AD-776 911

PROCEEDINGS OF THE CONFERENCE ON
THE DESIGN OF EXPERIMENTS IN ARMY
RESEARCH DEVELOPMENT AND TESTING
(18TH) HELD AT ABERDEEN PROVING
GROUND, MARYLAND, ON 25-27 OCTOBER 1 72.
PART 2

Army Research Office
Durham, North Carolina

October 1973

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

DOCUMENT CONTROL DATA - R & D *AD 776 911*

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*	2a. REPORT SECURITY CLASSIFICATION
U. S. Army Research Office	Unclassified
Box, CM, Duke Station	2b. GROUP
Durham, North Carolina 27706	N/A

3. REPORT TITLE

PROCEEDINGS OF THE EIGHTEENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS
IN ARMY RESEARCH, DEVELOPMENT AND TESTING

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Interim Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 1973	763	

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	ARO Report 73-2
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT

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

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	Army Mathematics Steering Committee on behalf of the Chief of Research and Development.

13. ABSTRACT

This is a technical report resulting from the Eighteenth Conference on the Design
of Experiments in Army Research, Development and Testing. It contains most of the
papers presented at that meeting. These treat various Army statistical and design
problems.

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

14. KEY WORDS:

exploratory data analysis
orientation analysis
frequence response
weighing designs
block designs
spurious correlations
penetration ballistics
suitability testing
stopping rules
helicopter lift
adaptive filtering
bivariate data
ionospheric data
reliability growth

propagation of errors
tail length analysis
bicellular communication
shock initiation
burn - through phenomenon
propagation loss data
Grubbs' estimators
confidence limits
gun tube forgings
factorial estimation
forecasting
continuous sampling
response surface analysis
performance evaluation
biased samples

Pages i, ii, and iii of this document are Blank.

367

DD FORM **1473** REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

iv

TABLE OF CONTENTS

v

viii

DPG WIND TUNNEL MODIFICATION AND EVALUATION

E.G. Peterson, E.E.Covert, D.L. Hansen
Deseret Test Center, Fort Douglas, Utah

ABSTRACT. The wind tunnel facility at Dugway Proving Ground, Utah, has been used to calibrate chemical samplers. The samplers were challenged with an agent generated by a particular munition. The wind tunnel has been modified to improve the reliability of the calibration procedure. Thus, the tunnel could be used primarily as a facility to calibrate sampling apparatus. The following steps were taken to eliminate or reduce as many of the uncontrolled variables as possible from the calibration procedure.

   a. Speed rings were installed to determine the wind speed in the test section.

   b. Nonuniformities in the wind stream were reduced by modifying the tunnel inlet section. It was extended 3 feet to hold a set of plastic tubes that serve as flow straighteners. Immediately downstream to dampen out small-scale disturbances. This installation consisted of two screens about an inch apart. These modifications reduced the fluctuations in the wind speed at the test section from 20 percent to less than 1 percent. The nonuniformities were reduced from 7 to about 1 percent.

   c. A turbulence grid was placed 6 inches downstream of the last copper screen to mix the agent with the air during its passage from the generator to the test section. The intensity of the turbulence produced by this grid is about 50 percent near the point where the agent is introduced and about 2-percent turbulence at the sampling plane.

   d. An isokinetic sampling probe was designed to operate at any wind speed in the tunnel. By controlling the pressure difference between the inlet to the isokinetic sampler and the free stream, the inlet velocity ratio of this probe is held at unity, within experimental error. The test results indicate that the isokinetic probe operates in the design mode. Consequently, this sampler was used as a standard to determine the sampling efficiency of all glass impinger (AGI), the British chemical impinger (CBI), the snoot sampler, and the chemical impinger (CI) using CS as the agent.

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

SECTION 1. SUMMARY

1.1 BACKGROUND

Prior to conducting field tests, chemical samplers must be calibrated and checked to determine their efficiency and their ability to collect the particular agent to be disseminated. The efficiency was based on the concept of a mass balance. In the past, the existing wind tunnel facility at Dugway Proving Ground (DPG), Utah, has been used to perform this function. The calibration data obtained from these tests have been analyzed. The results indicate that the tunnel facility had many uncontrolled factors that limited its usefulness. Nonuniformities in the wind stream and agent concentration were considered to be the main contributors to the uncertainty of the results. An attempt was made to average out the uncontrolled factors by using R.A. Fisher's latin square method. This method improved the calibration procedure, but only relative efficiencies could be determined. The procedure became more complex with the introduction of pyrotechnic munitions such as the XM-100.

In April 1970, a meeting was held to determine the future requirements of the DPG tunnel. It was decided that no testing would be done in the tunnel until it has been evaluated and modified to eliminate uncertainties caused by the tunnel flow itself. The effectiveness of the modification was assessed in two ways: (1) by the uniformity and steadiness of the flow in the wind tunnel and (2) through the use of an isokinetic sampler. The isokinetic sampler could also be used as a standard in determining the efficiencies of field samplers. The concepts involved in the isokinetic sampler are consistent with the work done at Suffield and Porton.[1,2]

1.2 OBJECTIVE

The objective of this project was to evaluate the existing DPG wind tunnel and, if necessary, modify it for use as a primary sampler calibration facility. The base line or reference to compare the performance of the modified tunnel was the existing tunnel. If necessary, procedures would be developed to determine collection efficiencies of chemical samplers in the tunnel.

1.3 CONCLUSIONS

1.3.1 The wind speed profile across the sampling plane in the test section can be made uniform to about 1 percent.

1.3.2 Fluctuations in the wind speed were reduced to below 1 percent. They were as high as 20 percent before the modifications.

1.3.3 No further modifications to improve the uniformity of the air stream are necessary until other uses of the facility require a lower level of turbulence in the air flow or until it becomes necessary to conduct tests in any kind of weather.

1.3.4 The isokinetic sampler can be used as a standard for calibrating field samplers. Care must be taken to make sure the pad used in the isokinetic sampler does not overload. This can be checked by recording the unbalance in the pressure difference between the inlet plane of the sampler and the static pressure in the test section during the calibration.

SECTION 2. DETAILS OF STUDY

2.1 SCOPE

The project was divided into two phases. The purpose of Phase II was to provide a standard probe and calibration. Phase I consisted of evaluating problem areas in the tunnel facility, making corrective modifications, and running control studies after each modification. Phase II included developing the isokinetic sampling probe, validating the isokinetic probe, installing a turbulence grid for mixing purposes, and comparing the standard probe to other samplers to determine sampler efficiencies.

2.2 PROCEDURES

2.2.1 Preliminary Evaluation

The preliminary evaluation was that the air flow in the tunnel could be improved enough to calibrate chemical samplers. The concept of a standard probe in a low-turbulence air flow was based on the use of an isokinetic sampler. This would be used as a reference in determining the efficiencies of field samplers. This concept is consistent with work by H.H. Watson of Suffield,[1] K.R. May and H.A. Druett of Porton,[2] and J.E. Mayhood.[1]

2.2.2 Modifications and Results

Each modification shown in Figure 1 will be discussed in detail.

2.2.3 Tunnel Inlet Modification

The wind shield was moved back 9 feet to allow for the construction of the flow-straightener section at the tunnel inlet. The flow-straightener and dampening section was designed to remove the uncontrolled large velocity fluctuations at the test section of the tunnel. As shown in Figure 2, this section was constructed as a separate unit and could be rolled back from the existing tunnel entrance. The louvres that were located on the old entrance were removed and relocated on the new entrance of the straightener section to provide limited control over test-section velocity profile. A set of plastic tubes with inside diameters of 1 inch and lengths of 30 inches were installed 4 inches downstream of the louvres. These tubes serve as flow straighteners and average out large gusts of wind. A layer of 18-mesh ordinary copper window screening was installed immediately downwind of the tubes. A second screen was located about an inch downstream from the first. These two screens are used to further break up the surviving eddys and

DPG WIND TUNNEL FACILITY
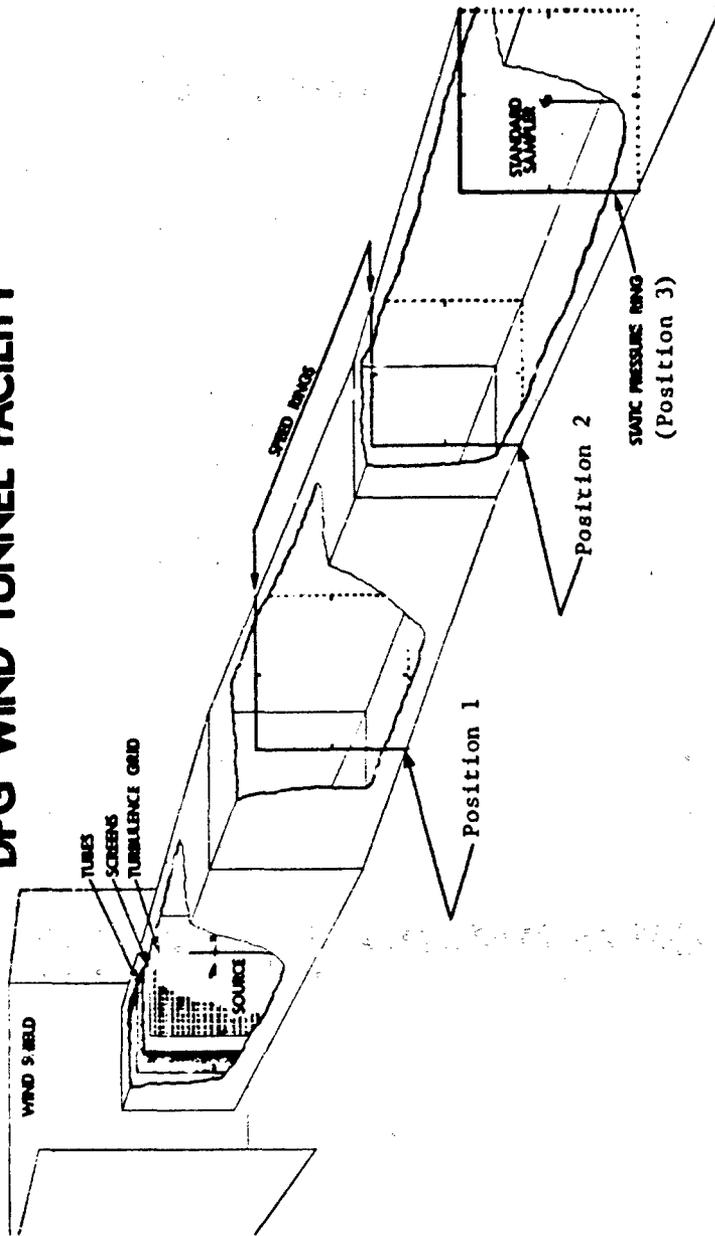
Figure 1. DPG Wind Tunnel Facility

Two copper screens, 1 inch apart, located 2 inches downstream from the plastic tubing

End flanges with quick release hooks should be airtight (i.e., rubber gasket)

Existing tunnel cone

Existing louvres

Iron track, 8 feet long

Approximate scale: ½ inch = 1 foot

Existing doors

1-inch inside diameter tubes, 30 inches long

Concrete pad
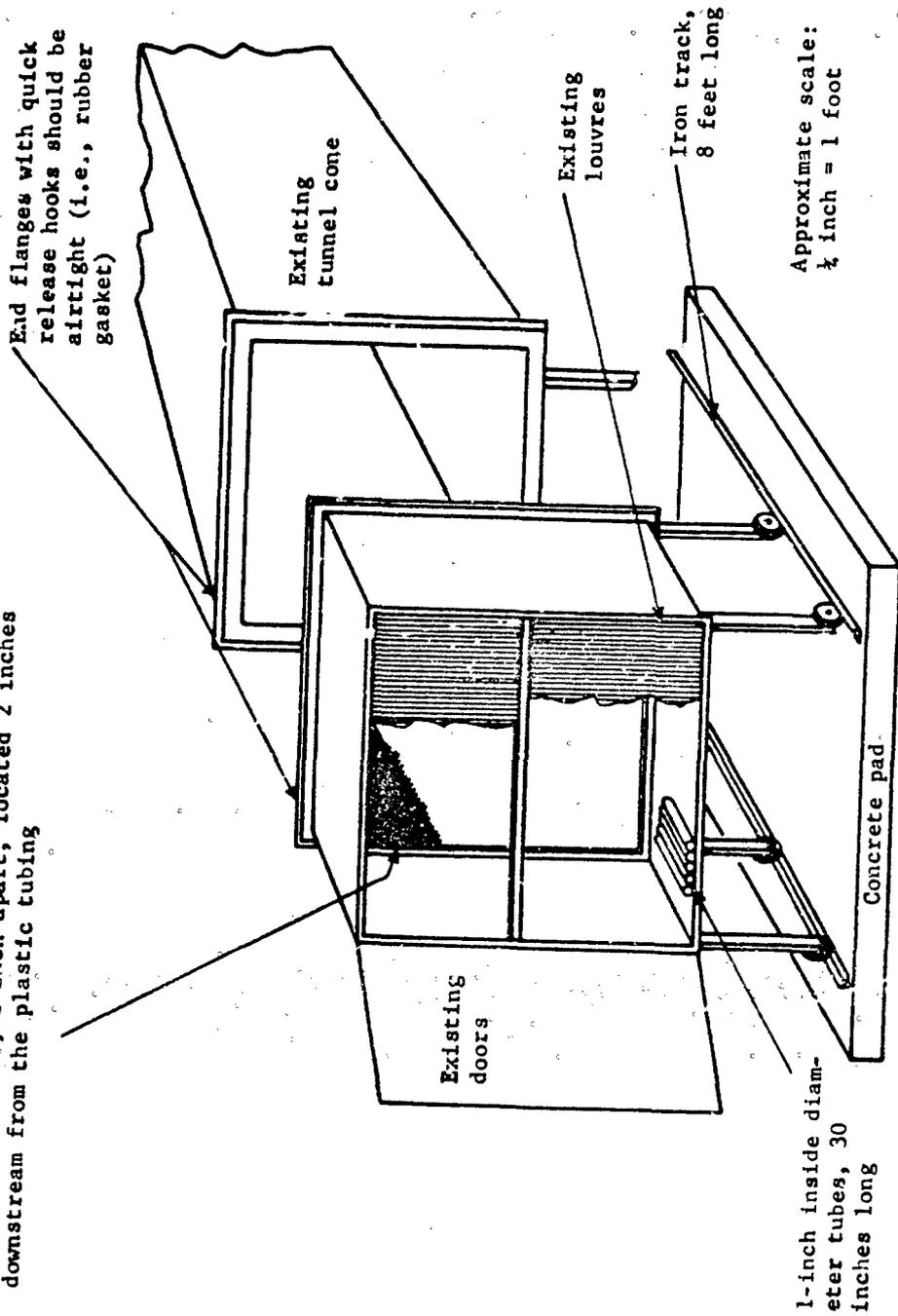
Figure 2. Straightener Section Modification

dampen the intensity of turbulence downstream from the plastic tubes according to the relation[3]

$$U_2'/U_1' = 1/(1 + k)^{n/2} \qquad\qquad \text{Eq. (1)}$$

where 1 and 2 refer to stations ahead of and behind the dampening screens, U' refers to the root mean square velocity fluctuation, n is the number of screens, and k is the pressure-drop coefficient for one screen. For 18-mesh screen, k has a value of 1.0 at 10 miles per hour.

Equation (1) can be used to show that the two screens have the effect of reducing the turbulence downstream from the tubes by a factor of 2.

2.2.4 Test Section Instrumentation

Sixteen pitot tubes were installed in the center of the test section. The test section is located at the standard sampler plane, as shown in Figure 1. The pitot tube configuration (Fig. 3) was used to measure dynamic pressure profiles across the test section before and after the new inlet section was added to the tunnel.
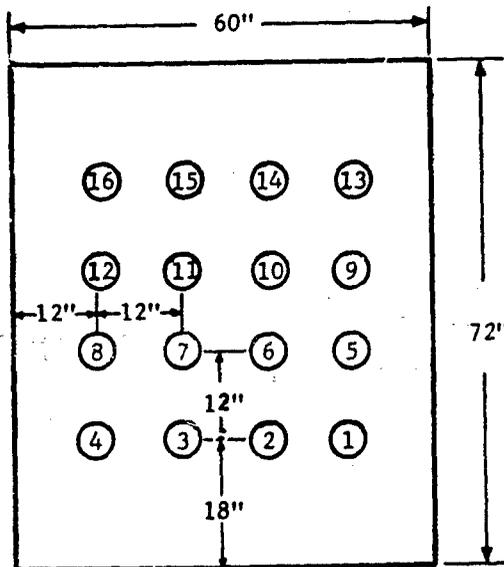
Figure 3. Pitot Tube Configuration at the Test Section (Downstream View)

-397-
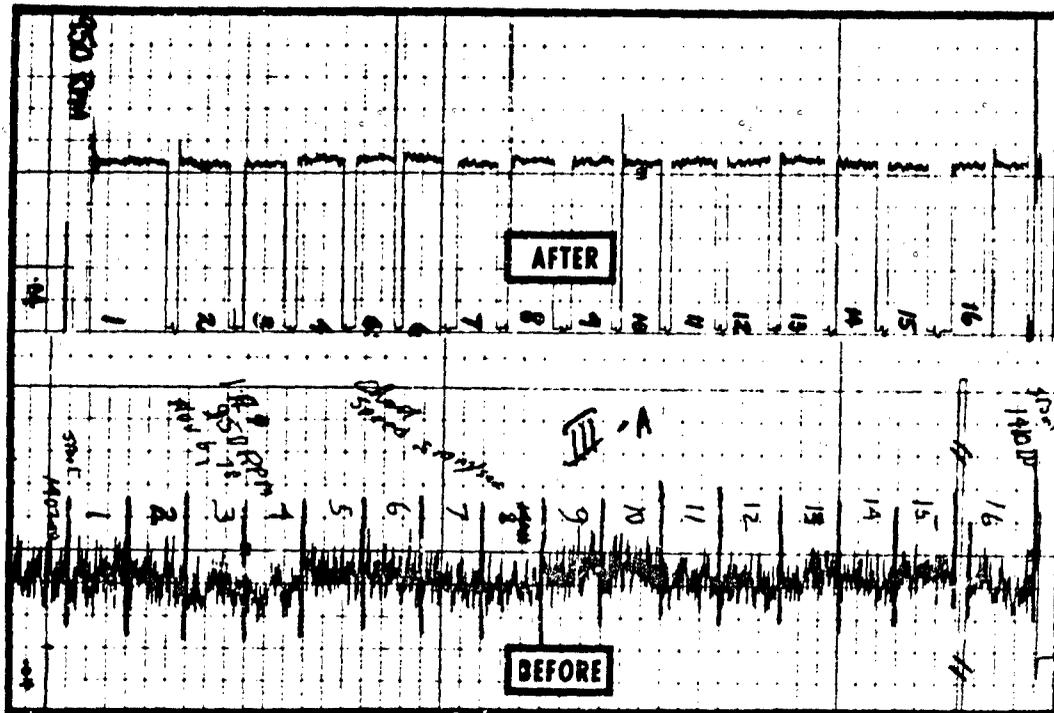
Figure 4. Dynamic Pressure Measurements Before and After Modification

2.2.5 Test Results Before and After Inlet Modification

Measurements taken before and after the straightener-dampening section was added to the tunnel are shown in Figures 4, 5, and 6. The results shown in Figure 4 are the dynamic pressures measured by the pitot tubes at a motor speed of 950 r.p.m. This corresponds to about 15 miles per hour. The number shown above each pitot tube reading in Figure 4 corresponds to the number shown in Figure 3. The data shown in Figure 4 were reduced to velocity, and the results were used to construct the contours of constant %ΔV shown in Figure 5. Figure 6 shows the fluctuations in velocity before and after the straightener section was added to the tunnel. The data indicates that the added section reduces the velocity fluctuations from 16 to 3 percent at about 4 miles per hour and 12 to 1 percent at 15 miles per hour. As shown in Figure 5, the %ΔV variations were reduced from 7 percent to 1 percent at about 15 miles per hour. The dynamic pressure data for 250 and 600 r.p.m. show the same trend as Figure 4.

-398-
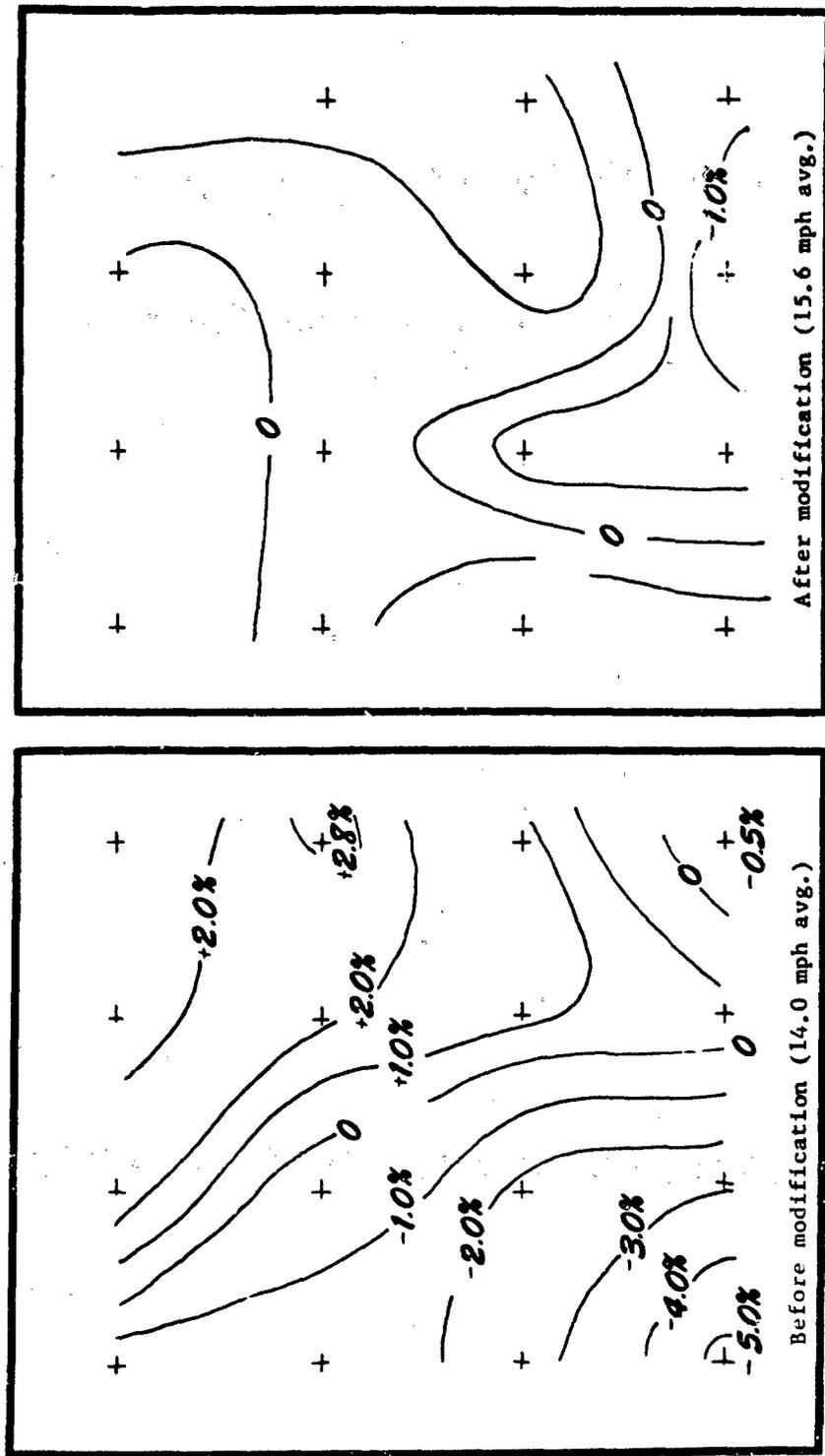
Figure 5. Contours of Constant Percent ΔV Variation
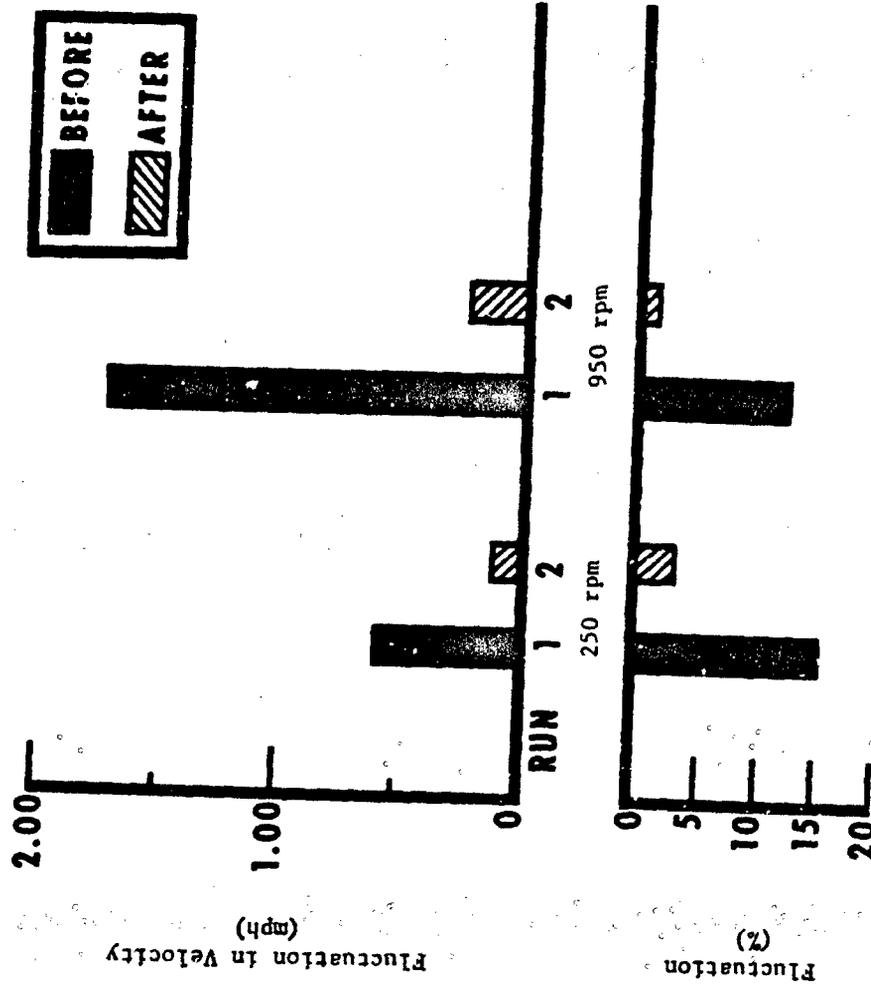
Figure 6. Trend of Fluctuation with Modifications

2.2.6 Turbulence Grid Modification

A 2-inch turbulence grid was installed about a foot downstream from the last copper screen. The downstream turbulence decay from a grid in low turbulence air flow is well known.[4] The turbulence grid was installed to introduce uniform large turbulent eddies to promote rapid mixing immediately downstream of the grid. The lattice was designed not only to accomplish the mixing by the eddies formed from the air stream passing through the lattice, but also to reduce the magnitude of these eddies at a suitable decay rate. This insures that the introduced turbulence has little effect on the air stream in the test section. The dimensions of the grid were determined using the graph in Figure 7.[4] The grid was designed using 2-inch lath slats with 2-inch square openings between the slats. In Figure 7, x refers to downstream distance from the grid, V_x' is the root mean square of the velocity, $\bar{V}$ is the mean velocity, b is the bar width of the grid, and the ratio $V_x/\bar{V}$ is defined as turbulence intensity. The distance (x) from the grid to the test section is about 480 inches, and the bar width of the grid is 2 inches. This gives an x/b ratio of 240, and when applied to the graph shown in Figure 7, the turbulence intensity is about 2 percent. The mixing effect can be estimated using the same value for b and a value of 8 inches for x. This gives a turbulence intensity of 40 to 50 percent near the point of release of the agent. To estimate the cloud spread from a point source near the grid, the Smith-Hays model[6] for isotropic turbulence was used in the form

$$\sigma = 3 \int_0^x \left(\frac{V_x'}{\bar{V}} \right)^2 dx \qquad \text{Eq. (2)}$$

Using Equation (2) and Figure 7, the cloud spread 480 inches downstream was calculated to be 1 foot for 1 σ if the munition were fired downstream. In the calibration trials, the munition was fired upstream to increase the cloud spread.

2.2.7 Speed Ring Description, Installation, and Calibration

The first speed ring was installed downstream from the turbulence grid as shown in Figure 1. The locations of the first speed ring, second speed ring, and test section will be referred to as positions 1, 2, and 3, respectively. Bernoulli's equation is used to obtain the relation between the static pressure at positions 1 and 2 and the dynamic pressure at position 3. The ring at position 2 is placed far enough upstream of the test section to avoid any effects on the static pressure measured at that point when samplers are removed or installed in the test section.

-401-
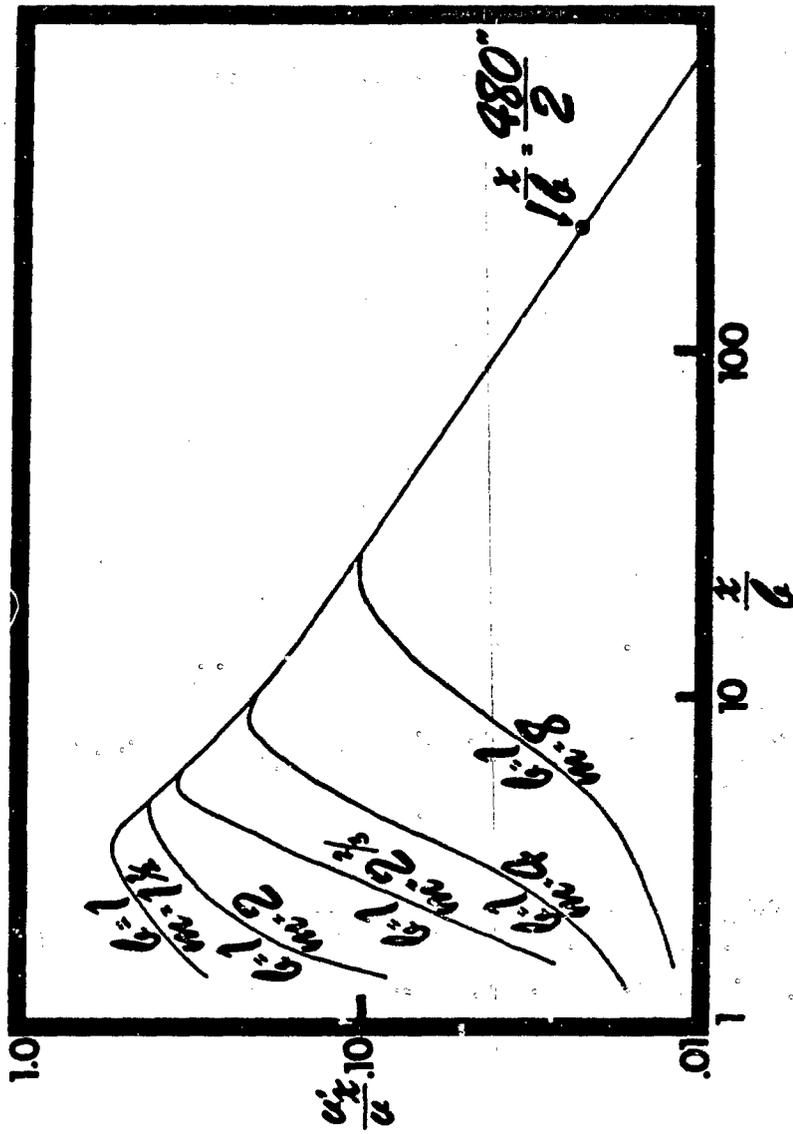
Figure 7. Variation in Intensity of Turbulence Produced by Lattice-Type Screens[4]

2.2.8 Theory of Speed Ring

Because of a small pressure drop between the 2 speed rings, the total pressure at position 2 will be slightly smaller than that at position 1. Using k_1 as the loss coefficient of the section between rings, Bernoulli's equation can be written[3]

$$p_1 + q_1 = p_2 - k_1 q_2 + q_2 \qquad \text{Eq. (3)}$$

where p and q are static and dynamic pressures, respectively. The mass flow rate at any location in the wind tunnel is given by

$$\dot{m} = \rho A V$$

where (ρ) is the density of the air, (A) is the section area, and (V) is the mean wind speed. The mass flow rate at positions 1 and 2 are the same, and the relationship between the velocity and cross-sectional area is given by

$$\frac{V_1}{V_2} = \frac{A_2}{A_1} \qquad \text{Eq. (4)}$$

The dynamic pressure is defined by

$$q = \frac{\rho V^2}{2} \qquad \text{Eq. (5)}$$

Using Equation (4), Equation (5) becomes

$$q_1 = k_2 q_2 \qquad \text{Eq. (6)}$$

where $k_2 = A_2^2 / A_1^2$

Likewise, $\qquad\qquad q_2 = k_3 q_1 \qquad\qquad$ Eq. (7)

where $k_3 = A_3^2/A_2^2$

Using Equations (6) and (7), Equation (3) becomes

$$q_3 = k_4 (p_1 - p_2) \qquad \text{Eq. (8)}$$

where $k_4 = \dfrac{1}{(1 - k_1 - k_2)k_3}$

2.2.9 Calibration of the Speed Ring

The wind tunnel was operated at various speeds with $p_1 - p_2$ and q_3 recorded using the Baratron output. The dynamic pressure (q_3) was obtained using the average of 16 pitot tubes located at the test section. The results of the calibration test are shown in Figure 8.
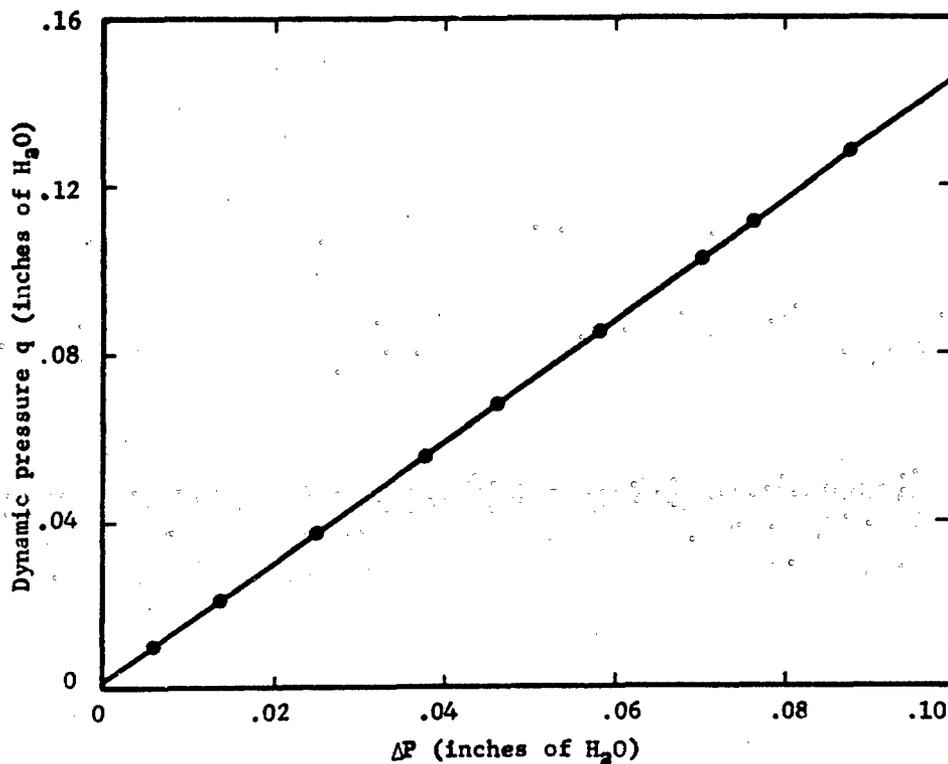
Figure 8. Calibration of Speed Rings

2.2.10 Velocity Determination

The mass flow rate through the test section remains the same with or without samplers. When samplers are placed in the test section, the cross-sectional flow area (A_3) decreases, and the average flow velocity at the sampler plane increases. This relation is given by

$$V_3' (A_3 - \Delta A_3) = V_3 A_3 \qquad \text{Eq. (9)}$$

In Equation (9), V_3' refers to the mean air velocity with samplers in the test section, V_3 is the mean air velocity without samplers in the test section, and ΔA_3 is the solid blocking area presented by the samplers to the air at the plane of the inlet to the samplers. The mean air velocity without samplers (V_3) can be determined by (1) using the speed rings and (2) using a standard pitot tube centered in the test section about a foot upstream from the samplers. The velocity V_3, using the speed rings, is given by

$$V_3^2 = \left(\frac{k_4}{2.521 \times 10^{-4}}\right) \left(\frac{T + 460}{P}\right) (\Delta p) \qquad \text{Eq. (10)}$$

where $\Delta p = p_1 - p_2$ (static pressure differential measured in inches of water)

P = barometric pressure in millibars

V_3 = air velocity at the test section in miles per hour

T = temperature in $^\circ F$

k_4 = slope of the line shown in Figure 8 (k_4 = 1.463).

If a pitot tube is used to set the velocity at the test section, Equation 11 is used.

$$V_3^2 = \left(\frac{1}{4.7097 \times 10^{-4}}\right) \left(\frac{T + 460}{P}\right) q_3 \qquad \text{Eq. (11)}$$

where q_3 is the dynamic pressure measured by the pitot tube in mm Hg, and P and T are as defined in Equation 10.

2.2.11 Operation Procedure to Set Tunnel Speed

The most convenient way to set the wind speed in the tunnel is as follows:

    a. Use the pitot tube to set the speed.

    b. Measure the barometric pressure in millibars.

    c. Measure the outside temperature in $^\circ F$.

    d. Use Eq. 11 to find the dynamic pressure in mm of Hg for the given air velocity.

    e. Set the dynamic pressure using the dial setting in the Baratron.

    f. Adjust the speed of the tunnel fan so that the Baratron output needle reads 0 (zero).

    g. The tunnel can be operated with outside gusts of wind that do not change the dynamic pressure more than ±10 percent. This corresponds to about ±0.5 miles per hour in the test section.

2.2.12 Isokinetic Sampler

2.2.12.1 Description of Sampler. The isokinetic sampler was based on a Massachusetts Institute of Technology design. The novel feature of this design is the static pressure tap at the inlet of the sampler. The static pressure measured at the inlet is compared with the ambient static pressure to insure isokinetic operation as closely as possible. The sampler is located in the test sections (Fig. 1). Figure 9 is a detailed sketch of the instrumentation required for the isokinetic sampler and Figure 10 shows the actual design. The relationship between the average stream velocity (V_0) and the inlet velocity (V) of the sampler (Fig. 9) is given by

$$V = V_0 \sqrt{1 + \frac{P_0 - P}{q_0}} \qquad\qquad \text{Eq. (12)}$$

where p_0 = average static pressure in the air stream at the test
section as measured by the static pressure ring shown in
Figure 9

    q_0 = dynamic pressure of the air stream

    p = static pressure just inside the entry orifice of the sampler.

