AD-A051 997 GEORGIA INST OF TECH ATLANTA SCHOOL OF INDUSTRIAL AN--ETC F/G 12/1
STUDIES IN SUPPORT OF THE APPLICATION OF STATISTICAL THEORY TO --ETC(U)
JUL 77 D C MONTGOMERY, H M WADWORTH DAAG39-76-C-0085

UNCLASSIFIED NL

1 OF 1
AD
A051997

END
DATE
FILMED
5-78
DDC

Final Summary Report

for the

United States Army Operational Test and Evaluation Agency
5600 Columbia Pike
Falls Church, Virginia 22041

"Studies in Support of the Application
of Statistical Theory to Design and
Evaluation of Operational Tests"
DAAG39-76-C-0085

Conducted by

The School of Industrial and Systems Engineering
Georgia Institute of Technology

Douglas C. Montgomery	Co-Principal Investigator
Harrison M. Wadsworth	Co-Principal Investigator
Leslie G. Callahan, Jr.	Faculty Associate
Russell G. Heikes	Faculty Associate

July, 1977

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 6	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Studies in Support of the Application of Statistical Theory to Design and Evaluation of Operational Tests. ~~(report + four annexes)~~		9 5. TYPE OF REPORT & PERIOD COVERED Final summary rept.,
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) 10 Douglas C./Montgomery Harrison M./Wadworth		8. CONTRACT OR GRANT NUMBER(s) 15 DAAG 39-76-C-0085 ew
9. PERFORMING ORGANIZATION NAME AND ADDRESS School of Industrial & Systems Engineering Georgia Institute of Technology Atlanta, GA		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 65101A
11. CONTROLLING OFFICE NAME AND ADDRESS US Army Operational Test & Evaluation Agency 5600 Columbia Pike Falls Church, VA 22041		11 12. ~~REPORT DATE~~ Jul 1977
		13. NUMBER OF PAGES 88
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 12 89p.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Evaluation	Training level	Sample size
Operational testing	Statistics	
Bayesian Theory	Multivariale statistics	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report is a summary report of four studies in support of the application of statistical theory to design and evaluation of operational tests. The four topics are: — over

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

409 390 — LB

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

a. "A Methodology for Determining the Power of MANOVA when the Observations are Serially Correlated," by Norviel R. Eyrich, CPT, Artillery.

b. "An Application of Multiple Response Surface Optimization to the Analysis of Training Effects in Operational Test and Evaluation," by Vernon M. Bettencourt, Jr., CPT, Artillery.

c. "A Cost Optimal Approach to Selection of Experimental Designs for Operational Testing under Conditions of Constrained Sample Size"; and by Sam W. Russ, MAJ, Signal Corps.

d. "An Application of Bayesian Statistical Methods in the Determination of Sample Size for Operational Testing in the US Army," by Robert M. Baker, CPT, Infantry.

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Final Summary Report

for the

United States Army Operational Test and Evaluation Agency
5600 Columbia Pike
Falls Church, Virginia 22041

"Studies in Support of the Application
of Statistical Theory to Design and
Evaluation of Operational Tests"
DAAG39-76-C-0085

Conducted by

The School of Industrial and Systems Engineering
Georgia Institute of Technology

Douglas C. Montgomery	Co-Principal Investigator
Harrison M. Wadsworth	Co-Principal Investigator
Leslie G. Callahan, Jr.	Faculty Associate
Russell G. Heikes	Faculty Associate

July, 1977

Table of Contents

I. Nature of the Research Program

A. Background: The School of Industrial and Systems Engineering of the Georgia Institute of Technology began to offer Operations Research/ Systems Analysis courses at the graduate level in the mid-1950's. A small number of officers and civilians from the Department of Defense who were pursuing graduate degrees in established areas enrolled in these courses. In 1969 the U.S. Army developed a core curriculum for a formal graduate program in OR/SA, and selected Georgia Tech as one of the two civilian institutions for concentrated use in meeting Army graduate educational needs in this area. In 1972 the School was authorized to award a graduate degree in operations research, MSOR. A number of joint reviews have been made in order to improve the Army OR/SA program requirement. The latest was in November 1976. Sixteen Army personnel entered the program in 1969, and by 1973, the program had peaked with 35 students in residence with approximately 20 graduating each year. Since the mid-50's over one hundred officers have received graduate degrees with heavy emphasis on OR/SA methodologies. At present 15 are in residence with a forecasted level of 30 in residence and an output of 15 a year.

B. The Theses Problem

At the academic instructional level, methodological course work is inextricably interwoven with application and research activities. For most Master's degree candidates, the identification and definition of a thesis topic of interest both to the student and to his research advisor requires a disproportionate amount of time when compared with the course requirements or actual thesis research. One of the important objectives to be realized in this program is the development of readily available research topics relevant to Army needs and objectives and potentially interesting to Army personnel, and to competent, involved research advisors. These availabilities are critical if the Army personnel are to complete an acceptable thesis within the time constraint of their tenure in the program.

During the 1960's and early 1970's a number of informal contacts were made between students, faculty and Army agencies to generate relevant theses research areas and reliable data sources. A host of agency "shopping lists" for proposed theses were made available to Army students. These efforts proved largely unsuccessful, and less than one-tenth of the theses completed by Army officers prior to 1974 were related to Army needs and problems. This situation was summarized in an October 1973 letter from Dr. Wilbur Payne, then Deputy Under Secretary of the Army, to Georgia Tech approving the revised curriculum programs when he stated:

> "I was very interested in the comments you received from the officer students in response to your Proposal Review memorandum. Of particular interest were their remarks concerning the lack of adequate communication between the Army and students, and the resulting scarcity of appropriate military related thesis topics. This has for some time also been a concern of mine. I believe that something can be done to improve this situation, and would be delighted to work with the Institute toward that goal."

C. Contract Support For Army Theses

The first Army sponsored research which supported Army graduate students at Georgia Tech was provided under a contract from the Army Research Office from Jan. 1970 to 31 March 1972. Under the title of "A Research Program in Operations Research and Management Sciences," the scope of work under this contract called for a general research program with emphasis on research, development and engineering administration, and mathematical programming theory and applications. Specific tasks required that Georgia Tech:

1. Construct, and find procedures for the solution of operations research models in areas important to the Army;
2. Identify potential thesis topics and provide experience in model building and analysis to participants in the Army Operations Research Program;
3. Study the application of the models and procedures of military oriented OR models to civilian life.

This contract was funded at a level of $40,000 from the Army Materiel Command, and supported five Army theses. Three of these theses were oriented towards theoretical extensions, and only two were directed at the application of theory to solve Army problems. Consequently there was still a need for a better means to bring together students, faculty and Army agencies.

During the Fall of 1973 and Spring of 1974 a number of conferences and seminars were held between Georgia Tech faculty, students and Army representatives to improve the relevancy of thesis research. In June 1974 the Army Materiel Systems Analysis Agency contracted to support three officers during the year ending in the Fall of 1975. The contract was renewed and supported three more officers during 1976. These AMSAA contracts supported the officer students by providing special office space, leased computer terminals, and other logistic support at Tech, TDY travel funds, and data sources within the sponsoring agency. In addition the contracts also covered approximately 1/4 time salaries, overhead and limited travel for faculty members for efforts beyond what would otherwise be required for their faculty duties. Actual thesis topics were developed between the individual student, the faculty and the sponsor to assure relevance and academic quality and are listed below:

"An Application of Multivariate Statistical Methods in Developing Operational Usage Patterns for U.S. Army Vehicles," by Randall B. Medlock, Captain, Infantry

"An Analysis of Computer Algorithms for Use in Design of Helicopter Control Panel Layouts," by Sam D. Wyman, Captain, Armour

"An Application of Multivariate Statistical Techniques to the Analysis of the Operational Effectives of a Military Force," by James T. Baird, Captain, Infantry

"An Application of Time-Step Simulation to Estimate Air Defense Site Survivability," by James M. Rowan III, Captain, Air Defense

*"A Mathematical Predictive Model of Arm Strength," by Robert S. Lower, Infantry

"Optimum Assignment and Scheduling of Artillery Units to Targets," by Everett D. Lucas, Captain, Artillery

*Partially supported by Human Engineering Labs thru AMSAA

Shortly after award of the AMSAA contract in June 1974 negotiations began with the U.S. Army Operational Test and Evaluation Agency to direct the research efforts of Army officer theses research into the general area of Decision/Risk Analysis applied to Operational Tests and Evaluation with initial emphasis on complex command and control systems. Two separate contracts were awarded in the Fall of 1974 in the following subject areas:

1. "Study to Evaluate the Results of Operational Tests and Evaluation of Complex Command and Control Systems" DA39-75-C-0095

2. "Application of Decision/Risk Analysis in Operational Tests and Evaluation" DA39-75-C-0097

Literature search and problem definition in the two areas began in the Summer of 1974 even though the contracts were not awarded until Dec. 1974. They were conducted on a parallel basis with strong interaction between three faculty members and seven graduate students supported under each contract. Frequent seminars and conferences were held throughout the period until individual thesis topics were developed in January 1975. After the Phase I briefing for OTEA at Georgia Tech in February 1975, the individual officers worked independently with their own thesis advisor and committee until graduation in June 1975. A final summary report was made by the faculty at OTEA headquarters in September 1975. This report in both written and oral form discussed the problem, approach, and results of the individual theses and presented results and recommendations in a more general manner than that presented in individual theses which are cited below:

"A Comparison of the Applicability and Effectiveness of ANOVA with MANOVA for Use in the Operational Evaluation of Command and Control Systems," by Thomas N. Burnette, Jr., Capt., Infantry

"An Application of Fault Tree Analysis to Operational Testing," by Gordon Lee Rankin, Capt., Signal Corps

"A Methodology to Establish the Criticality of Attributes in Operational Tests," by Gary S. Williams, Capt., Armor

"An Application of Multivariate Discriminant Analysis and Classification Procedures to Risk Assessment in Operational Testing," by Edward D. Simms, Jr., Capt., Infantry

"An Application of Simulation Networking Techniques in Operational Test Design and Evaluation," by E. L. Brown, Major, Ordnance

"An Application of Bayesian Analysis in Determining Appropriate Sample Sizes for Use in U.S. Army Operational Tests," by Robert L. Cordova, Capt., Ordnance

"Finding a Minimum Risk Path Through a Network Using Resource Allocation Techniques," by Lawrence G. O'Toole, Capt., Armor

At the conclusion of the first year OTEA contract in 1975 it became apparent that it was impossible to clearly delineate work under two separate contracts from the perspective of literature searches, methodological bases and student or faculty efforts. Consequently the current contract was negotiated for 1975-1976 under the broader scope of "Studies in Support of the Application of Statistical Theory to Design and Evaluation of Operational Tests" with four independently developed tasks. The second chapter discusses how each of these tasks were developed, and the final chapter the results of the research in each task area.

II. Development of OTEA Research Area

This research effort has a dual objective. The first objective is to conduct studies in the application of statistical methodology to designing operational tests and to evaluating the data generated from such tests. The second objective is to enhance the relevance of graduate thesis research undertaken by military officers, so that a higher correlation between their academic studies and the requirements of the Army will be obtained.

The research problem area was approached by first conducting a survey of the relevant technical literature. Both the current open scientific literature and reference material available through DDC and OTEA were evaluated. A series of group and individual meetings between project faculty and the officer-students involved in the program were conducted. The purpose of these meetings was to acquaint the officer-students with the general problem area, to discuss previous research effort both in related fields and conducted specifically for the DOD, and to develop specific proposals for current research related to the general project objectives. The officer-student research proposals must have three features:

1. They must be directed towards a problem area of interest to OTEA, as outlined in the project task statement.
2. They must describe a project that constitutes a reasonable contribution to the profession, so that the requirements of a Georgia Tech Master's thesis are satisfied.
3. They must be within the general area of interest of available faculty and other resources currently available.

Subject to these guidelines, the individual research proposals were then developed by the four officer-students involved in the project. They were approved by the project faculty, and by the Associate Director for Graduate Studies of the School of Industrial and Systems Engineering. These officer-student research proposals were also sent to OTEA for evaluation and feedback.

The general project objectives were realized through the creation of four specific tasks. Each task was investigated by one officer-student. Task I was to apply the principles of small sample size statistics to the design and analysis of operational tests characterized by limited sample size. This task was investigated by Captain S. W. Russ, who developed an economic model for sample size allocation in a class of factorial designs. The procedure allows direct incorporation of total sample size constraints on the problem, so that total test resource limitations will not be exceeded. This methodology would be useful in test designs where all treatment combinations are not of equal interest to the test designer and a cost of experimentation can be allocated to each cell in the test design.

Task II was to apply the principles of multivariate statistical analysis, decision theory, and risk analysis in specifying risk levels associated with the design of operational tests and the evaluation of operational test results. This task was studied by Captain N. R. Eyrich. He investigated the power of analysis of variance type tests in the multivariate case, demonstrating a relationship between power of the test and associated risk. He considered the case where successive observation vectors were autocorrelated, as would often be the case when operational test data are of a time series character.

Task III was to apply the principles of numerical analysis, training evaluation, regression analysis, and systems analysis to the currently subjective assessment of unit training levels during operational testing. This general problem area was studied by Captain V. M. Bettencourt, Jr. He described a general methodology whereby training effects in operational testing could be evaluated and optimized through computer simulation. He also discusses the general role of computer simulation in operational testing. The methodology is demonstrated by applying it to a hypothetical operational test of a new main battle tank.

Task IV was to apply the principles of Bayesian and classical statistics to determine optimal sample size over an entire operational test. This problem was investigated by Captain Robert M. Baker. He developed a method of selecting sample sizes in operational testing through Bayesian statistical analysis. His procedure incorporates the use of prior information at each stage to reduce the required sample size at that stage. The prior information can either be of a subjective or an objective nature.

There is a strong continuity to the overall research effort. Two of the tasks, II and IV, are direct extensions of research conducted during the FY 1975 contract.

III. Review of Theses

"An Application of Multiple Response Surface Optimization to the Analysis of Training Effects in Operational Test and Evaluation," by Vernon M. Bettencourt, Jr., Captain, Artillery

The Problem

The relationship between systems effectiveness and crew/unit training has recently begun to receive increased emphasis in the Department of the Army. There are a variety of reasons for this increased interest. Establishment of the U.S. Army Training and Doctrine Command (TRADOC) has institutionalized the importance of training and doctrine by fixing responsibility at a high level of the Army command. Without the troop and equipment demands of a belligerent theater, the main mission of the Army transforms to training for the next belligerency. The increasing cost of systems combined with a federal budget squeeze necessitates increased combat effectiveness from fewer weapons. The result of these factors is increased interest in training.

TRADOC is the major proponent of training in the Army. Within the last year, operations research analysts at TRADOC have been examining training and weapons system effectiveness. A general model of systems effectiveness has been derived;

$$E = f(w,p,t)$$

where E is combat effectiveness expressed as a function of w the performance capability of the system, p the proficiency of the crew/unit manning the system, and t the tactic or technique of employment. Development Test (DT) results can often be utilized to measure and quantify w.

Results of Operational Tests (OT) conducted by OTEA, can also be utilized in determining w.

Some inconsistencies arise in the consideration of p in the above equation. A Department of Defense directive states that Operational Test and Evaluation will be accomplished by operational and support personnel of the type and qualification of those expected to use and maintain the system when deployed. Most OT's are conducted with troops/units selected to satisfy this directive and then trained either by the unit or Equipment Training Team in accordance with a training package prepared by OTEA and/or TRADOC. Training is accomplished at home station, at the test site, and at Military Occupational Specialty (MOS) producing schools if required. Having undergone such well supervised and concentrated training, it is not unreasonable to assume that the test personnel are atypical of Army users in proficiency on the system.

Another inconsistency in the above equation is the effect of the learning-forgetting curve on proficiency. That is, the influence of a training season or a period of concentrated training in a specific area, on proficiency followed by a forgetting slump. The training cycles of most tactical units approximate such a curve.

