, let the probability density function of Θ be $g(\theta)$, such that $g(\theta) < k$, $0 \le \theta < \infty$. Then

$$E_\Theta\{R(t;\alpha,\theta)\ II\}$$

$$\equiv \int_0^\infty \exp\{-\theta t^\alpha\} g(\theta) d\theta$$

$$\cdot \int_N^\infty \frac{\theta^r}{\Gamma(r)} \frac{(y+\mu)^{r+\nu}}{(y+\mu+t^\alpha)^{r+\nu}} y^{r-1} \exp\{-\theta y\} dy$$

$$\le K \int_0^\infty \{\exp -\theta t^\alpha\} \Gamma(r,N\theta) d\theta$$

$$= K \frac{\Gamma(r)}{t^{\alpha}} \{1 - \frac{N^{r}}{(N+t^{\alpha})^{r}}\} \qquad (4.5)$$

where $\Gamma(r,x)$ is the *incomplete gamma function*,

$$\int_{t}^{\infty} \exp\{-t\}\, t^{r-1}\, dt.$$

From equation (4.5) it is clear that by choosing N sufficiently large, the right hand side of the inequality can be made smaller than ε. In fact it can be shown that if

$$N \geq K \frac{\{1 - \frac{\varepsilon t^{\alpha}}{\Gamma(r)}\}^{\frac{1}{r}}}{1 - \{1 - \frac{\varepsilon t^{\alpha}}{\Gamma(r)}\}^{\frac{1}{r}}} \qquad (4.6)$$

then

$$E_{\Theta}\{R(t;\alpha,\theta)\ II\} \leq \varepsilon.$$

Similarly, it can be shown that if N is chosen to satisfy the inequality (4.6), then

$$E_{\Theta}\{E_{\underline{t}}[\hat{R}^{2}(t)|\theta]\} \leq$$

$$E_{\Theta} \int_{0}^{N} \hat{R}_{c}^{2}(t)\ \frac{\theta^{r}}{\Gamma(r)}\ T_{r}^{\,r-1 - \frac{T_r}{\theta}}\ dT_{r} + \varepsilon. \qquad (4.7)$$

where $\mu^{(n)}$ is the nth ordinary (noncentral) moment of the prior probability distribution function of the stochastic variate Θ, $g(\theta)$.

Thus, inequality (4.9) gives an approximate value for the average mean square error in terms of the moments of the prior probability distribution.

Average Mean Square Error For The Conjugate Distribution

Using the definition of $\overline{\text{M.S.E.}}$, equation (4.1), the reliability function of the Weibull failure model, equation (2.2) and the conjugate prior distribution, equation (3.1) we have

$$E_{\Theta}\{R(t;\alpha,\theta)\ E_{\underline{t}}[\hat{R}_c(t)|\theta]\}$$

$$= \int_0^{\infty} \frac{\mu^{\nu}}{\Gamma(\nu)}\ \theta^{\nu-1} \exp[-\mu\theta]\ \exp[-\theta t^{\alpha}]$$

$$\cdot\{\int_0^{\infty} \frac{\theta^{r}}{\Gamma(r)}\ T_r^{\,r-1} \exp[-T_r\theta][\frac{T_r+\mu}{T_r+\mu+t^{\alpha}}]^{r+\nu} dT_r\}d\theta$$

$$= \frac{\mu}{\Gamma(\nu)}\ \frac{1}{\Gamma(r)} \int_0^{\infty} (\frac{T_r+\mu}{T_r+\mu+t^{\alpha}})^{r+\nu}\ T_r^{\,r-1}\ \frac{\Gamma(r+\nu)}{(T_r+\mu+t^{\alpha})^{r+\nu}}\ dT_r$$

$$= \frac{\mu^{\nu}\,\Gamma(r+\nu)}{\Gamma(r)\Gamma(\nu)}\ \mu^{r}\ B(\nu,r)\ {}_2F_1[2(r+\nu),\ r;\ r+\nu;\ -\frac{t^{\alpha}}{\mu}]$$

$$= \mu^{\nu+r}\ B(r,\nu)\}^{2}\ {}_2F_1[2(r+\nu),\ r;\ r+\nu;\ -\frac{t^{\alpha}}{\mu}]$$

Using inequalities (4.6) and (4.7) in equation (4.1) we can obtain an approximation to the average mean square error, that is,

$$\overline{\text{M.S.E.}} \leq E_\Theta \left\{\int_0^N \frac{\theta^r}{\Gamma(r)} \frac{(T_r+\mu)^{2(r+\nu)}}{(T_r+\mu+t^\alpha)^{2(r+\nu)}} T_r^{r-1}\{\exp -T_r\theta\}dT_r\right\}$$

$$+ 2 E_\Theta \left\{\int_0^N \exp[-\theta t^\alpha] \frac{\theta^r}{\Gamma(r)} \frac{(T_r+\mu)^{r+\nu}}{(T_r+\mu+t^\alpha)^{r+\nu}} T_r \exp\{-T_r\theta\}dT_r\right\}$$

$$- E_\Theta\{\exp[-2\theta t^\alpha]\} + 2\varepsilon. \qquad (4.8)$$

Let

$$A_k = \int_0^N \frac{(-1)^k}{\Gamma(r)} \frac{y^{k+r-1}}{k!} \left(\frac{y+\mu}{y+\mu+t^\alpha}\right)^{2(r+\nu)} dy,$$

$$B_k = \int_0^N \frac{(-1)^k}{\Gamma(r)} \frac{(t^\alpha+y)^{k+r-1}}{k!} \left(\frac{y+\mu}{y+\mu+t^\alpha}\right)^{r+\nu} dy$$

and

$$C_k = \frac{(-1)^k}{k!} (2t^\alpha)^k.$$

Using these definitions, inequality (4.8) can be written as follows

$$\overline{\text{M.S.E.}} \leq \sum_{k=0}^{\infty} \{A_k - 2 B_k + C_k\} \mu^{(r+k)} + 2\varepsilon \qquad (4.9)$$

where $B(x,y) = \int_0^1 t^{x-1} (1 - t)^{y-1} \, dt$

and

$$F(\alpha, \beta; V; z)$$

is a hypergeometric function. For some details see [7, p. 287].
Similarly,

$$E_\Theta\{R_c^{\,2}(t) \mid \theta\} = \frac{\mu^\nu}{\Gamma(\nu)} \frac{1}{\Gamma(r)} \int_0^\infty \left(\frac{T_r+\mu}{T_r+\mu+t^\alpha}\right)^{2(r+\nu)} T_r^{\,r-1}$$

$$\left\{\int_0^\infty \theta^{r+\nu-1} \exp[-(\mu+T_r)\theta] \, d\theta\right\} dT_r$$

$$= \mu^{\nu+r} \{B(\nu, r)\}^2 \; {}_2F_1[2(\nu+r), r; r+\nu; -\frac{t^\alpha}{\mu}].$$

Also,

$$E_\Theta\{\exp[-2t^\alpha \theta]\}$$

$$= \int_0^\infty \exp[-2t^\alpha \theta] \frac{\mu^\nu}{\Gamma(\nu)} \theta^{\nu-1} e^{-\mu\theta} \, d\theta$$

$$= \frac{\mu^\nu}{\Gamma(\nu)} \frac{\Gamma(\nu)}{(\mu+2t^\alpha)^\nu}$$

$$= \frac{1}{\left(1 + 2\,\frac{t^\alpha}{\mu}\right)^\nu} .$$

Hence, the average mean square error, $\overline{M.S.E.}$, for the conjugate prior distribution is given by

$$\overline{M.S.E}_c. = \frac{1}{(1+2\frac{t^\alpha}{\mu})^\nu} - \mu^{\nu+r}\{B(\nu,r)\}^2$$

$$\cdot {}_2F_1[2(r+\nu),\ r;\ r+\nu;\ -\frac{t^\alpha}{\mu}]. \tag{4.9a}$$

Thus, as we have mentioned above, a measure of the robustness, $R(t)$, is the ratio of the mean square error when the prior is different from the conjugate, $\overline{M.S.E.}$, to the average mean square error when the prior is assumed to be the conjugate, $\overline{M.S.E}_c.$ That is,

$$R(t) = \frac{\overline{M.S.E.}}{\overline{M.S.E}_c.} \tag{4.10}$$

For the conjugate prior we have

$$R(t) = \frac{\overline{M.S.E.}}{(1+2\frac{t^\alpha}{\mu})^{-\nu} - \mu^{\nu+r}\{B(\nu,r)\}^2 {}_2F_1[2(r+\nu),\ r;\ r+\nu,\ -\frac{t^\alpha}{\mu}]}.$$

Unfortunately, the $\overline{M.S.E.}$ for most priors other than the conjugate cannot be given in a closed mathematical form. Thus, we must rely on approximating the $\overline{M.S.E.}$ for other prior distributions using equation (4.9) in conjunction with electronic computers.

5. HAZARD RATE BEHAVIOR

In this section we shall propose to use the method we discussed in the previous section to investigate the robustness of the Bayesian estimate of the hazard rate using the prior conjugate with respect to other prior distributions. Namely, we shall employ the ratio of the average mean square error when the prior is different from the conjugate prior (and $\hat{\rho}_c(t)$ is used as the estimate of the hazard rate) to the average mean square error when the prior is assumed to be the conjugate (and $\hat{\rho}_c(t)$ is again used as the Bayesian hazard rate estimate).

The average mean square error, $\overline{M.S.E.}_H$ for the hazard rate is given by

$$\overline{M.S.E.}_H = E_\Theta\{E_{\underline{t}}[\rho(t;\alpha,\theta) - \hat{\rho}_c(t)]^2|\theta\}$$

$$= E_\Theta\{\rho(t;\alpha,\theta)\}^2 - 2E_\Theta\{\rho(t;\alpha,\theta)E_{\underline{t}}[\hat{\rho}_c(t)|\theta]\}$$

$$+ E_\Theta\{E_{\underline{t}}[\hat{\rho}_c(t)]^2|\theta\}. \qquad (5.1)$$

Recall that the Bayes estimate of the hazard rate is given by

$$\hat{\rho}_c(t) = \frac{\alpha t^{\alpha-1}(r+\nu)}{(T_r+\mu)}$$

and that T_r follows the gamma distribution, (2.7). Following the procedure we have employed in the previous section we can calculate the average mean square error for the Bayes estimate of the hazard rate. That is, from equation (5.1) we have

$$\overline{M.S.E.}_H = \alpha^2 t^{2(\alpha-1)}\{E(\Theta^2) - 2 \sum_{k=0}^{\infty} C_k E(\Theta^{k+r+1}) + \sum_{k=0}^{\infty} D_k E(\Theta^k)\} \tag{5.2}$$

where

$$C_k = \int_0^N (-1)^k \frac{(r+\nu)}{(r-1)!} \frac{y^k}{k!} \frac{1}{(y+\mu)} dy$$

and

$$D_k = \int_0^N \frac{(-1)^k}{(r-1)!} \frac{y^k}{k!} \frac{1}{(y+\mu)^2} dy.$$

Thus, for a specified (very) small ε we can obtain an N that satisfies inequality (4.6) and one can proceed to calculate the $\overline{M.S.E.}_H$ for a specified probability distribution for the parameter Θ using equation (5.2).

Average Mean Square Error of the Hazard Rate For the Conjugate Distribution

The average mean square error for the Bayesian hazard rate estimate for the conjugate prior, $\overline{M.S.E.}_{H_c}$; can be obtained to be

$$\overline{M.S.E.}_{H_c} = \alpha^2 t^{2(\alpha-1)}\{ \frac{\alpha(\alpha+1)}{\mu^2} - \frac{(r+\nu)^2}{\mu^{2(r+1)}} \frac{B(r, \nu+2)}{B(r,\nu)} \} \tag{5.3}$$

where B(x,y) is the beta function defined in Section 4.

Thus, one can employ a procedure similar to the one we discussed previously to study the robustness of the Bayesian hazard rate of a conjugate prior with respect to other prior probability distributions by forming the ratio of $\overline{M.S.E.}_H$ to $\overline{M.S.E.}_{H_c}$.

6. *COMPUTER SIMULATION*

As we have pointed out in the previous sections to investigate the robustness of the natural conjugate with respect to different prior probability distributions that characterize the behavior of the scale parameter we propose to study the ratio of the average mean square error with a non-conjugate prior, $\overline{\text{M.S.E.}}$ to the average mean square error using a conjugate prior, $\overline{\text{M.S.E.}}_C$. That is, the measure of robustness, $R(t)$ is given by

$$R(t) = \frac{\overline{\text{M.S.E.}}}{\overline{\text{M.S.E.}}_C} .$$

Our sensitivity analysis will be determined by a careful computer simulation. A brief description of the simulation procedure is given below:

i) A very small ε is specified and an N is obtained which satisfies inequality (4.6).

ii) For specific values of the parameters and the conjugate prior the $\overline{\text{M.S.E.}}_C$ is calculated using equation (4.9a).

iii) For the different prior probability distributions given in Table 6.1 the $\overline{\text{M.S.E.}}$ is calculated using equation (4.9).

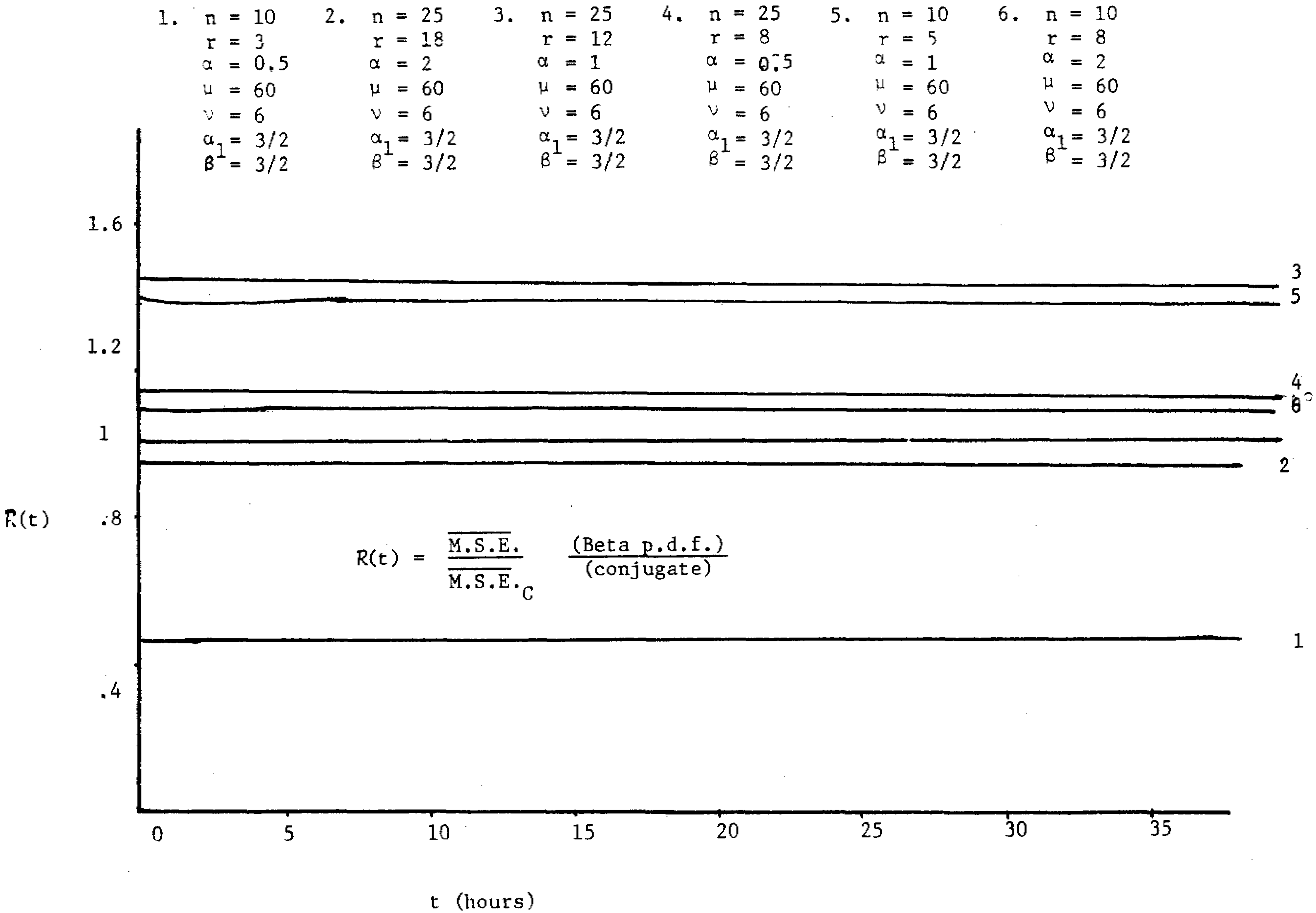
iv) The measure of robustness, $R(t)$ is calculated and plotted as a function of time for different configurations of the parameters.

Table 6.1 gives the different prior distributions along with the various values of the parameters that we have used in the simulation. These values are chosen so that the mean and variance of the specified prior is approximately close to the mean and variance of the natural conjugate prior. For the Weibull failure model the shape parameter was fixed at $\alpha = \frac{1}{2}$, 1, 2. The parameters of the natural conjugate were $\mu = 60$ and $\nu = 6$.

Figures 6.1 - 6.6 are representative samples of the proposed measure of robustness, $\mathcal{R}(t)$, of the conjugate prior distribution, with respect to the following different priors; *beta, Poisson, inverted gamma, truncated normal, log-normal* and the *extreme value*.

Table 6.1

Prior Distribution For θ	*Form Of Probability Density Function*	*Values of Parameters*
Beta p.d.f.	$g_1(\theta) = \frac{\Gamma(\alpha_1+\beta)}{\Gamma(\alpha_1)\Gamma(\beta)} \theta^{\alpha_1 - 1}(1-\theta)^{\beta-1}$; $0 < \theta < \infty$, $0 < \alpha_1, \beta$	$\alpha_1 = 3/2$, $\beta = 3/2$
Poisson p.d.f.	$g_2(\theta) = \frac{e^{-\lambda}\lambda^{\theta}}{\theta!}$; $\theta = 0,1,2,...$	$\lambda = 13$
Inverted Gamma p.d.f.	$g_3(\theta) = \frac{(\frac{\alpha_2}{\theta})^{\beta_1+1}}{\alpha_2\Gamma(\beta_1)} \exp\{-\frac{\alpha_2}{\theta}\}$; $0 < \theta < \infty$, $0 < \alpha_2, \beta_1$	$\alpha_2 = 4$, $\beta_1 = 3$
Truncated Normal p.d.f.	$g_4(\theta) = \frac{d}{\sigma\sqrt{2\pi}} \exp\{-\frac{(\theta - \mu_1)^2}{2\sigma^2}\}$; $-\infty < \mu_1 < \infty$, $0 < \sigma$, $0 < \theta < \infty$ (d is a truncation factor)	$\mu_1 = 12$, $\sigma = 6$
Log-Normal p.d.f.	$g_5(\theta) = \frac{1}{\theta\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2\sigma^2}(\ell n\theta - \alpha_3)^2\}$; $0 < \theta < \infty$, $-\infty < \alpha_3 < \infty$, $0 < \sigma$	$\alpha_3 = \ell n\ 12 - \frac{\sigma}{2}$ $\sigma = \ell n\ 5 - \ell n\ 4$
Extreme Value p.d.f.	$g_6(\theta) = \frac{1}{\lambda_1} \exp\{-\frac{e^{\theta}-1}{\lambda_1} + \theta\}$; $0 < \theta < \infty$, $0 < \lambda_1$	$\lambda_1 = e^{13}$



1. n = 10, r = 3, α = 0.5, μ = 60, ν = 6, α1 = 3/2, β1 = 3/2
2. n = 25, r = 18, α = 2, μ = 60, ν = 6, α1 = 3/2, β1 = 3/2
3. n = 25, r = 12, α = 1, μ = 60, ν = 6, α1 = 3/2, β1 = 3/2
4. n = 25, r = 8, α = 0.5, μ = 60, ν = 6, α1 = 3/2, β1 = 3/2
5. n = 10, r = 5, α = 1, μ = 60, ν = 6, α1 = 3/2, β1 = 3/2
6. n = 10, r = 8, α = 2, μ = 60, ν = 6, α1 = 3/2, β1 = 3/2
R(t)
1.6
1.2
1
.8
.4
R(t) = M.S.E. / M.S.E.C (Beta p.d.f.) / (conjugate)
0
5
10
15
20
25
30
35
t (hours)
3
5
4
6
2
1

Figure 6.1

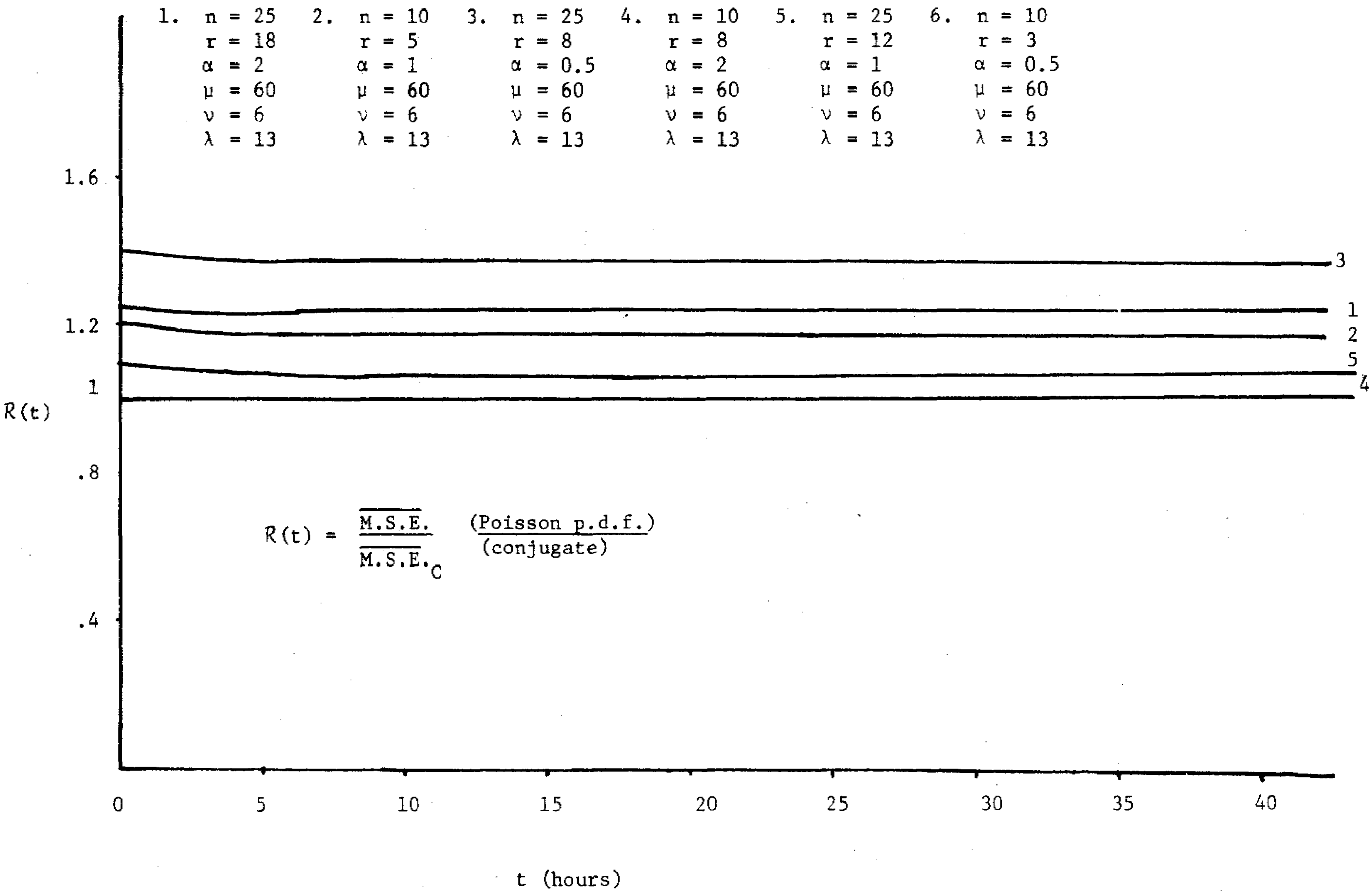

1. n = 25 r = 18 α = 2 μ = 60 ν = 6 λ = 13
2. n = 10 r = 5 α = 1 μ = 60 ν = 6 λ = 13
3. n = 25 r = 8 α = 0.5 μ = 60 ν = 6 λ = 13
4. n = 10 r = 8 α = 2 μ = 60 ν = 6 λ = 13
5. n = 25 r = 12 α = 1 μ = 60 ν = 6 λ = 13
6. n = 10 r = 3 α = 0.5 μ = 60 ν = 6 λ = 13
R(t)
1.6
1.2
1
.8
.4
R(t) = M.S.E. / M.S.E.C (Poisson p.d.f.) / (conjugate)
0
5
10
15
20
25
30
35
40
3
1
2
5
4
t (hours)

Figure 6.2

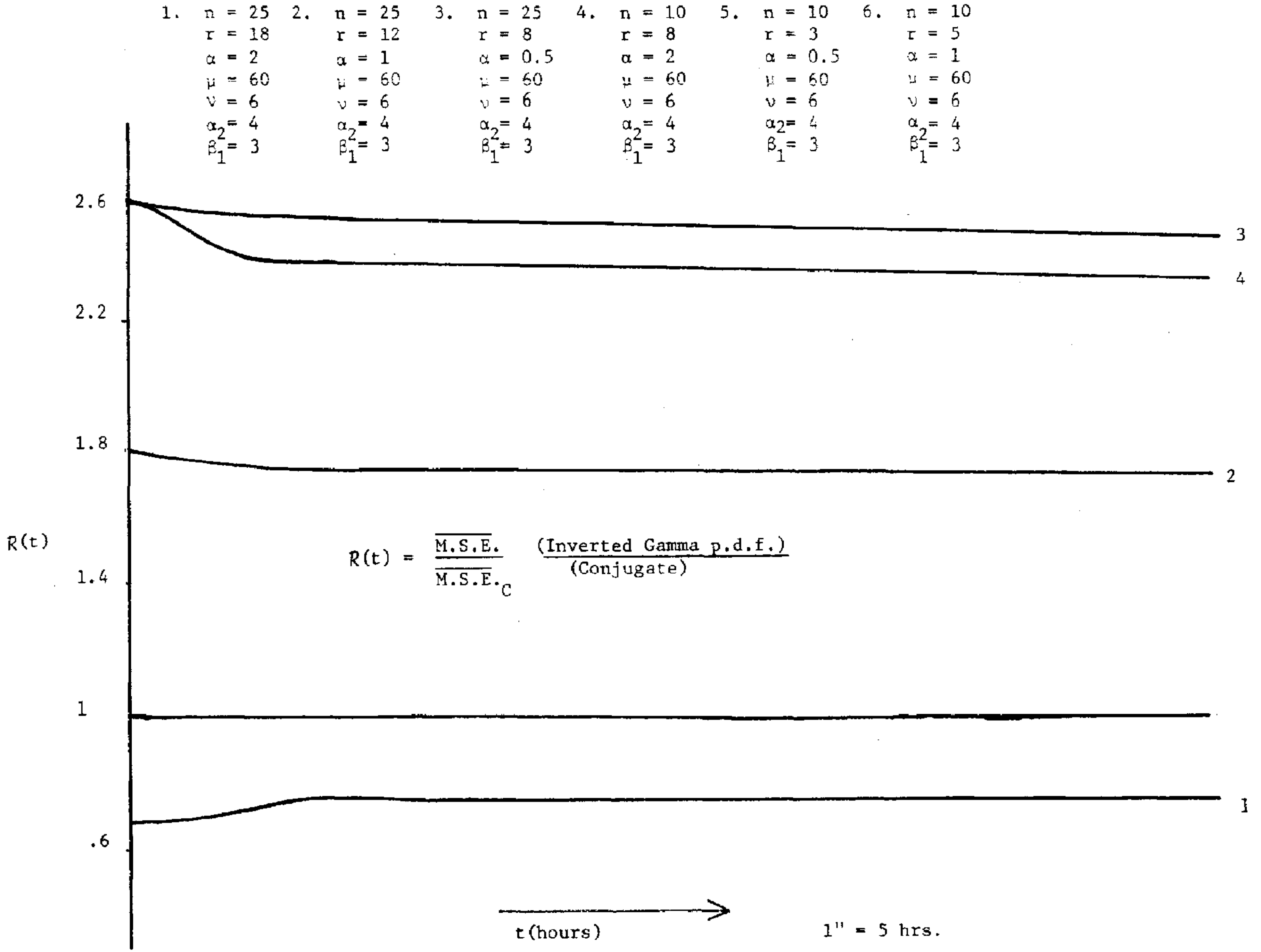
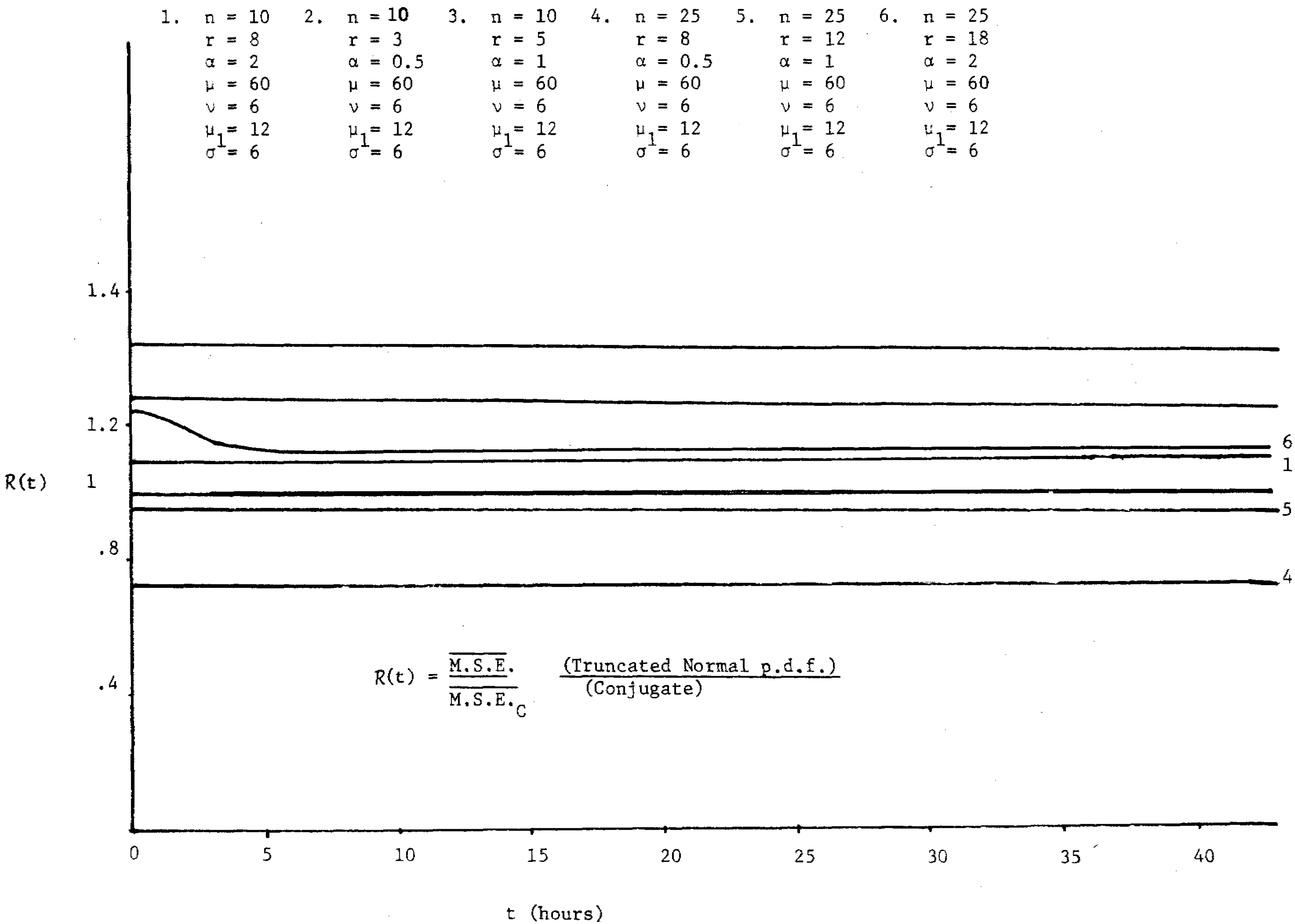

1. n = 25 r = 18 α = 2 μ = 60 ν = 6 α2 = 4 β1 = 3
2. n = 25 r = 12 α = 1 μ = 60 ν = 6 α2 = 4 β1 = 3
3. n = 25 r = 8 α = 0.5 μ = 60 ν = 6 α2 = 4 β1 = 3
4. n = 10 r = 8 α = 2 μ = 60 ν = 6 α2 = 4 β1 = 3
5. n = 10 r = 3 α = 0.5 μ = 60 ν = 6 α2 = 4 β1 = 3
6. n = 10 r = 5 α = 1 μ = 60 ν = 6 α2 = 4 β1 = 3
2.6
2.2
1.8
1.4
1
.6
R(t)
R(t) = M.S.E. / M.S.E.C (Inverted Gamma p.d.f.) / (Conjugate)
3
4
2
1
t(hours)
1" = 5 hrs.

Figure 6.3



1. n = 10, r = 8, α = 2, μ = 60, ν = 6, μ1 = 12, σ1 = 6
2. n = 10, r = 3, α = 0.5, μ = 60, ν = 6, μ1 = 12, σ1 = 6
3. n = 10, r = 5, α = 1, μ = 60, ν = 6, μ1 = 12, σ1 = 6
4. n = 25, r = 8, α = 0.5, μ = 60, ν = 6, μ1 = 12, σ1 = 6
5. n = 25, r = 12, α = 1, μ = 60, ν = 6, μ1 = 12, σ1 = 6
6. n = 25, r = 18, α = 2, μ = 60, ν = 6, μ1 = 12, σ1 = 6
R(t)
1.4
1.2
1
.8
.4
6
1
5
4
R(t) = M.S.E. / M.S.E.C (Truncated Normal p.d.f.) / (Conjugate)
0
5
10
15
20
25
30
35
40
t (hours)

Figure 6.4

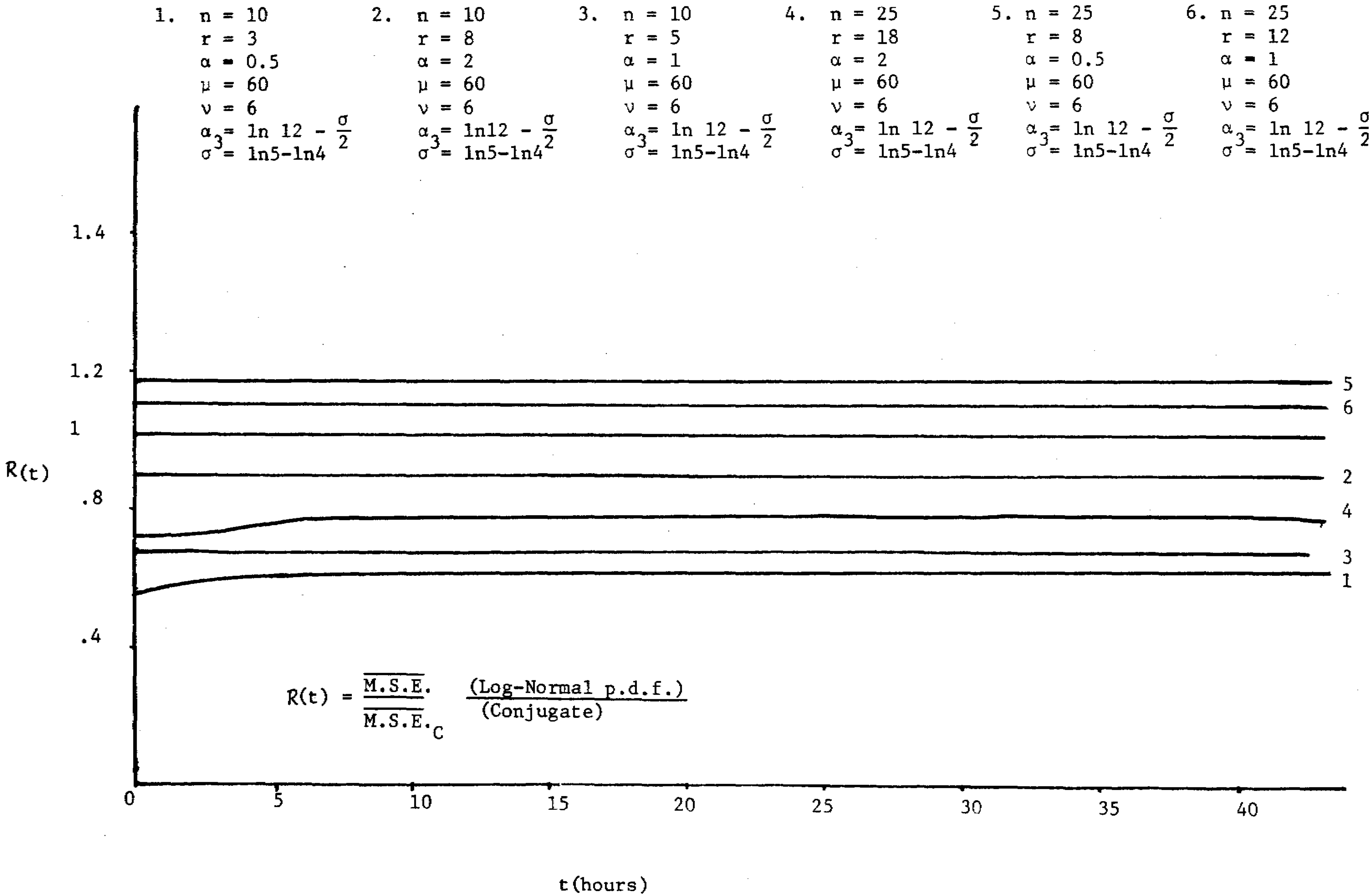

1. n = 10
r = 3
α = 0.5
μ = 60
ν = 6
α_3 = ln 12 - $\frac{\sigma}{2}$
σ = ln5-ln4
2. n = 10
r = 8
α = 2
μ = 60
ν = 6
α_3 = ln12 - $\frac{\sigma}{2}$
σ = ln5-ln4
3. n = 10
r = 5
α = 1
μ = 60
ν = 6
α_3 = ln 12 - $\frac{\sigma}{2}$
σ = ln5-ln4
4. n = 25
r = 18
α = 2
μ = 60
ν = 6
α_3 = ln 12 - $\frac{\sigma}{2}$
σ = ln5-ln4
5. n = 25
r = 8
α = 0.5
μ = 60
ν = 6
α_3 = ln 12 - $\frac{\sigma}{2}$
σ = ln5-ln4
6. n = 25
r = 12
α = 1
μ = 60
ν = 6
α_3 = ln 12 - $\frac{\sigma}{2}$
σ = ln5-ln4
R(t)
1.4
1.2
1
.8
.4
0
5
10
15
20
25
30
35
40
5
6
2
4
3
1
$R(t) = \frac{\overline{M.S.E.}}{\overline{M.S.E.}_C}$ $\frac{\text{(Log-Normal p.d.f.)}}{\text{(Conjugate)}}$
t (hours)

Figure 6.5

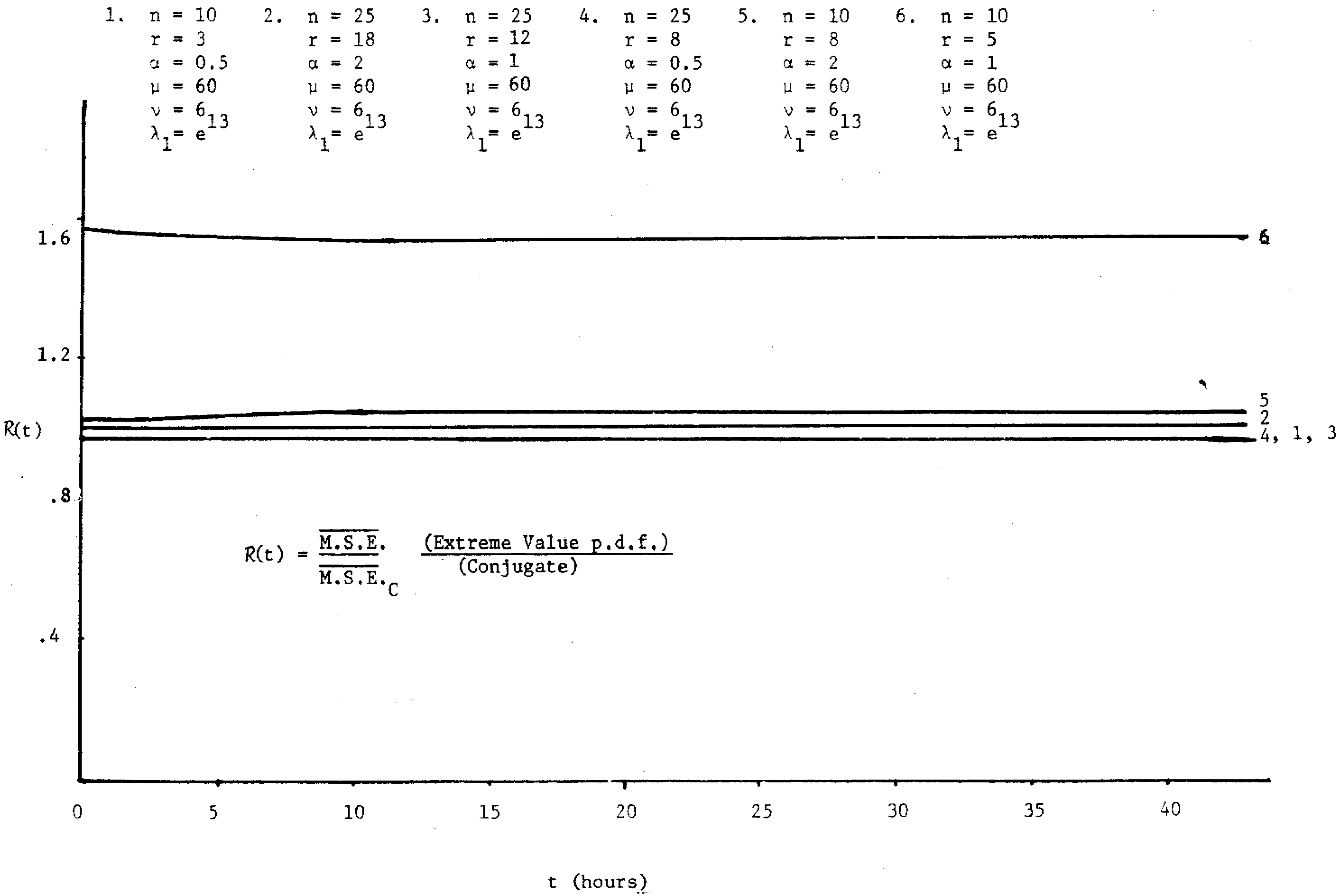

1. n = 10 r = 3 α = 0.5 μ = 60 ν = 6 λ1 = e13
2. n = 25 r = 18 α = 2 μ = 60 ν = 6 λ1 = e13
3. n = 25 r = 12 α = 1 μ = 60 ν = 6 λ1 = e13
4. n = 25 r = 8 α = 0.5 μ = 60 ν = 6 λ1 = e13
5. n = 10 r = 8 α = 2 μ = 60 ν = 6 λ1 = e13
6. n = 10 r = 5 α = 1 μ = 60 ν = 6 λ1 = e13
1.6
1.2
R(t)
.8
.4
6
5
2
4, 1, 3
R(t) = M.S.E. / M.S.E.C (Extreme Value p.d.f.) / (Conjugate)
0
5
10
15
20
25
30
35
40
t (hours)

Figure 6.6

7. SUMMARY AND CONCLUSIONS

Bayesian estimates of the scale parameter and reliability function of the Weibull failure model under the "popular" conjugate prior have been given. The employment of this prior results in analytically tractable forms of the Bayesian estimates is of interest. A procedure has been developed to investigate the consequence of the prior probability distribution being different from the assigned conjugate. The proposed method employs the ratio of the average mean square error when the prior is different from the conjugate prior (and its Bayes reliability estimate is used) to the average mean square error when the prior is assumed to be the conjugate distribution (and again the Bayes reliability estimate is used). That is, we shall assume that the level of the average mean square error of the Bayesian reliability estimate under the conjugate prior (which is assumed to be the true prior) is acceptable. The present study investigates $\overline{\text{M.S.E.}}$ under different priors and compares them to the accepted $\overline{\text{M.S.E.}}_C$.