MKS baratron

Static pressure ring (typical)

V_o
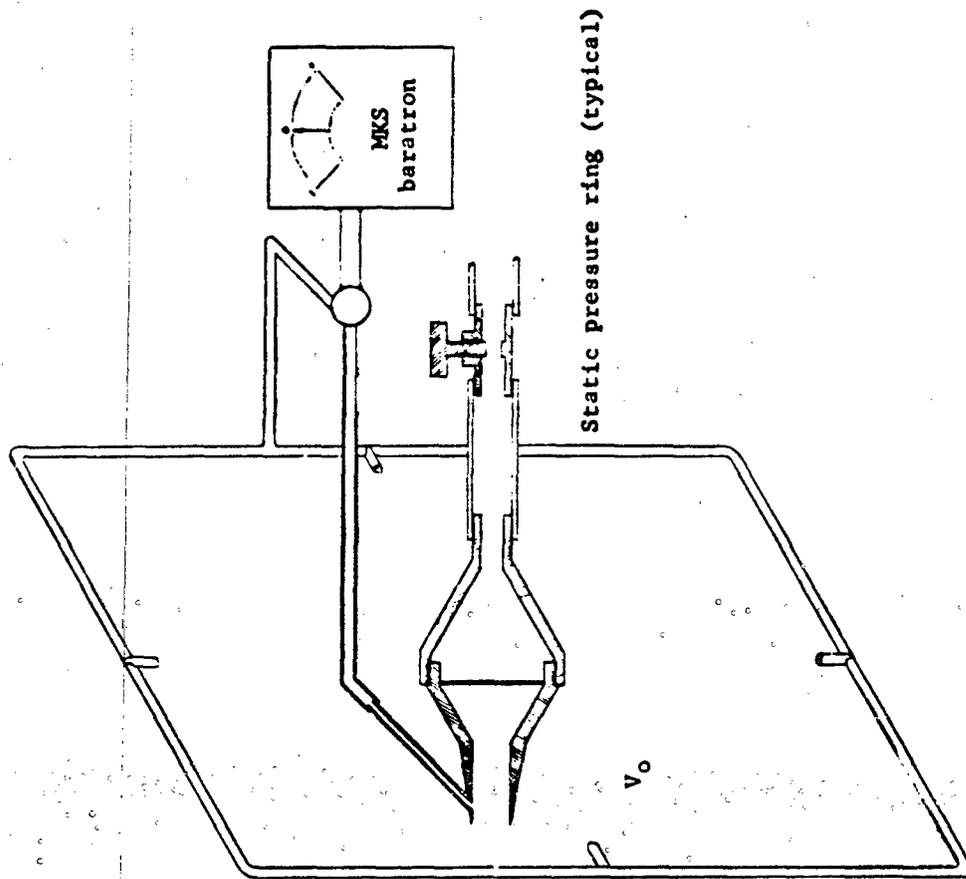
Figure 9. Isokinetic Probe
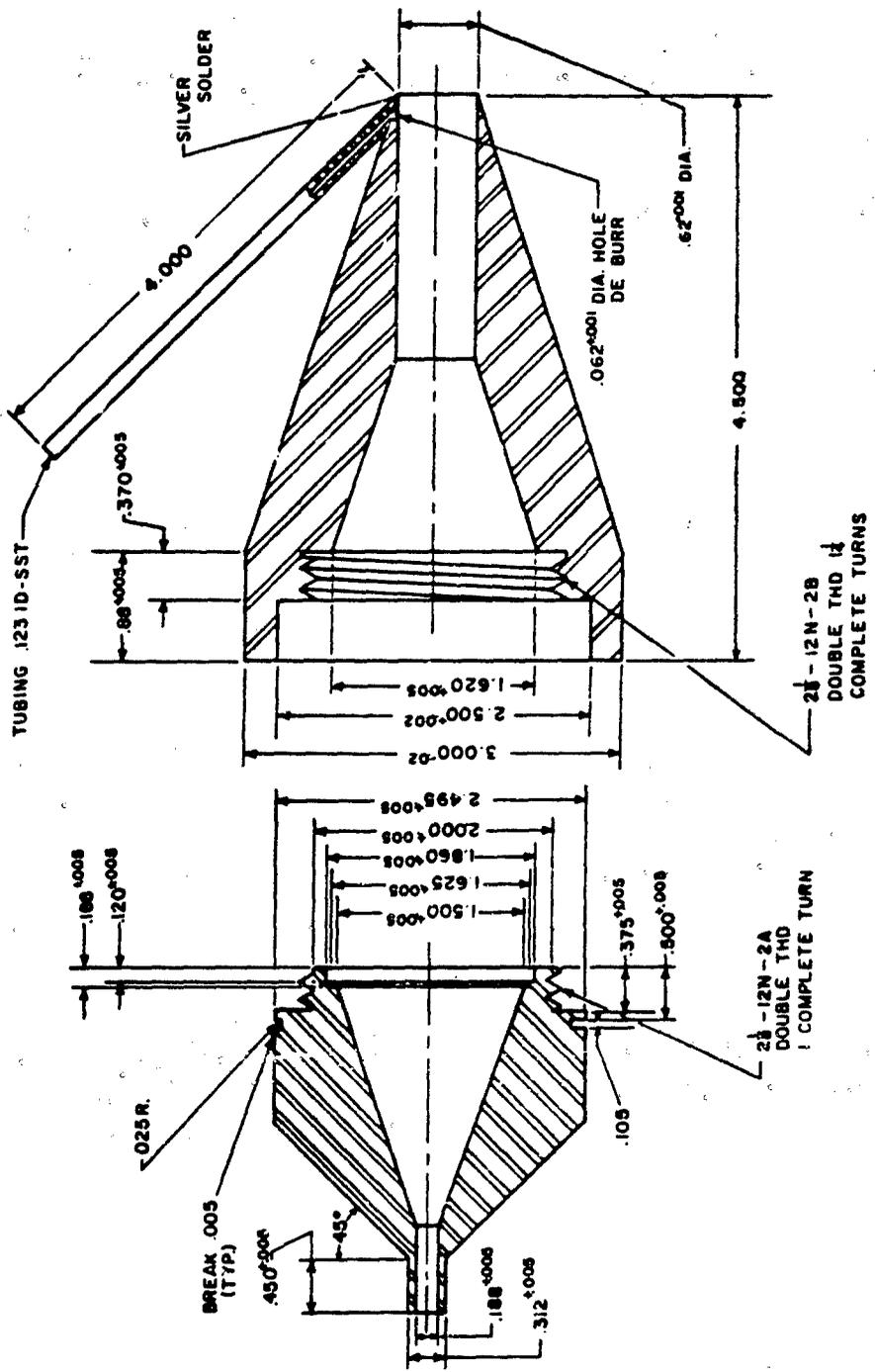
Figure 10. Design of Isokinetic Probe

The MKS Baratron was used to measure the static pressure difference ($p_0 - p$). The static pressure ring shown in Figure 9 failed to operate as constructed, so the static pressure (p_0) was measured using the static pressure tap of the no. 6 pitot tube shown in Figure 3. The sampler is isokinetic when the aspiration rate of the isokinetic sampler is adjusted so that $p_0 - p = 0$. This is done by adjusting the valve shown in Figure 9 and watching for the zero balance on the Baratron.

2.2.12.2 Results of Calibration of Isokinetic Sampler. The design of the isokinetic sampler was validated by measuring the flow rate through the sampler. The flow rate was measured using a Brooks rotameter. The wind tunnel was operated at a set wind velocity, and the sampler was balanced for the isokinetic conditions, using the Baratron as shown in Figure 9. The flow rate through the isokinetic sampler was then calculated using the product of the velocity in the tunnel and the inlet area of the sampler. The results are shown in Figure 11. At velocities above 11 miles per hour, critical flow develops in the vacuum lines and prevents isokinetic operation of the sampler. This can be overcome without difficulty by using larger tubing and a larger pump. The results indicate that the design of the isokinetic sampler is adequate and that it operates isokinetically over the range of interest for calibrating field samplers.

In all tests, the isokinetic sampler was checked for slippage through the sampler by placing an AGI sampler in the vacuum line during the sampler trials. This check showed no significant slippage.

2.2.12.3 Method Used to Determine Sampling Efficiencies. The physical model used to determine the comparative sampling efficiencies of field samplers versus the isokinetic probe is shown in Figure 12. The terms shown in the diagram of the physical model are defined below:

V_3 = the wind speed at the test section of the wind tunnel

C_0 = the unperturbed concentration of agent immediately in front of the sampler

V_s = the air speed immediately inside the orifice of the field sampler

A_s = the orifice area of the field sampler

A_r = the orifice area of the isokinetic probe.

-409-
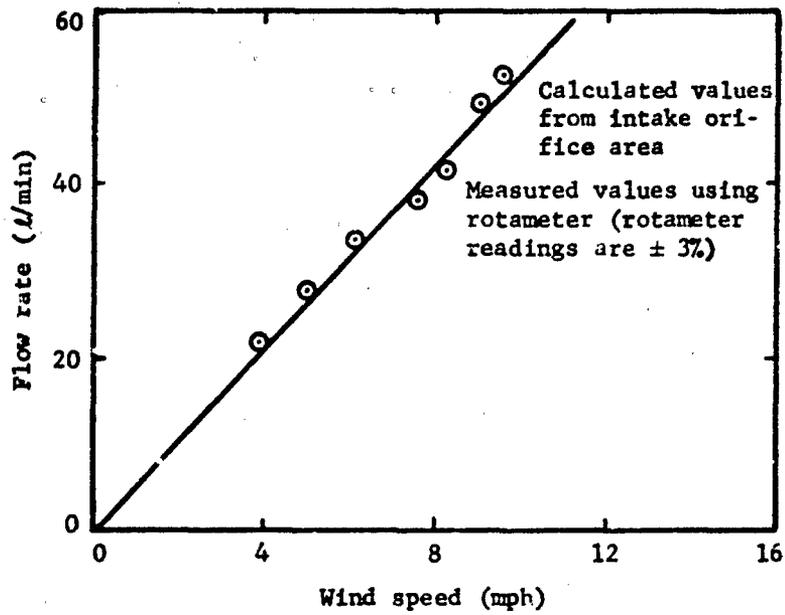
Figure 11. Flow Rate Through the Isokinetic Sampler
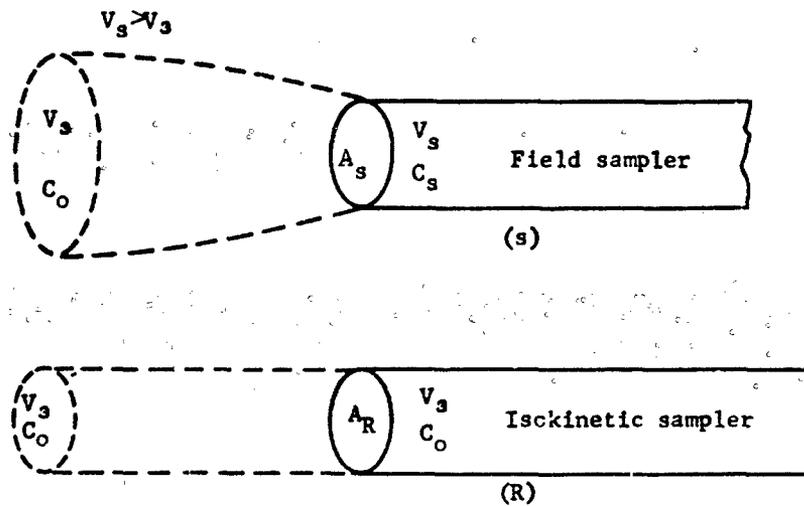
Figure 12. Model for Determination of the Collection Efficiency of Candidate Field Sampler

The masses collected by the two samplers are given by

$$m_s = \int V_s A_s C_s \, dt \qquad \text{Eq. (13)}$$

$$m_R = \int V_3 A_R C_0 \, dt \qquad \text{Eq. (14)}$$

Equation (13) can be rewritten

$$m_s = f_s \int \frac{C_s}{C_0} C_0 \, dt \qquad \text{Eq. (15)}$$

where f_s is the flow rate of the field sampler ($f_s = V_s A_s$). If the ratio C_s/C_0 is constant, Equation (15) becomes

$$m_s = fk \int C_0 \, dt \qquad \text{Eq. (16)}$$

and the efficiency of the field sampler is determined using Equations (14) and (16) in the form

$$K = \frac{\left(\dfrac{m_s}{f_s}\right)}{\left(\dfrac{m_R}{f_R}\right)} \qquad \text{Eq. (17)}$$

Equation (16) can be used to estimate the average concentration for a given sampling time by

$$\overline{C}_0 = \frac{m_s}{f_s K \Delta T} \qquad \text{Eq. (18)}$$

where ΔT is defined as the sampling time.

-411-

2.2.13 Sampler Calibration Analysis

The analysis is divided into two parts. The first analysis was conducted for the primary purpose of demonstrating the use of the isokinetic sampler as a standard. Once the isokinetic sampler was established as a standard, the second analysis was performed to determine the sampling efficiencies of various samplers, differences between samplers, and effect of wind speed and sampling times on sampler efficiency. The second analysis was also used in support of the work discussed in reference 8.

2.2.13.1 First Analysis. AGI and CI samplers were calibrated using isokinetic samplers as a standard. The purpose of this analysis was to demonstrate the technique of using the isokinetic sampler to calibrate field samplers. The testing array shown in Figure 13 was used in 21 trials. The munitions used for these trials were the pyrotechnic XM-100 cannister and the Mark-9 CS2 disseminator. The results of these trials are presented in Table 1. The sampling time for each trial was 30 seconds. The estimates of K are derived by the data reduction technique presented in paragraph 2.2.12.3. No statistical analysis was performed on these data.

O Field sampler

● Isokinetic sampler
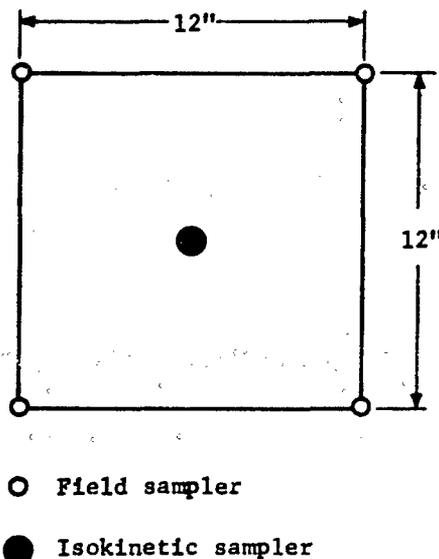
Figure 13. Sampling Array for First Analysis

-412-

Table 1. Preliminary Calibration of the AGI and CI Sampler

Trial	Sampler Type	Munition Type	Agent Released (gm)	Wind Velocity (mph)	$\dfrac{m_s}{m_R}$	K (effi- ciency)
1	AGI	XM-100	54.4	4.0	...	...
2	"	"	54.4	3.6	...	...
3	"	"	13.6	7.9	.14	.95
4	"	"	27.2	8.2	.14	.99
5	"	"	40.8	8.2	.13	.97
6	"	"	54.4	8.2	.14	.93
7	"	MK-9-CS2	5.0	8.1	.14	1.01
8	"	"	10.0	8.3	.17	1.20
9	"	"	15.0	8.1	.15	1.07
10	"	XM-100	54.4	11.1	.11	1.03
11	"	"	"	11.0	.11	
12	CI	"	"	11.1	.11	1.00
13	CI	"	"	11.1	.10	
14	AGI	"	"	7.2	.20	1.09
15	AGI	"	"	6.9	.20	
16	CI	"	"	7.1	.20	1.07
17	CI	"	"	7.1	.24	
18	AGI	"	"	5.1	.27	.98
19	AGI	"	"	5.1	.30	
20	CI	"	"	5.1	.32	.94
21	CI	"	"	5.1	.25	

2.2.13.2 Second Analysis. The data for the second analysis were obtained from a statistical design with four levels of sample time (5, 10, 15, and 30 seconds), three levels of wind speed (4, 8, and 11 miles per hour) and three samplers (CBI, CI, and AGI). The sampling efficiencies of the CBI, CI, and AGI samplers were determined by comparing the amount of agent collected by these samplers with the amount collected by isokinetic samplers. The sampler array is shown in Figure 14. The physical model used to determine the comparative sampling efficiencies of the CBI, CI, and AGI versus the isokinetic particulate samplers and the mathematical relationships involved are presented in paragraph 2.2.12.3. The average concentration for these tests was of the order of 1 to 3 x 10^{-4} milligrams per cubic centimeter. The concentrations were obtained by use of pyrotechnic XM-8 CS cannisters. The data obtained from these trials, together with the values obtained by the data-reduction technique explained in paragraph 2.2.12.3 are presented in Table 2. Statistical analysis performed on the collection efficiency estimates (Table 3) provided the following results:

a. Table 4 indicates the absence of significant differences for wind speeds or sampling times for trials of the CBI sampler.
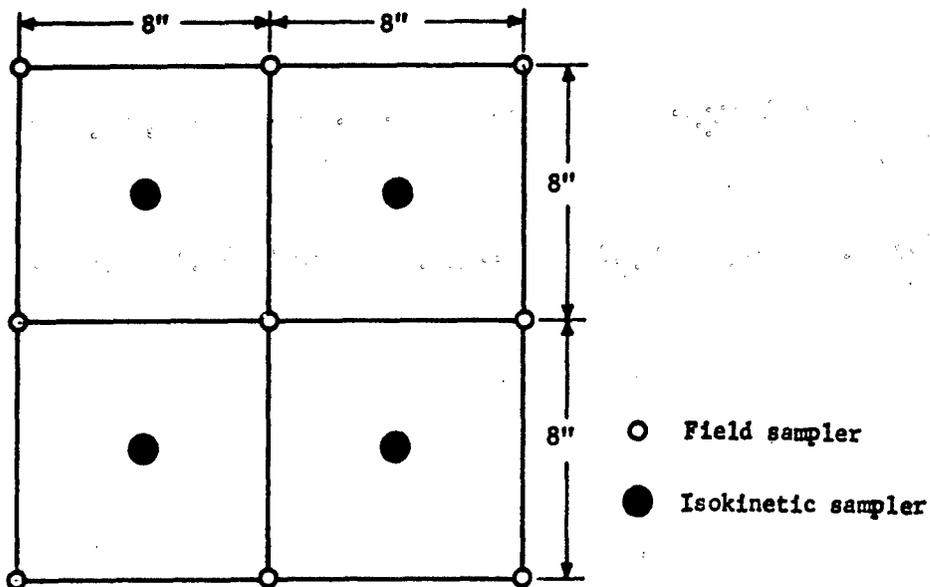
Figure 14. Sampling Array for Second Analysis

b. Table 5 indicates no significant differences for either wind speeds or sampling time for trials of the CI sampler.

c. Table 6 indicates highly significant differences in both wind speed and sampling times for the AGI sampler.

d. Table 7 presents a combined analysis of variance for the entire test. When all data are combined, no significant differences were attributable to any of the main effects or interactions.

e. Since CBI and CI samplers showed no significant differences among wind speeds or sampling times, all data over the range of trial conditions were combined, and the data were fitted, for each sampler, by means of regression analysis. Since the predicted lines should intersect at coordinates (0,0), the data were fitted to the equation form

$$y = Kx \qquad \text{Eq. (19)}$$

where y = sampler reading, mg-min/ℓ

K = slope (efficiency estimate)

x = isokinetic reading, mg-min/ℓ.

-414-

Table 2. Test Summary for Sampler Collection

Sample Time (sec)	Wind Speed (mph)	CBI (m/f)	Isokinetic (m/f)	K	Avg. K	CI (m/f)	Isokinetic (m/f)	K	Avg. K	ACI (m/f)	Isokinetic (m/f)	K	Avg. K
30	4	.175	.141	1.24		.136	.119	.98		.271	.258	1.05	
30	4	.179	.148	1.22		.134	.135	.99		.263	.233	1.13	
30	4	...	...	...		.190	.178	1.07		.265	.216	1.23	
30	4	.161	.176	.92	1.13	.165	.138	1.19	1.06	.269	.220	1.22	1.16
15	4	.120	.114	1.06		.036	.041	.90		.150	.143	1.05	
15	4	.107	.096	1.12		.028	.024	1.15		.159	.150	1.06	
15	4	.095	.082	1.17		.063	.065	.96		.167	.147	1.14	
.5	4	.088	.080	1.09	1.10	.041	.046	.89	.97	.179	.148	1.21	1.11
10	4	.066	.063	1.05		.038	.041	.91		.095	.085	1.11	
10	4	.072	.057	1.27		.036	.035	1.03		.093	.079	1.18	
10	4	.079	.071	1.11		.045	.047	.94		.107	.107	1.00	
10	4	.079	.076	1.05	1.12	.036	.030	1.19	1.02	.112	.097	1.15	1.11
5	4	.054	.050	1.08		.020	.015	1.37		.050	.046	1.09	
5	4	.045	.039	1.17		.015	.014	1.04		.048	.046	1.04	
5	4	.045	.038	1.19		.029	.020	1.44		.050	.043	1.16	
5	4	.040	.037	1.06	1.13	.020	.026	.78	1.16	.047	.046	1.03	1.08
30	8	.091	.117	.78		.194	.202	.96		.149	.140	1.06	
30	8	.076	.085	.90		.175	.165	1.06		.134	.115	1.16	
30	8	.146	.144	1.01		.176	.206	.85		.173	.168	1.03	
30	8	.097	.105	.93	.90	.137	.172	.80	.92	.135	.131	1.03	1.07
15	8	.042	.041	1.03		.090	.094	.96		.090	.100	.90	
15	8	.028	.026	1.09		.067	.061	1.09		.081	.076	1.06	
15	8	.070	.064	1.09		.103	.100	1.04		.105	.109	.97	
15	8	.044	.041	1.17	1.14	.072	.076	.95	1.01	.093	.085	1.10	1.01
10	8	.040	.041	1.00		.069	.074	.92		.049	.060	.82	
10	8	...	...	...		.052	.048	1.08		.048	.045	1.08	
10	8	.059	.056	1.06		.074	.061	1.21		.052	.049	1.06	
10	8	.037	.031	1.18	1.09	.046	.036	1.28	1.13	.050	.048	1.04	1.00
.5	8	.021	.024	.88		.031	.036	.87		.013	.017	.81	
5	8	.018	.018	1.00		.030	.031	.97		.010	.009	1.15	
5	8	.029	.029	1.00		.033	.028	1.17		.021	.023	.93	
5	8	.020	.016	1.28	1.04	...	...	...	1.04	.013	.012	1.04	0.98
30	11	.160	.164	.98		...	...	...		.139	.124	1.12	
30	11	.159	.141	1.13		.188	.195	.97		.142	.124	1.14	
30	11	.178	.166	1.07		...	...	...		.175	.156	1.12	
30	11	.135	.116	1.16	1.09	.218	.221	.99	.98	.144	.139	1.03	1.11
15	11	.067	.066	1.01		.111	.106	1.05		.035	.031	1.13	
15	11	.054	.052	1.06		.116	.096	1.24		.038	.038	.98	
15	11	...	...	...		.107	.107	1.00		.057	.058	.99	
15	11	...	...	...	1.04	.111	.109	1.02	1.08	.045	.041	1.09	1.05
10	11	.055	.053	1.05		.029	.028	1.04		.021	.019	1.09	
10	11	.048	.041	1.16		.043	.036	1.24		.022	.024	.92	
10	11	.070	.062	1.12		.046	.036	.83		.030	.033	.91	
10	11	.051	.042	1.21	1.13	.058	.071	.81	.97	.027	.026	1.04	0.99
5	11	.023	.026	.91		.036	.032	1.12		.021	.019	1.08	
5	11	.021	.019	1.14		.035	.029	1.23		.017	.017	1.00	
5	11	.032	.028	1.13		.027	.024	1.12		.031	.033	.93	
5	11	.023	.021	1.10	1.07	.027	.021	1.29	1.19	.021	.020	1.04	1.02

*In this table, m is defined as mass of CS in milligrams, and f is defined as flow rate in liters per minute. (For the isokinetic sampler, f equals 5.343 times the windspeed in mph.)

Table 3. Sampler Efficiency (Percent) for the Particulate Collectors
in Test C-136, Phase II, Wind Tunnel Trials

Sampling Time (sec)	Percent Efficiency for Indicated Sampler and Wind Speed								
	CBI			CI			AGI		
	4mph	8mph	11mph	4mph	8mph	11mph	4mph	8mph	11mph
5	113	104	107	116	104	119	108	98	102
10	112	109	113	102	113	97	111	100	99
15	110	114	104	97	101	108	111	101	105
30	113	90	109	106	92	98	116	107	111

Table 4. Analysis of Variance for the CBI Collector - Test C-136,
Phase II, Wind Tunnel Trials

Source	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Mean	1	140400.3333		
Wind speed (A)	2	120.1667	60.0834	[a]1.273
Sampling time (B)	3	86.3333	28.7778	[a]0.610
A x B	6	282.6667	47.1111	
Total	12	140890.0000		

[a]Not significant.

Table 5. Analysis of Variance for the CI Collector - Test C-136,
Phase II, Wind Tunnel Trials

Source	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Mean	1	130834.0833		
Wind speed (A)	2	22.1667	11.0834	[a]C.1669
Sampling time (B)	3	338.2500	112.7500	[a]1.6976
A x B	6	398.5000	66.4167	
Total	12	131593.0000		

[a]Not significant.

Table 6. Analysis of Variance for the AGI Collector - Test C-136,
         Phase II, Wind Tunnel Trials

Source	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Mean	1	134196.7500		
Wind speed (A)	2	213.5000	106.7500	[a]37.312
Sampling time (B)	3	139.5833	46.5278	[b]16.2621
A x B	6	17.1667		
Total	12	134567.0000		

[a]Significant at α = 0.0005.
[b]Significant at α = 0.005.

Table 7. Combined Analysis of Variance for the CBI, CI, and AGI
         Particulate Collectors - Test C-136, Phase II, Wind Tunnel
         Trials

Source	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Mean	1	405344.4444		
Sampler (A)	2	86.7222	43.3611	[a]2.299
Wind speed (B)	2	280.3889	140.1944	[a]2.816
A x B	4	75.4444	18.8611	[a]0.566
Sampling time (C)	3	49.1111	16.3704	[a]0.329
A x C	6	515.0556	85.8426	[a]2.575
B x C	6	298.7222	49.7870	[a]1.493
A x B x C	12	400.1111	33.3426	
Total	36	407050.0000		

[a]Not significant.

The K values (estimated percent efficiency for the respective sampler) with 95-percent confidence limits are presented in Figure 15.

The AGI data could not be analyzed in this manner because of the significant differences in both wind speeds and sampling times. Instead, these data were plotted in Figure 16 to indicate the significant differences between the two main effects: wind speeds and sampling time.

The first analysis showed a dependence of sensitivity of samplers on sampling time. The second analysis was performed in more detail and showed that the concentration of agent, rather than time, was the parameter on which sensitivity depends. For these tests, the dosage ranged from approximately 0.020 to 0.250 mg-min/l, sampling times ranged from 5 to 30 seconds, and wind speeds ranged from 4 to 11 miles per hour. The samplers were considered to measure concentration isokinetically with an average efficiency of 1.05.

2.3 REVIEW OF OTHER WORK

Previous experimentation has been conducted using isokinetic samplers and particles of various sizes. In a paper by H.H. Watson,[1] several significant relationships are presented. Figure 17 shows the relationship between C/C_0 and U_0/U as a function of particle size, where C = measured concentration, C_0 = true concentration, U = inlet air speed, and U_0 = stream velocity.

A similar graph (Fig. 18) depicts work done by Mayhood and Langstroth. The results are essentially the same as those obtained by Watson. The AGI and CI samplers were checked over a range of values U_0/U from .5 to 2.5. The values for C/C_0 were approximately 1 for all values of U_0/U. These data are consistent with Figures 17 and 18 for particles 1 micron in diameter. Figure 19 depicts the relationship between C/C_0 and the angle between the inlet tube and wind direction. This graph shows the effect of turning the sampler at various angles to the windstream. This has very little effect on the efficiency up to angles of 30 degrees.

Figure 20 is taken from a report by K.R. May and H.A. Druett[2] which shows how the intake efficiency of the AGI with a preimpinger changes with wind speed. Figure 20 shows that for 1-micron diameter particles, the efficiency is approximately 1 for all wind speeds, even though the preimpinger intake orifice is on a 45-degree angle with airstream. This is consistent with the results of Mayhood (Fig. 19).

2.3.1 Relationship of Present Results to Other Work

At U_0/U = 1.5 for 12-micron diameter particles, C/C_0 = 1.1 in Figure 17 and 1.15 in Figure 18. It can be concluded that a scatter of

-418-
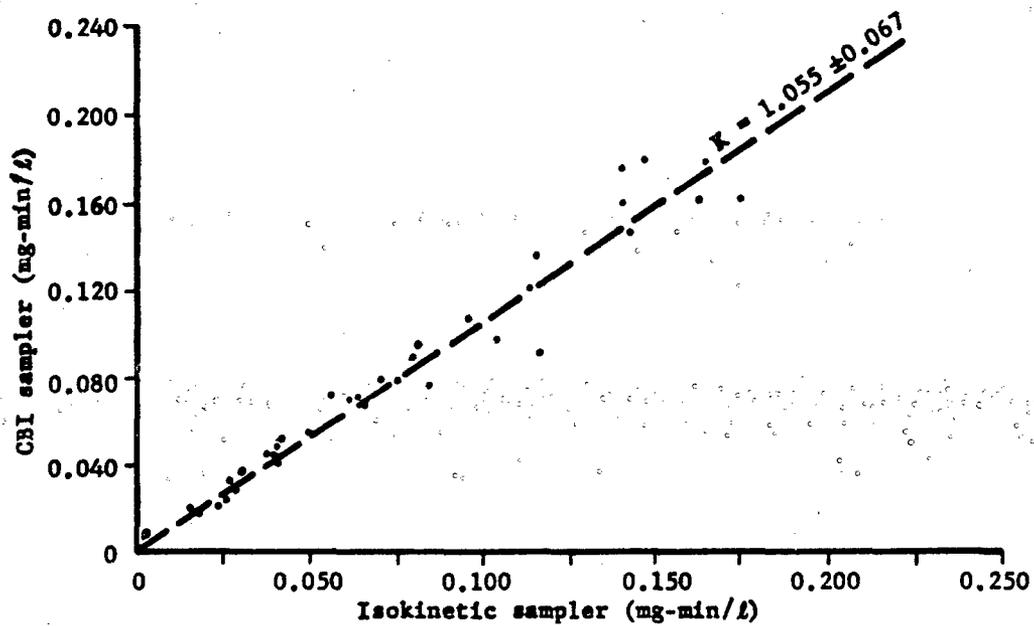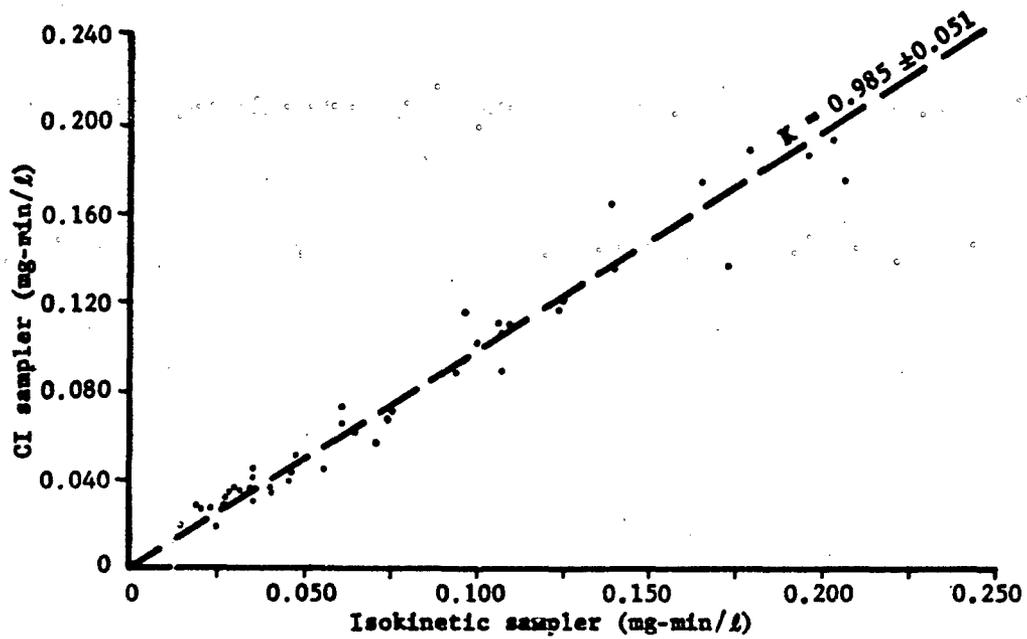
Figure 15. Isokinetic Sampler vs. CBI and CI Sampler (Regression Analysis)
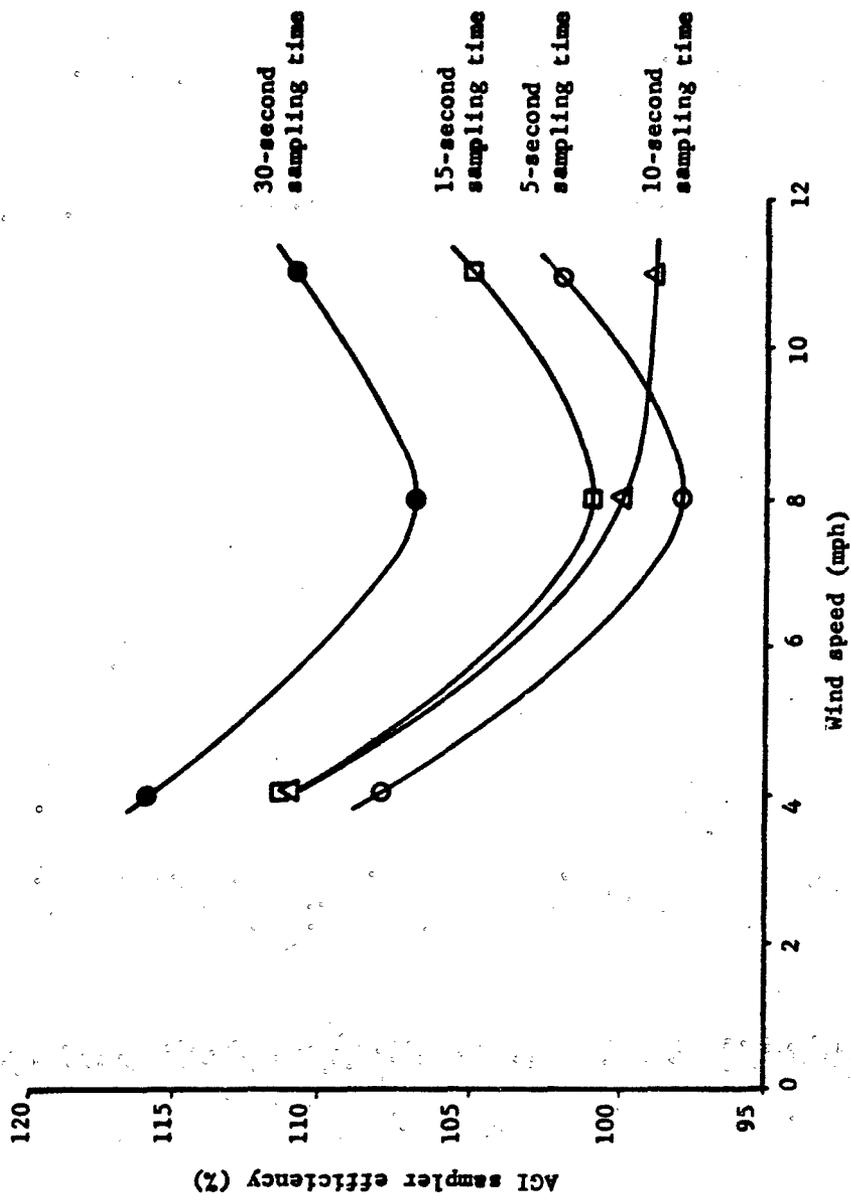
Figure 16. AGI Sampler Efficiency Versus Wind Speed and Sampling Time
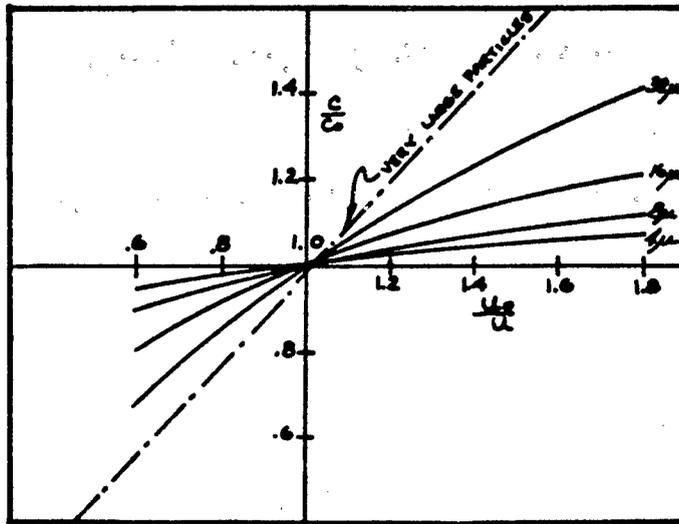
Figure 17. Work Performed by Watson[1]
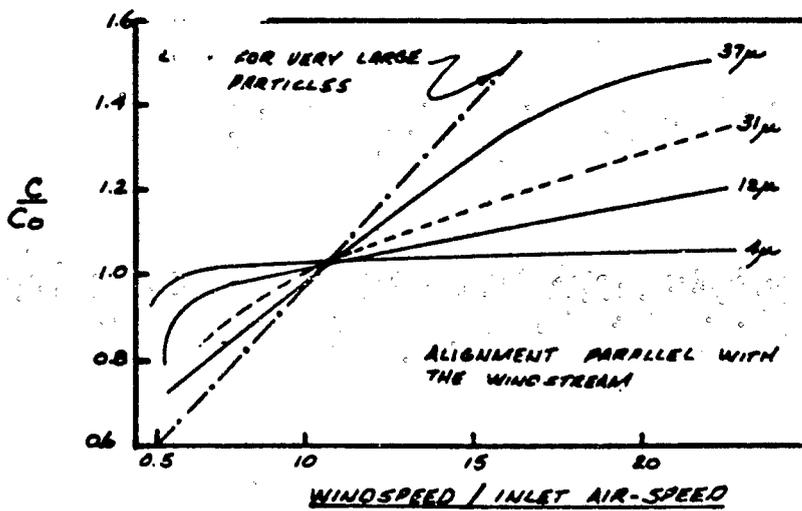
Figure 18. Work Performed by Mayhood and Langstroth[1]
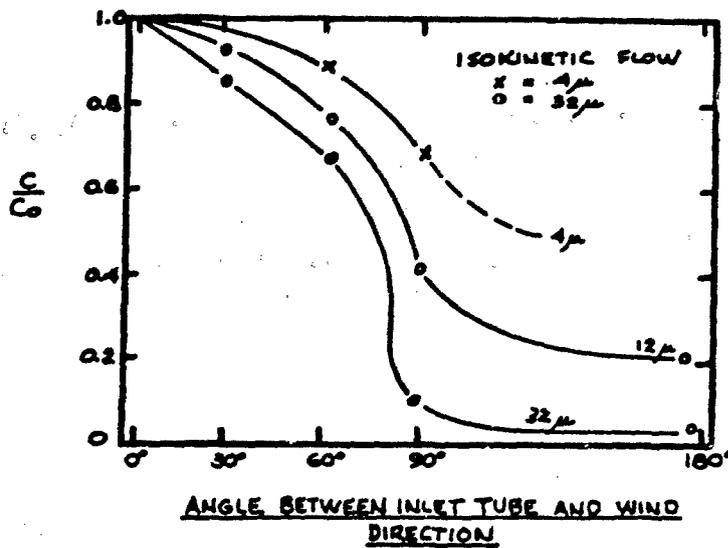
Figure 19. Isokinetic Sampling Factors with the
Inlet Tube Aligned at Various Angles
to the Windstream (after Mayhood and
Langstroth)[1]
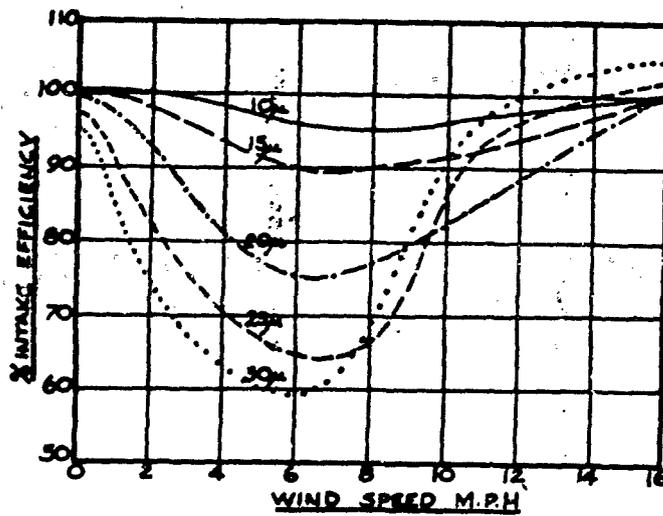
Figure 20. Preimpinger Data Collection by
May and Druett[2]

5 percent, at most, is likely in experiments of this kind. The values obtained for K in this analysis do not differ significantly from the previously reported values. (See Tables 1 and 2 for K estimates for this analysis.) Figure 19 indicates that the scatter in K is not caused by the irregularity in the air flow. Before the tunnel was modified, sampler efficiencies of 0.75 and 0.63 were determined for the CI and the AGI samplers.[7] In interpretating these results, it must be remembered that the airstream was very irregular and only three samplers were used; hence, it was difficult to get an accurate mass balance. In both the present tests and in those reported in reference 7, the source of the agent was a pyrotechnic munition. This procedure introduces added uncertainties in the experiment. Nevertheless, improving the quality of the air flow seems to lead to reproducible agent concentration profiles, as determined by repeated experiments.