The weapons system effectiveness utilized by the ASARC and DSARC is that obtained from the DT and OT. The above equation states that variation in actual user proficiency will cause variation in systems effectiveness. That is, there is a Performance Gap between AMSAA data (E_D) and actual performance in the hands of tactical troops (E_A) as predicted by the model above. This predicted Performance Gap has been verified in actual weapons test. In May 1974, the U.S. Army Infantry Board (USAIB) test fired the M72A2 Light Antitank Weapon (LAW) against moving targets at varying ranges. A significant Performance Gap was uncovered by this test. The major problem encountered by the troops was a lack of proper training on the graduated lead sight for a moving target.

The implications of these variations in combat effectiveness for the national defense posture are profound. It is imperative that OTEA, functioning as a major source of data on weapons systems effectiveness to high level decision bodies, account for training levels in their OT reports and analysis.

Approach and Methodology

The objective of this research was to develop an improved methodology for optimizing a set of operational test and evaluation performance measures which are functions of training. The research consisted of analysis and adaptation of response surface methodology, multiple response surface optimization, and multiple objective optimization to the problem. The Geoffrion-Dyer Interactive Vector Maximal algorithm was reviewed in detail and adapted to the multiple response problem. The adapted algorithm was applied to previously optimized multiple response surfaces to demonstrate its utility.

Multiple response surfaces and the adapted optimization algorithm are related to OTEA by use of a Tank Duel Model computer simulation. The military application will consider:

1. The extension of an OT through computer simulation.
2. The effect of training on tested system effectiveness.
3. The optimization of pre-test and tactical unit training programs concerning the tested system when confronted with multiple objectives or criteria.

4. The role of the military decision maker in the interactive optimization process.

Computer Simulation in Operational Testing

Computer simulation is finding wide application as a predictive and investigative tool. Most major defense systems undergo a computer simulation in a tactical environment both before and after the issuance of the required operational capability (ROC) report. Simulation can provide useful pre-test and post-test information for each OT. An important consideration is that computer simulations and OT's are mutually supporting. OT's provide verified data inputs for the simulation. In return the simulation provides predictions of input data for OT's or further investigates OT output data.

Pre-test computer simulation can enhance the OT in three basic areas:

1. Examine the identified critical operational issues to assess their significance.
2. Develop or discover critical operational issues that have been overlooked.
3. Provided a sensitivity analysis to indicate the accuracy required of each measurement.

This information will be obtained at relatively little cost and with the utilization of no test troops or equipment. The OT will be initialized with useful information and critical operational issues will be verified or identified. Data requirements in the test plan will be refined.

Post-test computer simulation can contribute to the success of an OT in the following four areas:

1. Constraining the scope of operational field tests to manageable proportions by providing analytical means for test extension.
2. Extending the OT into areas which are currently infeasible (such as two-sided combat).
3. Corroborating the impact of the OT results.
4. Supplying much needed operational performance inputs to other agencies utilizing simulation.

OT results can be combined with simulation results to fulfill the stringent requirements of statistical design of experiment methodology analysis.

Summary of Methodology

Response surface methodology is a branch of experimental design which is useful in the analysis of experiments where system optimization is the goal. Suppose that x_1 and x_2 are the independent variables in an experiment. The observed dependent variable or response y is a function of the levels of x_1 and x_2, say

$$y = f(x_1,x_2) + \varepsilon$$

where ε is a random error component. Usually the response y is the key measure of systems effectiveness. If we denote the expected response by E(y), then the surface represented by $E(y) = f(x_1,x_2)$ is called a response surface.

We may represent the two-variable case graphically by drawing the x_1 and x_2 axes in the plane of the paper. Then plotting contours of constant response yields the response surface shown in Figure 1. In the typical application of response surface methodology, search or "hill-climbing" techniques are used to move from an initial (usually poor) estimate of the optimal x_1, x_2 to a more precise final estimate of the optimal x_1, x_2.

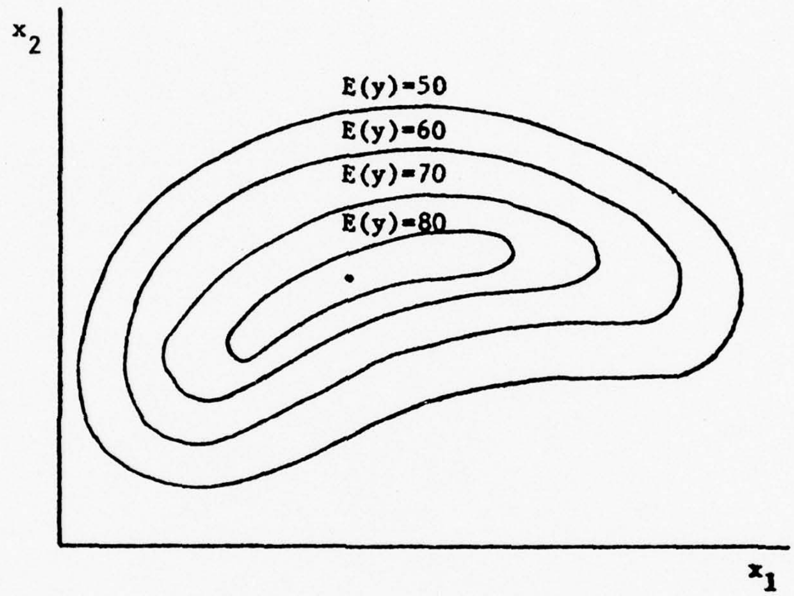

Figure 1. A Typical Response Surface

The true response surface is usually unknown. Therefore, the experimenter must find a suitable approximation for this unknown response surface. Graduating polynomials are the most widely used class of approximating function. These polynomials are fit to output data generated from the simulator. At the initial stages of a response surface study, when we are likely to be far from the optimum, first-order (linear) polynomials are usually employed. The method of steepest ascent is then applied, which allows the experimenter to move to a region more likely to contain the optimum. As we approach the optimum, a second-order (quadratic) polynomial is usually required to provide a satisfactory approximation to the true response surface. Optimization methods derived from the calculus are then used to obtain a more precise estimate of the optimal levels of the independent variables. For a detailed description of this methodology, see references [40] and [41] of the original thesis.

In most operational tests, the analyst is interested in several responses or measures of effectiveness. These problems can be structured as multiple objective or multiple response problems. This research surveys the literature on multiple response problems, classifying it into three general areas:

1. Graphical superposition methods
2. Adaptations of single-response mathematical programming methods
3. Interactive goal programming methods.

This latter approach is very new. An approach to the problem based extensively on the Geoffrion-Dyer Interactive Vector Maximal algorithm is given.

Description of the Methodology

Let $f_1(\underline{x})$, $f_2(\underline{x}),\ldots,f_n(\underline{x})$ be distinct response functions that represent the measures of effectiveness of interest in the operational test, and $\underline{x}$ is a vector of independent variables that are controllable by the test designer. The elements of $\underline{x}$ could include training variables or factors. The methodology maximizes

$$U = \alpha_1 f_1(\underline{x}) + \alpha_2 f_2(\underline{x}) + \ldots + \alpha_n f_n(\underline{x}) \tag{1}$$

U is viewed as a utility function formed by combining the individual response factions and $\{\alpha_i\}$ are a set of constants. If the $\{\alpha_i\}$ were known, any convenient nonlinear programming algorithm could be used to maximize U. However, the $\{\alpha_i\}$ are in general unknown.

The Geoffrion-Dyer algorithm is an interactive procedure whereby the test designer is presented a series of ordinal comparisons relative to the several measures of effectiveness in his particular problem. By his choice of prefered outcomes from this series of comparisons, the weights $\{\alpha_i\}$ are determined. The details of the ordinal comparison procedure are given in Bettencourt's thesis, and will be illustrated in the example to follow. He has also provided a computer program that performs the weight determination and optimization process.

It is important to realize that the test designer views the entire problem in objective faction space rather than in the more confusing decision variable space. He is making tradeoffs of objectives with no distractions from the decision variables. He is also seeing a multitude of alternate solutions as he progresses through the procedure This is an educational process for the decision maker in the implications of his tradeoffs among objectives. There is no requirement for the decision maker to be familiar with mathematical programming. Also, the algorithm converges to an optimal solution. The decision maker may subjectively terminate the algorithm once he feels further iterations would yield minimal improvement. The thesis also describes some modifications to the basic algorithm that make it suitable for the response surface environment.

Demonstration of the Methodology

The above methodology was demonstrated by applying it to a hypothetical operational test problem. Subsequent to the cancellation of the Main Battle Tank 1970 (MBT70) acquisition program, the Army began development of the less costly MBT76. As one means of cost reduction, all factors of system effectiveness were considered rather than exclusive consideration of the MBT76 technological capabilities. The Project Manager (PM) felt that crew training could be of utmost importance in overall MBT76 combat effectiveness. Prior to OT II, he directed an analysis of the effects of crew training utilizing a computer simulation of a combat situation indicative of the European environment. The laser ranging and optical tracking of the MBT76 were sophisticated enough to negate any effect of training on weapon accuracy. Consequently the PM directed that mean time to fire the first round, mean time between rounds, and probability of sensing be studied as system factors affected by crew training. In this initial stage, he also directed that one scenario, an engagement between two tanks in the open at a range of 1000 meters, be analyzed to establish feasibility of the methodology. This scenario was representative of tank combat in the European theater.

This hypothetical study utilizes a modified version of the tank duel simulation program developed by the U.S. Army Materiel Systems Analysis Agency. This is a small-scale, two-sided model used to simulate brief fire engagements between two armored vehicles. The model utilizes a stationary defending vehicle (blue) that fires first at a fully-exposed attacker vehicle (red). The engagement ends when a kill occurs or when a predetermined time limit expires. The deterministic and stochastic input variables to the model are shown in Tables 1 and 2, respectively. The time of flight was based on the use of high explosive anti-tank rounds with a muzzle velocity of 3800 feet per second for the Blue tank and 2800 feet per second for the Red tank. The fixed time to fire accounts for the mechanical actions between rounds such as recoil and breech operation. Thus the firing times analyzed are human actions such as issuing a fire order, loading the round, and tracking the target. A complete listing of the FORTRAN program of this model is in Bettencourt's thesis.

Table 1. Input Variables

Input Variable	Value
Engagement Time (sec)	120.0
Blue Time of Flight (sec)	.86
Blue Fixed Time to Fire (sec)	7.0
Range (meters)	1000.0
Blue Rd Reliability	.85
Red Time of Flight (sec)	1.17
Red Fixed Time to Fire (sec)	7.0
Red Rd Reliability	.825

Table 2. Stochastic Input Variables (Normal Distributions)

Input Variable	BLUE Mean	BLUE Variance	RED Mean	RED Variance
P(Hit 1st Rd)	.75	.0025	.60	.0025
P(Rehit)	.85	.0011	.75	.0011
P(Hit | Sensing 1st Rd Miss)	.80	.0011	.7	.0011
P(Hit | Loss of 1st Rd Miss)	.775	.0017	.625	.0017
P(Kill | 1st Rd Hit)	.5	.0011	.45	.0011
P(Kill | Rehit)	.85	.0003	.8	.0003
P(Kill | Hit ∩ Sensing 1st Rd Miss)	.5	.0011	.45	.0011
P(Kill | Hit ∩ Loss of 1st Rd Miss)	.5	.0011	.45	.0011
P(Sensing)			.525	.0006
Time to Fire 1st Rd (sec)			8.5	.6944
Time to Fire Subsequent Rd (sec)			10.5	.6944

The objective of the experiment is to study the effect of Blue crew training on combat effectiveness. Three independent variables were chosen; mean time to fire the first round (x_1), mean time between rounds (x_2), and the probability of sensing a round (x_3). Based on crew performance experience, realistic ranges were chosen for the independent variables. Mean time to fire the first round, human action component, ranged between 30 and 8 seconds. Mean time between rounds, human component, ranged between 30 and 5 seconds. Probability of sensing ranged between .0 and .6. The Red probability of sensing is somewhat higher since the Red round has a lower muzzle velocity and, consequently, is easier to sense. The dependent or response variables initially chosen were the probability of Blue victory (y_1) and the expected number of Blue rounds fired (y_2). One scenario, an engagement between Blue and Red at 1000 meters with both tanks in the open was analyzed. This scenario is representative of tank combat in the European theater.

Initial experiments with the model were in the region $20 \leq x_1 \leq 30$, $20 \leq x_2 \leq 30$, and $0 \leq x_2 \leq .2$. This produced observed probabilities of Blue victory of $.3 \leq y_1 \leq .45$ and expected number of Blue rounds fired of $.6 \leq y_2 \leq .8$. This region is obviously one of low combat effectiveness. The method of steepest ascent was used to move to a region of the factor space where higher combat effectiveness measures would be observed. During this phase of the study, it was noted from statistical analysis of the coefficients in the fitted first-order regression model that the probability of sensing, x_3, had no effect on the two responses. Therefore, x_3 was eliminated from further analysis and set at the mean of its practical range (i.e., $x_3 = .3$). Apparently, at the specified range and with the given probabilities of hit and kill, the ability to sense a round is not critical. The engagement seems to be won on the speed of firing the first round and a second round if required. Given another scenario, it is not unreasonable to expect that x_3 would be significant.

The method of deepest ascent indicated that the true optimum is in the vicinity of the point $x_1 = 12$ sec. and $x_2 = 10$ sec. To improve this estimate of the optimum, a second order response surface analysis was conducted. A rotatable central composite design, shown in Table 3, was used to fit the second-order surfaces. The second-order response surfaces are, for the probability of Blue victory

$$\hat{y}_1 = 0.629 + 0.014x_1 - 0.0062x_2 - 0.001x_1^2 - 0.00024x_2^2 + 0.00015x_1x_2 \qquad (2)$$

and for the expected number of Blue rounds fired,

$$\hat{y}_2 = 1.684 + 0.0215x_1 - 0.0234x_2 - 0.0002625x_1^2 - 0.00124x_2^2 + 0.00135x_1x_2 \qquad (3)$$

A canonical analysis indicated that both of these surfaces contain maximums which lie outside the experimental region.

Table 3. Central Composite Design

x_1	x_2	y_1	y_2
8	5	.669	1.635
16	5	.581	1.315
8	15	.538	1.235
16	15	.460	1.021
12	10	.577	1.337
12	10	.585	1.380
12	10	.581	1.366
12	10	.573	1.332
12	10	.609	1.426
6.344	10	.591	1.408
17.656	10	.518	1.148
12	2.93	.617	1.504
12	17.07	.533	1.092

Response surface equations relating the design variables to training were developed from interviews with experienced armored officers. The approximating relationship between x_1, x_2 and hours of dry (no live *firing*) *training* (y_3), in the region of experimentation for Equations (2) and (3) was found to be

$$\hat{y}_3 = 87.2009 - 2.5556x_1 - 2.1667x_2 \quad . \tag{4}$$

The approximating equation for live training rounds fired (y_4), in the region of experimentation for Equations (2) and (3) was found to be

$$\hat{y}_4 = 107.30015 - 2.611x_1 - 2.9167x_2 \quad . \tag{5}$$

The cost of training (y_5), in the region of experimentation for Equations (2) and (3), based mainly on cost of rounds and of petroleum, oil, and lubricants, was computed to be approximately

$$\hat{y} = 9667.5135 - 234.999x_1 - 262.503x_2 \tag{6}$$

The objective now is to maximize combat effectiveness (y_1,y_2) while simultaneously minimizing crew training parameters (y_3,y_4,y_5).

We note that regardless of the values of x_1 and x_2, the expected number of Blue rounds fired is between one and two. This response is of minimal interest in comparison to the probability of Blue victory and the training parameters, and was eliminated from further analysis. The four remaining response surfaces are illustrated in Figure 3.

The Interactive Vector Maximal algorithm was applied to this problem. At the outset of the optimization phase, it was determined that no more than 50 hours dry training per crew, no more than 55 training rounds per crew, and no more than $5500.00 training cost per crew could be expended. Figure 4 illustrates the four iterations of the algorithm which results in an optimum point of $x_1 = 10.7$ secs and $x_2 = 8.2$ secs. Typical output from the interactive optimization program is shown in Figure 5. The results of the optimization algorithm predicted that training to this proficiency would result in a probability of Blue victory of .6099. The predicted training effort to arrive at this level was 41.9 hours of dry training per crew, 55.2 live rounds fired per crew, and a cost of $4982.62 per crew. To confirm these results, the tank duel simulation was run at these levels and 12 replicates obtained. A 90% confidence interval on the mean probability of Blue victory is

$$.5377 \leq E(y_1) \leq .6547,$$

which is supportive of the conclusions drawn from the multiple response surface analysis.