The proposed measure of "goodness" of the Bayes reliability estimate under the conjugate prior is the ratio

$$\mathcal{R}(t) = \frac{E_{\Theta}\{E_{\underline{t}}[R(t;\alpha,\theta) - \hat{R}_C(t)]^2|\theta\}}{E_{\Theta}\{E_{\underline{t}}[R(t;\alpha,\theta) - \hat{R}_C(t)]^2|\theta\}_C} .$$

The closer this ratio is to one, the more robust is the conjugate distribution as a prior to obtain a Bayesian reliability estimate. Thus, there are three significant possibilities the proposed ratio as a function of time will assume.

(i) The ratio will be approximately equal to one, in which case we can conclude that the conjugate prior distribution is quite robust. Thus, since the conjugate prior is mathematically more attractive to work with, its implementation is recommended.

(ii) The ratio is significantly greater than one. In this case the use of the conjugate prior will result in considerable increase in the average mean square error. By inspecting the above ratio, R(t), it is clear that the numerator, which represents the average mean square error of the Bayesian reliability estimate under the influence of the conjugate prior, when in fact the true prior is different from the conjugate, is considerably larger than the denominator which represents the average mean square of the Bayesian reliability estimate under the conjugate prior when in fact the true prior is the conjugate. Thus, it is recommended that a more careful choice of the prior distribution be investigated. In fact, any non-conjugate prior which results in R(t) greater than one would be a better candidate for characterization of the stochastic behavior of the parameter.

(iii) The ratio of R(t) is significantly less than one. In this case the use of the conjugate prior is not objected because it yields an average mean square error which is less than the accepted level.

A similar approach can be employed to study the Bayesian estimate of the hazard rate under the influence of the conjugate prior versus different prior distributions.

A brief computer simulation has been given utilizing the six different prior probability density functions, namely, *beta, Poisson, inverted gamma, truncated normal, log-normal* and *extreme value*. The computer results indicate that there is a significant variation in the $\overline{\text{M.S.E.}}$ when the priors were chosen so that their first two moments approximately agreed with that of the conjugate. One would therefore conclude that it is the higher moments that cause the variation. This observation is in agreement with the analytical expressions for robustness which show that the $\overline{\text{M.S.E.}}$ depends on the higher order moments. Thus, it appears that in Bayesian reliability estimation one has to be very careful in choosing (estimating) the prior probability distribution to characterize the stochastic parameter of the given failure model.

REFERENCES

[1] Bhattacharya, S.K., Bayesian Approach to Life Testing and Reliability Estimation. Jour. Amer. Statistis. Assoc. Vol. 62, pp. 48-62, 1967.

[2] Canavos, G.C. and C.P. Tsokos, Bayesian Estimation of Life Parameters in the Weibull Distribution, Journal of Operations Research, Vol. 21, No. 3, pp. 755-763, 1973.

[3] Crelin, G.L. The Philosophy and Mathematics of Bayes Equation, IEEE Transactions of Reliability, Vol. R-21, No. 3, pp. 131-135.

[4] Drake, A.W. Bayesian Statistics for the Reliability Engineer, Proc. 1966 Annual Symposium on Reliability, pp. 315-320*.

[5] Evans, R.A. Prior Knowledge, Engineers Versus Statisticians, IEEE Transactions on Reliability, Vol. 18, Editorial.

[6] Feducia, J.A. A Bayesian/Classical Approach to Reliability Demonstration, RADC-TR-70-72, Technical Report, pp. 1-23, June 1970.

[7] Gradshteyn, I.S. and I.M. Ryzhik, Table of Integrals, Series and Products, Academic Press, New York, N.Y., 1965.

[8] Papadopoulos, A. and C.P. Tsokos, Bayesian Confidence Bounds for the Weibull Failure Model, IEEE Transactions on Reliability, to appear, 1975.

[9] Soland, R.M., Bayesian Analysis of the Weibull Process with Unknown Scale Parameters and Its Application to Acceptance Sampling, IEEE Transactions on Reliability, Vol. 17, pp. 84-90, 1968.

---

*Available from IEEE
345 East 47th St.
New York, N.Y. 10017

A BAYESIAN APPROACH TO RELIABILITY GROWTH ANALYSIS

John G. Mardo
Product Assurance Technology Division
Picatinny Arsenal
Dover, New Jersey

ABSTRACT: A Bayesian approach to assessment of reliability for systems under-going active development is presented. It is assumed, that at distinct stages of development, tests are conducted on similar items providing data for estimation of reliability and for the tracking and projection of reliability growth. While a parametric model is not assumed, it is assumed that reliability is non-decreasing from stage to stage. An approximation of the posterior conditional distribution of the nth stage reliability given that it is at least as great as the reliability in the (n-1)th stage is employed. Properties of estimates extracted from this distribution are examined. Illustrative numerical examples are provided.

INTRODUCTION: Reliability has for some time been regarded as one of the single most important characteristics by which adequacy of system design can be measured. Prior to the initiation of the development program general requirements for the system are established. One of these requirements is often the minimum reliability to be considered acceptable for the system design resulting as the end product of the development effort. An important facet of the development program is the systematic effort on the part of responsible engineers to see that these reliability goals are achieved. The level of engineering effort is determined by program managers who are responsible for the allocation of available funds, time, manpower, and facilities during development to insure that design goals are achieved. Program managers have a need for a consistent quantitative technique by which they can monitor and evaluate, through assessment of current status and forecasting of future trends, the progress of the development program toward attainment of the goals. It was this need that inspired the development of the concept of reliability growth analysis.

The phenomenon of reliability growth, as it occurs in Army materiel development programs, does not occur simply as a result of placing prototype systems on test with reliability growing in proportion to the time or number of tests occurring before a failure is observed. The ingredients of a development program which cause reliability growth are many and complex. First and foremost are the, hopefully well-designed, test programs which together become an iterative process of design, test, redesign, retest, and so on.

Following each test stage the redesigns are accomplished in part for the purpose of eliminating observed failure modes, as well as any potential failure modes or deficiencies which may have been recognized since the last design. Many reliability growth models appearing in the literature consider only this process, and as a result, fail to accurately describe reliability growth in the context of a realistic Army development program. Some of the other factors which influence end item reliability and which should be taken into account when constructing reliability growth models are advancements in the state-of-the-art allowing for the elimination of deficiencies considered inherent in the initial design, improved inspection and quality control procedures, identification of and concentration on important parameters, better methods of incorporating design changes, accelerated or overstress testing in combination with better understanding of the environmental stresses the system will experience in actual use, and many more.

That these factors and their interactions exist allows for the following important preliminary conclusions:

1. Reliability is increasing from stage to stage of testing.

2. A general representation of reliability growth by a smooth curve is unreasonable.

The first conclusion is reached because the influencing factors provide a "built-in" guarantee that the resultant system will be improved; the second, because the impact of interacting factors between stages will vary significantly.

The approach to be presented in this paper incorporates the first conclusion as an underlying assumption and the second conclusion through avoidance of parametric models [2] representing reliability as a function of development program time. The methods of estimating reliability at each stage are based on the given condition that the reliability has not decreased since the last stage. At each test stage similar items are placed on test providing binomial data in the form of successes and failures from which estimates of the current value of reliability are to be made. The unknown reliability at each stage is treated as a random variable which is the unknown parameter of the Bernoulli process generating the data in that stage. Prior and posterior distributions are formulated as beta distributions, the family of distributions conjugate to the Bernoulli process [1]. It will be apparent that conjugate priors are not requisite to application of the approach. In actual applications one could use personalized or evidential priors following essentially the same steps outlined herein. The prior distributions, whether they be conjugate or otherwise, and the likelihood functions based on the observed data in each stage are used to compute, via Bayes Theorem, the posterior conditional distribution of the nth stage reliability given that it is greater than or equal to the reliability at the (n-1)th stage. This distribution can then be used to make inferences about the nth stage reliability in the context of reliability growth. The details of the approach are provided in the next section.

PROCEDURE FOR TRACKING---INDEPENDENT BERNOULLI PROCESSES: The testing during the development program is conducted in m distinct stages. At the ith stage the results of testing are recorded as the observed number, x_i, of failures out of n_i trials. We denote by $s_i = n_i - x_i$, the number of successes observed at the ith stage. The ith stage reliability, that is, probability of success, is denoted by r_i. The factors listed previously, including elimination of failure modes observed in previous stages, lead to improvements in system reliability from stage to stage so that we may assume the r_i's are nondecreasing, or equivalently, $r_1 \leq r_2 \leq \ldots \leq r_m$. In utilizing the Bayesian approach we will consider the unknown reliabilities as independent random variables. When it becomes necessary to emphasize that the unknown ith stage reliability is a random variable in the Bayesian sense, it will be denoted by $\tilde{r}_i$. The likelihood for the ith stage event is given by

$$L(r_i; s_i, n_i) = \binom{n_i}{s_i} r_i^{s_i} (1 - r_i)^{n_i - s_i} .$$

To apply Bayes Theorem for the determination of the posterior probability density functions of the $\tilde{r}_i$ we require formulation of prior distributions for these random variables. As stated previously the conjugate beta family will provide the class of distributions from which a choice will be made. Our studies have shown that good estimation procedures in the reliability growth environment result when a uniform distribution on the unit interval is assumed as the prior distribution for each $\tilde{r}_i$. As members of the beta family the densities of these prior distributions are formulated as

$$f'_{\tilde{r}_i}(r_i) = \frac{\Gamma(a_{oi}+b_{oi})}{\Gamma(a_{oi})\Gamma(b_{oi})} r_i^{a_{oi}-1}(1 - r_i)^{b_{oi}-1} \qquad \begin{array}{l} 0 \leq r_i \leq 1 \\ i=1,\ldots,m \end{array}$$

where the parameters of the beta distributions are $a_{oi} = b_{oi} = 1$, $i=1,\ldots,m$, with the o subscript and prime used to signify that these are prior representations. To employ other members of the beta family (i.e. non-uniform) as priors all that is required is a redefinition of the a_{oi}'s and b_{oi}'s. It is suggested, however, that the impact on the properties of resulting estimates be thoroughly understood beforehand. Studies of prior representations for the reliability growth situation are now being undertaken.

Employing Bayes Theorem following each stage of testing, we determine the unconditional posterior distribution of the $\tilde{r}_i$ to have density

$$f''_{\tilde{r}_i}(r_i) = \frac{f'_{\tilde{r}_i}(r_i)L(r_i;s_i,n_i)}{\int_0^1 f'_{\tilde{r}_i}(r_i)L(r_i;s_i,n_i)dr_i}$$

$$= \frac{\Gamma(n_i + a_{oi} + b_{oi})}{\Gamma(s_i+a_{oi})\Gamma(n_i-s_i+b_{oi})} r_i^{s_i+a_{oi}-1}(1-r_i)^{n_i-s_i+b_{oi}-1} .$$

Letting $a_i = s_i + a_{oi}$ and $b_i = n_i - s_i + b_{oi}$, we can reformulate this density as

$$f''_{\tilde{r}_i}(r_i) = \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} r_i^{a_i-1}(1-r_i)^{b_i-1} . \qquad (1)$$

Having the probability densities for the independent random variables, $\tilde{r}_i$, we now wish to structure our procedure for assessing reliability at each stage within the context of reliability growth. We do this by taking advantage of our primary assumption, $\tilde{r}_1 \leq \tilde{r}_2 \leq \ldots \leq \tilde{r}_m$, considering at the kth stage that $\tilde{r}_1 \leq \tilde{r}_2 \leq \ldots \leq \tilde{r}_{k-1}$ and then introducing the additional assumption that $\tilde{r}_{k-1} \leq \tilde{r}_k$. The manner in which this is accomplished is by first formulating the unconditional joint posterior density of $\tilde{r}_1$ and $\tilde{r}_2$ as

$$h_{\tilde{r}_1,\tilde{r}_2}(r_1,r_2) = f''_{\tilde{r}_1}(r_1)f''_{\tilde{r}_2}(r_2) , \qquad 0 \leq r_1,r_2 \leq 1 .$$

To simplify notation we will for the remainder of this discussion drop the subscript of the density functions indicating the appropriate random variables unless ambiguity will result. Thus $h_{\tilde{r}_1,\tilde{r}_2}(r_1,r_2)$ will be simply represented as

$$h(r_1,r_2) = f''(r_1)f''(r_2)\,. \tag{2}$$

The conditional joint posterior distribution of $\tilde{r}_1$ and $\tilde{r}_2$ given $\tilde{r}_1 \leq \tilde{r}_2$ then has density

$$h(r_1,r_2|\tilde{r}_1\leq \tilde{r}_2) = \frac{h(r_1,r_2)}{\iint_S h(r_1,r_2)\,dr_1dr_2}$$

for $(r_1,r_2)\varepsilon S$, where $S=\{(r_1,r_2):\ 0\leq r_1\leq r_2\leq 1\}$.

Hence, from (2)

$$h(r_1,r_2|\tilde{r}_1\leq \tilde{r}_2) = \frac{f''(r_1)f''(r_2)}{\int_0^1\int_0^{r_2} f''(r_1)f''(r_2)\,dr_1dr_2} \quad ;\ 0\leq r_1\leq r_2\leq 1$$

$$= 0 \quad ;\ \text{elsewhere}\,.$$

To obtain the conditional marginal posterior density of $\tilde{r}_2$ given $\tilde{r}_1 \leq \tilde{r}_2$ we integrate with respect to r_1 over its range. Hence,

$$h(r_2|\tilde{r}_1\leq\tilde{r}_2) = [f''(r_2)\int_0^{r_2} f''(r_1)dr_1]/[\int_0^1\int_0^{r_2} f''(r_1)f''(r_2)dr_1dr_2] \tag{3}$$

The denominator of the right side of (3) is the posterior probability that $\tilde{r}_1 \leq \tilde{r}_2$. Substituting the unconditional marginal distributions given by (1) in (3) and cancelling the normalizing constants we have

$$h(r_2|\tilde{r}_1\leq\tilde{r}_2) = \frac{r_2^{a_2-1}(1-r_2)^{b_2-1}\int_0^{r_2} r_1^{a_1-1}(1-r_1)^{b_1-1}dr_1}{\int_0^1\int_0^{r_2} r_2^{a_2-1}(1-r_2)^{b_2-1}r_1^{a_1-1}(1-r_1)^{b_1-1}dr_1dr_2}\,. \tag{4}$$

To remain consistent with our choice of the beta family to summarize beliefs concerning an unknown reliability and to eliminate the difficulties expected in using the form of the density appearing in (4) we will approximate this density by a beta density. The beta fit will be accomplished through use of the method of moments employing only the first two moments of the distribution with density given by (4) to obtain the two parameters of the desired beta density. If we let μ_{12} and μ_{22} denote the first and second moments, respectively, of this distribution, we can compute their values from

$$\mu_{12} = \int_0^1 r_2 \; h(r_2 | \tilde{r}_1 \leq \tilde{r}_2) dr_2 \tag{5}$$

and

$$\mu_{22} = \int_0^1 r_2^2 \; h(r_2 | \tilde{r}_1 \leq \tilde{r}_2) dr_2 \; . \tag{6}$$

The required double integrations are performed using the binomial expansion for the interior integrals, in this case the incomplete beta integrals, and summations of the complete beta integrals forming a finite or convergent infinite series.

To determine the beta distribution with first moment μ_{12} and second moment μ_{22} we recall that a beta distributed random variable $\tilde{y}$ with density

$$f_\beta(y) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1} \quad ; \quad 0 \leq y \leq 1$$

has n^{th} moment given by

$$E(\tilde{y}^n) = \prod_{i=o}^{n} [(p+i-1)/(p+q+i-1)] \; .$$

Hence, two equations can be solved for p_2 and q_2, the parameters of the beta fit to $h(r_2 | \tilde{r}_1 \leq \tilde{r}_2)$ given in (4); these are

$$\mu_{12} = p_2 / (p_2 + q_2) \tag{7}$$

and

$$\mu_{22} = [\, p_2 \, (p_2 + 1) \,]/[\, (p_2 + q_2)(p_2 + q_2 + 1) \,] \, . \tag{8}$$

The solutions for p_2 and q_2 are

$$p_2 = [\mu_{12}^2(1-\mu_{12})/ \sigma_2^2] - \mu_{12} \tag{9}$$

and

$$q_2 = [\mu_{12}(1-\mu_{12})^2/\sigma_2^2] + \mu_{12} - 1 \tag{10}$$

where $\sigma_2^2 = \mu_{22} - \mu_{12}^2$ is the variance of the distribution with density given by (4). We will represent by $g_2(r_2)$ the beta density approximating $h(r_2|\tilde{r}_1 \leq \tilde{r}_2)$. Hence,

$$g_2(r_2) = \frac{\Gamma(p_2 + q_2)}{\Gamma(p_2)\Gamma(q_2)} \, r_2^{p_2-1}(1-r_2)^{q_2-1}$$

where p_2 and q_2 are given by (9) and (10), respectively. Obviously, the beta distribution summarizing our beliefs concerning $\tilde{r}_1$ has density given by $f''(r_1)$ so that we can define $g_1(r_1)$ by

$$g_1(r_1) = f''(r_1).$$

Point and Bayesian confidence interval estimates of r_1 can be obtained using g_1. Similarly, these estimates of r_2 can be obtained using g_2. The point estimates of r_1 and r_2 are the means of the distributions with densities approximated by g_1 and g_2, respectively. Hence,

$$\hat{r}_1 = \frac{a_1}{a_1 + b_1} \quad \text{and} \quad \hat{r}_2 = \frac{p_2}{p_2 + q_2} .$$

Lower $100(1 - \gamma)\%$ Bayesian confidence limits for r_1 and r_2 are obtained from solution for r_i of

$$\gamma = \int_0^{r_i} g_i(z_i)\,dz_i \qquad i = 1,2 .$$

To track reliability growth through the remaining stages we require the conditional marginal posterior distributions of the $\tilde{r}_k$, $k=3,\ldots,m$. Instead we will obtain the beta approximations to the densities of these distributions as we did for the conditional marginal posterior density of $\tilde{r}_2$. For $k = 3,\ldots, m$ we begin with the conditional joint posterior distribution of $\tilde{r}_{k-1}$ and $\tilde{r}_k$ given $\tilde{r}_1 \leq \tilde{r}_2 \leq ,\ldots, \leq \tilde{r}_{k-1}$. The density of this distribution is approximated by

$$h(r_{k-1}, r_k \mid \tilde{r}_1 \leq \tilde{r}_2 \leq \ldots \leq \tilde{r}_{k-1}) = f''(r_k)\,g_{k-1}(r_{k-1})$$

The conditional joint posterior distribution of $\tilde{r}_{k-1}$ and $\tilde{r}_k$ with the additional given condition that $\tilde{r}_{k-1} \leq \tilde{r}_k$ is then approximated by

$$h(r_{k-1}, r_k \mid \tilde{r}_1 \leq \tilde{r}_2 \leq \ldots \leq \tilde{r}_k) = \frac{f''(r_k)\,g_{k-1}(r_{k-1})}{\int_0^1 \int_0^{r_k} f''(r_k)\,g_{k-1}(r_{k-1})\,dr_{k-1}\,dr_k} .$$

Therefore, the conditional marginal posterior density of $\tilde{r}_k$ given $\tilde{r}_1 \leq \tilde{r}_2 \leq, \ldots, \leq \tilde{r}_k$ is

$$h(r_k|\tilde{r}_1 \leq \ldots \leq \tilde{r}_k) = \frac{f''(r_k)\int_0^{r_k} g_{k-1}(r_{k-1})\,dr_{k-1}}{\int_0^1\int_0^{r_k} f''(r_k)\, g_{k-1}(r_{k-1})\,dr_{k-1}dr_k} \quad . \tag{11}$$

To determine $g_k(r_k)$, the beta fit to (11), we again employ the method of moments. The first and second moments of the distribution with approximate density given by (11) are, respectively,

$$\mu_{1k} = \int_0^1 r_k h(r_k|\tilde{r}_1 \leq \ldots \leq \tilde{r}_k)\,dr_k$$

and

$$\mu_{2k} = \int_0^1 r_k^2 h(r_k|\tilde{r}_1 \leq \ldots \leq \tilde{r}_k)\,dr_k \quad .$$

Proceeding as we did for k=2, the values of μ_{1k} and μ_{2k} can be used to determine the parameters, p_k and q_k, of g_k from

$$p_k = [\,\mu_{1k}^2(1-\mu_{1k})/\sigma_k^2] - \mu_{1k} \tag{12}$$

and

$$q_k = [\,\mu_{1k}(1-\mu_{1k})^2/\sigma_k^2] + \mu_{1k} - 1 \quad . \tag{13}$$

So that the conditional marginal posterior density of $\tilde{r}_k$ is approximated by the beta density

$$g_k(r_k) = \frac{\Gamma(p_k + q_k)}{\Gamma(p_k)\Gamma(q_k)} r_k^{p_k-1}(1-r_k)^{q_k-1}$$

and the estimate of r_k is the mean of g_k; that is,

$$\hat{r}_k = p_k / (p_k + q_k) . \tag{14}$$

The lower $100(1-\gamma)\%$ Bayesian confidence limit of r_k is r determined from

$$\gamma = \int_0^r g_k(r_k)\,dr_k . \tag{15}$$

We can now employ (14) and (15) to track reliability growth through each stage of the development program testing.

<u>EXAMPLE OF APPLICATION</u>: The results in TABLE 1 represent the data recorded for seven stages of testing during a hypothetical development program.

<u>TABLE 1</u>

STAGE, i	1	2	3	4	5	6	7
SUCCESSES, s_i	6	5	7	6	8	9	8
FAILURES, x_i	4	5	3	4	2	1	2
TESTS, n_i	10	10	10	10	10	10	10

Suppose we are confident that the factors we listed earlier which influence reliability growth are in effect so that we, without question, assume that test items at each stage are improved over items tested in previous stages and, as a consequence, reliability is nondecreasing from stage-to-stage. We assume that the items are homogeneous within each stage.

Before proceeding with our analysis of this data, we must emphasize here that at all times our assumption of non-decreasing reliability should rest, not only on the subjective judgment that the influencing factors are obtaining, but also on careful consideration of the data. If, for example, the data up to the kth stage yields a very small posterior probability that $\tilde{r}_k \geq \tilde{r}_{k-1}$ we would seriously question our assumption and treat

the data in a manner different from that provided in the preceding section. It is important that this is kept in mind when using this approach.

The data in Table 1 was used to track reliability through the stages employing the approach outlined above with $a_{ok}=b_{ok}=1$

for k=1,...,7. In addition the maximum likelihood estimation procedure of Barlow and Scheuer [3] was employed for purposes of comparison. Barlow and Scheuer showed that the maximum likelihood estimates of the kth stage reliability, r_k, given $r_1 \leq \dots \leq r_k$ are given by

$$\hat{r}_k = 1 - \min_{j \leq k} \left(\sum_{i=j}^{k} x_i \Big/ \sum_{i=j}^{k} n_i \right).$$

They also suggested a method for computing conservative lower confidence limits for reliability at each stage. This method involved consideration of the data as though homogeneity existed between all stages as if no reliability growth were taking place. The usual technique for obtaining a one-sided lower confidence limit for a binomial parameter is then used with S_k successes in N_k trials where

$$S_k = \sum_{i=1}^{k} s_i \qquad \text{and} \qquad N_k = \sum_{i=1}^{k} n_i \, .$$

Table 2 provides the results of application of the two approaches which include point estimates of the reliability in each stage for both approaches, lower Bayesian 90% confidence limits, and Barlow and Scheuer's conservative 90% confidence bounds.

TABLE 2

STAGE	BAYESIAN ESTIMATE	M.L.E.	LOWER 90%CONF. LIM. BAYESIAN	CONSERVATIVE
1	0.583	0.600	0.401	0.354
2	0.607	0.550	0.463	0.385
3	0.728	0.700	0.596	0.467
4	0.729	0.650	0.613	0.486
5	0.819	0.800	0.713	0.540
6	0.891	0.900	0.808	0.594
7	0.884	0.850	0.808	0.619

The results included in TABLE 2 reflect immediately the relatively close agreement between the point estimates determined using the two approaches. They also demonstrate the extreme conservatism of the confidence bound suggested by Barlow and Scheuer and the improvement in this respect achieved by the method introduced here. These features are more readily apparent when reviewed graphically, as in FIGURE 1, where the tracking of reliability growth for this example is provided.

The improvement in the confidence bound obtained via the Bayesian approach can be viewed from a different aspect and said to result from an increase in the pseudo-sample size for each stage. This value for the kth stage is denoted by $\hat{N}_k$ and is simply the sum of the parameters of the beta approximation determined for $h(r_k | \tilde{r}_1 \leq \tilde{r}_2 \leq \ldots \leq \tilde{r}_k)$ reduced by the sum of the prior parameters for that stage, i.e. $a_{ok}+b_{ok}$. Hence,

$$\hat{N}_k = p_k + q_k - a_{ok} - b_{ok} .$$

If we define the number of pseudo-successes to be $\hat{S}_k = p_k - a_{ok}$ and the number of pseudo-failures to be $\hat{F}_k = q_k - b_{ok}$, then $\hat{N}_k = \hat{S}_k - \hat{F}_k$

The values of $\hat{N}_k$, $\hat{S}_k$, $\hat{F}_k$ for our example are provided in TABLE 3

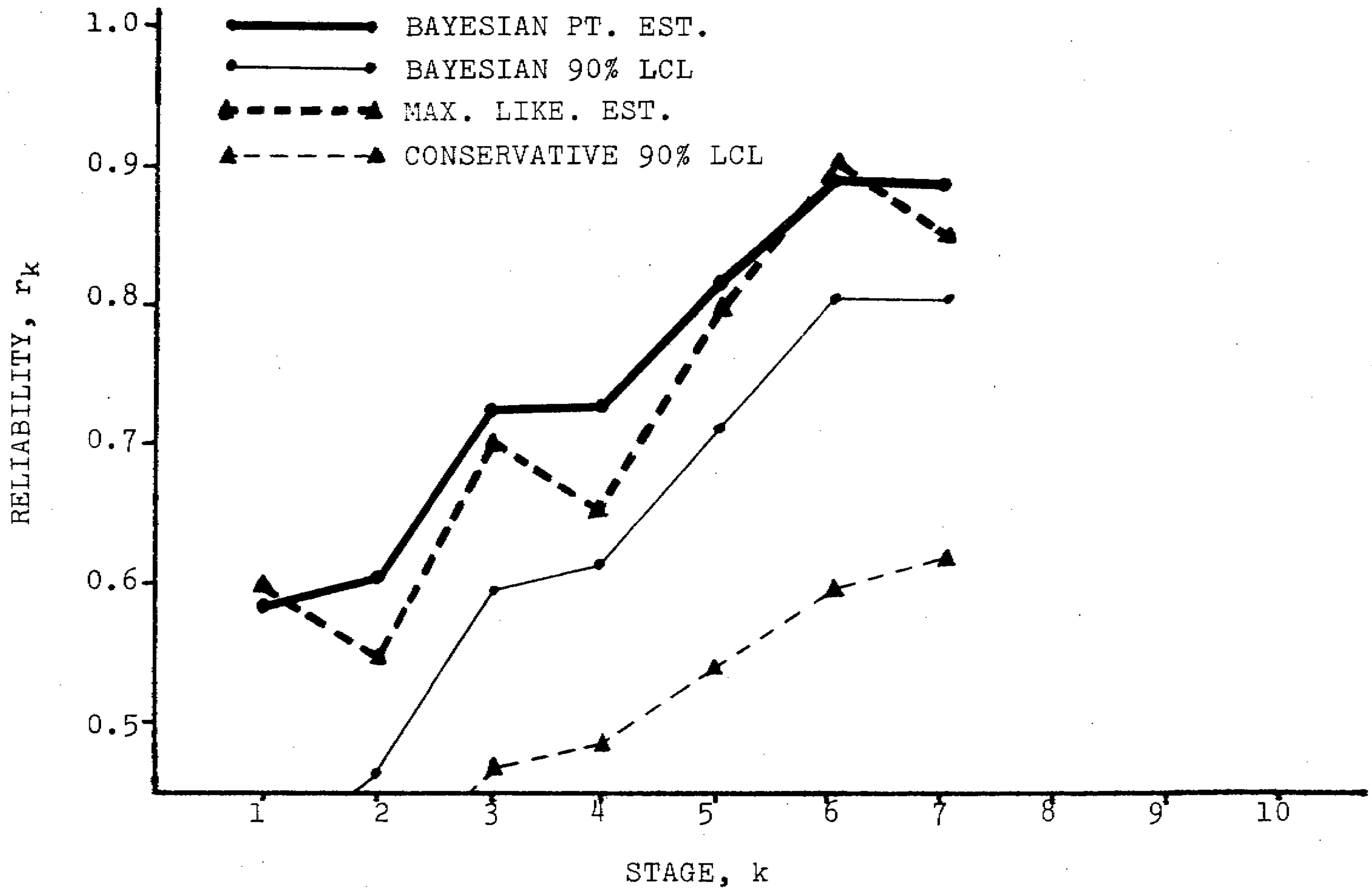

1.0
0.9
0.8
0.7
0.6
0.5
RELIABILITY, r_k
BAYESIAN PT. EST.
BAYESIAN 90% LCL
MAX. LIKE. EST.
CONSERVATIVE 90% LCL
1
2
3
4
5
6
7
8
9
10
STAGE, k

FIGURE 1

TABLE 3

STAGE, k	$\hat{S}_k = p_k - 1$	$\hat{F}_k = q_k - 1$	$\hat{N}_k = \hat{S}_k + \hat{F}_k$
1	6.0	4.0	10.0
2	10.6	6.5	17.1
3	13.2	4.3	17.5
4	17.3	5.8	23.1
5	18.0	3.2	21.2
6	20.9	1.7	22.6
7	27.1	2.7	29.8

The results in this table show the degree to which the pseudo-sample size was increased in each stage as compared to the actual sample size of 10 tested at each stage. This increase in pseudo-sample size and the use of $\hat{S}_k$, rather than the cumulative number of actual successes used by Barlow and Scheuer lead to the improved confidence limits in TABLE 2.

In addition to carrying out the necessary computation for many examples such as the one just presented, computer simulation and direct computations were made to determine the properties of the point estimates resulting from the suggested Bayesian approach. These investigations involved the effects of varying sample sizes within stages and from stage to stage. A broad spectrum of true reliability growth profiles were employed for which a determination was made of the bias and mean-square-error (MSE) of estimates of reliability at each stage. These values were determined for each of the profile/sample size cases for both the Bayesian approach and the maximum likelihood method of Barlow and Scheuer. In every case at each stage the MSE of the estimates resulting from the Bayesian approach were smaller than that for the MLE. In most cases, while the bias of the Bayesian estimates was larger than that of the MLE in the early stages, as the number of stages increased the difference decreased. In some of these cases the bias of the Bayesian estimates became smaller than that of the MLE in the later stages.

PROCEDURE FOR PREDICTION: At any intermediate stage of testing during a development program, the program managers would like to have a projection of the reliability through future stages of testing assuming the program will proceed without many unforeseen problems. In this respect, a reliability growth analysis approach which provides only a method of tracking the past and current status of system reliability, is incomplete. A means of predicting or forecasting reliability through later stages of testing based on the past experience is necessary.

Fortunately there is a natural and intuitively pleasing way to extend the Bayesian approach of tracking reliability presented above to provide a procedure for predicting reliability growth through future stages. Suppose that testing through stage k has been accomplished and the resulting data has been used along with data from previous stages to track reliability through stage k by the Bayesian approach. Prediction for future stages is accomplished by continuing the analysis conducted for stages 1 through k into stages k+1,k+2,...,m, by defining the prior beta parameters for these stages to be

$$a_{oi} = p_k \quad \text{and} \quad b_{oi} = q_k$$

for i=k+1,k+2,...,m, where p_k and q_k are the computed parameters of the beta approximation to $h(r_k|\tilde{r}_1 \leq \tilde{r}_2 \leq \ldots \leq \tilde{r}_k)$. Then for each stage i, i=k+1,...,m, since no actual data has been observed, we define s_i=o and x_i=o, so that the parameters of the unconditional marginal beta distribution of the $\tilde{r}_i$, i=k+1,...,m are

$$a_i = s_i + p_k = p_k \quad \text{and} \quad b_i = x_i + q_k = q_k$$

The conditional marginal distribution of the $\tilde{r}_i$ given $\tilde{r}_1 \leq \tilde{r}_2 \leq \ldots \leq \tilde{r}_i$ can then be approximated by a beta distribution by continuing the procedure used for stages 1 through k. Point and interval estimates of the predicted reliability for the future stages can then be obtained in the same way as for earlier stages.

This procedure was applied to the example presented in the previous section to predict reliability growth for stages 8, 9, and 10, since for that example k=7. The results are presented in TABLE 4 and include point estimates and lower 90% Bayesian confidence limits for stage 7, computed previously, as well as, for stages 8, 9, 10.

TABLE 4

STAGE	POINT ESTIMATE	LOWER 90% CONF. LIMIT
7	0.884	0.808
8	0.915	0.864
9	0.936	0.901
10	0.950	0.926

A graphical representation of the tracking and prediction of reliability growth for stages 1 through 10 of our example is provided in FIGURE 2.