2.3.2 Summary

The modifications that were made to improve the uniformity of the air stream are shown to be adequate for sampler calibration until the time that other uses of the facility require a lower level of turbulence in the air flow.

The isokinetic sampler functioned as designed. The preliminary calibration of the AGI, CI, and CBI samplers indicate all samplers have an efficiency of about one over a velocity range from 4 to 12 miles per hour. The results obtained for these samplers using the isokinetic sampler as a standard are consistent with the results of previous investigators.[1,2] It is therefore recommended that the isokinetic sampler developed in this study be used as a standard to calibrate field samplers at DPG. However, the AGI sampler collected significantly different amounts of agent as wind speeds changed and also as the sampling times changed. No such difference was noted with the CBI and CI samplers. The ratio of the amount of agent collected by the field sampler and that collected by the isokinetic sampler is used to determine the efficiency of the field samplers. The calibration then becomes independent of the amount of agent released, provided the samplers are not overloaded with agent. Thus, the concept of using an isokinetic sampler eliminates many of the uncontrolled variables that create difficulties in application of the mass-balance technique of sampler calibration.

2.4 REFERENCES

1. Watson, H.H., "Errors Due to Anisokinetic Sampling of Aerosols," *American Industrial Hygiene Association Quarterly Report*, 1969.

2. May, K.R., and Druett, H.A., "The Pre Impinger," *British Journal of Industrial Medicine*, 1953.

3. Pope, Alan, and Harper, John J., *Low-Speed Wind Tunnel Testing*, 1966.

4. Barnes, W.D., and Peterson, E.G., "An Investigation of Flow Through Screens," *Transactions of ASME*, July 1951.

5. Smith, F.B., and Hay, J.S., The Expansion of Clusters of Particles in the Atmosphere, *Quarterly Journal of Royal Meteorological Society*, 1961.

6. Batchelor, G.K., *Homogeneous Turbulence*, Cambridge University Press, 1956.

7. DF, STEPD-TT-JP(S), to STEPD-TT-CD(D), Deseret Test Center, Fort Douglas, Utah, 16 April 1970, subject: *Statistical Analysis of CT-26 Wind Tunnel Sampler Technology Trials*.

8. Deseret Test Center, Fort Douglas, Utah, *Determination of Minimum Sampling Intervals for Aerosolized Riot Control Agent CS, Final Report*, by Wilbert T. Taylor, Harry S. Sloane, E.G. Peterson, and William C. McIntyre, DTC-FR-73-C135(I, II), August 1972.

TECHNIQUES FOR TAIL LENGTH ANALYSIS

James J. Filliben
National Bureau of Standards
Washington, D.C.

This paper concerns itself with the problem of estimating from a set of data the tail length of the underlying distribution. A probability plot technique for such distributional analysis is developed which makes use of order statistic medians. The probability plot correlation coefficient r_D for a distribution D is introduced which gives a statistical measure of probability plot linearity. The output from a computerized version (written in machine-independent ANSI Fortran) of the proposed tail length analysis procedure is illustrated. Three examples are discussed.

Introduction

The purpose of this paper is to present various techniques for the problem of determining, for a given data set, what the appropriate underlying distribution is. Although we shall ostensibly be constraining ourselves to the univariate problem where

$$\text{response} = \text{unknown constant} + \text{random error}, \qquad (1)$$

the results obtained and the procedures advocated are applicable through residual analysis to a wide variety of more general response models.

Let us first of all review briefly the typical measurement process situation. We start of course with data. The goodness of the data in reflecting the phenomenon under study is a function of the experimental design and of the expertise of the experimenter. The n observations collected are our only real contact with the phenomenon under study. In effect our data says to us: "I want to tell you about such and such a phenomenon, and, incidentally, about such and such an experimenter." Considered in this light, our data is seen as something to be analyzed and not just summarized. Anyhow, let us assume at the beginning here that we have n independent (a problem unto itself) observations Y_1, Y_2, ..., Y_n; this constitutes our data set.

At this point the statistical approach is imposed which assumes that there exists a random variable Y with some unknown distribution which corresponds to the population of all possible responses for the phenomenon under study. This distribution has some location and scale, defined by the arbitrary parameters μ and σ, say.

Frequently at this stage the experimenter is content to concentrate on inquiries about the two parameter values only. He says in essence: "Look, μ and σ are the only things I am interested in; these are the only things I report; so who cares about other details."

The response to this twofold: 1) knowledge of μ and σ does not uniquely characterize a distriubtion -- we can have an infinity of different distributions, with a wide assortment of probability function or density function shapes -- but with the same values for μ and σ. Knowing only two pieces of information (location and scale) yields only partial knowledge of our measurement process. It is of interest in itself from both a physical and statistical point of view to know as much about our measurement process as possible. This is information which is presently obtainable -- all we have to do is take the time to ask, and to ask the right questions. 2) In general, of course, μ and σ will never be known exactly -- we would need an infinity of observations under constant experimental conditions to have this -- thus we must estimate μ and σ. But <u>not all estimators are equally good and the goodness of a particular estimator is a function of what the actual unknown underlying distribution is</u> (see

-427-

references 4 and 6). For example, as shown in exhibit 1, the commonly used sample mean is excellent for the moderate-tailed normal distribution, but is poor for the short-tailed uniform distribution and is very poor for the long-tailed Cauchy distribution. Similarly, the sample midrange is good for short-tailed distributions, but poorer for longer-tailed distributions, and the sample median is good for long-tailed distributions, but poorer for shorter-tailed distributions. Thus, if we do not know what the underlying distribution is, or we do not thoroughly inquire as to what it is, then a given estimator of location, say, the sample mean for example, may or may not efficiently estimate μ; based on no distributional information, it will simply be a pot luck situation. However, if we do know what the underlying distribution is, or if we can fairly accurately "estimate" what the underlying distribution is, then by making use of standard mathematical statistical estimation results, we can choose an estimator which we know will efficiently estimate μ (and σ).

Probability Plots and Probability Plot Correlation Coefficients

Having concentrated on the necessity of a detailed distribution/ tail length analysis, we now address ourselves to the problem of how to estimate the tail length of the underlying distribution. What we have had success with and recommend as a procedure for "estimating" the underlying distribution is based on probability plots (see references 2, 3, 5, 10, and 13), order statistic medians, and Tukey's lambda distribution family (see references 6, 7, and 9). As used in this paper, a probability plot for a given distribution D is defined as a plot of the ith ordered observation X_i (that is, the ith order statistic) versus some measure of location $loc(X_i;D)$ of the ith ordered observation from that given distribution D. The ith ordered X_i is, of course, a function of our data set; the location $loc(X_i;D)$ of the ith ordered observation from a given distribution D is <u>not</u> a function of our data set -- it is a number computed from mathematical statistical considerations and is dependent only on the value of i, the sample size n, and the hypothesized distribution D. If, in fact, our data set was generated from the hypothesized underlying distribution D, then aside from an unimportant location and scale factor, X_i will be approximately equal to $loc(X_i;D)$ for all i, and so the plot of X_i versus $loc(X_i;D)$ will be approximately linear. If our data set was generated from another underlying distribution $D' \neq D$, then X_i will not approximate $loc(X_i;D)$ for all i and so the plot will tend to be non-linear (but of course in this latter case, X_i will approximate $loc(X_i;D)$ and so the probability plot for D' will tend to be linear). Probability plots are thus obtainable for any distribution D in which $loc(X_i;D)$ is computable.

One reason why non-normal probability plots have not in the past been commonly used is because $E(X_i;D)$ (the expected value of the ith ordered observation for a distribution D) has been used for $loc(X_i;D)$ -- but the use of the expected value has the distinct disadvantage of requiring a special integration technique for each different distribution D. Also, $E(X_i;D)$ does not always exist for some of the longer-tailed distributions. These problems may be circumvented by the use of the median $med(X_i;D)$ of the ith ordered observation from a distribution D for $loc(X_i;D)$ rather than using the expected value. First of all, we have no existence problems with $med(X_i;D)$ -- the median exists for all i, n, and D. Secondly, to compute $med(X_i;D)$ -- for any given i, n, and D, we need only transform (via the percent point function (see references 6 and 7) $G_D(p) = F_D^{-1}(p)$ of the given distribution D) the corresponding order statisic medians of the uniform distribution on [0,1]; that is,

$$med(X_i;D) = G_D(med(X_i;U))$$

where $G_D(p)$ is the percent point function (the inverse cumulative distribution function) of the distribution D and where $med(X_i;U)$ is the median of the ith ordered observation from the uniform distribution on the interval of 0 to 1. Thus we have a theoretically unified, as well as a computationally simple, approach to forming probability plots for any distribution D.

These probability plot considerations have been implemented into a machine-independent ANSI FORTRAN subroutine. Probability plots for 4

-430-

basic symmetric distributions of widely differing tail length are generated: the uniform (short-tailed), the normal (moderate-tailed), the Tukey $\lambda = -0.5$ (moderate-long-tailed), and the Cauchy (long-tailed). In addition, in order to get a more detailed and global picture of the distribution space, we have made use of the probability plot correlation coefficient, r_D, which for a given distribution D is here defined as the product moment correlation coefficient of the ordered observations and the medians of the ordered observations from distribution D; more precisely,

$$r_D = Corr(X_i, m_i) \qquad (3)$$

$$= \frac{\Sigma(X_i - \bar{X})(m_i - \bar{m})}{\sqrt{\Sigma(X_i - \bar{X})^2 \Sigma(m_i - \bar{m})^2}} \qquad (4)$$

where X_i is the observed i^{th} ordered observation and where $m_i = med(X_i; D)$ is the median of the ith ordered observation from the distribution D. (Simpler expressions for r_D, making use of symmetries, do exist but are extraneous to the central theme and so will not here be presented.) The rationale behind the probability plot correlation coefficient is of course that probability plots indicate affirmative results when they are linear, and a simple measure of linearity is the correlation coefficient. In the above mentioned subroutine, the probability plot correlation coefficient has been computed for 44 symmetric distributions -- ordered by tail length from very short-tailed U-shaped distributions to extremely long-tailed distributions. "Standard" distributions such as the uniform, normal, logistic, double exponential, and Cauchy have been included; the remaining distributions are from the Tukey λ distribution family.

Examples

Example 1. Josephson Junction cryothermometry voltage counts.

700 counts were observed; each count was proportional to voltage output from a Josephson Junction in a cryothermometry experiment. To give the reader a "feel" for the data, a plot of the data versus time is included (exhibit 2). The principle feature exhibited in the data plot is the discretness of the data set — there are only 8 distinct values which the counts have taken on. The first page (exhibit 3) of the automatic output from the computerized tail length analysis focuses on the specific assumption of normality. Five commonly used test statistics are included. In addition, the normal probability plot correlation coefficient is also included. On the basis of the standardized test statistics (given in column 4), there are conflicting and inconclusive indications of normality and non-normality. However, we really should not be attempting to conclude that a data set was generated from some particular distribution in the first place; rather, we should be attempting to determine an admissible set of distributions (all of more or less the same tail length) from which it is plausible that the observed data set could have been generated. Hence, although the normal test statistic results are informative, they are not to be weighted as heavily as the following probability plot considerations.

Uniform, normal, $\lambda = -0.5$, and Cauchy probability plots are shown in exhibit 4. The discretness of the data is also evident here. More important, from the plots and from the associated probability plot correlation coefficients (given at the bottom of each plot), it is clear that the underlying distribution is apparently near-normal. This conclusion is substantiated

in exhibit 5 in which we see the probability plot correlation
coefficient remaining fairly stable in the short-tailed region,
maximizing itself in the moderate-tailed region, and then decreasing
in the moderate-long and long-tailed region. Although the maximum
correlation happens to be for the normal distribution, a more
reasonable conclusion than exact normality is that underlying
distribution is simply near-normal. This conclusion is important
in itself; further, applying this conclusion to the problem of
location and scale estimation, we may thus feel reasonably safe in
using the sample mean or sample trimmed mean (with only a small
amount of trimming) to efficiently estimate μ, and to use the sample
standard deviation to efficiently estimate σ. Better yet, if we
consider the set of 11 distributions from $\lambda = +0.7$ to $\lambda = -0.1$
as an admissible set (where the criterion of admissibility was the
probability plot correlation coefficient being in excess of 96
percent) then a valid alternative course of action which would
protect us even more is to use a robust estimator over the
admissible set (see reference 6).

Example 2. Wind Velocities.

1200 coded wind velocities taken at one tenth of a second intervals
were observed. The data plot (exhibit 6) indicates a possible drift
and/or a possible shift in location in the second half of the data.
Other analyses (not here described) detected the existence of non-
randomness and low frequency components; these findings will be ignored
for the present for the sake of illustrating the point at hand; viz.,
the tail length analysis procedure. We first note the normal test
statistics in exhibit 7 -- they are inconclusive. From the 4 prob-
ability plots of exhibit 8, we see that the data set as a whole appears
moderate-tailed in spite of the fact that the randomness assumption
is untenable. Also, since the uniform probability plot is "~" shaped,
whereas the λ = -0.5 probability plot is "S" shaped, this implies that
the linear crossover will occur in the set of distributions between
the uniform and the λ = -0.5 distribution. Exhibit 9 shows that this
is in fact the case: the probability plot correlation coefficient
is maximized by the λ = +0.3 distribution which is a moderate-tailed
distribution very close to the normal distribution. For this data
set, we therefore might conclude that the underlying distribution
is moderate-tailed or possibly short-moderate-tailed. The sample
mean and sample standard deviation will probably be fairly efficient
though not optimally efficient. For this particular data set, of
course, where there is obviously much inherent structure to be taken
into account, a more detailed analysis (on, say, the differences
$Z_i = Y_i - Y_{i-1}$) is recommended.

-434-

Fxample 3. Deflections of a Steel-Concrete Beam.

200 coded deflections of a steel-concrete beam when subjected to
a periodic force by an attached piston were observed. The data plot
is given in exhibit 10. Five of the 6 normal test statistics in
exhibit 11 are in excess of 4 standard deviations from their expected
value, thus suggesting non-normality. From exhibit 12 we see that the
uniform probability plot is more linear than any of the others; also,
we note that the provability plots are "S" shaped for all 4 distributions
indicating that no intermediate crossover will take place and that the
"true" underlying distribution is even shorter-tailed than the uniform.
Exhibit 13 confirms this conclusion; it is seen that the maximum prob-
ability plot correlation coefficient is given for the $\lambda = +1.5$
distribution which is a finite-domain, short-tailed, U-shaped distri-
bution. The use of the sample mean and sample standard deviation
in this case would have been extremely inefficient -- the sample midrange
and sample range are preferable. Other analyses (not here included)
indicated a strong non-random sinusoidal component; this corroborates
the tail-length analysis because when the generating model is cyclic,
as detected, then it is characteristic for the distribution of the
responses to be U-shaped.

Concluding Remarks

Data analysis techniques based on probability plots, order statistic medians, Tukey's lambda distibution, and probability plot correlation coefficients have been proposed. By use of such techniques, useful information may be obtained regarding the distributional structure of a given data set. Such distributional information provides important guidance in the choice of efficient estimators of location and scale. The implementation of these techniques into an internally-documented, machine-independent FORTRAN subroutine allows such an analysis to be easily and routinely performed.

References

1. Blom, G. (1958). _Statistical Estimates and Transformed Beta-Variables_. Almqvist and Wiksell, Uppsala, Sweden; Wiley, New York.

2. Chernoff, H. and Lieberman, G. J. (1954). Use of normal probability paper. _J. Amer. Statist. Assoc._ __49__, 778-785.

3. Chernoff, H. and Lieberman, G. J. (1956). The use of generalized probability paper for continuous distributions. _Ann. Math. Statist._ __27__, 806-818.

4. Crow, E. L. and Siddiqui, M. M. (1967). Robust estimation of location. _J. Amer. Statist. Assoc._ __62__, 353-389.

5. David, H. A. (1970). _Order Statistics_. Wiley, New York.

6. Filliben, J. J. (1969). Simple and robust linear estimation of the location parameter of a symmetric distribution. Ph.D. Dissertation, Princeton University.

7. Filliben, J. J. (1973). The percent point function. To appear.

8. Hastings, C., Jr., Mosteller, F., Tukey, J. W., and Winsor, C. P. (1947). Low moments for small samples: a comparative study of order statistics. _Ann. Math. Statist._ __18__, 413-426.

9. Joiner, B. L. and Rosenblatt, J. R. (1971). Some properties of the range in samples from Tukey's symmetric lambda distribution. _J. Amer. Statist. Assoc._ __66__, 394-399.

10. Kimball, B. F. (1960). On the choice of plotting positions on probability paper. _J. Amer. Statist. Assoc._ __55__, 546-560.

11. Sarhan, A. E. and Greenberg, B. G. (Eds.). (1962). _Contributions to Order Statistics_. Wiley, New York.

12. Shapiro, S. S., Wilk, M. B. and Chen, H. J. (1968). A comparative study of various tests for normality. _J. Amer. Statist. Assoc._ __63__, 1343-1372.

13. Wilk, M. B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. _Biometrika_ __55__, 1-17.

UNIFORM (short tailed)

		Releff to BLUE
Midrange	good	100%
Mean	poor	26%
Median	poor	10%

NORMAL (moderate tailed)

Midrange	poor	35%
Mean	good	100%
Median	fair	68%

CAUCHY (long tailed)

Midrange	poor	0%
Mean	poor	0%
Median	good	90%
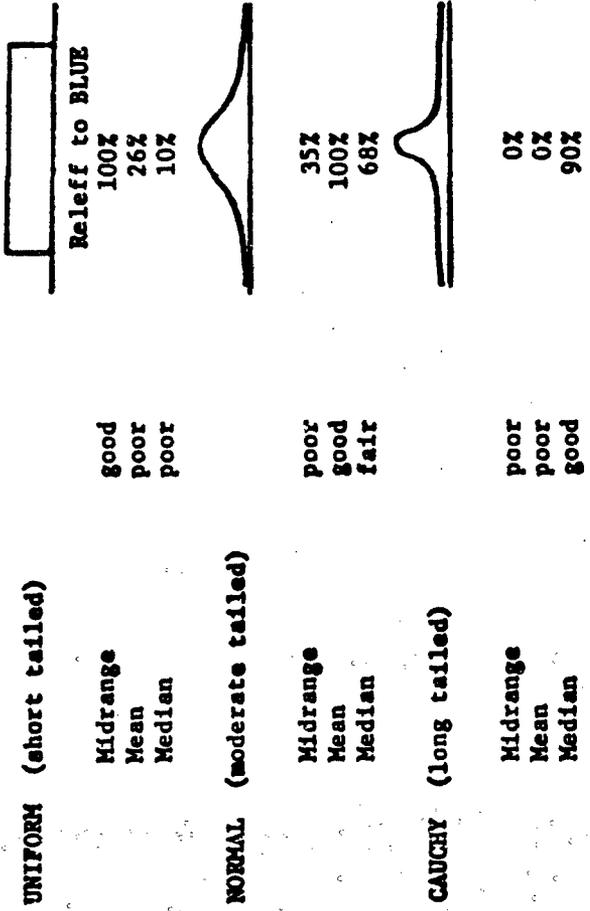
Exhibit 1. Location estimation and tail length
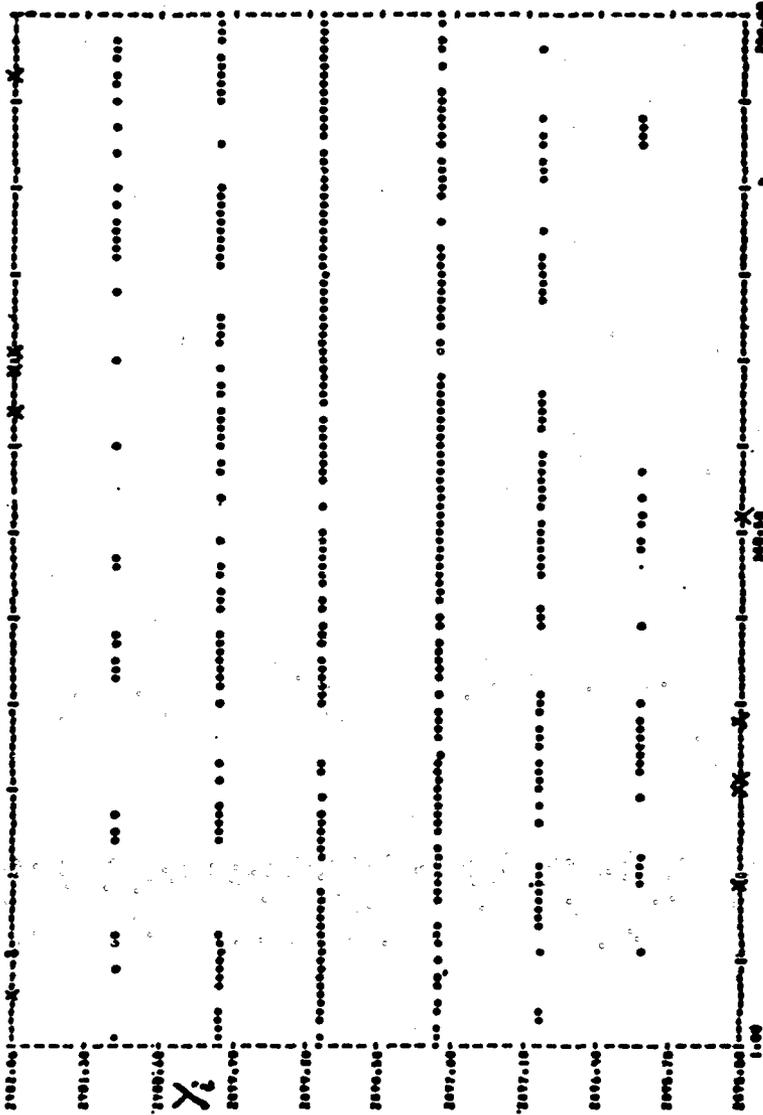
Exhibit 2. Voltage counts data plot

VOLTAGE COUNTS

Exhibit 3. Voltage counts tail length analysis
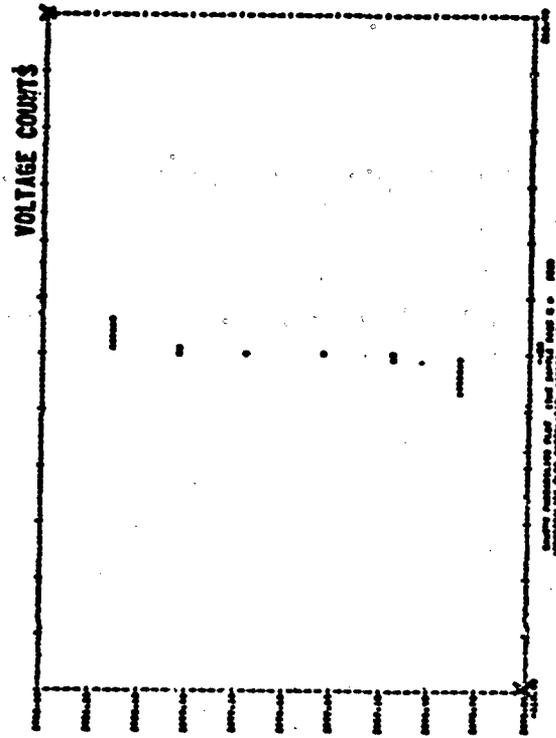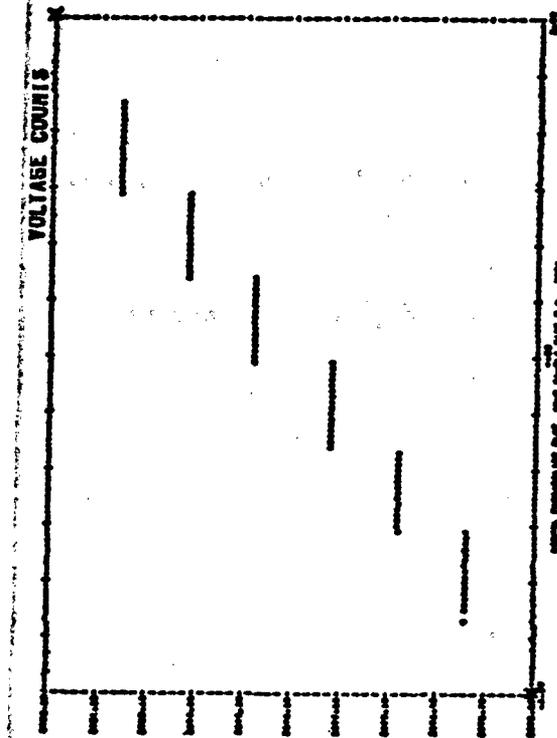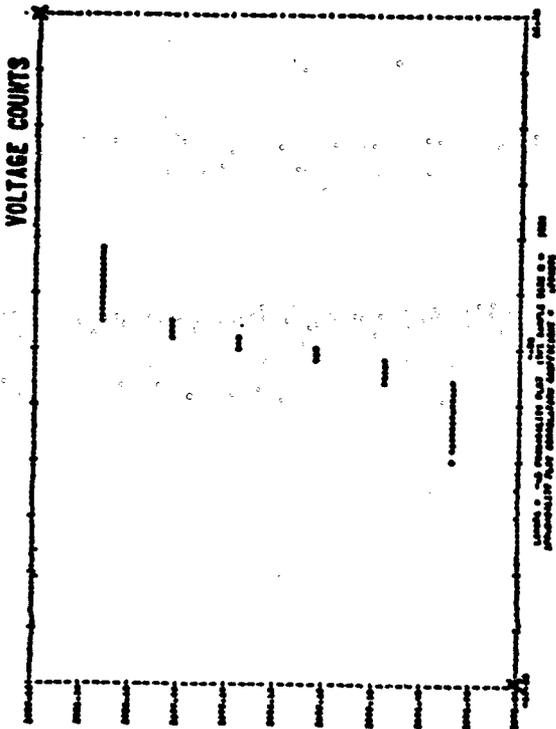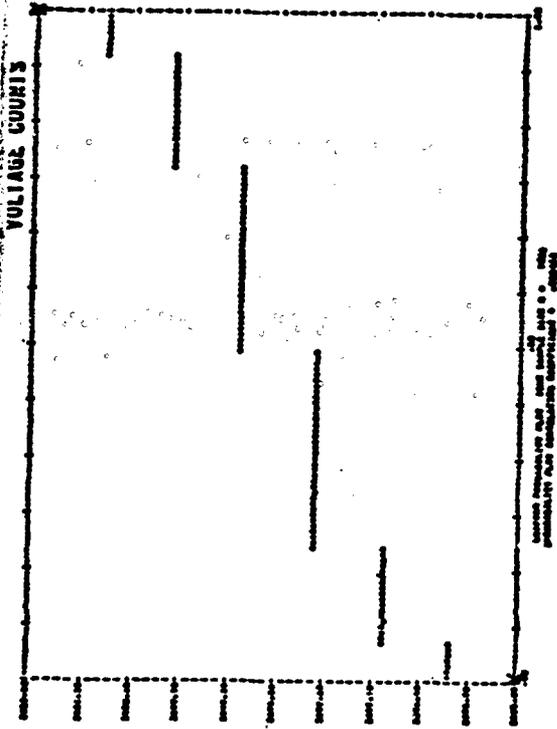
Exhibit 4. Voltage counts tail length analysis

VOLTAGE COUNTS

Exhibit 5. Voltage counts tail length analysis
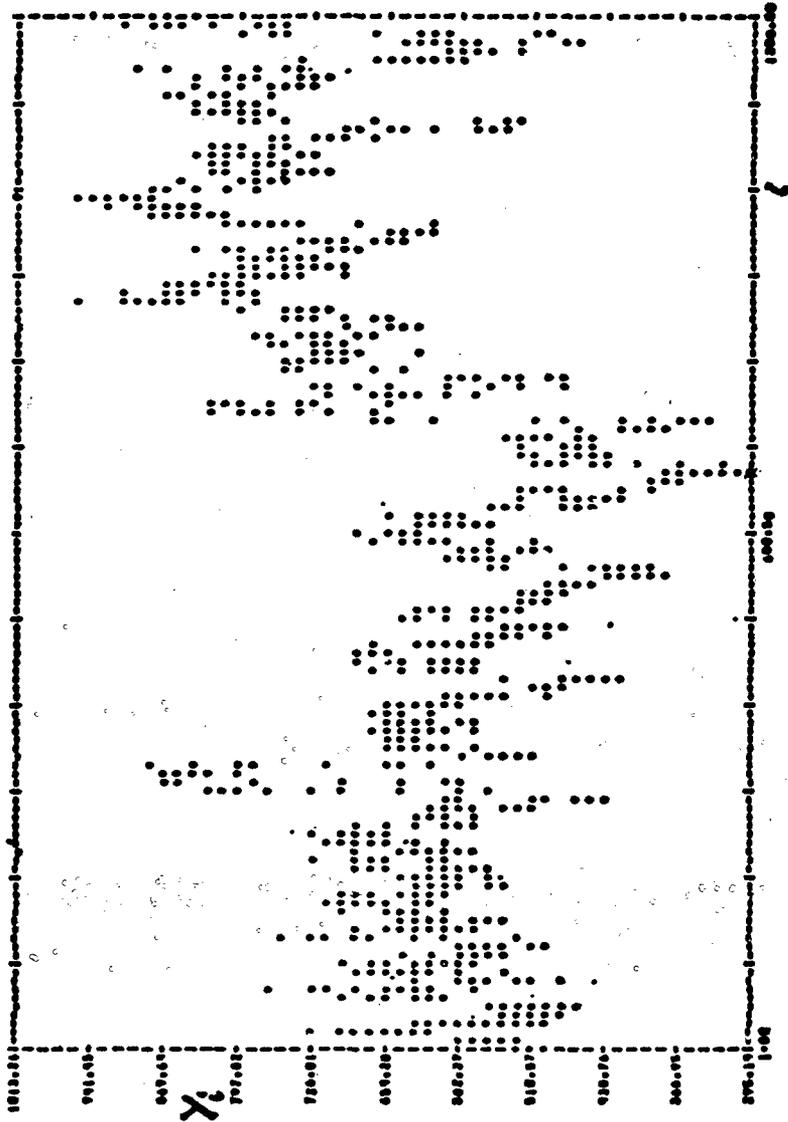
WIND VELOCITIES

Exhibit 6. Wind velocities data plot

-443-

WIND VELOCITIES

Exhibit 7. Wind velocities tail length analysis

WIND VELOCITIES

WIND VELOCITIES
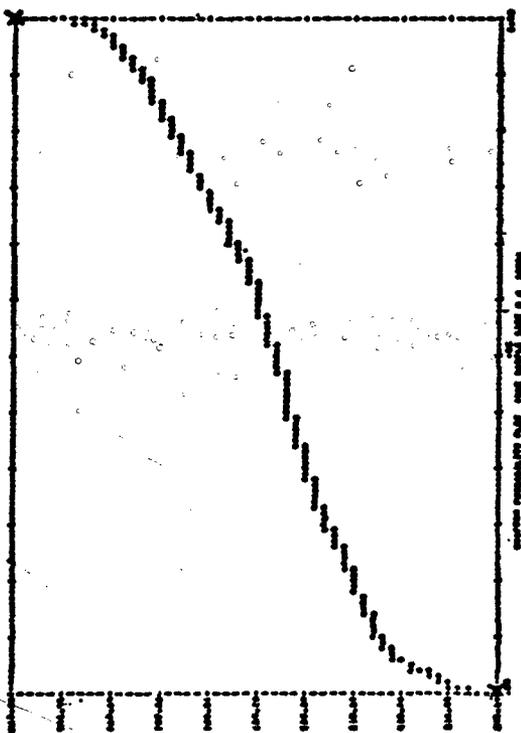
WIND VELOCITIES
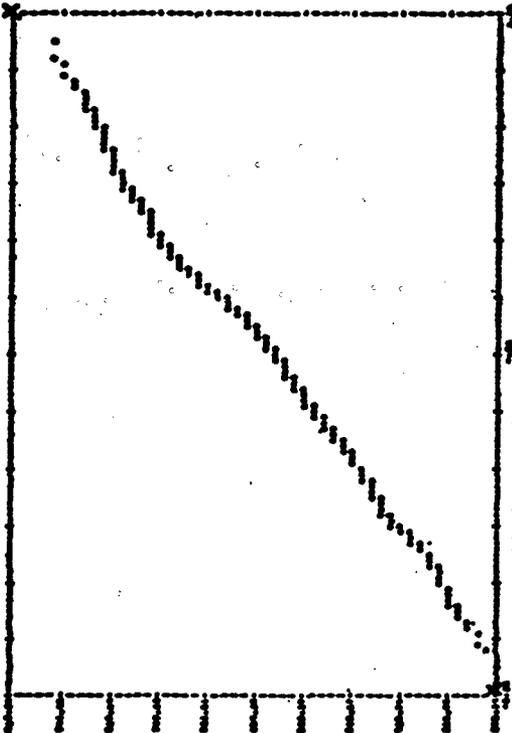
WIND VELOCITIES

Exhibit 8:   Wind velocities tail length analysis

-445-

WIND VELOCITIES

Exhibit 9. Wind velocities tail length analysis
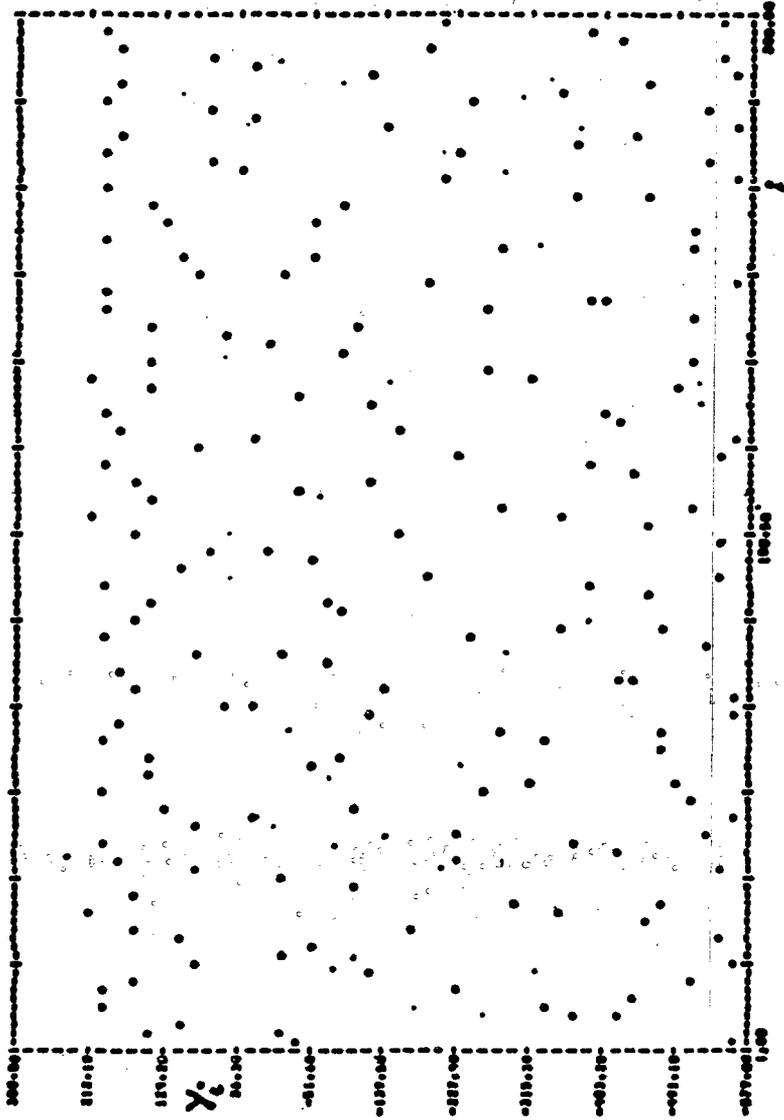
BEAM DEFLECTIONS

Exhibit 10. Beam deflections data plot

BEAM DEFLECTIONS

Exhibit 11. Beam deflections tail length analysis

Exhibit 12. Beam deflections tail length analysis

BEAM DEFLECTIONS

Exhibit 13. Beam deflections tail length analysis

CRITERIA FOR A BIOCELLULAR MODEL - BIOCELLULAR COMMUNICATION

George I. Lavin
Vulnerability Laboratory
Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

We are interested in the design, construction and testing of a model which could be used for the assessment of absorbed energy on biocellular systems. Previous reports to this Conference have had to do essentially with the methodology available for this purpose. However, from time to time an attempt has been made to describe a model.

We do not have to go far afield for the model criteria. Obviously, it is that particular animal or animal system under consideration. The difficulty of course is that we really know very little about the biochemistry of behavior. The analytical data available (on soft tissue) indicate the presence of proteins, nucleic acids, polysaccharides, lipoids, phosphorus containing cerebrosides....Also the general impression is that these materials "act" as sources of energy for the various physiological processes or as enzymes (catalysts).

It is obvious that no attempt is being made at the moment to apply the above to models. The present paper is designed to call attention to another aspect of performance which should be taken into consideration in the design of models. That is, Biocommunication (nerve transmission).

In order to have a dimensional concept on which to speculate, let us consider the example of a man sighting an object and then reaching for it. The following are involved: Brain, Nerve, Muscle, The Visual Systems......

We now have a dimensional ensemble which performs a definite task which we can use for the application of "Biocellular Numbers." That is to say we are in a position where we can make an assessment of the mechanism and utilization of absorbed energy....By a consideration of the effects on the task specific compounds.

To come back to our original subject, that is Biocommunication, we can make the assumption that the performance of a task is the result of the presence and intercommunication of a series of animal organs, the extent of performance being regulated by feedback processes - subject to specific energy interchanges.

It is obvious that the above is not a complete "recipe" for a Biocellular Model. It is intended, however, to be a step along in the hope that by such thinking a working simulation can eventually be attained...Only God can make a tree.

EQUATION-OF-STATE AND SHOCK INITIATION EXPERIMENTS
ON EXPLOSIVES USING PULSED ELECTRON BEAMS

L. Avrami and P. Harris
Picatinny Arsenal, Dover, New Jersey
J. Shea
Physics International Co.

ABSTRACT. Pulsed electron beam energy deposition experiments were carried out on porous granular primary and secondary explosive pellets. Experimental techniques and shock theory were developed to obtain energy-pressure coupling data and sound velocities. The deposition conditions for shock initiation also were investigated and the initiation levels for lead azide and KDNBF were determined.

Experimentally an effective Gruneisen parameter was obtained by rear surface response, and the measurements were correlated theoretically with pulse time and the structure of the porous explosive. The experiments revealed a strong dependence on density ratios.

INTRODUCTION. During a recent program studying the effects of a radiation pulse on explosive materials using pulsed-electron beams (Ref.1) experimental techniques were initiated (1) to obtain energy-pressure coupling data and sound velocities on porous granular explosives, and (2) to determine the energy deposition conditions for shock initiation of explosives.

Although a fair amount of information is available on the static pressure properties of materials the emphasis has shifted to determine the dynamic high pressure properties of materials, not only solid or liquid, but also porous. In the latter case, any material which has a mass density less than the maximum possible equilibrium mass density at a given pressure and temperature is called a porous solid. The example used here is the less than crystal density granular explosives pressed into pellets.

There are three ways to cause a shock wave to occur in a material. An already existing shock in one medium can be propagated into a second material simply by having the two materials be in contact. Solids can be caused to impact each other at high velocity resulting in shock waves propagating out from the point of impact. Lastly, thermal energy can be stored in a material in a spatially inhomogeneous manner; since thermal energy is equivalent to a pressure, and since $\partial P / \partial x$ is equivalent to a force per unit volume the inhomogeneous energy deposition combined with a nonlinear equation of state will result in a propagating shock. This paper describes the results utilizing the last method.

APPROACH. High intensity, high voltage, pulsed electron beams can be used to produce a sudden increase in energy (and pressure) throughout a substantial volume of a solid without shock compression. With this method the problem associated with the local thermal non-equilibrium aspects of shock loading of granular explosives need not be considered. Furthermore, the

Preceding page blank

states attained by sudden volume heating are well removed from the Hugoniot, thus affording the opportunity to evaluate the adequacy of the equation-of-state models under conditions not associated with shock loading.

In order to characterize the state of a material following constant volume heating, the pressure-energy coupling relationship must be known. This relationship is determined through the Gruneisen coefficient, Γ, of the material. This parameter appears in the Mie-Gruneisen equation of state (Ref. 2) and can be written in the form

$$P = f(V) + \frac{\Gamma E}{V} \tag{1}$$

where P is the pressure, V is the specific volume, and E is the specific internal energy.