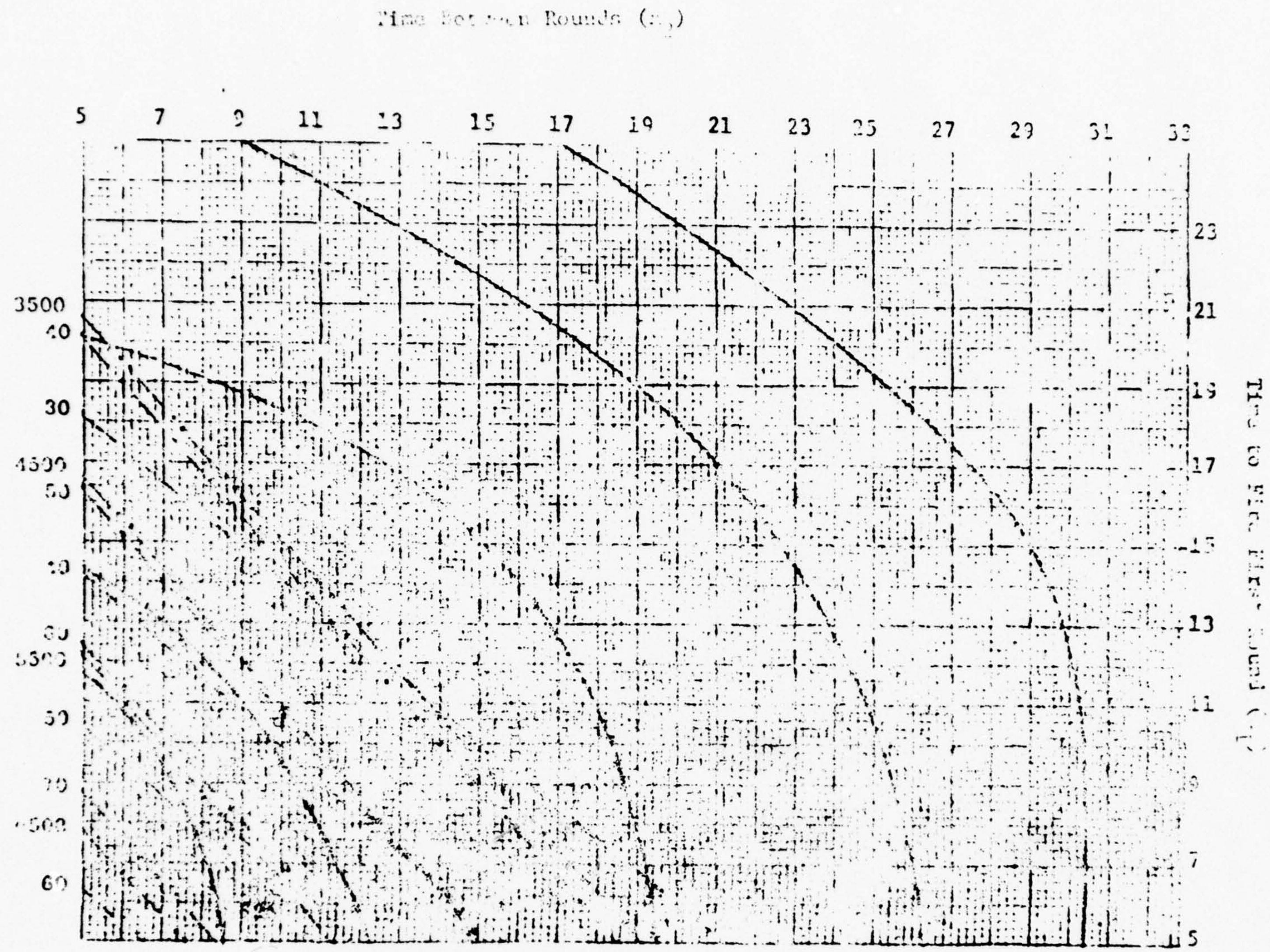

Figure 3. Response Surfaces for Tank Duel Simulation

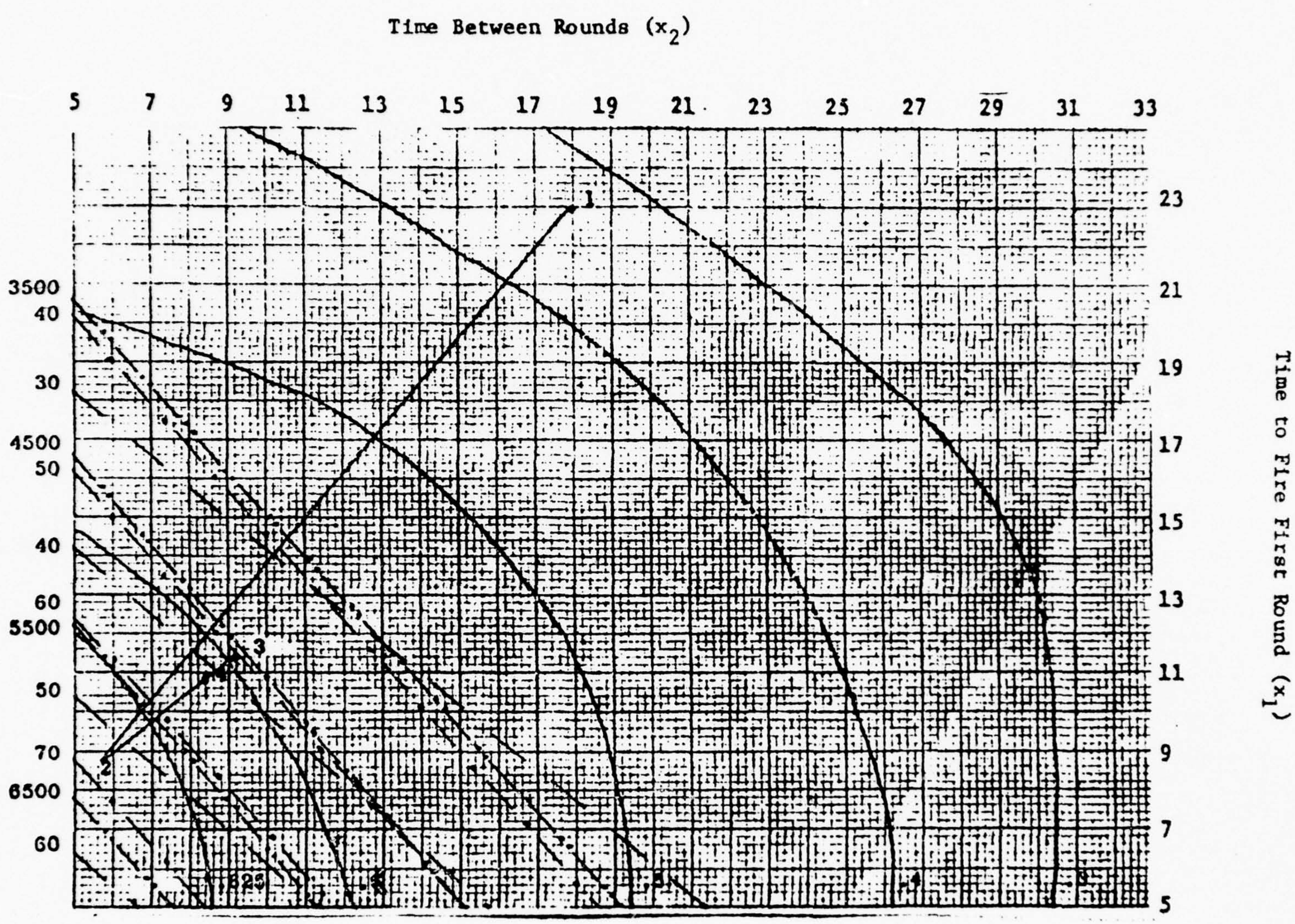

Figure 4. Four Iterations of the Algorithm in the Tank Duel Simulation

LEGEND FOR FIGURES 3 AND 4

$\hat{y}_1$ = probability of victory

$\hat{y}_3$ = training hours

$\hat{y}_4$ = training rounds

$\hat{y}_5$ = training cost in dollars

```
INPUT PERTURBATION OF F( 1), IN FAVORABLE DIRECTION
? .1
INPUT PERTURBATION OF F( 2), IN FAVORABLE DIRECTION
? .5
                              A                 B
P(VICTORY)                 .64600            .74600
E(BTL RDS)                1.59400           1.59400
  TRNG HRS               55.92260          55.92260
  TRNG RDS               76.25000          76.25000
 TRNG COST             6373.00000        6373.00000
WHICH DO YOU PREFER.   IF YOU ARE INDIFFERENT TYPE I.
? I
INPUT PERTURBATION OF F( 3), IN FAVORABLE DIRECTION
? -5.
                              A                 B
P(VICTORY)                 .64600            .74600
E(BTL RDS)                1.59400           1.59400
  TRNG HRS               55.92260          60.92260
  TRNG RDS               76.25000          76.25000
 TRNG COST             6373.00000        6373.00000
WHICH DO YOU PREFER.   IF YOU ARE INDIFFERENT TYPE I.
? I
INPUT PERTURBATION OF F( 4), IN FAVORABLE DIRECTION
? -5.
                              A                 B
P(VICTORY)                 .64600            .74600
E(BTL RDS)                1.59400           1.59400
  TRNG HRS               55.92260          55.92260
  TRNG RDS               76.25000          81.25000
 TRNG COST             6373.00000        6373.00000
WHICH DO YOU PREFER.   IF YOU ARE INDIFFERENT TYPE I.
? I
INPUT PERTURBATION OF F( 5), IN FAVORABLE DIRECTION
? -500.
                              A                 B
P(VICTORY)                 .64600            .74600
E(BTL RDS)                1.59400           1.59400
  TRNG HRS               55.92260          55.92260
  TRNG RDS               76.25000          76.25000
 TRNG COST             6373.00000        7373.00000
WHICH DO YOU PREFER.   IF YOU ARE INDIFFERENT TYPE I.
? I
```

Figure 5. Example of Interactive Optimization Computer Program.

BEST AVAILABLE COPY

```
INPUT PERTURBATION OF F( 1), IN FAVORABLE DIRECTION
? .1
INPUT PERTURBATION OF F( 2), IN FAVORABLE DIRECTION
? .5
                              A                B
P(VICTORY)                .64600           .74600
E(BTL RDS)               1.59400          1.09400
  TRNG HRS              55.92260         55.92260
  TRNG RDS              76.25000         76.25000
 TRNG COST            6373.00000       6373.00000
WHICH DO YOU PREFER.   IF YOU ARE INDIFFERENT TYPE I.
? I
INPUT PERTURBATION OF F( 3), IN FAVORABLE DIRECTION
? -5.
                              A                B
P(VICTORY)                .64600           .74600
E(BTL RDS)               1.59400          1.59400
  TRNG HRS              55.92260         60.92260
  TRNG RDS              76.25000         76.25000
 TRNG COST            6373.00000       6373.00000
WHICH DO YOU PREFER.   IF YOU ARE INDIFFERENT TYPE I.
? I
INPUT PERTURBATION OF F( 4), IN FAVORABLE DIRECTION
? -5.
                              A                B
P(VICTORY)                .64600           .74600
E(BTL RDS)               1.59400          1.59400
  TRNG HRS              55.92260         55.92260
  TRNG RDS              76.25000         81.25000
 TRNG COST            6373.00000       6373.00000
WHICH DO YOU PREFER.   IF YOU ARE INDIFFERENT TYPE I.
? I
INPUT PERTURBATION OF F( 5), IN FAVORABLE DIRECTION
? -500.
                              A                B
P(VICTORY)                .64600           .74600
E(BTL RDS)               1.59400          1.59400
  TRNG HRS              55.92260         55.92260
  TRNG RDS              76.25000         76.25000
 TRNG COST            6373.00000       7373.00000
WHICH DO YOU PREFER.   IF YOU ARE INDIFFERENT TYPE I.
? I
```

Figure 5. Example of Interactive Optimization Computer Program.

```
INPUT NUMBER OF POINTS TO SEE IN STEP SIZE
? 7
         .6460       1.5940       55.9226      76.2500     6873.3800

         .6279       1.5239       43.9040      65.9444     5945.4990

         .6096       1.4536       41.8853      55.6339     5217.9930

         .5850       1.3630       34.8667      45.3333     4290.4970

         .5560       1.2722       27.8481      35.3277     3162.9960

         .5232       1.1662       20.8294      24.7222     2235.4950

         .4360     _ 1.0500       13.8108      14.4166     1327.9940

INPUT NUMBER OF PREFERRED POINT
? 3
IF YOU WISH TO TERMINATE TYPE T.  OTHERWISE, TYPE C.
? -
OPTIMAL X
      10.6667      5.3333
       .256 CP SECONDS EXECUTION TIME
```

Figure 5. Continued

Discussion of Methodology

The methodology developed in this thesis is a very general set of techniques useful in the analysis and/or optimization of complex systems. While applied to a simulation model, the methodology is applicable to full-scale systems or processes as well. In general, experiments or tests are performed with one of two objectives; either (1) to learn how the factors of interest (independent variables) affect the output, or (2) to find the levels of the factors that optimize the output or response. This latter category of problems is addressed here.

The methodology would require that a simulation model of the system to be studied be available, and that the effect of training variables could be incorporated directly into this model. Alternatively, it could be applied to a live test, providing that resources to conduct training and optimize the test relative to the training variables were available. A limitation of the test is that it is difficult to deal with more than 5 or 6 independent variables. However, the problem of multiple measures of effectiveness is directly incorporated into the methodology.

There are a number of extensions and applications of this research that could be of interest in the operational testing environment. One possibility now currently under study is the use of nonlinear goal programming methods for the optimization or solution of problems involving multiple measures of effectiveness.

"A Cost Optimal Approach to Selection of Experimental Designs for Operational Testing Under Conditions of Constrained Sample Size," by Sam W. Russ, Jr., Major, Signal Corps

The Problem

The problem was that of selecting the specific design structure for an operational test under conditions of constrained sample size. The work was limited to univariate, quantitative, continuous, linear response models. The approach was to develop a mathematical model which has as its objective function, expected additional system cost (EASC). The EASC is defined as the sum of four cost elements. These are:

(a) Fixed cost of testing

(b) Sampling cost

(c) Expected cost due to a type I error

(d) Expected cost due to a type II error

Two classes of designs were considered, however the model would be applicable to any designs for which the above cost elements could be determined. The two classes of designs considered in this research were:

(a) Crossed, fixed factorial (including fractional factorial) designs

(b) Analysis of covariance designs.

Motivation of Research

The research was motivated by a problem of OTEA, stated by them and reported in the thesis as, "OTEA is continuously required to design and analyze the results of operational tests based upon small sizes whether the sample concerns numbers of prototypes, personnel, or trials. The effect (of a research project) would be directed at developing a methodology for designing, planning, and evaluating operational tests of limited sample size."

This problem motivated the researcher to develop a methodology for selecting the design of an OT based on a criterion of minimum expected additional system cost due to the entire testing procedure. The research thus addresses directly only the first part of the problem stated above. However, once the design is selected there is no particular difficulty in selecting the method of analysis. For the designs considered by this research, the method of analysis is well defined and well known.

Development of the Cost Model

The cost model developed is generally stated by the following equation

$$EASC = C_0 + \sum_{i=1}^{N} C_i + C_\alpha \alpha + C_\beta \beta$$

where EASC = Expected cost of additional testing

C_0 = Fixed cost of testing

N = Number of observations

C_i = Cost of sampling for observation i

C_α = Penalty cost of a type I error

C_β = Penalty cost of a type II error

α = Probability of a type I error

β = Probability of a type II error

The research considered the EASC necessary to make a decision regarding main effects only. That is the decision was of the type needed to determine the advisability of adopting a proposed device over a standard device or equipment. Thus the hypotheses tested were of the form:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 = d > 0$$

Here the null hypothesis, H_0, states that the proposed device is not significantly better than the standard for comparison (SFC). The alternative hypothesis states that the proposed device is better than the SFC by an amount d, the performance margin required for adoption of the proposed device. The required performance margin, d, must of course be stipulated in order to compute the probability of making a type II error, i.e. accepting the null hypothesis when the proposed device is better.