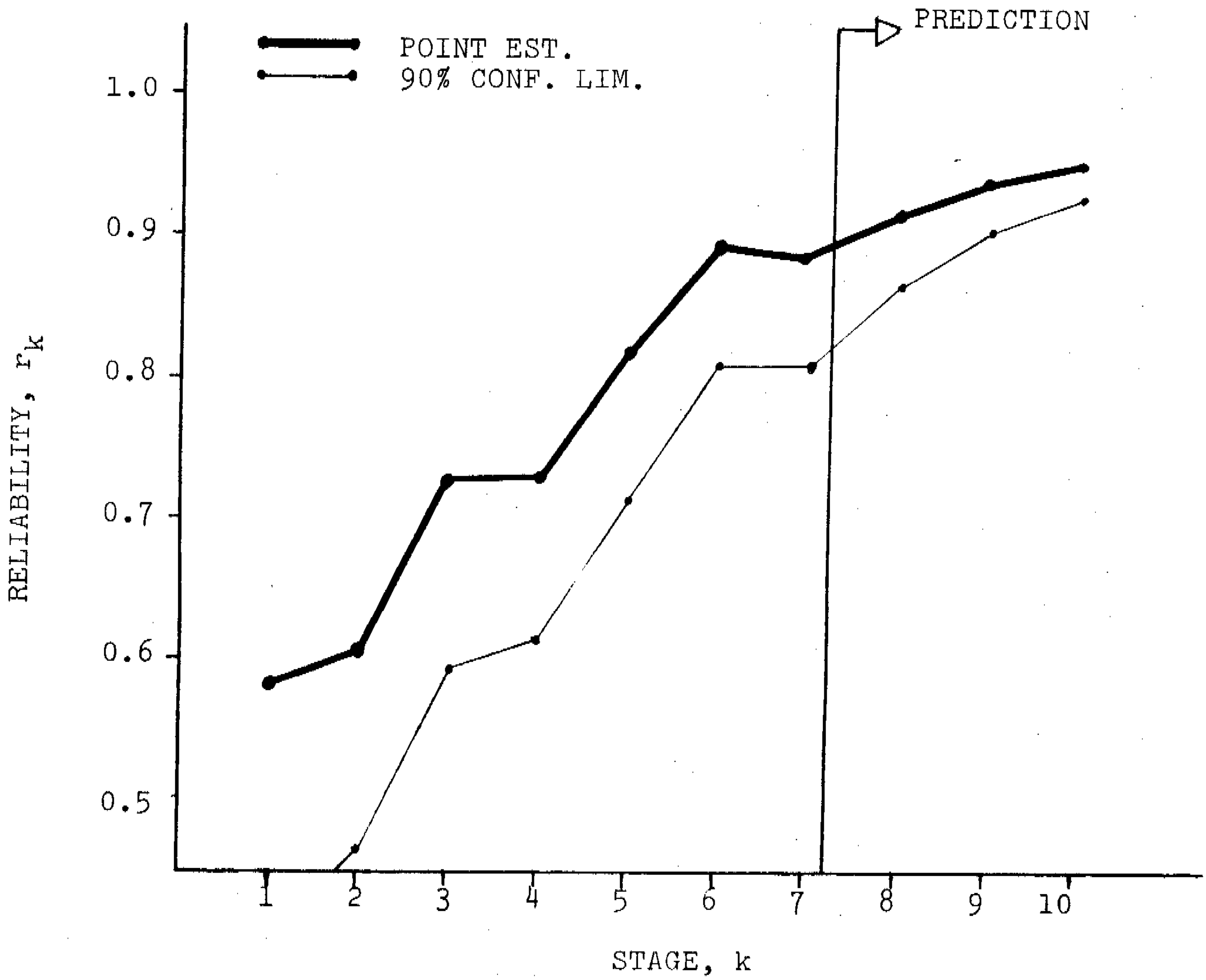


FIGURE 2

RELIABILITY GROWTH-INDEPENDENT POISSON PROCESSES:

We conclude this paper with a discussion of an extension of the method presented in the previous section for test stages defined by independent Bernoulli processes to the case where the test stages are defined by independent Poisson processes. In such cases data, instead of being provided as a number of failures in a given number of trials, is provided in terms of a number of failures occurring in some total test time at each stage. That is, at stage k, N_k items are placed on test, the test being terminated when either (a) all N_k items have failed, (b) a predetermined number, N_{ok}, of items have failed, or (c) a predetermined test time, T_{ok}, has been reached. Regardless of the manner in which the kth stage test is terminated, the time to failure of each failed item is recorded along with the total number of failures, and this information is used to define a set of sufficient statistics, $\{n_k, T_k\}$, for the unknown kth stage failure rate, λ_k. The definition of the elements of the sufficient statistic set depend on the manner in which the test was terminated.

When termination criterion (a) is used n_k and T_k are defined as

$$n_k = N_k \qquad \text{and} \qquad T_k = \sum_{i=1}^{n_k} t_i$$

where t_i is the ith greatest time to failure of the n_k items. For termination criterion (b) we have

$$n_k = N_{ok} \qquad \text{and} \qquad T_k = \sum_{i=1}^{n_k} t_i + (N_k - N_{ok})t_{n_k}$$

with the t_i defined as for (a). For criterion (c) we have

$$n_k = N_{ok} \qquad \text{and} \qquad T_k = \sum_{i=1}^{n_k} t_i + (N_k - N_{ok})T_{ok}$$

where N_{ok} is the observed number of failures occuring before or at T_{ok} and the t_i defined as for (a).

The probability distribution of the times to failure of the items tested at stage k has density

$$f(t|\lambda_k) = \lambda_k e^{-\lambda_k t} \qquad 0 \leqq t < \infty$$

This is the density of the exponential distribution with failure rate λ_k. It is easily shown that the set $\{n_k, T_k\}$, with definition depending on the termination criterion is a set of sufficient statistics for λ_k given the sample information at each stage. The family of distribution conjugate to this Poisson process is the family of gamma distribution characterized by the density

$$f_\gamma(\lambda|n,T) = \frac{(T\lambda)^n \exp[-T\lambda]\ T}{\Gamma(n+1)} \qquad 0 \leqq \lambda < \infty$$

Let stage k prior be $f_\gamma(\lambda_k|n'_k,T'_k)$

where $\{n'_k, T'_k\}$ is interpreted as the prior sufficient statistic set, then given the observed sufficient statistic set $\{n_k, T_k\}$ the posterior distribution of $\tilde{\lambda}_k$ has density $f_\gamma(\lambda_k|n''_k,T''_k)$ with

$$n''_k = n_k + n'_k$$

$$T''_k = T_k + T'_k$$

For the remainder of this discussion we will let $n_k' = T_k' = 0$, the situation analogous to $a_{ok} = b_{ok} = 1$ for the Bernoulli process. Then the posterior distribution of $\tilde{\lambda}_k$ is considered to have density $f_\gamma(\lambda_k | n_k, T_k)$.

To develop an approach to analysis of reliability growth for this case we first note that the reliability of each item placed on test at stage k in terms of a specified mission time, t_m, is given by

$$r_k = \Pr\{ t_k \leq t_m \} = e^{-\lambda_k t_m}$$

so that if we treat the unknown failure rate as a Bayesian random variable, $\tilde{\lambda}_k$, then the reliability is also such a random variable defined by

$$\tilde{r}_k = e^{-\tilde{\lambda}_k t_m}$$

with distribution depending on the distribution of $\tilde{\lambda}_k$. Instead of determining the distribution of the $\tilde{r}_k$ directly and proceeding with our analysis of reliability growth with the assumption that $\tilde{r}_1 \leq \tilde{r}_2 \leq \ldots \leq \tilde{r}_k$, we will approximate these distributions by members of the beta family using the method of moments. This is accomplished by first determining the first and second moments of $\tilde{r}_k$, denoted μ_{1k} and μ_{2k} , by

$$\mu_{1k} = E(\tilde{r}_k) = E(e^{-\tilde{\lambda}_k t_m}) = \int_0^\infty e^{-\lambda_k t_m} f_\gamma(\lambda_k | n_k, T_k) d\lambda_k$$

and

$$\mu_{2k} = E(\tilde{r}_k^2) = E(e^{-2\tilde{\lambda}_k t_m}) = \int_0^\infty e^{-2\lambda_k t_m} f_\gamma(\lambda_k | n_k, T_k) d\lambda_k$$

Performing the required integration yields

$$\mu_{1k} = [\ T_k/(t_m + T_k)\ ]^{n_k+1} = [\ \delta_k/(1 + \delta_k)\ ]^{n_k+1}$$

and

$$\mu_{2k} = [\ T_k/(2t_m + T_k)\ ]^{n_k+1} = [\ \delta_k/(2 + \delta_k)\ ]^{n_k+1}$$

where $\delta_k = T_k/t_m$ is the ratio of total test time to mission time.

We can then employ (12) and (13) with a_k and b_k substituted for p_k and q_k, respectively, to obtain the parameters of the beta fit to the distribution of $\tilde{r}_k$. The resulting beta distributions for each stage, $f_\beta(r_k|a_k, b_k)$, can then be used for reliability growth analysis proceeding in exactly the same way as for the Bernoulli process to obtain point and interval estimates of the reliability at each stage under the non-decreasing reliability assumption. It is expected that simulation studies would demonstrate that these estimates have properties almost as good as those demonstrated for the Bernoulli process.

REFERENCES

1. Raiffa, H., Schlaifer, R., Applied Statistical Decision Theory, Harvard University Press, Boston, Mass., 1961.
2. Bauman, M. F., "A Comparative Analysis of Reliability Growth Models," US Army Logistic's Management Center, USAMC Intern Training Center, Report No. 1-73-01, May 1973.
3. Barlow, R. E., Scheuer, E. M., "Reliability Growth During a Development Testing Program," Technometrics, Vol. 8, No.1, pp. 53-60, February, 1966.
4. Mardo, J. G., Cole, P.V.Z., Seibel, G., Stephenson, A. R., "A Tri-Service Bayesian Approach to Nuclear Weapons Reliability Assessment," Picatinny Arsenal Technical Report #4359, April 1972.
5. Mardo, J. G., McCarthy, T. G., "Bayesian Estimation of Reliability for Components with Independent, Normally Distributed Characteristics," Picatinny Arsenal Technical Report #4670, December, 1974.
6. Sackrowitz, H., Strawderman, W., "On the Admissibility of the M.L.E. for Ordered Binomial Parameters," Annals of Statistics, Vol. 2, No.4, pp. 822-828, 1974.

EXPERIMENTAL COLLECTION OF STATISTICS BY COMPUTER SIMULATION: THE AUTOVON NETWORK[1]

Egon Marx
Electromagnetic Effects Laboratory
Harry Diamond Laboratories
Adelphi, Maryland 20783

ABSTRACT. A computer simulation model of the AUTOVON Network has been constructed in GPSS-V. It includes all 70 switches and 60 multiple-homed PBX's in the continental United States. This model is used to predict the behavior of the network under normal operating conditions and under stress situations, such as a sudden accumulation of high-priority calls or the elimination of a number of switches or interconnecting links. The size of the network and its importance within the communications systems make it impractical to determine the reactions of the network by actual experiment. The large degree of complication brought about by the multiple interconnections, routing instructions and five levels of call priority preclude a purely theoretical calculation. Thus, the tool of computer simulation provides the best way to obtain the answers when the network is represented in the required detail. Questions that are answered in this manner include the effects of routine traffic on the high-priority calls under stress condition, and the transient effects of the elimination of a block of switches or interswitch links. The validity of the statistics gathered for the model depends on the length of the simulation run and the adequacy of the random number generators used. General results on these points are not known for GPSS.

1. INTRODUCTION. The AUTOVON network is part of the World-Wide Military Command and Communications System/Defense Communications System (WWMCCS/DCS). The Continental US (CONUS) part of the network covers the US and Canada and permits voice communications for routine traffic and priority matters.

The precise configuration of the network is constantly being altered as a result of improvements and changes in the nature of the load. The model that was used in the simulation consists of 70 switches and 60 multiple-homed Private Branch Exchanges (PBX). For this investigation, the changes introduced in the network do not affect the validity of the results.

The model is coded in GPSS-V, and it faithfully represents the processing of long-distance calls through the AUTOVON system. Minor changes are introduced to simulate the special conditions considered for each run.

The purpose of these runs was to study transient effects on the flow of traffic, which may occur under emergency conditions. This is

[1]This effort was initially formulated as part of DNA/DCA joint PREMPT Program.

in contradistinction to the more usual simulation programs primarily directed to the evaluation of steady state conditions of normal operations.

Section 2 presents a summary of the functional aspects of the network and section 3 describes the principal features of the model. Section 4 summaries a study of the effects of routine background traffic on surges of high-priority calls; and section 5 gives some results for the transient effects of the elimination of a block of switches or a part of the link between them on priority traffic. Some of the problems associated with the statistical aspects of simulations are discussed in section 6.

The effect of the routine background traffic is not large but clearly noticeable, about 10%. The transient effects of the disruption of switches or links are negligible, and the steady-state effects depend strongly on the selection of links to be eliminated. This reflects the importance of the role of the random-number generators in a GPSS model.

2. THE AUTOVON NETWORK. There are 70 switches that form the basic grid of AUTOVON, ideally laid out in a hexagonal pattern. Each switch then has 6 nearest neighbors and 6 second nearest neighbors, except for those on the edges of the network, and each switch is connected to these 12 neighbors that form the home grid of the switch. In practice this pattern is considerably distorted and the home grid can include either more or less than 12 switches. In addition, each switch is connected to a number of other switches outside the home grid by long lines. The 60 multiple-homed PBX's are smaller switches that are connected to two (in one case, three) main switches and have a single address. When the destination of a call is established, each switch has a table that shows how to route this call to a switch or PBX in the most expeditious manner. Outside the home grid, the routing instructions provide a direct route, when available, and three triples of switches that serve as an intermediate destination, a most direct triple, a first alternate triple, and a second alternate triple. Inside the home grid, there exist a direct route and a first and second alternate triples only. The home grid for a multiple-homed PBX consists of the home grids of the switches to which it is connected.

There are five priority levels for the calls, routine calls having the lowest priority. The routing instructions for routine calls restrict them in their ability to move laterally when forward routes are not available; for priority calls, a route-control digit is used to avoid shuttling and circular patterns in lateral moves. Calls can preempt trunks used by those of lower priorities after searching for an available one to the destination or in the most direct triple. The PBX's are normally capable of generating only routine traffic, and they can go to either switch first, whereas priority calls typically originate from subscribers directly connected to a switch.

The information to process a call is transmitted forward from one switch to the next as it uses tandem switches to reach its destination. Although part of this information is the priority, priority calls get no preferential treatment when queuing for common equipment used in call processing at a switch.

A more detailed description of the AUTOVON network and its simulation will be presented in an HDL Technical Report now in preparation.

3. THE NETWORK MODEL. The code for simulation of the AUTOVON network is written in GPSS-V. It uses the random number generators in the origination of calls, the selection of destination, the delays at switches, the length of conversation, the selection of a tandem switch within a triple for priority calls, the choice of a call to be preempted, etc., modified in the appropriate way for each case.

The background traffic is generated in such a way that the network is more or less uniformly loaded, as indicated by the number of calls blocked at each switch. To get good statistics for reasonable run times, the load is chosen rather heavy, so that an appreciable fraction of the calls is blocked or preempted; this is not intended to represent the normal traffic pattern. The priority profile of the calls is also chosen in a way that it facilitates the study of the special effects that are being investigated, and does not represent an average condition.

The model uses the actual routing instructions for the network, and the numbers of interswitch and PBX-to-switch trunks are close to the actual ones. It also follows the procedures used by the switch to preempt calls.

A switch is represented by a delay, which in this case is chosen to be a function of the load. This function was determined by a simulation of an AECo switch in a separate model; it was assumed that the differences between this and the ESS and 4W5 switches, which are the two other kinds in the network, would not affect the nature of the results. It is straightforward to include a more detailed model of the switch in the simulation if this is necessary. The local traffic that might exist and the possibility of the called line being busy were not taken into account. This could be included by a suitable extension of the code. Additional assumptions are detailed in the following sections.

4. EFFECTS OF BACKGROUND LOAD ON PRIORITY TRAFFIC. Although the ability of high-priority calls to preempt lower priority ones seems to indicate that routine traffic has no effect on the priority traffic, there are at least two reasons why a high load of background traffic makes the completion of priority calls more difficult.

The priority of a call is recognized only by a switch after this information has been relayed to it by the previous switch. Thus, while queuing up to be serviced for the first time by the switch marker, routine calls have the same probability of being taken as a priority one, delaying their processing. Even hot-line calls, which are checked first by the marker in the originating switch, receive no special preference when competing with other calls to be recognized at another switch. Furthermore, there is a time-out limit for the call to be accepted by the next switch. If this time is exceeded, the previous switch will try another trunk; if this happens again, it decides that the switch is out and tries another route.

Routine traffic also affects the routing of priority calls through the idle search. If a direct line is available but all trunks are busy, a priority call will normally search for a free trunk in the most direct triple of switches; only if none is available does it go back to preempt a lower priority call in the direct route. The time delay for the search itself is negligible; but when a route from the triple is chosen the number of tandem switches increases, the call has to be processed by the additional switches, more equipment is used, and the probability of its being preempted by a call of higher priority also increases.

To determine the extent of these effects, the system was heavily loaded with routine traffic. At a set time, it is assumed that an emergency occurs and that a very large number of priority calls is generated. At the same time, most of the routine traffic is excluded from AUTOVON; in the model, the remaining routine traffic being generated is represented by the 50% of blocked and preempted calls that try again. It is also assumed that an additional 40% of the priority traffic originates in the Washington, DC area, and that an additional 30% terminates there. No priority traffic originates from the multiple-homed PBX's, but 40% terminates there. Statistics are taken primarily for the priority calls. The surge of calls is of short duration, and the network is then allowed to go to an idle condition. After resetting the random number generators, the same priority calls are generated, this time on an empty network, and the statistics are collected separately and compared with the previous ones.

Before giving some of the results, a few additional details on the simplifying assumptions for the switch behavior are in order. The delay of the call at a switch is a function of the load, which was computed for the standard AECo switch with 24 registers, with the proper differences for the originator and destination switches. This value is then increased or decreased at random by up to 20% to give some expression of the arbitrary selection of a call by the marker. The switch can handle only 24 calls at a time, and new calls have to queue for the marker. It was assumed that a queue of 5 calls or less would make it try one trunk, of 10 calls or less, two trunks; and for more than 10 calls, it continues its search for another route. A limit of 56 calls queued was set for the originator switch, representing the patience of the caller to get a dial tone. These assumptions do not reflect precisely the variations in delay times, but a better solution would have demanded a rudimentary switch model to be included, with the corresponding increase in running time. All these additional waiting periods increase the delay time at the switch.

In each case, 5430 priority calls are generated with a uniform priority profile, and the number of calls started increases to 7146 and 7141 (with and without background of routine calls) due to retries from blocked and preempted calls. The calls that reached their destination were 2227 and 2366 respectively, an increase of approximately 6% when there was no interference from routine traffic. The numbers of completed calls were 3573 and 2085 respectively, but the first number includes routine calls; in the second case, 38% of the calls were eventually completed. It should be noted that this percentage does not correspond to a realistic scenario, but to an overloaded network. The average number of tandem switches used by priority calls that were completed is 1.561 and 1.474 respectively, showing a decrease of 6%; the maximum number of tandem switches was 6 in both cases. The average time for a priority call to reach its destination was 120 and 111 seconds, showing a decrease of 8%. An indication of the overloading of the network is the average number of calls active or queued for a switch when a new call arrives, which is 25.3 and 24.7 calls respectively, with a maximum of 81 as should be expected from the patience assigned to the caller.

The number of blocked priority calls shows a decrease of 10%, from 2930 to 2650, while the corresponding number of preempted calls shows an increase of 9% from 856 to 928. This last result, which at first sight seems to go against the general trend of improvement with the absence of background traffic, probably results from the lack of preemptable routine calls and the increase in the number of established priority calls. The largest effect appears in the number of highest priority calls that were blocked, with an 11% decrease from 681 to 613. For the next lower priority, the decrease is a 7% from 753 to 697, accompanied by a 10% increase in preempted calls, from 137 to 151.

The conclusion that can be drawn from the above results is that the effects of routine background traffic are small but not negligible.

5. <u>SWITCH OR LINK ELIMINATION</u>. To test the transient effects of the elimination of a block of 12 switches in the center of the network, three sets of statistics are taken. The first set is for normal conditions, the second one is right after the switches have been eliminated, and the third one is taken after the network has reached a steady state without the switches. The network is heavily loaded with routine traffic and priority traffic of the three lower priorities, and superimposed is coast-to-coast traffic of the highest priority used to test more vividly the effects of the switch elimination. The statistics are gathered for a simulated time of 6 minutes each time.

The numbers of special calls (highest priority, coast to coast) for the three intervals are 580, 571, and 635 respectively, and the numbers of calls that were blocked during these periods were 23, 177 and 172. It is difficult in this case to establish a clear relationship between the two sets of numbers, since the former set does not include calls that try again and there is a variable time delay between the initiation of the call and its reaching the destination switch or PBX. The average time to connect the special calls was 13.1, 13.3, and 13.3 seconds; the average numbers of tandem switches used by these calls were 1.425, 1.415, and 1.441, with maximum numbers of 4, 5, and 5 respectively.

There is a considerable increase in the number of blocked calls (again it should be emphasized that the load is not realistic but chosen large enough to bring out these effects), but there is no sizable effect on those calls that get through either in the period immediately following the perturbation or later on. No appreciable queuing occurs at the switches, where the maximum number of active calls being processed goes from 12 to 15 to 18 in the three periods, compared with a maximum capacity for 24 simultaneous calls of the standard AECo switch.

The number of "preempted" calls, all of lower priority, goes from 720 to 1272 to 570; the second number includes those calls that were disconnected due to the failures.

A variation of the same problem was then tried, where the links between the same 12 switches and other switches were failed with a probability of .3, as chosen by a random number generator.

The special calls were 580, 537, and 561; those blocked were 23, 62, and 108; the average times to connection were 13.1, 13.3, and 13.3 seconds; the average number of tandem switches 1.425, 1.467, and 1.494, with a maximum of 4, 5, and 6 switches respectively.

There is a significant reduction in the number of blocked calls when the switches remain in service, but otherwise the behavior of the network is not too different from the former case or the normal network. The large change in the number of blocked calls from the transient state to the steady state of the damaged network is due to a change in the sequence of random numbers to determine the elimination of links. In the run to determine the transient effects, 71 links were eliminated, while in the one for steady-state conditions 72 were eliminated, but they were not the same ones.

The numbers of "preempted" calls were 720, 1014, and 629 respectively, and the maximum active loads on the switches were 12, 13, and 12 calls.

6. STATISTICAL CONSIDERATIONS. In conducting computer "experiments" such as these, it is important to keep in mind the requirements and limitations of the simulation language used.

There always exists a problem about the length of the simulation run required to gather the information sought. It is desirable to ascertain that the results are reasonably stable, which would suggest repeating the runs with different time spans and random number generators. But on a big model such as that on the AUTOVON network, which uses between 600 and 800K of core and about one-half the simulated time in CPU time on an IBM 360-95, there are strong reasons to limit the length of runs to a minimum. (The CPU time can vary significantly with the simulated load of the system, the type of"experiment", and even the fortune of running in a faster portion of core.) Experience then dictates the length of simulation runs, together with the characteristic times in the model such as the delays at the switches, the length of conversations, and the intercall arrival time.

Questions also arise about the influence of a particular sequence of numbers from a random number generator, which can be considerable as shown by the results from the elimination of links discussed in section 5. This problem can be associated with the limitations of a random number generator or with the actual uncertainties of the system itself when chance plays an important role.

Another problem arises when statistics gathered under different conditions are compared. It appears wasteful to extend the length of the runs until the averages are well established, but the manipulation of the random number generators to reproduce the same sequences requires a considerable amount of programming with the associated drawbacks.

At the present level of understanding of these characteristics of a simulation language, these decisions have to be left to the judgement of the developer of the model both in terms of the program characteristics and the interpretation of the results.

7. SUMMARY AND CONCLUSIONS. A detailed GPSS-V model of the CONUS AUTOVON network was developed, and it was used to determine some load effects and transient effects under abnormal conditions.

It was found that routine background traffic has a small but significant obstructing role on a surge of priority traffic. The most critical phase of processing the call occurs after a trunk to another switch is taken and the call waits for a response from this switch. When it is overloaded, a call might not be recognized soon enough to avoid a time-out which delays the calls or results in its being blocked when trunks are really still available. Other calls are significantly delayed or abandoned because they get no dial tone. Another cause of delays is the idle search for trunks in the most direct triple when calls of lower priority are using the direct trunks. The deterioration of the service to priority calls due to the presence of background traffic is of the order of 10%.

The elimination of a block of 12 switches in the center of the network reduces markedly the probability of a coast-to-coast high priority call being completed during heavy traffic, but it does not increase by a large amount the time it needs for processing or the number of tandem switches it uses. The transient effects due to the traffic that is disrupted and tries again is not a very significant factor in the processing of the high priority calls. Similar results are true when about 30% of the links between the switches are eliminated without affecting the performance of the switches themselves; in this case, an important factor is apparently which of the links are eliminated.

These are just a few special conditions that can be simulated with the computer model, where practical considerations preclude actual testing of the network under abnormal circumstances. Furthermore, the model can be refined to include local traffic and a switch model, either the same one for all switches or a different one in each case.

AN ANALYSIS OF BUFFERS IN A PRODUCTION SYSTEM

ANTON HAUSCHILD
MANUFACTURING TECHNOLOGY DIRECTORATE
FRANKFORD ARSENAL
PHILADELPHIA, PENNSYLVANIA 19137

ABSTRACT: BUFFERS CAN BE USED TO INCREASE THE AVERAGE THROUGHPUT OF A PRODUCTION SYSTEM. THE USE OF BUFFERS HAS THE EFFECT OF SMOOTHING THE PROCESS FLOW THROUGH THE SYSTEM BY PROVIDING FOR IN-PROCESS INVENTORIES. IF A BUFFER EMPTIES, IT CAUSES THE SHUT-DOWN OF OTHER MACHINES IN THE PRODUCTION LINE. THESE SHUT-DOWNS CAN BE MINIMIZED WITH PROPERLY SIZED BUFFERS.

INITIALLY THE USE OF BUFFERS WAS INVESTIGATED TO DETERMINE THE EFFECTS THEY WILL HAVE ON A PRODUCTION SYSTEM. RULES WERE ESTABLISHED TO IMPROVE THE EFFECTIVENESS AND INCREASE THE BENEFIT DERIVED FROM PLACING BUFFERS IN A PRODUCTION LINE. THESE RULES TAKE INTO ACCOUNT THE DISTRIBUTION OF THE MTTR AND THE PRODUCTION SPEED FOR EACH MACHINE IN THE PRODUCTION LINE. THESE RULES WERE THEN APPLIED TO THE SCAMP PRODUCTION SYSTEM. A MONTE CARLO SIMULATION OF THE SCAMP SYSTEM WAS USED TO DETERMINE WHAT EFFECT VARIOUS SIZE BUFFERS AND SMALL CHANGES IN THE PRODUCTION SPEED WILL HAVE ON THE THROUGHPUT OF THE SYSTEM, AND THUS MINIMIZING THE NUMBER OF SHUT-DOWNS CAUSED BY THE EMPTYING OF A BUFFER.

1. INTRODUCTION: BUFFERS CAN BE USED TO INCREASE THE AVERAGE THROUGHPUT OF A PRODUCTION SYSTEM. THE USE OF BUFFERS HAS THE EFFECT OF SMOOTHING THE PRODUCT FLOW THROUGH THE SYSTEM BY PROVIDING FOR IN-PROCESS INVENTORIES. IF A BUFFER EMPTIES, IT CAUSES THE SHUTDOWN OF OTHER MACHINES IN THE PRODUCTION SYSTEM. THESE SHUTDOWNS CAN BE MINIMIZED WITH PROPERLY SIZED BUFFERS.

INITIALLY, THE USE OF BUFFERS WAS INVESTIGATED TO DETERMINE THE EFFECTS THEY WILL HAVE ON A PRODUCTION SYSTEM. RULES WERE ESTABLISHED TO IMPROVE THE EFFECTIVENESS AND INCREASE THE BENEFIT DERIVED FROM PLACING BUFFERS IN A PRODUCTION LINE. THESE RULES TAKE INTO ACCOUNT THE DISTRIBUTION OF THE MEANTIME TO REPAIR (MTTR) AND THE PRODUCTION SPEED FOR EACH MACHINE IN THE PRODUCTION SYSTEM, AND STATE GENERAL CONCLUSIONS AND RECOMMENDATIONS CONCERNING THE SIZE OF THE BUFFERS WHICH SHOULD BE EMPLOYED. THESE RULES WERE THEN APPLIED TO THE SMALL CALIBER AMMUNITION MODERNIZATION PROGRAM (SCAMP) PRODUCTION SYSTEM. A MONTE CARLO SIMULATION OF THE SCAMP SYSTEM WAS USED TO DETERMINE WHAT EFFECT VARIOUS SIZE BUFFERS AND SMALL CHANGES IN THE PRODUCTION

SPEED WILL HAVE ON THE AVERAGE THROUGHPUT OF THE SYSTEM, AND THUS REDUCE THE NUMBER OF SHUTDOWNS CAUSED BY THE EMPTYING OF A BUFFER.

THE SCAMP PRODUCTION SYSTEM IS DESIGNED TO PRODUCE SMALL CALIBER AMMUNITION AT A MAXIMUM RATE OF 1200 PIECES PER MINUTE. THE SYSTEM IS A SERIES OF MACHINES KNOWN AS SUBMODULES THAT PERFORM UNIQUE PROCESSES THAT ULTIMATELY YIELD A COMPLETED CARTRIDGE. THE MACHINES CONSIST OF A CASE SUBMODULE THAT PRODUCES A COMPLETE BRASS CARTRIDGE CASE; A PRIMER INSERT SUBMODULE THAT INSERTS THE PRIMER INTO THE JUST MANUFACTURED CASE; A BULLET SUBMODULE THAT PRODUCES THE COMPLETED BULLET; AND A LOAD AND ASSEMBLE SUBMODULE THAT LOADS THE CASE WITH PROPELLANT AND INSERTS THE BULLET INTO THE CASE, THUS PRODUCING A COMPLETED CARTRIDGE.

BECAUSE OF THE SERIAL NATURE OF THIS PRODUCTION SYSTEM, IT WAS FELT THAT SOME IN-PROCESS INVENTORIES BETWEEN SUBMODULES WOULD BE REQUIRED IN ORDER TO OBTAIN A SYSTEM THROUGHPUT THAT REFLECTED THE CAPABILITIES OF THE EQUIPMENT. THE ADDITION OF BUFFERS REDUCES THE SERIAL EFFECT AND SMOOTHS THE PRODUCT FLOW THROUGH THE SYSTEM. WITH BUFFERS, A SUBMODULE FAILURE WILL NOT AUTOMATICALLY SHUTDOWN THE WHOLE SYSTEM, BUT WILL ALLOW CONTINUED PRODUCTION BY THE OTHER SUBMODULES.

THE USE OF ARBITRARILY LARGE OR INFINITE BUFFERS WILL PRODUCE THE MAXIMUM SYSTEM THROUGHPUT, WHICH WILL BE THE THROUGHPUT OF THE LAST SUBMODULE IN THE SYSTEM. FAILURES IN ANY OF THE SUBMODULES WOULD NOT AFFECT SYSTEM THROUGHPUT OR CAUSE SHUTDOWN OF ANY OTHER SUBMODULE BECAUSE THE INPUT BUFFER TO THE FAILED SUBMODULE CAN STORE AN ARBITRARILY LARGE NUMBER OF ITEMS FROM THE PRECEDING SUBMODULE, THUS THE PRECEDING SUBMODULE WILL NEVER NEED TO BE SHUTDOWN DUE TO A FULL BUFFER. SIMILARLY, THE SUCCEEDING SUBMODULE WILL NOT BE SHUTDOWN SINCE AN INFINITE AMOUNT OF IN-PROCESS WORK PIECES ARE AVAILABLE AS INPUT TO THAT SUBMODULE.

OBVIOUSLY, INFINITELY LARGE BUFFERS ARE IMPRACTICAL. THIS PAPER ATTEMPTS TO ESTABLISH HOW BUFFERS CAN BE SIZED IN ORDER TO MINIMIZE THE SERIAL EFFECT OF A PRODUCTION SYSTEM AND THUS INCREASE THE AVERAGE THROUGHPUT FOR THAT SYSTEM.

2. METHODOLOGY:

INITIALLY, A STATISTICAL APPROACH WAS ATTEMPTED IN ORDER TO OBTAIN NUMERICAL PROBABILITIES AND ESTABLISH CONFIDENCE INTERVALS FOR THE EMPTYING OF A BUFFER FOR A GIVEN SIZE. THIS METHOD PROVED VERY CUMBERSOME, COMPLEX, AND MATHEMATICALLY TEDIOUS. AN ATTEMPT AT AN EXACT SOLUTION ALSO ELIMINATED THE POSSIBILITY OF INVESTIGATING VARIOUS ALTERNATIVES, THEREFORE, THIS APPROACH WAS ABANDONED.

A SECOND METHOD OF ATTACHING A COMPLEX PROBLEM IS TO MAKE AN ABSOLUTE MINIMUM NUMBER OF ASSUMPTIONS INVOLVING OPERATING POLICIES AND

TIME DISTRIBUTIONS OF TIME TO FAIL AND REPAIR. THIS APPROACH MORE ACCURATELY REPRESENTS THE SYSTEM TO BE MODELED AND PROVIDES A GREAT AMOUNT OF FLEXIBILITY TO VARIOUS TYPES OF ASSUMPTIONS THAT NEED TO BE INVESTIGATED. THE DISADVANTAGE OF THIS APPROACH IS THAT THE GENERALITY AND FLEXIBILITY INCORPORATED INTO THE ANALYSIS PREVENT A COMPLETE AND GENERAL SOLUTION FOR THE DETERMINATION OF THE OPTIMUM BUFFER SIZE.

THE ACTUAL METHOD OF ANALYSIS USED WAS A COMBINATION OF THE TWO ABOVE MENTIONED TECHNIQUES. INITIALLY, A GENERALIZED ANALYSIS OF THE EFFECT OF BUFFERS IN THE SCAMP PRODUCTION SYSTEM WAS CONDUCTED. THIS RESULTED IN THE ESTABLISHMENT OF A SET OF RULES TO IMPROVE THE EFFECTIVENESS AND INCREASE THE BENEFIT DERIVED FROM PLACING BUFFERS IN A PRODUCTION LINE. THESE RULES TAKE INTO ACCOUNT THE DISTRIBUTION OF TIME TO REPAIR, PRODUCTION SPEEDS, AND AVAILABILITY FOR EACH SUBMODULE IN THE SYSTEM, AND WERE USED AS GUIDELINES AND HELPED TO LIMIT THE NUMBER OF ALTERNATIVE SOLUTIONS.

IN ORDER TO DETERMINE NUMERICAL SOLUTIONS FOR THE BUFFER SIZES, A MONTE CARLO SIMULATION OF THE SCAMP PRODUCTION SYSTEM WAS WRITTEN. USING THE RULES FOR SIZING BUFFERS IN A PRODUCTION SYSTEM AS A GUIDE, NUMERICAL SOLUTIONS FOR SEVERAL ALTERNATIVE OPERATING MODES WERE DETERMINED.

THE RULES CONCERNING THE PLACEMENT OF BUFFERS ARE OF SUCH A GENERAL NATURE THAT THEY CAN BE APPLIED TO ANY PRODUCTION SYSTEM. HOWEVER, IN ORDER TO DETERMINE NUMERICAL VALUES, A SPECIFIC MATHEMATICAL OR SIMULATION MODEL MUST BE AVAILABLE FOR USE.

3. <u>BUFFER RULES</u>:

THE RULES OBTAINED FOR THE EFFECTIVE PLACEMENT OF BUFFERS WERE DERIVED FROM GENERAL OBSERVATION OF THE EFFECT A BUFFER WOULD HAVE ON THE THROUGHPUT OF THE PRODUCTION SYSTEM. THESE RULES ALSO INCLUDE FACTS THAT ARE TRUE FOR ANY PRODUCTION SYSTEM, AND CANNOT NECESSARILY BE CHANGED BY THE ADDITION OF BUFFERS. SOME OF THESE RULES MIGHT SEEM ELEMENTARY, BUT THEIR KNOWLEDGE IS REQUIRED TO INSURE THAT A COMPLETE AND COMPREHENSIVE REVIEW OF THE FACTORS PERTAINING TO BUFFERS IN A PRODUCTION SYSTEM IS INCLUDED.

THE RULES THAT FOLLOW ARE BASED ON THE SCAMP SYSTEM DIAGRAM SHOWN IN FIGURE 1.

A. A SYSTEM WITH NO BUFFERS WILL YIELD THE MINIMUM POSSIBLE AVERAGE THROUGHPUT FOR THE SYSTEM.

B. THE USE OF INFINITELY LARGE BUFFERS WILL YIELD THE MAXIMUM AVERAGE THROUGHPUT FOR THE SYSTEM.

C. WHEN USING BUFFERS, THE AVERAGE THROUGHPUT OF A SYSTEM CAN BE NO GREATER THAN THE AVERAGE THROUGHPUT OF THE SLOWEST SUBMODULE.

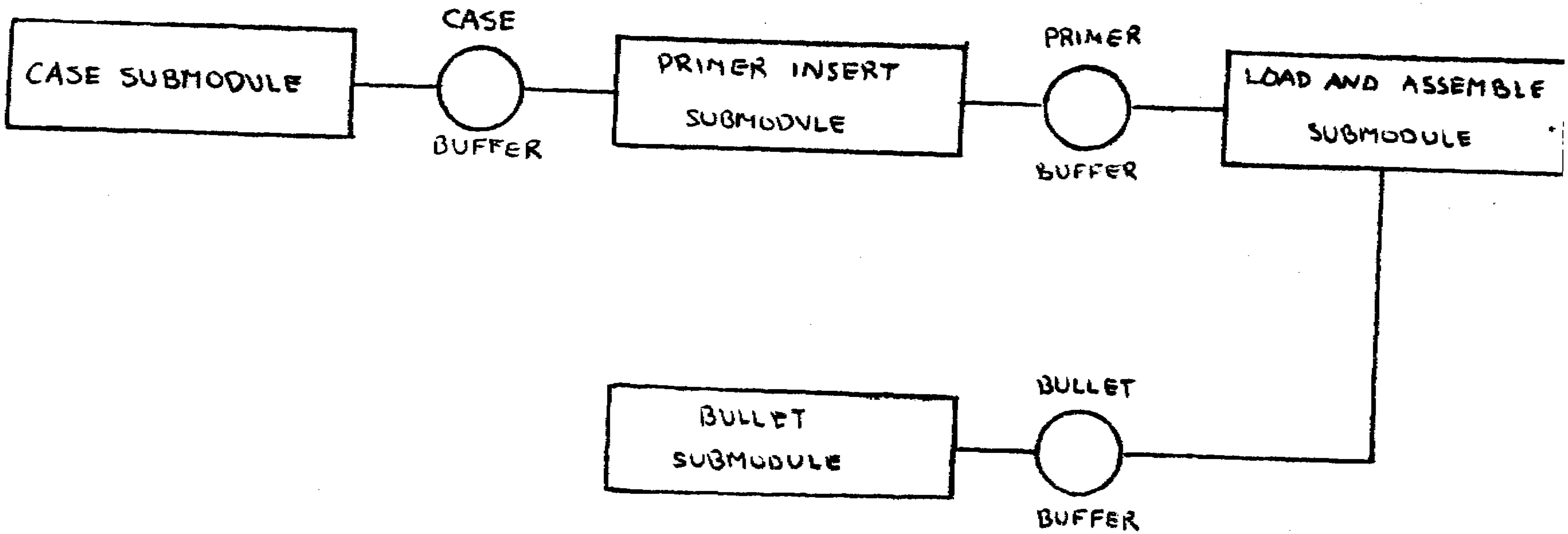


FIGURE 1: SCAMP SYSTEM DIAGRAM

D. THE ADDITION OF BUFFERS SMOOTHS THE PRODUCTION PROCESS BY REDUCING THE INTERACTION OF SUBMODULES THAT RESULT WHEN FAILURES OCCUR.

E. IF THE AVERAGE THROUGHPUT OF THE PRECEDING SUBMODULE IS LESS THAN THE AVERAGE THROUGHPUT OF THE SUCCEEDING SUBMODULE, THE BUFFER BETWEEN THE TWO WILL EVENTUALLY BECOME EMPTY.

F. THE CLOSER THE BUFFERS ARE TO THE END OF THE SYSTEM, THE MORE IMPORTANT IT IS TO INSURE THAT THEY DO NOT EMPTY.

G. IF A BUFFER FILLS, IT WILL CAUSE THE SHUTDOWN OF A MACHINE, BUT IT WILL NOT DECREASE THE AVERAGE THROUGHPUT FOR THE SYSTEM.

H. RUN THE SLOWEST SUBMODULE AT FULL SPEED AT ALL TIMES IN ORDER TO GET MAXIMUM PRODUCTION.

I. RUN THE LAST SUBMODULE IN THE SERIES TO ITS MAXIMUM CAPACITY IN ORDER TO GET THE MAXIMUM POSSIBLE THROUGHPUT.

4. SCAMP SIMULATION:

THE SIMULATION OF THE SCAMP SYSTEM USED IN THIS ANALYSIS IS A RELATIVELY SIMPLE PROGRAM USING A MONTE CARLO TECHNIQUE. IT IS VERY GENERAL IN NATURE SO THAT DIFFERENT OPERATING POLICIES AND DISTRIBUTIONS OF TIME TO FAILURE AND REPAIR CAN BE USED. THE PROGRAM GENERATESATES TIMES TO FAILURE AND REPAIR BASED ON AN EXPONENTIAL DISTRIBUTION. THE QUANTITIES IN THE BUFFERS ARE TABULATED EACH MINUTE AND TESTED TO SEE IF THEY HAVE EMPTIED OR FILLED. OPTIONS ALSO EXIST THAT WILL CHANGE THE OPERATING SPEEDS WHEN BUFFERS REACH CERTAIN VALUES.