The Mie-Gruneisen equation of state is generally employed in the finite difference computer codes that are currently used to calculate dynamic response of homogeneous material (Ref. 3). However, when materials with porosity are encountered, one cannot continue to consider them as homogeneous media. On the other hand, the convenience of the finite difference computer codes is such that it is desirable to define constitutive relationships in such a way that they are treated as if they were homogeneous, and average values for pressure, volume and energy are given correctly. Such formulations have been developed by Hermann (Ref. 4) and Seaman and Linde (Ref. 5).

An effective Gruneisen coefficient can be defined for solid materials, including porous substances such as granular explosives, by

$$\Gamma = V \left(\frac{\partial P}{\partial E}\right)_V \tag{2}$$

where the volume is considered in a macroscopic sense, i.e. bulk material plus any voids. The coefficient thus defined is expected to be energy dependent for porous materials, since internal pressure relief (still at constant volume) can occur due to the collapse of voids, especially as the shear strength of the material generally decreases with increasing temperature.

If the pressure in a substance is determined as a function of energy under constant volume conditions, value of the Gruneisen coefficient can be obtained as a function of energy by differentiation (Equation 2).

High-intensity pulsed electron beams are well suited for producing sudden volume heating, so that the Gruneisen coefficient can be evaluated. When the irradiated sample (slab geometry) is thick enough to absorb all of the incident electrons, a compressive stress wave propagates from the heated, high-pressure region toward the rear (away from the source) of the sample.

A transducer, such as a piezoelectric material or manganin gauge, located just beyond the deposition depth, can be used to record the stress pulse. Alternatively, the displacement or velocity of the rear surface may be observed using optical techniques. A relatively straightforward procedure involving material response calculations can be used to infer the initial pressure distribution from the experimentally measured stress history, provided the constitutive relation is reasonably well known. A knowledge of the energy deposition profile is then sufficient to determine the Gruneisen coefficient.

By locating the transducer just beyond the deposition region, the effects of uncertainties in the Hugoniot, dispersion, or attenuation are minimized. Moreover, since the entire stress profile is obtained, serious errors in the constitutive relation, if present, will be detected.

EXPERIMENTAL RESULTS. Experiments to determine the pressure-energy coupling were performed on PETN (pentaerythritol tetranitrate), KDNBF (Potassium dinitrobenzcfuroxan), lead azide (PbN_6), and RDX. The PETN wafers were pressed from a very fine powder (grain diameter .003mm) to densities ranging from 1.48 to 1.57 grams/cm^3 (Ref. 6). KDNBF pellets of 5/8 inch diameter pressed with 4 tons (ρ=1.67gm/cc) and 5 tons (ρ=1.60gm/cc) and PbN_6 pellets pressed with 3 tons (ρ=2.92gm/cc) and 5 tons (ρ=3.27gm/cc) were tested. The nominal thickness of the specimens was 0.070 inch. Electron beam irradiations of the specimens were performed with both the Model 312 Pulserad and on the Model 738 Pulserad. Stresses generated in the specimens were measured with piezoelectric (X-cut) quartz stress gauges. The gauges have a nominal read time of 0.5 μsec before internal reflections perturb the output. The active area of the gauges is 1/2 inch in diameter.

In the initial series of tests that were performed with the Physics International Model 312 Pulserad, specimens were bonded directly onto the quartz gauge. In order to deep the duration of the stress pulse to less than 0.5 μsec., the electron range, R, in the specimen was limited to a value less than c x 0.5 μsec, where c is the specimen sound speed. For preliminary calculations, c was assumed to be 0.2 cm/μsec; thus R < 0.1 cm. This criterion dictated a maximum mean energy of 0.5 MeV for the least dense of the materials under test. The electron beam was controlled by a low pressure background gas. Fluences were monitored between data shots by inserting a graphite calorimeter into the test chamber at the specimen location. Normalized dose profiles were obtained at intervals with an aluminum depth dosimeter. A typical normalized dose profile in aluminum is given in Figure 1.

A quartz transducer record obtained for the pressed PETN is shown in Figure 2. For KDNBF and lead azide the records are shown in Figures 3 and 4. The first signal is a short pulse due to the electron bremsstrahlung radiation incident upon the cables for the transducer. This provides a precise fiducial point for the energy deposition. The compressive stress pulse arrives shortly thereafter and is recorded well within the 0.5 μsec reading time of the transducer. With this record a measurement of the scund speed in the material can be obtained when the sample thickness and energy deposition profile is known.

The record stress levels were comparatively low: further, a consequence of the Seaman and Linde model for porous solids (Ref. 5) is that the propagated stress pulse resulting from energy deposition in a porous solid will not exceed half the crush strength, provided the voids are not filled and no vaporization takes place. Therefore, linear equations were used to compute the material response.

The results for all the PETN samples are summarized in Figure 5 where the peak initial pressure deduced from the measured stress is shown as a function of peak dose. Linear relationships are found for samples with densities of 1.59 ± 0.01 gram/cm^3 and 1.54 ± 0.01 gram/cm^3. The Gruneisen coefficients for these materials are computed to be 0.51 and 0.15, respectively. Greater uncertainties are associated with the data of the low density samples, 1.47 gram/cm^3, and 1.54 gram/cm^3, as a result of the lower densities and sound velocities producing stress pulses that slightly exceed the nominal read time of the quartz gauges. Therefore an extrapolation of the records was necessary.

The pressure that was observed from the 1.47 gram/cm^3 dense sample exposed to a dose of 16 cal/gram falls within an expected region. The other 1.47 gram/cm^3 dense sample, however, produced a pressure that is higher than anticipated. It falls slightly above the data for the 1.54 gram/cm^3 dense samples. This discrepancy was not resolved.

The datum point for material with a density of 1.67 gram/cm^3 indicates a value for the Gruneisen coefficient of 1.2 ± 0.2 which is considerably higher than that of the lower density materials. Since the Gruneisen coefficient is expected to be sensitive to density, this result was anticipated.

The sound speeds obtained from the oscilloscope traces for the 1.59 gram/cm^3 and 1.54 gram/cm^3 samples were found to be 2.4 ± 0.1 mm/μsec and 1.8 ± 0.3 mm/μsec, respectively. These agree with data reported by Roth and Blackburn (Ref. 12) for PETN in the same density ranges. The measured sound speeds for the 1.47 gram/cm^3 and the 1.67 gram/cm^3 samples were 1.7 ± 0.3 mm/μsec and 2.8 mm/μsec, respectively. The complete results of the measurements are summarized in Table 1.

The measured stress profile is compared with a computed stress profile based upon the energy deposition profile, a sound speed of 2.4 mm/μsec and an assumed value for the Gruneisen coefficient of 0.51 (Figure 2) and shows satisfactory agreement.

For the lead azide and KDNBF, similar work is now in progress. Besides using the technique describes above a second series of tests are being performed employing a somewhat different quartz gauge technique. In this technique a buffer of fused quartz is sandwiched between the sample and the quartz gauge. With this configuration, the mean electron energy is chosen high enough so that deposition occurs through the entire specimen and into the buffer. Ideally, the specimen-buffer interface should lie at about half-range (or less). Since fused quartz has a very low Gruneisen coefficient ($\Gamma \leq 0.035$, Ref 7) compared to most materials, it produces very little stress. The stress pulse that is obtained from the combined package thus will usually exhibit a large step corresponding to arrival of stress from the rear surface of the specimen. In any case, the equation of state for fused silica is well known (References 3 and 4) and the arrival time for the pressure signal from the test material can be uniquely determined.

This technique has several advantages. First, gauge read times are of little importance since the step, which is the signal of interest, is generally the first detectable signal to be read by the gauge. Second, one-dimensional read times and stress reverberation time (i.e. multiple reflections between the sample front surface and the buffer) do not affect the amplitude of the step. And third, the normalized dose at about half-range is least affected by slight variations in the electron energy spectrum.

Preliminary results obtained using both techniques are listed in Table 2. For KDNBF the results indicate so far that for the two types of samples the sound speeds and effective Gruneisen parameters are the same at the two densities. Additional tests are being conducted to determine if a density ratio actually does exist. For lead axide the sound speed value of 0.35 ± 02 cm/μsec and a probably Gruneisen constant of 0.25 ± 03 were obtained for the 5 ton pellets with a density of 3.29 grams/cm^3.

DISCUSSION OF GRUNEISEN RESULTS. There are two possible definitions of the Gruneisen parameter, one microscopic and the other macroscopic. The microscopic involves a model for interactions on an atomic level, while the macroscopic definition involves measurable parameters.

In the macroscopic definition, the Gruneisen constant, Γ, is defined by:

$$\Gamma = \frac{1}{\rho} \left(\frac{\partial p}{\partial E}\right)_\rho , \qquad\qquad (2)$$

where p is pressure, ρ is mass density, and E is internal energy. Equation 2 says that if one deposits thermal energy in a time small compared to the time necessary (defined below) for significant mass motion, then the pressure in the solid changes. The quantity Γ determines the change in pressure, and is a property of the structure of the material.

The time criterion mentioned above is usually taken as the time necessary for an acoustic wave to traverse an electron scattering mean free path in the solid. Let such a time be t_1. If the pulse duration is T, then the constant volume definition of Eq. 2 requires:

$$T < t_1. \qquad\qquad (3)$$

Physically, this means if Eq. 3 holds, then any energy inhomogenities introduced by the electron beam are not relieved by acoustic signals before the energy is deposited. In other words, the energy can be considered as deposited instantaneously if Eq. 3 holds.

Considering that the average dimensions of the granular explosive particles are small compared to the appropriate radiation mean free path, it can be readily seen that the time criteria of Eq. 3 is, in general, not valid for a porous material. In the case of small particles, the energy deposition- induced thermal stresses are relieved by thermal expansion prior to completion of the energy deposition process. Table 3 below lists some electron energy absorption mean free paths for a variety of materials and energies (Ref. 8).

As contrast to the typical dimension of $10^3\mu$, i.e., 0.1 cm, of Table I, an acoustic wave propagating at 3×10^5 cm/sec can relieve the stress to the center of a 180μ particle in 30 nanoseconds. Consequently, an electron beam machine having a pulse width of 30 nanoseconds (FWHM) would not be able to utilize the criteria of Equation 3 for experiments with porous solids having particle sizes of 180μ or less. It is interesting to note that the 30 nanosecond figure is typical of the faster machines in use today. The average grain size of the PETN was observed to be 250μ. The electron beam was at 0.4 Mev with a pulse width of 25 nanoseconds (FWHM). Thus, if the PETN in question was porous enough so that a significant number of grains were not in intimate contact along a grain boundary, then the time criteria of Equation 3 would be invalid.

Consequently, the Γ measured was not the true constant volume Gruneisen parameter as defined by Equation 2. Exactly what detailed microscopic physics corresponds to the measured Γ is a very difficult question in porous solid mechanics which remains unanswered. While the measured "effective Γ" is a worthwhile experimental variable in that it characterizes the gross behavior of the material, unfortunately it does not allow for distinguishing between all possible porous materials as a function of material parameters.

The same conclusions were evident in flyer plate experiments (Ref. 9) where Harris analyzed the PETN results and the flyer plate experiments of Boade (Refs 8, 9). Wave propagation in porous materials was considered in detail and Harris concluded that in such materials macroscopic measurements can not always be used to determine microscopic parameters. In support of this, the porous material is considered to act as a diffraction grating with respect to a shock wave, and that unidirectional one-dimensional strain does not hold. Also the role of sur-

face energy effects was considered since the measured pressures were in the fraction of a kilobar range and the loading pressures for the pellets were in the one to ten kilobar range.

Experiments are scheduled to conduct the same type of experiments with large single crystals of RDX and PETN. Mazzella et al (Ref. 11) have proposed a Gruneisen parameter for a porous material given by

$$\Gamma = \frac{c^2}{c_s^2} \Gamma_s \qquad (4)$$

where Γ_s is the parameter for the voidless undeformed solid, with c_s being the propagation velocity corresponding to a pure bulk modulus wave in the voidless undeformed material. c is the corresponding velocity for the porous solid.

Whether a correlation can be made is not known, but it seems that Equation 4 is valid for those experimental situations where the energy deposition is so large in time that the porous medium is able to continuously adjust its density on a microscopic scale without affecting the macroscopic density. Other factors such as relaxation time will have to be considered.

Conparisons are also being made with the relationship

$$\Gamma = \frac{\beta k}{\rho c_p} \qquad (5)$$

where β is the volume expansion coefficient, k is the adiabatic bulk modulus and c_p is the specific heat at constant pressure. Values are being determined for single crystals and pressed pellets and the Gruneisen parameter compared to those obtained with the energy deposition technique.

SHOCK INITIATION RESULTS. Shock initiation experiments were performed on lead azide and KDNBF. Also some preliminary results were obtained on lead styphnate and LMNR (lead mononitroresorcinate).

The lead azide wafers were 1/4 inch diameter and 0.040 inch thick with a density of 4.0 grams/cm^3. These wafers were bonded with Silastic 732 RTV adhesive to aluminum overlays that were 3/4 inch diameter and 1/8 inch thick. The lead azide samples were irradiated with a pulsed 900 kev electron beam; the energy deposition was entirely in the explosive.

The stress pulse in the aluminum could be computed from the energy deposition profile, the beam intensity, and the equation of state for aluminum. The stress transmitted to the explosive depends upon the relative acoustic impedances of aluminum and the explosive. In addition, the bond layer must be quite thin to avoid significant attenuation.

A test was conducted to check the experimental configuration, and, in particular, the effect of the adhesive bond on the transmitted stress pulse. In this test a quartz gauge was bonded to an aluminum sample using the Silastic 732 RTV adhesive, the sample was irradiated, and a comparison was made between the predicted and measured stress hisotires. There appeared to be a 20 percent attenuation of the stress pulse due to the bond. This factor was used in the subsequent analysis to estimate the stress transmitted to the explosives.

The experimental results obtained for lead azide are summarized in Table 4 and in Figure 6. Seventeen samples were irradiated; six samples initiated and

eleven samples did not. The temperatures indicated in column 4 of the table are the final temperatures that the aluminum overlay achieved due to the energy deposition. These values are upper bounds; any thermal losses were neglected. The highest temperature achieved at the aluminum-lead azide inter- face was approximately 120 C. This value is significantly below the lowest value of 297 C for thermal initiation threshold of the lead azide. The data indicate a stress initiation threshold of 3.6 kbar for the lead azide, assuming a sound velocity of 0.25 cm/μsec for the explosive. The stress pulse width was approximately 0.2 μsec. However in the equation-of-state work on lead azide the sound speed for lead azide with a density of 3.29 grams/cm^3 was .35 cm/μsec. If this value is used the initiation threshold is 4.4 kbar.

These results are in reasonable agreement with some small-scale gap tests which obtained threshold initiation levels of 4 to 6 kbar for lead azide (Ref- erence 13); they are also consistent with an experiment performed by Roth (Ref- erence 14), who observed an initiation in 95 percent lead azide, 5 percent Teflon (ρ=2.7 gram/cm^3) subjected to an 8.5 to 10 kbar shock loading.

Only a limited amount of data was obtained on lead styphnate and LMNR. One initiation was achieved in lead styphnate. Assuming a sound speed of 0.25 cm/μsec and a density of 2.75 gram/cc the stress transmitted to the lead styphnate that produced initiation is estimated at 4.3 kbar, while the highest stress that did not produce initiation was 3.1 kbar. Pulse duration was in the order of 0.4 μsec. With LMNR with the same type of pulse no initiations were achieved with the highest stress transmitted to the LMNR being in the order of 8 kbar.

The shock initiation experiments were performed on KDNBF pellets of 1/4 inch diameter and 0.044 ± 0.003 inch thickness. The density of the pellets was 1.75 ± 0.03 grams cm^3. Thermomechanical stress pulses were produced by rapid electron energy deposition in aluminum disks, 3/8 inch in diameter and 1/8 inch think. The stress pulses propagated into KDNBF pellets, which were bonded to the rear surface of the aluminum disks with phenol salicylate (salol).

Due to the different acoustical impedances of aluminum and KDNBF, only part of the shock wave was transmitted into the KDNBF and the remainder was reflected. The transmission coefficient, T, can be calculated from the formula

$$T = \frac{2}{1 + Z_1/Z_2} \qquad (6)$$

where Z_1 and Z_2 are the acoustic impedances of aluminum and KDNBF, respectively. For aluminum,

$$Z_1 = \rho u_s = 2.71 \text{ grams/cm}^3 \times 0.57 \text{ cm/μsec}$$

$$= 1.54 \text{ grams/cm}^2 \text{ μsec}$$

where u_s is an average value of the shock velocity in aluminum for the range of pressures of interest (about 50 kbars). For KDNBF, using the experimentally determined sound speed of 0.30 cm/μsec,

$$Z_2 = \rho c = 1.74 \text{ grams/cm}^3 \times 0.30 \text{ cm/μsec}$$

$$= 0.525 \text{ gram/cm}^2 \text{ μsec}$$

Thus T = 0.51.

The electron beam environment for this experiment was provided by the Model 738 Pulserad. Machine parameters were set to produce an electron beam of approximately 1 MeV mean energy and approximately 50 nanosecond duration. Such a beam produces a peak normalized dose of about 4.5 (cal/gram)/(cal/cm^2) and a range in aluminum of about 0.085 inch. Since the aluminum overlays were 0.125 inch thick, no electrons were deposited into the KDNBF. The electron beam was controlled by gas focussing in a low pressure background gas; a metal beam guide consisting of a copper pipe was used for the fluences below 50 cal/cm^2 and a copper cone for the higher fluences. Normalized deposition profiles in aluminum were obtained with an aluminum depth-dosimeter. Fluences for each data shot were monitored by a set of graphite calorimeters surrounding the speci-men in the case of the copper pipe, and by a Rogowski coil net current monitor in the case of the copper cone.

Peak stresses generated in aluminum for each data shot were calculated using the formula

$$\sigma_p = 1/2 \ \rho \Gamma \epsilon \ (1 - \frac{c\tau}{2R}) \ (\text{kbars}) \tag{7}$$

where ρ is the sample density, Γ is the Gruniesen coefficient for aluminum, ϵ is the peak dose (0.04186 E if E is in cal/gram), $c\tau$ is the relief depth, i.e. the product of the sound speed, c, and the energy deposition time, τ, and R is the range in aluminum. For this experiment,

$$\rho = 2.71 \ \text{grams/cm}^3$$
$$\Gamma = 2.0$$
$$E = 4.5 \ \phi\text{cal/gram}$$
$$c = 0.57 \ \text{cm/}\mu\text{sec}$$
$$\tau = 0.05 \ \mu\text{sec}$$
$$R = 0.085 \ \text{inch}$$

hence,

$$\sigma_p = 0.477 \ \phi(\text{kbars})$$

where ϕ is the fluence in cal/cm^2.

Table 5 contains a summary of the results obtained in the experiments. The third column contains the peak stress in aluminum calculated for each shot using Equation 7. In the fourth column, the peak stress in KDNBF has been calculated from the stress in aluminum utilizing the transmission coefficient [Equation 6]. The last column indicates the observed effect on the explosive. The results are presented graphically in Figure 7. The data indicate that shock initiation may occur at pressures of 20 to 30 kbars. This conclusion should be regarded as ten-tative since one specimen clearly did not initiate in this pressure range and another specimen apparently exhibited only a low-order initiation or burning at a pressure of approximately 50 kbars. It is conceivable that at higher fluences, initiation could have resulted from late-time heating of the specimen by thermal conduction through the aluminum overlay. However, front surface spall probably removed much of the deposited energy. Late-time heating effects in future experi-ments could be detected by instrumenting the overlay with a thermocouple. If pre-sent the heating could be eliminated by including a second disk of aluminum in the sample package with a "strengthless" bond to the first disk of aluminum.

SUMMARY AND RECOMMENDATIONS. In summary using pulsed electron beams we have experimentally obtained an "effective" Gruneisen coefficient for PETN which reveals a strong dependence on density ratios. Preliminary data has been obtained also on KDNBF and lead azide.

The elastic sound velocities for each of these materials have been found concurrently.

Further work has to be done to explain theoretically the results obtained since the time dependence criteria is not being met; only the gross behavior of the explosive is characterized.

In the shock initiation work a technique has been developed to determine the shock initiation of explosives, especially in the 0-20 kilobar range. Results for lead azide are consistent; in the range above 20 kilobars modifications have to be considered for temperature effects and spallation. The results for KDNBF seem to bear this out.

REFERENCES

1. Shea, J.H., Mazzella, A., Avrami, L., Jackson, H.J., "Effects of a Radiation Pulse Upon Primary Explosives (U)" Picatinny Arsenal Tech. Report 4230, May 1971 (SRD).

2. Rice, M.H., McQueen, R.G., and Walsh, J.M., Solid State Physics, 6 (1958)

3. Bakken, L.H., and Anderson, P.D., "An Equation of State Handbook", Rept. SCL-DR-68-123 (1969), Sandia Corporation, Livermore, California.

4. Hermann, W., "Constitutive Equation for the Dynamic Compaction of Ductile Porous Materials", J. Appl. Phys. 40, 2490 (1969).

5. Seaman, L., and Linde, R.K., "Distended Material Model Development", AFWL-TR-68-143, Vol. 1, Air Force Weapons Laboratory, Albuquerque, New Mexico, (May 1969).

6. Shea, J.H., Mazzella, A., and Avrami, L., "Equation-of-State Investigation of Granular Explosives Using a Pulsed Electron Beam", Fifth Symposium (International) on Detonation, Paper No. 5071-P, Pasadena, Calif. August 1970.

7. Data to be published by Physics International Company.

8. Harris, P., "The Gruneisen Constant of Porous Materials in Energy Deposition Experiments", Picatinny Arsenal Tech. Report 4255, August 1971.

9. Boade, R.R., "Experimental Shock Loading Properties of Porous Materials and Analytic Method to Describe These Properties", SC-DC-70-5052, Sandia Laboratories 1970.

REFERENCES (cont'd)

10. Harris, P., "A Survey of the Physics of Shock Waves in Solids",
    Picatinny Arsenal Tech. Report 4345, February 1972.

11. Mazzella, A., Shea, J., and Stefansky, T., "Dynamic Response of Porous
    Materials to Electron Beam Deposition", Physics International Co.
    Report PIFR-216, Oct. 1971.

12. Roth, J., and Blackburn, J., "Effect of Initial Temperature on the Shock
    Sensitivity of Granular Explosives", Rept. SC-CR-67-2805, Sandia Cor-
    poration, Albuquerque, New Mexico, August 1967.

13. Standardization of the Small Gap Test Used to Measure the Sensitivity
    of Explosives, NAVWEPS 7342, Jan. 1961.

14. Personal communication - J. Roth, Stanford Research Institute.

NORMALIZED DOSE PROFILE IN ALUMINUM PULSE #1901 ON 312 PULSERAD
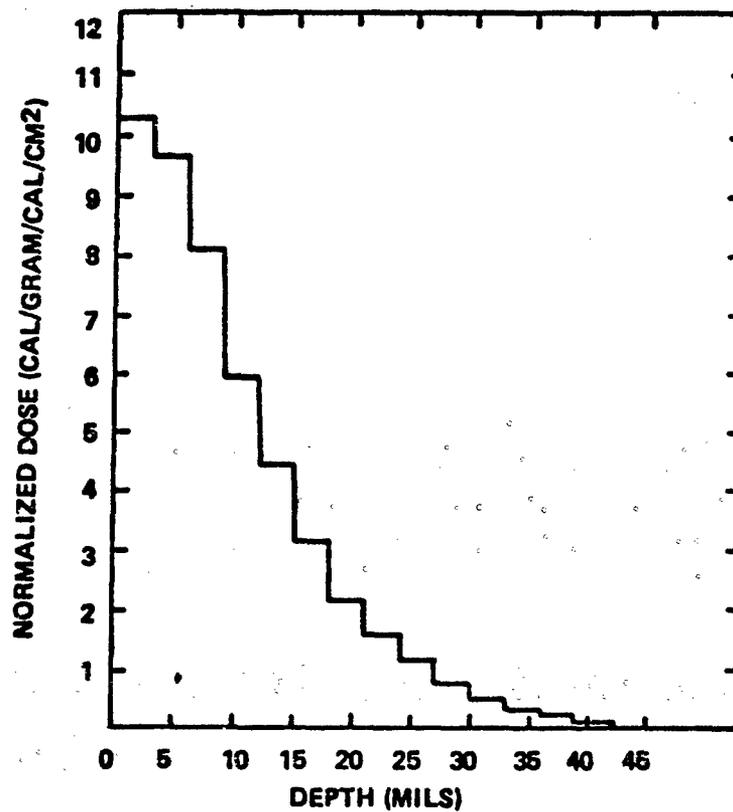
Figure 1

QUARTZ GAUGE RECORD
OBTAINED FOR A PETN SAMPLE
WITH A DENSITY OF 1.60 G/CM^3,
SHOT 683

UPPER TRACE
0.2 V/MAJOR
DIVISION

LOWER TRACE
0.5 V/MAJOR
DIVISION

SWEEP SPEED
0.2 μsec/MAJOR
DIVISION
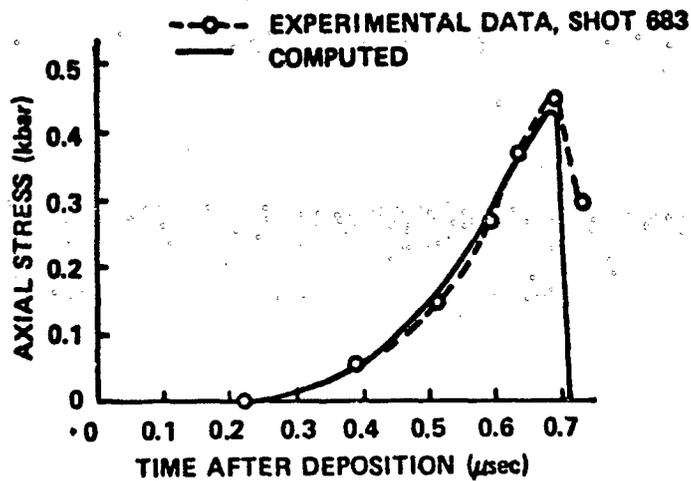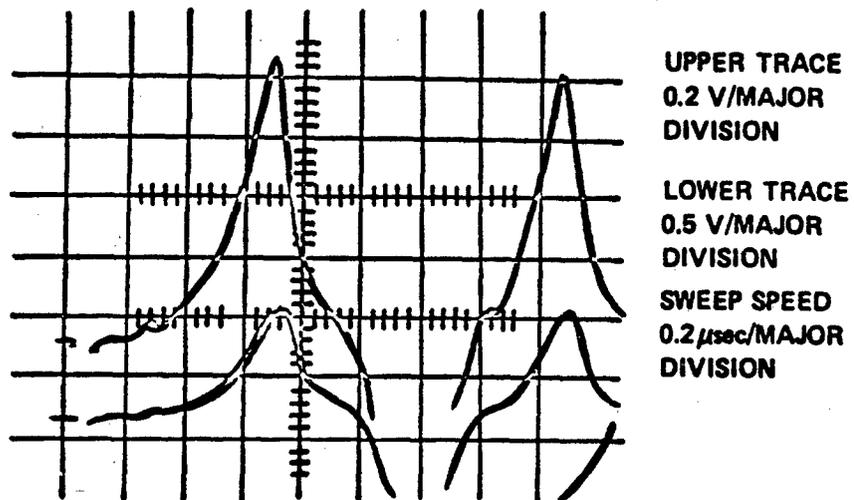
COMPUTED AND MEASURED STRESS RECORDS
FOR PETN SAMPLE

Figure 2

QUARTZ GUAGE RECORD
SHOT 1910 5 TON KDNBF
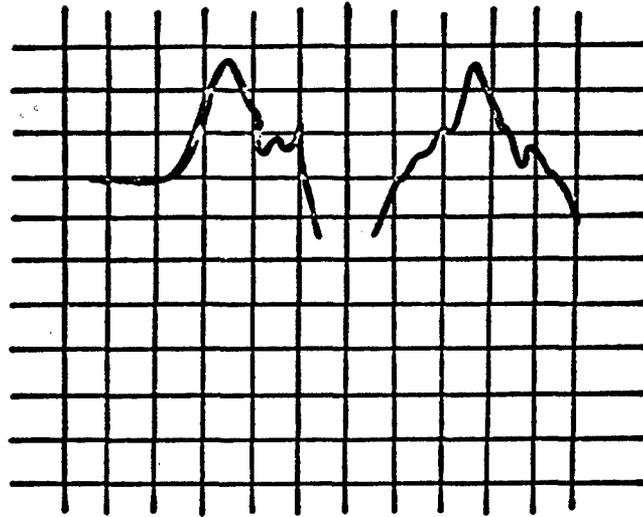
0.1 V/cm, 0.2 μsec/cm

Figure 3A

QUARTZ GAUGE
RECORD SHOT
1913 5 TON KDNBF
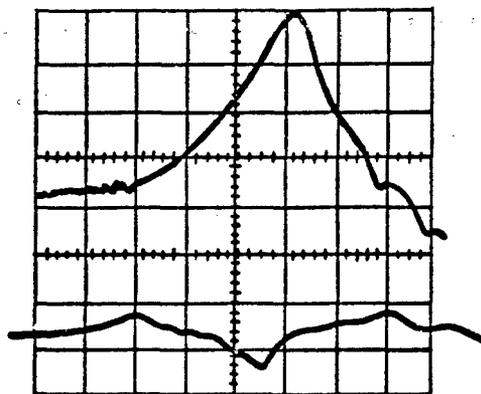
0.1 V/cm, 0.1 μsec/cm
1 V/cm, 0.2 μsec/cm

Figure 3B

QUARTZ GAUGE RECORD SHOT
21995 5 TON PBN$_6$
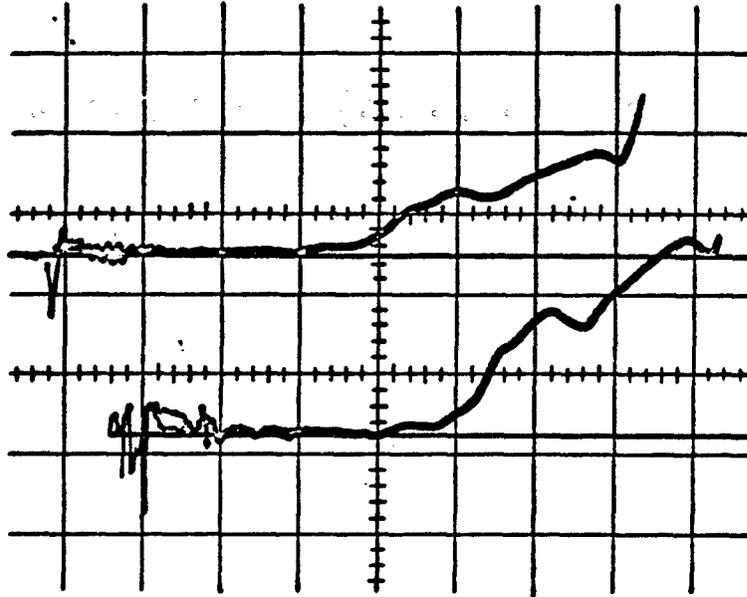
2 V/cm, 0.2 μsec/cm
1 V/cm, 0.2 μsec/cm

QUARTZ GAUGE
RECORD SHOT
22004 5 TON PbN$_6$

1 V/cm, 0.2 μsec/cm
0.5 V/cm, 0.2 μsec/cm

Figure 4

-466-
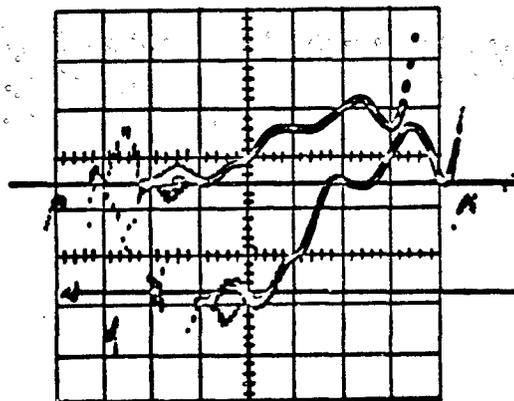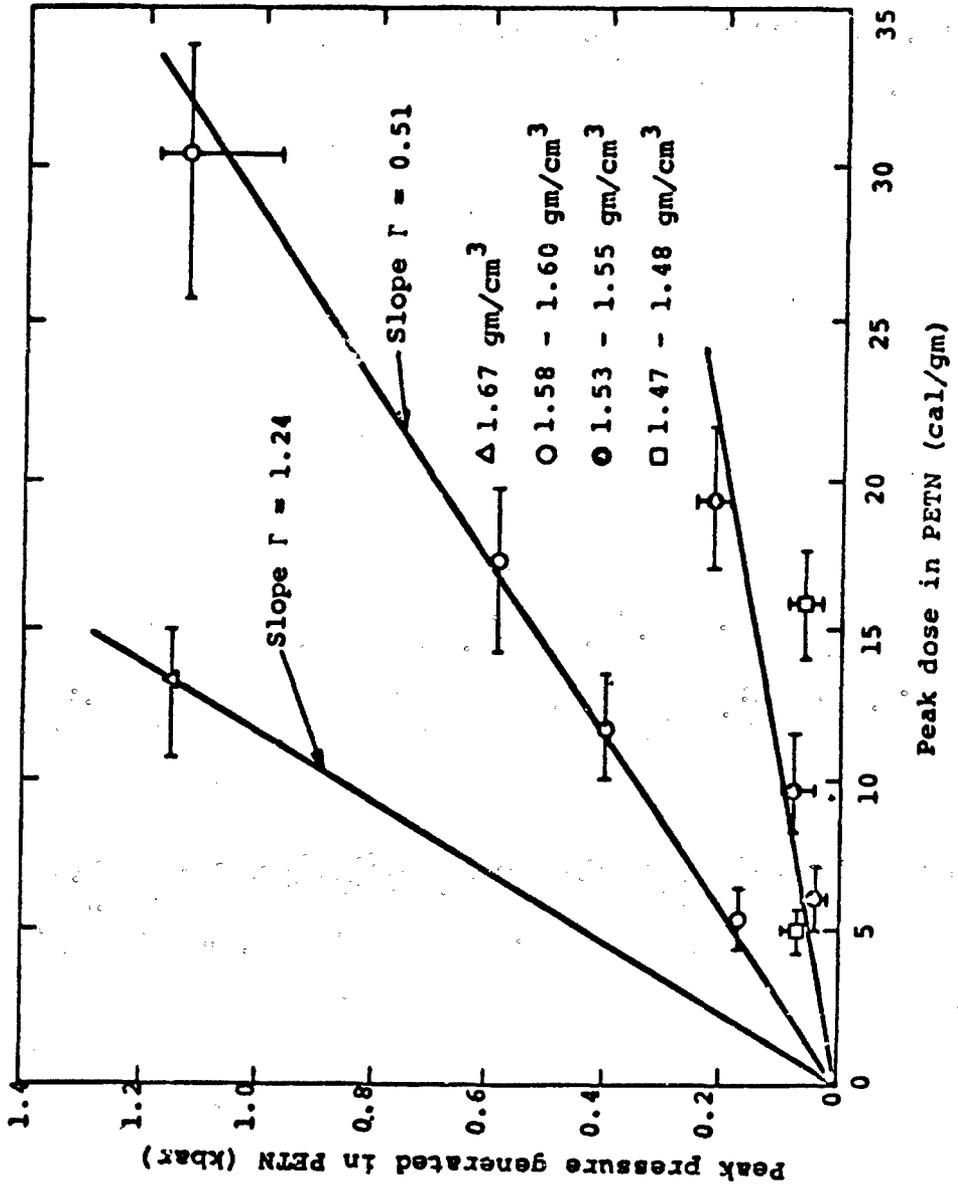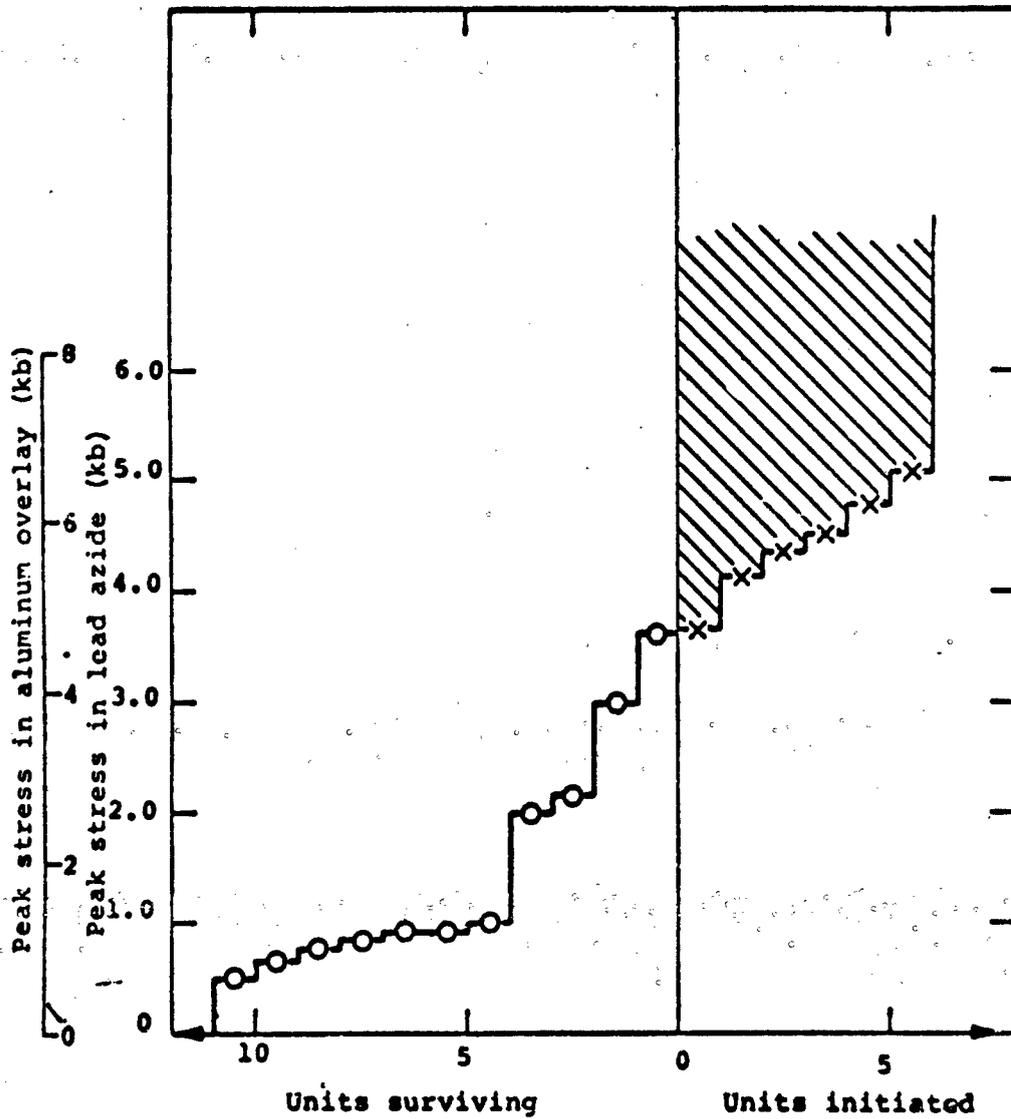
Figure 5

Initiation of lead azide pellets by a 0.2-μsec thermomechanical stress pulse in an aluminum overlay

Figure 6

-468-

INITIATION OF KDNBF PELLETS BY A 0.37 μSEC THERMOMECHANICAL STRESS PULSE IN ALUMINUM OVERLAY
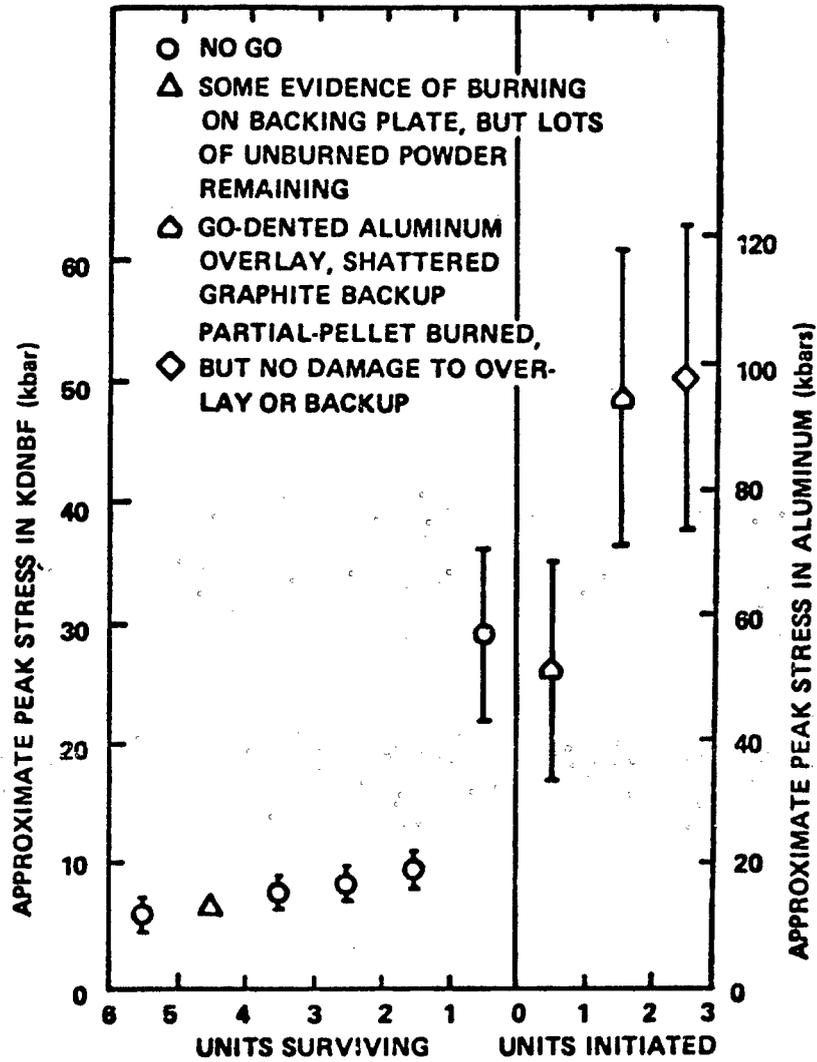
Figure 7

TABLE 1. Summary of results on pressure energy coupling experiments for PETN (U)

Density (gm/cm^3)	Relative Density, ρ/ρ_0 (percent)	Sound Speed (mm/µsec)	Γ
1.67	94.5	2.8 ± 0.1	1.2
1.59	90.0	2.4 ± 0.1	0.51
1.54	87.0	1.8 ± 0.3	0.15
1.48	83.5	1.7 ± 0.3	(0.07 to 0.23)

TABLE 2

SUMMARY OF PRESSURE-ENERGY COUPLING EXPERIMENTAL RESULTS

(Preliminary)

Explosive	Density (gram/cm^3)	Sound Speed (cm/μsec)	Probable Gruneisen
5-ton KDNBF	1.67	0.30 ± 0.02	0.4 ± 0.15
4-ton KDNBF	1.60	0.30 ± 0.02	0.4 ± 0.15
5-ton PbN$_6$	3.29	0.35 ± 0.02	0.25 ± 0.03
3-ton PbN$_6$	2.92	Undetermined	≤ 0.06 (?)