Figure 1 illustrates the errors and penalty costs in operational tests required for evaluation of the last two terms in the cost model. The other terms are self explanatory and would likely be well known for any specific test situation.

The cost model developed for a factorial design is given in the following equation,

$$\text{EASC} = C_0 + \sum_{\varepsilon_1=1}^{L_1} \cdots \sum_{\varepsilon_K=1}^{L_K} n_{\varepsilon_1 \cdots \varepsilon_K} C_{\varepsilon_1 \cdots \varepsilon_K} +$$

$$C_\alpha \alpha + C_\beta \beta(\alpha, \lambda, \nu_t, \nu_e)$$

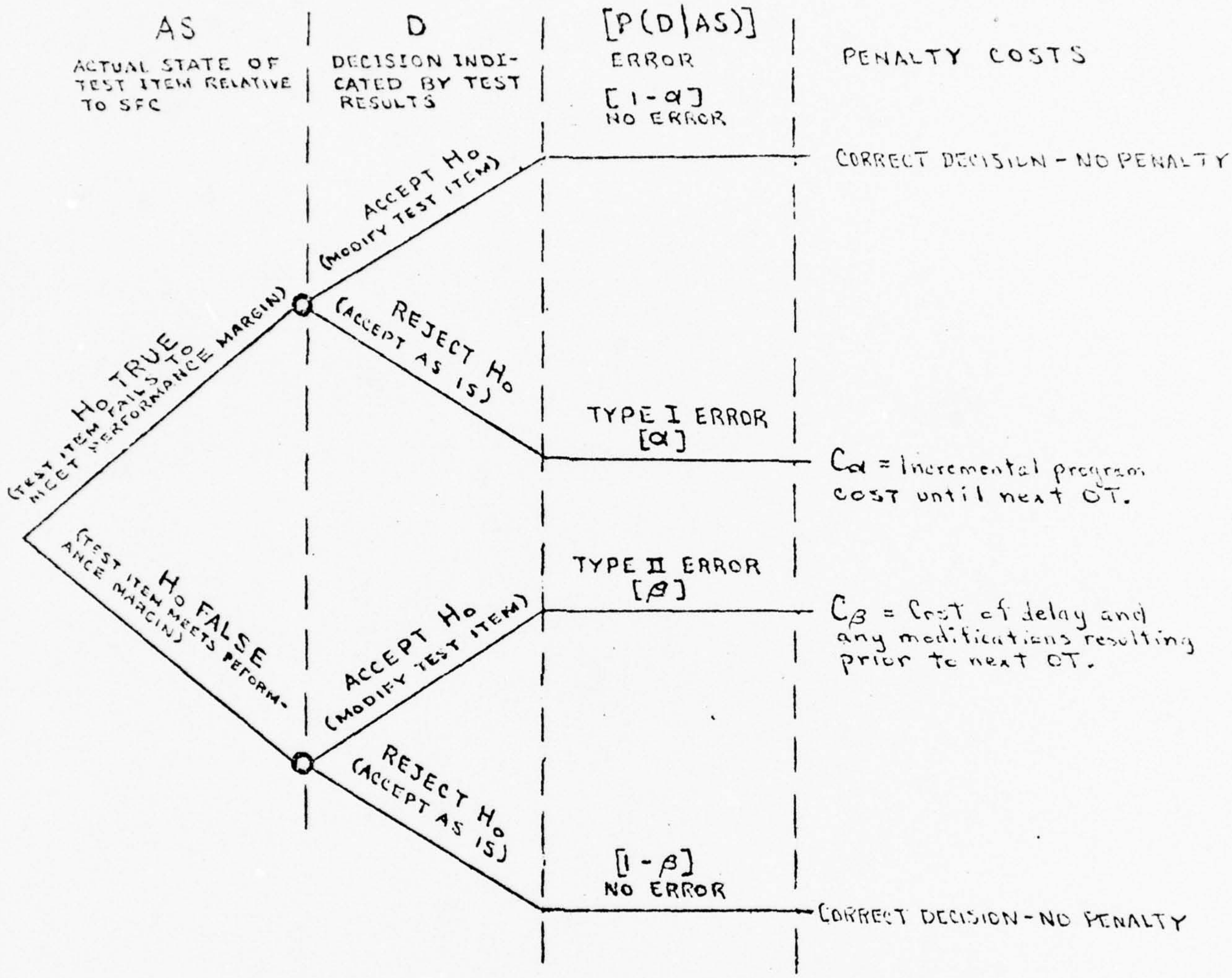

Figure 1. Errors and Penalty Costs in Operational Testing.

where

K = number of factors,

L_i = number of levels of the i^{th} factor, X_i,

ε_i = ε^{th} level of the i^{th} factor,

$n_{\varepsilon_1 \ldots \varepsilon_K}$ = number of observations in the $\varepsilon_1 \ldots \varepsilon_K^{th}$ cell,

$C_{\varepsilon_1 \ldots \varepsilon_K}$ = cost of an observation in the $\varepsilon_1 \ldots \varepsilon_K^{th}$ cell,

α = significance level,

λ = noncentrality parameter

ν_t = degrees of freedom between treatments,

ν_e = degrees of freedom for error.

The form of λ and ν_e will be determined by the specific type of factors involved and the pattern in which they are combined.

Parameter Estimates Needed. The following parameter estimates are needed prior to the design of OT-I, the first stage operational test. Their values would usually come from developmental tests conducted on the devices or from similar operational tests conducted previously. They may also be obtained from a series of pre-tests if this is feasible.

1. All cost coefficients
2. Error variance for the response variable in a completely random design
3. Correlation coefficients between the response variable and each covariate as well as all control factors
4. The ratio of the average variation of each factor about its fixed level to its population variance.

The estimates for subsequent test phases (OT-II, etc.) would be obtained from the first phase (OT-I).

The Optimization Problem. A 2^K completely crossed factorial design with all factors fixed and with a single covariate, Z, was used for illustration purposes. The cost optimization problem would thus be formulated specifically as the problem of selecting a design structure for operational tests with limited sample size. It was formulated as a constrained nonlinear optimization problem with EASC as the objective function and with sample size restrictions as the constraints.

The EASC Algorithm

An algorithm based on the derivation described in detail in the thesis was developed and is discussed in the thesis. This algorithm was programmed in FORTRAN IV for the Georgia Institute of Technology's CDC CYBER 70 computer. A complete listing of this program and description of the output options is contained in the Appendices of the thesis.

The algorithm was used to generate data for a 2^3 completely crossed design with one covariate based on hypothetical values of the cost coefficients and the primary parameters in order to test the program and empirically investigate the functional relationships between the objective function and the decision variables, $\bar{n}_1$, $\bar{n}_2$, N, and α. With the exception of Figure 3, all remaining illustrations in this section are based on these data.

Figure 2 illustrates, for two different values of α, the probability of a type II error, β, plotted as a function of the noncentrality parameter, λ, and the error degress of freedom.

Figure 3 shows several cost factors and rates of change of cost factors plotted as functions of $(T_1, T_2|\alpha,N)$, the individual treatment sample sizes when the total sample size and α are fixed. Since T_1

is bounded (due to sample size restrictions), only a portion of Figure 3 will actually occur. Also, since T_1 takes on only integer values, only discrete points within that segment can occur. Figure 4 illustrates these segments of the EASC curve which are obtained from the simulated data for several different values of N, the total sample size. It should be noted that increasing the value of N shifts the segment of the EASC curve from right to left with respect to Figure 3.

Figure 5 illustrates the effect of increasing the significance level, α. The figure shows that as α increases all of the curves in Figure 3 are compressed to the left. This is because as α increases, for fixed N, the rate of change of β with respect to T_1 increases.

EASC as a Function of N for Optimal $(\bar{n}_1, \bar{n}_2 | \alpha, N)$

Selecting for each value of N the optimal allocation of observations, $(\bar{n}_1, \bar{n}_2)$, results in the EASC values shown in Figure 7. Note that as the significance level increases, the optimal number of observations initially increases, then decreases. This is the result of the variations in the rate of change of β with respect to N for given values of α and N. Where this rate is high enough to off-set the increase in sampling cost, increasing N will reduce EASC. Once this rate decreases to the point where

$$C_\beta \frac{\Delta\beta}{\Delta N} < \frac{\Delta SC}{\Delta N}$$

then increasing N will increase EASC.

Summary of Procedure

The basic procedure for the design of an OT developed by this research is summarized by the following 14 steps.

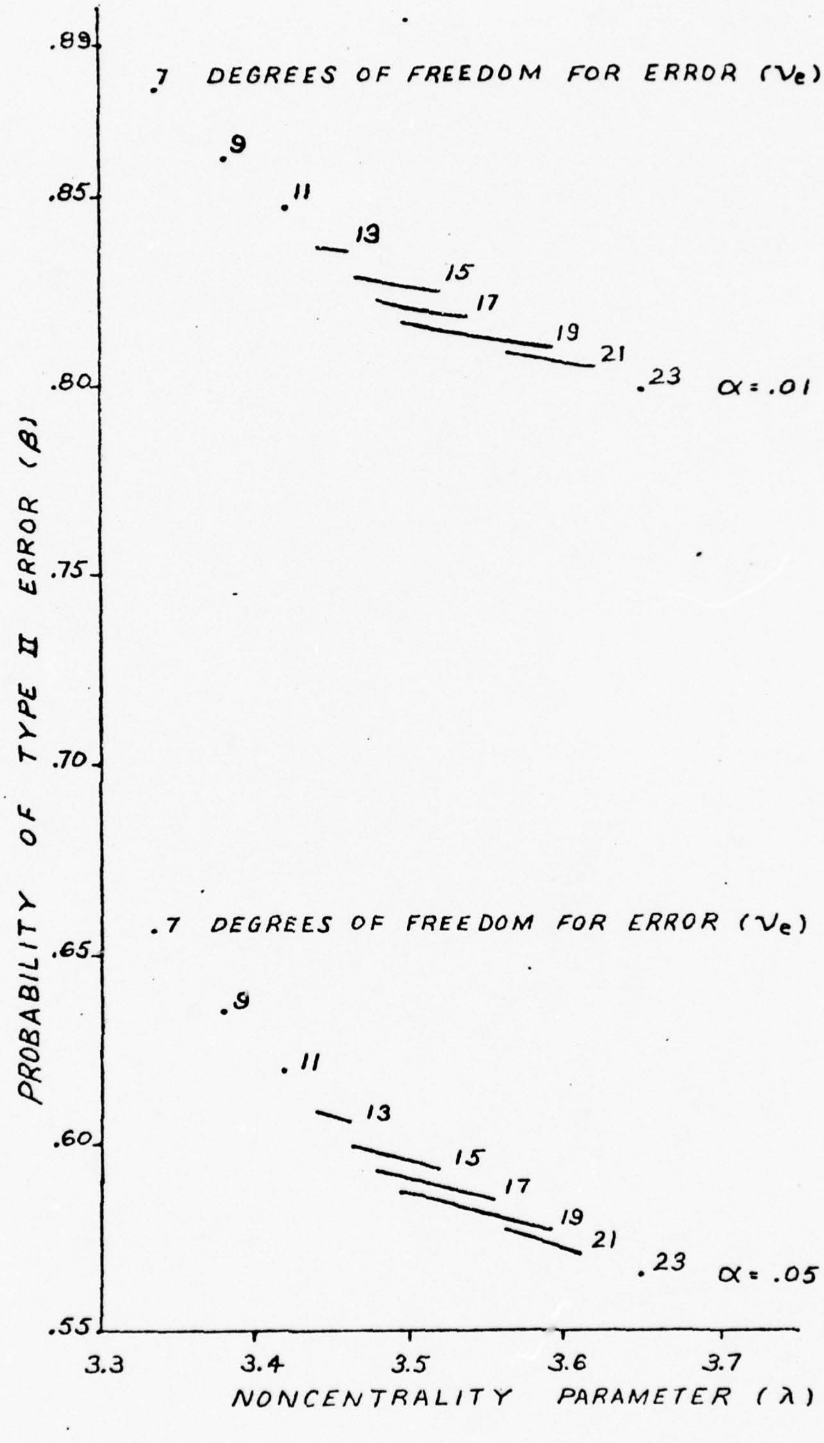

Figure 2. β as a Function of λ.

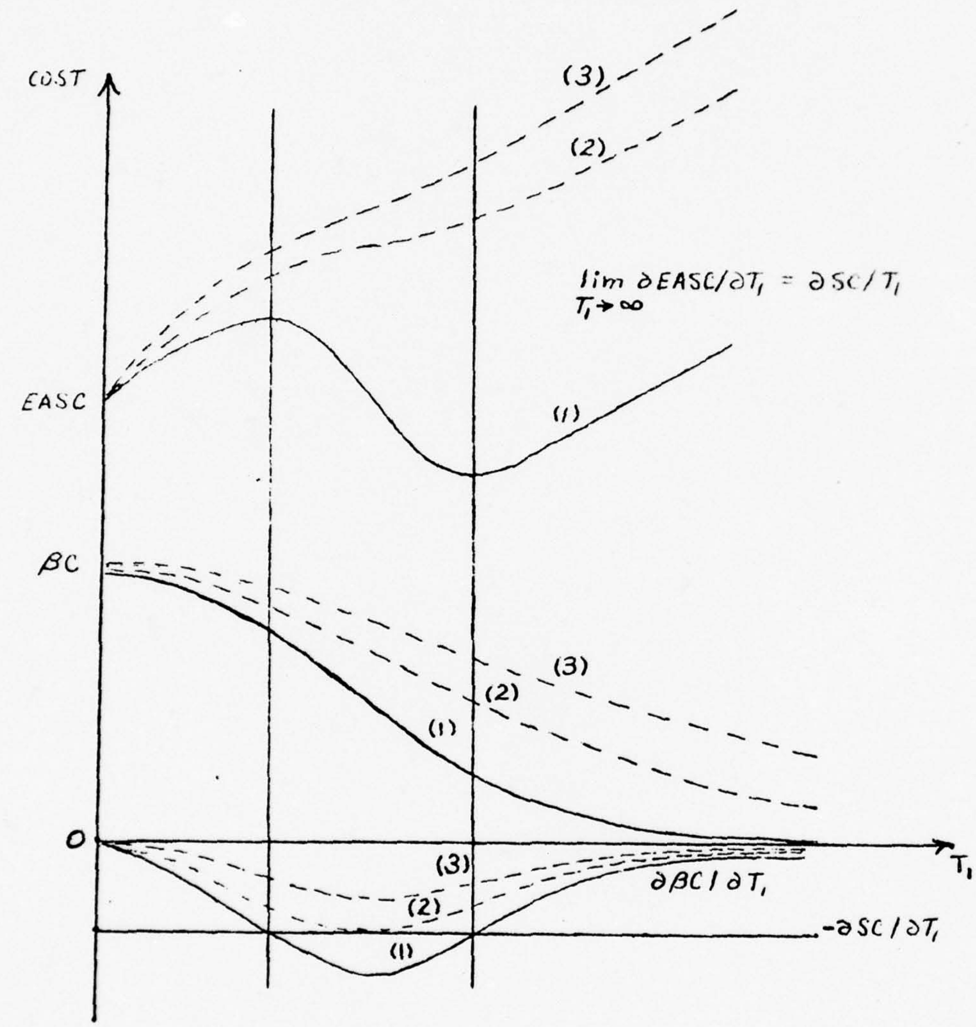

Figure 3. EASC as a Function of $(T_1, T_2|\alpha, N)$ - General Form.