5. ALTERNATIVES INVESTIGATED:

NOW THAT THE GROUND RULES FOR THE BUFFER ANALYSIS HAS BEEN ESTABLISHED, THE DIFFERENT OPERATING POLICIES THAT NEED TO BE ANALYZED MUST BE DETERMINED. THESE POLICIES REPRESENT FEASIBLE ALTERNATIVES THAT CAN LEAD TO DIFFERENT VALUES FOR THE AVERAGE THROUGHPUT. IT IS NECESSARY TO INVESTIGATE THESE POLICIES, BECAUSE OF THE NEWNESS OF THE SYSTEM, AND IN ORDER TO DETERMINE BUFFER CHARACTERISTICS FOR EACH OF THE POLICIES.

TWO OPERATING POLICIES ARE INVESTIGATED AND PRESENTED IN THIS PAPER. FIRST, THE PRODUCTION RATE OF THE SUBMODULES IN THE UP-MODE WERE MATCHED. THE OPERATING CONDITIONS THAT WERE USED WERE:

	PRODUCTION RATE	MTBF	MTTR	QUALITY
CASE SUBMODULE	1080	75	25	98%
PRIMER SUBMODULE	1080	40	8	100%
BULLET SUBMODULE	1080	90	30	98%
LOAD/ASSEMBLE SUBMODULE	1080	40	20	100%

THE SECOND POLICY THAT IS INVESTIGATED MATCHES THE AVERAGE THROUGHPUT OF THE SUBMODULES. THE CONDITIONS USED WERE:

	PRODUCTION RATE	MTBF	MTTR	QUALITY
CASE SUBMODULE	1080	75	25	98%
PRIMER SUBMODULE	952	40	8	100%
BULLET SUBMODULE	1080	90	30	98%
LOAD/ASSEMBLE SUBMODULE	992	40	10	100%

SUBMODULE SPEED CHANGES ARE COMPARED FOR EACH OF THE ABOVE OPERATING POLICIES. FIRST SPEEDS ARE HELD CONSTANT. NEXT, THE SPEED OF THE PRECEDING SUBMODULE IS INCREASED WHEN THE BUFFER LEVEL DROPS TO ONE QUARTER FULL AND IS DECREASED WHEN THE BUFFER LEVEL REACHES THREE QUARTERS FULL.

FINALLY, THE OPERATING SPEEDS ARE ONLY INCREASED AND THEN ONLY DECREASED WHEN THE BUFFER LEVELS REACH THE ABOVE MENTIONED LEVELS.

FOR EACH SPEED CONDITION AND OPERATING POLICIES, TWO BUFFER LEVELS WERE INVESTIGATED. THESE LEVELS WERE DETERMINED USING THE MTTR OF THE PRECEDING SUBMODULE. THE MAXIMUM QUANTITY ALLOWED IN THE BUFFER WAS THAT QUANTITY WHICH WOULD SUPPLY THE SUCCEEDING SUBMODULE WITH INPUT FOR TWICE ITS MTTR OR THREE TIMES ITS MTTR.

IN ORDER TO VALIDATE THE POLICY COMPARISONS, THEY MUST BE TESTED AGAINST THE SAME CONDITIONS. THIS IS DONE BY RE-INITIALIZING THE RANDOM NUMBER GENERATOR SO THAT THE SAME CONDITIONS WILL BE RE-CREATED FOR EACH POLICY THAT IS TESTED. THE DURATION OF ALL SIMULATION RUNS WERE FOR 9600 MINUTES OR TWENTY DAYS OF PRODUCTION AND BUFFERS WERE INITIALIZED TO THE HALF FULL LEVEL.

THESE OPERATING POLICIES REPRESENT ONLY A FEW OF THE POSSIBLE CONDITIONS THAT COULD BE INVESTIGATED, THEREFORE, EVERY EFFORT HAS BEEN MADE TO GUARANTEE THAT THE SELECTION REPRESENTS A GOOD CROSS-SECTION OF ANTICIPATED OPERATING ENVIRONMENTS.

ALTHOUGH, IT MIGHT NOT SEEM OBVIOUS, THE GENERAL ANALYSIS OF BUFFERS HELPED TO LIMIT THE NUMBER OF ALTERNATIVES THAT NEEDED TO BE INVESTIGATED. THE RULES LIMITED THE SCOPE OF THE ANALYSIS AND HELPED PREVENT THE INVESTIGATION OF UNFEASIBLE ALTERNATIVES. THE STUDY OF THE ALTERNATIVES ALSO SUCCEEDED IN PRODUCING ADDITIONAL FACTS PERTAINING TO THE EFFECTIVENESS OF BUFFERS IN A PRODUCTION SYSTEM. THESE WILL BE LISTED IN THE RESULTS SECTION OF THIS PAPER.

6. <u>RESULTS</u>:

SINCE TWO MAJOR OPERATING POLICIES WERE INVESTIGATED, THE RESULTS OF EACH SHALL BE PRESENTED SEPARATELY.

FIRST, CONSIDER THE MATCHING OF PRODUCTION RATES IN THE UP-MODE. THE CONTROLLED CONFIGURATION FOR THE FIRST POLICY CONSISTED OF AN UNBUFFERED PRODUCTION SYSTEM AND ONE WITH INFINITE CAPACITY BUFFERS. FOR THESE CONFIGURATIONS, IT WAS OBSERVED THAT:

AVERAGE THROUGHPUT WITH NO BUFFERS - 358,47/PIECES/MINUTE
AVERAGE THROUGHPUT WITH INFINITE BUFFERS - 887.40 PIECES/MINUTE

THE RESULTS OF THE SIMULATION WHEN BUFFERS WERE SIZED TWICE THE MTTR WERE:

AVERAGE THROUGHPUT WITH NO SPEED CHANGES - 760.07/PIECES/MINUTE
AVERAGE THROUGHPUT WITH INCREASE AND DECREASES IN SPEED - 768.90/PIECES/MINUTE
AVERAGE THROUGHPUT WITH SPEED INCREASES ONLY -768.60/PIECES/MINUTE
AVERAGE THROUGHPUT WITH SPEED DECREASES ONLY -737.66/PIECES/MINUTE

THE POLICY WAS REPEATED FOR BUFFER CAPACITIES THAT WERE THREE TIMES THE MTTR:

AVERAGE THROUGHPUT WITH NO SPEED - 795.61/PIECES/MINUTE
AVERAGE THROUGHPUT WITH INCREASES AND DECREASED IN SPEED - 791.24/PIECES/MINUTE
AVERAGE THROUGHPUT WITH INCREASES ONLY - 807.29/PIECES/MINUTE
AVERAGE THROUGHPUT WITH DECREASES ONLY - 783.98/PIECES/MINUTE

THE CONTROLLED CONFIGURATION FOR THE POLICY WHERE THE AVERAGE THROUGHPUTS WERE MATCHED WAS THE SAME AS THE FIRST POLICY. FOR THIS CONFIGURATION, IT WAS OBSERVED THAT:

AVERAGE THROUGHPUT WITH NO BUFFERS - 344.8 PIECES/MINUTE
AVERAGE THROUGHPUT WITH INFINITE BUFFERS - 790.29 PIECES/MINUTE

WHEN BUFFER CAPACITIES WERE SIZED TWICE THE MTTR:

AVERAGE THROUGHPUT WITH NO SPEED CHANGES - 691.14/PIECES/MINUTE
AVERAGE THROUGHPUT WITH INCREASES AND DECREASED IN SPEED - 703.07/PIECES/MINUTE
AVERAGE THROUGHPUT WITH SPEED INCREASES ONLY - 737.29/PIECES/MINUTE
AVERAGE THROUGHPUT WITH SPEED DECREASES ONLY - 697.52/PIECES/MINUTE

WHILE RESULTS OBTAINED WITH BUFFER CAPACITIES SIZED THREE TIMES THE MTTR WERE:

AVERAGE THROUGHPUT WITH NO SPEED CHANGES - 739.50/PIECES/MINUTE
AVERAGE THROUGHPUT WITH INCREASES AND DECREASES IN SPEED - 746.35/PIECES/MINUTE
AVERAGE THROUGHPUT WITH SPEED INCREASES ONLY - 753.47/PIECES/MINUTE
AVERAGE THROUGHPUT WITH SPEED DECREASES ONLY - 722.70/PIECES/MINUTE

FROM THESE RESULTS, SOME GENERAL CONCLUSIONS CAN BE DRAWN:

A. INCREASING THE OPERATING SPEED WHEN THE BUFFER LEVEL DROPS IS THE MOST EFFICIENT WAY TO INCREASE THE AVERAGE THROUGHPUT AND INCREASE THE EFFECTIVE SIZE OF THE BUFFERS.

B. DECREASING THE SPEED WHEN A BUFFER STARTS FILLING DOES NOT INCREASE THE AVERAGE THROUGHPUT.

C. MATCHING THE PRODUCTION RATES IN THE UP-MODE GIVES A HIGHER AVERAGE THROUGHPUT THAN MATCHING THE AVERAGE THROUGHPUT RATE.

D. ON THE AVERAGE, A BUFFER CAPACITY OF TWICE THE MTTR HAS AN EFFICIENCY OF 91% COMPARED TO AN INFINITELY LARGE BUFFER.

E. A BUFFER CAPACITY OF THREE TIMES THE MTTR HAS AN EFFICIENCY OF 95% COMPARED TO AN INFINITELY LARGE BUFFER.

BECAUSE THE SCAMP SYSTEM IS STILL UNDERGOING DEVELOPMENTAL TESTING, MUCH MORE WORK ON BUFFER ANALYSIS AND MANY MORE OPERATING POLICIES NEED TO BE INVESTIGATED. ALSO, AS SYSTEM OPERATING CHARACTERISTICS CHANGE AND COME TO LIGHT, NEW ANALYSES SHOULD BE MADE AND HOPEFULLY, THIS ANALYSIS IS OF SUCH A GENERAL NATURE THAT ALL NEW CONDITIONS CAN BE ANALYZED USING THIS TECHNIQUE.

STATISTICAL MODEL FOR CONTROLLER PERFORMANCE MEASURES FOR AN AIR TRAFFIC AUTOMATED SYSTEM (ATMAC)

Erwin Biser
Avionics Laboratory
U. S. Army Electronics Command
Fort Monmouth, New Jersey

ABSTRACT

The Avionics Laboratory is presently engaged in the construction of a computer queuing model to evaluate and to compare candidate system configurations for an Air Traffic Automated Center (ATMAC).

It is necessary to verify that the model generates and operates on data that are truly representative of the Air Traffic System configurations being simulated.

The ATMAC queuing model simulates Army aircraft operations, the approach control, enroute control and departure control functions in a division area, data processing and communications functions in regulating and controlling Army Air Traffic consisting of 100 to 300 aircraft of various types.

In order to validate the ATMAC queuing model, an Interactive Simulator was developed to integrate the various data processing and display equipment and to integrate the software program packages required to generate the desired realistic controller responses. The data recorded and collected from the Interactive Simulator include:

Controller time response data: elapsed time from displayed alert to acknowledge action; from presentation of flight blank display to completed flight plan entrance, etc. These are response time probability distributions associated with each of the sequence of tasks performed by an operator (controller).

The enroute control function for instance, includes a measurement of the time required for a controller to acknowledge a conflict alert. A scenario is developed to verify that the alert acknowledged time measurements are representative of realistic (actual) events. Based on a designed realistic scenario discrete action required by the Controller are determined. From the distribution of operator response time measurements, time intervals, the mean, the standard deviations and the cumulative probability distribution are computed. This process is repeated for all operator task time statistics gathered by the interactive simulator.

In addition to providing valid data (from the actions of live experienced operators) as input for the ATMAC queuing model, the interactive simulator is being used for the validation of the ATMAC queuing model: the data collected from the interactive simulator are compared with the results of other system simulations. These data include controller response time distributions associated with sequences of tasks and the system performance for those functions incorporated in the interactive simulator. The system performance data consist of: number of flights in a sector, number of conflicts, number of near misses, separation distance between any two planes, etc.

The basic statistical method employed in the analysis of the interactive simulator data is that of a two-way analysis of variance to analyze the approach control, Departure Control, and Handover data. It identifies simultaneously differences in operator performance between individual operators and between levels of automation. For each operator, the mean response time, the number of measurements, the standard error of the mean, the 90 percent confidence interval of the time mean are printed for each task and for each mode of operation.

The Scheffé method was used to compare the mean response times of the semi-automated and max-automated modes for the same task. The data of operators which consistently yielded response times significantly different from the majority was deleted from the data set. The analysis was then repeated for the remaining data. Differences in the response times resulting from pairs of automation levels were identified. The response times from different levels of automation represented the same or at least similar tasks.

The inputs to the queuing model consist, among other items, of the response time distribution for (one) representative operator for each task, function or subfunction of the ATMAC.

1. INTRODUCTION

The basic objective of the Air Traffic Management System is to facilitate safe, orderly, and efficient movement of aircraft throughout the combat area with minimum constraints on mission accomplishments.

The following are some of the typical missions flown by Army rotary wing and fixed wing aircraft.

Tactical airlift: landing, airdrop, and maximum load
Logistical airlift: troops, normal cargo, and maximum cargo
Electronic warfare
Battlefield illumination
Search and rescue
Medical rescue
Emergency resupply
Delivery of critical personnel and supplies
Radar surveillance
Infrared surveillance
Armed reconnaissance, etc.

One of the principal objectives of ATMAC is to accomplish safe movement of aircraft within the volume of responsibility. Each aircraft must arrive at its destination safely before it can carry out its mission.

Since timeliness is an important factor in mission accomplishment, the orderly movement of aircraft is essential. With an orderly movement of aircraft in the terminal area, each aircraft is able to depart the airfield close to the pilot's filed estimated time of arrival.

The ATMAC must be capable of handling aircraft operations 24-hours a day. It must handle high traffic densities with minimum traffic delays.

The combat area represents the tactical environment for combat, combat support, and combat support units that are directly engaged with the enemy.

2. INTERACTIVE SIMULATOR (COMPUTER PROGRAM) DESCRIPTION

The Interactive Simulator (I/A Sim) is a real-time system operating under the Varian VORTEX operating system. In order to provide validation data for the ATMAC Queuing Model, an air traffic environment is generated simulating the movement of aircraft and performing the functions of console input/output, track state maintenance, maneuvering, conflict prediction and resolution, and approach/departure control.

The I/A Sim operates under the Varian VORTEX operating system as two multiprogrammed tasks; 1) an event processing task consisting of Scenario Input, Track State Maintenance, Approach/Departure Control, Conflict Prediction/Resolution, Data Recording and Reduction, and 2) a Console Input/Output Task consisting of Switch Action Processing and Display Generation. Both tasks run concurrently. The functions in the event processing task are called in from rotating memory disc storage as overlays by a system scheduler program, as shown in the following diagram.

The I/A Sim has three basic functions in common with the Queuing Model: Approach/Departure Control, Track State Maintenance, and Conflict Prediction/Resolution. Approach/Departure Control monitors flights within a designated radius of a terminal checkpoint, determines patterns of approach and departure, and calculates spacing of aircraft for take-off and landing. Track State Maintenance monitors the flight enroute, updates its position and heading, and calculates the necessary parameters for maneuvering. Conflict Prediction/Resolution monitors flights with respect to other flights and restricted areas. When a conflict is detected, a display alert is generated for the console operator.

Flights and fixed points are initiated through the Scenario Input Function. A preformatted Scenario input file is read and data is stored in the appropriate location in the data base. This data is then processed by the other system functions, thereby generating flights, terminals, boundaries, and navigational aids.

The five display consoles of the I/A Sim are controlled by the Console Input/Output function. Updated displays are output to each console at periodic intervals. All input from the alphanumeric and function keyboards and joy/pressure stick is interpreted, processed, and appropriate data is stored in the data base.

The following paragraphs describe the I/A Sim program functions in more detail. The I/A Sim Program Description Document (CDRL item G001) provides a still more detailed description of the program.

<u>Console I/O</u> – This function processes inputs from the display consoles and outputs updated displays to them. This function is core resident and operates every 200 Milliseconds. It performs the following functions:

- Output updated displays to the console or consoles. Whenever displays have been updated (changed) a flag is set indicating that they are to be output to the display consoles. Displays are updated when track positions change, when ARO's have been modified, or when new categories of display data have been made up to alert the console operator of conditions requiring his attention or in response to switch actions requesting such data.
- Input from console alphanumeric keyboard. Each console is examined to determine whether an input is ready (pending) from the alphanumeric keyboard. If so, the message is input, buffered, and the

console input (switch action) processor scheduled to operate in the next cycle of the I/A Sim.

- Input from the function keyboard. Each console is examined to determine whether an input is ready (pending) from the function keyboard. If so, the input is read, buffered, and the console input (switch action) processor scheduled to operate in the next cycle of the I/A Sim.
- Input from the joy/pressure stick. If the joy/pressure stick has been activated it is read and the value input used to update the position of the hook symbol on the console display. Display update to the console from which the input was received is set to cause console I/O to output displays (including the updated position of the hook symbol) to that particular console the next time it is operated (200 milliseconds later).

System Scheduler – This function schedules the operation of the programs in the I/A Sim. This function is core resident and is designated as a root segment under VORTEX. The functional program segments of the I/A Sim are designated as overlay segments and reside on disc. They are loaded into memory and operated on a cyclic basis by the system scheduler.

Scenario Input Function – This function is read into memory and operated on a cyclic basis. Its task is to read scenario input data (simulation data) from either the simulation file on disc storage or from the simulation file on magnetic tape. This program reads such data into memory and update the I/A Sim data base (i.e. tracking tables, etc.) as required.

Track State Maintenance – This function is read into memory and operated on a cyclic basis. Its purpose is to update the X, Y, Z, velocity and heading for each active track in the I/A Sim system depending upon the particular flight's characteristics.

Location Monitor – This function is read into memory and operated on a multiple cycle basis (i.e. every other cycle, every third cycle, etc.). Its task is to monitor maneuvering flights (i.e. those following flight plans) and to determine at what point they should be required to maneuver to maintain course along a given flight path. Predicted maneuvers are communicated to the display update function for the making up of displays to alert the console operator of suggested course changes.

Track Maneuvering – This function is called when it is desired to maneuver a track in the I/A Sim system. This function determines the necessary parameters required to maneuver a flight to another heading, another altitude, another speed, or to a fixed location in the I/A Sim system. The parameters output by this function are communicated to the display track update function for formatting displays to the console operator indicating the required change to the track. Changes to tracks are communicated to the ghost pilot console operator who then implements the change by an appropriate switch action.

Conflict Prediction – This function is read into memory and operated on a multiple cyclic basis (i.e. every other cycle, every third cycle, etc.). Its function is to monitor the position of all flights within the I/A Sim system (a) in respect to all other flights and (b) in respect to system boundaries and restricted areas. When a conflict is detected, such data is communicated to the display update function for the making up of displays to alert the console operator of the condition. The conflict prediction algorithm computes suggested course changes for display to the console operator to avoid the conflict.

Approach/Departure Control – The Approach/Departure Control routine is entered when 1) an approaching flight enters within a specified radius of its terminal checkpoint, or 2) whenever a flight is scheduled to depart. A/D Control monitors the flight, determines patterns of approach and departure, calculates spacing of aircraft for takeoff and landing, and provides vectoring instructions for formation link up and break up.

Display Generation – This function is read into memory and operated on a cyclic basis. Its purpose is to convert the positional data for the tracks in the I/A Sim system into display data for output to the consoles. In addition, this routine makes up displays for a console to indicate conflicts and suggested maneuvers for flights in the system. When this function has operated the display update flag is set indicating that the Console I/O function should output updated displays to the consoles.

Switch Action Processing – This routine is read into memory and operated only upon receipt of an input from a console. Its function is to process the input for legality and, if valid, provide the proper processing. Functions performed by this routine include the following:

1. Format display for output to console (e.g. boundaries, navigational aids, corridors, etc.)
2. Input or modify flight plans
3. Change track speed, altitude, or heading (i.e. vector flight)
4. Accept flight from outside sector
5. Handover flight to outside sector
6. Handover flight to tower controller
7. Accept flight from tower controller
8. Break up or link flights
9. Change display scale or display coordinates
10. Format data for ARO display
11. Initiate or drop flights within system
12. Start or terminate simulation processing

Once switch action processing has been completed the display update flag will be set for Console I/O indicating the updated displays are to be output to the consoles.

Data Recording – This routine is read into memory and operated on a cyclic basis when data recording has been requested. Its function is to format data for output to the recording output device.

Data Reduction Program – The I/A Sim data reduction program operates off-line from the rest of the I/A Sim program. It operates upon the data collected during a simulation and computes statistics describing operator task times and parametric function effectiveness. The output data provided by this program are described in the next topic.

3. TYPES OF DATA PROVIDED BY INTERACTIVE SIMULATOR

The I/A Sim collects operator response time and algorithm effectiveness data for all major ATMAC controller tasks at up to 3 automation levels.

The data to be collected from the I/A SIM for validation of the ATMAC Model consists of controller response data and system and controller performance data. Controller data includes such items as alert response time, alert resolution time, flight plan entry/modification time, flight plan clearance time, and handover coordination time. System and controller performance data consists of elapsed time of the simulation, number of flights in a sector, number of conflicts, near misses, restricted zones active, penetrations into restricted zones, separation alerts, arrivals, departures, flight plans files, modified plans, original and modified flight plan errors, plans cleared, plans rejected, unassigned flights, and late handovers. The collected data is output after reduction and consists of cumulative probability distributions including means and standard deviations for all time-dependent categories of recorded data.

To facilitate the collection of controller response data with the I/A SIM, validation test procedures have been developed for each of the ATMAC functions implemented on the I/A SIM as follows:

- Flight Plan Entry and Modification
- Flight Plan Clearance
- Departure Control Functions
- Approach Control Functions
- Enroute Air Traffic Operation Functions

The validation test procedures are contained in the Validation Test Plan (CDRL D001) and consist of two parts. The first part is a detailed description of the actions required by the test subject controller and the supporting controllers. The second part is a series of flow diagrams which parallel the first part and include interactions between the exercise participants and the I/A SIM computer. Since most of the I/A SIM functions are being evaluated for more than one automation level, test procedures are included for each level as required.

The automation levels provided in the I/A SIM are as follows:

- Flight Plan Entry – Semi-automated
- Flight Plan Modification – Semi-automated
- Flight Plan Clearance – Semi-automated
- Departure Control Operations – Min. Auto and Max. Auto
- Approach Control Operations – Min. Auto and Max. Auto
- Handover Operations – Min. Auto and Max. Auto
- Conflict Prediction – Automated
- Conflict Resolution – Min. Auto, Semi-automated and Max. Automated
- Flight Plan Monitor – Min. Auto and Max. Automated

The types of operator tasks involved in the various modes of the I/A SIM listed above include the following:

Conflict Prediction/Resolution Data – Data collected for conflict prediction and resolution include controller response times for conflict alert acknowledge, conflict resolution determination, conflict resolution coordination, and flight vectoring. In addition, the program measures the effectiveness of the conflict resolution decisions and the effectiveness of aircraft separation performed by the controller.

Approach/Departure Control Data – Data measurements collected for approach/departure control include controller participation time and effectiveness of control for flight breakup and flight linkup. Flight breakup is an approach control task where an incoming flight of several aircraft is divided into individual aircraft tracks to permit final approach and landing. Flight linkup is a departure control task which combines aircraft which depart separately into one flight. Data measurements also are made for the approach control tasks associated with aircraft vectoring, sequencing and metering. Both controller times and control effectiveness will be recorded for these tasks.

Flight Plan Entry/Modification Data – Controller time required to enter and update flight plans are collected for this function. For this operation, the controller enters flight plan data with a keyboard entry device. The computer records entry time and also measures accuracy of data entry.

Flight Plan Clearance Data – Controller time required to perform flight plan clearance is collected for this function. Flight plan clearance involves examining the plan parameters and comparing them with other cleared flights for possible conflicts.

Handover Data – Data collected for handover situations include controller response times for handover alert acknowledgment and handover resolution. These data items will be collected for the following handover situations:

- Enroute to enroute
- Enroute to approach
- Approach to tower
- Approach to GCA
- Approach to TLS
- Tower to departure

The general form of the output data consists of horizontal cumulative probability distributions with units on the top line with the cumulative percents below, as shown in the table on the facing page. Non-time-dependent data is also output horizontally; for example: Number of Flights = 54. The figure on the opposite page is an example of the output for enroute conflict resolution. In the figure, d represents a decimal digit. Detailed descriptions of the I/A SIM output formats are contained in Vol. IV of the I/A SIM Program Description Document.

Table I. EXAMPLE OF THE OUTPUT FORMATS PROVIDED BY THE I/A SIM COMPUTER PROGRAM

ENROUTE CONFLICT RESOLUTION MEASURES

AIR – AIR

ACKNOWLEDGE ALERT MEAN = 7.5 STANDARD DEVIATION = 2.7

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	Seconds
00	03	09	15	25	36	47	51	60	65	72	80	83	90	92	93	94	96	96	98	99	00	00	00	00	00	00	00	00	Percent

RESOLVE CONFLICT MEAN = 23.1 STANDARD DEVIATION = 5.0

5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	Seconds
00	00	25	49	65	75	84	90	91	95	98	98	98	98	00	00	00	00	Percent

NUMBER OF FLIGHTS = 54 CONFLICTS = 18 NEAR MISS = 5 ELAPSED TIME = 185 MIN

RESTRICTED ZONES

ACKNOWLEDGE ALERT MEAN = d.d STANDARD DEVIATION = d.d SEC

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	Seconds
dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	Percent

RESOLVE CONFLICT MEAN = dd.d STANDARD DEVIATION = d.d SEC.

5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	Seconds
dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	dd	Percent

NUMBER OF ZONES = dd CONFLICTS = dd PENETRATIONS = dd ELAPSED TIME = dd MIN

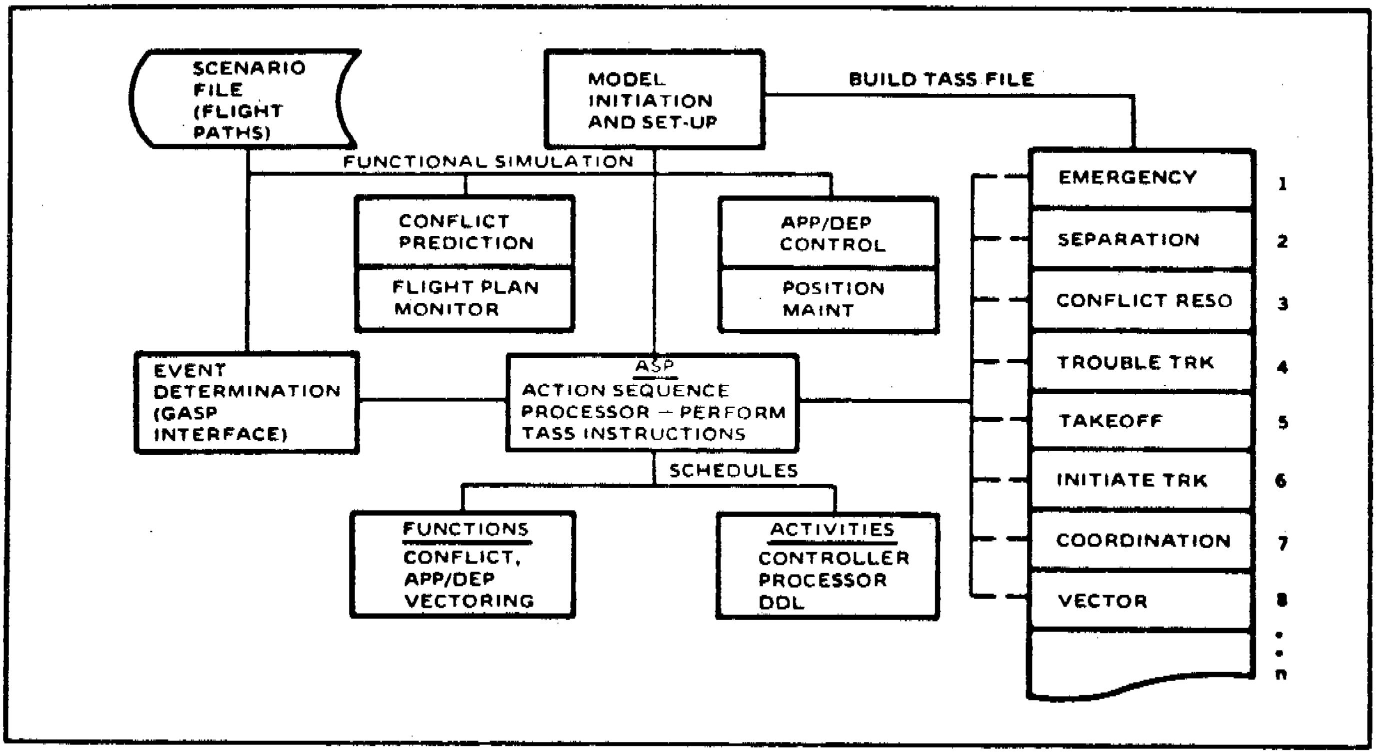
The two primary activities performed by the ATMAC Queuing Model are the functional application of air traffic control algorithms and the determination of system events in simulated time based upon these algorithms. The functional simulations respond to the status of system conditions and determine that a sequence of simulated system activity is required (e.g., a controller must vector a flight in response to a conflict alert). This need is communicated to ASP which then initiates and processes a Table of Action Sequences and Schedules (TASS), modeling the desired chain of simulated events. ASP simulates the utilization of controllers and system hardware such as the central data processor, (computer), G/G DDL and G/A/G DDL facilities. The TASS contains control and instruction parameters relating to this simulation capability as discussed above. Data concerning the utilization of system resources (controllers, data processor and DDL) are collected by ASP and are stored for later reduction by the Model Data Reduction Program.

A single TASS may be in process for more than one flight during the same simulated time period. In cases where two or more flights are assigned to the same controller and have the same TASS in process, ASP provides the queuing priority and maintains the queued tasks until the controller becomes free to handle the next task.

ASP, residing at the center of these activities, provides the primary model control mechanism. As such it receives and processes requests from external sources (functional modules) to initiate a sequence of TASS processing steps and from internal sources such as requests encountered through processing TASS parameters. The external source interface is through the GASP event chain. Consider the example of the ASP interface processing which occurs when the Conflict Prediction algorithm determines that a flight is in conflict with a restricted zone, as described in the preceding sample TASS. The Air-Air Conflict Prediction/Resolution TASS must be processed to simulate the controller's efforts in attempting to resolve the conflict in accordance with the sequence of events described in the next paragraph.

The Conflict Prediction algorithm passes the appropriate parameter data (the flight's identification and control parameters which request ASP processing) to GASP. The GASP routines store this data in the GASP event chain. When the Conflict Prediction algorithm is finished testing for conflicts, it returns control to GASP. As GASP processes the event chain, it removes the next event in time sequence from the event chain. The control parameters placed in the event chain by the Conflict Prediction algorithm are interpreted as a request for ASP processing and control is passed to ASP. ASP receives control through its external source interface using the control parameters associated with the event to initiate the interpretation of the appropriate TASS. Processing the TASS request begins immediately and the internal source interface is exercised. Some requests such as branches and ASP services are performed wholely within the ASP module. Other requests, such as simulating usage of a controller, cause ASP to place parameters into the GASP event chain which request future ASP processing.

During ASP processing of the Conflict Prediction/Resolution TASS two requests for execution of functional application algorithms are encountered. The ASP processing is the same for both even though the control parameters are different. ASP places the control parameters into the GASP event chain and continues processing the TASS. After the next time delay event is processed ASP returns control to GASP. GASP again processes the event chain removing the next event in time sequence from the chain. The control parameters placed in the event by ASP are interpreted to be a request for one of the two prediction/resolution functional applications and GASP passes control to the requested routine. After operating on the data base, the application algorithm returns control to GASP, completing the sequence of simulated activities caused by the detection of the conflict situation.



Queuing Model Block Diagram: TASS Interface Relationship

4. OUTPUT DATA PROVIDED BY QUEUING MODEL

Output data from the queuing model provides information for performing analysis of ATMAC system effectiveness by providing data on the most significant parameters which have sensitivity to scenario factors and to differences between the ATMAC configurations being modeled.

The queuing model output and reduction functions provide data for analysis of the system effectiveness measures of an ATMAC configuration. These system effectiveness measures provide the means for comparison of alternate ATMAC configurations.

During the execution of the queuing model, performance data is continuously written to disk for post-processing. This performance data consists of events that are about to occur or have just been completed, the time of day, and other information about the event.

After the execution of the queuing model simulation, offline processing and reduction of the performance data begins. The performance data is read into the Varian computer from disk and is processed by a postprocessing program. This program groups the data into many different sequences to provide the required data reduction and concludes by printing data reduction reports.

The reduction reports consist of performance measures for the preceding simulation of a specific ATMAC configuration. They consist of Arrival/Departure rates, conflict rates, conflict resolution effectiveness, flight plan monitor effectiveness, approach/departure control effectiveness, flight/aircraft densities, aircraft/flight delays and thruput, and processor, digital data link, and controller utilization. The type of information printed for each of these categories consists of "time-line" reports showing the actual time sequences in which events occurred, mean and standard deviations of event occurrence times and a tabular report of the event giving a percentage of occurrence (i.e., a distribution) of each type of category. Accumulations of the occurrence of the total time required for an event in each category are also printed. A detailed description of the queuing model output formats is given in Vol. V of the ATMAC Model Program Description Document.

By relating these reports of performance measures for each ATMAC configuration simulated it is possible to evaluate ATMAC system performance among the candidate configurations.

5. PLAN FOR ANALYSIS OF INTERACTIVE SIMULATOR RESULTS

Each of the ATMAC functions simulated on the Interactive Simulator generates data which represent controller response times, switch actions and efficiency of operation.

To insure that these performance measures are representative and can be used on the ATMAC Queueing Model, the following preparation and analyses will be performed.

- Prepare Plan of Experiments
- Statistical Analysis of Results
- Statistically Compare Results with Previous Studies

6. PLAN OF EXPERIMENTS

The Plan of Experiments defines a series of tests which are designed to gather human performance data for ATMAC functions implemented in the I/A SIM. The personnel involved in the tests will include test subjects, test conductors, ghost pilots and ghost controllers. Test subjects with sufficient experience to be classified as Air Traffic Controllers will be used. Each test subject will be made familiar with the I/A SIM and the ATMAC functions through a training program prior to collection of test data.

The experiments will present the test subject with air situations based on approved scenarios. The test subject and the support personnel (test conductor, ghost controllers) will perform the experiment using the procedures described.

7. STATISTICAL ANALYSIS OF RESULTS

The Plan of Experiments test program has been designed for eight subjects who are assumed to have essentially the same level of skill in air traffic control. Precautions shall be taken, however, to avoid the contaminating effects of different experience levels.

In order to prevent biasing the experimental response data by individual differences of skill, a training session of at least two days with appropriate testing procedures will be used.

The focus of data evaluation shall center about the purpose of the I/A Simulator and the ATMAC Model. The primary goal is to gather meaningful human response data which can be applied to the ATMAC Queuing Model.

The principal statistical methods which will be employed to analyze the data are an analysis of variance for a two-way classification together with the standard Tukey and Scheffe' methods (denoted T-method and S-method, respectively) for multiple comparisons. The analysis of variance model proposed in the following section results in the standard F-test for accepting or rejecting hypotheses concerning the mean response times for the three different levels of automation.

8. EXPERIMENTAL DESIGN*

This section describes briefly the two-way classification model for the analysis of variance to be performed. It is assumed that the response time of each operator for each class of tasks under each level of automation will be sampled N_{ij} times. The sampled response times will be denoted X_{ijk} where $i = 1, 2, \ldots n$ denotes the operator, $j = 1, 2, \ldots m$ denotes the level of automation, and $k = 1, 2, \ldots N_{ij}$ denotes the sample size.

The basic assumption for the subsequent analysis is that the random variable X_{ijk} may be represented as follows (see H. Scheffe': The Analysis of Variance, Wiley and Sons, 1959):

$$Q: \begin{cases} X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + E_{ijk}, \\ \sum_{i=1}^{n} \alpha_i = 0, \quad \sum_{j=1}^{m} \beta_j = 0, \\ \sum_{j=1}^{m} \gamma_{ij} = 0, \quad \sum_{i=1}^{n} \gamma_{ij} = 0, \\ E_{ijk} \text{ is } N(0, \sigma^2). \end{cases}$$

In the future, the foregoing set of assumptions will be referenced as hypothesis Q. The components of X_{ijk} are described in the following list:

μ = general mean,

α_i = variation in the mean due to operator i,

β_j = variation in the mean due to the level j of automation,

*The Analysis of Variance, H. Scheffé, J. Wiley & Sons, 1959.

γ_{ij} = interaction between operator i and automation level

E_{ijk} = random sample error from a $N(O,\sigma^2)$ distribution.

Note that the expected value, $E(X_{ijk})$, of X_{ijk} is, therefore,

$$E(X_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij}.$$

In order to be specific, β_1 will represent the variation in the mean due to the min-automated mode of operation, β_2 the variation due to the semi-automated mode of operation, and β_3 the variation due to the max-automated mode of operation.

There are three separate hypotheses which will be of interest in analyzing the experimental data; they are:

$$H_o:\gamma_{ij} = O \text{ for } i=1,2,\ldots n \text{ and}$$

$$j = 1,2,\ldots m,$$

$$H_1:\beta_j = O \text{ for } j=1,2,\ldots m,$$

$$H_2:\alpha_i = O \text{ for } i=1,2,\ldots n.$$

The hypothesis H_o is of no direct interest; however, acceptance or rejection of H_o determines the specific statistics used to test H_1 and H_2 and influences significantly any interpretation of the results of those tests. In particular, if the hypothesis H_o is accepted, by statistical inference or otherwise, then inferences about the α_i's and the β_j's will be sufficient to summarize the experiment. Thus, if it is concluded that $\beta_3 < \beta_1$, then the difference in the response times between levels of automation is the same for each operator. If H_o is rejected, however, then $\beta_3 < \beta_1$ implies only that operator response, when averaged over all operators, is greater in the min-auto mode than in the max-automated mode of operation.

The important hypothesis with respect to the Interactive Simulator is $\tilde{H}_1$, the alternative hypothesis for H_1. Note that rejection of H_1 automatically implies acceptance of $\tilde{H}_1$. In order to facilitate the computations, H_1 will be tested rather than testing $\tilde{H}_1$ directly. If H is rejected, then the T-method will be used to identify significant differences between the β_j's under the hypothesis H_o. If H_o is rejected (that is, interactions are present), the T-method cannot be used because the β_j's may not have equal variances.

Hypotheses H_2 is included in order to verify that the operators have, at least approximately, the same level of skill. If the hypothesis H_2 is rejected, the T-method (or the S-method if H_o is rejected) will be used to determine which operator (or operators) contributed to the rejection. The analysis then can be repeated after deletion of the relevant operator data or the experiments can be repeated by the appropriate operator (or operators) after the appropriate training.

9. COMPUTATIONAL PROCEDURE

In this section the statistics which are necessary to test the hypotheses outlined in the previous section are defined. Recall that X_{ijk} represents the K^{th} sample response time of operator i in the operating mode j.