TABLE 3

ELECTRON MEAN FREE PATH AS A FUNCTION OF ENERGY

Material	Mean Free Path	Electron Energy
Al	100 μ	0.4 Mev
PETN	500 μ	0.4 Mev
Be	1,000 μ	0.6 Mev
Quartz Crystal	4,000 μ	2 Mev
PETN	1,000 μ	4 Mev
Lead Styphnate	1,000 μ	4 Mev

TABLE 4

SHOCK INITIATION OF LEAD AZIDE

Shot No.	Average Fluence (cal/cm²)	Peak Dose (cal/gram)	Final Temp. of Aluminum Overlay (C)	Peak Propagating Stress in Aluminum (kbar)	Peak Stress in Lead Azide (kbar)	Response[a]
17522	17	65	105	6.2	4.8	I
17523	13	50	88	4.7	3.6	NI
17524	16	60	103	5.7	4.4	I
17525	15	57	98	5.4	4.2	I
17526	19	70	118	6.6	5.1	I
17527	17	61	108	5.8	4.5	I
17530[b]	10	36	73	3.5c	4.3d	
17531	11	41	78	3.9	3.0	NI
17532	14	50	93	4.7	3.6	I
13785	3.2	8.8	38	.84	.65	NI
	5.5	15	51	1.4	1.1	NI
	6.2	17	55	1.6	1.2	NI
13786	6	17	53	1.6	1.2	NI
	13	36	87	3.4	2.6	NI
	14	38	94	3.6	2.8	NI
13787	4.3	12	45	1.1	.85	NI
	5.1	14	49	1.3	1.0	NI
	6.4	18	55	1.7	1.3	NI

a I = Initiated; NI = Non-initiated

b Calibration shot - aluminum/bond/quartz transducer

c Measured stress in quartz

d Computed stress in quartz

TABLE 5

INITIATION OF KDNBF PELLETS
BY A 0.37 μSEC THERMOMECHANICAL
STRESS PULSE IN ALUMINUM OVERLAY

SHOT NUMBER	FLUENCE (cal/cm^2)	STRESS IN AL (kbar)	STRESS IN KDNBF (kbar)	GO/NO-GO
20410	25.5(10%)*	12.1(15%)*	6.2(15%)*	NO GO
20411	33.2(10%)	15.8(15%)	8.1(15%)	NO GO
20412	40.4(10%)	19.3(15%)	9.8(15%)	NO GO
20414	36.2(10%)	17.3(15%)	8.8(15%)	NO GO
20415	28.0(10%)	13.4(15%)	6.8(15%)	PARTIAL?
20420	205(20%)	98(25%)	50(25%)	PARTIAL
20421	198(20%)	94(25%)	48(25%)	GO
20423	107(30%)	51(35%)	26(35%)	GO
20424	119(25%)	57(25%)	29(25%)	NO GO

* ESTIMATED UNCERTAINTY

AN ANALYSIS OF 5.56mm ALUMINUM CARTRIDGE CASE "BURN-THROUGH" PHENOMENON

Walter H. Squire and Reed E. Donnard
Frankford Arsenal, Philadelphia, Pennsylvania

ABSTRACT. The use of aluminum alloys as small arms cartridge case material is precluded by a catastrophic failure process previously identified as "burn-through". To determine engineering solutions which would enable the use of aluminum alloys, instead of the conventional cartridge brass, a combination experimental and theoretical program was designed and conducted to identify this failure process.

Postmortem analyses of cartridge cases which evidenced "burn-through" lead to the premise that this phenomenon occurs only when there is a path for the high energy propellant gases to exit any unsealed opening in the cartridge case. Since the natural occurrence of this process is reduced significantly by advanced processing and super strength alloys, it was necessary to provide gas paths to study the failure dynamics. The experimental program was designed to allow a parametric investigation relating size and location of the induced orifice and propellant gas pressure and flow time to the amount of damage inflicted to the cartridge case.

An analytical model was constructed from existing solutions of the physical processes occurring during "burn-through". The gas dynamics of flow in a small bore, boundary layer effects, transient heat conduction, and change of phase processes were all considered. Utilization of the analytical model allowed determination of the local temperature profiles in selected regions of an aluminum cartridge case and permitted the generalization that the failure process is ablation. Given gas flow, it was easily shown that portions of the aluminum cartridge case undergo melting. However, further use of the model demonstrated that the total amount of aluminum ablated from the cartridge case was not correctly predicted by a classical melting theory.

High speed motion pictures taken at the surface of the cartridge case showed the existence of a localized exothermic chemical reaction. When this additional energy source is considered together with the classically derived heat flux, the agreement between theory and experiment is excellent.

INTRODUCTION. The United States Army - SASA, MUCOM, and Frankford Arsenal - is engaged in an exploratory development program to determine the engineering parameters required to utilize aluminum cartridge cases in high-pressure, small-caliber ammunition systems. This work is based on the need to create lightweight ammunition/weapon systems in near future applications. Conservation of copper resources is an added benefit to be gained from success in this effort.

To date, a significant fact which has precluded the acceptance of aluminum cartridge cases in the logistic system is the nature of the failure process - heretofore identified as "burn-through". When a mechanical

case failure is encountered during the firing of aluminum-cased ammunition, particularly with high-pressure, high-performance weapon systems, the failure of an aluminum cartridge case is characterized by a large efflux of very luminous gases at the breech of the weapon, the serious erosion of the cartridge case, and often, the inability of the weapon to function properly thereafter. Thus, this failure may result in serious harm to the rifleman and damage the weapon.

In order to investigate the failure dynamics, as a basis for finding a solution or solutions to this problem, it was found that a small hole drilled in the head region of an aluminum cartridge case, or a four thousands deep longitudinal scratch would, upon firing, result in "burn-through". As a result of this simulation of the failure process, it was concluded that a gas path must be available for the otherwise unrestricted flow of propellant gases from the interior of the cartridge case through the path in the case wall and to the atmosphere. Figure 1 is a photograph of the results of firing brass and aluminum cartridge cases with a 0.0135 inch (diameter) hole in the head region of each case. The two brass cases seem unaffected after firing. However, the two aluminum cartridge cases show the typical erosion in the head region; the unfired aluminum cartridge case can be used to compare the damage after firing with its initial, drilled condition. This paper is addressed to a discussion concerning the peculiar results when propellant gases are allowed to pass through an induced fissure in an aluminum cartridge case.

STATEMENT OF THE PROBLEM. Using a pre-placed gas path in an aluminum cartridge case as a vehicle to study this phenomenon, the mathematics which follow describe the heat flux resulting from the fast moving propellant gases through the induced orifice and allow determination of the temperature profiles in selected regions of the cartridge case. Emphasis is placed on the gas/solid interface – the interior surface of the induced orifice. To test the accuracy of the mathematical model, the appropriate physical constants of the case materials under consideration, along with values obtained from an experimental program, were substituted into the equations. Temperature profiles and total erosion values were thus calculated. This analytical tool was used to investigate the premise that portions of the aluminum cartridge case undergo melting and that the total amount of aluminum ablated from the cartridge case is correctly predicted by a classical melting theory. The same mathematics, when applied to the flow through a brass cartridge case, should indicate minimal metal removal.

In order to facilitate a closed-form solution of the mathematical model, thereby enabling a parametric investigation of the critical factors affecting "burn-through" and verification with experimental data, the following assumptions were employed:
(1) The dimensions of the induced orifice remain constant until melting,
(2) Propellant gas behaves as an ideal gas,
(3) Classical heat conduction equation describes the dissipation of energy in the solid until melting,

(4) The physical properties of the cartridge case material are constant during the experiment,

(5) It is possible to identify an average operating pressure ($P_O = 25,000$ psi), and

(6) The gas temperature is the adiabatic flame temperature ($T_O = 5040°R$).

EXPERIMENTAL. To obtain pressure vs time data, a thick-walled test barrel was drilled and threated to accept a 607A Kistler Gauge. This gauge was positioned midway along the longitudinal axis of the cartridge case to measure the gas pressure. Figure 2 shows the pressure-time history of the propellant gases inside a 5.56mm cartridge case which contained the charge of 27.0 grains of WC846 propellant.

In addition to the pressure-time curve, it was desired to monitor the initiation and the duration of the gas flow from the orifice. Hence, positioning a photoelectric cell in the same horizontal plane and at right angles to the induced orifice provided the desired information. That is, by recording the output of the photoelectric cell on the screen on an oscilloscope, it is possible to determine when the gases first exit the orifice and for how long the flow continues.

An experiment was designed which allowed a parametric investigation of peak chamber pressure, initial hole size, and the effect of these parameters on the damage sustained by an aluminum cartridge case as a result of "burn-through". In order to relate quantitatively the damage inflicted to the cartridge case, the cases were weighed (to four decimal place accuracy) before and after firing; the amount of metal lost is indicative of the severity of the erosion. The initial hole sizes ranged from 0.0135 to 0.0625 inch in diameter. To obtain peak chamber pressures which ranged from 22 to 58K psi, the charge of WC846 propellant was varied from 21.0 to 28.0 grains. Figure 3 shows the results of this work - each data point is the average of five firings per condition of the experiment. As is shown in the plot, for a given particular pressure level, the least amount of damage - that is, cartridge case weight loss - occurs with the smallest hole size (0.0135 inch). Examining the data, a trend is established for increasing hole sizes of 0.0250, 0.0312 and 0.0400 inch in diameter. Namely, it is apparent that as the hole size is increased, the damage is correspondingly increased. However, the curve for the 0.0625 inch hole falls within the envelope defined by the 0.040 and 0.0135 inch holes. This is perplexing until one realizes that the peak pressures experienced during firing of these cases (0.0625 inch diameter hole) are lower than those resulting from firing the same charge with a smaller hole size. This brings up the second important consideration gained from this study. Namely, for any particular hole size, the effect of increasing chamber pressure is to increase the damage sustained by the cartridge case.

ANALYSIS. The mathematical analysis is directed to that portion of the cartridge case shown in Figure 4. Since the hole is drilled in the most massive region of the case, and since the length to diameter (L/D)

ratio of the hole is approximately 5, a reasonable approximation to the
region of the case where the hole is drilled is that of a small hole in a
medium of infinite extent in the radial direction. Distance along the
axis of the bore is specified by the variable Z, and distance from the
center line of the bore in the radial direction is given by the variable r.

It is first necessary to describe the flow of propellant gases from
the interior of the cartridge case through the induced fissure. To do so,
it is important to select the most representative gas dynamic model, based
upon bore size. With a L/D ratio of approximately 5, the relatively small
opening of 0.0135 inch in diameter, and the rapid rate of pressurization,
the flow may be assumed to be choked. The additional assumption is made
that the flow is developed to the point where it consists of a well de-
fined boundary layer and a central, core region. Figure 5 is a drawing
depicting this flow process. The temperature, pressure, density, and
velocity as a function of position in the core region can be determined
with the appropriate flow model. Determination of the nature of the
boundary layer, laminar or turbulent, can be accomplished by an examina-
tion of the magnitude of the local Reynolds number.

Lee and Sears (1) suggest an adiabatic treatment of the flow be con-
sidered providing the bore is not too long (less than ten bore diameters).
Admittedly, the flow process should account for the substantial transfer
of energy to the bore's sidewalls since melting has been hypothesized to
occur. The adiabatic assumption of the flow does not contradict the hypo-
thesized fact that there is substantial energy transfer. The analysis of
the core region merely permits a determination of the gas conditions – i.e.
temperature, pressure, density and velocity – external to the boundary
layer. It must be remembered that it is across the boundary layer that
the energy is transferred to the bore's sidewalls. Thus, the gas dynamics
in the core will be treated as adiabatic flow with friction in a duct of
constant area.

Shapiro (2) has developed a series of working formulas to describe
such a flow process. Momentum, energy, and mass equations are written
for the flow of a perfect gas through an elemental control volume.
Since the reservoir (chamber) conditions are available and since choking
is assumed to occur at the exit, these formulas may be used to predict
the temperature, pressure, velocity, and density at any position along
the bore's axis.

The Mach number as a function of position along the bore's axis must
first be determined. Lee and Sears (1) give a relationship for the bore
length L, required for the flow to pass from a Mach number, M_1, to a Mach
number, M_2, as

$$\frac{4fL}{D} = \frac{4f}{D} (L_{max})_{M_1} - \frac{4f}{D} (L_{max})_{M_2} \qquad (1)$$

where f is the friction factor, and D is the diameter of the bore.

The friction factor f, is defined by the Reynolds analogy as

$$f = \frac{2h_{cz}}{\rho v c_p} \qquad (2)$$

where h_{cz} is the heat transfer coefficient, ρ is the gas density, v is the gas velocity, and c_p is the specified heat as constant pressure.

Since at this point in the analysis, values for h_{cz}, ρ and v are all unknown, a value for the friction factor must be assumed. An iterative technique can be used later to identify a correct friction factor, once more information on the fluid properties within the central core region is available. Shapiro (2) states that friction factors in the range of 0.001 to 0.004 are realistic for the type of flow with which this analysis is concerned. Keenan and Kaye's (3) gas tables for the adiabatic flow of a perfect gas through a constant area duct with friction, in conjunction with equation (1), are used to determine the Mach number as a function of position throughout the bore for friction factors of 0.001, 0.002, and 0.004. These data are presented in Figure 6.

Now that the local Mach number is available, it is possible to evaluate both the free stream and stagnation values for the temperature, pressure, and density at any point along the bore's axis. The chamber conditions – those describing the combustion of the propellant grains inside the cartridge case – to be used through-out this analysis are:

$$T_o = 5040°R, \quad P_o = 2.5x10^4 \text{ psi}, \quad R = 64.372 \text{ ft-lb}_f/\text{lb}_m °R$$

$$\rho_o = 11.09 \text{ lb}_m/\text{ft}^3, \text{ and } \gamma = 1.24$$

To obtain the gas properties at the beginning of the bore – the inlet, an isentropic process is thought to be valid. Shapiro (2) provides the following governing equations for an isentropic process:

$$T_\infty = \frac{T_o}{1 + \frac{\gamma-1}{2} M^2} \tag{3}$$

$$P_\infty = \frac{P_o}{(1 + \frac{\gamma-1}{2} M^2)^{\gamma/\gamma-1}} \tag{4}$$

$$\rho_\infty = \frac{\rho_o}{(1 + \frac{\gamma-1}{2} M^2)^{1/\gamma-1}} \tag{5}$$

where the free stream conditions and the stagnation states are identified by the subscripts ∞ and o respectively. The local value for the gas velocity may also be determined from formula:

$$v = M\sqrt{\gamma RT_\infty}. \tag{6}$$

In order to determine the axial dependency of the free stream temperature, pressure, density, and velocity, equations (3), (4), (5), and (6) are again used. The appropriate value for the local Mach number is obtained from Figure 6. These calculations are performed for a friction factor of 0.002 at positions where the Mach numbers are 0.09 and 0.95, and 1.00 and are shown in Figure 7. Although a detailed treatment of the fluid mechanics has been addressed, Figure 7 shows that the free stream conditions do not vary appreciably along the axis of the bore.

The problem which is being studied is that of flow and heat transfer in a small bore. Kreith (4) states that for very short tubes or rectangular ducts with initially uniform velocity and temperature distribution, the flow conditions along the wall approximate those along a flat plate. Hence, the original two dimensional cylindrical geometry (r,Z) can be replaced with a two dimensional cartesian system (x, y). Further justification for the flat plate treatment of the problem may be obtained by showing that the boundary layer displacement thickness is small compared to the bore's radius. This fact implies that the boundary layer is essentially localized near the surface of the bore.

To determine the nature of the boundary layer, turbulent or laminar, the local Reynolds number is needed. The Reynolds number, based on length, is given by $Re_Z = v \rho Z / \mu$ where the values of the velocity, v, and the gas density, ρ, for a particular value of Z are obtained from Figure 7 and the absolute viscosity, μ, is 4.72×10.0^{-5} lb_m/ft-sec. Kreith (4) reports that the flow over a flat plate is turbulent where the local Reynolds number is approximately 3×10^5. Comparing 3×10^5 with the calculated values of 1.26×10^6 ($Re_Z = 0.030$ inch) and 2.40×10^6 ($Re_Z = 0.062$ inch), it is concluded that the flow is indeed turbulent.

It is now possible to calculate the boundary layer thickness at any point along the axis of the bore. Rohsenow and Choi (5) report that the boundary layer thickness, δ, for turbulent flow is

$$\delta = \frac{0.37}{(Re_Z)^{1/5}} \qquad (7)$$

For the flat plate assumption to be valid it is necessary to show

$$\frac{\delta^*}{a} \ll 1 \qquad (8)$$

where a is the bore's radius and δ^* is the boundary layer displacement thickness. Olson (6) states that the boundary layer displacement thickness for turbulent flow is given by $\delta/8$; hence, equations (7) and (8) may be combined and evaluated. For example, at Z = 0.031 inch-the bore's midpoint- δ^*/a is 0.0128, which is very much less than one, thereby justifying the flat plate treatment of the bore's surface.

The energy transport to the bore's sidewall occurs principally by forced convection. The effect of radiation can be shown to contribute little to the net heat flux. It is therefore important to define the Nusselt number in terms of the Prandtl and Reynolds numbers. With turbulent flow over a flat plate, Olson (6) states that a reasonable value for the Nusselt number may be obtained from

$$Nu_Z = 0.0288 (Pr)^{1/3} (Re_Z)^{4/5} \tag{9}$$

where Nu_Z implied that the Nusselt number is a function of position along the bore's longitudinal axis. Combining equation (9) with the classical definition of the Nusselt number, $Nu_Z = h_{cZ} Z/kg$, where kg is the thermal conductivity of propellant gas, it is possible to identify the convective heat flux to the bore's sidewall as

$$q = \frac{-0.0288 (Pr)^{1/3} (Re_Z)^{4/5} kg \ (Ts - Tg)}{Z} \tag{10}$$

where Tg is a characteristic gas temperature and Ts is the surface temperature of the solid. The gas properties used in the evaluation of the Reynolds and Prandtl numbers must be evaluated at a reference temperature, T*, determined empirically be Eckert (7) to be

$$T^* = T_\infty + 0.50 \ (T_0 - T_\infty) + 0.22 \ (T_{a.w.} - T_\infty).$$

Consideration of the recovery factor permits substitution of the adiabatic wall temperature, $T_{a.w.}$, by the stagnation temperature so that the above equation becomes

$$T^* = T_\infty + 0.72 \ (T_0 - T_\infty).$$

The heat flux described by equation (10) is incident to the bore's sidewalls. By showing that the thermal layer does not penetrate the solid to any significant depth during the interior ballistic cycle, it is possible to treat the conduction problem as a one dimensional slab instead of as a region exterior to a cylindrical hole of diameter, D, and extending to infinity in the radial direction. The heat conduction is then reduced to a one dimensional, time-dependent problem subject to a Newton's Law of Cooling boundary condition of the third kind. Rohsenow and Choi (5) provide a solution to such a problem and further point out that the heating of a thick body by a hot fluid at the surface approximates this case during the early stages of the transient. Their solution is as follows

$$\frac{T_s - T_1}{T_g - T_1} = [erfc \ (\frac{x}{2\sqrt{\chi t}}) - exp (\frac{x h_{cZ}}{k_s} + \frac{\chi t}{(k_s/h_{cZ})^2}) erfc \ (\frac{x}{2\sqrt{\chi t}} + \frac{\sqrt{\chi t}}{k_s/h_{cZ}})] \tag{11}$$

where χ and k_s, are the thermal diffusivities and conductivities of the metal in question, T_i is the initial temperature of the solid, and t is the time variable.

DISCUSSION. The analytical model, developed previously and culminating in equation (11), allows determination of the bore surface temperature given gas flow. To investigate the performance of brass and aluminum alloy cartridge case, it is a simple matter to substitute the appropriate physical parameters -χ and k_s - and calculate the surface temperature at several axial positions as was done to prepare Figures 8 and 9. Figure 8 implies that the surface of an aluminum bore (x = 0) reaches the melting temperature after approximately 0.35 milliseconds; the brass bore's melting temperature (Figure 9) is approached only late in the interior ballistic cycle.

To predict the total amount of metal removed during a "burn-through," equation (11) is solved using a forward difference-in-time technique with the additional condition that for any point to be removed, it must realize a local temperature equal to the melting temperature plus an additional amount of heat to cause an elemental mass, Δm, to liquify. Therefore, the total heat flux required to cause melting, hence metal removal, will be

$$q_{removed} = \Delta m C_\rho (T_{melt} - T_i) + \Delta m H_f \qquad (12)$$

where H_f is the heat of fusion. Figure 10 shows the results of such an exercise (labeled classical melting) and the results of an experimental effort wherein the amount of metal lost during "burn-through" has been correlated with peak chamber pressure. A comparison shows that the experimental results are not predicted entirely by a classical melting theory.

Since this analysis has been initiated upon the premise that the case degradation is the result of thermal energy transport, it is necessary to determine if any additional energy sources are present. This search is based upon the knowledge that aluminum is a very chemically reactive metal and that there is a large, bright gaseous discharge attending this event. It violates our intuitive understanding of the phenomenon, however, to accept that there is sufficient energy feedback to the case from the exterior cloud to account for the additional damage.

In order to determine if any additional energy sources were present during an aluminum cartridge case "burn-through", high speed motion pictures and still photographs were taken of this phenomenon. Figure 11(a) shows the characteristic plume resulting from propellant gases, generated in a combustor, passing through a 0.0135 inch (diameter) hole in an aluminum test specimen. This discharge has been channeled into a plexiglas cylinder and produces a bright incandescence throughout the entire cylinder. However, when the cylinder was flushed with nitrogen, the major portion of gaseous discharge has been quenched as is shown in Figure 11(b). All that remains as a result of the inert atmosphere is a small localized region in close proximity to the specimen. It is interesting to note that

-482-

although a major portion of the discharge has been eliminated by the nitrogen atmosphere, the damage (metal removal) inflicted to each test specimen is the same. Also, it is apparent that the "burn-through" plume occurs in two separate phases - a large secondary cloud existing exterior to the aluminum specimen (identified as the oxidation of aluminum to aluminum oxide) and a localized primary reaction zone existing in close proximity to the specimen.

Since the ignition and combustion of aluminum has been well investigated, the literature provides some insight into the nature of the primary reaction zone. A doctoral thesis by Brzustowski (8) of Princeton University enables the identification of this zone as a vapor phase reaction between freshly exposed aluminum and the combustion gases. Also, Brzustowski (8) has correlated the ignition of the protective oxide coating on the wire. This fact will be used later in the discussion. Hence, it is concluded that the localized primary reaction zone is responsible for additional metal removal and must be considered in the mathematical model.

To assure that the understandings developed in a combustor are applicable to a gun environment, a test weapon was placed in an air tight box. The test weapon was modified so that a direct observation, with a high speed motion picture camera, could be made of the exit plane of a drilled hole in an aluminum cartridge case. It is possible to use the fact that discharging the secondary plume into an inert atmosphere will quench and reduce the intensity of the secondary cloud. The elimination of the external cloud would permit a close observation of the hole where the localized primary reaction zone is occurring. Figure 12 is a photograph of the test weapon in the box. After the lid is put in place and clamped down, the interior of the box is filled with helium. This gas is allowed to flow continually throughout the box assuring a one-hundred percent atmosphere of the test gas.

Figure 13 shows a series of selected frames abstracted from a high speed motion picture study of the firing of an aluminum cartridge case. In this photograph, the induced hole is observed at the left of the frame and the discharge throughout a viewing slot in the test weapon. In the flames identified t = 0.000+ to t = 0.286 milliseconds, propellant gases- as indicated by their orange glow - are observed exiting the orifice This is the heat-up period. The flame identified t = 0.357 milliseconds shows the first evidence of our localized primary reaction zone. This region of exothermic reaction continued for the remainder of the experiment time shown. However, the presence of this zone in a gun environment has been experimentally verified together with the correlation of Brzustowski (8). Namely, that in order to initiate the vapor phase reaction, the aluminum must undergo melting. The time of 0.357 milliseconds, from the high speed motion picture study, compares quite favorably with the model's prediction of 0.35 milliseconds. It is therefore concluded that as the propellant gas flows throughout the induced hole, the bulk aluminum situated below the protective oxide coating, which lines the surface of the hole, is melting. Due to the continuing efflux of propellant gases, the protective oxide coating can no longer maintain its structural integrity and gets washed away because the bulk aluminum immediately below has melted. Once fresh aluminum is exposed to the propellant gases, the protective oxide layer is unable to form and a vapor phase reaction is initiated.

-483-

To determine the magnitude of the heat flux resulting from the primary reaction zone, a double disk arrangement of aluminum and brass test specimens was used in the combustor. There were two variations of this experiment. First, the aluminum was placed closest to the combustor and the brass faced the atmosphere. In the second experiment, the orientation of the two test specimens was reversed. The data obtained from this experiment are shown in Figure 14. It is possible to account for the drastic differences in these data. If the aluminum disk first witnesses flow of propellant gas, the localized exothermic reaction will be exposed to the brass specimen. The final result being that the brass specimen is exposed to much more heat flux than would be expected if the aluminum specimen were not present. If, on the other hand, the propellant gas first passes through the brass specimen and then the aluminum, the exothermic reaction associated with the aluminum will be carried (by the fast moving propellant gas stream) to the atmosphere. There will be no material on which the exothermic reaction can act. By measuring the amount of metal removed from the brass specimen under both cases of the experiment and by using equation (12), it is possible to calculate the additional energy flux to the brass specimen as a result of the exothermic reaction. The brass specimen, in a sense, is being used as a calorimeter for the primary reaction occurring with the aluminum specimen. Once determined, this additional energy flux is used as a corrective factor to the melting theory. Figure 15 shows a comparison between the experimental results and the theoretical predictions when augmented by the heat flux from the primary reaction zone. The agreement is favorable.

CONCLUSIONS. This work was aimed at understanding the "burn-through" problem that has impended orderly engineering development and application of aluminum alloy cartridge cases in high-performance ammunition since the 1890's. It has been shown that a gas path through the wall of an aluminum case, and through which propellant gas can flow during the internal ballistic cycle, is a precursor to the "burn-through" phenomenon. Once this gas path has been established in an aluminum cartridge case, melting of the path surface starts early in the flow cycle. On the other hand, the melting point of a similar surface in a brass case is only approached late in the cycle. Hence, the key factor in an aluminum cartridge case, is the onset of melting early in the interior ballistic cycle. This melting is almost instantaneously followed by primary, exothermic chemical reactions of propellant gas molecules with available aluminum alloy material. This exothermically aggravates the ablation of the aluminum case surface and adjacent steel weapon surfaces over which the conglomerate molten and reacting material flows. This primary reaction is followed by a secondary reaction which consists of the oxidation of unreacted case material that is blown into the atmosphere.

Armed with this knowledge, solutions to this problem have been found that either prevent propellant gas flow through a path in the case that develops unintentionally during firing of the ammunition, or alter the effect of propellant gas flow through such a gas path. Since an engineering understanding of the "burn-through" phenomenon is available, work is currently underway to demonstrate the feasibility of aluminum cartridge cases thereby advancing the program from exploratory development to advanced engineering development.

REFERENCES

1. Lee, J. F., and Sears, F. W., Thermodynamics, Addison-Wesley, Cambridge, Mass., 1955.

2. Shapiro, A. H., The Dynamics and Thermodynamics of Compressible Fluid Flow, Vol. I, Ronald Press, New York, 1953.

3. Keenan, J.H., and Kaye, J., Gas Tables, Wiley, New York, 1956.

4. Kreith, F., Principles of Heat Transfer, International Textbook Company, Scranton, Penna., 1963.

5. Rohsenow, W. M., and Choi, H., Heat, Mass, and Momentum Transfer, Prentice-Hall, Englewood Cliffs, New Jersey, 1961.

6. Olson, R.M., Essentials of Engineering Fluid Mechanics, International Textbook Company, Scranton, Penna., 1962.

7. Eckert, E.R.G., Transactions of the American Society of Mechanical Engineers, Vol. 78, 1956.

8. Brzustowski, T.A., "Vapor-Phase Diffusion Flames in the Combustion of Magnesium and Aluminum," Ph.D. Thesis, Princeton University, Dept. of Aeronautical Engineering, (1963).

Brass   Aluminum   Unfired
                              Aluminum
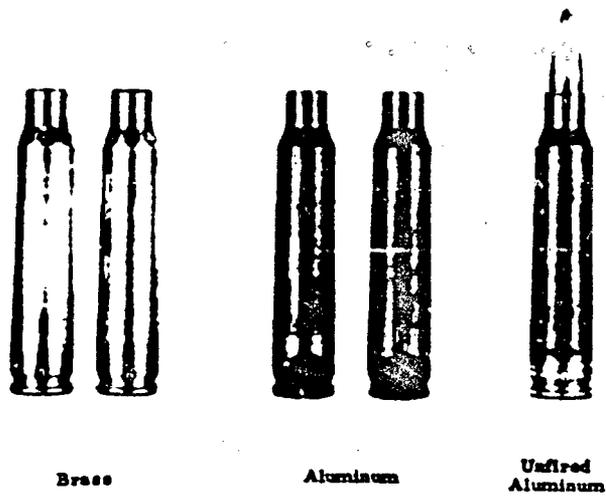
Figure 1. Drilled Hole Experiment
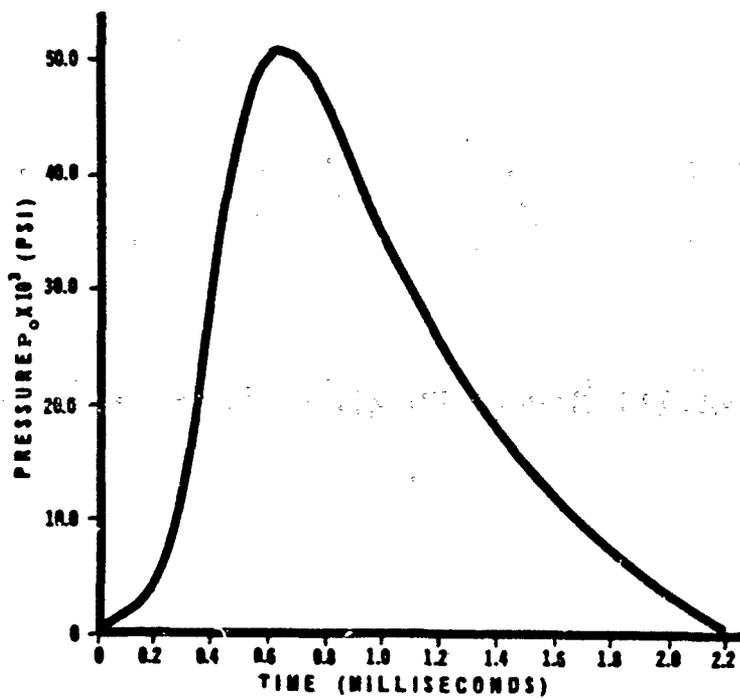
Figure 2. Pressure vs Time for Standard Charge in Aluminum Case
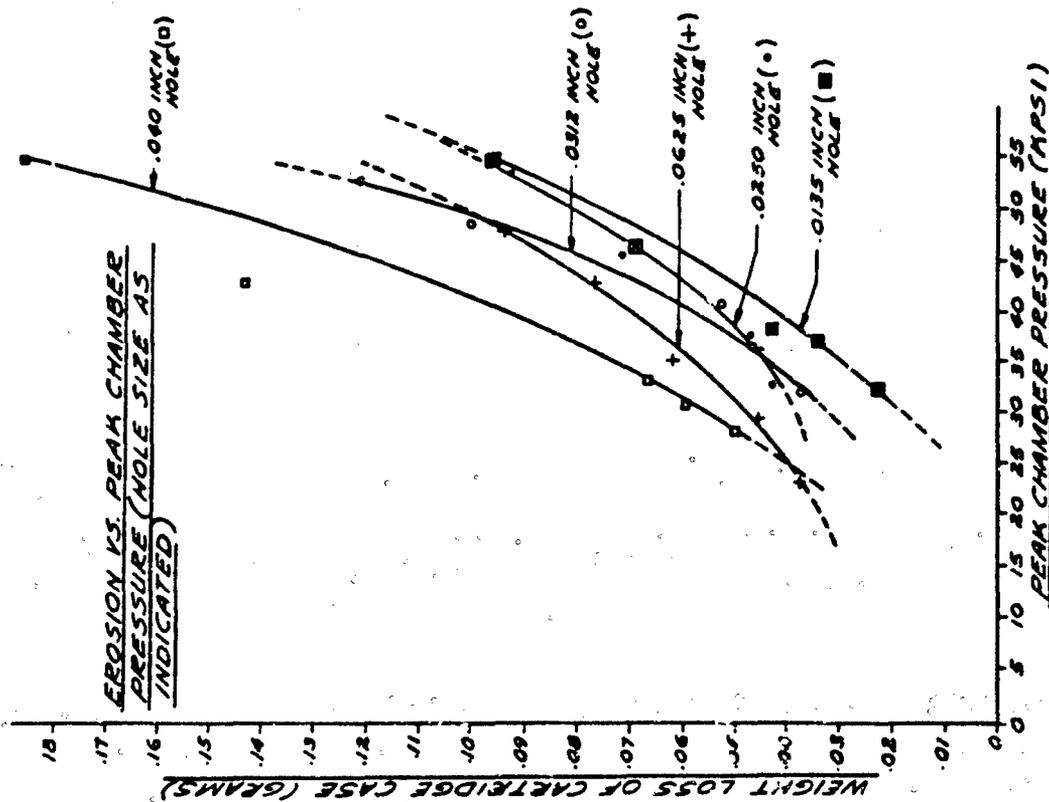
Figure 3. Results of Parametric Investigation --
Drilled Hole Experiment