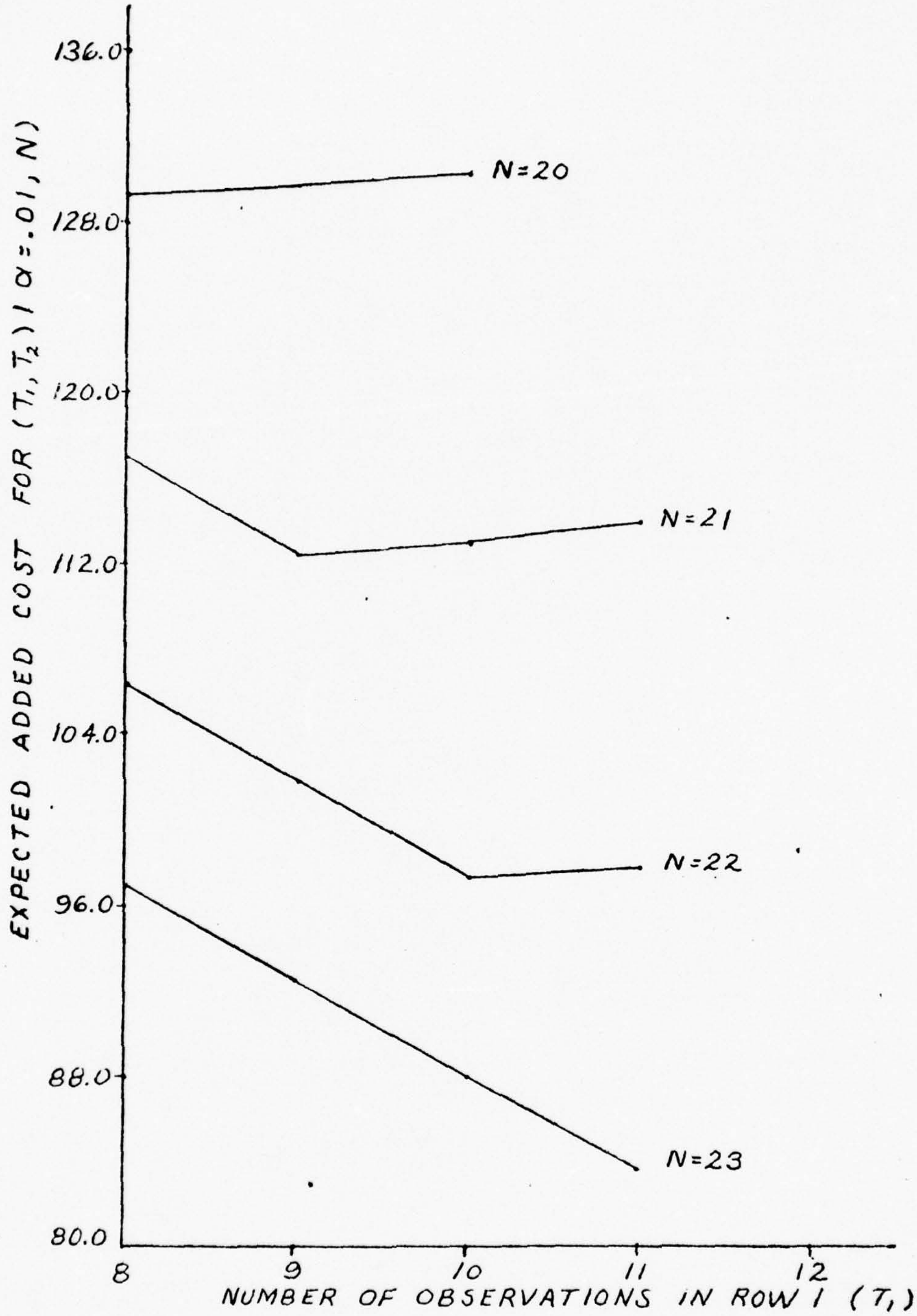

Figure 4. EASC for $(T_1, T_2 | \alpha = .01, N)$.

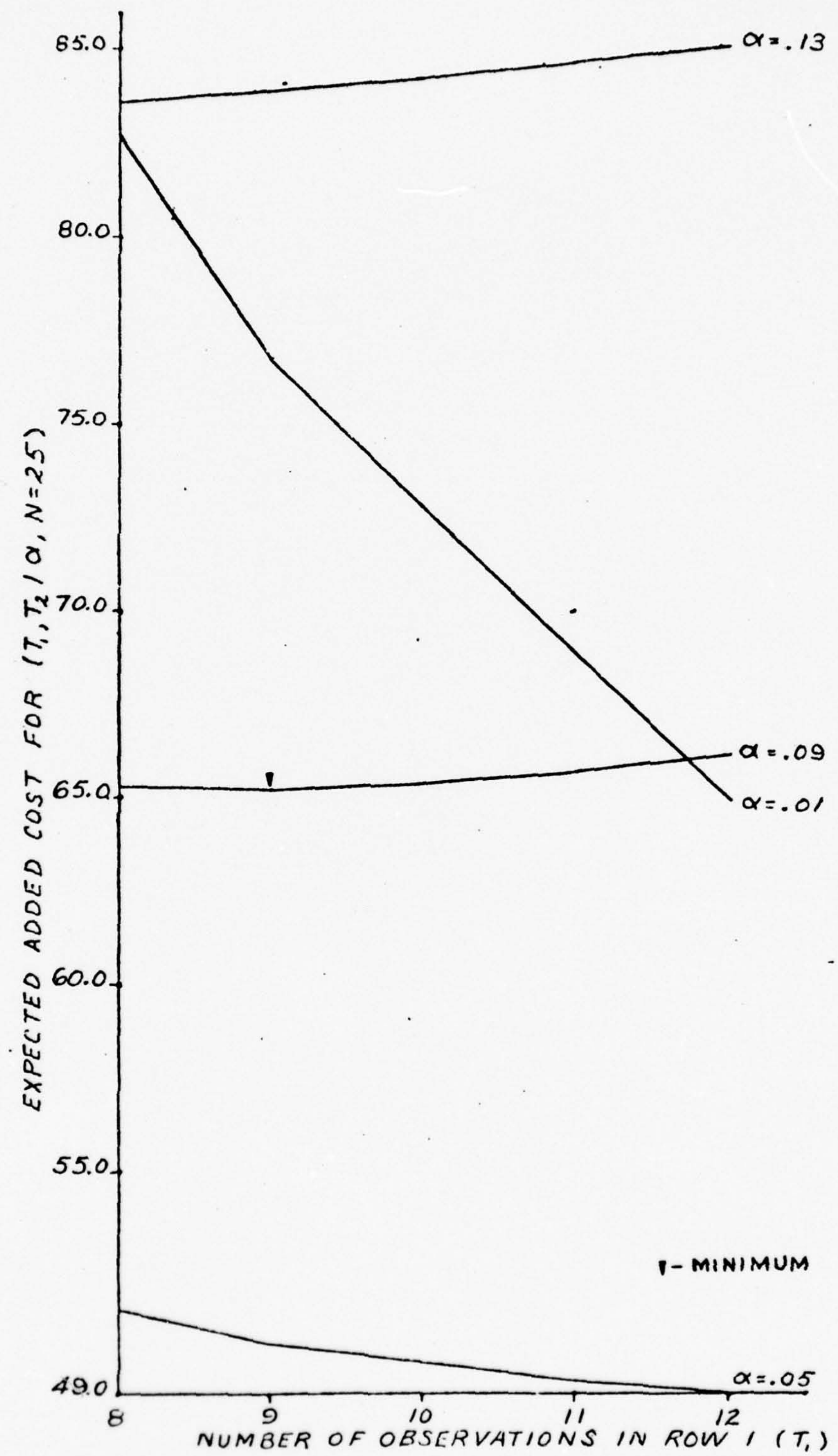

Figure 5. EASC as a Function of $(T_1, T_2 | \alpha, N)$ - Computed Example.

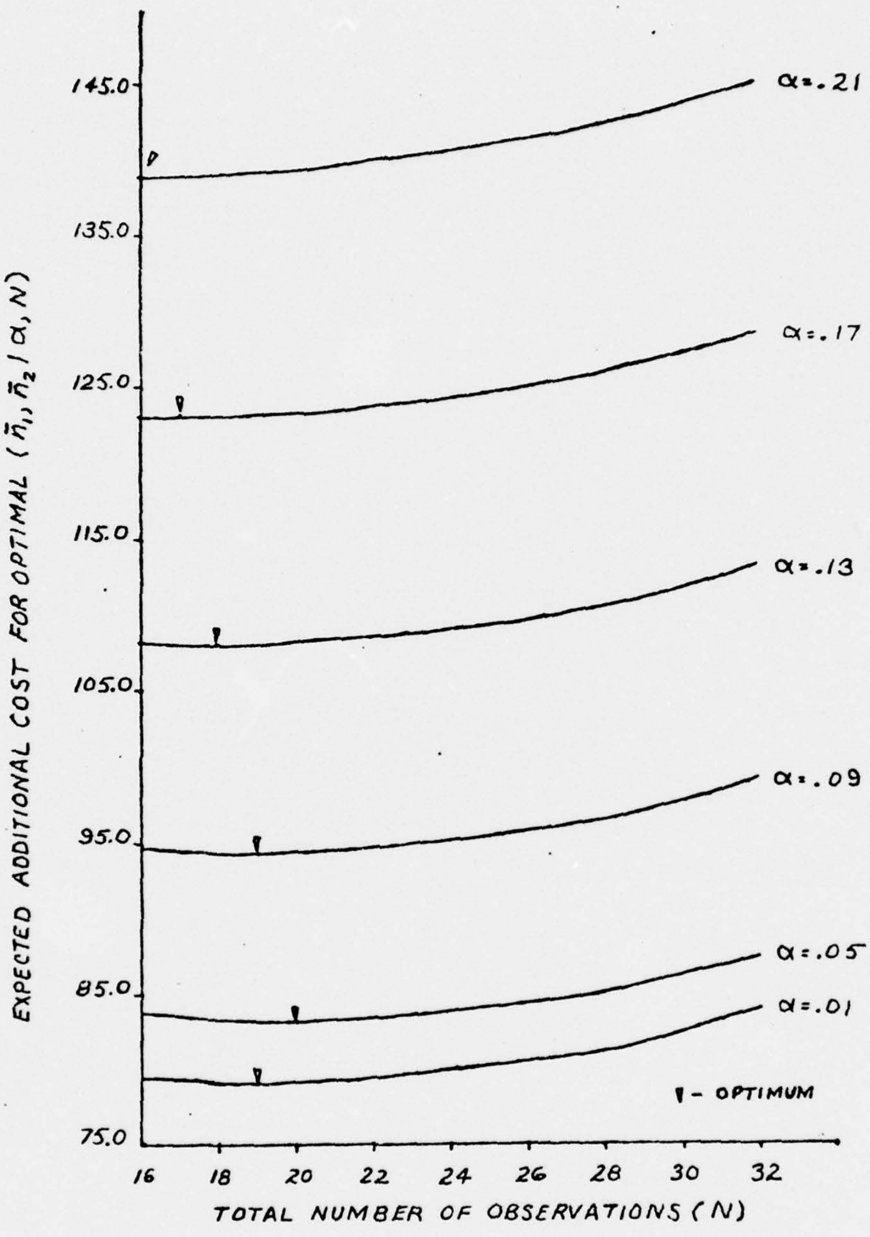

Figure 6. EASC as a Function of N for Optimal $(T_1, T_2 | \alpha, N)$.

1. Determine minimum number and type of factors to be considered and how they are to be combined to determine the conditions under which observations will be taken. The minimum number of factors will generally be dictated by the test issues.

2. Determine response variable to be measured (MOE). This must be a continuous variable.

3. Formulate the appropriate response model based on Steps 1 and 2.

4. Select the set of exact hypotheses to be used as the basis for optimizatation. Normally, this will be the null hypothesis of no treatment effect versus an exact form of the alternate hypothesis: the tested system exceeds the SFC by the required performance margin.

5. Determine the cost model to include estimates of all cost coefficients and primary parameters.

6. Formulate the optimization problem to include all constraints.

7. Apply the EASC algorithm to determine the number of observations to be taken in each row and their distribution, the level of significance, and the power of the test.

8. Use a random process to assign observations to specific cells and to determine the sequence in which observations are to be taken.

9. Vary the control limits on the levels of factors to determine the optimum control required if control is an-

ticipated to become a problem.

10. Repeat Steps 5, 6, and 7 for any alternatives which may be of interest to the experimenter such as addition of a blocking factor or covariate; an increase in the number of observations, if the previous optimal solution occurred at the upper limit of this constraint for one or both treatments; or fractional replication.

11. Select the optimal feasible alternative.

12. Begin experimentation.

13. Correct estimates of input parameters as test data becomes available.

14. Repeat Step 7 and other steps as necessary to determine the effect, if any, of the corrected parameter estimates on the optimal solution.

Demonstration of the Algorithm

The algorithm was demonstrated by a hypothetical example in which operational tests were to be designed to evaluate the overall military worth of a new ground-to-air tactical missile system, TAAMS, which is under development as a replacement for the HAWK missile system. The specific illustration concerns tests for the guidance system.

The critical issue for evaluation is the accuracy of the guidance system. Ambient temperature, altitude of target, and speed of the target are the most likely factors to have a significant effect on the accuracy. The maximum numbers of TAAMS and HAWK missiles that may be fired in each phase of the OT to evaluate the guidance system are 12 and 20 respectively. The measure of effectiveness (MOE) is stated as the mean miss distance from the target.

A 2^3 completely crossed factorial design was selected with ambient temperature, Z, treated as the covariate. The two independent variables were altitude of the target, X_2, and speed of the target, X_3. These two variables are treated as control variables while ambient temperature was considered a covariate since it could not be controlled. Factor X_1 is the missile type.

The test designer then uses the proposed procedure to determine the number of firings to be used for each missle type and their distribution among the 2^3 cells of the design. Estimates of cost coefficients and variability estimates required for use of the procedure are first obtained. These are shown in Table 1.

Figure 7 shows the results of the use of the EASC program with the input values listed in Table 1. The optimal values shown in Figure 8 were found to be, $\alpha = 0.29$, $N = 16$, $T_1 = 8$, $T_2 = 8$ and $\beta = 0.2207$. This resulted in an EASC of \$8.907 M.

Table 1. Initial Input Data for OT I

Cost Coefficients (million dollars)	Primary Parameters
$C_o = 1.000$	$\sigma_Y^2 = 4.000$
$C_\alpha = 10.000$	$d = .200$
$C_\beta = 10.000$	$\rho^2_{X_2Y} = .500$
$c_1 = .250$	$\rho^2_{X_3Y} = .500$
$c_2 = .100$	$\rho^2_{ZY} = .500$
	$\bar{\sigma}^2_{x_2} = 2.000$
	$\bar{\sigma}^2_{x_3} = 2.000$
	$\sigma^2_{X_2} = 20.000$
	$\sigma^2_{X_3} = 10.000$

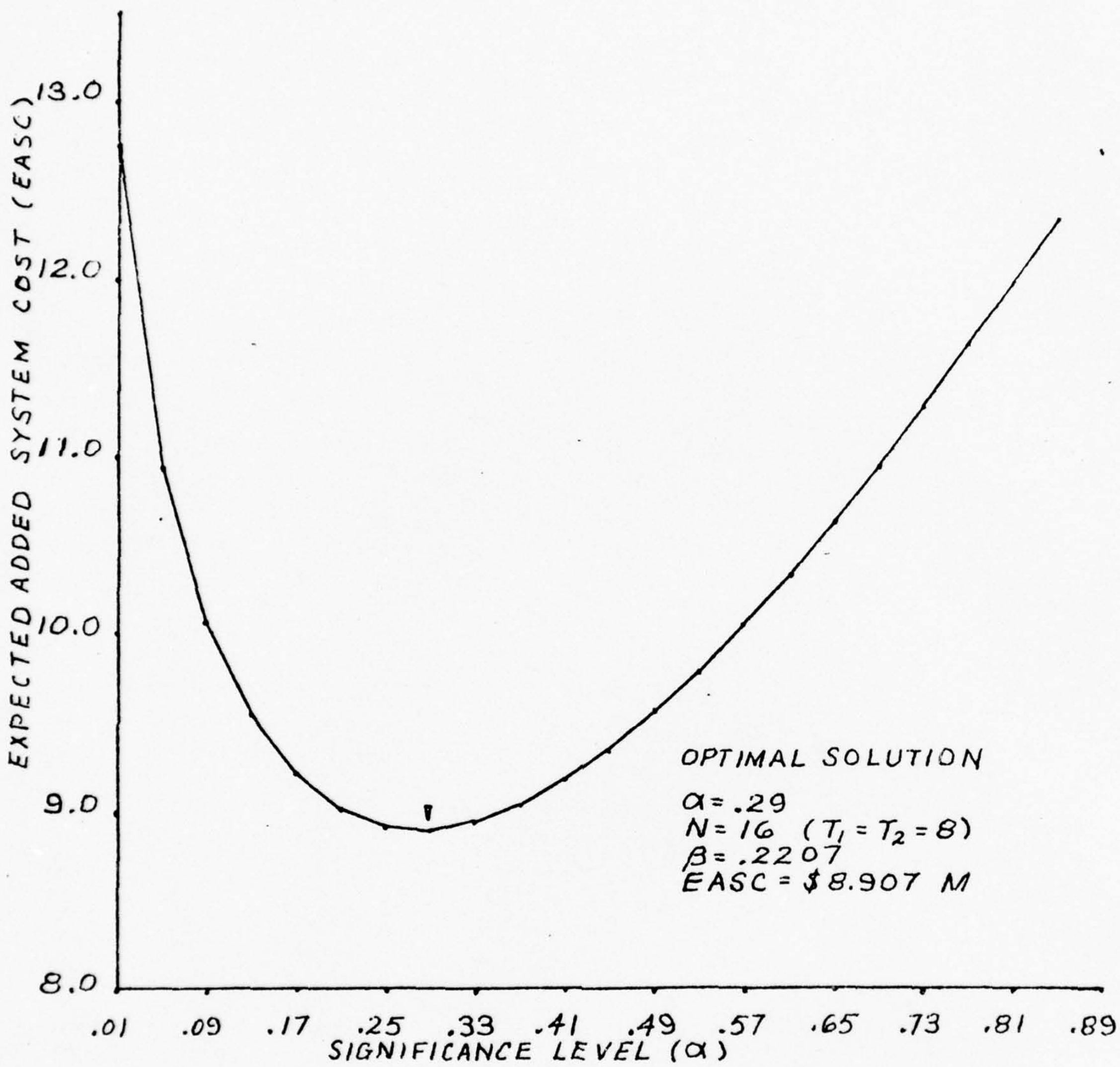

Figure 7. Optimal (EASC/α) for Initial OT I Design.