Assume that the data set X_{ijk} is given. In order to test the hypotheses mentioned previously, it will be necessary to compute the following sample means and estimates (denoted collectively as the Means):

$$\lambda_{ij} = \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} X_{ijk},$$

$$N_t = \sum_{i=1}^{n} \sum_{j=1}^{m} N_{ij},$$

$$g_i = \sum_{j=1}^{m} \sum_{k=1}^{N_{ij}} X_{ijk} = \sum_{j=1}^{m} \lambda_{ij} N_{ij},$$

$$h_j = \sum_{i=1}^{n} \sum_{k=1}^{N_{ij}} X_{ijk} = \sum_{i=1}^{n} \lambda_{ij} N_{ij},$$

$$G_i = \sum_{j=1}^{m} N_{ij}, \quad \text{: row sums of cell numbers}$$

$$A_i = \frac{1}{m} \sum_{j=1}^{m} \lambda_{ij},$$

$$H_j = \sum_{i=1}^{n} N_{ij}, \quad \text{: column sums of cell numbers}$$

$$B_j = \frac{1}{n} \sum_{i=1}^{n} \lambda_{ij}.$$

Following Scheffe': The Analysis of Variance, the maximum likelihood estimates of β_j, denoted by $\hat{\beta}_j$, are the solutions of the following set of linear equations:

$$\sum_{l=1}^{m} b_{jl} \hat{\beta}_l = R_j, \; j=1,2,\ldots m,$$

where

$$b_{jl} = \delta_{jl} H_j - \sum_{i=1}^{m} \left(\frac{N_{ij} N_{il}}{G_i} \right),$$

$$R_j = h_j - \sum_{i=1}^{m} \left(\frac{N_{ij} g_i}{G_i} \right).$$

The sums of squares (denoted collectively by SS) which are necessary for the analysis of variance are:

$$SS_A = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{N_{ij}} X_{ijk}^2 - \sum_{j=1}^{m} \left(\frac{h_j^2}{H_j} \right),$$

$$SS_A' = \sum_{i=1}^{n} W_i A_i^2 - \left[\sum_{i=1}^{n} W_i \right]^{-1} \left[\sum_{i=1}^{n} W_i A_i \right]^2,$$

$$SS_B = \sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{k=1}^{N_{ij}} X_{ijk}^2 - \sum_{i=1}^{n} \frac{g_i^2}{G_i},$$

$$SS_B' = \sum_{j=1}^{m} U_j B_j^2 - \left[\sum_{j=1}^{m} U_j \right]^{-1} \left[\sum_{j=1}^{m} U_j B_j \right]^2,$$

$$SS_C = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{N_{ij}} X_{ijk}^2 - \sum_{j=1}^{m} R_j \hat{\beta}_j - \sum_{i=1}^{n} \frac{g_i^2}{G_i},$$

$$SS_E = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{N_{ij}} (X_{ijk} - \lambda_{ij})^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} N_{ij} \sigma_{ij}^2,$$

where the weighting factors U_j and W_i are defined:

$$W_i = m^2 \left[\sum_{j=1}^{m} N_{ij}^{-1} \right]^{-1},$$

$$U_j = n^2 \left[\sum_{i=1}^{n} N_{ij}^{-1} \right]^{-1}.$$

Table II. ANALYSIS OF VARIANCE FOR A TWO-WAY CLASSIFICATION; SAMPLE MEANS

	Operators				
	1	2	....	n	Row Sums ($\hat{\alpha}_i$)
Min-Auto. (1)	X_{11k}, N_{11}	X_{12k}, N_{12}		X_{1nk}, N_{1n}	$H_1 = \sum_{j=1}^{n} N_{1j}$ $\quad B_1 = \sum_{j=1}^{n} \sum_{k=1}^{N_{1j}} \frac{X_{1jk}}{n N_{1j}}$ $h_1 = \sum_{j=1}^{n} \sum_{k=1}^{N_{1j}} X_{1jk}$
Semi-Auto. (2)	X_{21k}, N_{21}	X_{22k}, N_{22}		X_{2nk}, N_{2n}	$H_2 = \sum_{j=1}^{n} N_{2j}$ $\quad B_2 = \sum_{j=1}^{n} \sum_{k=1}^{N_{2j}} \frac{X_{2jk}}{n N_{2j}}$ $h_2 = \sum_{j=1}^{n} \sum_{k=1}^{N_{2j}} X_{2jk}$
Max-Auto. (3)	X_{31k}, N_{31}	X_{32k}, N_{32}		X_{3nk}, N_{3n}	$H_3 = \sum_{j=1}^{n} N_{3j}$ $\quad B_3 = \sum_{j=1}^{n} \sum_{k=1}^{N_{3j}} \frac{X_{3jk}}{n N_{3j}}$ $h_3 = \sum_{j=1}^{n} \sum_{k=1}^{N_{3j}} X_{3jk}$
Modes of Operations Column Sums ($\hat{\beta}_j$)	$G_1 = \sum_{j=1}^{m} N_{1j}$ $g_1 = \sum_{j=1}^{m} \sum_{k=1}^{N_{1j}} X_{1jk}$ $A_1 = \sum_{j=1}^{m} \sum_{k=1}^{N_{1j}} \frac{X_{1jk}}{mN_{1j}}$	$G_2 = \sum_{j=1}^{m} N_{2j}$ $g_2 = \sum_{j=1}^{m} \sum_{k=1}^{N_{2j}} X_{2jk}$ $A_2 = \sum_{j=1}^{m} \sum_{k=1}^{N_{2j}} \frac{X_{2jk}}{mN_{2j}}$		$G_n = \sum_{j=1}^{m} N_{nj}$ $g_n = \sum_{j=1}^{m} \sum_{k=1}^{N_{nj}} X_{njk}$ $A_n = \sum_{j=1}^{m} \sum_{k=1}^{N_{nj}} \frac{X_{njk}}{mN_{nj}}$	$N_t = \sum_{j=1}^{3} H_j = \sum_{j=1}^{n} G_j$

Table III. ANALYSIS OF VARIANCE FOR A TWO-WAY CLASSIFICATION; SUMS OF SQUARES

Source	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (SS/df = MS)	F-STATISTIC
Operators	$SS_A = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{N_{ij}} X_{ijk}^2 - \sum_{j=1}^{m} \frac{h_j^2}{H_j}$	n-1	$\frac{SS_A}{(n-1)}$	$\frac{SS_A - SS_C}{SS_C} \frac{N_t - m - n + 1}{n-1}$
	$SS'_A = \sum_{i=1}^{n} W_i A_i^2 - \left\{\sum_{i=1}^{n} W_i\right\}^{-1} \left\{\sum_{i=1}^{n} W_i A_i\right\}^2$	n-1	$\frac{SS'_A}{(n-1)}$	$\frac{SS'_A}{SS_E} \frac{N_t - nm}{n-1}$
Levels of automation	$SS_B = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{N_{ij}} X_{ijk}^2 - \sum_{i=1}^{n} \frac{g_i^2}{G_i}$	m-1	$\frac{SS_B}{(m-1)}$	$\frac{SS_A - SS_C}{SS_C} \frac{N_t - m - n + 1}{m-1}$
	$SS'_B = \sum_{j=1}^{m} U_j B_j^2 - \left\{\sum_{j=1}^{m} U_j\right\}^{-1} \left\{\sum_{j=1}^{m} U_j B_j\right\}^2$	m-1	$\frac{SS'_B}{(m-1)}$	$\frac{SS'_B}{SS_E} \frac{N_t - nm}{m-1}$
Interactions	$SS_C = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{N_{ij}} X_{ijk}^2 - \sum_{j=1}^{m} R_j \beta_j - \sum_{i=1}^{n} \frac{g_i^2}{G_i}$	N_t-m-n+1	$\frac{SS_C}{N_t - m - n + 1}$	$\frac{SS_C - SS_E}{SS_E} \frac{N_t - nm}{(m-1)(n-1)}$
Errors	$SS_E = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{N_{ij}} (X_{ijk} - \lambda_{ij})^2$	N_t-mn	$\frac{SS_E}{N_t - mn}$	
Total	$SS_T = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{N_{ij}} (X_{ijk} - \mu)^2$	N_t-1	$\frac{SS_T}{N_t - 1}$	

Table IV. QUEUING MODEL AUTOMATION LEVELS

	Manual Level*	Min-Automated Level	Semi-Automated Level	Max-Automated Level
Facility				
Computer	No	Yes	Yes	Yes
G-G DDL	No	No	Yes	Yes
G-A DDL**	No	No	No	Yes
A-G DDL	No	Yes	Yes	Yes
Function				
Track Acquisition/Tracking	CONT	DP	DP	DP
FP Processing/Approval	CONT	DP**	DP	DP
FP Conformance	CONT	CONT	DP	DP
Approach/Departure Control	CONT	CONT	DP	DP
Formation Linkup/Breakup	CONT	CONT	DP	DP
Handover	CONT	CONT	DP	DP
Conflict Prediction/Resol.	CONT	CONT	DP	DP

*Uses radar video
**Controller approves FP

NOTE: CONT means controller performs function
DP means data processor performs function with controller review, approval and/or override.

When these computations have been completed, the hypotheses H_o, H_1, and H_2 will be tested by the procedures enumerated below:

Step 1: Test H_o against Q at the significance level σ (= 0.05). In order to perform this test, define the statistic:

$$F = \frac{N_t - mn}{(m-1)(n-1)} \cdot \frac{SS_C - SS_E}{SS_E}$$

The random variable F has an F-distribution with parameters (m-1)(n-1) and (N_t-mn). The F-test for H_o is defined as follows: Reject H_o if $F > F_o$ where F_o satisfies:

$$P(f \leq F_o) = 1-\alpha.$$

Step 2: Test H_1. If H_o is accepted, test H_1 against $\Omega \cap Q$; that is, set:

$$F = \frac{N_t - n - m + 1}{m-1} \cdot \frac{SS_B - SS_C}{SS_C}.$$

The random variable F has an F-distribution with parameters (m-1) and (N_t-m-n+1). An F-test similar to that of Step 1 may be used to accept or reject H_1.

If H_o is rejected, then H_1 should be tested against Q rather than $\Omega \cap Q$; the appropriate statistic in this case is:

$$F = \frac{N_t - mn}{m-1} \cdot \frac{SS'_B}{SS'_E}$$

which has an F distribution with parameters (m-1) and (N_t-mn).

As noted previously, rejection of H_1 is equivalent to acceptance of the hypothesis that there are significant differences in the β_j's, that is, differences other than those produced by sampling errors. The significance of accepting or rejecting H_o upon H_1 was discussed in the preceding section. If H_o is accepted, the T-method will be used to determine the significant differences in the β_j's whenever H_1 is rejected. If H_o is rejected, then the S-method will be used rather than the T-method to determine significant differences in the β_j's when H_1 is rejected.

Step 3: Test H_2. If H_o is accepted, test H_2 against $\Omega \cap Q$ with the statistic:

$$F = \frac{N_t - m - n + 1}{n-1} \cdot \frac{SS_A - SS_C}{SS_C}$$

which has an F distribution with parameters (n-1) and (N_t-m-n+1). The standard F-test will be employed (see Step 1).

If H_0 is rejected, H_2 should be tested against Q. In this case,

$$F = \frac{N_t - mn}{n-1} \cdot \frac{SS'_A}{SS_E}$$

Acceptance of H_2 is equivalent to acceptance of the hypothesis that all of the operators have an equal level of skill. Conversely, rejection of H_2 implies that the operators have unequal levels of skill; in this case, the S-method or the T-method (as above) will be used to determine which operators were responsible for rejecting H_2.

10. EXPECTED DATA RESULTS

All communications shall be recorded for detailed analysis. Operator workloads shall be recorded throughout the entire run. The operation workload for the purpose of this study is defined as the sum of all aircraft being displayed or under his control plus all communications, verbal or otherwise, that he is engaged in at any discrete period of time.

Data analysis of cumulative probability distributions, summary totals, and flights per time displayed coupled with statistical test applications will provide knowledge of differences in human operator responses which will be attributable to mode manipulations i.e., max-automated or min-automated rather than sampling error.

The results of this comprehensive data collection and analysis will provide inputs to the ATMAC Model which will, within the limitations of the I/A Simulator, be both realistic and meaningful.

The data to be gathered for each test subject and the treatment of the data for each ATMAC function to be examined on the I/A SIM are contained in the following sections.

a. Flight Plan Entry/Modification/Clearance.

The following calculations will be made.

	Means	SD	C.P.D.
Flight plan entry time	xx	xx	xx
Flight plan modification time	xx	xx	xx
Flight plan clear-reject time	xx	xx	xx

The data output i.e., number of flight plans entered, etc. is expected to be large. For this reason no unusual statistical test of significance is envisioned; however, if the sample size should diminish or be affected by widely discrepant subject responses, the Student "t" test shall be employed between subject means. Cumulative probability distributions and sample size shall determine the need for the use of the Student "t".

b. Approach Control.

The following calculations shall be made for each mode, max-automated and min-automated.

	Means	SD	C.P.D.
Accept hand-in time	xx	xx	xx
Approved vector instructions time	xx	xx	xx
Resolve separation alert time	xx	xx	xx
Closest approach distances	xx	xx	xx

Data shall be compiled for each subject. A total mean shall be calculated for each measure in each mode. The hypothesis to be tested shall be:

$$H_1: \beta_1, \beta_2 \neq O$$

The implied null hypothesis is:

$$H_1: \beta_1 = \beta_2 = 0$$

The significance level shall be set at 0.05.

The F-test shall be used to test the differences between sample means. In specific terms the research hypothesis is that the mean responses in the min-automated mode will be different than the mean responses in the max-automated mode and that such a difference is not due to sampling error.

c. Departure Control.

The following calculations shall be made for each mode, max-automated and min-automated.

	Means	SD	C.P.D.
Departure alert acknowledge time	xx	xx	xx
Approve vector instructions time	xx	xx	xx
Resolve Separation alert time	xx	xx	xx
Closest approach distances	xx	xx	xx

Data shall be compiled for each subject. A total mean shall be calculated for each measure in each mode. The hypothesis to be tested shall be:

$$\tilde{H}_1: \beta_1, \beta_2 \neq O$$

The implied null hypothesis is:

$$H_1: \beta_1 = \beta_2 = O$$

The significance level shall be set at 0.05.

The F-test shall be used to test the differences between sample means. In specific terms the research hypothesis is that the mean responses in the manual mode will be different than the mean responses in the automated mode and that such a difference is not due to sampling error.

d. Enroute Control.

The following calculations shall be made for each mode, Max-Automated, Semi-Automated, and Min-Automated for both Air-to-Air and Air-to-Restricted Zone conflicts.

	Means	SD	C.P.D.
Conflict alert acknowledge time	xx	xx	xx
Conflict resolution time	xx	xx	xx

A total mean shall be calculated in each mode. The hypothesis to be tested shall be:

$$\tilde{H}_1: \beta_1, \beta_2, \beta_3 \text{ not all zero}$$

The implied null hypothesis is:

$$H_1: \beta_1 = \beta_3 = \beta_3 = O.$$

The significance level shall be set at 0.05.

The F-test shall be used to test the differences between sample means. In specific terms the research hypothesis is that the subject mean responses will be different across each mode and that such a difference is not due to sampling error.

e. Handover.

Handovers will occur in the Approach, Departure, and Enroute functions. The event (handover/handin) will be controlled by the scenario. The following calculations shall be made for handovers in each function, for min-auto and max-automated modes.

	Means	SD	C.P.D.
Handover alert acknowledge time	xx	xx	xx
Handover coordination time	xx	xx	xx

The F-test shall be applied for testing a difference between means in each of the three functions. The hypothesis to be tested shall be:

$$\tilde{H}_1: \beta_1, \beta_2 \neq O$$

The implied null hypothesis is:

$$H_1: \beta_1 = \beta_2 = O$$

The significance level shall be set at 0.05. In specific terms the research hypothesis is that the mean responses in the min-automated mode will be different than the mean responses in the max-automated mode and that such a difference is not due to sampling error.

11. RESULTS COMPARISON

The controller and system performance distributions obtained from the tests will be used to validate the Queuing Model. Validation of specific functions will be performed by comparison with the results of previous studies or (when there are no previous results for a function) by comparison with predicted results.

The previous studies to be considered are:

- Final Report on the Evaluation of a Semiautomatic Flight Operations Center
- Army Tactical Air Space Regulation System (ATARS)
- Design Plan Supplement for Basic Semiautomatic Flight Operations Center
- MTACCS
- FAA Controller task time data

The methods of implementing air traffic control functions and the scenarios used vary between the above studies and the ATMAC. It is necessary, therefore, to determine which study corresponds best for a particular function, and to make allowance for these differences in making the comparisons with the data collected on the I/A SIM.

For those ATMAC functions where there are no previous data (or the degree of correspondence is not acceptable), the test results will be compared with predicted results. The predicted results will be based on analysis of the modeling techniques used to generate the simulation and the test procedures developed for the test subjects. The analysis will separate the data processing actions from the test subject actions and assign elapsed time periods to each. The data processing times will be derived from the same or similar processing techniques used in ATMAC which have correspondence with one of the previous studies. The test subject times will be obtained with real time observations of the test subject performing the ATMAC functions.

To validate the data, controller response distributions from the previous studies or from the predicted results analysis will be statistically compared with the corresponding distributions from the Interactive simulator. The data will be considered validated if the comparison shows that both distributions are within the confidence interval of the selected population parameter. If the comparison is false, further attempts will be made to justify the differences by analysis. In the event that adequate explanations for the differences are found, the data obtained from the I/A SIM will be used as input parameters to the Queuing Model. If, on the other hand, inadequate reasons for the discrepancies are found, the I/A SIM data will be adjusted until the remaining differences can be justified.

GLOSSARY OF SYMBOLS

X_{ijk} = k th time response of operator i in mode j

λ_{ij} = mean response time for operator i in mode j (cell ij mean)

μ = general mean

α_i = main effect of operator i

β_j = main effect of mode j

γ_{ij} = interaction of operator i with mode j

E_{ijk} = sample error, $N(0, \sigma^2)$

H_0 = hypothesis: $\gamma_{ij}=0$ for all i, j

H_1 = hypothesis: $\beta_j=0$ for all j

H_2 = hypothesis: $\alpha_i=0$ for all i

α = significance level for the F-test

$\hat{\lambda}_{ij}$ = sample mean response time for operator i in mode j

$\hat{\mu}$ = general sample mean

$\hat{\alpha}_i$ = sample main effect of operator i

$\hat{\beta}_j$ = sample main effect of mode j

$\hat{\gamma}_{ij}$ = sample interaction between operator i and mode j

SS_A = sum of squares for main effects of the operators (α_i)

SS_B = sum of squares for main effects of the modes (β_j)

SS_C = sum of squares for the interactions (γ_{ij})

SS_E = sum of squares for the errors

ψ_{j1} = contrast of main effects of the operating modes ($\beta_j - \beta_1$)

$\hat{\psi}_{j1}$ = least square estimate of ψ_{j1} from the sampled values of X_{ijk}

θ_{i1} = contrast of the main effects of the operators ($\alpha_i - \alpha_1$)

$\hat{\theta}_{i1}$ = least square estimate of θ_{i1}

N = no. of observations from the ij-cell

n = no. of operators

m = no. of modes of operation

Means = sample means

SS = sum of squares

Q = general hypothesis for the two-way classification analysis of variance

CPD = cumulative probability distribution

12. DATA COLLECTION FROM THE INTERACTIVE SIMULATOR

Time measurements of controller interaction with simulated situations are collected from the Interactive Simulator and used as part of the validation data for the ATMAC Queuing Model. The collected data, as applied to the ATMAC Queuing Model, are in the form of distributions resulting from several trials of each simulated situation. Section 3 describes the data to be collected and its output format.

13. DEVELOPMENT OF OTHER MODEL VALIDATION DATA

The model validation data obtained from the Interactive Simulator represents only a portion of the data which is required to assure a totally valid queueing model. Such additional validation data include, but are not limited to, additional controller response time statistics derived from other sources, usage rate of various classes of subsystems, and flight durations and durations of discrete events as a function of the scenarios. Descriptions of these other validation techniques are included in Section 14.

14. ATMAC QUEUING MODEL VALIDATION

The model validation process includes the following:

- Use of valid scenarios by using Government furnished and acceptable documents.
- Validation of the ATMAC Model structure.
- Validation of the results of certain simulation runs using data obtained from the Interactive Simulator.
- Validation of simulation measures using empirical data.
- Indirect validation of other data which have not been previously developed.

The plan for Queuing Model validation is described in this section.

Whenever either of the hypotheses H_1 or H_2 is rejected under the hypothesis Q (that is, Ho is rejected), the S-method will be used to determine which factors contributed to the rejection and to estimate the size of the differences between the parameters. In particular, if H_1 is rejected, then the functions

$$\psi_{\ell j} = \beta_\ell - \beta_j, \; 1 \le \ell < j \le n,$$

will be tested for significant differences from 0 by use of the least square estimate $\hat{\psi}_{\ell j}$ of $\psi_{\ell j}$. Note that

$$\hat{\psi}_{\ell j} = B_\ell - B_j.$$

The basis of the S-method is the following fact: the probability is $1-\alpha$ that

$$\left| \psi_{\ell j} - \hat{\psi}_{\ell j} \right| < S\hat{\sigma}\left(\hat{\psi}_{\ell j}\right)$$

where

$$\hat{\sigma}^2\left(\hat{\psi}_{\ell j}\right) = \frac{SS_E}{N_t - nm}\left\{\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{N_{i\ell}} + \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{N_{ij}}\right\}$$

and S is determined from an F-distribution with parameters (m-1) and $(N_t - mn)$ such that

$$P_r\left(F \le \frac{S^2}{(m-1)}\right) = 1-\alpha.$$

The statement above implies that ψ_{ij} is significantly different from 0 if

$$0 < \hat{\psi}_{\ell j} - S\hat{\sigma}\left(\hat{\psi}_{\ell j}\right).$$

If H_2 is rejected under the hypothesis Q, then the functions

$$\theta_{i\ell} = \alpha_i - \alpha_\ell, \; 1 \le i < \ell \le n,$$

will be tested for significant differences from 0. The procedure is similar to that outlined above with

$$\hat{\theta}_{i\ell} = A_i - A_\ell,$$

and

$$\hat{\sigma}^2 (\hat{\theta}_{i\ell}) = \frac{SS_E}{N_t - nm} \left\{ \frac{1}{m^2} \sum_{j=1}^{m} \frac{1}{N_{ij}} - \frac{1}{m^2} \sum_{j=1}^{m} \frac{1}{N_{\ell j}} \right\} .$$

THE T-METHOD (TUKEY'S METHOD)

Whenever either of the hypotheses H_1 or H_2 is rejected, Tukey's Method (hereafter, simply the T-method) will be used to determine which factors contributed to the rejection and to estimate the size of the differences. In particular, if H_1 is rejected, then the functions

$$\psi_{j\ell} = \beta_j - \beta_\ell, \; 1 \le j < \ell \le m$$

will be tested for significant differences from 0 by use of the least square estimate $\hat{\psi}_{j\ell}$ of $\psi_{j\ell}$. Note that

$$\hat{\psi}_{j\ell} = \hat{\beta}_j - \hat{\beta}_\ell = \beta_j - \beta_\ell$$

Let α satisfy $0 < \alpha < 1$. The T-method is based upon the following theorem (due to Tukey): The probability is $1 - \alpha$ that

$$\left| \hat{\psi}_{j\ell} - \psi_{j\ell} \right| \le Ts \text{ for all } 1 \le j < \ell \le m,$$

where

$$s^2 = \frac{S_E}{N_t - mn}$$

and T is chosen to satisfy

$$P(t < b\,T) = 1 - \alpha,$$

$$b^2 = n\left(U_j^{-1} + U_\ell^{-1}\right)$$

where t has a student t-distribution with parameters m and $(N_t - mn)$. Thus, there are significant differences in the β_j's if

$$0 < \hat{\psi}_{j\ell} - TS$$

for some j and ℓ.

If the hypothesis H_2 is rejected, then the functions

$$\theta_{i\ell} = \alpha_i - \alpha_\ell, \; 1 \leq i < \ell \leq n$$

will be tested for significant differences from 0. The procedure is the same as that outlined above except that T is chosen to satisfy

$$P(t \leq a\,T) = 1 - \alpha,$$

$$a^2 = m\left(W_i^{-1} + W_\ell^{-1}\right)$$

where t has parameters n and $(N_t - mn)$.

15. INTERACTIVE SIMULATOR DATA ANALYSIS

The following paragraphs contain a description and summary of the results from the statistical analysis of the Interactive Simulator data proposed in the Validation Test Plan (VTP). As stated in the VTP, each of the ATMAC functions simulated on the Interactive Simulator generates data which represent controller response times, switch actions and efficiency of operation. The goal of the statistical analysis of this data is to gather meaningful data concerning the human response times which can be applied to the ATMAC Queueing Model.

a. Description of the Data.

The primary method of the statistical analysis of the I/A data was the two-way analysis of variance scheme outlined in the VTP; also see Henry Scheffé: The Analysis of Variance (New York: John Wiley & Sons, Inc.; 1959). The analysis of variance scheme identifies simultaneously differences in operator performance between the individual operators and between levels of automation. The data for those operators which consistently displayed response times significantly different from the majority was deleted from the data set; the analysis then was repeated for the remaining data. In this way it was possible to obtain information concerning the distribution of operator response times (the mean, standard deviation and cumulative distribution function) in each level of automation and to identify the differences in the response times which were a result of mode manipulation rather than sampling errors.

b. Interactive Simulator Data.

The data from the Interactive Simulator in the max-automated automation level consists of an average response time together with the number of samples, standard deviation, and cumulative distribution function for each operator with respect to each task.

In order for the results of the two factor analysis to be meaningful, it is necessary that the response times from the different levels of automation represent performance of the same (or at least similar) tasks. Unfortunately, in many cases the data from the interactive simulator in the max-auto mode do not correspond directly with the data from the interactive simulator in the min-auto mode of operation. Consequently it was necessary in some cases to group or identify two separate tasks as one task; in other cases it was necessary to add the response times. The formulae for these manipulations are given in the following paragraph.

As in the VTP, let the operators be indexed by the subscript i and the level of automation by the subscript j. For a given task, N_{ij} represents the number of times operator i performed the task in level j; λ_{ij} and σ_{ij} represent the corresponding mean and standard deviation of the response times. Whenever it was necessary to group or identify two tasks, the new statistics N_{ij}, λ_{ij}, σ_{ij} were computed from the original statistics $N_{ij}^{(1)}$, $\lambda_{ij}^{(1)}$, $\sigma_{ij}^{(1)}$ and $N_{ij}^{(2)}$, $\lambda_{ij}^{(2)}$, $\sigma_{ij}^{(2)}$ as follows:

$$N_{ij} = N_{ij}^{(1)} + N_{ij}^{(2)},$$

$$\lambda_{ij} = \frac{1}{N_{ij}}\left\{N_{ij}^{(1)}\,\lambda_{ij}^{(1)} + N_{ij}^{(2)}\,\lambda_{ij}^{(2)}\right\},$$

$$\sigma_{ij}^{2} = \frac{1}{N_{ij}}\left\{N_{ij}^{(1)}\left[\left(\lambda_{ij}^{(1)}\right)^2 + \left(\sigma_{ij}^{(1)}\right)^2\right] + N_{ij}^{(2)}\left[\left(\lambda_{ij}^{(2)}\right)^2 + \left(\sigma_{ij}^{(2)}\right)^2\right]\right\} - \lambda_{ij}^{2}.$$

Whenever it was necessary to add response times, the new statistics were computed according to:

$$N_{ij} = \text{minimum}\left[N_{ij}^{(1)},\ N_{ij}^{(2)}\right],$$

$$\lambda_{ij} = \lambda_{ij}^{(1)} + \lambda_{ij}^{(2)},$$

$$\sigma_{ij}^{2} = \left(\sigma_{ij}^{(1)}\right)^2 + \left(\sigma_{ij}^{(2)}\right)^2.$$

In addition to the discrepancy between the min-auto and max-auto modes with respect to the definition of the tasks, there is another fundamental difference in the data. A response time in the max-auto mode includes the time necessary to perform two distinct phases of the task; namely a decision-making process in which the appropriate action is determined followed by the physical or mechanical processes (i.e., switch actions) necessary to implement the decision. In contrast to the max-auto mode, the min-auto mode response times do not include the times for the decision making process because only times for voice communications are available. In order to compensate for this difference between the data from the two modes, the manual mode response times were increased by 25 percent. No attempt was made either to justify the magnitude of the increase or to investigate the sensitivity of the results to the magnitude of the increase.

c. Analysis of Variance: Data.

The two-factor analysis of variance scheme was used to analyze the Approach Control, Departure Control, Enroute Control and Handover data. Valid Flight Plan data was collected from two operators only; consequently the means and standard deviations for the response times were computed with no additional data evaluation.

When the analysis of variance methods were used to analyze the data for a particular ATMAC function, a computer program (a listing is included with the data) was executed to perform the necessary computations. Each time the program was executed, the data described in the following sections was printed.

(1) For each operator, the mean response time λ_{ij}, the number of observations N_{ij}, the standard error of the mean S_{ij} and a 90 percent confidence interval for the true mean μ_{ij} of the distribution is printed for each method of operation. The standard error of mean S_{ij} is

$$S_{ij} = \frac{\sigma_{ij}}{\sqrt{N_{ij} - 1}}.$$

The confidence interval is computed by considering

$$Pr(-b < T < b) = 0.90$$

where

$$T = \frac{\lambda_{ij} - \mu_{ij}}{S_{ij}}$$

has a student's t - distribution with parameter $N_{ij} - 1$. The variable μ_{ij} represents the true mean. The probability is 0.90 that

$$\lambda_{ij} - bs_{ij} < \mu_{ij} < \lambda_{ij} + bs_{ij}.$$

(2) For each mode of operation under consideration, the mean operator response time, the standard error of the mean, and a 90 percent confidence interval for the true mode mean are printed. The first value of the mode mean λ_i is computed according to

$$\lambda_j = \frac{1}{n} \sum_{i=1}^{n} \lambda_{ij}$$

where n is the number of operators considered. The second value of the mode mean $\bar{\lambda}_j$ is given by

$$\bar{\lambda}_j = \frac{1}{N_j} \sum_{i=1}^{n} N_{ij} \lambda_{ij}$$

where

$$N_j = \sum_{i=1}^{n} N_{ij}.$$

Similarly, two values of the standard error, s_j and $\bar{s}_j$, are given;

$$s_j^2 = \frac{1}{(n-1)^2} \sum_{i=1}^{n} \sigma_{ij}^2,$$

$$\bar{s}_j^2 = \frac{1}{(N_j-1)^2} \sum_{i=1}^{n} N_{ij} \left[\lambda_{ij}^2 + \sigma_{ij}^2\right] - \bar{\lambda}_j^2.$$

Two confidence intervals for the true mode mean μ_j are computed by considering the student t - statistics

$$T = \frac{\lambda_j - \mu_j}{s_j}$$

and

$$T = \frac{\bar{\lambda}_j - \mu_j}{\bar{\bar{s}}_j}$$

which have parameters n-1 and N_j-1, respectively.

(3) The analysis of variance parameters are printed also. These include the general mean μ, the variations α_i in the mean due to the operators (A-main effects), the variations β_j in the mean due to the automation levels (B-main effects), and the interactions γ_{ij}. Recall that the underlying assumption of the analysis of variance scheme is that the individual response times X_{ijk} satisfy

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon,$$

where ϵ is a random sample from a normal distribution with mean O and variance σ^2. In addition to these parameters, five F-statistics are printed. (See Table V).

16. SUMMARY OF RESULTS

In the following sections, the mean response times together with the corresponding standard deviations are given for the various ATMAC functions. Except for the Flight Plan data, the analysis of variance scheme was used to isolate those operators who consistently displayed response times significantly different (as determined by the S or T method of multiple comparison). The mean and standard deviation of the response times were computed by deleting the data of the operators

Table V. F-STATISTICS

Statistic	Degrees of Freedom	Use
F(1)	$mn - n - m + 1,\ N_t - mn$	Test H_0 against Q (Additivity: $\gamma_{ij} = 0$)
F(2)	$n - 1,\ N_t - m - n + 1$	Test H_2 against $\Omega \cap Q$ ($\alpha_i = 0$)
F(3)	$m - 1,\ N_t - m - n + 1$	Test H_1 against $\Omega \cap Q$ ($\beta_j = 0$)
F(4)	$n - 1,\ N_t - mn$	Test H_2 against Q ($\alpha_i = 0$)
F(5)	$m - 1,\ N_t - mn$	Test H_1 against Q ($\beta_j = 0$)

m = number of automation levels

n = number of operators

N_t = total number of observations

mentioned above. A computer print-out of the original data sets as well as the final data sets is included.

As was stated previously, the Flight Plan means and standard deviations were computed by using all of the available data.

17. FLIGHT PLAN ENTRY/MODIFICATION/CLEARANCE

The following is based upon the data collected from operators Olson and Wilson.

Table VI. FLIGHT PLAN DATA

Function	Mean	Std. Dev. of Population	Std. Dev. of Mean	No. of Observations
Entry Time (Seconds)	209.50	62.58	11.43	30
Modification Time	52.05	21.75	4.44	24
Clear-Reject Time	130.12	67.19	14.66	21

18. APPROACH CONTROL

The analysis of variance program was used to compare the max-automated mode of operation with the min-automated mode of operation for two data sets. In the first data set, which is labeled "Accept Hand-In", the automatic mode data consists of the I/A SIM data bearing the same label; the min-auto mode data consists of the sum of the following tasks:

- Establish contact with adjacent sector controller
- Complete handover
- Pilot establishes contact

The second data set, which is labeled "Approve Vector Instructions", consists of the I/A SIM data with the same name for the max-auto mode together with the following grouped min-auto mode data:

- Provide vector instructions
- Issue formation breakup command.

In both cases, the analysis showed that the data did not satisfy the criterion for additivity; that is, H_0 was rejected. Again, in both cases, all but two operators were eliminated from consideration; for both cases, the final mean response time in the max-auto mode was significantly (by the S-method) less than the mean response time in the min-auto mode.

Table VII. APPROACH CONTROL DATA: MAX-AUTOMATED MODE

Function	Mean	Std. Dev. of Population	Std. Dev. of Mean	No. of Observations
Accept Hand-In	7.34 sec.	5.15	0.72	51
Approve Vector Instructions	3.38 sec.	4.45	0.58	59
Resolution of Separation Alert	42.33 sec.	29.85	8.62	12
Minimum Approach Separation Distance	553.57 meters	310.83	89.73	12

Table VIII. APPROACH CONTROL DATA: MIN-AUTOMATED MODE

Function	Mean (sec.)	Std. Dev		No. of Observations
		of Pop.	of Mean	
Provide Vector Instructions	5.67	2.53	0.15	294
Formation Backup Command	4.72	4.21	0.84	25
Establish Contact with adj. sector	1.55	1.71	0.41	17
Complete Handover	10.94	5.10	1.20	18
Pilot establishes contact	3.76	1.44	0.29	24

19. DEPARTURE CONTROL.

The analysis of variance program was used to compare the max-auto mode of operation with the min-auto mode of operation for two sets of data. In the first data set, which is labeled "Alert Acknowledge", the max-auto mode data consists of the I/A SIM data with the same name; the min-auto mode data consists of the sum of the following steps:

- Receive call from tower
- Request and receive departure data

In the second data set, which is labeled "Approve Vector Instructions", the max-auto mode data consists of the I/A SIM data with the same name; the min-auto mode data consists of:

- Provide Vector Instructions
- Provide Departure Instructions

For the "Alert Acknowledge" data, the hypothesis of additivity, that is H_0, was accepted. The mean response time for the max-auto mode was significantly less than that of the min-auto mode (as determined by the S or T method). In the second data set, the hypothesis H_0 was rejected; however the max-auto mode response time also was found to be significantly less than the min-auto mode response time.

20. ENROUTE CONTROL

The analysis of variance program was used to compare the operator response times for conflict resolution in the max-automated, semi-automated, and min-automated modes of operation. The max-auto mode data consists of the I/A SIM data for the resolution time of air-to-air conflicts and air-to-restricted zone conflicts. The semi-automated mode data consists of all data for: "Provide vector instructions". The min-auto mode data is similar to that of the semi-auto mode.

Table IX. DEPARTURE CONTROL DATA: MAX-AUTOMATED MODE

Function	Mean	Std. Dev.		No. of Observations
		of Pop.	of Mean	
Approve Vector Instructions	2.80 sec.	3.13	0.28	122
Departure Alert Acknowledge	4.96 sec.	4.18	0.47	79
Resolution of Separation Alert	41.52 sec.	26.03	8.23	10
Minimum Departure Separation Distance	362.44 meters	440.92	139.43	10

Table X. DEPARTURE CONTROL DATA: MIN-AUTOMATED MODE

Function	Mean (sec.)	Std. Dev.		No. of Observations
		of Pop.	of Mean	
Provide Vector Instructions	5.63	2.65	0.16	262
Receive Call from Tower	1.81	1.43	0.19	55
Request and Receive Departure Data	6.49	2.81	0.38	56
Provide Departure Instructions	6.56	4.79	0.56	74

Successive application of the analysis of variance scheme eliminated all but three operators. On the basis of the data for these three, the hypotheses H_0 and H_1 were rejected. By the S-method, the response times in the semi-automated mode were found to be significantly less than those of the other two modes; however there was no significant difference between the max-auto and min-auto mode response times.

In addition to the conflict resolution data, the analysis of variance program was exercised to compare the operator response times for the max-auto mode against the response times for the semi-auto mode for the Air-to-Air and the Air-to-Restricted Zone Conflict Alert Acknowledge Time data. Both data sets were taken directly from the I/A SIM output data.

The analysis of variance indicates that, in both data sets, H_0 is accepted and H_1 is rejected. Relative to the S-method, the mean response time in the max-auto mode is significantly less than that of the semi-automated mode.

Table XI. ENROUTE CONTROL DATA: MAX-AUTOMATED AND SEMI-AUTOMATED MODES (DATA COLLECTED BY COMPUTER PROGRAM)

Function	Mean (sec.)	Std. Dev.		No. of Observations
		of Pop.	of Mean	
Conflict Alert Ack.				
Air-Air (Max-Auto)	15.31	8.93	0.49	332
(Semi-Auto)	19.15	9.56	0.52	339
Air-Restricted Zone				
(Max-Auto)	14.94	9.04	1.05	74
(Semi-Auto)	19.11	10.07	1.28	62
Conflict Resolution				
Air-Air	7.69	4.20	0.26	271
Air-Restricted Zone	7.20	3.00	0.95	10

Table XII. ENROUTE CONTROL DATA: MIN-AUTOMATED AND SEMI-AUTOMATED MODES (DATA COLLECTED FROM VOICE TAPES)

Function	Mean (sec.)	Std. Dev.		No. of Observations
		of Pop.	of Mean	
Provide Vector Instructions (Min-Auto)	6.69	3.37	0.26	172
Provide Vector Instructions (Semi-Auto. Mode)	5.30	2.55	0.17	225
Controller calls adjacent facility and completes handover (Semi-Auto)	3.07	0.92	0.09	111

21. HANDOVER

For each of the three general functions – Approach, Departure and Enroute Control – data was collected for the associated handover functions. The analysis of variance program was exercised in order to compare the max-automated with the min-automated mode response times for the Approach and Departure Control functions; the program was used to compare the response times for all three levels of automation for the Enroute Control function. In all of the three data sets bearing the label "Handover", the max-auto data (and the semi-auto mode data for the Enroute Control function) consists of the sum of the response times labeled "Handover Alert Acknowledge" and "Handover Coordination Time". In the data set for the Approach Control function, the min-auto mode data is the sum of the following Steps:

- Establish contact with adjacent sector controller
- Complete Handover
- Pilot establishes contact

together with the sum of:

- Initiate tower handover
- Complete tower handover.

For the Departure Control function, the min-auto mode data is the grouped data for:

- Pilot establishes contact
- Contact adjacent sector controller
- Complete handover.

Finally, for the Enroute Control function, the min-auto mode data consists of the sum of the data for:

- Establish contact with adjacent sector controller
- Complete handover
- Pilot establishes contact

together with the sum of the data for:

- Contact adjacent sector controller
- Complete handover.

For both the Approach Control function and the Departure Control function, the hypothesis H_0 was accepted and the hypothesis H_1 was rejected. In both cases, the S-method indicated that the mean response time for the max-auto mode is significantly less than the mean response time for the min-auto mode. On the other hand, the analysis of variance of the Enroute Handover data resulted in the rejection of H_0 and the acceptance H_1, that is, there is no significant difference in the response times for the three modes of operation when tested by the S-method.