-487-
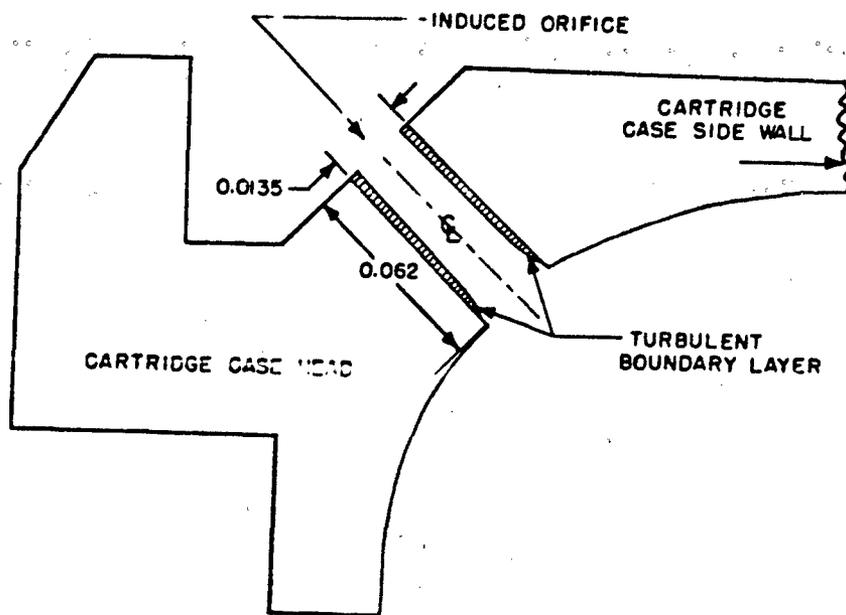
Figure 4. Head Region of 5.56mm Aluminum Alloy Cartridge Case
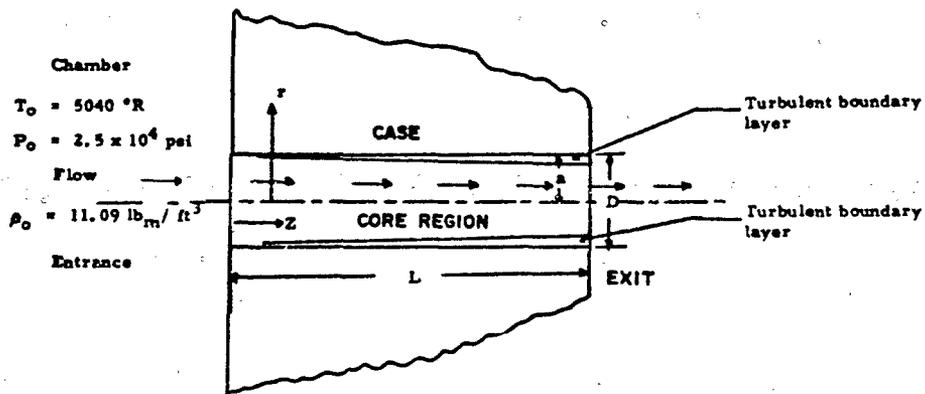
Figure 5. Schematic of Flow Process
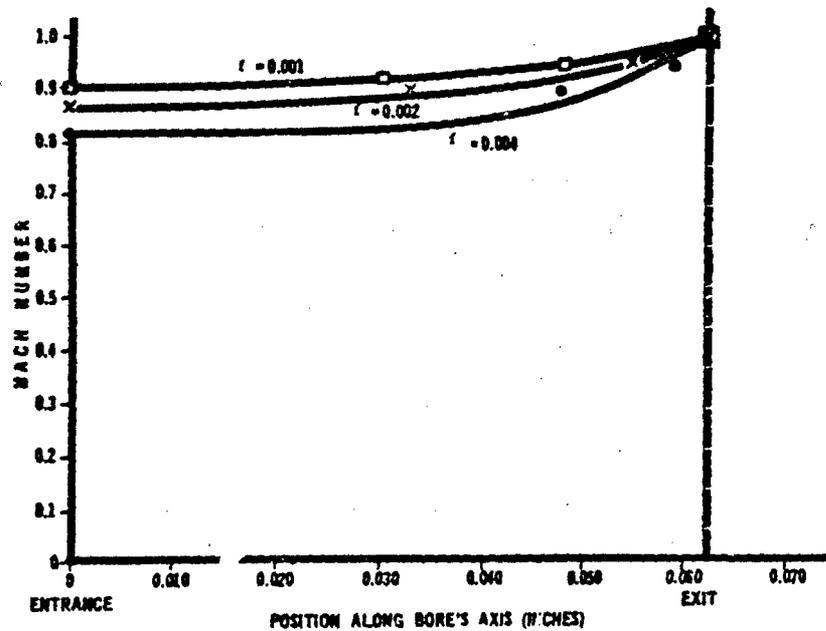
Figure 6. Mach Number as a Function of Position Along the Bore (Friction Factor as Indicated)
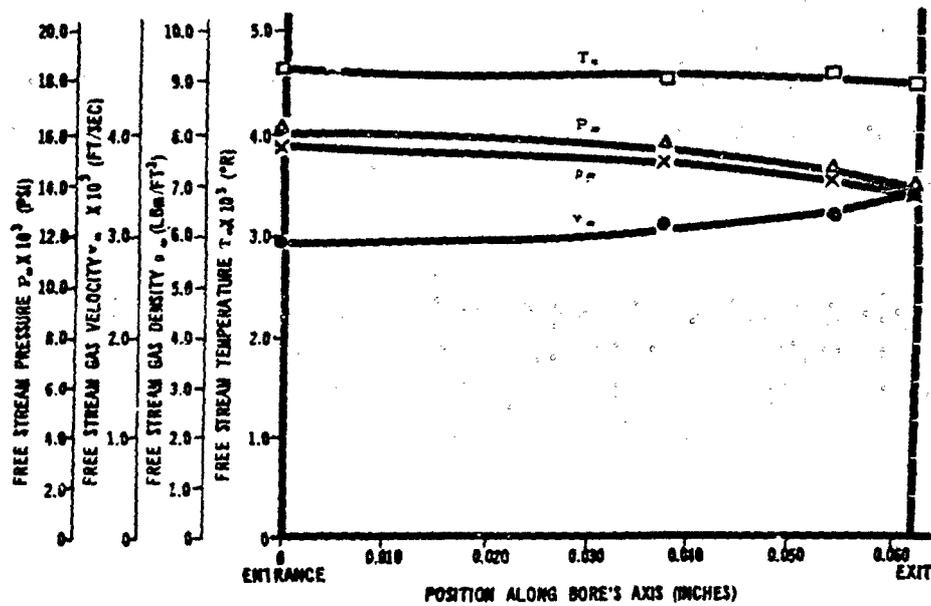
Figure 7. Free Stream Conditions vs Position Along Bore's Axis (0 = 0.002)
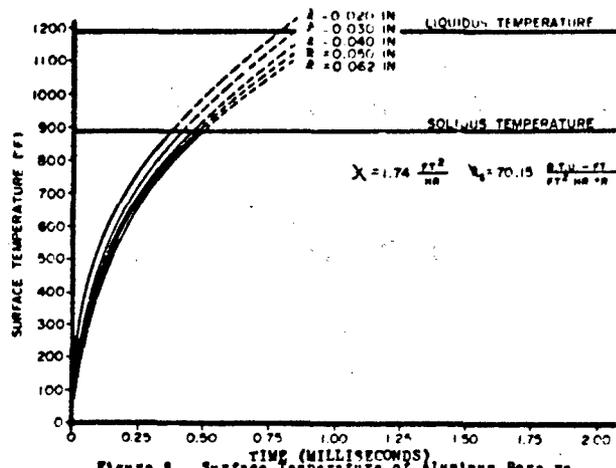
Figure 8. Surface Temperature of Aluminum Bore vs Time (Position as Indicated)
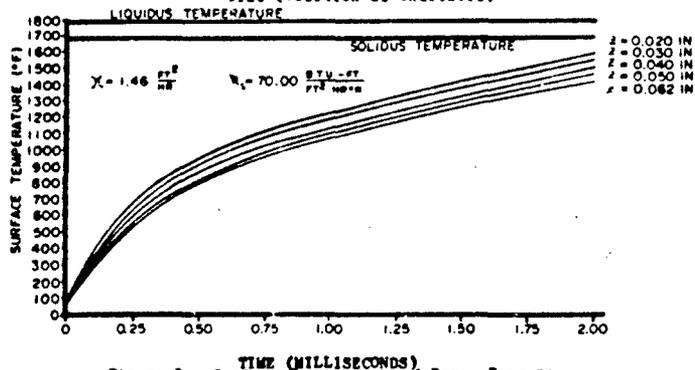
Figure 9. Surface Temperature of Brass Bore vs Time (Position as Indicated)
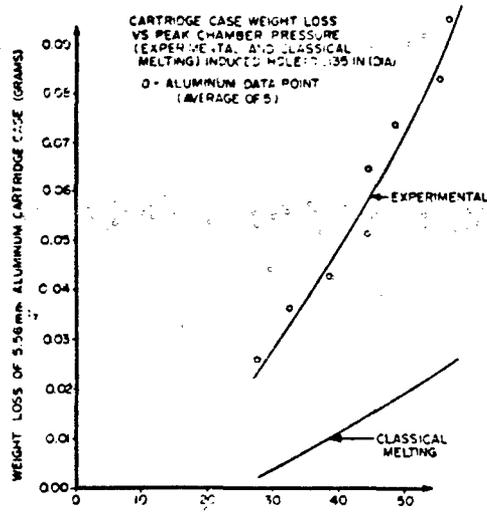
Figure 10. Comparison of Classical Melting Theory and Experiment
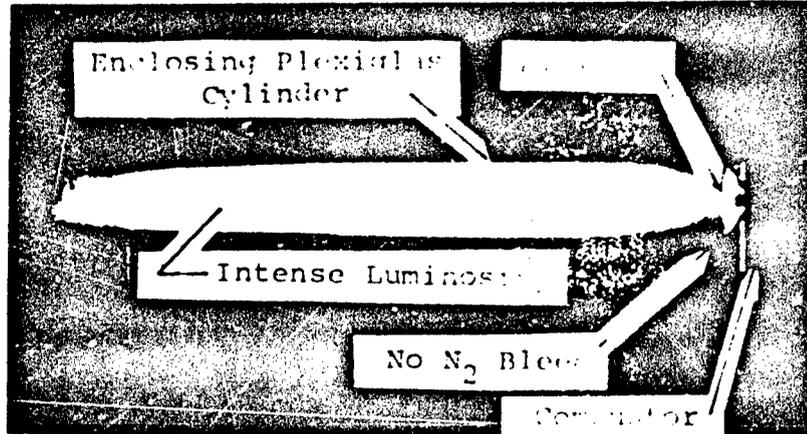
(a) Discharge into Air Environment
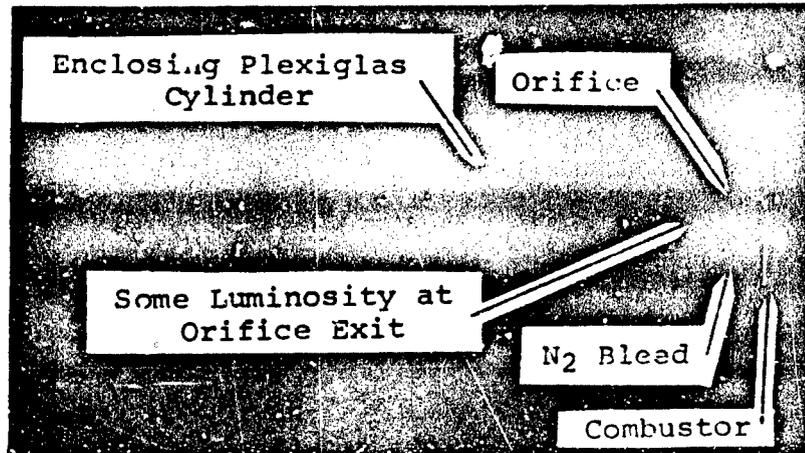
(b) Discharge into Nitrogen Environment

Figure 11. Time Exposed Photographs of the "Burn-Through"
Plume into Air and Nitrogen Environments

-491-
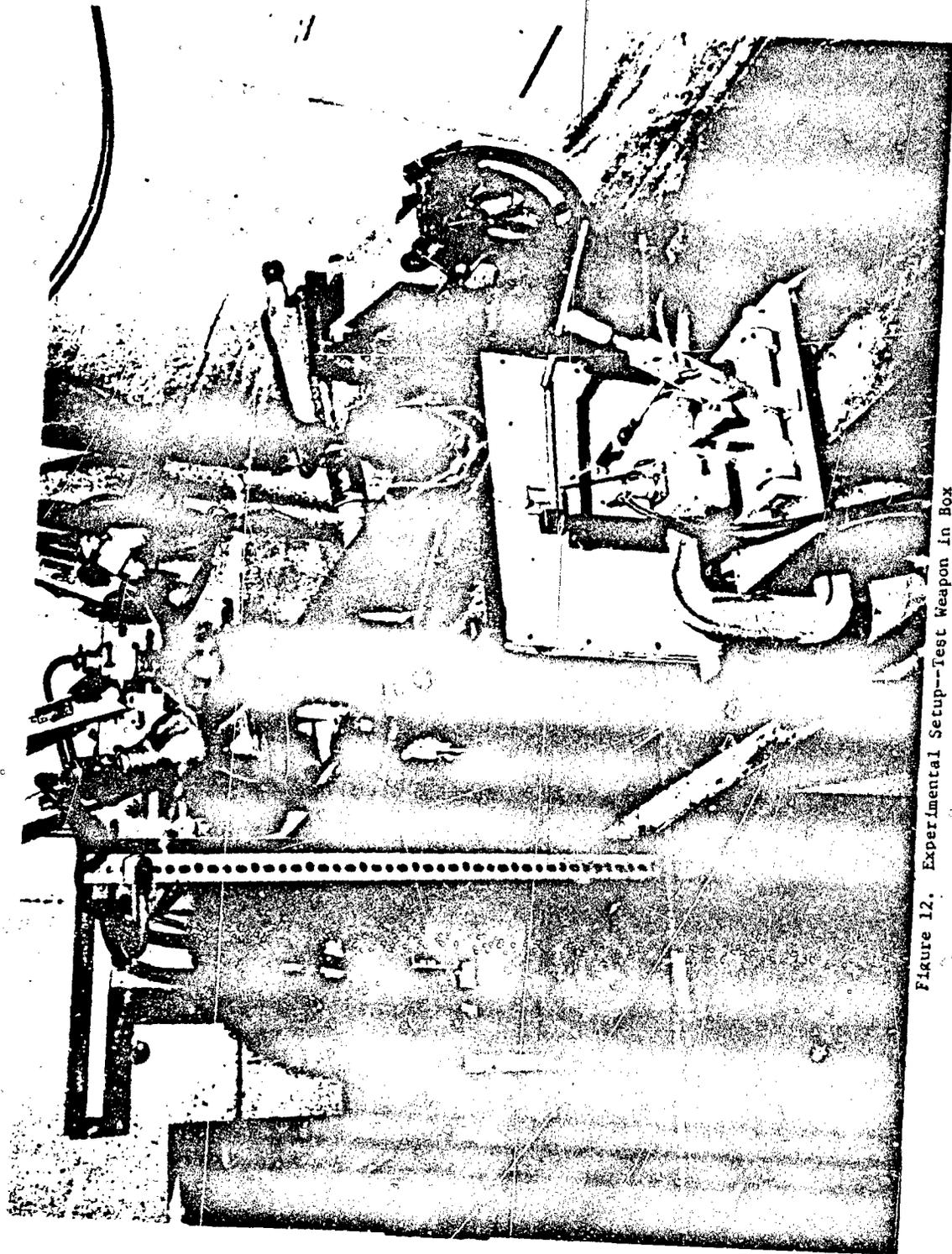
Figure 12. Experimental Setup--Test Weapon in Box

Figure 13.

Initial Exposure

t = 0.003m Sec

t = 0.071m Sec

t = 0.143m Sec

t = 0.214m Sec

t = 0.286m Sec

t = 0.357m Sec

t = 0.428m Sec

t = 0.571m Sec

t = 0.642m Sec

t = 0.714m Sec

t = 0.785m Sec

t = 0.857m Sec
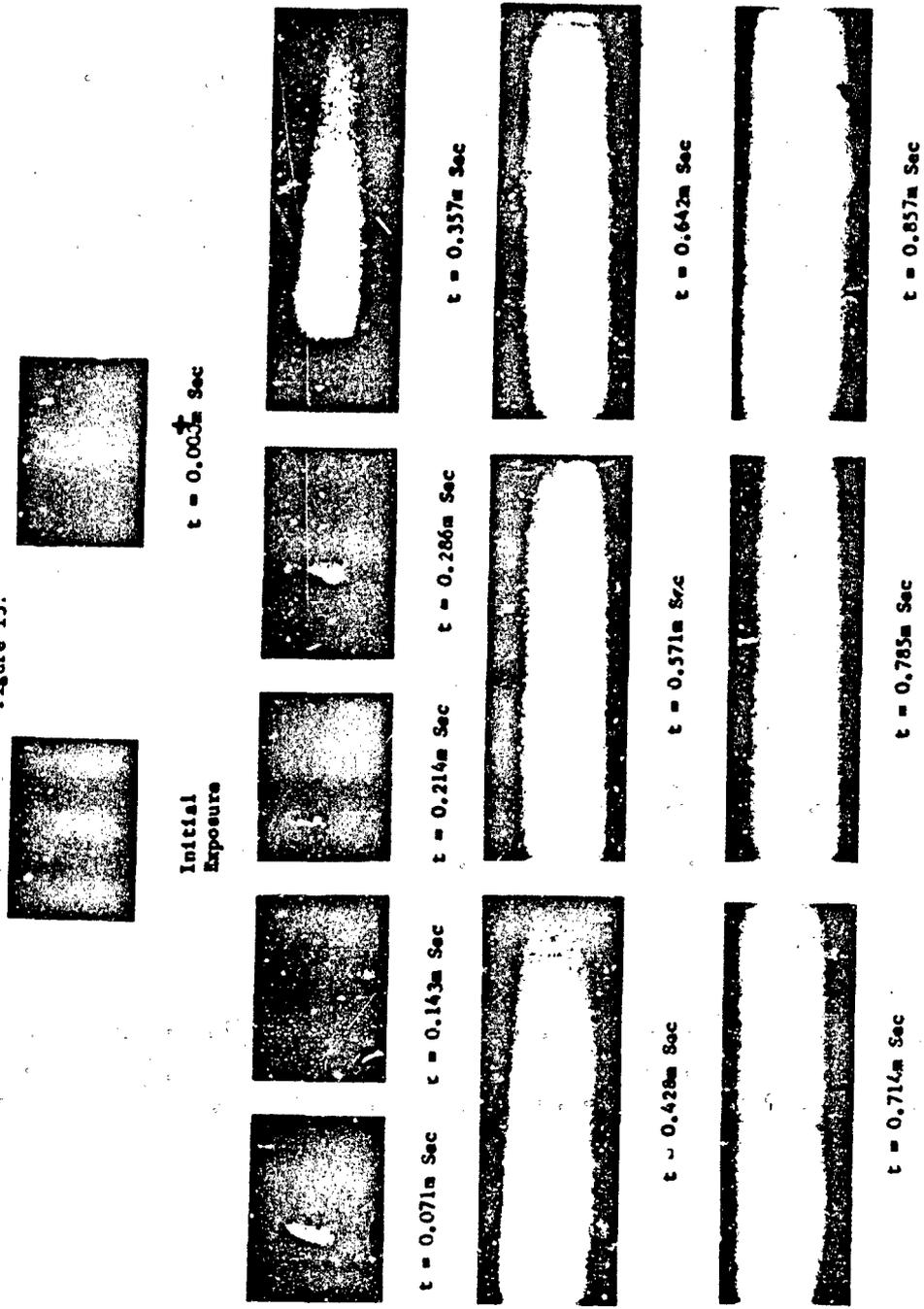
5.56MM ALUMINUM CARTRIDGE CASE

Framing Rate: 14000 Frames/Sec   Propellant Charge: 27.4 Grains

Atmosphere: Helium   Hole Size: 0.0625 In. (dia.)
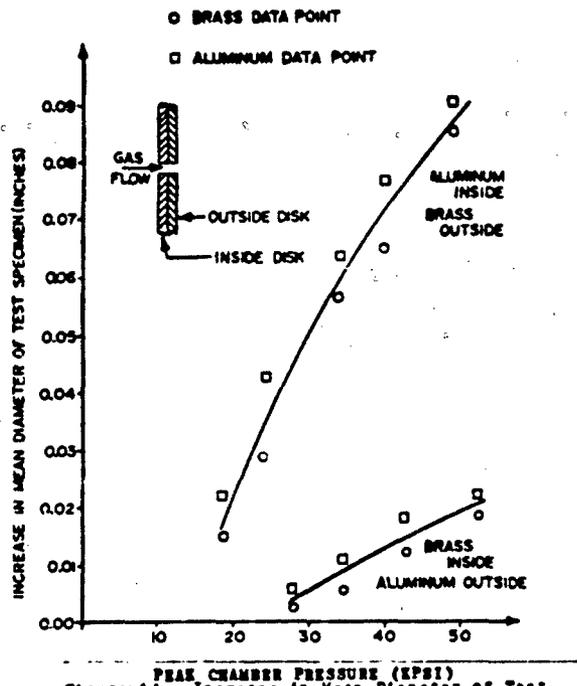
Figure 14. Increase in Mean Diameter of Test
Specimens in Double Disk Experiment
vs Peak Chamber Pressure
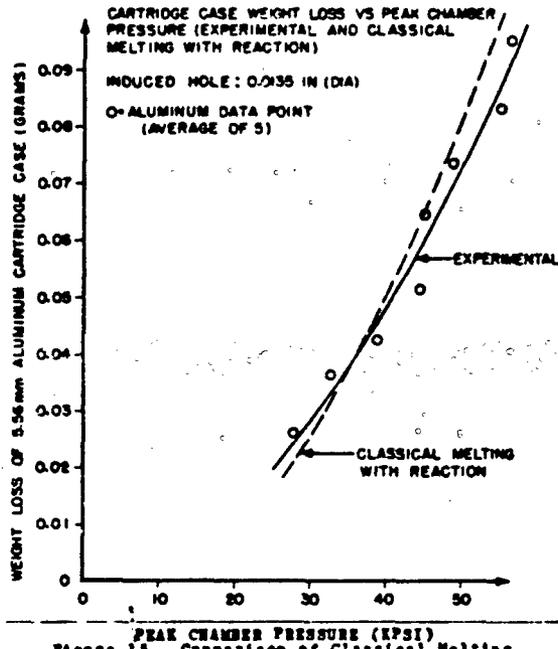
Figure 15. Comparison of Classical Melting
Theory with Reaction and Experiment

STATISTICAL MODELING OF PROPAGATION LOSS DATA

M. Acker, R. D'Accardi, D. Dence and C. Tsokos*

U. S. Army Electronics Command, Fcrt Monmouth, New Jersey

ABSTRACT. Since 1962 under Project SEACORE the US Army Electronics Command and the Advanced Research Projects Agency have supported investigations and measurements in Southeast Asia and the United States to improve communication electronics system performance in heavily forested environments. The present report deals with one aspect of these investigations, namely, the statistical characterization of path loss data which was collected by Jansky and Bailey. The object of the present study is threefold: First, we briefly discuss some of the findings which are contained in previous presentations of Southeast Asia propagation data. Secondly, we compare some of the previous findings which viewed the data as a deterministic phenomena with a more realistic stochastic formulation. Thirdly, we set forth recommendations for further defining the statistical character of propagation data.

The study shows that the propagation loss data exhibits strong random fluctuations, especially above 100 MHz; and the deterministic equations do not properly describe the statistical character of the path loss data. In our investigation we rely solely on statistical modeling of the propagation loss data to account for these random fluctuations. Non-parametric statistical techniques are employed to analyze the data for specific configurations of transmitter antenna height, receiver antenna height, distance, frequency, and polarization.

The information obtained from the statistical modeling of the data has been extremely useful in the feasibility and the design of communication-electronics equipment and systems.

1. INTRODUCTION. In recent years extensie investigations and measurements have been made in Southeast Asia and the United States to determine the communication conditions that exist in forest environments. Studies were initiated in 1962, spcnsored by the Advanced Research Projects Agency and performed under the direction of the U. S. Army Electronics Command as part of the Southeast Asia Communications Research (SEACORE) Program. The overall aim was to help over-come severe radio communications problems occurring in Southeast Asia. The present report deals with one aspect of these investigations, namely, the measurement and analysis of path loss data. This path loss information was obtained by Jansky and Bailey (a division of Atlantic Research Corporation), one of the prime contractors engaged in the SEACORE Program, and involved making extensive measurements at various locations in Thailand. The path loss measurements covered the frequency range from 100 kHz to 10 GHz encompassing a wide range of antenna heights, locations and seasonal rainfall variations.

---

*Department of Mathematics and Statistics, University of South Florida, Tampa, Florida

The aim of our presentation is threefold: First, we shall mention briefly some of the findings which are contained in the presentation of the SEACORE path loss data, contained in references (1,2). Secondly, we shall be concerned with spot check statistical analysis of the information in the frequency range 2-400 MHz the results of which are then compared with previous results. Thirdly, we shall set forth recommendations for futher defining the statistical character of propagation loss data consistent with the original objectives of the SEACORE Project.

In Section 2 we shall briefly discuss the presentation of some very important and basic concepts given by Jansky and Bailey. Primarily, we shall be concerned with the manner in which the data was normalized to a common distance.

A more realistic presentation of the analysis of the propagation loss data from a statistical point of view will be given in Section 3. In addition, we shall compare our spot check statistical analysis of the path loss information with that obtained by J & B.

In Section 4, we shall present a summary of our preliminary findings, including recommendations for further analysis.

2. COMMENTS ON PREVIOUS STATISTICAL ANALYSIS OF PATH LOSS DATA. There are many points which one should take into consideration with respect to the manner in which the data was presented in the reports by J & B. There was limited resources available to J & B to delineate statistical techniques for processing and analyzing the data. Therefore only very limited statistical analysis was performed from a rigorous point of view. To increase the data base from which statistical observations could be drawn, J & B normalized all the data to a distance of one mile.

To accomplish this normalization they utilized the analytical expression,

$$Y_i = X_i - 40 \log d_i \qquad\qquad (1)$$

where: Y_i = normalized path loss,

           X_i = measured path loss over distance d_i,

           d_i = path length in miles, $i = 1,2,\ldots\ldots n$,

which supposedly describes the path loss as a function of distance. This analytical approach is applicable when one views the jungle as a deterministic environment. Previous analytical studies (3,4) by various researchers have shown that in the frequency range 2-100 MHz the main mechanism which guides the energy from the transmitter to the receiver is a lateral wave which exhibits a distance power loss of 40 log d. This particular theoretical behavior occurs when one characterizes the electrical characteristics of the jungle from a deterministic viewpoint. The fact that the data exhibit

-496-

large random fluctuations indicates that a statistical normalization procedure is more desirable than the deterministic approach. This statistical procedure will be discussed in the subsequent sections. It should be noted, however, that differences of 10 dB and larger occurred quite frequently when using the deterministic characterization of the path loss data.

In an attempt to determine whether or not the propagation loss varied from a wet season to a dry season, J & B categorized the data into two effects, namely, wet and dry. This classification was based on less than scientific guidance obtained in Thailand as to what months constitute the different seasons. Certain minor discrepancies were noted in that the data was not correctly categorized using this procedure. It was noted that the wet season contained many periods with little rainfall and vice versa, and it was almost impossible to resolve the question without further study.

It should be mentioned that it was not a specific objective of the measurement program to determine whether or not the propagation loss in wet vegetation is the same as that in dry vegetation, though attempts were made to impact available data on as many questions as possible. In this case all the variables were not controlled sufficiently to answer this question. For example, if a slightly different path were used to measure the path loss under wet conditions than under dry conditions, the effect of rainfall could be easily masked by the difference in losses caused by the different paths themselves. Further analysis, utilizing a more realistic classification, will be presented in subsequent sections.

3. STOCHASTIC CHARACTERIZATION OF THE PATH LOSS DATA. One of the major and most important points in the analysis of the SEACORE data is the manner in which one attempts to normalize the propagation loss to a common distance. As mentioned previously, J & B used x = 40 log d as the normalizing procedure. This is the theoretical distance behavior of the propagation loss when one considers the electromagnetic environment to be deterministic in nature. In our analysis, we have found a strong indication that the information dictates otherwise, especially above 100 MHz.

The information collected by J & B was logarithmically transformed prior to their normalizing the data to a common distance. This transformed data was obtained by ECOM and analyzed (with the results set forth in this paper). It should be recognized, however, that a fundamental question which must be considered is the extent to which this transformation altered the statistics of the data. Time did not permit this question to be fully considered. Some preliminary estimates yielded results which indicated that the logarithmic transformation did not significantly alter the statistics. Nevertheless, this question should be considered in the future by performing a more complete analysis.

Therefore, in our preliminary analysis, for specific configurations of frequency, transmitter antenna height, receiver antenna height, polarization, distance and wet and dry classifications, i.e. (f, T, R, P, d, c) we

obtained an estimate of the path loss distance dependency, α_1, in such a way that the variance associated with the estimate will be minimum. The expression that we obtained for the estimate of "α_1" is given by:

$$\hat{\alpha}_1 = \frac{\sum\limits_{i=1}^{N} x_i \log (d_i)}{\sum\limits_{i=1}^{N} \log^2 (d_i)} \tag{2}$$

where: $\hat{\alpha}_1$ = estimate of the distance dependency coefficient,

X_i = measured path loss data in dB,

d_i = distance corresponding to the path loss data.

We made spot checks of the $\hat{\alpha}_1$'s for radial A data to determine the best estimates for the following frequencies: 2, 6, 12, 25.5, 50, 100, 250*, and 400* MHz, and for various combinations of transmitter and receiver antenna heights at distances of .2-2.0 miles. The data were classified as wet or dry by two criteria: (i)   rainfall greater or less than 3 inches per month
(ii)   rainfall greater or less than 6 inches per month.

The calculated best estimates, $\hat{\alpha}_1$, for the selected configurations are given in Tables Ia, and Ib.

---

*The samples containing the measurements for 250 MHz and 400 MHz may be somewhat biased, (1).

TABLE Ia: SUMMARY OF PATH LOSS DISTANCE DEPENDENCY, ALPHA 1 FOR RADIAL A* - VERTICAL POLARIZATION FOR 3" - 6" RAINFALL CRITERIA**

FREQ (MHz)	T (Ft)	R (Ft)	DISTANCE (Mi)	ALPHA 1***	
				WET	DRY
25.5	40	20	0.2-2.0	37.9	31.4
		42		30.9	31.6
		79		31.4	33.6
50	40	20	0.2-2.0	23.7	21.5
		42		33.0	28.5
		79		32.0	29.7
100	40	20	0.2-2.0	49.4	32.5
		42		50.9	31.1
		79		43.1	31.7
250	40	20	0.2-2.0	16.0	27.5
		42		34.4	23.5
		79		38.4	28.5
400	40	20	0.2-2.0	31.2	21.2
		42		14.2	15.0
		79		36.0	34.4

*   Radial B data was not examined.
**   The results for the 3" and 6" rainfall criteria were the same.
*** To achieve the proper distance dependency, the data was referenced to a common distance, which for our purposes was chosen to be 0.2 mile.

TABLE Ib: SUMMARY OF PATH LOSS DISTANCE DEPENDENCY, ALPHA 1 FOR RADIAL A* - HORIZONTAL POLARIZATION FOR THE 3" - 6" RAINFALL CRITERIA**

FREQ (MHz)	T (Ft)	R (Ft)	DISTANCE (Mi)	ALPHA***	
				WET	DRY
2	40	23	0.2-2.0	39.0	41.5
		45		38.0	40.1
		76		38.0	40.3
6	40	23	0.2-2.0	42.7	42.9
		45		52.5	39.1
		76		42.4	40.2
12	40	23	0.2-2.0	32.1	38.7
		45		33.3	39.3
		76		32.4	33.9
25.5	40	20	0.2-2.0	33.5	30.6
		42		37.0	29.7
		79		37.4	31.9
50	40	20	0.2-2.0	38.6	37.0
		42		45.2	41.5
		79		39.9	35.3
100	40	20	0.2-2.0	40.2	39.3
		42		35.1	31.7
		79		38.3	34.3
250	40	20	0.2-2.0	28.3	27.8
		42		15.9	17.6
		79		21.8	32.5
400	40	20	0.2-2.0	23.4	31.2
		42		13.3	33.3
		79		36.3	44.7

* Radial B was not examined.

** The results for the 3" and 6" rainfall criteria were the same.

*** To achieve the proper distance dependency, the data was referenced to a common distance, which for our purposes was chosen to be 0.2 mile.

We have found that for specific configurations of the parameters involved, there is a fluctuation from 13.3 to 52.5 dB. It is clear from Tables Ia and Ib that the estimates, $\widehat{\alpha_1}$, behave as a random variable. These preliminary findings indicate that the media should definitely not be considered as a deterministic phenomena and that a deterministic presentation of the data would give misleading results with respect to further characterization of the propagation loss as a function of the various independent parameters. When one utilizes the deterministic formula one can not help but force the data to accept the theoretical behavior of a deterministic phenomena. Thus, the resulting conclusions would be consistent with previous theoretical efforts. That is, the resulting conclusions are forced to be consistent to those which the deterministic theory dictates.

Results shown in Table II reveal a mean $\overline{\alpha_1}$ of 35.7 dB for the dry classification with a standard error of 6 dB, and for the wet classification a sample mean of 34.8 dB with a standard error of 9.0 dB. These values of standard error indicate that the respective averages for dry and wet classifications are not adequate. Furthermore, if we combine the wet and dry data, (we will discuss the feasibility of combining the data in subsequent remarks) we obtain a sample mean $\overline{\alpha_1}$, of 35.3 dB, and a standard deviation of approximately 7.6 dB. This simply indicates that if we are allowed to combine the wet and dry propagation losses, that is, if there is no significant difference between the two data sets, we should be utilizing, as a rough estimate of α_1 , approximately 35 dB to normalize the data with respect to distance. Since previous investigations employing a deterministic formulation (3,4) have shown that the distance dependency of the path loss varies as 40 log d in the frequency range 2-200 MHz, the data was separated into two sets, namely,

TABLE II: MINIMUM VALUE, MEAN VALUE, MAXIMUM VALUE, AND STANDARD DEVIATION OF ALPHA 1 FOR WET, DRY, AND WET PLUS DRY CONDITIONS WITH 3" - 6" RAINFALL CRITERION*, HORIZONTAL POLARIZATION

FREQ (MHz)	SEA	MINIMUM ALPHA 1	MEAN ALPHA 1	MAXIMUM ALPHA 1	STANDARD DEVIATION ALPHA 1
2-400	Wet	13.3	34.8	52.5	± 9.0
2-400	Dry	17.6	35.7	44.7	± 6.0
2-400	Wet + Dry	13.3	35.3	52.5	± 7.6
2-100	Wet + Dry	29.7	37.9	52.5	± 4.6
250-400	Wet + Dry	13.3	27.2	44.7	± 9.2

* The results for the 3 inch and 6 inch criteria were the same.

2-100 MHz and 250-400 MHz, and examined. Table II shows $\bar{\alpha}_1$ = 37.9 dB with a standard error = 4.6 dB for the 2-100 MHz range and $\bar{\alpha}_1$ = 27.2 dB with a standard error = 9.2 dB for the 250-400 MHz range. This indicates that the deterministic 40 log d more closely fits the 2-100 MHz range. In view of the above remarks one should consider the following equation for normalizing the propagation losses referenced to a common distance:

$$Z_i = X_i - \alpha_1 \log d_i, \quad i = 1, 2, \ldots\ldots n, \qquad (3)$$

where Z_i is the normalized path loss data.

The second most important phase of obtaining a complete statistical analysis of the data, having now obtained an acceptable normalizing procedure, is to investigate the type of probability distribution function which characterizes the normalized experimental information. Since we are confining our analysis from .2-2.0 miles, we are restricting ourselves to the amount of data available for analysis. In most cases, this is around 10 measurements for each configuration. Presently, we have no evidence that the normalized information of the wet and dry classification can be combined to increase our sample size, thus facilitating the realization of a specific probability distribution function, pdf, to characterize our path loss data, Z_i. With respect to obtaining a specific pdf, we have conducted a goodness-of-fit test to selected normalized data transformed by 40 log d, that is, according to the J&B approach, and also to the data which we transformed using our statistical approach. The results are shown in Table III. In the J&B case, the hypothesis that the path loss information can be characterized by the log-normal distribution was rejected at the 5% level, whereas for our α_1 approach H_o was accepted. It is noted that a random

-503-

TABLE III: SELECTED GOODNESS-OF-FIT DECISIONS FOR THE ALPHA 1, AND JANSKY AND BAILEY LOGARITHMIC DATA AT THE α = .05 LEVEL OF SIGNIFICANCE (TWO TAILED) FOR 3"/6" RAINFALL CRITERION

RADIAL A FREQ (MHz)	T (feet)	P	R (feet)	DISTANCE (miles)	DECISIONS BASED UPON THE NULL HYPOTHESIS THAT VARIATIONS DUE TO CHANCE FLUCTUATIONS DO NOT DEPART FROM NORMALITY			
					ALPHA 1 DATA		JANSKY & BAILEY DATA	
					WET	DRY	WET	DRY
6	40	V	23	0.2-2.0	Accept	Accept	Reject	Reject
			45		Accept	Accept	Reject	Reject
			76		Accept	Accept	Reject	Reject
50	40	V	20	0.2-2.0	Accept	Accept	Reject	Reject
			42		Accept	Accept	Reject	Reject
			79		Accept	Accept	Reject	Reject
100	40	V	20	0.2-2.0	Reject	Reject	Reject	Reject
			42		Reject	Accept	Reject	Reject
			79		Accept	Accept	Reject	Reject
250	40	V	40	0.2-2.0	Accept	Accept	Reject	Reject
400	40	V	42	0.2-2.0	Accept	Reject	Reject	Reject
6	40	H	23	0.2-2.0	Accept	Accept	Reject	Reject
			45		Accept	Accept	Reject	Reject
			76		Accept	Accept	Reject	Reject
50	40	H	20	0.2-2.0	Accept	Accept	Reject	Reject
			42		Accept	Accept	Reject	Reject
			79		Accept	Accept	Reject	Reject
100	40	H	20	0.2-2.0	Reject	Reject	Reject	Reject
			42		Accept	Accept	Reject	Reject
			79		Accept	Accept	Reject	Reject
250	40	H	42	0.2-2.0	Accept	Reject	Reject	Reject
400	40	H	42	0.2-2.0	Accept	Accept	Reject	Reject

variable is said to be log-normally distributed if the pdf of the logarithm of the variable is normal.

The rejection of the hypothesis that the 40 log d transformed information did not follow the log-normal distribution is not really surprising, since this undesirable normalizing factor was used. We should mention that in the selected configurations we investigated using our distance normalizing approach, we accepted the hypothesis that the information can be characterized by the log-normal distribution. However, this decision may also be somewhat shaky due to a sample size of less than 20 observations.

In view of the above comments, we have no alternative but to increase the sample size so as to be more elegant in our presentation of the analysis. It is extremely important to establish whether or not there is a significant difference between the wet and dry propagation loss, keeping in mind that the original experiment was not intended for this purpose. If there is no sig- nificant difference in the propagation loss between the two classifications, then one can combine the information, thereby increasing the sample size, enabling stronger conclusions to be obtained regarding the behavior of propagation loss under different configurations. If the propagation loss during wet seasons is not the same as that during dry seasons, we may not want to combine our wet and dry information to analyze the path loss data from a non-parametric approach, (i.e. a distribution free analysis). At this point, we utilized non-parametric statistical analysis to answer the question: "Is there any significant difference between the mean propagation loss for the wet classification and the mean propagation loss for the dry classification?", which is important primarily for the systematic presenta- tion of our proposed statistical modeling.

Therefore, in order to analyze data and gain maximum information, hypotheses were formulated under the following assumptions:

(1) The population distributions were "normal" (gaussian) and that the individual observations are independent.

(2) The observations are identically distributed, the nature of which is <u>not assumed</u> normal.

Under the first assumption, the parametric small sample "t" test and associated confidence intervals were used, Brownlee (5).

Since there was an extensive amount of data under various classifications, simpler non-parametric (distribution free) methods were employed under the second assumption, namely the Mann-Whitney-Wilcoxon rank test, along with the associated confidence intervals on the <u>medians</u> (6). It is to be noted that if the underlying distribution is <u>indeed normal</u>, the non-parametric tests may be somewhat poorer, in that the probability of error of the second kind, β , is somewhat larger for a given level of significance, α , and population size, n.

To test the classifications of wet vs dry, that is, whether or not the difference between wet and dry data was significant, the following hypotheses were formulated for ten selected parameter configurations (see Table IVa, IVb):

$$H_o: \mu_{wet} = \mu_{dry} \quad vs. \quad H_1: \mu_{wet} \neq \mu_{dry}$$

where the respective estimates of the true parameters μ_{wet} and μ_{dry} were:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{and} \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i \text{ for each configuration,}$$

TABLE IVa: SELECTED NON-PARAMETRIC DECISIONS ON THE SIGNIFICANT DIFFERENCE BETWEEN THE PROPAGATION LOSS FOR WET PLUS DRY CLIMATE AT THE α = 0.05 LEVEL OF SIGNIFICANCE USING ALPHA 1 DATA, VERTICAL POLARIZATION

RADIAL A FREQ (MHz)	TH (Ft)	P	RH (Ft)	DISTANCE (Mi)	DECISIONS BASED UPON NULL HYPOTHESIS WET = DRY	
					3"-6" RAINFALL CRITERION	SEASONAL RAINFALL CRITERION
6	40	V	23 45 76	0.2-2.0	Accept Accept Accept	Accept Reject Accept
50	40	V	20 42 79	0.2-2.0	Accept Accept Accept	Reject Accept Reject
100	40	V	20 42 79	0.2-2.0	Reject Reject Reject	Accept Accept Accept
250	40	V	40	0.2-2.0	Reject	-
400	40	V	40	0.2-2.0	Accept	-

TABLE IVb: SELECTED NON-PARAMETRIC DECISIONS ON THE SIGNIFICANT DIFFERENCE BETWEEN THE PROPAGATION LOSS FOR WET PLUS DRY CLIMATE AT THE $\alpha = 0.05$ LEVEL OF SIGNIFICANCE USING ALPHA 1 DATA, HORIZONTAL POLARIZATION

RADIAL A FREQ (MHz)	TH (Ft)	P	RH (Ft)	DISTANCE (Mi)	DECISIONS BASED UPON NULL HYPOTHESIS WET = DRY	
					3"-6" RAINFALL CRITERION	SEASONAL RAINFALL CRITERION
6	40	H	23	0.2-2.0	Accept	Accept
			45		Accept	Accept
			76		Accept	Accept
50	40	H	20	0.2-2.0	Accept	Accept
			42		Accept	Accept
			79		Accept	Accept
100	40	H	20	0.2-2.0	Accept	Accept
			42		Accept	Accept
			79		Accept	Accept
250	40	H	40	0.2-2.0	Accept	-
400	40	H	40	0.2-2.0	Accept	-

where: X_i = propagation loss for wet conditions,

Y_i = propagation loss for dry conditions.

Our decisions were based upon the three classifications of rainfall listed below:

(i) 3" rainfall criterion,

(ii) 6" rainfall criterion,

(iii) seasonal rainfall criterion.

The "t" statistic used at the $\alpha = .05$ level of significance is given by:

$$t = \frac{\overline{X} - \overline{Y}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\overline{X} - \overline{Y}}{S_p \sqrt{2/n}} \text{ , where } n_1 = n_2 = n, \qquad (4)$$

and:

$$S_p = \sqrt{\frac{(n_1-1)S_x^{\,2} + (n_2-1)S_y^{\,2}}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{S_x^{\,2} + S_y^{\,2}}{2}} \text{ ,}$$

where: $S_x^{\,2} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$, the estimate of true variance for the wet classification,

and: $S_y^{\,2} = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2$, estimate of the true variance for the dry classification.