During a planning meeting a new control unit costing \$7,000 was proposed for the target drones. This control unit would reduce altitude variations by 50%. The new value of the control variance for altitude, $\bar{\sigma}^2_{X_2}$, was then inputed to the EASC program. All other parameters were left the same. This gave a new optimal solution of \$8.897 M, a reduction of \$10,000. This was used to justify the purchase of the new control unit and the first test phase was conducted.

The results of the first phase are used to revise the parameter estimates for subsequent phases. The input data for OT II are shown in Table 2 and Figure 8 illustrates the results of this run of the EASC program. It is to be noted that the error costs, C_α and C_β, are changed for the OT II tests. Following the evaluation shown in Figure 8, the performance margin, d, was reduced from 0.200 to 0.150. This necessitated a new program run and resulted in a new set of values. The new values were:

$\alpha = 0.21$

$\beta = 0.2583$

$N = 18$

$T_1 = 8$

$T_2 = 10$

EASC = \$12.074 M

For OT III, new estimates of the input data were determined. These included significant increases in C_α and C_β since an error would now become critical. Results of OT III will be used to decide whether to put the TAAMS missile into production. The new data are shown in Table 3 and the output is graphed in Figure 9.

Table 2. Initial Input Data for OT II

Cost Coefficients (million dollars)	Primary Parameters
C_o = 1.000	σ_Y^2 = 2.500
C_α = 20.000	d = .200
C_β = 15.000	$\rho_{X_2Y}^2$ = .700
c_1 = .250	$\rho_{X_3Y}^2$ = .600
c_2 = .100	ρ_{ZY}^2 = .650
	$\bar{\sigma}_{x_2}^2$ = .800
	$\bar{\sigma}_{x_3}^2$ = 1.400
	$\sigma_{X_2}^2$ = 20.000
	$\sigma_{X_3}^2$ = 10.000

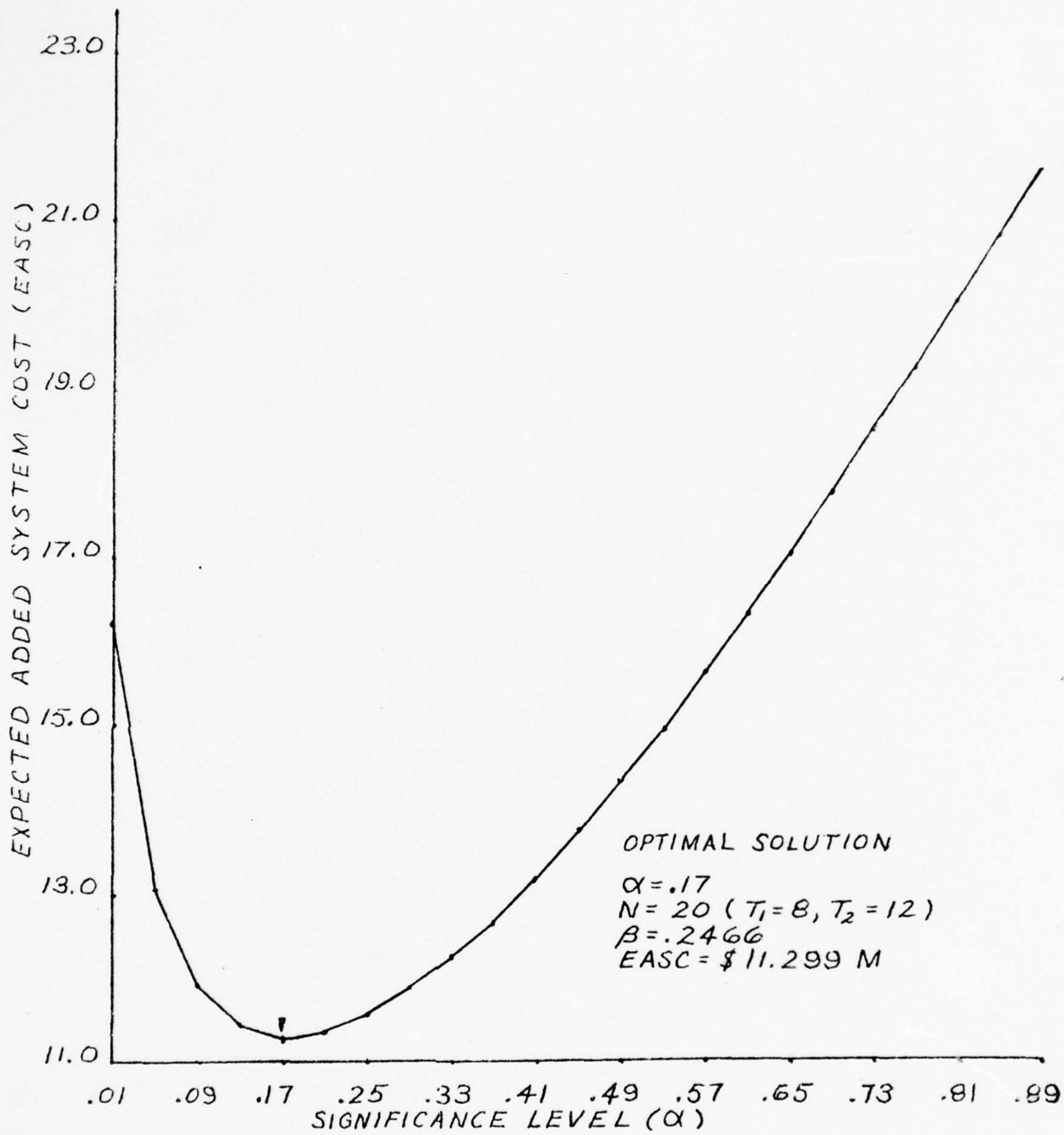

Figure 8. Optimal (EASC/α) for Initial OT II Data.

Table 3. Initial Input Data for OT III

Cost Coefficients (million dollars)	Primary Parameters
$C_o = 1.000$	$\sigma^2 = 2.500$
$C_\alpha = 500.000$	$d = .150$
$C_\beta = 150.000$	$\rho^2_{X_2 Y} = .600$
$c_1 = .350$	$\rho^2_{X_3 Y} = .600$
$c_2 = .100$	$\rho^2_{ZY} = .550$
	$\bar{\sigma}^2_{x_2} = .800$
	$\bar{\sigma}^2_{x_3} = 1.400$
	$\sigma^2_{X_2} = 20.000$
	$\sigma^2_{X_3} = 10.000$

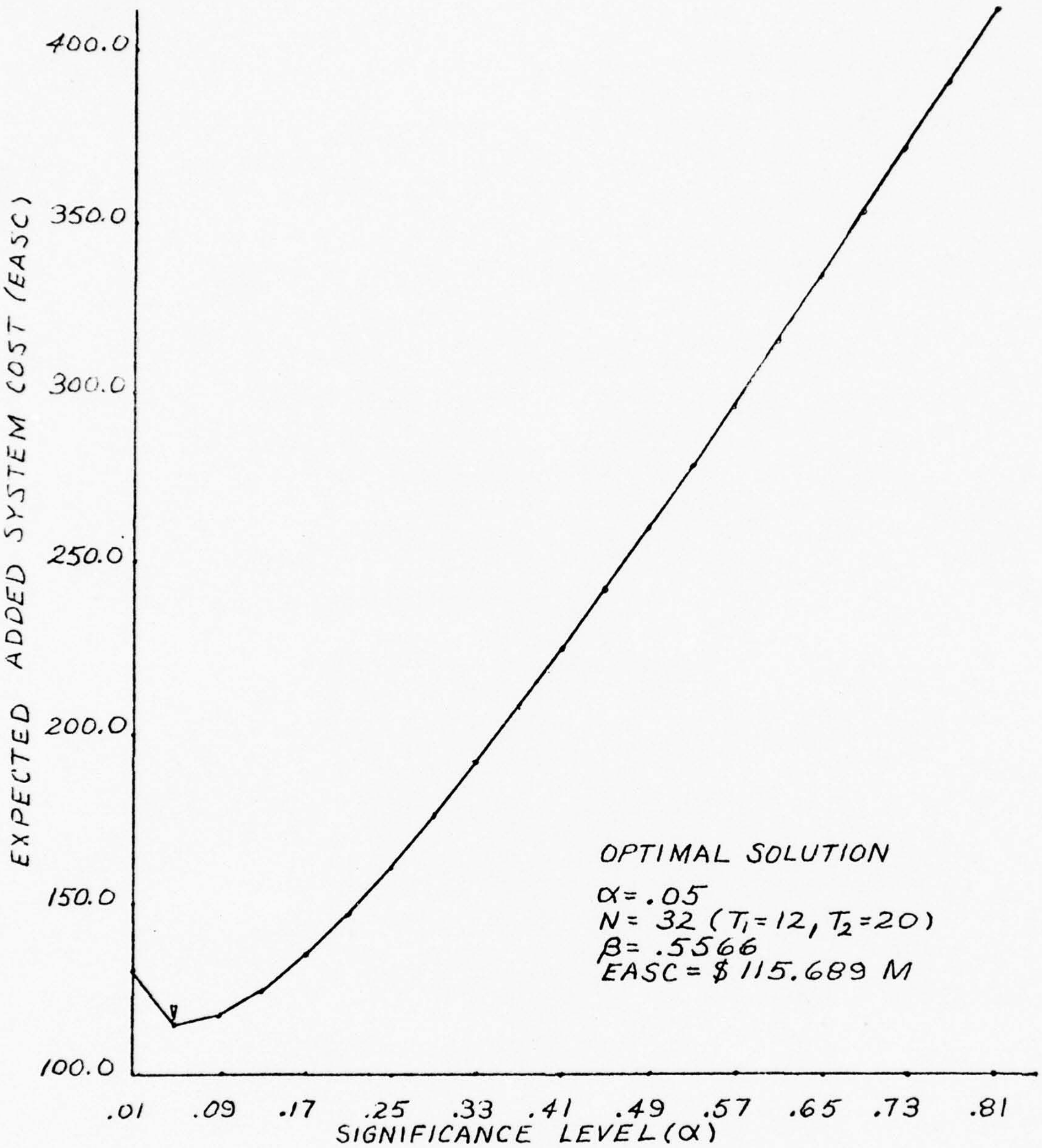

Figure 9. Optimal (EASC/α) for Initial OT III Data.

Prior to testing, a new speed control device is introduced on the target drones which reduces the variance in speed, $\bar{\sigma}^2_{X_3}$, by 28.5%. This new value is then used for the program and a new optimal solution is obtained. This is:

α = 0.05

β = 0.5527

N = 32

T_1 = 12

T_2 = 20

EASC = \$115.107 M

This reduced the expected cost by \$582,000. The cost of the 32 new drones is \$320,000 and therefore the new drones were justified.

The results above indicate using the maximum number of firings for both missile systems. Because of this result the program was run again to determine the effect on EASC of increasing the allowable number of HAWK missiles to 21. The results were observed to be:

α = 0.05

β = 0.5502

N = 33

T_1 = 12

T_2 = 21

EASC = \$114.831 M

This reduction of \$276,000 in EASC could be obtained by an expenditure of \$100,000 for the additional missile and thus the additional HAWK could be justified.

Evaluation of the Research

Use of the algorithm requires reasonably accurate estimates of the many required input parameters. This could be viewed as a disadvantage of the procedure. However some knowledge of these parameters must be obtained prior to the design of the test procedures by any method. Use of the EASC procedure would perhaps force the test designer to be more careful in his estimation procedure. In fact, by using the model with slight variations in these parameter estimates, he can evaluate the sensitivity of these initial estimates.

Extensions of this work should include a thorough study of the sensitivity of the parameter estimates. It might also include the introduction of multiple measures of effectiveness into the model. Also the use of discrete or qualitative MOE might be studied. The possibility of a more accurate objective function using a nonlinear model might also be studied.

However, without all of these extensions it is still recommended that OTEA adopt the EASC approach on a trial basis to evaluate their test design procedures.

"An Application of Bayesian Statistical Methods in the Determination of Sample Size for Operational Testing in the U.S. Army," by Robert M. Baker, Captain, Infantry

The Problem

The impetus for this study was provided by the interest of the U.S. Army Operational Test and Evaluation Agency (OTEA) in investigating the possible application of Bayesian statistical analysis and decision theory to sample size determination for operational testing. In the OTEA environment, the sample size problem becomes one of determining the minimum number of replicates required for each set of experimental conditions in order to produce sufficient sample information upon which to base statistically valid inferences concerning two competing systems. This problem can become quite complex since a single operational test may involve as many as a hundred measures of effectiveness (MOE).

In reviewing OTEA procedures, two areas of possible modification were identified. The first is concerned with making efficient use of all available data. The operational testing program is sequential in nature and, many times, the same measure of effectiveness may be examined in more than one test. When this occurs, the data from the previous test is sometimes used in the design of the subsequent test in that it serves as a basis for the formulation of hypotheses and as a source of variance estimates for sample size calculations. This data is not, however, being combined with the data obtained during later tests in the final statistical analysis. By not doing this, it is felt that valuable information is being wasted. It is believed that, if this information were used to its fullest extent, a reduction in the required sample size would be possible. One method of combining prior information with sample results is provided by Bayes' theorem.

The second area identified for possible improvement is concerned with the economics involved in experimentation. Presently the costs associated with proposed experiments are not directly considered in sample size calculations. Additionally, there is no evidence of a quantitative assessment of the expected value of the sample information to be obtained from a particular experiment. Considering this, it is doubtful that the money available for testing is being allocated to the various experiments in an optimal fashion.

Objectives

(1) To determine the sample size required to satisfactorily estimate the difference between the means of a measure of effectiveness for two competing systems when Bayesian analysis is used.

(2) To develop a procedure for the optimal allocation of resources to various experiments in the investigation of a system.

Methodology

The research associated with the first objective involved identifying the distribution of the difference between the means, $\tilde{u}$, of a MOE for two competing systems. It is assumed that the MOE follows a normal distribution with unknown mean and variance, and that the prior information concerning the difference of the means is in the form of a normal-gamma distribution. In this situation the combined information about the difference in the means is described by the Student-t distribution.