Table XIII. HANDOVER: MAX-AUTOMATED AND SEMI-AUTOMATED MODES

Mode	Function	Mean (sec.)	Std. Dev. of Pop.	Std. Dev. of Mean	No. of Observations
Approach Control	1	6.63	7.20	0.45	259
(Max-Automated)	2	7.91	7.20	0.44	268
Departure Control	1	6.35	4.52	0.39	135
(Max-Auto)	2	7.15	4.54	0.39	135
Enroute Control	1	9.25	10.80	0.71	234
(Max-Auto)	2	9.33	7.83	0.51	234
(Semi-Auto)	1	9.50	11.89	0.79	228
	2	9.54	9.52	0.63	228

Function 1 – Handover Alert Acknowledge

Function 2 – Handover Coordination Time

Table XIV. HANDOVER: MIN-AUTOMATED MODE

Function	Mean	Std. Dev. of Pop.	Std. Dev. of Mean	No. of Observations
Approach Control				
Initiate Tower Handover	1.71	2.77	0.42	43
Complete Tower Handover	9.77	3.89	0.62	39
Departure Control				
Pilot establishes contact	4.05	1.86	0.19	101
Contact adjacent sector controller	3.64	2.87	0.29	100
Complete Handover	5.73	3.82	0.38	100
Enroute Control				
Establish contact with adjacent controller	2.44	2.28	0.23	96
Complete handover	9.95	5.16	0.53	95
Contact adjacent sector controller	1.96	0.75	0.10	62
Complete handover	11.24	5.42	0.69	61
Pilot establishes contact	4.19	2.11	0.22	89

22. DETAILED ANALYSIS OF THE DATA FOR APPROACH CONTROL: ACCEPT HAND-IN

The original data set for the Approach Control: Accept Hand-in function is contained in the Appendix (A1.1). The results of the application of the analysis of variance program are given in the Appendix (A1.2). A brief inspection of the F-statistics in the Appendix (A1.2) will indicate that the hypothesis H_z will be rejected by the F-test. Table XV contains the values of the contrasts

$$\hat{\theta}_{il} = \hat{\alpha}_i - \hat{\alpha}_l$$

for $1 \leq i < l \leq 7$; the value of $\hat{\sigma}\ (\theta_{il})$ appears immediately below $\hat{\theta}_{il}$.

According to Scheffé's method, $\hat{\theta}_{il}$ will be significantly different from 0 if

$$(1) \quad |\hat{\theta}_{il}| > S\ \hat{\sigma}\ (\theta_{il})$$

where S is defined by

$$(2) \quad P_r\ (F \leq S^2/(n - 1) = 1 - \alpha = .95$$

for an F-distribution with (n - 1) and (N_t - mn) degrees of freedom. In this case S = 3.55. From Table XV it is clear that the inequality (1) is valid for l = 4, i - 1, 2, 3, 5, 6, 7.

The Appendix (A1.3 and A1.4) contains the results of the analysis of variance with the data for operator 4 deleted. The value of S in equation (2) is 3.33; from Table XVI it is clear that inequality (1) is valid for l = 4, i - 1, 2, 3.

The Appendix (A1.5 and A1.6) contains the results of the analysis of variance with the data for operators 4 and 5 deleted. Since the value of F(1) is 7.12, the hypothesis H_o (Additivity) is rejected; thus F(4) is used to test H_z. The degrees of freedom for F(4) are 4 and 128. Now

$$P_r\ (F > f) = \alpha = .05$$

implies that f is approximately 2.45. Since the value of F(4) is 2.84, H_z is rejected again. The value of S in (2) is 3.10. Table XVII shows that

$$|\hat{\theta}_{ij}| \leq S\,\sigma\ (\hat{\theta}_{ij})$$

for all i, j. Consequently, the rejection of the hypothesis H_z is due to a linear combination of the α_i's other than a simple contrast (that is, the θ_{ij}'s). However, the value of the sample standard deviation $\hat{\sigma}_{ij}$ for operator 4 (6 in the Appendix (A1.1)) is 21.4; the values of $\hat{\sigma}_{ij}$ for the other operators in the Appendix (A1.5) are 3.8, 4.9, 5.1, and 6.5, respectively. Therefore, the data for Operator 4 was deleted from the data set and the analysis was repeated.

The results of the analysis are given in the Appendix (A1.7 and A1.8). The hypothesis H_z was rejected again. The value of S in (2) is 2.87; the inequality (1) holds for θ_{14} and θ_{34}. The data for operator 4 (7 of the original) was deleted and the analysis repeated. The results are given in the Appendix (A1.9 and A1.10). The appropriate value of S in (2) is 2.49; from Table XIX it is clear that (1) holds for $\hat{\theta}_{12}$. The data for operator 2 was deleted.

The analysis of variance program was applied to the data for the remaining two operators, the results are given in the Appendix (A1.11 and A1.12). The F statistic in the Appendix (A1.12) indicate that H_0 and H_z will be accepted and that H_1 will be rejected. The value of ψ_{12} is - 9.01 where

$$\psi_{12} = \beta_1 - \beta_2.$$

The value of S in (2) (where m replaces n) is approximately 1.6. Since the value of $\hat{\sigma}$ (ψ_{12}) is 1.12, β_1 is significantly larger than β_2.

Table XV . MULTIPLE COMPARISON BY THE S-METHOD

		2	3	4	5	6	7
1	θ_{12} $\hat{\sigma}(\theta_{12})$	3.34 5.05	1.03 4.70	32.16 5.61	11.32 4.84	5.11 5.38	5.38 4.97
2			- 2.31 5.01	28.83 5.87	7.98 5.15	1.77 5.65	2.04 5.27
3				31.13 5.57	10.29 4.81	4.08 5.35	4.35 4.94
4					-20.84 5.69	-27.05 6.16	-26.78 5.81
5						- 6.21 5.47	- 5.94 5.08
6							.27 5.58

Note: Line 1 is $\theta_{il} = \alpha_i - \alpha_l$ for $i < l$.

Line 2 is $\hat{\sigma}(\theta_{il})$.

Table XVI. MULTIPLE COMPARISON BY THE S-METHOD

		2	3	4	5	6
1		3.38 2.20	1.03 2.05	11.32 2.11	5.11 2.34	5.38 2.17
2			2.31 2.18	7.98 2.24	1.77 2.46	2.04 2.29
3				10.29 2.09	4.08 2.33	4.35 2.15
4					- 6.21 2.38	- 5.94 2.21
5						.27 2.26

Table XVII. MULTIPLE COMPARISON BY THE S-METHOD

		2	3	4	5
1		3.33 2.01	1.03 1.87	5.11 2.14	5.38 1.98
2			- 2.31 1.19	1.77 2.25	2.04 2.09
3				4.08 2.13	4.35 1.96
4					.27 2.22

Table XVIII. MULTIPLE COMPARISON

		2	3	4
1		3.34 1.40	1.03 1.30	5.38 1.38
2			- 2.31 1.39	2.04 1.46
3				4.35 1.43

Table XIX. MULTIPLE COMPARISON

		2	3
1		3.34 1.30	1.03 1.21
2			- 2.31 1.29

ACKNOWLEDGMENTS

The author gratefully acknowledges the aid, assistance, and cooperation of:

Dr. C. E. Kastenholtz, Technical Advisor to the ATMAC Project, Hughes Aircraft Co., Fullerton, California, Ground Systems Group.

Dr. Martin Dana, Systems Analysis Division, Hughes Aircraft Company for the programing and computations of the analysis of variance data.

The Air Traffic Regulation Advanced Team Technical Area, Avionics Laboratory, USAECOM, Fort Monmouth, N.J., Messrs. H. Mencher, E. Hansen, A. Coppola, J. Parker, and H. Tanzman.

Miss Ann McGough for the aid in the preparation and correction of the manuscript.

APPROACH CONTROL: ACCEPT HAND-IN

A1.1

MANUAL MODE

OPERATOR	1	2	3	4	5	6	7
OPERATOR MEANS	16.37	24.35	17.04	28.60	37.11	16.90	27.34
NUMBER OF OBSERVATIONS	14	9	11	5	11	10	9
ST. ERROR OF THE MEANS	1.03	2.18	1.39	4.41	2.55	1.15	1.56
90% CONFIDENCE INT.	14.15	19.31	13.95	16.36	31.44	14.30	23.73
	18.60	29.39	20.12	40.84	42.78	19.50	30.94

MODE MEAN	23.96	23.22
ST. ERROR OF THE MEAN	2.50	0.89
90% CONFIDENCE INT.	17.85	21.45
	30.07	24.99

AUTOMATIC MODE

OPERATOR	1	2	3	4	5	6	7
OPERATOR MEANS	7.00	5.70	8.40	59.10	8.90	16.70	6.80
NUMBER OF OBSERVATIONS	18	20	31	35	21	10	24
ST. ERROR OF THE MEANS	0.92	1.17	0.89	7.31	2.64	7.13	1.36
90% CONFIDENCE INT.	5.06	3.25	6.57	44.25	3.40	0.56	4.00
	8.94	8.15	10.23	73.95	14.40	32.84	9.60

MODE MEAN	16.09	19.41
ST. ERROR OF THE MEAN	4.40	2.56
90% CONFIDENCE INT.	5.32	14.35
	26.85	24.47

GENERAL MEAN = 20.0223

STANDARD ERROR = 18.8084 , $= \sqrt{\frac{SSE}{N_t - mn}}$

A MAIN EFFECTS =	-8.33479	-4.99728	7.30353	23.8277	2.98395	-3.22229	-2.95354

STANDARD ERROR(ADDITIVE) = 103.509

STANDARD ERROR(NON-ADD.) = 49.6484

B MAIN EFFECTS = 3.937 -3.937

STANDARD ERROR(ADDITIVE) = 69.1122

STANDARD ERROR(NON-ADD.) = 51.8188

INTERACTIONS =	0.7509	5.388	0.3821	-19.19	10.17	-3.837	6.332
	-0.7509	-5.388	-0.3822	19.19	-10.17	3.837	-6.332

STANDARD ERROR = 43.3141

F STATISTICS	5.3034	27.1055	612.0840	6.9680	7.5905

NUMBER OF OPERATORS = 7 = n

NUMBER OF MODES = 2 = m

TOTAL NUMBER OF OBSERVATIONS = 228 = N_t

$$\hat{\theta}_{12} = \hat{\alpha}_1 - \hat{\alpha}_2$$

APPROACH CONTROL: ACCEPT HAND-IN

MANUAL MODE

OPERATOR	1	2	3	4	5	6
OPERATOR MEANS	16.37	24.35	17.04	37.11	16.90	27.34
NUMBER OF OBSERVATIONS	14	9	11	11	10	9
ST. ERROR OF THE MEANS	1.03	2.18	1.39	2.55	1.15	1.56
90% CONFIDENCE INT.	14.15	19.31	13.95	31.44	14.30	23.73
	18.60	29.39	20.12	42.78	19.50	30.94

MODE MEAN	23.19	22.80
ST. ERROR OF THE MEAN	1.90	1.05
90% CONFIDENCE INT.	18.31	20.70
	28.06	24.89

AUTOMATIC MODE

OPERATOR	1	2	3	4	5	6
OPERATOR MEANS	7.00	5.70	8.40	8.90	16.70	6.60
NUMBER OF OBSERVATIONS	18	20	31	21	10	24
ST. ERROR OF THE MEANS	0.92	1.17	0.89	2.64	7.13	1.36
90% CONFIDENCE INT.	5.06	3.25	6.57	3.40	0.56	4.00
	8.94	8.15	10.23	14.40	32.84	9.60

MODE MEAN	8.92	8.21
ST. ERROR OF THE MEAN	3.54	0.79
90% CONFIDENCE INT.	-0.19	6.64
	18.02	9.77

GENERAL MEAN = 16.0510

STANDARD ERROR = 8.18752

A MAIN EFFECTS =	-4.36353	-1.02603	-3.33228	6.95520	0.748962	1.01772

STANDARD ERROR(ADDITIVE) = 17.1635

STANDARD ERROR(NON-ADD.) = 21.9849

B MAIN EFFECTS =	7.134	-7.134

STANDARD ERROR(ADDITIVE) = 93.5568

STANDARD ERROR(NON-ADD.) = 90.1149

INTERACTIONS =	-2.447	2.191	-2.816	6.972	-7.034	3.134
	2.447	-2.191	2.816	-6.972	7.034	-3.134

STANDARD ERROR = 25.3402

F STATISTICS	9.5789	3.5526	105.5556	7.2101	121.1403

NUMBER OF OPERATORS = 6

NUMBER OF MODES = 2

TOTAL NUMBER OF OBSERVATIONS = 188

APPROACH CONTROL: ACCEPT HAND-IN

A1.5

MANUAL MODE

OPERATOR	1	2	3	4	5
OPERATOR MEANS	16.37	24.35	17.04	16.90	27.34
NUMBER OF OBSERVATIONS	14	9	11	10	9
ST. ERROR OF THE MEANS	1.03	2.18	1.39	1.15	1.56
90% CONFIDENCE INT.	14.15	19.31	13.95	14.30	23.73
	18.60	29.39	20.12	19.50	30.94

MODE MEAN	20.40	19.83
ST. ERROR OF THE MEAN	1.70	0.56
90% CONFIDENCE INT.	15.69	18.70
	25.11	20.95

AUTOMATIC MODE

OPERATOR	1	2	3	4	5
OPERATOR MEANS	7.00	5.70	8.40	16.70	6.80
NUMBER OF OBSERVATIONS	18	20	31	10	24
ST. ERROR OF THE MEANS	0.92	1.17	0.89	7.13	1.36
90% CONFIDENCE INT.	5.06	3.25	6.57	0.56	4.00
	8.94	8.15	10.23	32.84	9.60

MODE MEAN	8.92	6.06
ST. ERROR OF THE MEAN	3.73	0.79
90% CONFIDENCE INT.	-1.44	6.49
	19.28	9.64

```
GENERAL MEAN =   14.6600

    STANDARD ERROR =   7.47661

A MAIN EFFECTS =  -2.97250     0.365003     -1.94125     2.14000     2.40875

    STANDARD ERROR(ADDITIVE) =   9.46994

    STANDARD ERROR(NON-ADD.) =   12.5958

B MAIN EFFECTS =    5.740      -5.740

    STANDARD ERROR(ADDITIVE) =   68.0205

    STANDARD ERROR(NON-ADD.) =   65.6680

INTERACTIONS =    -1.052       3.585      -1.421      -5.640       4.529
                   1.053      -3.585       1.421       5.640      -4.529

    STANDARD ERROR =   19.9436

F STATISTICS         7.1154      1.3794      71.1641      2.8382      77.1431

   NUMBER OF OPERATORS =  5

   NUMBER OF MODES =  2

   TOTAL NUMBER OF OBSERVATIONS = 156
```

APPROACH CONTROL: ACCEPT HAND-IN A1.7

MANUAL MODE

OPERATOR	1	2	3	4
OPERATOR MEANS	16.37	24.35	17.04	27.34
NUMBER OF OBSERVATIONS	14	9	11	9
ST. ERROR OF THE MEANS	1.03	2.18	1.39	1.56
90% CONFIDENCE INT.	14.15	19.31	13.95	23.73
	18.60	29.39	20.12	30.94

MODE MEAN	21.27	20.51
ST. ERROR OF THE MEAN	1.84	0.52
90% CONFIDENCE INT.	15.41	19.47
	27.14	21.55

AUTOMATIC MODE

OPERATOR	1	2	3	4
OPERATOR MEANS	7.00	5.70	8.40	6.80
NUMBER OF OBSERVATIONS	18	20	31	24
ST. ERROR OF THE MEANS	0.92	1.17	0.89	1.36
90% CONFIDENCE INT.	5.06	3.25	6.57	4.00
	8.94	8.15	10.23	9.60

MODE MEAN	6.97	7.14
ST. ERROR OF THE MEAN	1.27	0.58
90% CONFIDENCE INT.	2.93	5.98
	11.02	8.29

GENERAL MEAN = 14.1250 A1.8

STANDARD ERROR = 5.20123

A MAIN EFFECTS = -2.43750 0.900002 -1.40625 2.94375

STANDARD ERROR(ADDITIVE) = 8.25836

STANDARD ERROR(NON-ADD.) = 12.8292

B MAIN EFFECTS = 7.150 -7.150

STANDARD ERROR(ADDITIVE) = 73.5125

STANDARD ERROR(NON-ADD.) = 76.1624

INTERACTIONS =	-2.462	2.175	-2.831	3.119
	2.463	-2.175	2.831	-3.119

STANDARD ERROR = 16.4669

F STATISTICS	10.0233	2.0893	165.5515	6.0839	214.4221

NUMBER OF OPERATORS = 4

NUMBER OF MODES = 2

TOTAL NUMBER OF OBSERVATIONS = 136

APPROACH CONTROL: ACCEPT HAND-IN

MANUAL MODE

OPERATOR	1	2	3
OPERATOR MEANS	16.37	24.35	17.04
NUMBER OF OBSERVATIONS	14	9	11
ST. ERROR OF THE MEANS	1.03	2.18	1.39
90% CONFIDENCE INT.	14.15	19.31	13.95
	18.60	29.39	20.12
MODE MEAN	19.25	18.70	
ST. ERROR OF THE MEAN	1.97	0.63	
90% CONFIDENCE INT.	10.78	17.43	
	27.73	19.97	

AUTOMATIC MODE

OPERATOR	1	2	3
OPERATOR MEANS	7.00	5.70	8.40
NUMBER OF OBSERVATIONS	18	20	31
ST. ERROR OF THE MEANS	0.92	1.17	0.89
90% CONFIDENCE INT.	5.06	3.25	6.57
	8.94	8.15	10.23
MODE MEAN	7.03	7.25	
ST. ERROR OF THE MEAN	1.23	0.63	
90% CONFIDENCE INT.	1.75	6.00	
	12.32	8.51	

A1.10

```
GENERAL MEAN =  13.1437

    STANDARD ERROR =   4.84068

A MAIN EFFECTS =  -1.45625      1.88125    -0.424997

    STANDARD ERROR(ADDITIVE) =   5.70259

    STANDARD ERROR(NON-ADD.) =   8.90652

B MAIN EFFECTS =    6.110      -6.110

    STANDARD ERROR(ADDITIVE) =   55.1203

    STANDARD ERROR(NON-ADD.) =   57.1693

INTERACTIONS =     -1.423       3.215       -1.792
                    1.423      -3.215        1.792

    STANDARD ERROR =   14.4658

F STATISTICS        8.9504      1.1962     111.7563      3.3853     139.4799

   NUMBER OF OPERATORS =  3

   NUMBER OF MODES =  2

   TOTAL NUMBER OF OBSERVATIONS = 103
```

APPROACH CONTROL: ACCEPT HAND-IN

MANUAL MODE

OPERATOR	1	2
OPERATOR MEANS	16.37	17.04
NUMBER OF OBSERVATIONS	14	11
ST. ERROR OF THE MEANS	1.03	1.39
90% CONFIDENCE INT.	14.15	13.95
	18.60	20.12
MODE MEAN	16.71	16.67
ST. ERROR OF THE MEAN	1.73	0.81
90% CONFIDENCE INT.	-5.24	15.00
	38.65	18.33

AUTOMATIC MODE

OPERATOR	1	2
OPERATOR MEANS	7.00	8.40
NUMBER OF OBSERVATIONS	18	31
ST. ERROR OF THE MEANS	0.92	0.89
90% CONFIDENCE INT.	5.06	6.57
	8.94	10.23
MODE MEAN	7.70	7.89
ST. ERROR OF THE MEAN	1.28	0.71
90% CONFIDENCE INT.	-8.62	6.45
	24.02	9.32

A1.12

```
GENERAL MEAN =   12.2031

     STANDARD ERROR =    4.48648

A MAIN EFFECTS = -0.515625      0.515628

     STANDARD ERROR(ADDITIVE) =    4.78156

     STANDARD ERROR(NON-ADD.) =    4.12453

B MAIN EFFECTS =      4.503       -4.503

     STANDARD ERROR(ADDITIVE) =   35.9959

     STANDARD ERROR(NON-ADD.) =   36.0141

INTERACTIONS =     0.1844      -0.1844
                  -0.1844       0.1844

     STANDARD ERROR =   1.47016

F STATISTICS        0.1074      1.1503      65.1911      0.8452      64.4366

    NUMBER OF OPERATORS =  2

    NUMBER OF MODES =  2

    TOTAL NUMBER OF OBSERVATIONS = 74
```

A FLEXIBLE, GENERAL PURPOSE COVARIANCE COMPUTER PROGRAM

Clifford J. Maloney
and
Lucille A. Carver
Food and Drug Administration
Bethesda, Maryland 20034

I. Introduction: Heraclitus, five centuries before the birth of Christ, maintained that no one steps in the same river twice; yet "enduring as a mountain" is an ancient simile of permanence though rivers are older than the mountains they drain. Life is lived in a multi-faceted, ever changing milieu; yet, if the past is to be our guide to the future, that can only be because the future is like the past. It is the invariant features, not the ephemeral, that matter. Constancy adheres to substance, properties, and relations. Father or son denotes no particular individual, but the father-son relation is fixed.

Strictly speaking, substance is neither accessible to man, nor needed in the conduct of affairs. The scientific revolution in essence was an appreciation of that fact. Substance is an intellectual construct whose constancy or lack of it is inferred from constancy or variation in properties or relations. The latter two alone receive the professional attention of scientists, though scientists have varying appreciation of the validity of that fact.

In turn, properties and relations are qualitative or quantitative. Each is often directly apparent, and where that is so, have been known since antiquity. But where the qualitative is inferred from quantitative observations it was, in practice, necessary to await: (1) the discovery of which quantitative relations throw light on which qualitative properties or relations and (2) the development of appropriate quantitative scales and devices.

Until recently this latter process was carried out informally, each quantitative construct being worked out anew with no attention to accumulated experience until Helmholtz (5) developed a general theory of mensuration. An adequate elementary treatment is now available (2).

A complementary advance proceeds in the opposite direction. While it is often necessary to discover which "instrumental" variables to observe in order to measure, control, or understand a quantity of interest, it is equally imperative to learn which plausible but irrelevant extraneous variables to ignore. The phases of the moon, the conjunctions of the planets, the services of soothsayers (except in economics) are easily recognized as such. But choice of treatment or prognosis for a diseased individual is a constant reminder that the problem remains in more difficult circumstances. In essence, statistical method is the form in which this process is most efficiently carried out.

In statistical terms, we express constancy of a property by some "measure of central tendency"; often a mean. And for many purposes this is not only "good enough" but the best we can do. Thus, as we seek to dimension a doorway, we cannot vary its height as each person passes through, but must adopt some one height convenient for almost everyone, but seldom for all, unless monumental considerations determine the answer and not the heights of individuals.

For clothing this doesn't work--and fortunately another solution is feasible, though far from ideal. Here a multidimensional complex of continuously varying measures are reduced to a limited set of "sizes". This same procedure, not always recognized as such, lies back of most if not all applications of analysis of variance. By treating one variate or cluster of variates as a qualitative categorization, a multivariate problem is made univariate within classes. Much, though not all, of the advantage in doing so lies in the resultant simplification of the arithmetic required for an effective analysis of the resulting data. The advent of computers, where available, greatly reduces the attraction from this point of view. In addition, data must frequently be analyzed which does not admit the arithmetic simplification of analysis of variance.

Table 1 sets out the four situations. In line 1 <u>all</u> factors are held constant except the variate of interest. This corresponds to the "vary one factor at a time" prescription formerly regarded as <u>the</u> scientific method, and still appropriate in some circumstances. The second applies wherever all but one factor varies by jumps or can be made to do so--the analysis of variance situation. The contribution of American Experiment Station scientists, in particular, J. Arthur Harris, to the development of the concept of generalized variance, and to the now universally applied calculational algorithm, has dropped from sight. Moreover, the extension of the number of class means above the two treated by Student has opened the Pandora's box of multiple comparisons. My own contribution to this debate is given in (6).

But so soon as regression coefficients are calculated the usual plethora of statistical questions arises. Is each significant? Do they differ? Is there a significant regression of the group means? How is it related to the within group regressions? And so on. It is a curious historical accident that the usual textbook presentation of covariance tend to regard these questions as rarely, if ever of interest and the use of the covariance technique as helpful primarily (1) to place group means on a comparable basis, and (2) to reduce the estimate of residual error. This attitude is in marked contrast to the insistence on the examination of the parallelism of regression in the closely related, but presentationally and computational separate, technique of bioassay. The textbook proclivity to spare the novitiate "unnecessary" complexity by treating only "the most common" cases subjects him, in the next stage, to a seemingly endless succession of unconnected variations in technique neither mutually organized nor suggesting what further developments are likely or would exhaust the possibilities. Yet to do so only requires setting each of the parameters in the last column of the fourth row of Table 1 to (a) known values, (b) known relations or, (3) unrestricted freedom.

The following section presents from first principles, and somewhat more detail than is usual, those aspects of covariance provided for in the computer program which is the principle topic of this paper.

II. Covariance: Covariance analysis is a standard statistical technique widely explained in statistical texts (Snedecor and Cochran,(10), Steel and Torrie, (11)), though admitting elaborations for special circumstances for which reference must be made to the journal literature. An experiment is never, or almost never, conducted in complete ignorance of every aspect of the behavior of the material under study. The use of linear covariance does not assume that the relation between the dependent and independent variate is strictly linear, but only that it is good enough for the purpose at hand. Use of a higher order polynomial or a monotone function of the readings differs only in computational complexity. It is neglected here in the interest of simplicity while covering the most common case. There is also an implied assumption that accidental unobserved factors did not disturb the progress of the study (i.e., that departures from the assumed linear relation are homogeneous in the probability sense). That such departures from the linear relation are normally distributed (perhaps after transformation) is required for tests of significance but this requirement is not normally a stringent one.

Further discussion will be facilitated by reference to Tables 2, 3, 4, and 5. Table 2 gives, purely for illustration, the linear regression of day three weights of mice on their day zero weight in one laboratory of a twelve laboratory collaborative trial. Full details of the experiment, the statistical analysis and the findings are reported elsewhere (7), (8). The mouse numbers are for identification and supply no information. It is conceivable that, were observational data available on individual mice, such information could suggest explanations for the anomalous weight gain behavior of specific mice. The initial weights are shown in Column 2. Column 3 gives each measured three day weight and Column 4 gives what the weight of each mouse would have been had it been exactly what the regression line computes for it. Column 5 shows how large the discrepancy between the observed weight (Column 3) and calculated weight (Column 4) is and Column 6 shows the contribution of that one discrepancy to the uncertainty in every deduction from the data. The sum of Column 6 is 25.3186. From Column 6 it is seen that mouse 2 and mouse 5 each contribute about a third and together two-thirds of the total. Mouse 5 lost weight and mouse 2 gained well above expected.

If we assume that the day three weights are linearly dependent on the initial weights, then the twenty weight observations can be summarized by three quantities, the mean initial weight, the mean final weight, and the slope of the line relating initial and final weight. Of course, the observed mean and slope are only estimates of what would be observed if a very large study were to be carried out. How good an estimate is measured by the deviations in Column 5 of Table 2. Provided all of the usual assumptions in least square analysis are valid, this measure requires only the sum of the deviation squares of Column 6 and their number, 10. Again a study of the adequacy of conformance of the actual assay to the theoretical ideal involves the observed deviations, or residuals, of Column 5. Failure to conform to the model may be due to (a) momentary loss or gain in weight due to crowding, fighting, inequalities in food sharing, errors in weighing, and so on, (b) to real differences in rate of gain by individual mice, (c) to heterogeneity in the supply of mice, (d) to non-normality in the distribution of weight deviations (Column 5, Table 2) and (e) to many other possible factors.

In this experiment weights of each of ten mice were measured before treatment and three days later in each of six treatment classes in each of twelve laboratories. The six treatments were three vaccines in two replications. Table 3 gives the initial, observed three-day, and adjusted three-day mean weight, and the slope for each of the six vaccine-replication combinations for the same laboratory as that of Table 2. It will be noticed that the mean initial and day three weights of Row 1 agree with the means of Table 2. The adjusted means of Table 3 are our estimates of what the day three means for each lot would have been had all of the mice weighed 15.18 grams initially. The slope of the line relating the three-day and zero-day weights is the line giving the best fit to the data of the ten mice within each treatment class. These slopes and means can be viewed as the essential inputs to the covariance analysis of the one laboratory's three-day weights, Table 4.

It is possible that the slope relating a variate of interest, in this case the three-day weight, and an unavoidable nuisance variate, in this case the initial or zero-day weight, may be known, or known so accurately that it is considered legitimate to treat it as known. Doing an analysis of variance on weight gains is equivalent to doing a covariance analysis with a slope constrained to be unity. In other applications a different scaling factor might be more appropriate. The appropriateness of using a known regression coefficient and, if so, its specification is not discussed in Table 4. It is provided for in the covariance program. However, the complete covariance analysis over the twelve laboratories yielded a regression coefficient indistinguishable from unity.

There is a striking variation in the observed slopes for the six treatment groups for this laboratory (Table 3). The mean of these slopes, 0.523, is only half the value, unity, assumed when analysis is carried out in terms of weight gains. The reality of this variation in slopes is examined in Tables 4 and 5.

Further, the data of an experiment where covariance is possible can always be analyzed to yield three different types of slope estimates. Table 4 supplies the information needed to check the adequacy of each of these methods of adjusting three-day weights to allow for variation in initial weights. Depending upon which procedure is chosen, a different estimate of residual variation is obtained. These are also supplied in Table 4. First, all of the data can be entered into a single overall regression, yielding a common regression estimate of the slope. This would be appropriate if the treatment classes, in our case the six vaccine-replication combinations, were without effect on either the means or the slopes (section 14.7, Snedecor and Cochran, 1967). Second, the data of each treatment class on its own can be used to supply a separate slope estimate varying for each class (Column 6, Table 3). A compromise, and the method normally meant when the technique of analysis of covariance is used, is to assume that the various treatments might well affect the mean level of the response in each class--it is usually the object of the experiment to study that question--but to assume (subject to test) that the treatments (in this study, vaccines) will not appreciably affect the slope; which consequently will remain constant over the classes. For each of these three alternatives a different set of adjusted means, supposedly removing the effect of variation in the nuisance variate, initial weight, will result. Of course, a fourth alternative is simply to neglect the covariate (or to show that it is without effect).

It is normally assumed in a covariance analysis that the residual mean squares within classes vary no more than would be expected by chance. Hence the residual mean squares of the six vaccine-replication classes, each with eight degrees of freedom, are pooled in line 1 of Table 4 to give a single estimate of 1.4817 based on 48 degrees of freedom. The legitimacy of this step will be discussed later. Six of the lines of Table 4 are concerned with slopes (lines 2, 3, 4, 6, 7, 8, and 9). Line 5 is the only line of the table not yet listed. It is this line that answers the question: Was there a mean effect on mouse weight gain by either the vaccines or the repetition of the trial?

Even if the adjusted weights are significantly different, the slopes for each different vaccine in each of the two replications may or may not be the same. This is tested in line 2 of Table 4. The non-significance in Column 7 implies that a given change in three-day weight for a given change in initial weight is the same in all three vaccines and in both replications. The mean square of Column 5, line 3, 1.5403, hence is an estimate of the same quantity that the 1.4817 of line 1 is. Line 4 of the table tests whether there is a regression line with non-zero slope relating initial and three-day weight after allowing for a possible variation in the subclass means; the fourth possibility listed above. That the relation tests non-significant (Column 7, line 4) shows that the variability is high compared to the range of initial weights so that any such variation does not clearly show a dependence between three and zero-day weights. Of course, there was in fact little variation in the initial weights themselves. Since the adjusted weights (line 5) also test non-significantly different, a single regression encompassing all 60 readings seems justified. The residual mean squares from this regression is given in Column 5, line 6, and the reality of the non-zero slope of the regression in line 7. Again, such a line is not clearly established.

Lines 8 and 9 have a different meaning. Each of the six treatment classes (three vaccines in two replications) yields a mean initial weight, $\bar{x}$ and a mean final weight, $\bar{y}$. The six such pairs of readings could fall on a line, which might or might not have the same slope as the average within-class lines. Whether such a line meets the standard test of significance is tested in line 9. Whether the scatter of the six class means depart from that line more than can be accounted for by chance is tested in line 8 of Table 4. Neither tests significant in this one laboratory. If line 9 were significant and line 8 not, a presumption would be shown that the adjusted treatment means differed only because the average initial weight of the mice assigned to each class did so, and hence that the nature of vaccine treatment was ineffective. For this conclusion to be appealing, the between-class slope would have to agree with the within-class slope. This topic is discussed more fully in Section 14.7 of Snedecor and Cochran (10).

A clearer insight into the meaning of the lines of Table 4 can be obtained by reference to Table 5. Line 1 of Table 5 gives the pooled corrected sum of squares and products in all six treatment classes after adjusting for possible treatment effects on the mean values of the initial and three day weights (i.e., assuming that the dependence (regression) of three day weight on initial weight is not affected by the treatments even though the three day expected weights themselves might well be). Notice that the corrected sum of squares for three day weights, y^2=87.325, is the sum of the line 3 and line 4 sums of squares of Table 4, repeated in Columns 6 and 7 of Table 5. The 81.634 then is the variability left over after fitting the best possible common slope line through the mean values of the variates in each of the six classes whereas the 5.691 is the benefit obtained by doing so. Next, assume for the moment that the weights of all mice within each vaccine-replication class were identical. Then we get the same result if we calculate a regression on the six group means as if we use all the data in a single regression. The necessary corrected sums of squares and products are shown in line 2 of Table 5. Again the regression sum of squares and the residual sum of squares agree with the entries in Column 4 of lines 8 and 9 of Table 4. But, of course, the initial and final weights for every mouse within each vaccine-replication class are not identical. Ignoring that variation and using all the data in a single overall regression yields the sums of squares and products of the third and last line of Table 5. It is easy to verify that the entries in the first four numerical columns of the last line of Table 5 are the sums of the two entries just above each.

In effect then, the first four numerical columns of Table 5 divide the 59 degrees of freedom for a single regression of y on x into 54 for a pooled slope regression allowing for a different mean for $\bar{x}$ and $\bar{y}$ from subclass to subclass and 5 degrees of freedom for a regression of the $\bar{y}$'s on the $\bar{x}$'s. Now, examine the columns of Table 5. The three entries in Columns 6 and 7, Table 5, each appear in Table 4. The entries in Column

5, Table 5, are the sums of those in Columns 6 and 7 on the same line. Hence, the corrected sum of squares for y in each line is divided into two parts, one in Column 7 representing failure of a line through the overall mean (lines 2 and 3) or a line through each subclass mean but with a common slope (line 1) to account for all of the variation in y. The entries in Column 6 are the amounts of sums of squares which are accounted for in each case. The slopes calculated on the basis of Columns 3 and 4 are shown in Column 8. If treatment class were without effect, and if there were no chance variation, the slope calculated from each of the three lines of Table 5 would agree with those of the other two. The observed within-class pooled slope is 0.625 and the slope of the regression of $\bar{y}$ on $\bar{x}$ is 0.194. Of course neither is close to unity, required for use of mean gains. Moreover, they do not test significantly different from each other.

A distinction should be noted between the five degrees of freedom in line 2 of Table 5 and the five degrees of freedom in line 5 of Table 4. The former dissects the observed treatment class sum of squares for final weight 17.087 into two parts, one, 0.353, measuring how well the six class means for final weight fall on a straight line and the other, 16.734, how much they fail to do so. The entry in line 5 of Table 4, 17.8023, measures the overall agreement of the adjusted class mean final weight, that is, after making allowance for variations in initial weights within the one vaccine-laboratory-replication grouping.

The detailed exposition of the covariance analysis above is only to make the procedure clear. Such a clear understanding is required to exploit the output of the covariance program which is the principal subject of this paper.

Table 1

Types of Statistical Models

Case	Mean	Deviation	Sum of Squares
Univariate	$\bar{x} = (\Sigma x_i)/n$	$d_i = x_i - \bar{x}$	$\Sigma x_i^2 - n\bar{x}^2$
Analysis of Variance	$\bar{x}_{i.} = (\Sigma x_{ij})/n_i$	$d_{ij} = x_{ij} - \bar{x}_{i.}$	$\Sigma\Sigma(x_{ij}-\bar{x}_{i.})^2$
Regression	$\bar{y} = (\Sigma y_i)/n$ $\bar{x} = (\Sigma x_i)/n$	$d_i = y_i - \hat{y}$	$\Sigma(y_i-a-bx_i)^2$
Analysis of Covariance	$\bar{y}_{i.} = (\Sigma y_{ij})/n_i$ $\bar{x}_{i.} = (\Sigma x_{ij})/n_i$	$d_{ij} = y_{ij} - \hat{y}_{i.}$	$\Sigma\Sigma(y_{ij}-a_i-bx_{ij})^2$

Notes:

1. Analysis of Variance involves K>1 classes.
2. Regression involves regressor variable.
3. Covariance involves both.
4. For further discussion see text.

TABLE 2

Least Square Fit
Day Three on Initial Weight
One Vaccine in One Laboratory

Mouse	Weight Initial	Weight Day 3	Expected Day 3 Weight	Deviation (3)-(4)	Deviation Squared
1	15	18	16.97	1.0336	1.0682
2	14.5	20	17.04	2.9597	8.7600
3	16	15	16.82	-1.8188	3.3080
4	16	17	16.82	0.1812	0.0328
5	14.5	14	17.04	-3.0403	9.2432
6	15	16	16.97	-0.9664	0.9340
7	15.5	17	16.89	0.1074	0.0115
8	16	16.5	16.82	-0.3188	0.1016
9	16	18	16.82	1.1812	1.3953
10	16	17.5	16.82	0.6812	0.4640
Mean	15.45	16.9	16.9	0.0000	25.3186*

* Sum

TABLE 3

Group Means for One Laboratory
Day Three Weights (grams)

Lot	Rep	Mean Weights Initial	Mean Weights Final	Mean Weights Adjusted	Slope
3	1	15.45	16.9	16.73	-0.148
7A	1	15.85	16.8	16.38	2.762
Saline	1	15.30	17.6	17.53	0.578
3	2	14.80	16.8	17.04	-0.364
7A	2	15.00	15.8	15.91	2.000
Saline	2	14.70	17.1	17.35	0.406
Mean		15.18	16.82	16.82	0.523

TABLE 4

Analysis of Covariance*
Day Three Weights for One Laboratory

Line	Effect	df	Sum Squares	Mean Square	F	
1	Within	48	71.1231	1.4817		
2	Bet. Coeff.	5	10.5105	2.1021	1.4187	
3	CM Sl Res SS	53	81.6336	1.5403		
4	CM Sl Red SS	1	5.6914	5.6914	3.8411	
5	Adj. Means	5	17.8023	3.5605	2.4029	NS
6	CM Rg Res SS	58	99.4359	1.7144		
7	CM Rg Red SS	1	4.9766	4.9766	3.3586	
8	Dev Bet Cl Sl	4	16.7344	4.1836	2.8235	
9	Bet Class Sl	1	.3531	.3531	.2383	
10	Total	59	104.4123	1.7697		

* For explanation, see text.

TABLE 5

Regression Variability*
Day Three and Initial Weights for One Laboratory

Slope	df	x^2	xy	y^2	Reg	Dev	Slope
Pooled	54	14.55	9.100	87.325	5.691	81.634	0.625
Between	5	9.43	1.825	17.087	0.353	16.734	0.194
Overall	59	23.98	10.925	104.412	4.976	99.436	0.455

* For explanation, see text.

Bibliography

1. Bargmann, Rolf E. [1969] Principles and design of statistical computer languages. Chapter 8 of Progress in Operation Research, Vol. III. Julius S. Aronofsky, editor. Wiley Publications No. 16.

2. Blakers, Albert L. [1967] Mathematical concepts of elementary measurement. SMSG. Studies in Mathematics. Vol XVII, Stanford Univ.

3. Crary, F. D. and Ladner, T. D. [1970] A simple method of adding a new data type to Fortran. MRC Tech. Summary Report No. 1065.