The relationship which describes the probability of accepting H_o is given by:

$$\Pr \left\{ t_{(n-1);\, \alpha/2} \leq t \leq t_{(n-1);\, 1-\alpha/2} \right\} = 1 - \alpha . \qquad (5)$$

The associated confidence interval which affords guidance as to how closely the sample values $\bar{X}$, $\bar{Y}$ estimate the true parameters μ_{wet} and μ_{dry}, given H_o is true, is:

$$\left\{ (\bar{X}-\bar{Y}) \pm t_{(n-1)\ \alpha/2}\ S_p \sqrt{\frac{2}{n}} \right\} \tag{6}$$

Now, we shall give a brief description of the <u>non-parametric</u> analysis (6,7) which was performed on the propagation loss data for selected parameter configurations. Since the data consists of independent random samples, let $Z_1, Z_2, \ldots Z_n$ denote the random sample of size n from the <u>wet classification data</u>:

$$Z_i = X_i - \alpha_i \log d_i , \tag{7}$$

and let $Z_1{}^*, Z_2{}^*, \ldots Z_m{}^*$ denote the random sample of size m from the <u>dry classification data</u>:

$$Z_j{}^* = X_i{}^* - \alpha_i \log d_i , \tag{8}$$

the variables of which were previously defined.

If there is a difference between the population distribution functions of the $Z_j{}^*$ and Z_i, that difference will be in the location of the distribution. Therefore, we again test for the significant difference in the true states of nature. The following hypothesis was formulated for the selected parameter configurations of the path loss data:

$$H_o : \mu_{wet} = \mu_{dry} \quad vs. \quad H_1 : \mu_{wet} \neq \mu_{dry}.$$

First we assign ranks 1, 2, 3, ..., n + m to the <u>combined samples.</u> That is, assign rank 1 to the samllest, rank 2 to the next larger and so on. Let $R(Z_i)$ and $R(Z_j{}^*)$ denote the rank assigned to the Z_i and $Z_j{}^*$ for all i and j.

Next, the Mann-Whitney test statistic, T, is computed using the following
equation:

$$T = S - \frac{n(n+1)}{2} , \qquad (9)$$

where:
$$S = \sum_{i=1}^{n} R(Z_i),$$

(if $S^* = \sum_{j=1}^{n} R(Z_j^*)$) is used, the same result can be expected with respect

to accepting H_o).

The decision rule for the two tailed test is to accept H_o at the level of
significance $\alpha = .05$ if:

$$W_{\alpha/2} < T < W_{1-\alpha/2}$$

where $W_{\alpha/2}$ and $W_{1-\alpha/2}$ are the upper and lower quantiles of the table of the
Mann-Whitney test statistic.

The distribution free <u>confidence interval</u> estimate procedure was for an
unknown population median, M. This is related to the binomial probability

$$\sum_{k=k_{\alpha/2}}^{N} \binom{N}{k} 0.5^N \leq \alpha/2 \quad \text{for the upper bound}, \qquad (10)$$

and:

$$\sum_{k=0}^{k'_{\alpha/2}} \binom{N}{k} 0.5^N \leq \alpha/2 \quad \text{for the lower bound}, \qquad (11)$$

where:

$$k'_{\alpha/2} = N - k_{\alpha/2} ,$$

N = population size,

$k_{\alpha/2}$ = C from the binomial table which corresponds
to $P(c, n, p) = \alpha/2$.

The interval is: $Z'_{\alpha/2} + 1 < M < Z_{\alpha/2}$

from $$P(X'_{\alpha/2} + 1 < M < X_{\alpha/2}) = 1 - \alpha \ , \qquad (12)$$

which is based upon: $$\Pr\left\{Z_{(r)} < M < Z_{(s)}\right\} = \Pr\left\{r \leq K \leq s - 1\right\} \qquad (13)$$

That is, we need K positive numbers among N differences: $Z_i - M$, $i = 1, 2,$ ....,N where the Z_i's are ordered. r and s are the positions of Z_r and Z_s of the ordered Z_i's, or $r = k'_{\alpha/2}$, and $s = k_{\alpha/2} + 1$. Thus, in an ordered array of observations Z_i, the confidence interval end-points are those numbers which are in the $(k'_{\alpha/2} + 1)^{st}$ positions from either end.

The non-parametric decisions on the significant differences between the propagation losses Z_i for wet and dry classifications at the $\alpha = .05$ level of significance are shown in Tables IVa and IVb for selected parameter configurations. The null hypothesis, H_o, was accepted without doubt for all tests where the antennas were horizontally polarized. Both the 3"/6" and seasonal criteria were equally accepted. However, for the vertical polarization, the 3"/6" criteria show that for the chosen operating frequencies greater than 100 MHz, the differences between wet and dry were significant, that is, H_o was rejected. However, since the path loss coefficients α_i for wet and dry classification were different above 100 MHz (as compared to $2 < f < 100$ MHz) for horizontal polarization, and also for selected spot checks of the vertical polarization data, it doesn't necessarily imply that H_o should be rejected. In general, one can conclude that for frequencies less than 100 MHz, the wet and dry criteria do not greatly effect the path loss, Z_i. For vertical polarization (Table IVa), the seasonal rainfall criteria is questionable with respect to the actually measured 3"/6" criterion.

A point estimate of a parameter, i.e., its mean value, is a random variable distributed in some way around the true state of nature. In the analysis, no indication is provided as to how closely the sample means estimate the true state of nature. Therefore, to afford some guidance we computed intervals which we are confident will actually include the true value of the parameters. Tables Va, Vb, VIa, VIb, show the selected 95% confidence bounds in which we may conclude that the true parameters is contained, providing H_o is true. Tables Va, b, show the non-parametric 95% confidence intervals for wet and dry classifications separately, (and some parametric spot checks) for each respective rainfall criterion. It can be seen that generally they are narrow for all parameter configurations, the smallest being 2.8 dB at (50, 40, H, 79) for the wet seasonal rainfall classification. The largest is 18.8 dB at (400, 40, V, 42) for 3"/6" dry classification. Generally, they were less than 8 dB overall. Tables VIa and VIb show the 95% confidence intervals for the combined wet and dry classification of data. As expected, they are generally narrower than wet and dry taken separately because of the small number of samples involved, the largest being 19.0 dB at (100, 40, V, 42) for the 3"/6" rainfall criterion. The smallest is 2.1 dB at (50, 40, H, 20) for the seasonal criterion. Parametric spot checks of the confidence intervals for the sample means, $\bar{Z}_i$, are also included for comparison. From the overall groupings, the majority of intervals were less than 5 dB. The intervals clearly show that the sample estimates, $\bar{Z}_i$, are reliable, and that because of their dependence on α_i, the estimates $\hat{\alpha}_i$ as shown (see Tables Ia and Ib) are reliable propagation loss coefficients.

TABLE Va: SELECTED NON-PARAMETRIC CONFIDENCE INTERVALS ON THE PROPAGATION LOSS FOR WET AND DRY CLIMATE AT THE **α = 0.05** LEVEL OF SIGNIFICANCE USING ALPHA 1 DATA, VERTICAL POLARIZATION

RADIAL FREQ (MHz)	TH (Ft)	RH (Ft)	DISTANCE (Mi)	3"-6" RAINFALL CRITERION				SEASONAL RAINFALL CRITERION			
				WET		DRY		WET		DRY	
				LB	UB	LB	UB	LB	UB	LB	UB
6	40	23	0.2-2.0	93.3 (93.2)	97.6 (98.6)	90.6 (91.3)	96.8 (95.7)	92.0 (92.6)	99.7 (98.1)	89.0 (90.9)	96.1 (100.2)
		45		91.9 (92.6)	98.1 (99.5)	89.6 (90.7)	96.8 (95.7)	92.0	100.0	88.3	96.1
		76		90.3	93.6	89.0	95.3	89.0	96.9	87.6	93.4
50	40	20	0.2-2.0	103.2	113.2	103.5	111.5	101.0	112.7	105.2	114.1
		42		95.5	104.0	95.6	102.8	94.0	99.2	95.7	104.7
		79		89.0 (84.7)	96.0 (95.5)	87.9 (87.2)	94.7 (96.6)	87.4 (85.1)	91.0 (93.2)	90.6 (90.3)	96.0 (97.0)
100	40	20	0.2-2.0	123.0*	135.0*	108.2**	122.0**	120.0**	124.3**	108.4*	122.7*
		42		113.0*	126.0*	101.6**	111.8**	107.0**	120.3**	100.0	112.6**
		79		101.0*	111.0*	96.0**	98.5**	95.6**	105.2**	96.0*	99.8*
250	40	42	0.2-2.0	129.9	127.3	112.1	123.0				
		79		108.3	119.4	100.4	115.9				
400	40	42	0.2-2.0	-	-	110.2	129.0				
		79		114.0	117.5	115.3	118.3				

* - α = .06
** - α = .07

Values in ( ) are parametric spot checks.

TABLE Vb: SELECTED NON-PARAMETRIC CONFIDENCE INTERVALS ON THE PROPAGATION LOSS FOR WET AND DRY CLIMATE
<u>AT THE α = 0.05 LEVEL OF SIGNIFICANCE USING ALPHA 1 DATA, HORIZONTAL POLARIZATION</u>

RADIAL A FREQ (MHz)	TH (Ft)	RH (Ft)	DISTANCE (Mi)	3"-6" RAINFALL CRITERION				SEASONAL RAINFALL CRITERION			
				WET		DRY		WET		DRY	
				LB	UB	LB	UB	LB	UB	LB	UB
6	40	23	0.2-2.0	87.3	96.3	86.7	92.7	86.9**	95.9**	87.9	92.5
		45		81.4	90.4	81.5	87.6	81.2**	90.2**	83.1	87.1
				(82.9)	(89.8)	(81.7)	(88.5)	(76.1)	(95.5)	(82.9)	(87.5)
		76		77.5	86.5	78.5	81.4	77.2**	86.2**	78.9	81.8
50	40	20	0.2-2.0	96.6	100.9	95.0	107.7	98.0*	100.9*	95.8***	99.6***
		42		93.7	101.3	92.3	104.5	94.4*	101.9*	93.0***	99.0***
		79		81.0	93.0	88.0	97.6	90.2*	93.0*	88.6***	92.0***
100	40	20	0.2-2.0	111.6*	121.3*	109.0	118.5	110.0	124.4	109.0	119.9
		42		106.0*	113.6*	104.0	110.9	105.8	117.7	104.0	114.1
				(106.7)	(113.8)	(99.3)	(115.4)				
		79		99.0*	104.5*	95.0	100.9	95.0	104.2	95.0	101.3
250	40	42	0.2-2.0	115.5	124.5	110.2	127.7				
		79		103.0	106.6	101.0	116.2				
400	40	42	0.2-2.0	119.5	127.0	122.5	129.0				
		79		115.0	121.0	108.9	120.5				

* - α = 0.07
** - α = 0.12
*** - α = 0.06

Values in ( ) are parametric spot checks.

TABLE VIa: SELECTED NON-PARAMETRIC CONFIDENCE INTERVALS ON THE PROPAGATION LOSS FOR WET PLUS DRY CLIMATE AT THE α = 0.05 LEVEL OF SIGNIFICANCE USING ALPHA 1 DATA, VERTICAL POLARIZATION

RADIAL A FREQ (MHz)	TH (Ft)	RH (Ft)	DISTANCE (Mi)	3"-6" RAINFALL CRITERION WET + DRY		SEASONAL RAINFALL CRITERION WET + DRY	
				LB	UB	LB	UB
6	40	23	0.2-2.0	91.8 (92.2)*	96.5 (96.1)*	92.0 (92.1)*	96.1 (96.3)*
		45		91.8 (91.7)*	96.8 (96.3)*	91.7	96.1
		76		89.3	93.5	89.3	93.0
50	40	20	0.2-2.0	103.5	111.2	104.1	112.7
		42		95.6	102.5	95.7	101.8
		79		89.0 (88.0)*	94.0 (93.8)*	88.7 (90.5)*	94.8 (93.1)*
100	40	20	0.2-2.0	108.2	128.1	108.4	123.3
		42		101.6	120.6	100.0	116.3
		79		96.0	106.1	95.6	102.2
250	40	42	0.2-2.0	114.6	123.0		
400	40	42	0.2-2.0	110.2	125.9		

* - Values in ( ) are parametric spot checks.

TABLE VIb: SELECTED NON-PARAMETRIC CONFIDENCE INTERVALS OF THE PROPAGATION LOSS FOR WET PLUS DRY CLIMATE AT THE α = 0.05 LEVEL OF SIGNIFICANCE USING ALPHA 1 DATA, HORIZONTAL POLARIZATION

RADIAL A FREQ (MHz)	TH (Ft)	RH (Ft)	DISTANCE (Mi)	3"-6" RAINFALL CRITERION WET + DRY		SEASONAL RAINFALL CRITERION WET + DRY	
				LB	UB	LB	UB
6	40	23	0.2-2.0	87.3	92.7	87.9	92.3
		45		82.4 (83.7)*	87.3 (87.7)*	83.0 (83.4)*	87.0 (87.4)*
		76		78.5	82.4	78.9	81.7
50	40	20	0.2-2.0	96.5	100.7	97.9	100.0
		42		93.7	101.3	93.4	99.0
		79		88.0	92.6	88.6	92.9
100	40	20	0.2-2.0	111.3	117.3	111.6 (112.3)*	117.4 (117.6)*
		42		106.0 (106.7)*	111.2 (111.3)*	106.0	112.0
		79		97.4	101.3	98.8	101.4
250	40	42	0.2-2.0	112.0	121.7		
400	40	42	0.2-2.0	122.5	127.3		

* - Values in ( ) are parametric values.

This, of course, indicates the need for providing a model, α, to define Z_1, over the 6 MHz through 400 MHz range rather than accept $^1\alpha = 40$ as proposed by Jansky and Bailey.

4. SUMMARY. In Section 2, we gave a brief discussion of the manner in which the SEACORE data was presented in the Final Report by Jansky and Bailey. We have obtained enough information from the above statistical analysis of the SEACORE data to make the following conclusions:

(1) It is evident that the data should be given a more sophisticated statistical analysis for firm decisions on the various questions raised with respect to behavior of the propagation loss as a function of the various independent variables (i.e., antenna heights, polarization, frequency, distance.)

(2) We have obtained enough evidence to show that the deterministic approach used to normalize the data to a common distance utilizing equation (1) is not acceptable to relate the propagation loss to a common distance.

(3) A statistical approach to estimating an α so as to minimize the variance of the estimate, using the logarithmic data, has been presented (see equation 2) and evaluated for specific parameter configurations of the SEACORE Project. It is clear that in normalizing the path loss data, using this technique, significantly different results are obtained which effect the statistical decisions one needs to make with respect to the behavior of propagation loss in a jungle environment.

(4) It was shown that due to the small amount of information available we cannot accept a specific distribution, such as the gaussian or lognormal, to completely characterize the physical situation. Thus, any statistical analysis which one may perform in the subject area should be done through non-parametric statistical analysis, that is, using distribution free statistics.

(5) It was also concluded that in some cases there is a significant difference in the mean propagation loss between the wet and dry classifications. This may be due to path differences rather than rainfall influence, and should be examined.

(6) To adequately answer the original objectives of the SEACORE program, the importance of the path loss data presented in the reports by J & B cannot be overlooked. Our preliminary investigations employing statistical analysis have answered a number of questions and indicate the importance of a thorough statistical analysis of the propagation loss data. It should also be mentioned that there is very little work, if any, which has been done in the subject area from a sophisticated statistical analysis approach. We feel that such an approach is much more realistic to the problem at hand than a deterministic investigation, and the preliminary findings certainly justify this point of view.

(7) Based on our findings, the following guidance should be considered with respect to the complete statistical modeling of the SEACORE path loss data:

-518-

(a) In the statistical analysis of the SEACORE data a very basic question must be answered. That is, should one perform the statistical analysis on the data in dB, (i.e. having it logarithmically transformed) or should the anti-log data be used?

(b) A through classification of the data with respect to wet-dry conditions should be made. The problem of propagation loss should be investigated under three classification categories:

(i) daily basis (if present data permits),
(ii) monthly basis,
(iii) seasonal basis.

Furthermore, the effect of path differences should be determined if possible.

(c) Having chosen the proper method for normalizing propagation loss to a common distance we need to develop a super alpha, i.e., α^*, for specific sets of parameter configurations that will provide a realistic approach which can be used to predict path loss as a function of distance. Since there is a significant difference among the α_1's measured for each parameter configuration, one should not treat the estimate of the α's as a deterministic parameter but rather as a random variable. To this effect we need to formulate statistical techniques through the empirical Bayes approach, i.e., to group these α's into a common one. For a certain group of frequencies, transmitter and receiver antenna height, and polarization, we should have a statistical estimate of an α^* which is made up of a group of α's which can be easily applied to a physical situation for communication which we believe will be of significant importance in obtaining this α^* which will give a realistic characterization of the sequence of the α's as random variables.

(d) Having path loss as the main variable, we need to classify the contributing variables, that is, independent variables such as transmitter antenna height, receiver antenna height, distance, polarization and frequency, according to their importance as contributing factors. This investigation can be done through multiple correlation analysis. It's importance lies in the fact that if a certain independent variable, such as changing the antenna height, does not contribute significantly to a change in propagation loss, then we should not consider it as one of the important variables in the statistical modeling. In other words, we should be concentrating on the independent variables which contribute most to the dependent variables. Such a classification of the independent random variables will be extremely helpful in accomplishing the succeeding recommendations.

(e) Having classified the importance of the random variables as contributing factors to the propagation loss, it is recommended that a non-linear regression model be developed. Once a non-linear regression model has been formulated, with path loss as the main objective, which is a function of transmitter antenna height, receiver antenna height, distance, polarization, and frequency, one can specify an acceptable propagation loss

-519-

and obtain the proper combinations of the independent variables required
to attain this loss. An important factor in this recommendation is that
if you are not willing to accept more than, say, 150 dB for path loss at
a specific frequency and distance , what combination of antenna heights
will not violate the proposed specification. One can proceed in formula-
ting such a non-linear regression model through an elimination procedure,
that is, consider a model which will consist, first, of the dependent
variable being a function of the antenna heights. Thus, with a specific
frequency and distance, the possible combinations of antenna height will
be determined so that we can maintain a specified propagation loss. Secondly,
with this approach one can increase the size of the model by having the
dependent variable as a function of the antenna heights and distance.

(f) In our preliminary investigation of the SEACORE data, our
statistical analysis was restricted primarily to distances between .2-2.0
miles. It is also recommended that the above recommendations be considered
for longer distances, that is, 2.0 - 15.0 miles.

(g) For selected sets of the control variables, taking into
consideration the above decisions, confidence intervals should be obtained
for the mean path loss (true state of nature) on the basis of the experimental
evidence. Specifically, confidence intervals should be obtained for the
following cases:

(i)   on the path loss parameter of radial A-wet
(ii)   on the loss of radial A-dry
(iii)   on the loss of radial A-wet + radial A-dry
(iv)   on the loss of radial B wet
(v)   on the loss of radial B dry
(vi)   on the loss of radial B wet + radial B dry
(vii)   on the loss of radial A wet + radial B wet
(viii) on the loss of radial A dry + radial B dry

It may not be necessary to calculate all of the above confidence intervals
if we accept certain hypotheses with respect to the behavior of wet, dry,
radial A, and radial B data.

In summary, the preliminary findings of our statistical analysis are
quite evident, and as a result, the above recommendations constitute some
of the essential elements for the final aspects of the SEACORE Project.
The value of this data, which has been collected and descriptively presented
at a great cost to the U.S. Government, should be fully utilized to attain
the answers to the questions posed above. The results of sophisticated
analysis will make significant contributions in aiding communications
engineers to determine transmission reliability and, ultimately, better
communication systems.

REFERENCES

1. Tropical Propagation Research, Final Report, Vol I, Jansky and Bailey Research Engineering Department, Alexandria, Virginia.

2. Tropical Propagation Research, Final Report, Vol II, Jansky and Bailey Research Engineering Department, Alexandria, Virginia.

3. Tamir, T., On Radio Wave Propagation in Forest Environments, IEEE Trans. AP-15 (1967), 806-817.

4. Dence, D., and T. Tamir, Radio Loss of Lateral Waves in Forest Environments, Radio Science, 4 (1969), 307-318.

5. Brownlee, K.A., Statistical Theory and Methodology, J. Wiley, 1960.

6. Conover, W. J., Practical Non-Parametric Statistics, J. Wiley, 1971.

7. Natrella, M. G., Experimental Statistics, National Bureau of Standards, U. S. Department of Commerce, 1963.

GRUBBS' ESTIMATORS TO DATE

Clifford J. Maloney
Bureau of Biologics
Food and Drug Administration
Rockville, Maryland

I. <u>Introduction</u>: Two at least partially compensating trends affect the Army research investigator's effectiveness in the prosecution of his endeavors. On the one hand a steady stream of new and improved procedures, techniques, principles, laws, and devices are made available and explained in an ever burgeoning technical literature--with a consequent diversion of "productive" effort into that devoted to locating and assimilating the innovations. This latter, in turn, is greatly ameliorated by simplifications of procedures, theories, and laws making them easier to acquire, to appreciate, and to retain in memory and, often, of wider, more accurate scope. One further device, in use since the earliest times and widely appreciated in mathematical circles, though, it seems, less so by statisticians, is the device of devising a classification of the corpus of findings on one class of topics. The periodic table in chemistry, and the Linnean classification in systematics are scarcely unknown to any literate person. Possibly the best known system in mathematics is Klein's classification of geometries, but many others are known, including Wedderburn's classification of associative algebras and Post's classification of logics.

The subject of this paper is the evaluation of the precision of two instruments, techniques, or procedures subject to two conditions. First, both can be simultaneously applied, so that the "true" quantity, though

unknown, is the same for both. Second, that only one reading for each instrument or procedure is available for one fixed value of the underlying true quantity. When another pair of readings is made, the true value has shifted—and by an unknown amount. Of course, in practice it will often be desired to simultaneously compare three or more instruments or techniques and a number of the papers in the bibliography treat these cases. The present paper is less interested in extending the theory or widening the field of its applicability than in providing an insight into the exact nature of the phenomenon by means of a geometric interpretation.

II. Earlier Work: Throughout our discussion it will be assumed that we are dealing exclusively with two (possibly) correlated normally distributed random variables with unspecified, hence possibly different means. All the points we wish to make are involved in this model. The pioneer paper in the field is that of Grubbs [16], though an earlier study by Pearson had been forgotten [49]. The same problem was encountered by Thompson, who has dealt with it in a series of papers [21], [46-48]. A further paper is in preparation by him for presentation at the Nineteenth Army Design Conference. The problem arose more recently in a biological context at the Bureau of Biologics of the Food and Drug Administration. While Grubbs' paper gives estimates and moments of the distribution, he does not directly present tests of significance. This latter problem was the subject of a clinical paper at the Twelfth Army Design Conference. The possibility of employing the test of Morgan [35] and Pitman [36] was

-524-

suggested by members of the panel of experts.[*]   This suggestion proved fruitful and a formal proof of the applicability of the method was later published [Maloney and Rastogi, 1970]. The appearance of this paper was followed by a series of elaborations and extensions [Jaech, 23; Makowski, 28; Shukla, 41 and 42].

While the work of Morgan and Pitman was available at the time Grubbs did his work, the relevance of their results to his problem only became apparent with the publication of [29]. That line of development stems originally from Bose [4]. A simpler approach to Bose's result was derived approximately by Finney [14] and exactly by Morgan and Pitman.

Another line of research close to the principal interest of this paper arose somewhat diffusely prior to 1940, but had become explicit by 1950 [13]. This is the development of a test for equal variances in the several categories of an analysis of variance. Contacts of that topic with that of this paper seem, however, to have reached print only in Shukla [42]. Finally in Tukey [49] the problem of Grubbs is related to that of regression with errors in both variates.

Over the years, then, disparate statistical problems whose simi- larities escaped notice for as much as one year to decades, were eventually seen to have such relations as would throw light on possible approaches to the others. These are: Grubbs' problem, the Morgan-Pitman

---

[*] The panelists were: Bernard Greenberg, Frank Grubbs, William Kruskal, Henry Lucas, Henry Mann, and Albert Parks.

test (Bose's problem), regression with error in both variates (structural relations), heterogeneous error variances in analysis of variance, and factor analysis. It would seem appropriate to add the Behrens-Fisher problem to this list.

The several recent papers treating extensions and special cases [23, 28, 33, 41, 42] suggest the desirability of attempting some sort of classification of the possibilities; so that new cases are revealed and/or it can be determined when all cases have been treated. The next section introduces a geometric model intended to serve this purpose.

III. Geometric Model: It is well known that isopleths of constant probability in the case of the normal distribution are coaxial ellipses of constant eccentricity. Hence, the characteristic of a particular normal frequency distribution can be exhibited graphically by choosing just one of these ellipses. The obvious candidate is the ellipse with semi-major axis σ_a and semi-minor axis σ_b. Such an ellipse is illustrated in figure 1. The location of the center of the distribution is at the mean X, μ_x, and mean of Y, μ_y. The orientation of the ellipse depends on the product term in its equation. The ellipse becomes a circle if $\sigma_a = \sigma_b$ and a true ellipse otherwise. The dotted line through the origin in figure 1 does not pass through the center of the ellipse so that in a large sample the Behrens-Fisher test should be significant.

While attention is focused on figure 1, we may call attention to a trivial relation, but one which nevertheless will be helpful in later

discussions. In figure 1 a bivariate distribution is assumed. But suppose only one measure were available and it is just repeated as X and Y. Then, whatever its value, it would fall exactly on the dotted line in the figure through the origin. Such a line bisecting a quadrant of the coordinate axes can be thought of as a degenerate (singular) bivariate distribution. In case the absolute values of the variates are equal but they differ in sign, the singular distribution would be the line bisecting quadrants II and IV. This, of course, holds whatever the numerical value of the joint mean. Again, this linearity follows if X (or Y) is any linear function of Y (or X). In these cases the correlation coefficient is equal to unity.

The algebraic equation of the general ellipse is a quadratic with all terms present and with a positive discriminant. To reduce the equation to normal form, (1) the center of the ellipse, C in the figure, is translated to the origin of axes, O, and (2) the ellipse is rotated through the angle θ so that the major and minor axes of the ellipse coincide with the coordinate axes. It is not normally appropriate to do so, but we shall wish to apply one more transformation, a rescaling of one axis of the coordinate axes, thus reducing the ellipse to a circle, in which case $\sigma_a^2 = \sigma_b^2$.

All analytic geometry textbooks demonstrate that these successive transformations are achieved by simple formulas provided the coefficients in the equation of the ellipse are known. The statistical problem arises

-527-

precisely because these are not known. For our model to be fully
serviceable it should now be possible to forget all about the statistical
origin of the problem and think wholly in terms of devising a transformation
or series of transformations which will take an ellipse in general position
and reduce it to a circle centered at the origin of coordinates. All of
the constants in the equation of the general ellipse are nuisance parame-
ters unless we can effect an appropriate transformation irrespective of
their actual value. By fixing the nuisance parameters we get the various
special statistical tests of significance in the literature.

IV. Digression: At the outset it is appropriate to deal with a
persistent confusion concerning the correlation coefficient in a bivariate
normal distribution. Curiously the invention of the coefficient of
correlation initiated the burgeoning application of statistics outside of
the narrow field of the reduction of observations which began just before
1890; yet a misunderstanding of its signification still persists; against
which protests are felt necessary from time to time.

Quoting from the excellent textbook of Professor Allen Edwards
(page 144) " . . . the numerical value of the correlation coefficient
is related to the scatter of the plotted points about the line representing
their trend." As is abundantly clear from the subsequent discussion, this
statement is not wrong; it is misleading. The student visualizes that, as
the correlation coefficient increases from a value of zero to unity, the
"scatter of the plotted points" shrinks from equality in all directions to

-528-

a concentration along a mathematical line; though again, a sufficiently close reading of Edwards' actual text establishes that he makes no such assertions. Indeed, Edwards (and all other writers) is only saying that segments cut by ellipse are shorter than orthogonal projections of the whole onto either axis, segment $Y_1 Y_2$ in figure 1. Again, in every bivariate normal distribution as in every ellipse, there is a transformation, a rotation, which will free the equation of the product term; there is a set of orthogonal coordinate axes, those parallel to the principal axes of the ellipse, such that the correlation in the new axes is zero. The actual situation is clearer from a sentence in Yule and Kendall (14th edition, page 241, §10.7) " . . . a normal surface for two correlated variables may be regarded merely as a certain surface for which r is zero turned around through some angle . . . ." Their equation 10.10, page 242, gives this angle as

$$\tan 2\theta = 2 r \sigma_x \sigma_y / \sigma_x^2 - \sigma_y^2)$$

This rotation changes the coordinate axes to a pair of orthogonal axes parallel to the principal axes of the ellipse.

In other words, the correlation coefficient has nothing to do with the scatter of the points per se, but only with the orientation of the principal axes of the normal distribution with respect to the coordinate axes. A rotation of axes that eliminates the product term from the equation of a contour ellipse of the distribution removes the correlation coefficient and vice versa.

What does describe the relative scatter of the points, the "fatness" of the bivariate normal distribution, is the eccentricity of the ellipse; just as in analytic geometry.

This situation invites an aside on an aspect of the goal of achieving a rational model of a body of knowledge. The mathematician's milieu impels him to view the coordinate system as arbitrary and without essential significance since he is interested in the internal and/or mutual relations of his figures and not in their external relations with outside elements, which latter, in effect, dictate the coordinate axes. But those who apply rather than develop mathematics are faced continually with exactly this problem. Failure to appreciate adequately this distinction underlies much discussion concerning the application of theoretical results in nature.

V. <u>Classification of Bivariate Tests</u>: Exposition will be easier if we start with the simplest cases and progress to the more general rather than the other way around.

A. <u>Quadrant bisector</u>. This is the dotted 45° quadrant bisector in figure 1. Of course, the case is degenerate and would apply only if the column of x's reproduced the column of y's. However, when compounded with the later cases, we get geometric models having considerable interest. Since the situation is essentially univariate, all estimation, tests of significance, and confidence limit problems of univariate theory apply without essential change.

B. A Circle Centered at the Origin: (Figure 2). This case, too, is degenerate in that the two variates are in effect pairs drawn independently from the same normal distribution with zero mean. It is also the null hypothesis for case E below.

C. A Circle Centered on the Quadrant Bisector: (not shown). A circle centered on a line of known slope through the origin is equivalent. Again, this situation is highly artificial, but our model if it is to be complete must cover all cases. It is the null hypothesis for case F below.

D. A Circle in General Position: (not shown). This is Student's problem. The question to be tested is: is the center of the circle on the quadrant bisector (or on a ray of specified slope)? This case is also the null hypothesis for case G below.

E. An Axial Ellipse Centered at the Origin: (Figure 3). By axial ellipse is meant one whose principal axes are parallel to the coordinate axes. While now commonplace, its solution by means of the F (actually z) test was one of Fisher's earliest successes.

F. An Axial Ellipse Centered on the Quadrant Bisector: (not shown). This case is included only for completeness.

G. An Axial Ellipse in General Position: (not shown). This shift brings us into the realm of unsolved problems, for this is the

-531-

Behrens-Fisher problem. We wish to test: is μ_x equal to μ_y? To do so, we test $t = \bar{x} - \bar{y}$. Now, if we arbitrarily put $u = \bar{x} + \bar{y}$, we have performed a rotation of axes, and the condition $t = o$ asks: is the ellipse center on the new axis? But in rotating the ellipse we have lost the alignment of the axes of the ellipse with the coordinate axes. Presumably this is the dilemma of the Behrens-Fisher problem. We $\underline{must}$ rotate to test $\mu_x = \mu_y$ but can't and retain the axial position of the ellipse. That a rotation has no such effect on a circle seems to be the geometric explanation of why Student's problem presents no such difficulty.

H. <u>A Symmetric Ellipse Centered at the Origin</u>: (Figure 4). By symmetric ellipse is meant that the major axis of the ellipse lies along the quadrant bisector, hence makes a 45° angle with the coordinate axes. By symmetry on the figure we see that the ellipse defines equal projections on the two axes, i.e., $\sigma_x^2 = \sigma_y^2$. While again a highly specialized case, the problem has in fact been dealt with by DeLury [11]. But what is much more interesting, the model is the basic model of analysis of variance.

If it is <u>known</u> that the two variances are equal, then there is no need to test for it. What remains uncertain, however, is whether the model is a circle or a symmetric ellipse, i.e., whether the correlation coefficient is different from zero (and, of course, to estimate it if it is). This was DeLury's problem. It is the converse of Grubbs' whenever bias is absent, since here we assume that $\sigma_x^2 = \sigma_y^2$, i.e.,

-532-

that the projections of the ellipse onto the two axes yield segments equal in length. Hence, the model is also the null hypothesis for Grubbs' problem where it is known that neither instrument is biased.

I. General Ellipse Centered at the Origin: (Figure 5). "General" means at an unspecified angle to the coordinate axes. Same as H except, that here the angle is unspecified. This is Grubbs' problem when bias is absent. In terms of our model then, his problem is: given that we have an ellipse (known not to be, or at least possibly not a circle) with center at the origin, is the angle of the major axis of the ellipse at an angle of 45° to the coordinate axes? This shows that the device used originally by Pitman (though, of course, with no claim to originality) is not just a trick and shows why we test for equality of variances by testing a correlation coefficient. It is because, if a 45° rotation angle transforms the ellipse into an axial ellipse, it was a symmetric ellipse before rotation.

J. Fairfield Smith's Problem: (no figure). Reference [29]. This problem is the same as Grubbs' (item I) except that here the scales on the two axes differ and by an unknown factor. Smith's solution was to solve the scaling problem first, by a non-parametric procedure, thus reducing his problem to that of Grubbs. The geometric model of this paper could possibly throw some light on other approaches to the problem. Also, it makes plausible why he went to a non-parametric method to adjust the scale since in analytic geometry it should come last.

K. **Symmetric Ellipse in General Position:** (no figure). This is
the general Model I analysis of variance situation. The mean of
the ellipse expresses the treatment effects. The orientation of
the major axis at a 45° angle to the coordinate axes expresses the
equality of variances within classes.

L. **Unspecified Orientation and Position of Ellipse.** (no figure).
This is Grubbs' general problem when biases are present: It <u>also</u>
involves analysis of variance where errors <u>are</u> allowed to vary
within arrays; to which a number of the papers in the bibliography
are devoted.

M. **General Ellipse with Unspecific Ratio Between Scales of the**
Two Coordinate Axes. (no figure). This is the problem of regression
with errors in both variables. There are, hence, 12 special cases
according as (a) the ellipse is a (1) circle, (2) an axial,
(3) symmetric, or (4) arbitrarily oriented true ellipse, and
(b) is centered (1) at the origin of axes, (2) on a quadrant bisector,
or (3) arbitrarily.

VI. **An Aside on Scales:** Smith's problem is reduced (by him) to Grubbs'
by specifying the slope of the line relating the two variables. It can be
viewed as a change of the scale of one coordinate axis. In the same way,
viewing as is often done, the variances as representing the <u>scaling</u> of
the two orthogonal projections of the ellipse parallel to the coordinate

axes, we can sort the several special cases on the basis of assumptions regarding the relations between the four scales. If a variance is known, this is equivalent to knowing the scale factor between that axis of coordinates and that of the ellipse parallel to it. If the slope of the regression line is known, that is equivalent to knowing the scale factor between the two axes of coordinates. If the ratio of the variances is known, that is equivalent to knowing that the ellipse is a circle. It seems that the difficult cases relate to an unknown relation within one of the two sets, coordinate and elliptic axes, and not between them. Can we summarize by saying: scale coupling uncertainty within coordinate sets (coordinate axes versus principal axes of the ellipse) prevents solutions; scale coupling uncertainty between sets presents (solvable) tasks. The result is possibly due to the fact that, despite appearances to the contrary, tests of means, like tests of correlation coefficients, are rotations of axes. This is a (the?) major difference between tests of independence in statistics and simplification of the general equation of an ellipse in analytic geometry. Tests of variances are likewise, except in the special case of independence, accomplished by rotations.

VII. _Final Remarks_: The preceding catalogue of cases illustrates one point. The proposed geometric model of the bivariate normal special cases-- a reduction of the general ellipse in orthogonal coordinates to a circle with center at the origin--passes one test; it provides a niche for a rich assortment of special cases which one or another author has found to be

-535-

related. The addition of the Behrens-Fisher problem to this list has not been previously located in the literature. Whether the model can subsume additional cases has not been seriously investigated, though none are immediately obvious.

As remarked earlier, no acquaintance with statistical theory is appropriate for any consideration internal to the model itself. The function of the model is to make the nature and the relationship of the several special cases intuitive. Supposedly, the geometric transformations are to be expressed in a set of transformation equations by which, in the spirit of analytic geometry the actual transformations are to be carried out. A proposed such set of equations is given in Figure 6. Each variate is expressed in terms of one constant, m, and two chance variables, t and e, each with zero expected value. If

$$a_1 = a_2 = b_1 = b_2$$

and
$$t_1 = t_2$$

we get the simplest form of Grubbs' problem, treated in [20]. Other cases are obtained by specifying certain of the quantities in various ways.

Bibliography on
COMPARISON OF CORRELATED VARIANCES

Clifford J. Maloney
Food and Drug Administration
Rockville, Maryland

1.   Aitken, A. C. [1950]. "On the Statistical Independence of
     Quadratic Forms in Normal Variates," Biometrika, Vol. 37,
     Pts. 1 and 2, pp. 93-96.

2.   Aspen, Alice A. [1949]. "Tables for Use in Comparisons Whose
     Accuracy Involves Two Variances, Separately Estimated,"
     Biometrika, Vol. 36, Pts. 3 and 4, pp. 290-296.

3.   Berkson, Joseph [1950]. "Are There Two Regressions?,"
     J. Amer. Stat. Assoc., Vol. 45, No. 250, pp. 164-180.

4.   Bose, S. [1935]. "On the Distribution of the Ratio of Variances
     of Two Samples Drawn from a Given Normal Bivariate Correlated
     Population," Sankhya, Vol. 2, Pt. 1, pp. 65-72.

5.   Chanda, K. C. [1951]. "Comparative Efficiencies of L-Test and
     Pitman's Test in Testing for Equality of Variances," Bull.
     Int'l. Stat. Inst., Vol. 33, Pt. 2, pp. 215-218.

6.   Choi, S. C. and Wettle, E. [1972]. "A Test for the Homogeneity
     of Variances among Correlated Variables," Biometrics, Vol. 28,
     No. 2, pp. 589-592.

7.   Cochran, W. G. [1943]. "The Comparison of Different Scales of
     Measurement for Experimental Results," Am. Math. Stat., Vol. 14,
     pp. 205-216.

8.   Cochran, W. G. [1968]. "Errors of Measurement in Statistics,"
     Technometrics, Vol. 10, No. 4, pp. 637-666.

9.   Curnow, R. M. [1957]. "Heterogeneous Error Variances in Split-
     Plot Experiments," Biometrika, Vol. 44, Pts. 3 and 4, pp. 378-383.

10. Dar, S. N. [1962]. "On the Comparison of the Sensitivity of
     Experiments," J. Roy. Stat. Soc., B. Vol. 4, No. 2, pp. 447-453.

11. DeLury, D. B. [1938]. "Note on Correlations," Ann. Math. Stat.,
     Vol. 9, pp. 149-151.

12. Durbin, J. and Kendall, M. G. [1951]. "The Geometry of Estimation," Biometrika, Vol. 38, Pts. 1 and 2, pp. 150-158.

13. Ehrenberg, A. S. C. [1950]. "The Unbiased Estimation of Hetero-geneous Error Variances," Biometrika, Vol. 37, pp. 347-357.

14. Finney, David J. [1938]. "The Distribution of the Ratio of Estimates of the Two Variances in a Sample from a Normal Bivariate Population," Biometrika, Vol. 30, pp. 190-192.

15. Graybill, F. [1954]. "Variance Heterogeneity in a Randomized Block Design," Biometrics, Vol. 10, No. 4, pp. 516-520.

16. Grubbs, F. E. [1948]. "On Estimating Precision of Measuring Instruments and Product Variability," J. Amer. Stat. Assoc., Vol. 43, No. 243, pp. 243-264, 564.

17. Grubbs, F.E. "Error of Measurement, Precision, Accuracy, and the Statistical Comparison of Measuring Instruments" (to appear).

18. Hahn, J. H. and Nelson, W. [1970]. "A Problem in the Statistical Comparison of Measuring Devices," Technometrics, Vol. 12, No. 1, pp. 95-102.

19. Han, C. P. [1968]. "Testing the Homogeneity of a Set of Correlated Variances," Biometrika, Vol. 55, No. 2, pp. 317-326.