The criteria used to specify the acceptability of an estimate were

a) That the variability of $\tilde{\mu}$ be sufficiently small. This variability, $\breve{\mu}''$, was expressed as a fraction, s, of the variance of the prior distribution, $\breve{\mu}'$.

b) That $(1 - \alpha)\%$ of the probability distribution of $\tilde{\mu}$ fall within an interval of expected length, d'', which is centered at the expected value of $\tilde{\mu}$.

These criteria are equivalent but are both discussed as there may be differences in the conceptual attractiveness of each in the OTEA environment. Using criterion (a) and Stirling's first approximation the required sample size was found to be

$$n = \left(\frac{1}{s^2} - 1\right) n' \qquad (1)$$

where n' is a parameter of the normal-gamma prior distribution. This parameter value can be interpreted as the equivalent sample size of a previous experiment which generated the information contained in the prior distribution.

Using Stirling's second approximation a somewhat more complex relationship between n and s was developed; however an iterative procedure for solution was required. The percent difference in the solutions using the first and second approximations was investigated for various values of $\nu' = n' - 1$ and n. Results indicate that there is little difference when ν' is 35 or greater.

When criteria (b) is used the required sample size is found to be

$$n = \left[\frac{2t(\alpha/2,\ \nu'')}{E(d'')}\right]^2 \breve{\mu}' n' - n' \qquad (2)$$

where $t(\alpha/2, \nu'')$ is the percentage point of the Student-t distribution with ν'' degrees of freedom such that $P(t > t(\alpha/2, \nu'')) = \alpha/2$. This solution makes use of Stirling's first approximation. It also requires an iterative solution.

Criterion (a) and (b) are equivalent in that specifying a desired posterior variance is equivalent to specifying a length which contains $(1 - \alpha)\%$ of the distribution.

Sample Size Illustrations

The procedures developed were applied to OT II for the Lightweight Company Mortar System (LWCMS). The purpose of the test was to provide comparative data on the two types of mortars for assessing the relative operational performance and military utility of the LWCMS. One of the MOE under consideration in this test was the time required for an individual to complete the gunner's examination.

This MOE was previously examined during OT I. In that test, 14 individuals were given the gunner's exam using the 81mm mortar. They were then presented with two weeks of instruction on the LWCMS, after which they once more took the gunner's exam, this time using the LWCMS. The results of this test were available. The format for the experiment in OT II is the same. The sample size problem is to determine the number of individuals to be used in that experiment. The first solution procedure to be illustrated will use criterion (a).

The initial step in the procedure is to determine the value of the prior standard deviation of $\tilde{\mu}$. For notational purposes, the sample data relevant to the 81mm mortar will be denoted by X_{1i}, i = 1, 2, ..., 14 and that associated with the LWCMS by X_{2i}, i = 1, 2, ..., 14. To compute the value of $\sqrt{\breve{\mu}'}$ it is necessary to know n', v', and ν', the parameters of the prior distribution. Since this MOE was examined previously, the prior distribution for OT II may be equated to the posterior distribution of OT I. However, prior to OT I there was no internally generated data available; therefore, a diffuse prior distribution was appropriate. Thus, the posterior distributions associated with OT I are based solely on sample information. Considering this, the posterior parameters relative to OT I are computed using the OT I data as

$$m'' = m \equiv \frac{\Sigma_i D_i}{n} = 17.6 \text{ sec.}$$

$$v'' = v \equiv \frac{\Sigma(D_i - m)^2}{n''-1} = 2040.5 \text{ sec.}^2$$

$$n'' = n = 14$$

$$\nu'' = n''-1 = n-1 = 13$$

where

$$D_i = X_{1i} - X_{2i} \text{ .}$$

The above values may now be used as the parameters of the prior distribution relative to OT II.

The next step, then, is to calculate the value of the prior variance of $\tilde{\mu}$.

$$\breve{\mu}' = \frac{v'}{n'} \; \frac{\nu'}{\nu' - 2}$$

$$= \left(\frac{2040.5}{14}\right)\left(\frac{13}{13-2}\right)$$

$$= 172.25 \text{ sec}^2 \text{ .}$$

This produces a prior standard deviation of

$$\sqrt{\breve{\mu}'} = 13.12 \text{ sec} \ .$$

The fact that this MOE is again being considered in OT II implies the above standard deviation is too large to formulate meaningful conclusions regarding $\tilde{\mu}$. What specific value of the posterior standard deviation would be acceptable is something which must be determined by the OTEA test designers. To assist in this decision, Table 1 depicts the sample sizes required to produce various expected values for the posterior standard deviation.

Table 1. Required Sample Sizes for Values of the Expected Posterior Standard Deviation (in seconds)

$E(\sqrt{\breve{\mu}''})$	12.0	11.0	10.0	9.0	8.0	7.0	6.0	5.0	4.0	3.0	2.0	1.0
n	3	6	11	16	24	36	53	83	137	254	589	2396

The values of n were found by using equation (1) with

$$s = \frac{E(\sqrt{\breve{\mu}''})}{13.12} \ .$$

All that remains is for the analyst to select the desirable value for the expected posterior standard deviation and obtain the required sample size from Table 1.

Now consider the solution procedure which uses criterion (b), a

Bayesian interval on the posterior distribution. Based on the prior distribution, the length of an interval, centered on the mean, containing 90% of the probability is given by

$$d' = 2t_{\alpha/2,\nu''}\sqrt{\tilde{\mu}'}$$

$$= 2(1.761)(13.12)$$

$$= 46.21 \text{ sec} \ .$$

Suppose that it is desired to have the expected width of the Bayesian interval, with respect to the posterior distribution, be equal to

$$E(d'') = 20.00 \text{ sec} \ ,$$

then

$$E(d'')^2 = 400.00 \text{ sec}^2 \ .$$

Using equation (2)

$$n = \frac{(25_{.05,\nu''})^2(172.25)}{400}(14) - 14 \ .$$

To obtain a first approximation for n, $Z_{.05}$ is substituted for $t_{.05,\nu''}$ where Z follows the standard normal distribution. This gives

$$n = \frac{4(1.645)^2(172.25)}{400}(14) - 14$$

$$n = 51.26 \ .$$

Rounding this up to the next greatest integer gives an initial value for n of 52. Using this sample size, n'' would equal 66, with the corresponding value of $t_{.05,65}$ being 1.6686. Using these values and solving for n gives

$$n = \frac{4(1.6686)^2(172.25)}{400}(14) - 14$$

$$= 53.14 \ .$$

From this result it appears that the optimal n will lie somewhere between 52 and 54. Setting n equal to 53 and using the appropriate value for $t_{\alpha/2,\nu}$" gives

$$n = \frac{4(1.6683)^2(172.25)}{400}(14) - 14$$

$$n = 53.12 \ .$$

Therefore, a sample of size 54 would reduce the expected width of a 90% Bayesian prediction interval to 20.

Economic Considerations

In an environment where cost constraints become active it is necessary to make decisions as to where to allocate resources. For any particular MOE it is desirable to increase the sample size to the point where the incremental value of the last data point is equal to the cost of obtaining that data point. This implies that it is possible to define the value or utility, say U(-), of having a posterior distribution on $\tilde{\mu}$ with certain characteristics. The characteristic chosen for use in this study was s, the ratio of the prior variance to the posterior variance. It was also assumed that the cost of sampling, K_s, can be represented by a fixed portion, K_f, and a variable portion, K_r, so that

$$K_s = K_f + K_r n$$

where n is the sample size. The utility of the cost of sampling is then

$$U(K_s) = -K_s$$

The utility of any experiment, say e_n, is given by

$$U(e_n) = U(s) - K_s$$

where U(s) is the utility of achieving a given value of s.

Two different forms for U(s) were investigated. When s and utility are related linearly, we have

$$U(s) = as + b \ .$$

Using the relationship found between n and s in the previous section in U(s), differentiating with respect to n, and setting the result equal to zero yields

$$n = \left[K_r \left(- \frac{2}{a} \right) \left(n' \right)^{-1/2} \right]^{-2/3} - n' \qquad (3)$$

where a is negative.

Alternatively suppose that U(s) is of the form

$$U(s) = (1 - s)^C K_t$$

where K_t is some maximum allowable dollar amount for this MOE. Then

$$U(e_n) = (1 - s)^C K_t - K_s$$

Substituting for s and K_s and differentiating with respect to n gives

$$\frac{CK_t}{2} n'^{(1/2)} [1 - (n')^{1/2} (n' + n)^{-1/2}]^{C-1} (n' + n)^{-3/2} - K_r \qquad (4)$$

Search methods are necessary for funding the optimal value of n in this case.

Economic Examples

The solution procedure is illustrated using both types of utility functions described above. The same experiment used previously will be used for this illustration. In order to do this, however, several additional inputs are necessary, specifically, the budget constraint, K_t, the sampling costs, K_f and K_r, and the utility function, U(s).

To think of a budget constraint and a cost of sampling associated with a single MOE may be somewhat unrealistic. In practice, a single experiment will produce data on many different MOE. Most of the time, the only budget and cost figures associated with the test are aggregate amounts in the form depicted in Table 2. Therefore, rather than attempting to determine the sampling cost for a specific MOE and the total money available for testing that MOE, it may be much more realistic to allocate to each MOE some proportion of the aggregate budget and estimated costs. This is not currently being done, so it was necessary to approximate these values.

It is suggested that the proportion of the aggregate budget to be assigned to a specific MOE be commensurate with that MOE's relative importance. The OTEA already assesses the relative importance of MOE in qualitative terms. All that is required then is to quantify this assessment, perhaps through a series of weighting functions. It is not anticipated that this requirement would represent a major problem to OTEA test design personnel who have detailed information on the relationship between the data requirements and the operational issues being examined.

Since this type of information is not presently available, a very simplistic approach was taken to the allocation problem. Each of the MOE

Table 2. Total Cost Estimates (Direct Costs) [14]

Elements of Cost	Estimated Cost (In Thousands of Dollars)
1. Test Directorate Operating Costs	19.1
2. Player Participants	22.1
3. Test Facilities	30.0
4. Items to Be Tested	.5
5. Data Collection, Processing and Analysis	6.4
6. Ammunition	145.4
7. Pre-Test Training	2.1
8. Photographic Support	15.0
9. Other Costs	4.5
Total	245.1

was weighted equally in determining the individual budget constraint. Based on an imposed test budget constraint of $250,000.00, the individual budget constraint for each MOE, K_t, was derived to be $1,724.00.

The derivation of values for the fixed and variable costs was accomplished in a slightly different manner. The aggregate estimated fixed cost was defined to be the sum of all those costs in Table 2 except the costs of player participants and ammunition. This resulted in a total figure of $77,600.00. This figure was then divided by the length of the test in weeks to yield a fixed cost per week of $5,969.00. Using this weekly cost estimate, each phase of the test was assigned a fraction of the total estimated fixed cost based on the time required to conduct that particular phase. The fixed cost associated with each phase was then distributed equally among the MOE being examined in that phase. Table 3 presents the results of this process.

The variable costs are of two types, those associated with a sample size requirement for a certain number of <u>different</u> individuals and those associated with the requirement for the expenditure of a specified number of rounds of ammunition. Both of these variable costs were approximated by dividing the appropriate total estimated cost figures presented in Table 3 by the total estimated requirements for that resource. This resulted in a variable cost for personnel of $57.00 per week per man and a cost of ammunition of $13.00 per round.

Table 3. Allocation of Estimated Fixed Costs

Phase	Length of Phase (weeks)	Fixed Cost for Phase ($)	No. MOE Examined	Fixed Cost per MOE ($)
1. Training	2	11,938	28	426
2. Pilot Test	1	5,969	0	0
3. Field Exercise	3	17,908	73	245
4. Live Fire	6	35,815	36	995
5. Parachute Delivery Demonstration	1	5,969	8	746

The MOE of interest in this illustration is to be examined during the training phase so the fixed cost, K_f, is $426.00. The test design calls for using the same number of individuals throughout the training phase. Therefore, the variable cost, K_r, was derived by multiplying the cost per man per week by the number of weeks required to complete the training phase and then dividing the result by the number of MOE examined during this phase. This process resulted in a value of \$4.00 for K_r.

The above methods for approximating budget constraints and sampling costs are not necessarily being advocated for use by OTEA; they were used here to provide a starting point for the demonstration. This being accomplished, it remains to select an appropriate function for U(s).

The first case to be considered is that of a linear utility function. The form of this function is

$$U(s) = as + b \qquad \begin{array}{c} a \leqq 0 \\ 0 < s \leqq 1 \end{array}$$

Consider Figure 1 below, by varying the values of the parameters a and b, it is possible to represent U(s) by any negatively sloped straight line which intersects the s-axis between zero and one. This provides the decision maker with a rich family of linear functions from which to choose. The one chosen for this illustration is the one depicted in Figure 1.

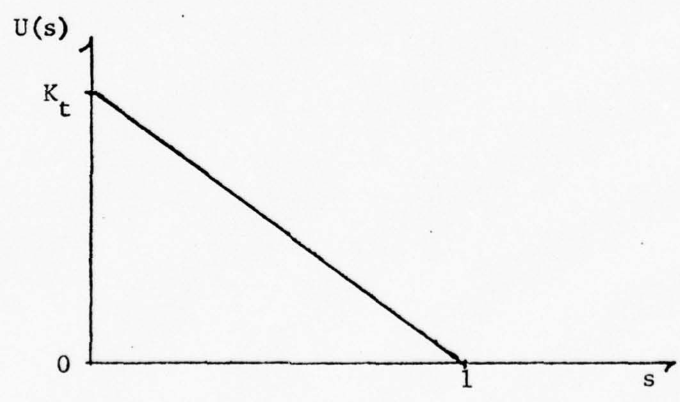

Figure 1. Linear Utility Function

The equation for this function is

$$U(s) = -K_t s + K_t = K_t(1-s)$$

Using this utility function and the budget constraint and sampling cost previously derived, the objective function becomes

$$U(e_n) = K_t[1-(n')^{-1/2}(n'+n)^{-1/2}] - K_f - K_r n .$$

The optimal value of n is found from

$$n = \left[K_r\left(-\frac{2}{a}\right)\left(n'\right)^{-1/2}\right]^{-2/3} - n'$$

$$= \left[4.00\left(\frac{2}{1,724}\right)\left(14\right)^{-1/2}\right]^{-2/3} - 14$$

$$= 88.55 - 14$$

$$= 74.5$$

This same analysis will now be conducted using two power function utilities. The first will be defined by

$$U(s) = (1-s)^{1/2} K_t \qquad 0 < s \leqq 1$$

Using this utility, the objective function is

$$U(e_n) = (1-s)^{1/2} K_t - K_f - K_r n \qquad 0 < s \leqq 1$$

This function was entered into a computer program which performed a golden section search giving the results shown in Table 4. As seen from this table, the economically optimal sample size is 52. This is a smaller sample size than obtained by using the linear utility function. This result is to be expected since this power function gives more weight to larger values of s.

Table 4. Computer Analysis Using Power Function with c = 1/2

Lower Limit	Upper Limit	N1	N2	U(N1)	U(N2)
0.00	324.50	123.93	200.54	.501	.259
0.00	200.54	76.61	123.93	.611	.501
0.00	123.93	47.33	76.61	.631	.611
0.00	76.61	29.28	47.33	.589	.631
29.28	76.61	47.33	58.56	.631	.631
47.33	76.61	58.56	65.38	.631	.625
47.33	65.38	54.15	58.56	.632	.631
47.33	58.56	51.74	54.15	.632	.632
47.33	54.15	49.73	51.74	.632	.632
49.73	54.15	51.74	52.14	.632	.632
51.74	54.15	52.14	53.74	.632	.632
51.74	53.74	52.14	53.34	.632	.632
51.74	53.34	52.14	52.94	.632	.632
51.74	52.94	52.14	52.54	.632	.632

The second power function utility to be considered has the parameter c equal to 1.5. Since this particular function is not guaranteed to be unimodal over all n, the method of subdividing the interval of uncertainty into a number of smaller intervals was employed. The interval of uncertainty, based on the budget constraint, is (0.00, 324.50). This interval was searched using subintervals of length 20. The results are shown in Table 5. As can be seen from this table, the optimal sample size is 83. Note that the utility of the experiment steadily increases until the optimal sample size is reached and then steadily declines over the remaining values of n. Thus, it is reasonably certain that a sample of size 83 is, in fact, a global optimal.