4. DeLury, D. B. [1948] The analysis of covariance. Biometrics, Vol. 4, No. 3, pp 153-170.

5. Helmholtz, Herman [1930] Counting and measuring. Translated by C. L. Bryan. Van Nostrand, Princeton, N. J.

6. Maloney, C. J. [1971] Multiple comparisons revisited. Biometrics Society Meetings. (ENAR-WNAR). Fort Collins, Colorado. (Available from the author).

7. Manclark, C. R., Dunnett, C. W., Maloney, C. J., Cooper, M. S., and Pittman, M. [1975] A collaborative study on laboratory assays of Pertussis vaccine. (In review).

8. Manclark, C. R., Maloney, C. J., Dunnett, C. W., and Cooper, M. S. [1975] Nonreproducibility of the mouse weight gain test for Pertussis vaccine toxicity. (In review).

9. Nelson, Wayne B., Phillips, M. C. and Thumhart, L. [1973] How to design computer programs to serve users--lessons learned from Statpac, a general program for data analysis. General Electric Report No. 73CRD083, Technical Information Series Class 1.

10. Snedecor, George and Cochran, William [1967] Statistical Methods. Iowa State University Press, Iowa.

11. Steel, Robert and Torrie, James [1960] Principles and Procedures of Statistics. McGraw Hill.

12. Whole issue [1957] Biometrics, Vol. 13, No. 3.

13. Yohe, J. M. [1974] An overview of programming practices. MRC Tech. Summary Report No. 1440. To appear in Computing Surveys.

<u>Note</u>: Recent literature on covariance is extensive but scattered. I know of no comprehensive treatment from a user's point of view.

BASIC PROGRAM DOCUMENTATION

The REM (remark) statements are located at the end of this program. If the number of items in each data set is fixed, the symbol M1, line 55 must be set to that count. If the number of entries per set varies then 9E9 is used to separate each set. If the number of items in every entry (or data set) are the same then 9E9 should be omitted. Missing data are listed as 8E8. If data is read as 8E8 then the program skips over the entry and omits it from the count and degrees of freedom. The symbol 7E9 in output indicates that that quantity was not calculated. Program data entry should begin in line 2235 (this deletes program example data).

Select a code number from the four tables at the end of the program. The (+) sign is for "yes" and the (-) sign is for "no". The following list will assist you in using the tables:

	SYMBOL	CODE RANGE	LINE
1. How many input data sets are there?	K		50
2. How many items are there in each group (if fixed)?	M1		55
3. How will the data be entered? P8=1 Converts to logs(see S2,6)	S2	1-8	1900
4. Is the transformed or untransformed input data to be printed?	C5	1-6	2005
5. (a) Will the printed output include the input data? (b) Will the entire output be printed, or only certain parts?	B5	1-6	2080
6. Which slope will form the adjusted mean?	L5	1-4	2145

NOTES.

1. Two excellent mutually clarifying texts are: Snedecor, G. W. and Cochran, W. G., "Statistical Methods", 6th edition 1967, Chapter 14, and Steele, R. G. D. and Torrie, J. H., "Principles and Procedures of Statistics", 1960, Chapter 15.
2. The program computes (virtually) all of the procedures involving one criterion of classification and one regression variable.
3. In addition, the program permits a choice of regression coefficient for adjusted means including a prior assignment and a test of significance for the latter.
4. The program also permits a wide choice of (a) source of data, (b) transformation of variates and (c) print out format.
5. If the classification of the data sets is without effect on either variate the common regression is the appropriate one. If not, the references or other authorities should be consulted.
6. While the program has been carefully checked, absolute accuracy and provision for all contingencies cannot be guaranteed.
7. Suggestions for corrections of program or documentation errors or infelicities will be appreciated.

C O V A R I A N C E
BASIC PROGRAM
SYMBOL INDEX

SYMBOL	MEANING	*LINE NUMBER WHERE FIRST DEFINED OR READ	USED OR PRINTED
A(J)	Uncorrected sum of X squared for Jth data set.	580	935
A	Total uncorrected sum of X squared.	935	970
A$(I)	Program headings.	75	80
B(J)	Uncorrected sum of products.	585	655
B	Total uncorrected sum of products.	940	975
B5	Code for selecting printouts.	50	140
C(J)	Uncorrected sum of Y squared.	590	660
C	Total uncorrected sum of Y squared.	945	980
C1	Effective error mean square for regression.	1530	1805
C2	Centering constant for program headings.	80	85
C5	Code selecting data to be printed.	50	145
C6	Centering constant for data set headings.	320	325
D(J)	Sum of X for Jth data set.	595	645
D	Overall (total) sum of X.	950	970
D1	Effective error mean square for difference.	1535	1805
D2	Prior slope (given).	65	1040
E(J)	Sum of Y for Jth data set.	600	640
E	Overall (total) sum of Y.	955	975
E1	Reduction sum of squares for between classes regression.	1555	1560

*Symbol occurrences in quotes are disregarded.

SYMBOL INDEX continued.

SYMBOL	MEANING	LINE NUMBER WHERE FIRST DEFINED OR READ	USED OR PRINTED
F(J)	Covariance degrees of freedom.	1390	1420
F	Overall (total) within degrees of freedom.	1440	1460
G(J)	Sum of deviations squared for Jth data set.	1395	1420
H(J)	D.F. for corrected sum of squares and products.	665	1210
H1	Sum of D.F. for corrected sum of squares and products.	685	1255
H9	Between coefficients degrees of freedom.	920	1580
H$(I)	Data set headings.	120	325
I(I)	Transformation symbol for Y.	375	770
J(I)	Transformation symbol for X.	425	430
K(I)	Expected value of Y.	705	710
K	Number of data sets.	50	115
L(I)	Individual deviation squares.	715	735
L	Overall (pooled) sum of squares within lines.	1445	1460
L1	Effective error mean square for Y only.	1540	1750
L2	Treatment mean square for Y only.	1545	1750
L5	Code for slope in adjusted mean calculation.	50	935
M(J)	Deviation mean square.	1420	1460
M	Residual sum of squares for common slope.	1510	1595

SYMBOL INDEX continued.

SYMBOL	MEANING	LINE NUMBER WHERE FIRST DEFINED OR READ	LINE NUMBER WHERE FIRST USED OR PRINTED
M1	Number of items in equal data sets.	55	355
N(J)	Actual count (8E8s not counted) for Jth data set.	350	510
N	Overall (total) count.	925	960
O	Sum of deviation squared for common regression residual sum of squares.	1515	1640
P(J)	Corrected sum of X squared for Jth data set.	650	670
P	Overall corrected sum of X squared.	690	915
P7	Number of lines of program heading.	50	70
P8	Code for converting X to logs.	50	415
Q(J)	Corrected within sum of XY for Jth data set.	655	675
Q	Overall corrected within sum of XY.	695	915
Q5	Slope for adjusted means (see code L5).	990	1020
R(J)	Corr. within data set sum of Y sqd.	660	900
R	Pooled sum of corrected Y squares.	900	1255
R1	Sum of deviation squared from between classes regression.	1560	1580
R2	Mean square for deviation from between classes regression.	1580	1670
S(J)	Within class slope for Jth data set.	685	990
S	Reduction sum of squares for common slope.	1505	1510
S1	Value for prior slope (see Code L5).	60	1020
S2	Code for input data sources and transformations.	50	340

SYMBOL INDEX continued.

SYMBOL	MEANING	LINE NUMBER WHERE FIRST DEFINED OR READ	LINE NUMBER WHERE FIRST USED OR PRINTED
T$	Type of study.	50	150
U(J)	Mean of X for Jth data set.	645	705
U	Corr. total sum of X squared.	970	1290
U8	Overall mean square for treatment for X.	1525	1530
U9	Overall mean square for residual for X.	1520	1725
V(I)	Deviation of Y from expected value for Jth data set.	710	735
V	Corrected total sum of XY.	975	1010
W(J)	Mean of Y for Jth data set.	640	705
W	Corrected total sum of Y squares.	980	1290
X(I)	Ith item of data for X.	365	430
X$(I)	Designation for Ith element of data.	615	735
Y(I)	Ith item of data for Y.	365	575
Z(J)	Adjusted mean of Y for Jth data set.	1030	1065
7E9	Quantity not calculated.	685	1855
8E8	Missing data.	470	1850
9E9	End of one treatment data set.	475	1845

OK. DATE FILED: 07/31/74.

LIST 5,245

```
5 DIM A(35),B(35),C(35),D(35),E(35),F(35),G(35),H(35)
10 DIM I(35),J(35),K(35),L(35),M(35),N(35),O(35),P(35)
15 DIM Q(35),R(35),S(35),T(35),U(35),V(35),W(35),X(35)
20 DIM Y(35),Z(35),H$(10),X$(35)
25 IREAD ON
30 OPEN 1,"CHRNX"
35 OPEN 2,"CHRNY"
40 OPEN 3,"HEDCHRN"
45 DATA 2,3,0,6,1,1,2
50 READ P7,K,P8,S2,C5,B5,L5,T$
55 LET M1=1000
60 LET S1=1
65 LET D2=1
70 FOR I=1 TO P7
75 READ A$(I)
80 C2=GTL(A$(I))
85 PRINTTABINT(36-C2/2);A$(I)
90 NEXT I
95 DATA"TRIAL"
100 DATA"SPECIAL TEST FROM COCHRAN:CLIFFORD J. MALONEY,PH.D.,STAT."
105 DATA"COCHRAN EXAMPLE"
110 DATA"DRUG A","DRUG B","DRUG F"
115 FOR I=1 TO K
120 READ H$(I)
125 NEXT I
130 PRINT
135 PRINT
140 IF B5>4 THEN 305
145 ON C5 GO TO 150,170,190,210,230,250,270,290
150 PRINTTAB4;T$;TAB16;"X(I)";TAB26;"Y(I)";TAB39;"DEVIATION";TAB59;
155 PRINT"DEV.SQD."
160 PRINTTAB25;"UNTRANS.X(I) AND Y(I)"
165 GO TO 305
170 PRINTTAB4;T$;TAB16;"X(I)";TAB26;"I(I)";TAB39;"DEVIATION";TAB59;
175 PRINT"DEV.SQD."
180 PRINTTAB14;"UNTRANSFORMED X(I),I(I).  (Y(I)=I(I) SQD. IN PROGRAM)"
185 GO TO 305
190 PRINTTAB4;T$;TAB16;"X(I)";TAB26;"Y(I)";TAB39;"DEVIATION";TAB59;
195 PRINT"DEV.SQD."
200 PRINTTAB17;"UNTRNS X(I),TRNSD Y(I)  (Y(I)=I(I) SQD.)"
205 GO TO 305
210 PRINTTAB4;T$;TAB16;"J(I)";TAB26;"Y(I)";TAB39;"DEVIATION";TAB59;
215 PRINT"DEV.SQD."
220 PRINTTAB14;"UNTRANSFORMED J(I),Y(I).  (X(I)=LOG J(I) IN PROG.)"
225 GO TO 305
230 PRINTTAB4;T$;TAB16;"X(I)";TAB26;"Y(I)";TAB39;"DEVIATION";TAB59;
235 PRINT"DEV.SQD."
240 PRINTTAB20;"TRANSFORMED X(I).  (X(I)=J(I) IN LOGS)"
245 GO TO 305
```

```
LIST 250,515

250 PRINTTAB4;T$;TAB16;"X(I)";TAB26,"Y(I)";TAB39;"DEVIATION";TAB59;
255 PRINT"DEV.SQD."
260 PRINTTAB24;"TRANSFORMED X(I) AND Y(I)"
265 GO TO 305
270 PRINTTAB4;T$;TAB16;"J(I)";TAB26;"I(I)";TAB39;"DEVIATION";TAB59;
275 PRINT"DEV.SQD."
280 PRINTTAB10;"UNTRANS.J(I),I(I). (X(I),Y(I)=LOG OF J(I),I(I) IN PROG)"
285 GO TO 305
290 PRINTTAB4;T$;TAB16;"X(I)";TAB26,"Y(I)";TAB39;"DEVIATION";TAB59;
295 PRINT"DEV.SQD."
300 PRINTTAB10;"TRANS. J(I) TO X(I), I(I) TO Y(I). (CONV. TO LOGS)"
305 FOR J=1 TO K
310 IF B5>4 THEN 340
315 PRINT
320 C8=GTL(H$(J))
325 PRINTTABINT(36-C8/2);H$(J)
330 PRINT
335 PRINT
340 ON S2 GO TO 350,350,345,345,350,350,345,350
345 REWIND 1
350 LET N(J)=0
355 FOR I=1 TO M1
360 ON S2 GO TO 365,375,385,400,400,415,440,455
365 READ(1,0) X(I),Y(I)
370 GO TO 500
375 READ(1,0) X(I),I(I)
380 GO TO 470
385 READ(1,0) X(I)
390 READ I(I)
395 GO TO 470
400 READ(1,0) X(I)
405 READ Y(I)
410 GO TO 500
415 IF P8=1 THEN 425
420 GO TO 440
425 READ(1,0) J(I)
430 LET X(I)=LOG(J(I))/LOG(10)
435 GO TO 445
440 READ(1,0) X(I)
445 READ(2,0) Y(I)
450 GO TO 500
455 READ(1,0) J(I),I(I)
460 LET X(I)=LOG(J(I))/LOG(10)
465 LET Y(I)=LOG(I(I))/LOG(10)
470 IF I(I)=8E8 THEN 515
475 IF I(I)=9E9 THEN 520
480 GO TO 510
485 IF J(I)=8E8 THEN 515
490 IF J(I)=9E9 THEN 520
495 GO TO 510
500 IF Y(I)=8E8 THEN 515
505 IF Y(I)=9E9 THEN 520
510 LET N(J)=N(J)+1
515 NEXT I
```

LIST 520,805

```
520 LET A(J),B(J),C(J),D(J),E(J)=0
525 IF S2=7 THEN 535
530 GO TO 540
535 REWIND 3
540 FOR I=1 TO M1
545 ON S2 GO TO 550,565,565,550,550,550,550,565
550 IF Y(I)=8E8 THEN 635
555 IF Y(I)=9E9 THEN 640
560 GO TO 580
565 IF I(I)=8E8 THEN 635
570 IF I(I)=9E9 THEN 640
575 LET Y(I)=I(I)^2
580 LET A(J)=A(J)+X(I)^2
585 LET B(J)=B(J)+X(I)*Y(I)
590 LET C(J)=C(J)+Y(I)^2
595 LET D(J)=D(J)+X(I)
600 LET E(J)=E(J)+Y(I)
605 IF S2=6 THEN 630
610 IF S2=7 THEN 630
615 READ(2,0) X$(I)
620 IF S2=7 THEN 630
625 GO TO 635
630 READ(3,0) X$(I)
635 NEXT I
640 LET W(J)=E(J)/N(J)
645 LET U(J)=D(J)/N(J)
650 LET P(J)=A(J)-(D(J)^2)/N(J)
655 LET Q(J)=B(J)-(D(J)*E(J))/N(J)
660 LET R(J)=C(J)-(E(J)^2)/N(J)
665 LET H(J)=N(J)-1
670 IF P(J)<0.0001 THEN 685
675 LET S(J)=Q(J)/P(J)
680 GO TO 690
685 LET S(J)=7E9
690 IF B5>4 THEN 830
695 PRINTTAB32;"COUNT= ";N(J)
700 FOR I=1 TO M1
705 LET K(I)=W(J)+S(J)*(X(I)-U(J))
710 LET V(I)=Y(I)-K(I)
715 LET L(I)=(Y(I)-K(I))^2
720 ON C5 GO TO 745,770,790,725,745,755,810,755
725 IF Y(I)=8E8 THEN 825
730 IF Y(I)=9E9 THEN 830
735 PRINT USING 760, X$(I),J(I),Y(I),V(I),L(I)
740 GO TO 825
745 IF Y(I)=8E8 THEN 825
750 IF Y(I)=9E9 THEN 830
755 PRINT USING 760, X$(I),X(I),Y(I),V(I),L(I)
760 FIELD(4X,A5,4X,F6.3,4X,F6.3,3X,F16.8,3X,F16.8)
765 GO TO 825
770 IF I(I)=8E8 THEN 825
775 IF I(I)=9E9 THEN 830
780 PRINT USING 760, X$(I),X(I),I(I),V(I),L(I)
785 GO TO 825
790 IF I(I)=8E8 THEN 825
795 IF I(I)=9E9 THEN 830
800 PRINT USING 760, X$(I),X(I),I(I)^2,V(I),L(I)
805 GO TO 825
```

```
LIST 810,1075

810 IF J(I)=8E8 THEN 825
815 IF J(I)=9E9 THEN 830
820 PRINT USING 760, X$(I),J(I),I(I),V(I),L(I)
825 NEXT I
830 NEXT J
835 ON B5 GO TO 840,840,875,875.840,840,875,875
840 PRINT
845 PRINT
850 PRINT
855 PRINTTAB29;"SUMS AND MEANS"
860 PRINT
865 PRINTTAB2;"X-SUM";TAB17;"X-MEAN";TAB31;"Y-SUM";TAB48;
870 PRINT"MEAN";TAB61;"ADJUST.MEAN"
875 LET H1,H9,P,Q,R=0
880 FOR J=1 TO K
885 LET H1=H1+H(J)
890 LET P=P+P(J)
895 LET Q=Q+Q(J)
900 LET R=R+R(J)
905 NEXT J
910 PRINT
915 LET K9=(S1-Q/P)^2*P
920 LET H9=K-1
925 LET A,B,C,D,E,N=0
930 FOR J=1 TO K
935 LET A=A+A(J)
940 LET B=B+B(J)
945 LET C=C+C(J)
950 LET D=D+D(J)
955 LET E=E+E(J)
960 LET N=N+N(J)
965 NEXT J
970 LET U=A-(D^2)/N
975 LET V=B-(D*E)/N
980 LET W=C-(E^2)/N
985 ON L5 GO TO 990,1000,1010,1020
990 LET Q5=S(J)
995 GO TO 1025
1000 LET Q5=Q/P
1005 GO TO 1025
1010 LET Q5=V/U
1015 GO TO 1025
1020 LET Q5=S1
1025 FOR J=1 TO K
1030 LET Z(J)=E(J)/N(J)-Q5*(D(J)-D/N)
1035 NEXT J
1040 IF D2=0 THEN 1050
1045 GO TO 1055
1050 LET D2=E/D
1055 ON B5 GO TO 1060,1060,1125,1125,1060,1060,1125,1125
1060 FOR J=1 TO K
1065 PRINT USING 1070,D(J),U(J),E(J),W(J),Z(J)
1070 FIELD(1H ,F7.2,2X,F16.8,3X,F8.3,1X,2(1X,F16.8))
1075 NEXT J
```

```
LIST 1080,1320


1080 PRINT
1085 PRINT
1090 PRINT
1095 PRINTTAB26;"TOTAL SUMS AND MEANS"
1100 PRINT
1105 PRINTTAB5;"SUM-X";TAB22;"MEAN-X";TAB37;"SUM-Y";TAB53;"MEAN-Y";
1110 PRINTTAB63;"COUNT"
1115 PRINT USING 1120,D,D/N,E,E/N,N
1120 FIELD(1H ,4X,F6.2,1X,3(1X,F15.4),4X,F3.)
1125 ON B5 GO TO 1130,1225,1225,1225,1130,1225,1225,1225
1130 PRINT
1135 PRINTTAB17;"UNCORRECTED SUMS OF SQUARES AND PRODUCTS"
1140 PRINT
1145 PRINTTAB1;"SOURCE";TAB19;"SUM X-SQ";TAB35;"SUM PROD.";TAB52;
1150 PRINT"SUM Y-SQ.";TAB63;"COUNT"
1155 PRINT
1160 FOR J=1 TO K
1165 PRINT USING 1170,H$(J),A(J),B(J),C(J),N(J)
1170 FIELD(A8,3X,3(1X,F15.4),4X,F3.)
1175 NEXT J
1180 PRINT
1185 PRINTTAB18;"CORRECTED SUMS OF SQUARES AND PRODUCTS"
1190 PRINT
1195 PRINTTAB1;"D.F.";TAB11;"SUM X-SQ.";TAB29;"SUM XY";TAB45;
1200 PRINT"SUM Y-SQ.";TAB65;"SLOPE"
1205 FOR J=1 TO K
1210 PRINT USING 1215,H(J),P(J),Q(J),R(J),S(J)
1215 FIELD(1X,F3.,4(1X,F16.8))
1220 NEXT J
1225 ON B5 GO TO 1230,1345,1230,1345,1230,1345,1230,1345
1230 PRINT
1235 PRINTTAB21;"********        COMMON        ********"
1240 PRINT
1245 PRINTTAB1;"D.F.";TAB11;"SUM X-SQ.";TAB29;"SUM XY";TAB45;
1250 PRINT"SUM Y-SQ.";TAB65;"SLOPE"
1255 PRINT USING 1215,H1,P,Q,R,Q/P
1260 PRINT
1265 PRINT TAB19;"******** BETWEEN CLASSES ********"
1270 PRINT
1275 PRINTTAB1;"D.F.";TAB11;"SUM X-SQ.";TAB29;"SUM XY";TAB45;
1280 PRINT"SUM Y-SQ.";TAB65;"SLOPE"
1285 IF U-P<0.0001 THEN 1300
1290 PRINT USING 1215,(N-1)-H1,U-P,V-Q,W-R,(V-Q)/(U-P)
1295 GO TO 1310
1300 PRINT USING 1305,(N-1)-H1,U-P,V-Q,W-R
1305 FIELD(1X,F3.,3(1X,F16.8))
1310 PRINT
1315 PRINTTAB21;"********        TOTAL        ********"
1320 PRINT
```

```
LIST 1325,1585

1325 PRINTTAB1;"D.F.";TAB11;"SUM X-SQ.";TAB29;"SUM XY";TAB45;
1330 PRINT"SUM Y-SQ.";TAB65;"SLOPE"
1335 PRINT USING 1215,N-1,U,V,W,V/U
1340 PRINT
1345 PRINTTAB30;"*************"
1350 PRINT
1355 PRINTTAB23;"DEVIATIONS FROM REGRESSION"
1360 PRINT
1365 PRINT"EXPERIMENTS";TAB17;"D.F.";TAB26;"SUM DEV.SQ.";TAB44;
1370 PRINT"MEAN SQUARE";TAB62;"....F...."
1375 PRINT
1380 FOR J=1 TO K
1385 IF P(J)<0.0001 THEN 1405
1390 LET F(J)=H(J)-1
1395 LET G(J)=C(J)-E(J)^2/N(J)-Q(J)^2/P(J)
1400 GO TO 1420
1405 LET G(J)=C(J)-E(J)^2/N(J)
1410 LET F(J)=H(J)
1415 LET H9=H9-1
1420 LET M(J)=G(J)/F(J)
1425 NEXT J
1430 LET F,L=0
1435 FOR J=1 TO K
1440 LET F=F+F(J)
1445 LET L=L+G(J)
1450 NEXT J
1455 FOR J=1 TO K
1460 PRINT USING 1465,H$(J),F(J),G(J),M(J),M(J)*F/L
1465 FIELD(A8,8X,F3.,3(1X,F16.8))
1470 NEXT J
1475 PRINTTAB15;"-------------------------------------------------"
1480 PRINT
1485 PRINTTAB25;"ANALYSIS OF COVARIANCE"
1490 PRINT
1495 PRINT USING 1500,F,L,L/F
1500 FIELD("WITHIN          ",3X,F3.,2(1X,F16.8))
1505 LET S=Q^2/P
1510 LET M=R-S
1515 LET O=W-V^2/U
1520 LET U9=P/H1
1525 LET U8=(U-P)/((N-1)-H1)
1530 LET C1=L/F*(1+U8/P)
1535 LET D1=(R-2*D2*Q+P*D2^2)/H1
1540 LET L1=R/H1
1545 LET L2=(W-R)/((N-1)-H1)
1550 IF U-P<0.0001 THEN 1570
1555 LET E1=(V-Q)^2/(U-P)
1560 LET R1=W-R-E1    #SUBTRACTS POOLED WITHIN COR.S.SQS.FR COR.TOT.S.S.
1565 GO TO 1575
1570 LET R1=W-R
1575 IF H9<2 THEN 1590
1580 LET R2=R1/(H9-1)
1585 GO TO 1595
```

```
LIST 1590,1855

1590 LET R2=R1
1595 PRINT USING 1600,H9,M-L,(M-L)/(H9),(M-L)*F/(H9*L)
1600 FIELD("BET.COEFF.    ",3X,F3.,3(1X,F16.8))
1605 IF L5<4 THEN 1620
1610 PRINT USING 1615,1,K9,K9,K9*F/L
1615 FIELD("PRIOR SLOPE   ",3X,F3.,3(1X,F16.8))
1620 PRINT USING 1625,N-K-1,M,M/(N-K-1),M*F/((N-K-1)*L)
1625 FIELD("CM SL RES SS ",3X,F3.,3(1X,F16.8))
1630 PRINT USING 1635,1,S,S,S*F/L
1635 FIELD("CM SL RED SS ",3X,F3.,3(1X,F16.8))
1640 PRINT USING 1645,K-1,O-M,(O-M)/(K-1),(O-M)*F/((K-1)*L)
1645 FIELD("ADJ.MEANS     ",3X,F3.,3(1X,F16.8))
1650 PRINT USING 1655,N-2,O,O/(N-2),(O*F)/((N-2)*L)
1655 FIELD("CM RG RES SS ",3X,F3.,3(1X,F16.8))
1660 PRINT USING 1665,1,W-O,W-O,((W-O)*F)/L
1665 FIELD("CM RG RED SS ",3X,F3.,3(1X,F16.8))
1670 PRINT USING 1675,H9-1,R1,R2,R2/(L/F)
1675 FIELD("DEV BET CL SL",3X,F3.,3(1X,F16.8))
1680 PRINT USING 1685,1,E1,E1,E1/(L/F)
1685 FIELD("BET CLASS SL ",3X,F3.,3(1X,F16.8))
1690 PRINTTAB15;"-----------------------------------------------"
1695 ON B5 GO TO 1700,1700,1700,1815,1700,1815,1700,1700
1700 PRINT
1705 PRINTTAB19;"SEPARATE VARIATE ANALYSES OF VARIANCE"
1710 PRINT
1715 PRINTTAB4;"SOURCE";TAB17;"D.F.";TAB26;"SUM SQUARE";TAB43;
1720 PRINT"MEAN SQUARE";TAB62;"....F...."
1725 PRINT USING 1730,(N-1)-H1,U-P,U8,U8/U9
1730 FIELD("TREATMENT(X) ",3X,F3.,3(1X,F16.8))
1735 PRINT USING 1740,H1,P,U9
1740 FIELD("RESIDUAL (X) ",3X,F3.,2(1X,F16.8))
1745 PRINT
1750 PRINT USING 1755,(N-1)-H1,W-R,L2,L2/L1
1755 FIELD("TREATMENT(Y) ",3X,F3.,3(1X,F16.8))
1760 PRINT USING 1765,H1,R,R/H1
1765 FIELD("RESIDUAL (Y) ",3X,F3.,2(1X,F16.8))
1770 PRINTTAB15;"-----------------------------------------------"
1775 PRINT
1780 PRINTTAB21;"ALTERNATIVE DESIGN COMPARISONS"
1785 PRINT
1790 PRINTTAB11;"ERROR MEAN SQUARE FOR";TAB46;"RELATIVE EFFICIENCY"
1795 PRINTTAB4;"Y ONLY";TAB16;"REGRESSION";TAB30;"DIFFERENCE";TAB43;
1800 PRINT"  Y ONLY  ";TAB56;"DIFFERENCE"
1805 PRINT USING 1810,L1,C1,D1,L1/C1,D1/C1
1810 FIELD(5(F12.8,2X))
1815 PRINT
1820 PRINTTAB30;"************"
1825 PRINT
1830 PRINT
1835 PRINT
1840 PRINT
1845  REM    9E9 USED AT END OF EACH DATA SET TO SIGNIFY END OF SET.
1850  REM    8E8 USED IN DATA TO SIGNIFY MISSING DATA.
1855  REM    7E9 USED IN OUTPUT (QUANTITY WAS NOT CALCULATED).
```

LIST 1860,2105

```
1860  REM   ----------------------------------------------------------------
1865  REM                                  TABLE I
1870  REM   S2=DATA SOURCE CODE FROM TABLE BELOW:
1875  REM
1880  REM                  --DATA READ FROM--
1885  REM   CODE           FILE         PROGRAM           TRANSFORMATIONS
1890  REM   1         X(I) Y(I)*                          NONE
1895  REM   2         X(I) I(I)*                          I(I) SGD BECOMES Y(I).
1900  REM   3         X(I)**             I(I)             I(I) SGD BECOMES Y(I).
1905  REM   4         X(I)**             Y(I)             NONE
1910  REM   5         X(I)               Y(I)             NONE
1915  REM   6         X(I) Y(I)***                        IF P8=0 NO TRANS.
1920  REM             J(I) Y(I)***                        IF P8=1,X(I)=LOG J(I)
1925  REM   7         X(I)**Y(I)***                       HEADING FILE RESTORED.
1930  REM   8         J(I) I(I)*                          X(I)=LOG J(I)
1935  REM                                                 Y(I)=LOG I(I)
1940  REM
1945  REM   *         DATA FOR X AND Y READ FROM THE SAME FILE.
1950  REM   **        X(I) FILE IS RESTORED.
1955  REM   ***       TWO SEPARATE FILES FOR X AND Y.
1960  REM
1965  REM   HEADING FILE:NO.3 IF X(I) AND Y(I) ARE READ FROM SEP.FILES.
1970  REM   ----------------------------------------------------------------
1975  REM                                  TABLE II
1980  REM   C5=TRANSFORMATION PRINT CODE FROM TABLE BELOW:
1985  REM     CHANGE AS NEEDED TO DESCRIBE DATA
1990  REM   CODE        TRANSFORMED          UNTRANSFORMED
1995  REM   1                                  X(I) Y(I)
2000  REM   2                                  X(I) I(I)
2005  REM   3           I(I) TO Y(I)              X(I)
2010  REM   4                                  J(I) Y(I)
2015  REM   5           J(I) TO X(I)              Y(I)
2020  REM   6           X(I)     Y(I)
2025  REM   7                                  J(I) I(I)
2030  REM   8     J(I)TO X(I),I(I) TO Y(I)
2035  REM   ----------------------------------------------------------------
2040  REM                                  TABLE III
2045  REM   B5=PRINT CHOICE CODE FROM TABLE BELOW:
2050  REM      YES (+) OR NO (-) FOR PRINT OUT OF RESULTS.
2055  REM
2060  REM   CODE  X,Y,   SUMS;AJ.   UNC.S.SQ. COM.BTW. REGR.   SEP.VAR.
2065  REM         DEV.   MN;TOTS.   C.S.SQ.   TOTALS   AOFCOV  ALT.DSGN.
2070  REM   1      +        +          +         +        +        +
2075  REM   2      +        +          -         -        +        +
2080  REM   3      +        -          -         +        +        +
2085  REM   4      +        -          -         -        +        -
2090  REM   5      -        +          +         +        +        +
2095  REM   6      -        +          -         -        +        -
2100  REM   7      -        -          -         +        +        +
2105  REM   8      -        -          -         -        +        +
```

```
LIST 2110


2110  REM     --------------------------------------------------------
2115  REM                                   TABLE IV
2120  REM     L5=ADJUSTED MEAN CODE FROM TABLE BELOW:
2125  REM
2130  REM     CODE              SELECTION
2135  REM     1                 INDIVIDUAL SLOPES
2140  REM     2                 POOLED SLOPE
2145  REM     3                 OVERALL SLOPE
2150  REM     4                 PRIOR
2155  REM     --------------------------------------------------------
2160  REM     P7      NUMBER OF LINES OF PROGRAM HEADING.
2165  REM
2170  REM     K       NUMBER OF DATA SETS.
2175  REM
2180  REM     P8      (EQUAL TO 1) CODE FOR CONVERTING X TO LOGS.
2185  REM
2190  REM     I$      ITEM NAME (EXAMPLE...TRIAL,VACCINE,ETC.).
2195  REM
2200  REM     A$(I)   PROGRAM HEADINGS.
2205  REM
2210  REM     H$(I)   DATA SET HEADINGS.
2215  REM
2220  REM     X$(I)   DESIGNATION FOR ITH ELEMENT OF DATA.
2225  REM     ----------------------------------
2230  REM
2235 DATA 6,0,2,8,11,4,13,1,8,0,9E9
2240 DATA 0,2,3,1,18,4,14,9,1,9,9E9
2245 DATA 13,10,18,5,23,12,5,16,1,20,9E9    #REM  THIS DATA IS Y ONLY.
2250 END
```

```
OLD HEDCHRN
OK. DATE FILED:  11/23/73.


LIST


10 DRGA1
12 DRGA2
14 DRGA3
16 DRGA4
18 DRGA5
20 DRGA6
22 DRGA7
24 DRGA8
26 DRGA9
28 DGA10
30 DRGD1
32 DRGD2
34 DRGD3
36 DRGD4
38 DRGD5
40 DRGD6
42 DRGD7
44 DRGD8
46 DRGD9
48 DGD10
50 DRGF1
52 DRGF2
54 DRGF3
56 DRGF4
58 DRGF5
60 DRGF6
62 DRGF7
64 DRGF8
66 DRGF9
68 DGF10
```

OLD CHRNY
OK. DATE FILED: 08/05/74.

LIST

0	6.000000
3	0
6	2.000000
9	6.000000
12	11.00000
15	4.000000
18	13.00000
21	1.000000
24	8.000000
27	0
28	9E9
30	0
33	2.000000
36	3.000000
39	1.000000
42	18.00000
45	4.000000
48	14.00000
51	9.000000
54	1.000000
57	9.000000
58	9E9
60	13.00000
63	10.00000
66	18.00000
69	5.000000
72	23.00000
75	12.00000
78	5.000000
81	16.00000
84	1.000000
87	20.00000
88	9E9

OLD CHINA
OK. DATE FILED: 02/21/74.

```
LIST



0       11.00000
3       8.000000
6       5.000000
9       14.00000
12      19.00000
15      6.000000
18      10.00000
21      6.000000
24      11.00000
27      3.000000
28   9E9
30      6.000000
33      6.000000
36      7.000000
39      8.000000
42      18.00000
45      8.000000
48      19.00000
51      8.000000
54      5.000000
57      15.00000
58   9E9
60      16.00000
63      13.00000
66      11.00000
69      9.000000
72      21.00000
75      16.00000
78      12.00000
81      12.00000
84      7.000000
87      12.00000
88   9E9
```

OLD COCHRAN
OK. DATE FILED: 11/23/73.

LIST

```
100  11,  6,
110   8,  0,
120   5,  2,
130  14,  8,
140   19, 11,
150   5,  4,
160  10,  13,
170  6,  1,
180  11,  8,
190  3,  0,
200  9E9,     9E9,
210  6,  0,
220  6,  2,
230  7,  3,
240  8,  1,
250  16,  16,
260  8,  4,
270  19,  14,
280  8,  9,
290  5,  1,
300  15,  9,
305  9E9,     9E9,
310  16,  13,
320  13,  10,
330  11,  18,
340  9,  5,
350  21,  23,
360  16,  12,
370  12,  5,
380  12,  16,
390  7,  1,
400  12,  20,
410  9E9,     9E9,
```

RUN
16A

SPECIAL TEST FROM COCHRAN:CLIFFORD J. MALONEY,PH.D.,STAT.
COCHRAN EXAMPLE

TRIAL	X(I)	Y(I)	DEVIATION UNTRANS.X(I) AND Y(I)	DEV.SQD.
			DRUG A	
			COUNT= 10	
DRGA1	11.000	6.000	-.56687398	.32135178
DRGA2	8.000	0	-4.33121019	18.75938172
DRGA3	5.000	2.000	-.09554140	.00912816
DRGA4	14.000	8.000	-.80254777	.64408292
DRGA5	19.000	11.000	-1.52866242	2.33680879
DRGA6	6.000	4.000	1.15923567	1.34382734
DRGA7	10.000	13.000	7.17834395	51.52862185
DRGA8	6.000	1.000	-1.84076433	3.38841332
DRGA9	11.000	8.000	1.43312102	2.05383580
DGA10	3.000	0	-.60509554	.36614061
			DRUG D	
			COUNT= 10	
DRGD1	6.000	0	-1.82580645	3.33356920
DRGD2	6.000	2.000	.17419355	.03034339
DRGD3	7.000	3.000	.10564516	.01116090
DRGD4	8.000	1.000	-1.94290323	3.77379553
DRGD5	18.000	18.000	3.35161290	11.23330905
DRGD6	8.000	4.000	.03709677	.00137617
DRGD7	19.000	14.000	-1.71693548	2.94786746
DRGD8	8.000	9.000	5.03709677	25.37234391
DRGD9	5.000	1.000	.24274194	.05892365
DGD10	15.000	9.000	-2.44274194	5.96698610
			DRUG F	
			COUNT= 10	
DRGF1	16.000	13.000	-3.00283889	9.01704142
DRGF2	13.000	10.000	-2.41944642	5.85372096
DRGF3	11.000	18.000	7.96946190	63.51234178
DRGF4	9.000	5.000	-2.64153973	6.97779657
DRGF5	21.000	23.000	1.02484031	1.05029757
DRGF6	16.000	12.000	-4.00283889	16.02271920
DRGF7	12.000	5.000	-6.22498226	38.75040410
DRGF8	12.000	16.000	4.77501774	22.80079445
DRGF9	7.000	1.000	-4.25260140	18.08512951
DGF10	12.000	20.000	8.77501774	77.00093639

SUMS AND MEANS

X-SUM	X-MEAN	Y-SUM	MEAN	ADJUST.MEAN
93.00	9.30000000	53.000	5.30000000	6.71496346
100.00	10.00000000	61.000	6.10000000	6.82393479
129.00	12.90000000	123.000	12.30000000	10.16110174

TOTAL SUMS AND MEANS

SUM-X	MEAN-X	SUM-Y	MEAN-Y	COUNT
322.00	10.7333	237.0000	7.9000	30

UNCORRECTED SUMS OF SQUARES AND PRODUCTS

SOURCE	SUM X-SQ	SUM PROD.	SUM Y-SQ.	COUNT
DRUG A	1069.0000	645.0000	475.0000	10
DRUG B	1248.0000	875.0000	713.0000	10
DRUG F	1805.0000	1755.0000	1973.0000	10

CORRECTED SUMS OF SQUARES AND PRODUCTS

D.F.	SUM X-SQ.	SUM XY	SUM Y-SQ.	SLOPE
9	204.10000000	152.10000000	194.10000000	.74522293
9	248.00000000	265.00000000	340.89999999	1.06854839
9	140.90000001	168.30000001	460.10000000	1.19446416

******** COMMON ********

D.F.	SUM X-SQ.	SUM XY	SUM Y-SQ.	SLOPE
27	592.99999999	585.40000001	995.09999999	.98718381

******** BETWEEN CLASSES ********

D.F.	SUM X-SQ.	SUM XY	SUM Y-SQ.	SLOPE
2	72.86666667	145.79999998	293.60000000	2.00091491

******** TOTAL ********

D.F.	SUM X-SQ.	SUM XY	SUM Y-SQ.	SLOPE
29	665.86666668	731.19999999	1288.69999997	1.09811774

****

DEVIATIONS FROM REGRESSION

EXPERIMENTS	D.F.	SUM DEV.SQ.	MEAN SQUARE	....F....
DRUG A	8	80.75159236	10.09394905	.00935714
DRUG B	8	57.73467741	7.21683468	.43566990
DRUG F	8	259.07168201	32.38396025	1.95497296

ANALYSIS OF COVARIANCE

WITHIN	24	397.35795179	16.56491466	
BET.COEFF.	2	18.64464515	9.32232258	.59295944
CM SL RES SS	26	417.20259695	16.04625373	.96868919
CM SL RED SS	1	577.89740303	577.89740303	34.68683256
ADJ.MEANS	2	68.55371062	34.27685531	2.06924430
CM RG RES SS	28	485.75630757	17.34843956	1.04730027
CM RG RED SS	1	802.94359241	802.94359241	48.47255232
DEV BET CL SL	1	1.86660577	1.86660577	.11268430
BET CLASS SL	1	291.73339422	291.73339422	17.61152413

SEPARATE VARIATE ANALYSES OF VARIANCE

SOURCE	D.F.	SUM SQUARE	MEAN SQUARE	....F....
TREATMENT(X)	2	72.86666667	36.43333334	1.65885329
RESIDUAL (X)	27	592.99999999	21.96296296	
TREATMENT(Y)	2	293.60000000	146.80000000	3.98311727
RESIDUAL (Y)	27	995.09999999	36.85555555	

ALTERNATIVE DESIGN COMPARISONS

ERROR MEAN SQUARE FOR			RELATIVE EFFICIENCY	
Y ONLY	REGRESSION	DIFFERENCE	Y ONLY	DIFFERENCE
36.85555555	17.58264663	15.45555555	2.09613242	.87902327

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

2250, NORMAL EXIT FROM PROG.