Table 5. Results of Computer Analysis Using Power Function Utility with c = 1.5

Subinterval	Optimal Sample Size for Subinterval	Utility of Experiment
0 - 20	20	-.144
20 - 40	40	.004
40 - 60	60	.065
60 - 80	80	.083
80 - 100	83	.084
100 - 120	100	.076
120 - 140	120	.053
140 - 160	140	.019
160 - 180	160	.022
180 - 200	180	-.069
200 - 220	200	-.121
220 - 240	220	-.176
240 - 260	240	-.234
260 - 280	260	-.356
300 - 320	300	-.420

Summary

The greatest limitation to the methodology developed in this study is that it is applicable only to the case of sizing an experiment for a single MOE. The logical extension of this is to the case of multiple MOE. There are at least two approaches to analyzing this case. One would be to apply multivariate Bayesian statistical theory combined with multi-dimensional nonlinear programming algorithms. A second approach would be to view the money required to perform each of the experiments involved in an operational test as a capital investment and the utility of each of the experiments as the return on that investment. Formulated in this manner the problem might be solved utilizing capital budgeting techniques. If it is possible to extend the methodology to include multiple MOE, then it may be possible to use it in multifactor experimental design problems.

Aside from extending the methodology, several other areas warrant further investigation. First, is the assumption that the normal process may be used as a reasonable model for a large number of operational testing problems. Closely associated with this would be an investigation of the variation in results when the sampling process is not normal.

The economic analysis assumes that certain costs relative to the conduct of OTEA's data collection and analysis can be determined. OTEA personnel must judge whether this information can be collected at a reasonable cost or whether adequate estimates can be made where actual data is not available so that the results of this methodology will provide additional information for the test planners.

As a final recommendation, it is suggested that the procedures outlined in this study be utilized in designing a number of operational tests and that these results be compared to the results obtained using the presently employed methods.

"A Methodology for Determining the Power of MANOVA When the Observations are Serially Correlated," by Norviel R. Eyrich, Captain, Artillery

The Problem

In recent years the U.S. Army has expended a great deal of money and time to develop and deploy sophisticated tactical command and control systems. Measures of effectiveness employed in the evaluation of command and control systems vary; however, the measures of effectiveness are rarely independent. For instance, the fraction of available time passed to subordinate echelons and time required to prepare staff actions, two possible measures of effectiveness, are highly correlated.

Both analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) appear to be appropriate statistical methods to be used for analysis of command and control experimental data. Recent research has developed a methodology for determining which statistical method, or combination of methods, is most appropriate for a particular system. This past research has not, however, considered that in addition to the various measures being correlated, that in the case of computer assisted systems they may also constitute a multivariate time series. A promising area of research appeared to exist in developing a methodology for identifying, analyzing, and incorporating this additional information into the methodology developed by Burnette for determining the appropriateness and effectiveness of ANOVA and MANOVA in the analysis of command and control systems.

Objective

(1) To investigate the effects of a multivariate time series on the multivariate analysis of variance power function.

(2) To develop a methodology for incorporating time series information into the MANOVA power generator previously developed by Burnette. This will enable test designers to determine the sample size required to achieve a given power when tests of competing systems yield multivariate time series data.

Methodology

Previous research on the MANOVA power function on data that was not serially correlated indicated the following:

1. Power is a decreasing function of the dimension of the multiresponse.
2. Power is an increasing function of the size departure from the null hypothesis.
3. Power is an increasing function of sample size.
4. Power is an increasing function of the probability of Type I error.
5. Power is an increasing function of $-\log |\underset{\sim}{P}|$, where $\underset{\sim}{P}$ is the correlation matrix of the multiresponse.

It was decided that an appropriate method to simultaneously investigate the above effects along with the serial correlation effect would be to use a factorial design and analyze the results by ANOVA. Prior to selecting the design, either a 2^k or a 3^k, it was necessary to determine if the main effects were linear or of some higher order. Thus, six individual experiments were conducted to determine the nature of the main effects. In each experiment the effect under investigation was varied over the range of interest while the other effects were held constant. In each case there appeared to be a linear trend in the main effect, with the exception of the response dimension, and thus, it was felt that a 2^k experimental design would be appropriate.

The effect of the dimension of the response was investigated by the procedure described above. It was found that the dimension of the response could not be separated from the other factors and thus could not be included as a factor. It was then decided to run two full 2^5 factorial experiments with the dimension of the response, p, set at 2 in the first and 3 in the second. Appropriate high and low levels of each of the other factors were selected (these are reported in the thesis).

Data for each of the experimental combinations was generated by the computer routines to simulate the power function which was developed by Burnette. These routines were modified to generate serially correlated multivariate data. The experiments were not replicated since the number of replications of the MANOVA power generator (500 replications) results in little or no variation in the responses. The effects in each experiment were plotted on normal probability paper, and the fourth and fifth order interactions fall along that portion of the plot where the effects may be represented by a straight line. Thus the error sums of squares was estimated using the fourth and fifth order interactions and a complete ANOVA was run.

The analysis of both experimental designs verify that all main effects are highly sugnificant. The results indicate a number of second order interactions are significant. However, if the percentage of total variation explained by the main effects, their mean square, and the amount of total variation explained by the second order interactions is examined, we may infer that some of the second order interactions are not significant. The $\lambda \times |\underset{\sim}{P}|$, $D_2 \times |\underset{\sim}{P}|$, $\lambda \times D_2$ and the $\lambda \times n$ interactions appear significant in this perspective, where λ is the auto correlation coefficient, $|\underset{\sim}{P}|$ is the euclidian norm, D_2 is the departure, and n is the sample size.

Additional information on the second order interactions was acquired through their graphical representation. The graphical results confirmed the interaction of the autocorrelation coefficient with the other factors and also indicated that the autocorrelation coefficient had its greatest effect on the other factors when they were at their low levels. This result is not surprising since we would expect the greatest increase in the MANOVA power to occur when the MANOVA power is low; that is, when the other factors are at their low levels.

Several general statements concerning the factors which influence the MANOVA power function were made. They are:

1. All five factors considered in the experimental design significantly affect the MANOVA power function.
2. The numerous second order interactions make an interpretation of the effects of the factors on the MANOVA power function extremely difficult.
3. The autocorrelation coefficient, λ, the determinant of the correlation matrix, $|\underset{\sim}{P}|$, and the departure, D_2, appear to have a very significant effect on the MANOVA power function through second order interactions.

4. The power of the MANOVA test statistic decreases with the dimension of the response.

5. The autocorrelation coefficient, λ, has a greater effect on the MANOVA power function when the other factors are at their low levels.

It is noted that power was an increasing function of the autocorrelation structure of the response vector. That is, power increases as the significance of the multivariate time series increases. It was also noted that the large number of significant second order interactions make an interpretation of the response difficult; however if subjective estimates are to be made for either λ or $|\underset{\sim}{P}|$ great care must be exercised due to their impact on the MANOVA power function.

An Application to Operational Testing

The methodology developed above was applied to an operational testing problem. The hypothetical command and control system used by Burnette was used so that the results could be compared. The hypothetical command and control system, known as the Brigade Anti-armor Command and Control System (BACCS), will be described now. Two competing forms of BACCS were under consideration for acquisition and are designated BACCS-I and BACCS-II.

For OT II, the commander, U.S. Army Operational Test and Evaluation Agency (OTEA), had approved a comparative operational test of the two systems consisting of three scenarios. The commander had also approved seven measures of effectiveness designated MOE-1 through MOE-7. In addition, the commander had approved a completely crossed two-factor experiment with equal numbers of observations per cell. He desired to determine for which MOE MANOVA would be most effective, powerwise, than ANOVA.

An objective estimate of the correlation structure of the MOE correlation matrix was:

	1	2	3	4	5	6	7
1	1.0	.00	-.06	-.12	.00	-.17	.16
2	.00	1.0	.01	-.11	.01	-.04	.76
3	-.06	.01	1.0	.68	-.49	.56	.07
4	-.12	-.11	.68	1.0	-.21	.72	-.04
5	.00	.01	-.49	-.21	1.0	-.26	[illegible]
6	-.17	-.04	.56	.72	-.26	1.0	[illegible]
7	.16	.76	.07	-.04	-.11	-.08	1.0

OT I test results indicated that each response vector was related to the previous response vector. However, insufficient information was available to obtain an objective estimate; therefore, a subjective estimate of the autocorrelation coefficient, $\lambda = 0.3$, was made by the BACCS project manager and the U.S. Army Training and Doctrine Command.

Based upon a knowledge of BACCS, it was felt that MOE-1 was independent of all other MOE. We test this hypothesis. The hypothesis that MOE-1 is independent of the other MOE is not rejected. MOE-1 is assigned to the set of mutually independent measures, I.

Knowledge of BACCS indicates that MOE-2 and MOE-7 were correlated, but independent of the other MOE. It was also felt that MOE-3, MOE-4, MOE-5, and MOE-6 were correlated but independent of the other MOE. Thus MOE-2 and MOE-7 were assigned to correlated set C_1. And MOE-3, MOE-4, MOE-5, and MOE-6 were assigned to correlated set C_2. Thus, the correlation matrix for the set C_1 was the 2 x 2 matrix

2	1.0	.76
7	.76	1.0

and the correlation matrix for set C_2 was the 4 x 4 matrix

3	1.0	.68	-.49	.56
4	.68	1.0	-.21	.72
5	-.49	-.21	1.0	-.26
6	.56	.72	-.26	1.0

It was desired to test the hypothesis that set C_1 and set C_2 were mutually independent using the appropriate test statistic with $\alpha = 0.05$. The test statistic is

$$\chi_0^2 = 4.1630$$

and the critical value of the test

$$\chi_{.05,8}^2 = 15.5072$$

The test statistic is less than the critical value of the test; hence, the hypothesis of independence was not rejected and it was concluded that C_1 and C_2 were independent. It was necessary to determine if the MOE within the mutually independent sets C_1 and C_2 were independent.

Set C_1 had only two MOE and thus has a bivariate normal distribution. The Fisher Z-transformation was used to test the hypothesis

$$H_{10}: \ \rho_{27} = 0$$

against

$$H_{11}: \rho_{27} \neq 0 .$$

This gave

$$Z = \tanh^{-1} (.76) = 0.638$$

and the test statistic was

$$|Z| \sqrt{N - 3} = 0.638 \sqrt{42 - 3} = 3.984 .$$

The critical value of the test with $\alpha = .05$ is $Z_{.05} = 1.96$. The test statistic exceeded the critical value of the test; hence, H_{10} was rejected and it was concluded that MOE-2 and MOE-7 were correlated.

To test the following hypothesis

$$H_{20}: \underset{\sim}{P}_{c2} = \underset{\sim}{I}$$

against

$$H_{21}: \underset{\sim}{P}_{c2} \neq I$$

to determine if MOE-3, MOE-4, MOE-5, and MOE-6 were correlated, the test statistic

$$\chi_0^2 = -\left(N - 1 - \frac{2k + 5}{6}\right) \text{Log } |\underset{\sim}{R}|$$

$$= -\left(42 - 1 - \frac{2 \cdot 4 + 5}{6}\right) \text{Log } |\underset{\sim}{R}|$$

$$= 65.81137$$

was used. With $\alpha = .05$ the critical value of the test is

$$\chi^2_{.05,6} = 12.59120 .$$

The test statistic exceeded the critical value of the test; hence, we concluded the members of C_2 were correlated.

The above procedures separated the MOE into three mutually independent sets:

$$I = \text{MOE-1}$$

$$C_1 = \text{MOE-2, MOE-7}$$

$$C_2 = \text{MOE-3, MOE-4, MOE-5, MOE-6} .$$

ANOVA was appropriate for MOE-1, the sole member of set I; therefore, MOE-1 was not used for a comparison of the effectiveness of MANOVA with ANOVA.

The Commander of OTEA had specified the following probability levels be used for BACCS OT-II:

Probability of Type I error, -.05

Power of the test $(1 - \beta)$ -.75.

These parameters were applied to both ANOVA and MANOVA. In addition, the maximum sample size, n_{max}, and the departure to be detected, D, were specified for each MOE. These parameters are shown in Table 5.

Using the information in Table 1 the minimum sample size, n_{ANOVA}, for each MOE required to achieve the desired power was computed. This was accomplished by using the results from Burnette's work. The results are shown in Table 2.

Table 1. MOE Maximum Sample Sizes and Departures

MOE	Maximum Sample Size n_{max}	Departure to Detect D
1	6	1.5
2	6	1.5
3	4	2.0
4	6	1.5
5	6	1.5
6	7	1.0
7	6	1.5

Table 2. MOE Sample Sizes for Required Power

MOE	Maximum Sample Size n_{max}	Departure to Detect D	Minimum Sample Size n_{anova}
1	6	1.5	5
2	6	1.5	5
3	4	2.0	4
4	6	1.5	5
5	6	1.5	5
6	7	1.0	7
7	6	1.5	5

For the two sets of correlated measures, C_1 and C_2, it was necessary to determine for which members of these sets MANOVA was more effective than ANOVA from the standpoint of power. The Commander of OTEA had approved a ratio R = 2 for use in setting the random levels of the MOE in the sets other than those under consideration.

For set C_1 = {MOE-2, MOE-7} it was found that $n_{min} = \min\{n_{ANOVA\ 2}, n_{ANOVA\ 7}\} = 5$. The two-factor MANOVA computer program was used with levels of factor A = 2, levels of factor B = 3, D = 1.5, sample size = $n_{min} = 5$, $\lambda = .3$, R = 2, Monte Carlo iterations = 500, and correlation matrix $\underset{\sim}{P}_{c_1}$. The results are tabulated in Table 3 with the results of Burnette's research for ease of comparison.

Table 3. MOE Power 1

MOE	MANOVA Sample Size n_{manova}	Departure to Detect D	Power Achieved by Burnette	Power Achieved by this Research
2	5	1.5	.762	.866
7	5	1.5	.824	1.000

The MANOVA power was greater than the ANOVA power with sample size n_{min}; thus, MANOVA was more effective than ANOVA for members of set C_1.

For set C_2 = {MOE-3, MOE-4, MOE-5, MOE-6} the same two factor MANOVA power program was used. The results are shown in Table 4 for this research and Burnette's for ease of comparison of results.

Table 4. MOE MANOVA Power 2

MOE	MANOVA Sample Size n_{manova}	Departure to Detect D	Power Achieved by Burnette	Power Achieved by This Research
3	4	2.0	.614	.850
4	4	1.5	.482	.824
5	4	1.5	.496	.776
6	4	1.0	.452	.994

It was noted that again the MANOVA power exceeded the power of the ANOVA for all components, therefore, MANOVA was more effective than ANOVA for all members of the set C_2. It was shown that MANOVA was superior to ANOVA for both set C_1 = {MOE-2, MOE-7} and set C_2 = {MOE-3, MOE-4, MOE-5, MOE-6}. This information would be used to aid in the design of BACCS OT II.

Although the example presented was hypothetical the methodology as demonstrated may be applied to any system so long as an estimate of the structure of the response is available. Note that the introduction of autocorrelated vectors greatly influence the MANOVA power function. Burnette was able to achieve joint inference on only two MOE in set C_2 at the specified power. This analysis, using the systems information, achieved joint inference on all four MOE of set C_2 at the specified power level greatly enhancing the analysis of the test results.

Summary

It was found that the incorporation of the time series into the MANOVA power function significantly increased the MANOVA power for a given sample size. It was also noted that a reduction in sample size, for a given power, could be achieved when the time series information is incorporated in the MANOVA power function.

This research has been limited by the initial assumptions of two-factor, fixed-effects, crossed models, equal sample sizes per cell, and no effects due to operators. In addition, it was assumed that an estimate of the correlation structure of the measure of effectiveness and the autocorrelation coefficient or all the parameters of a multi-variate time series are available.

One recommendation for further research is to develop an exact statistical test for a multiresponse system when the responses are time dependent. An experiment could then be designed using the exact test and the current procedure to determine if MANOVA is robust to independence of observations. Another recommendation is to extend the MANOVA power program so that it may handle nested, multi-factor designs.