CRUS: 67.76

COMPUTATION OF MOMENTS OF A LOG RAYLEIGH DISTRIBUTED RANDOM VARIABLE

William L. Shepherd
Research Projects Office
Instrumentation Directorate
White Sands Missile Range, New Mexico

ABSTRACT. A log Rayleigh distributed random variable is considered to be one for which there are numbers a and b, $a \neq 0$, $b > 0$, such that its probability density function is

$$f(x, a, b) = b(a\,sgn\,a)\,exp(ax - b\,exp\,ax) \quad , \quad -\infty < x < \infty \quad .$$

Formulas for recursive computation of the n^{th} moment are obtained and a convenient table of constants whose use facilitates the computation is given.

1. INTRODUCTION. In [1] and [2], a certain translation property of the log Rayleigh distribution was presented and used to obtain a simple formula for computing the moments from a small number of numerical constants and parameters of the distribution. In this paper, we derive alternate formulas to those in [1] and [2], which offer a somewhat better method for computing a second set of numerical constants related to the aforementioned set. The second set is also computed recursively, is more tractable to machine computation, and less subject to accumulated errors. A brief table of the altered constants is included. The derivation of the required formulas is self-contained in this presentation and the reader need only refer to [1] and [2] for background information. References [4], . . . , [6] cover related material in recent engineering literature; [7] is a paper on extreme value order statistics which touches on a special case of a log Rayleigh distributed random variable.

(1) A random variable, X, is a Rayleigh distributed random variable if there is a constant $b > 0$ such that the probability density function (pdf) for X is

$$p(x, b) = 2bxe^{-bx^2} \quad , \quad \text{for } 0 \leq x$$
$$= 0 \quad , \quad \text{for } x < 0 \quad .$$

(2) Y is a log Rayleigh distributed random variable if X is a Rayleigh distributed random variable and there is a constant $a \neq 0$ such that

$$Y = \frac{2}{a} \ln X \quad .$$

It can be shown that the pdf for Y is

$$f(x, a, b) = b(a\,sgn\,a)\,exp(ax - b\,exp\,ax) \quad , \quad -\infty < x < \infty \quad . \tag{1.1}$$

(3) The moments are

$$m_n(a, b) = \int_{-\infty}^{\infty} x^n f(x, a, b)dx \quad , \quad n = 0, 1, 2, \ldots . \tag{1.2}$$

Formulas for computing the moments are obtained in Section 3. Section 2 presents some prerequisite material on the Riemann zeta function and on the gamma function and its derivatives.

2. SOME REQUIRED FORMULAS FOR COMPUTING THE MOMENTS. The derivations require a number of equations from [3], which are referred to in the notation of that book. With $\psi^{(o)}(x)$ and $\Gamma^{(o)}(x)$ referring to the digamma and gamma functions, Equation 6.3.1 of [3] is

$$\psi^{(o)} = \frac{\Gamma^{(1)}(x)}{\Gamma^{(o)}(x)} \quad ,$$

or

$$\Gamma^{(1)}(x) = \psi^{(o)}(x)\Gamma^{(o)}(x) \quad . \tag{2.1}$$

Differentiating n-1 times, we obtain

$$\Gamma^{(n)}(x) = \sum_{j=o}^{n-1} \binom{n-1}{j} \psi^{(n-1-j)}(x)\Gamma^{(j)}(x) \quad , \quad n = 1, 2, \ldots .$$

$$\Gamma^{(n)}(1) = \sum_{j=o}^{n-1} \binom{n-1}{j} \psi^{(n-1-j)}(1)\Gamma^{(j)}(1) \quad . \tag{2.2}$$

$$\Gamma^{(n)}(1) = (n-1)! \sum_{j=o}^{n-1} \frac{1}{(n-1-j)!} \psi^{(n-1-j)}(1) \frac{1}{j!} \Gamma^{(j)}(1) \quad ,$$

from which

$$\frac{(-1)^n \Gamma^{(n)}(1)}{n!} = \frac{(-1)^n}{n} \sum_{j=o}^{n-1} \frac{1}{(n-1-j)!} \psi^{(n-1-j)}(1) \frac{1}{j!} \Gamma^{(j)}(1) \quad . \; . \; . \; . \qquad (2.3$$

Equations (6.3.2) and (6.4.2) of [3] are

$$\psi^{(o)}(1) = -\gamma \quad , \qquad (2.4)$$

$$\psi^{(j)}(1) = (-1)^{j+1} j!\zeta(j+1) \quad , \quad j = 1, 2, \ldots \qquad (2.5)$$

We will use the notation

$\zeta_1 = \gamma$ (Euler's constant)

$\zeta_n = \zeta(n)$, $n = 2, 3, \ldots$ (Riemann zeta function).

Substituting (2.4) and (2.5) into (2.3) yields, after manipulating,

$$\frac{(-1)^n \Gamma^{(n)}(1)}{n!} = \frac{1}{n} \sum_{j=o}^{n-1} \zeta_{n-j} \frac{(-1)^j \Gamma^{(j)}(1)}{j!} \quad , \quad n = 1, 2, \ldots . \qquad (2.6)$$

With

$$d_o = 1 \quad , \qquad (2.7)$$

$$d_j = \frac{(-1)^j \Gamma^{(j)}(1)}{j!} \quad , \quad j = 1, 2, \ldots , \qquad (2.8)$$

(2.6) is

$$d_n = \frac{1}{n} \sum_{j=o}^{n-1} \zeta_{n-j} \, d_j \quad , \quad n = 1, 2, \ldots . \qquad (2.9)$$

Since accurate values for the Riemann zeta function, $\zeta(n)$, are available in Table 23.3 of [3], Equations (2.9) and (2.8) permit a recursive computation of $\Gamma^{(n)}(1)$.*

Another formula for $\Gamma^{(n)}(1)$, needed in (3.3), can be derived from Equation 6.1.1 of [3] by repeated differentiation.

$$\Gamma(Z) = \int_0^\infty t^{Z-1}e^{-t}dt$$

$$\Gamma^{(n)}(Z) = \int_0^\infty (\ell n t)^n t^{Z-1}e^{-t}dt$$

$$\Gamma^{(n)}(1) = \int_0^\infty (\ell n t)^n e^{-t}dt \quad . \qquad (2.10)$$

3. FORMULAS FOR THE MOMENTS, $m_n(a, b)$. From (1.1) and (1.2),

$$m_n(a, b) = \int_{-\infty}^{\infty} x^n (a\,sgn\,a)exp[(ax + \ell n b) - exp(ax + \ell n b)]dx \; . \; . \; . \; . \qquad (3.1)$$

Letting $x = (\ell n t - \ell n b)/a$, (3.1) becomes

$$m_n(a, b) = a \int_0^\infty \left(\frac{\ell n t - \ell n b}{a}\right)^n te^{-t} \frac{1}{at}\, dt \quad .$$

With $0^\circ \overset{D}{=} 1$,

$$m_n(a, b) = \left(\frac{1}{a}\right)^n \sum_{j=o}^{n} \binom{n}{j}(-\ell n b)^{n-j} \int_0^\infty (\ell n t)^j e^{-t}dt \; . \; . \; . \; . \qquad (3.2)$$

*From (2.7), (2.8) and the Taylor expansion, about $x = 1$, of $\Gamma(x)$, we also have

$$\Gamma(x) = \Gamma(1) - d_1(x - 1) + d_2(x - 1)^2 + \ldots$$

$$= 1 + d_1(1 - x) + d_2(1 - x)^2 + \ldots , \quad 0 < x < 2 \quad .$$

Combining (2.10) and (3.2) results in

$$m_n(a, b) = n!(-\frac{1}{a})^n \sum_{j=o}^{n} \frac{1}{(n-j)!} (\ell n b)^{n-j} \frac{(-1)^j \Gamma^{(j)}(1)}{j!} \quad . \ . \ . \ . \qquad (3.3)$$

Using (2.8), this becomes

$$m_n(a, b) = n!(-\frac{1}{a})^n \sum_{j=o}^{n} \frac{1}{(n-j)!} (\ell n b)^{n-j} d_j \ , \quad n = 0, 1, \ . \ . \ . \ . \qquad (3.4)$$

In particular,

$$m_n(-1, 1) = n! d_n = (-1)^n \Gamma^{(n)}(1) \quad , \quad n = 0, 1, \ . \ . \ . \ . \qquad (3.5)$$

4. SUMMARY AND CONCLUSIONS . With $\zeta_1 = \gamma$, $\zeta_n = \zeta(n)$, $(n = 2, 3, \ . \ . \ . \ )$, $d_o = 1$, and

$$d_n = \frac{(-1)^n \Gamma^{(n)}(1)}{n!} \quad , \quad n = 0, 1, \ . \ . \ . \ , \qquad (4.1)$$

we have derived the formulas

$$d_n = \frac{1}{n} \sum_{j=o}^{n-1} \zeta_{n-j} d_j \quad , \quad n = 1, 2, \ . \ . \ . \ , \qquad (4.2)$$

$$m_n(a, b) = n!(-\frac{1}{a})^n \sum_{j=o}^{n} \frac{1}{(n-j)!} (\ell n b)^{n-j} d_j \quad , \quad n = 0, 1, \ . \ . \ . \ , \qquad (4.3)$$

$$m_n(-1, 1) = n! d_n \quad . \qquad (4.4)$$

Formulas (4.3) and (4.4) were presented in disguised form in [2]. The forms presented above offer advantages over [1] and [2] in computing $m_n(a, b)$. Also, (4.1) and (4.2) together offer a way to compute $\Gamma^{(n)}(1)$ which has not been found in the literature.

We have $d_1 = \gamma = .5772156649015327$. For $j = 1, 2, \ . \ . \ . \ , 50$, Table I gives numerical results in computing d_j.

TABLE I

j	d_j
1	.5772156649015327
2	.9890559953279726
3	.9074790760808863
4	.9817280868344002
5	.9819950689031452
6	.9931491146212762
7	.9960017604424315
8	.9981056937831289
9	.9990252676219549
10	.9995156560727774
11	.9997565975086013
12	.9998782713151333
13	.9999390642064443
14	.9999695177634821
15	.9999847526993770
16	.9999923744790732
17	.9999961865894733
18	.9999980930811309
19	.9999990464689111
20	.9999995232106057
21	.9999997615973444
22	.9999998807960191
23	.9999999403971250
24	.9999999701982676
25	.9999999850990354
26	.9999999925494849
27	.9999999962747315
28	.9999999981373621
29	.9999999990686798
30	.9999999995343395
31	.9999999997671696
32	.9999999998835847
33	.9999999999417923
34	.9999999999708961
35	.9999999999854481
36	.9999999999927240
37	.9999999999963620
38	.9999999999981810
39	.9999999999990905
40	.9999999999995452
41	.9999999999997726
42	.9999999999998863
43	.9999999999999431
44	.9999999999999715
45	.9999999999999857
46	.9999999999999928
47	.9999999999999964
48	.9999999999999982
49	.9999999999999991
50	.9999999999999995

The numerical results presented in the Table support the conjecture that $\{d_n\}_{n=0}^{\infty}$ has the limit 1. If this is true, it is easily shown that the relative error in either of the formulas

$$m_n(-1, 1) \simeq n!$$

or

$$(-1)^n \Gamma^{(n)}(1) \simeq \Gamma(n+1)$$

is zero at $n = \infty$.

REFERENCES.

[1] W. L. Shepherd and P. Milnarich, Jr., "Basic Relations Between a Rayleigh Distributed Random Varying Voltage and a Decibel Record of the Voltage," Proceedings of the IEEE (Letters), Vol. 61, No. 12, pp. 1765-66, December 1973.

[2] W. L. Shepherd and P. Milnarich, Jr., "Moments of Log Rayleigh Distributions," Proceedings of the IEEE (Letters), Vol. 62, No. 7, pp. 1168-69, August 1974.

[3] M. Abramowitz and I. A. Stegun, Eds., "Handbook of Mathematical Functions," (Applied Mathematics Series 55), Washington, DC, National Bureau of Standards, 1964.

[4] M. Peritsky, "Statistical Estimation of Mean Signal Strength in a Rayleigh-Fading Environment," IEEE Transactions on Communications, Vol. COM-21, No. 11, pp. 1207-1213, 1973.

[5] H. J. Hinder, "Poisson Transform Pair Evaluation of Distribution Moments," Proceedings of the IEEE (Letters), Vol. 61, No. 10, pp. 1501-1502, October 1973.

[6] I. Sugai and P. F. Christopher, "Comments on Logarithmic Compression of Rayleigh and Maxwell Distributions," Proceedings of the IEEE (Letters), Vol. 58, No. 2, February 1970.

[7] R. A. Fisher and L. H. C. Tippett, "Limiting Forms of the Frequency Distribution of the Largest and Smallest Member of a Sample," Proceeding of the Cambridge Philosophical Society, Vol. 24, Pt. 2, pp. 180-190, 1928. (Also published in R. A. Fisher, "Contributions to Mathematical Statistics," John Wiley and Sons, 1950.)

[8] H. A. David, "Order Statistics," Chapter 9 (especially Sections 9.1 and 9.3), John Wiley and Sons, 1970.

TYPE II ERROR OF THE 2X2 CONTINGENCY TABLE CHI-SQUARE STATISTIC

Robert L. Launer
Procurement Research Office
Fort Lee, VA 23801

ABSTRACT. Around the turn of the century, K. Pearson introduced the "mean square contingency", ϕ^2, which is a measure of the degree of association between 2 discrete variables. When applied to m x n contingency tables, ϕ^2 is identical to the usual chi-square statistic divided by the sample size, and in the 2 x 2 case, ϕ^2 was identified as the square of the correlation coefficient, ρ . [2;282].

For this note, a certain bivariate Bernoulli distribution is developed which explicitly involves the correlation coefficient. The square of the maximum likelihood estimate (MLE) of ρ is found to be equal to ϕ^2. This allows the use of known large sample distributional properties of MLE's to develop the general distribution of the chi-square statistic and also to develop the type II error of the contingency table test of independence as a function of ρ . Several interesting properties of the bivariate Bernoulli distribution are related.

INTRODUCTION: A BIVARIATE BERNOULLI DISTRIBUTION. Let X and Y represent Bernoulli random variables. That is,

$P[X = 1] = p_1$, $P[X = 0] = 1-p_1$, $P[Y = 1] = p_2$, and $P[Y = 0] = 1-p_2$.

If X and Y are independent, the joint probabilities $\pi xy = P[X = x, Y = y]$ are given by $p_1^x \; q_1^{1-x} \; p_2^y \; q_2^{1-y}$, where $q_1 = 1-p$, and $q_2 = 1-p_2$.

If X and Y are not independent, then at least one of the cell probabilities (say π_{11}) must differ from the marginal product probability by an amount (say K.) Thus $\pi_{11} = p_1p_2 + K$, but to keep the marginal probabilities unchanged the remaining cell probabilities must be altered so that

$\pi_{10} = p_1q_2 - K$, $\pi_{01} = q_1p_2 - K$ and $\pi_{00} = q_1q_2 + K$.

This may be written in one of the two forms:

(1) $\pi xy = (1 - 2x)(1 - 2y)\,K + p_1^x \; q_1^{1-x} \; p_2^y \; q_2^{1-y}$;

$$\pi xy = (1 - 2x)(1 - 2y)\,K + [1-x-p_1\,(1 - 2x)][1 - y - p_2\,(1 - 2y)].$$

$$x,y = 0,1$$

The two cases are given in tabular form below.

	Y = 0	Y = 1
x = 0	q_1q_2	q_1p_2
x = 1	p_1q_2	p_1p_2

Independent

	Y = 0	Y = 1
x = 0	$q_1q_2 + K$	$q_1p_2 - K$
x = 1	$p_1q_2 - K$	$p_1p_2 + K$

Non-Independent

Since $E(XY) = p_1p_2 + K$, $E(X) = p_1$, and $E(Y) = p_2$, it is immediately apparent that:

(2) $\text{Cov}(X, Y) = K$

Notice that since $\text{Var}(x) = p_1q_1$ and $\text{Var}(Y) = p_2q_2$, the correlation coefficient, ρ, is:

(3) $\rho = K / \sqrt{p_1p_2q_1q_2}$

If πxy are given by (1) and yield the marginal probabilities p_1 and p_2 for X and Y, then by forming the sum $\pi_{11}+\pi_{00} - \pi_{10}-\pi_{01}$, we obtain:

(4) $K = \pi_{11}\,\pi_{00} - \pi_{10}\,\pi_{01}$ and

(5) $\rho = (\pi_{11}\,\pi_{00} - \pi_{01}\,\pi_{10}) / \sqrt{(\pi_{11}+\pi_{10})(\pi_{11}+\pi_{01})(\pi_{00}+\pi_{01})(\pi_{00}+\pi_{10})}$

An examination of the illustration above, indicates that K may never exceed any one of the four quantities $1-p_1p_2$, $1-q_1q_2$, p_1q_2 or q_1p_2 and may never be less than $-p_1p_2$, $-q_1q_2$, $-(1-p_1q_2)$ or $-(1-q_1p_2)$ since all cell probabilities are bounded by 0 and 1. Then $p_1q_2 = p_1 - p_1p_2 < 1-p_1p_2$, $q_1p_2 = q_1 - q_1q_2 < 1 - q_1q_2$, etc., so that

(6) $-\min(p_1p_2,\ q_1q_2) \leq K \leq \min(p_1q_2,\ q_1p_2)$.

Now when $p_1 > p_2$ then $0 < p_1 - p_2 = p_1 - p_1p_2 - (p_2 - p_1p_2) = p_1q_2 - p_2q_1$ so that $\min(p_1q_2,\ q_1p_2) = q_1p_2$. Also, when $p_1 < q_2$, then $0 < p_1 - q_2 = p_1p_2 - (1-p_2-p_1+p_1p_2) = p_1p_2 - q_1q_2$ so that $\min(p_1p_2,\ q_1q_2) = q_1q_2$.

Then

$$
(7) \quad K \le \begin{cases} q_1 p_2 , & p_1 > p_2 \\ p_1 q_2 , & p_1 < p_2 \end{cases}
\qquad
K \ge \begin{cases} -p_1 p_2 , & p_1 < q_2 \\ -q_1 q_2 , & p_1 > q_2 \end{cases}
$$

The joint moment generating function (MGF) of X and Y is easily seen to be

$$M_{x,y}(\theta_1, \theta_2) = [1-p_1(1-e^{\theta 1})][1-p_2(1-e^{\theta 2})] + K(1-e^{\theta 1})(1-e^{\theta 2}).$$

It follows directly that if X_i and Y_i are distributed as X and Y respectively, if $W = X_1 + X_2 + \ldots + X_n$, and $X = Y_1 + Y_2 + \ldots Y_n$, then the joint MGF of $(W-np_1)/\sqrt{np_1q_1}$ and $(Z-np_2)/\sqrt{np_2q_2}$ is (apart from terms of order 1/n in the exponent)

$$\exp\ [1/2(\theta_1^2 + 2\ \ \theta_1\theta_2 + \theta_2^2)],$$

and therefore W and Z are approximately jointly normally distributed with mean np_1 and np_2 and correlation coefficient ρ .

<u>MAXIMUM LIKELIHOOD ESTIMATES OF p_1, p_2 and K.</u> The likelihood equation $L = \Pi^n{}_{i=1}\ \pi_{x_iy_i}$ is

$$(8) \quad L = \Pi^n{}_{i=1}\ [(1-2x_i)(1-2y_i)K + p_1{}^{xi} q_1{}^{1-xi} p_2{}^{yi} q_2{}^{1-yi}].$$

Differentiation of ln L with respect to p_1, p_2 and K yields the following set of equations:

$$
(9) \quad
\begin{aligned}
\frac{n_{00}}{\pi_{00}} - \frac{n_{10}}{\pi_{10}} &= 0 \\
\frac{n_{00}}{\pi_{00}} - \frac{n_{01}}{\pi_{01}} &= 0 \\
\frac{n_{00}}{\pi_{00}} - \frac{n_{11}}{\pi_{11}} &= 0
\end{aligned}
$$

where n_{xy} represents the number of observations for which X = x, Y = y. The solution of (9) is:

$$\text{(10)} \quad \hat{p}_1 = \bar{x} \qquad \hat{p}_2 = \bar{y}$$

$$\hat{K} = (n_{11}n_{00} - n_{10}n_{01})/n^2$$

If equations (10) are consistent, then from a result of Chanda [1] they are the unique MLE of p_1, p_2 and K. We will demonstrate consistency. Let

$$I_i^{xy} = \begin{cases} 1, & X_i=x,\ Y_i=y \\ 0, & \text{otherwise} \end{cases}$$

Then $E\,(n_{00}n_{11}) = E[\Sigma_i I_i^{00} \Sigma_j I_j^{11}] = n(n-1)(p_1p_2 + K)(q_1q_2 + K)$ and $E(n_{10}n_{01}) = n(n-1)(p_1q_2 - K)(q_1p_2 - K)$. Thus $E(n_{11}n_{00} - n_{01}n_{10})/n^2 = (n-1)K/n$ so $\hat{K}$ is asymptotically unbiased. Similarly, (the algebra is inordate) $V(\hat{K}) = K^2\ (n-1)(2-n)/n^3 + K(q_1-p_1)(q_2-p_2)(n=1)^2/n^3 + p_1p_2q_1q_2(n-1)/n^2$,

and neglecting terms of order greater than 1/n,

$$\text{(11)} \quad V(\hat{K}) = [-K^2 + K(q_1 - p_1)(q_2 - p_2) + p_1p_2q_1q_2]/n.$$

Therefore; $\hat{K}$ is a consistent estimate of K. It follows from [1] also that $\hat{p}_1$, $\hat{p}_2$ and $\hat{K}$ possess a multivariate normal distribution. We may then, for large samples, assume that $\hat{K}$ is normally distributed with mean K and variance $V(\hat{K})$ given by (11). More precisely, $\sqrt{n}(\hat{K} - K)/\sqrt{nV(\hat{K})}$ is asymptotically N(0,1).

THE ASYMPTOTIC DISTRIBUTION OF $\hat{\rho}$. To obtain the MLE for ρ we note that $K = \rho\sqrt{p_1p_2q_1q_2}$ is a one to one transformation from K to ρ since the Jacobian of the transformation $u_1(\theta_1) = p_1$, $u_2(\theta_2) = p_2$, and $u_3(\theta_3) = K/\sqrt{p_1p_2q_1q_2}$ is $\sqrt{\theta_1\ \theta_2\ (1-\theta_1)(1-\theta_2)} > 0$. Therefore from (9) we have,

$$\text{(12)} \qquad \hat{\rho} = \frac{n_{11}n_{00} - n_{01}n_{10}}{n^2\sqrt{\hat{p}_1\hat{p}_2\hat{q}_1\hat{q}_2}}$$

That (12) converges stochastically to ρ follows from a theorem of Slutsky [2;255] (since $\sqrt{\hat{p}_1\hat{p}_2\hat{q}_1\hat{q}_2}$ converges stochastically to $\sqrt{p_1p_2q_1q_2}$) and from the result of the last section.

It is interesting to note that $\hat{\rho}$ may be written

(13) $$\hat{\rho} = \frac{n_{11}n_{00} - n_{01}n_{10}}{\sqrt{(n_{11}+n_{10})(n_{10}+n_{00})(n_{00}+n_{01})(n_{01}+n_{11})}} ,$$

and that $|\hat{\rho}|$ is the well known phi coefficient which is used to estimate the "degree of association" in 2 x 2 contingency tables [2].

To obtain the distribution of $\hat{\rho}$ note first that from the result of the preceding section, approximately,

(14) $$\sqrt{n}\left[\frac{n_{11}n_{00} - n_{10}n_{01}}{n^2\sqrt{p_1q_1p_2q_2}} - \rho\right] \sim N\left(0, \frac{n\ V\ (\hat{K})}{p_1p_2q_1q_2}\right).$$

Since $\sqrt{\bar{x}\ (1-\bar{x})\ \bar{y}\ (1-\bar{y})}/\sqrt{p_1p_2q_1q_2}$ converges stochastically to 1, then combining this with (14) it follows from a result of Cramer [2; 254] that, asymptotically,

$$\sqrt{n}\,(\hat{\rho}-\rho) \sim N\ (0,\ n\ V\ (\hat{K})/p_1p_2q_1q_2)$$

Thus for large samples, $\hat{\rho}$ is approximately normally distributed with mean ρ and variance,

(15) $$V\ (\hat{\rho}) = [1-\rho^2 + \rho\ (q_1-p_1)(q_2-p_2)/\sqrt{p_1p_2q_1q_2}]/n$$

THE LARGE SAMPLE DISTRIBUTION OF THE 2x2 STATISTIC. If we denote the 2x2 chi-square statistic by X^2_{2x2}, then from (13) we have

(16) $$n\,\hat{\rho}^2 = X^2_{2x2}.$$

From (14) and (16) it follows that approximately for large samples,

(17) $$X^2_{2x2} \sim [n\ V\ (\hat{\rho})].\ X'^2_1\ (\rho^2/V\ (\hat{\rho})),$$

where $X'^2_\nu\ (\theta)$ represents a non-central chi-square random variable with ν degrees of freedom and non-centrality parameter θ.

APPROXIMATION FOR THE TYPE II ERROR OF THE 2x2 TEST. The type II error of the 2x2 contingency table test may be computed for large samples from (15) and (17), but it is easier to use the normal distribution of $\hat{\rho}$. For example, the .05 critical value for the chi-squared test with 1 degree of freedom is approximately 3.841. Then, for moderately large samples, we may write

$$\beta = P\,[\chi^2_{2x2} < 3.841 \mid \rho \neq 0], \text{ or}$$

$$\beta = P\,[\sqrt{n}\,\hat{\rho} < \sqrt{3.841}] - P\,[\sqrt{n}\,\hat{\rho} < \sqrt{3.841}].$$

Notice that if $p_1 = 1/2$, $p_2 = 1/2$ then $V(\hat{\rho}) = (1-\rho^2)/n$. This is an interesting special case because the variance is not dependent on p_1 or p_2 and because K has the largest range of possible values (see (6) and (7)). For this case we have $-1/4 < K < 1/4$, $-1 < \rho < 1$, and

$$\beta = P[N(0,1) < (\sqrt{3.841} - \rho\sqrt{n}\,/\sqrt{1-\rho^2})] -$$

$$P[N(0,1) < (-\,3.841 - \rho\sqrt{n})/\sqrt{(1-\rho^2)}].$$

In order to indicate the amount of error associated with the approximation given in this section, the following tables compare the actual type II error obtained from (1) and

$$\binom{n}{n_{00}, n_{01}, n_{10}, n_{11}}\ \pi_{00}^{n_{00}}\ \pi_{01}^{n_{01}}\ \pi_{10}^{n_{10}}\ \pi_{11}^{n_{11}},$$

with the approximate values obtained from (18) and (15) for several cases. Notice that by changing the values of p_1 and p_2 to $1-p_1$ and $1-p_2$, respectively or by changing only one of p_1 or p_2 along with the sign of ρ as given below, the tables may be used for 4 separate cases each (the symmetric case $p_1 = p_2 = 1/2$ excepted). This may be seen from (15) and (7).

PROBABILITY OF ACCEPTING Ho: $\rho = 0$

COMPARISON OF APPROXIMATE VALUES WITH EXACT VALUES

$p_1 = 1/2$; $p_2 = 1/2$; $\underline{n = 25}$

		$\alpha = .1$		$\alpha = .05$		$\alpha = .01$	
$\lvert K \rvert$	$\lvert \rho \rvert$	APPROX	EXACT	APPROX	EXACT	APPROX	EXACT
0	0	.900	.884	.950	.948	.990	.989
.025	.1	.860	.841	.922	.920	.984	.979
.050	.2	.741	.717	.835	.831	.946	.942
.075	.3	.560	.531	.685	.678	.870	.860
.100	.4	.349	.327	.483	.474	.735	.714
.125	.5	.162	.156	.266	.266	.535	.508
.150	.6	.045	.051	.097	.108	.298	.283
.175	.7	.005	.010	.016	.026	.098	.106
.200	.8	.000	.001	.000	.003	.009	.019
.225	.9	.000	.000	.000	.000	.000	.001

PROBABILITY OF ACCEPTING Ho: $\rho = 0$

COMPARISON OF APPROXIMATE VALUES WITH EXACT VALUES

$p_1 = 1/2$; $p_2 = 1/2$; $\underline{n = 50}$

		$\alpha = .1$		$\alpha = .05$		$\alpha = .01$	
$\lvert K\rvert$	$\lvert\rho\rvert$	APPROX	EXACT	APPROX	EXACT	APPROX	EXACT
0	0	.900	.890	.950	.943	.990	.990
.025	.1	.818	.804	.892	.881	.969	.971
.050	.2	.592	.573	.711	.692	.882	.885
.075	.3	.309	.295	.433	.413	.683	.686
.100	.4	.098	.097	.171	.165	.391	.397
.125	.5	.015	.017	.034	.038	.134	.145
.150	.6	.001	.001	.002	.004	.019	.025
.175	.7	.000	.000	.000	.000	.000	.002
.200	.8	.000	.000	.000	.000	.000	.000
.225	.9	.000	.000	.000	.000	.000	.000

PROBABILITY OF ACCEPTING Ho: $\rho = 0$

COMPARISON OF APPROXIMATE VALUES WITH EXACT VALUES

$p_1 = 3/4$; $p_2 = 3/4$;

<u>$n = 25$</u>

$p_1 = 1/4$; $p_2 = 1/4$;

K*	ρ*	α = .1 APPROX	α = .1 EXACT	α = .05 APPROX	α = .05 EXACT	α = .01 APPROX	α = .01 EXACT
-.05625	-.3	.580	.634	.740	.848	.934	.985
-.03750	-.2	.780	.815	.875	.936	.971	.995
-.01875	-.1	.882	.900	.939	.967	.987	.997
0	0	.900	.906	.950	.955	.990	.991
.01875	.1	.838	.848	.906	.900	.973	.971
.03750	.2	.731	.712	.803	.795	.922	.924
.05625	.3	.547	.543	.659	.642	.826	.836
.07500	.4	.380	.363	.486	.460	.688	.697
.09375	.5	.236	.206	.325	.281	.525	.516
.11250	.6	.129	.095	.193	.139	.362	.320
.13125	.7	.061	.034	.100	.052	.221	.154
.15000	.8	.024	.009	.044	.014	.117	.050
.16875	.9	.008	.002	.016	.003	.051	.008

*Change the sign of ρ to obtain the table of values for: $p_1 = 3/4$; $p_2 = 1/4$ or $p_1 = 1/4$; $p_2 = 3/4$

PROBABILITY OF ACCEPTING $H_0: \rho = 0$

COMPARISON OF APPROXIMATE VALUES WITH EXACT VALUES

$p_1 = .9; \ p_2 = .9$

n = 25

$p_1 = .1; \ p_2 = .1$

K*	ρ *	α = .1 APPROX	α = .1 EXACT	α = .05 APPROX	α = .05 EXACT	α = .01 APPROX	α = .01 EXACT
-.009	-.1	.985	.998	.997	.995	.999	.998
0	0	.900	.908	.950	.938	.990	.972
.009	.1	.760	.800	.839	.853	.935	.926
.018	.2	.619	.776	.705	.746	.836	.857
.027	.3	.497	.547	.580	.623	.722	.764
.036	.4	.398	.423	.472	.493	.609	.651
.045	.5	.317	.316	.382	.378	.507	.523
.054	.6	.252	.229	.307	.274	.418	.392
.063	.7	.200	.165	.246	.191	.342	.271
.072	.8	.158	.121	.196	.132	.278	.172
.081	.9	.125	.092	.155	.094	.224	.105

*Change the sign of ρ to obtain the table of values for:

$p_1 = .9; \ p_2 = .1$ or

$p_1 = .1; \ p_2 = .9$

DANIEL AWARDED THE 1974 SAMUEL S. WILKS MEMORIAL MEDAL

Introductory Remarks Made by Frank E. Grubbs, Conference Chairman

The Twentieth Design of Experiments Conference in Army Research, Development and Testing marks a very significant milestone for Statistical Methods in the Army. As we look back over all the conferences, it is easy to see that we owe the greatest of debts to Sam Wilks for his vision in getting Army statisticians together on a yearly basis for the common good of all. Indeed, we have benefited much from our twenty conferences, and I don't see now how we could have gone through the years without them. Do you? Every time, I have gone home with a new view of statistics in the Army, the papers by our University friends have kept us up to date and provided good summaries of timely accomplishments, and we have been motivated to either attack old problems with new vigor or to address new pressing problems. We have not stuck to the title, "Design of Experiments," in all detail, but that is good. The field changes fast and we must always move on to new things or areas, for example, reliability, to mention one. I could go on and on concerning the good these conferences have done, and Churchill Eisenhart has covered much of that so well anyway, so I'll stop on this point now. But I must mention that the success of these conferences would not have been so great were it not for our most dedicated friend, Francis Dressel, who as we all know deserves a vote of thanks at this time for his effective, continuing contributions.

Now I would like to not introduce but name people at the head table that we know so well and have enjoyed association with over the years in our design conferences.

We now turn to the Samuel S. Wilks Memorial Medal.

The Samuel S. Wilks Memorial Medal Award, initiated jointly in 1964 by the U. S. Army and the American Statistical Association, is administered by the American Statistical Association, a non-profit, educational and scientific society founded in 1839. The Wilks Award is given each year to a statistician and is based primarily on his contributions to the advancement of scientific or technical knowledge in Army statistics, ingenious application of such knowledge, or successful activity in the fostering of cooperative scientific matters which coincidentally benefit the Army, the Department of Defense, the U. S. Government, and our country generally.

The Award consists of a medal, with a profile of Professor Wilks and the name of the Award on one side, the seal of the American Statistical Association and name of the recipient on the reverse, and a citation and honorarium related to the magnitude of the Award funds. The annual Army Design of Experiments Conferences, at which the Award is given each year, are sponsored by the Army Mathematics Steering Committee on behalf of the Office of the Chief of Research, Development and Acquisition, Department of the Army.

Previous recipients of the Samuel S. Wilks Memorial Medal include John W. Tukey of Princeton University (1965), Major General Leslie E. Simon retired (1966), William G. Cochran of Harvard University (1967), Jerzy Neyman of the University of California, Berkely (1968), Jack Youden (1969) retired from the National Bureau of Standards and deceased, George W. Snedecor (1970) retired from Iowa State University, Harold Dodge (1971) retired from Bell Telephone Laboratories, G.E.P. Box (1972), University of Wisconsin, and H. O. Hartley (1973), Texas A. and M. University.

With the approval of President Jerome Cornfield of the American Statistical Association, the Samuel S. Wilks Memorial Medal Committee consisted of Bob Bechhofer (Cornell University), George Box (University of Wisconsin), Joe Cameron (National Bureau of Standards), Fred Frishman (Internal Revenue Service), Oscar Kempthorne (Iowa State University), Albert Madansky (Great Neck, N. Y.), Bill Pabst (Retired from the Navy), Les Simon (Major General, Retired) and of course the esteemed chairman, Stu Hunter (Princeton University). What an array of intelligentsia to select the 1974 Wilks Medalist!

The 1974 Wilks Medalist was born in 1904 in Williamsport, Pennsylvania. He secured a Bachelor of Science degree and a Masters degree in Chemical Engineering from the Massachusetts Institute of Technology in 1925 and 1926. He later studied briefly at the University of Berlin and at Harvard. After a short stint as a teacher, he became a research associate in the Evaluation of School Broadcasts at Ohio State University and later a research associate with the Princeton Office of radio Research. It was during this period that he began his acquaintance with S. S. Wilks and the statistical fraternity at Princeton. During World War II, he was employed as a statistician with the Manhattan Project in Oak Ridge, Tennessee. From 1947 onwards he has been a private consultant to the food, steel, chemical and pharmaceutical industries. For a period he was also a consultant to the U. S. Army at the Signal Research Laboratories at Fort Monmouth, New Jersey. He has also been a consultant for Consumers' Union, helping with organizing, planning and analysis of its comparative experiments. He is a member of the Cancer Clinical Investigations Review Committee of the National Cancer Institute and has served as a consultant the the Office of Air Pollution. He was Chairman of the Gordon Conference on Statistics in Chemistry and Chemical Engineering in 1954, and collaborated with W. J. Youden and S. Lee Crump in the early establishment of these annual conferences. He was Chairman of the Section of Physical and Engineering Sciences of the ASA in 1959. He has published widely in the JASA, Biometrics and Technometrics, has been a contributor to the Berkeley Symposia and to journals in the engineering professions. He co-Authored with F. S. Wood the text "Fitting Equations to Data" (1971) and he is currently completing a manuscript on the design and analysis of industrial experiments. This latest manuscript, although still unpublished, is already recognized for its novel contributions to the organization and interpretation of factorial experiments. He was the R. A. Fisher Memorial lecturer to the joint meetings of the ASA and IMS in 1971, and the W. J. Youden memorial lecturer for the American Society for Quality Control in 1974. He was one of the original associate editors of Technometrics. He is a fellow of the American Statistical Association and of the Institute of Mathematical Statistics. He has attended and contributed to many of these Army Design of Experiments Conferences. He is well known to us for his unique talks and commentaries, and listerners are assured of an entertaining as well as education experience whenever he speaks at a meeting. He is most adept at combining theory to the positive solution of practical problems, and is perhaps the Nation's most outstanding statistical consultant to the engineering and industrial sciences.

Now, the winner of the 1974 Wilks Medalist will be announced by Jerry Cornfield.

CUTHBERT DANIEL RECEIVES THE 1974 SAMUEL S. WILKS MEMORIAL MEDAL

The Presentation of the Award Made by Jerome Cornfield,
President of the American Statistical Association

The following citation was read:

> "To Cuthbert Daniel in recognition of his outstanding contributions to the applications of statistics in the sciences, for his researches in the novel uses and interpretations of factorial designs, and for his vigor in stimulating rewarding colloguy at statistical meetings."

ACCEPTANCE REMARKS OF CUTHBERT DANIEL ON RECEIVING THE SAMUEL S. WILKS MEMORIAL MEDAL FOR 1974

Mrs. Wilks, President Cornfield and fellow Statisticians: -

Like many others, I was helped by Sam Wilks from the moment of my first meeting him. He arranged for my first real contact with statistics in 1944 by asking a young member of his gorup at Princeton (A.Mood) to tutor me. Although younger than I, Sam was my Dutch Uncle and advisor many times over the next 25 years.

Wilks was the first of a long line of statisticians who have given me aid. I think that for my case the award should have added to its present inscription, the words "with a little bit of help from his friends." A great deal, actually. Allan Birnbaum (who both named and linearized the half-normal plot), Henry Scheffé (whose memoranda over 22 years I have just found in four file folders all labelled Scheffé, Current), Gus Haggstrom (who has worked manfully to keep me more honest than I had planned to be), and Fred Wood (who is a difficult colleague because of his insistence on getting everything right), are only four of the most helpful of my colleagues.

I thank you, then, and accept with full recognition of my past and continuing dependence on my statistical friends.