20. Han, C. P. [1969]. "Testing the Homogeneity of Variances in a Two-Way Classification," Biometrics, Vol. 25, No. 1, pp. 153-158.

21. Hanumara, R. C. and Thompson, W. A. [1968]. "Percentage Points of the Extreme Roots of a Wishart Matrix," Biometrika, Vol. 55, pp. 327-334.

22. Hirschfeld, H. O. [1937]. "The Distribution of the Ratio of Covariance Estimates in Two Samples Drawn from Normal Bivariate Populations," Biometrika, Vol. 29, Pt. 1, pp. 65-79.

23. Jaech, J. L. [1971]. "Further Tests of Significance for Grubbs' Estimators," Biometrics, Vol. 27, No. 4, pp. 1097-1101.

24. John, S. [1972]. "The Distribution of a Statistic Used for Testing Sphericity of Normal Distributions," Biometrika, Vol. 59, Pt. 1, pp. 169-173.

25. Kendall, Maurice G. [1951]. "Regression, Structure, and Functional Relationships," Biometrika, Vol. 38, Pts. 1 and 2, pp. 11-25.

26. Ku, Harry H., Ed. [1969]. "Precision Measurement and Calibration," Spec. Publ. 300, Vol. 1, National Bureau of Standards.

27. Madansky, A. [1959]. "The Fitting of Straight Lines When both Variables are Subject to Error," J. Amer. Stat. Assoc., Vol. 54, No. 285, pp. 173-205.

28. Makowski, Gary. "Note on a Significance Test from Grubbs' Estimators" (to be published in Biometrics).

29. Maloney, C. J. and Rastogi, S. C. [1970]. "Significance Test for Grubbs's Estimators," Biometrics, Vol. 26, No. 4, pp. 671-676.

30. Mathai, A. M. and Rathie, P. N. [1971]. "The Exact Distribution of Wilks' Criterion," Ann. Math. Stat., Vol. 42, No. 3, pp. 1010-1019.

31. Mathai, A. M. and Rathie, P. N. [1971]. "The Problem of Testing Independence," 38th Session, Inter. Stat. Inst., Washington, D.C. Proceedings, pp. 254-258.

32. Mauchly, John W. [1940]. "Significance Test for Sphericity of a Normal n-variate Distribution," Ann. Math. Stat., Vol. 11, No. 2, pp. 204-209.

33. McIntyre, G. A. [1972]. "Homogeneity of Variance in a Two-way Classification," Biometrics, Vol. 28, No. 1, p. 246.

34. Mehta, J. S. and Gurland, J. [1969]. "Testing Equality of Means in the Presence of Correlation," Biometrika, Vol. 56, pp. 119-126.

35. Morgan, W. A. [1939]. "A Test for the Significiance of the Difference between Two Variances in a Sample from a Normal Bivariate Population," Biometrika, Vol. 31, Pt. 1, pp. 13-19.

36. Pitman, E. J. G. [1939]. "A Note on Normal Correlation," Biometrika, Vol. 31, Pt. 1, pp. 9-12.

37. Putter, J. [1967]. "Orthonormal Bases of Error Spaces and Their Uses for Investigating the Normality and Variances of Residuals," J. Amer. Stat. Assoc, Vol. 62, No. 319, pp. 1022-1036.

38. Richardson, David H. and Wu, De-Min. [1970]. "Least Squares
    and Grouping Method Estimators in the Error in Variables Model,"
    J. Amer. Stat. Assoc., Vol. 65, No. 330, pp. 724-748.

39. Roy, S. N. and Bose, R. C. [1953]. "Simultaneous Confidence
    Interval Estimation," Ann. Math. Stat., Vol. 24, No. 4,
    pp. 513-36.

40. Russell, T. S. and Bradley, R. A. [1958]. "One-way Variances in
    a Two-way Classification," Biometrika, Vol. 45, pp. 111-129.

41. Shukla, G. "Some Exact Tests of Hypotheses about Grubbs's
    Estimators" (to be published in Biometrics).

42. Shukla, G. [1972]. "An Invariant Test for the Homogeneity of
    Variance in a Two-way Classification," Biometrics, Vol. 28,
    No. 4, pp. 1063-1072.

43. Sichel, Herbert S. [1949]. "The Method of Frequency-moments and
    Its Application to Type VII Populations," Biometrika, Vol. 36,
    Parts 3 and 4, pp. 404-425.

44. Smith, H. Fairfield [1950]. "Estimating Precision of Measuring
    Instruments," J. Amer. Stat. Assoc., Vol. 45, No. 251, pp. 447-451.

45. Theil, H. and Yzeren, J. van [1956]. "On the Efficiency of Wald's
    Method of Fitting Straight Lines," Revue Inst. de Stat., Vol. 24,
    pp. 17-26.

46. Thompson, William A., Jr. [1962]. "Estimation of Dispersion
    Parameters," J. of Res. Nat. Bur. Stds., 66B No. 4, pp. 161-164.

47. Thompson, William A., Jr. [1962]. "The Problem of Negative Estimates
    of Variance Components," Ann. Math. Stat., Vol. 33, No. 1,
    pp. 273-289.

48. Thompson, William A., Jr. [1963]. "Precision of Simultaneous
    Measurement Procedures," Proc. Eighth Conf. Des. of Exp. in Army
    Res. Dev. and Tstg., pp. 175-185.

49. Tukey, John W. [1951]. "Components in Regression," Biometrics,
    Vol. 7, No. 1, pp. 33-69.

50. Villars, D. S. and Anderson, T. W. [1943]. "Some Significance
    Tests for Normal Bivariate Distributions," Ann. Math. Stat.,
    Vol. 14, No. 2, pp. 141-148.

51. Votaw, D. F., Jr. [1948]. "Testing Compound Symmetry in a
    Normal Multivariate Distribution," Ann. Math. Stat., Vol. 19,
    pp. 447-73.

52. Wald, Abraham [1940]. "The Fitting of Straight Lines if Both
    Variables are Subject to Error," Ann. Math. Stat., Vol. 11,
    No. 3, pp. 284-300.

53. Whitwell, John C. [1951]. "Estimating Precision of Textile
    Instruments," Biometrics, Vol. 7, No. 1, pp. 102-112.

54. Wilks, S. S. [1946]. "Sample Criteria for Testing Equality of
    Means, Equality of Variances, and Equality of Covariances in
    a Normal Multivariate Distribution," Ann. Math. Stat., Vol. 17,
    No. 3, pp. 257-81.

55. Williams, E. J. [1970]. "Comparing Means of Correlated Variates,"
    Biometrika, Vol. 57, pp. 459-61.
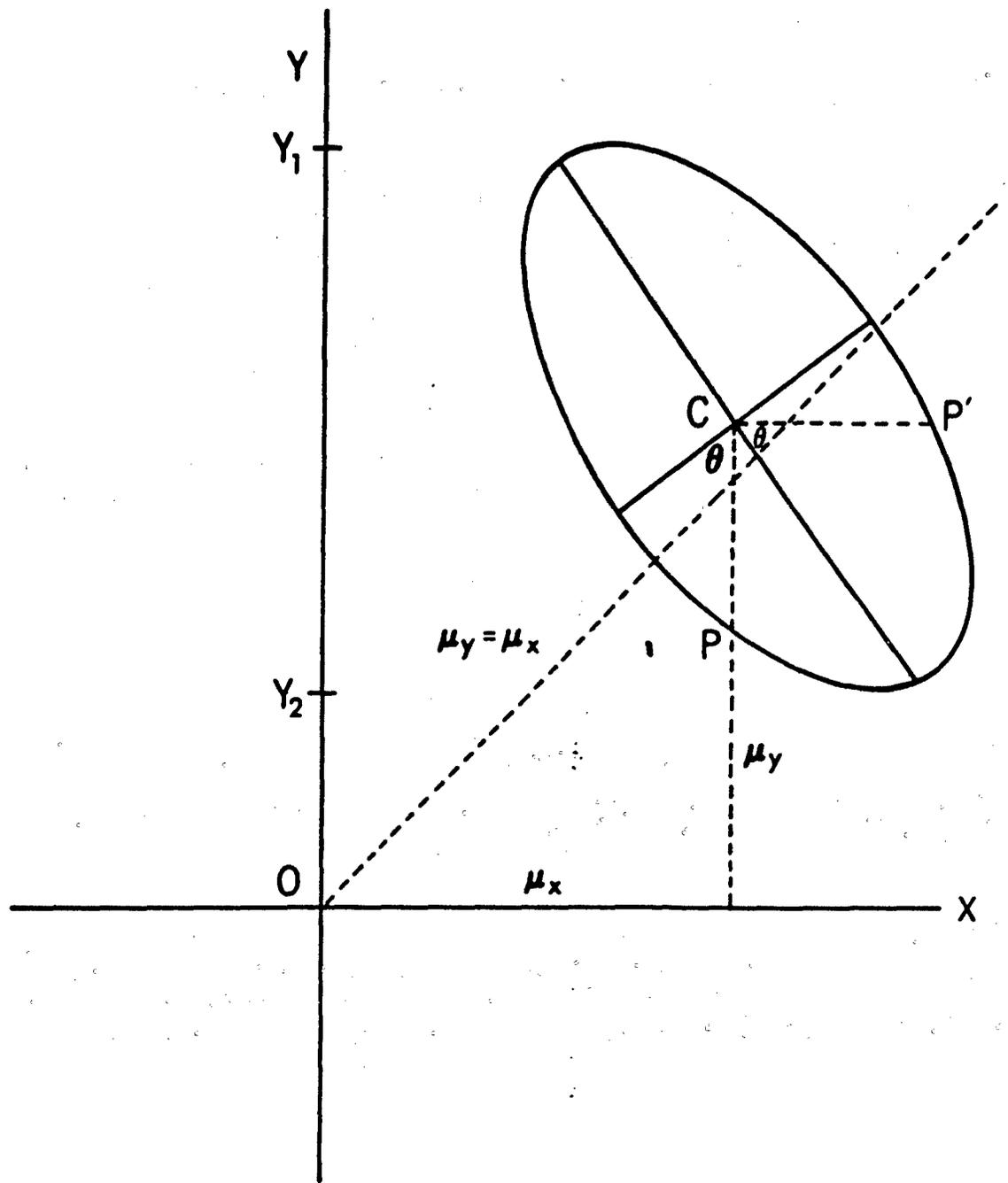
General Bivariate Normal Distribution

Figure 1
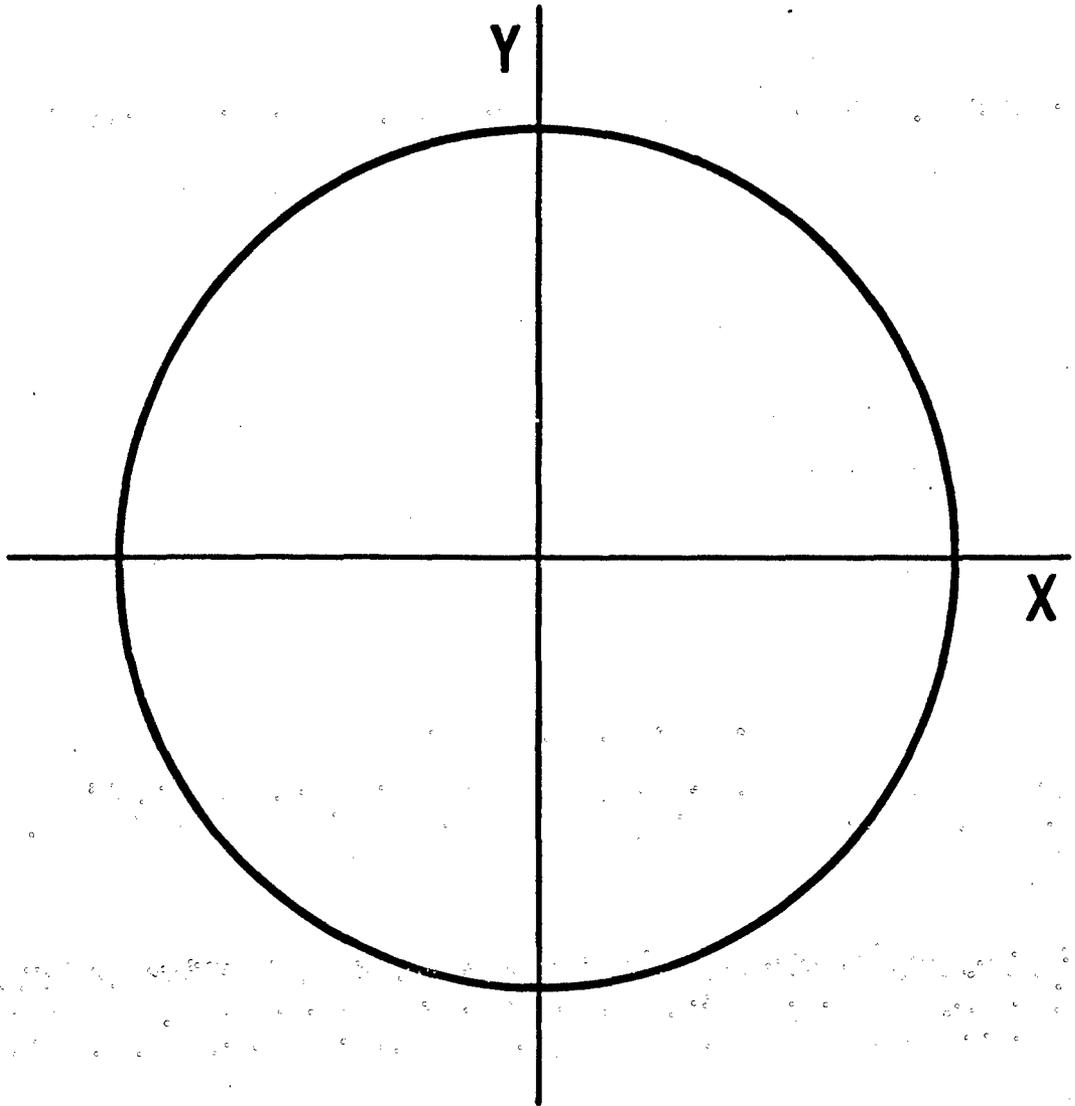
EQUAL INDEPENDENT VARIANCES

Figure 2
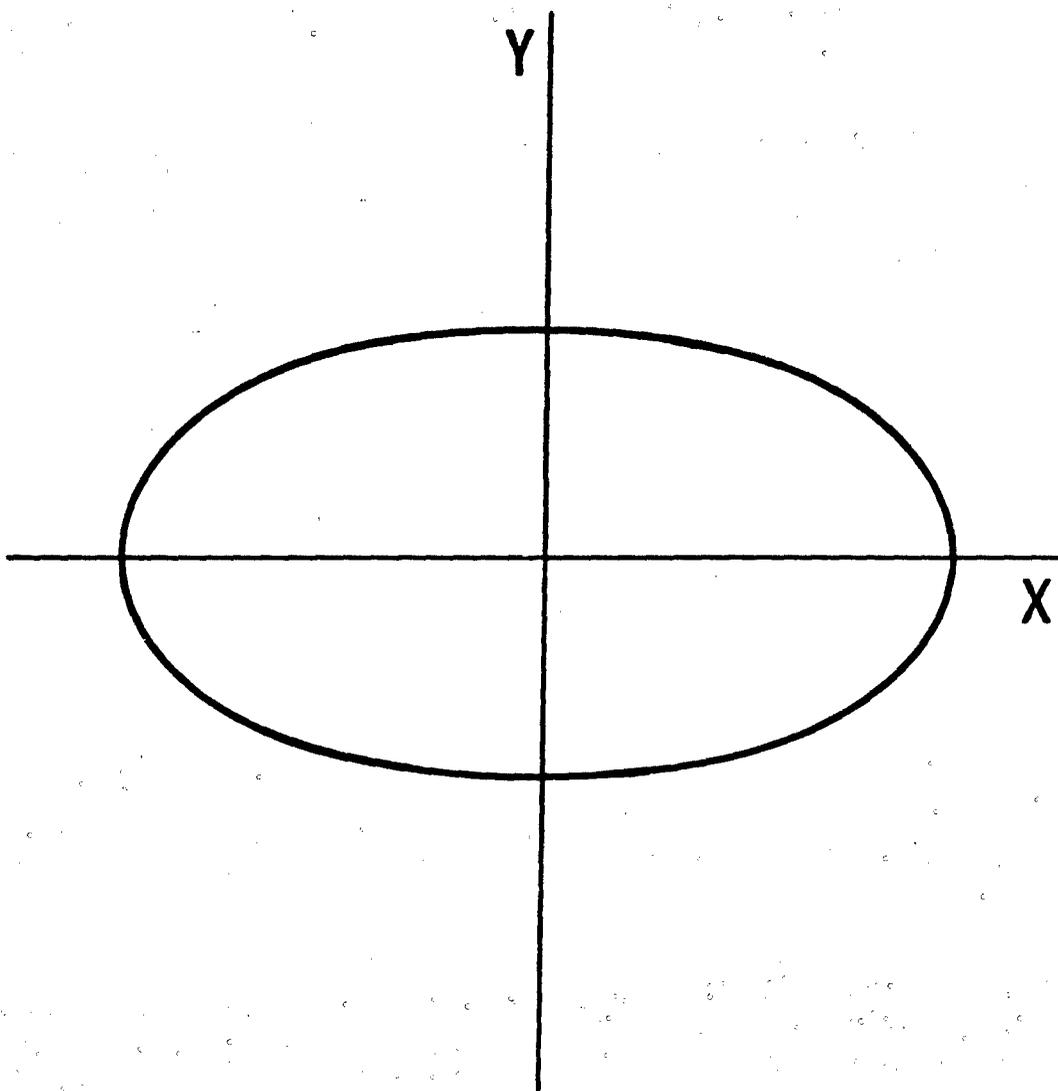
UNEQUAL BUT INDEPENDENT VARIANCES

Figure 3
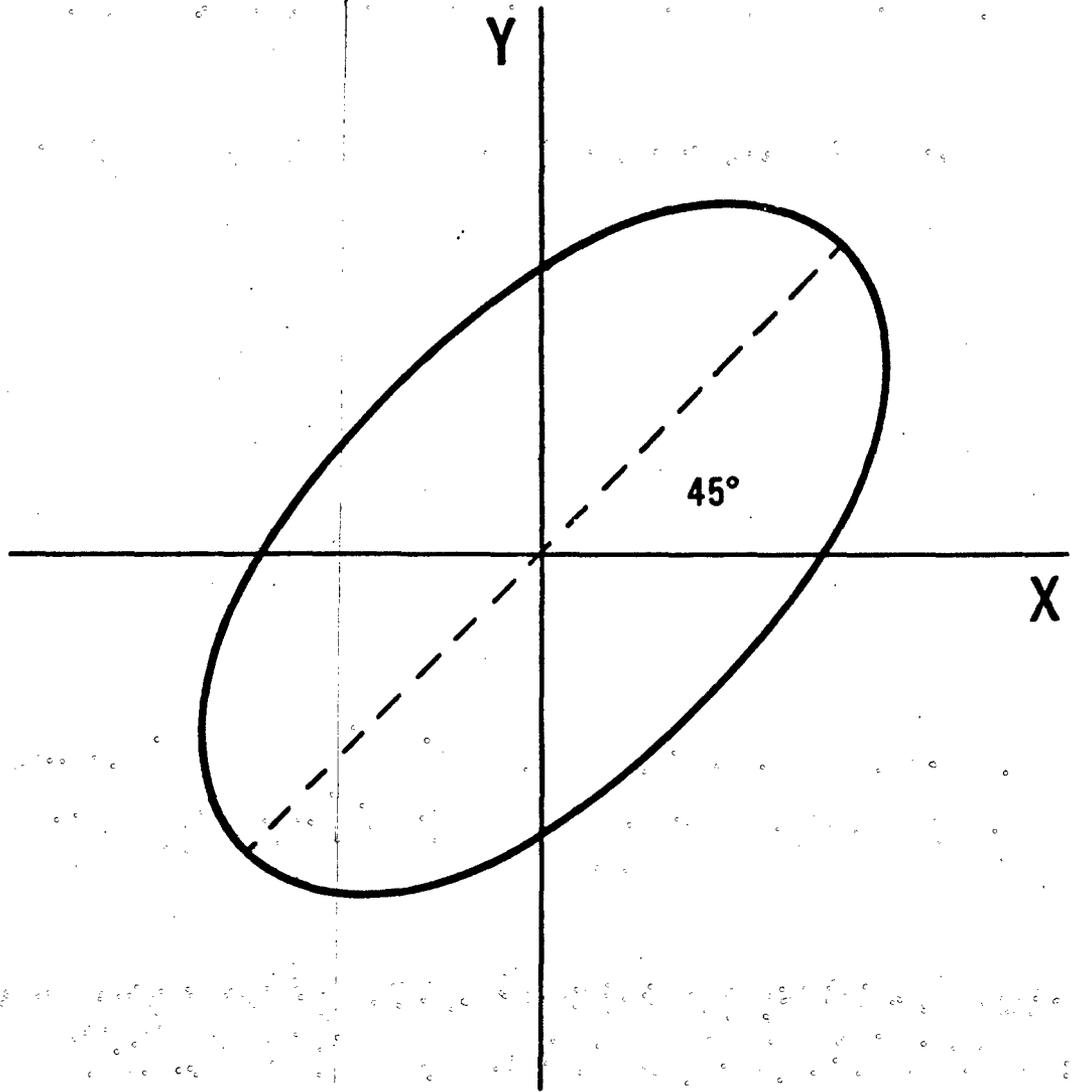
EQUAL DEPENDENT VARIANCES

Figure 4
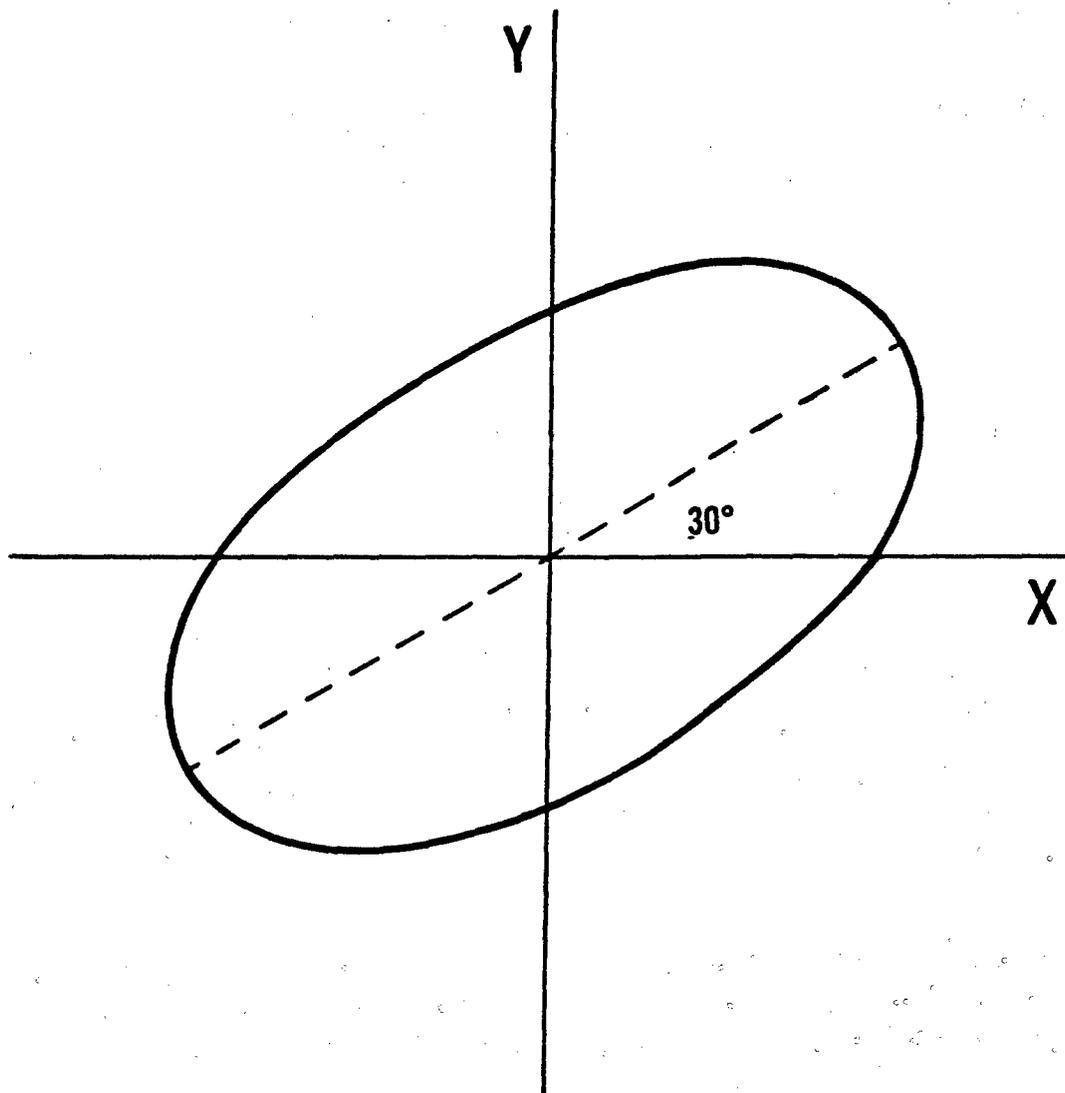
UNEQUAL DEPENDENT VARIANCES

Figure 5

$$X = m_1 + a_1 t_1 + b_1 e_1$$

$$Y = m_2 + a_2 t_2 + b_2 e_2$$

$$V(x) = a_1^2 V(t_1) + b_1^2 V(e_1)$$

$$V(y) = a_2^2 V(t_2) + b_2^2 V(e_2)$$

Figure 6

-547-

A SYSTEM FOR POSITION-LOCATION BASED ON RANGES *

Richard H. F. Jackson, James A. Lechner, and David J. Sookne

Applied Mathematics Division
National Bureau of Standards
U. S. Department of Commerce

ABSTRACT. The system in question is intended to employ range-only infor-
mation, or range-plus-altitude information, to track the positions of up
to hundreds of users (only some of which measure ranges to other users)
in three dimensions or on the earth's surface. This paper describes the
structure, use and results of a simulation study which focused, within a
larger analysis, on the absolute and relative adequacies of various
mathematical position-estimation algorithms. The discussion will include
comparisons among the six different algorithms investigated, considering
both accuracy and computer time required.

1. INTRODUCTION

The position location system under consideration utilizes one frequency
channel which is shared on a time division basis by a number, n, of users
of the system. Those users (hereafter called "units') are all synchronized
by a suitable electromagnetic signal at least once during each time inter-
val ΔT (the cycle time), and within every cycle each unit is assigned
one or more time slots during which it emits an electromagnetic signal.
This latter signal is received by some or all of the other units, who then
measure the time of arrival of that signal. From knowledge of the times
of arrival and the assignment of slots (emission times) to the sending
unit, the distance between that unit and the receiving units can be
estimated. These ranges, alone or in concert with altitude estimates,
can be used to estimate the location of the sending unit.

The system is intended to employ this approach to track the positions of
up to hundreds of units in three dimensions or on the earth's surface.
This paper describes the structure, use and results of a simulation study
which focused, within a larger feasibility analysis, on the absolute and
relative adequacies of various mathematical algorithms used in performing
that position estimation and tracking. The discussion will include
comparisons among the six different algorithms proposed as well as between
the final algorithm proposed and a Kalman filter approach developed
elsewhere.

The problem, specifically, was to determine the feasibility of such a
system in the presence of measurement errors. If all measurements were
exact, three units in the plane or four units in three-space would suffice
to locate another (or all other) units in the system. However, in the
presence of error, more measurements are needed and the "best" position

---

(the one that minimizes total error in some mathematical sense) must be found. The fundamental question was whether (with a suitable algorithm) the overall error could be kept within tolerable bounds as the system operates through time.

The various algorithms proposed were tested, compared, and evaluated in a number of different ways. One was to evaluate the algorithms analytically wherever possible, and in this vein several of mathematical studies of their properties were conducted. For example, a "simplest case" analysis was performed on a series of one dimensional location operations involving only two units, with Gaussian range errors assumed and with maximum likelihood estimation used as the method of position location. This situation was "solved" in closed form and it was shown that the distribution of the two units' position location errors (distances between true and estimated positions) stabilized in time.

Another method of evaluation can be described as "one-shot" accuracy tests. These were controlled tests in which the effects of random variation were minimized, by using identical sets of range and position errors for each algorithm being compared. They were designed to determine, first, how accurately each algorithm could locate a unit (the _locatee_) given range measurements from units (the _locators_) in known positions, and, second, how accurately they could locate a unit when the assumed locator-positions are also in error.

These one-shot tests involved a set of stationary locators that was used to locate each of several stationary locatees. Corresponding to the real situation in which errors exist in both ranges and positions, the "true" inter-unit distances and/or locator positions were perturbed with random errors to represent the information available to the system. The position of each locatee was then estimated by each of the algorithms being tested, and the position location errors were compared and tabulated.

While the on-shot tests described above do indicate which algorithms are "better" than others, they provide no information about error propagation in the system (i.e., the cumulative effect of thousands of location operations in which the positions of the locators are in error, since they themselves served as locatees a fraction of a second before). Simulation model of the position location system was designed and programmed. This simulation program (WHERSM) provides a "real-life" framework for the system through its capability for moving all units along prescribed or randomly generated (by WHERSM) paths that conform to the local terrain, which is available to the program either as a continuous function or as a digitized "terrain map". WHERSM checks intervisibilities among units to determine which ranges are available, computes the inter-unit distances required, perturbs these distances according to specified error distribution laws to produce "range measurements", and transmits these ranges to the position location algorithm being tested. The simulation program also provides a facility for monitoring the operation of the system under a variety of movement scenarios. It further provides outputs, useful for checking out the feasibility of the system, which include: the successive locations of all units, the errors between true and estimated positions, and measures of confidence derived from the available measurements.

-550-

2. DESCRIPTION OF ALGORITHMS AND RESULTS OF ONE-SHOT TESTS

Six position-location algorithms were programmed and analyzed, using one-shot accuracy tests and the simulation model described above. The results of the simulation tests are described in section 3.

The first algorithm investigated was the Linear Method (LM), whose two-dimensional version uses three ranges selected on the basis of geometrical considerations. Let r_i, $i = 1,2,3$ be the measured ranges, and let (x_i, y_i), $i = 1,2,3$ be the estimated positions of the locators. For $i = 1,2,3$, define C_i by

$$C_i = C_i(x,y) = (x - x_i)^2 + (y - y_i)^2 - r_i^2 .$$

Note that the equation $C_i = 0$ is that of a circle with center (x_i, y_i) and radius r_i; we call this a locator circle. Now for $i \neq j$, the equation $C_i - C_j = 0$ reduces to

$$2x(x_j - x_i) + 2y(y_j - y_i) + x_i^2 - x_j^2 + y_i^2 - y_j^2 + r_j^2 - r_i^2 = 0, \qquad (1)$$

i.e., the equation is linear in x and y, so $C_i - C_j = 0$ represents a line in the plane, the "radical axis" of the two locator circles. (We exclude the degenerate case when locators coincide, in which case $x_i = x_j$, $y_i = y_j$.) If the three locators are not collinear, the three lines represented by $C_1 - C_2 = 0$, $C_1 - C_3 = 0$, and $C_2 - C_3 = 0$ (also called radical axes) must intersect in a point, since if $C_1 - C_2 = 0$ and

$C_1 - C_3 = 0$, then $C_2 - C_3 = (C_1 - C_3) - (C_1 - C_2) = 0$. (If the locators are collinear, the three lines are parallel, and no solution exists.) The linear method chooses the point of intersection as the estimated position of the locatee. Note that if the x_i, y_i, and r_i are all error-free, then the locatee lies on each circle $C_i = 0$, and hence on each line $C_i - C_j = 0$, so the locatee is at the intersection point, and the solution is exact. However, the one-shot tests showed the linear method to be very sensitive to errors in the x_i, y_i, and r_i. It was rejected, and the three-dimensional version was not programmed for testing.

The second algorithm, that of the Smallest Tangent Circle (STC), also uses three ranges in two dimensions. Conceived by K. Goldberg, it calculates the centers and radii of the (up to eight) circles each of which is tangent to all three locator circles $C_i = 0$, $i = 1,2,3$, and chooses the center of the smallest of the circles as the estimated position of the locatee. The algorithm sets

$$A_i = (x_j - x_k)^2 + (y_j - y_k)^2 \quad \text{and} \quad W_i = A_j + A_k - A_i, \quad i = 1,2,3,$$

where i,j,k go over all cyclic permutations of $\{1,2,3\}$. The program then proceeds to calculate

$$B = 2(A_1 A_2 + A_2 A_3 + A_1 A_3) - A_1^2 - A_2^2 - A_3^2$$

and

$$F = A_1 A_2 A_3 + W_1(.5(r_2^2 - r_3^2)^2 - A_1 r_1^2) + W_2(.5(r_1^2 - r_3^2)^2 - A_2 r_2^2) + W_3(.5(r_2^2 - r_1^2)^2 - A_3 r_3^2).$$

Next, we let $(\epsilon_1, \epsilon_2, \epsilon_3)$ vary over all 8 combinations that result from $\epsilon_i = \pm 1$. For each combination, we substitute $R_i = \epsilon_i r_i$; $i = 1,2,3$, and set

$$c_i = W_i(R_j - R_k)^2 \quad i,j,k \text{ as above};$$

$$P = 2(c_1 + c_2 + c_3) - B$$

$$Q = W_1 A_1 R_1 + W_2 A_2 R_2 + W_3 A_3 R_3 - C_1(R_2 + R_3) - C_2(R_1 + R_3) - C_3(R_2 + R_1) ,$$

and the equation

$$Pr^2 + Qr + F = 0$$

is solved for r, the radius of the tangent circle. The eight combinations of ϵ_i will each yield two values of r, but there is duplication; if $(\epsilon_1, \epsilon_2, \epsilon_3)$ yields r-values s_1 and s_2, then $(-\epsilon_1, -\epsilon_2, -\epsilon_3)$ will yield $-s_1$ and $-s_2$. Thus only four ϵ-combinations need be tried. Let the smallest in magnitude of the eight r-values be denoted by s. The center of this circle is found by setting

$$t_i = r_i + \epsilon_i s, \quad i = 1,2,3,$$

and applying the linear method, with r_i replaced by t_i.

One-shot accuracy tests showed that while this method was less sensitive than the linear method to errors in x_i, y_i, and r_i, it was more sensitive than the LSL method described below, so that the three-dimensional version, involving the smallest tangent sphere, was not programmed for testing.

The linear method and smallest tangent circle methods both suffer from other defects as well. Since they use only three locators, bad data cannot easily be eliminated, and there is no way to give higher weight to locators whose positions are better known than others. Thus, for example, if ten locators report ranges, the best elimination technique available would involve choosing some subset of the 120 triples of locators, calculating the solution for each triple, averaging all solutions, eliminating those that are far from the average, and re-averaging. This procedure is quite cumbersome and time-consuming.

The next method, called the Least Squares Linear (LSL) method, finds a global minimum of the sum of the squares of the distances from the locatee to all the radical axes $C_i - C_j = 0$, defined above. Expressing one such line explicitly,

$$C_i - C_j \equiv a_k x + b_k y - c_k = 0 ,\qquad (2)$$

where the coefficients have been normalized so that

$$a_k^2 + b_k^2 = 1 ,\qquad (3)$$

the distance from a point (x,y) to the line is simply $\left| a_k x + b_k y - c_k \right|$.

-554-

With k running over all radical axes, the function to be minimized is as follows:

$$F = F(x,y) = \sum_k (a_k x + b_k y - c_k)^2 =$$

$$\sum_k (a_k^2 x^2 + 2a_k b_k xy + b_k^2 y^2 - 2a_k c_k x - 2b_k c_k y + c_k^2) . \tag{4}$$

Setting $\dfrac{\partial F}{\partial x}$ and $\dfrac{\partial F}{\partial y}$ to zero yields

$$\frac{\partial F}{\partial x} = \sum_k (2a_k^2 x + 2a_k b_k y - 2a_k c_k) = 0$$

$$\frac{\partial F}{\partial y} = \sum_k (2a_k b_k x + 2b_k^2 y - 2b_k c_k) = 0 ,$$

or

$$(\sum_k a_k^2)x + (\sum_k a_k b_k)y = \sum_k a_k c_k$$

$$(\sum_k a_k b_k)x + (\sum_k b_k^2)y = \sum_k b_k c_k \tag{5}$$

These linear equations are solved for x and y. In three dimensions, the equations are

$$a_k x + b_k y + c_k z - d_k = 0 \tag{2'}$$

$$a_k^2 + b_k^2 + c_k^2 = 1 \qquad\qquad (3')$$

$$F = \sum_k (a_k x + b_k y + c_k z - d_k)^2 \qquad\qquad (4')$$

$$\left.\begin{array}{l}
(\sum_k a_k^2)x + (\sum_k a_k b_k)y + (\sum_k a_k c_k)z = \sum_k a_k d_k \\[2mm]
(\sum_k a_k b_k)x + (\sum_k b_k^2)y + (\sum_k b_k c_k)z = \sum_k b_k d_k \\[2mm]
(\sum_k a_k c_k)x + (\sum_k b_k c_k)y + (\sum_k c_k^2)z = \sum_k c_k d_k \;.
\end{array}\right\} \qquad (5')$$

Two- and three-dimensional versions of this algorithm were programmed.

The remaining three algorithms utilize penalty functions which measure lack of fit of a position estimate to the data. Each algorithm searches for an estimate that minimizes a particular penalty function.

The <u>MINMAX</u> algorithm as suggested by W.A. Horn, was programmed in 2 dimensions. It finds

$$\min_{(x,y)} \; \max_i \; |d_i - r_i| \;,$$

where the index i runs over all locators, and

$$d_i = \sqrt{(x-x_i)^2 + (y-y_i)^2}$$

is the calculated distance between the estimated positions of locator and locatee. It can be proved that the solution lies at the center of a

circle tangent to 2 or 3 of the locator circles C_i, and that circle intersects or contains all the C_i. The algorithm, then, examines all pairs or triples of circles C_i, calculates the center and radius of each tangent circle, and checks whether that circle intersects or contains all the C_i. This procedure is quite time-consuming because it must consider so many tangent circles. For example, with 12 locators, it finds eight tangent circles for each of the 220 triples of C_i plus one tangent circle for each of the 66 pairs of C_i, or 1826 tangent circles in all. Further, it was not as accurate as the remaining two methods so it was rejected.

The Least Squares (LS) algorithm minimizes locally

$$E = \sum_{i=1}^{N} w_i (d_i - r_i)^2 \,,$$

where N is the number of locators, and the w_i are weights (see section 3). Given a starting point, the algorithm attempts to set $\frac{\partial E}{\partial x}$, $\frac{\partial E}{\partial y}$ (and $\frac{\partial E}{\partial z}$ in the 3-dimensional case) to zero, using the Newton-Raphson iteration technique. In three dimensions, let $p = (x_o, y_o, z_o)$ be the starting point. The algorithm finds a new point $p' = (x_o + \Delta x, y_o + \Delta y, z_o + \Delta z)$ by solving for $\Delta x, \Delta y, \Delta z$ in the first order equations

$$\sum_{i=1}^{N} w_i \left(\frac{r_i(x_o-x_i)^2}{d_i^3} + \frac{d_i-r_i}{d_i} \right) \Delta x + \sum_{i=1}^{N} w_i \frac{r_i(x_o-x_i)(y_o-y_i)}{d_i^3} \Delta y$$

$$+ \sum_{i=1}^{N} w_i \frac{r_i(x_o-x_i)(z_o-z_i)}{d_i^3} \Delta z + \left. \frac{\partial E}{\partial x} \right|_p = \left. \frac{\partial E}{\partial x} \right|_p = 0.$$

$$\sum_{i=1}^{N} w_i \frac{r_i(x_o-x_i)(y_o-y_i)}{d_i^3} \Delta x + \sum_{i=1}^{N} w_i \left(\frac{r_i(y_o-y_i)^2}{d_i^3} + \frac{d_i-r_i}{d_i} \right) \Delta y$$

$$+ \sum_{i=1}^{N} w_i \frac{r_i(y_o-y_i)(z_o-z_i)}{d_i^3} \Delta z + \left. \frac{\partial E}{\partial y} \right|_p = 0$$

$$\sum_{i=1}^{N} w_i \frac{r_i(x_o-x_i)(z_o-z_i)}{d_i^3} \Delta x + \sum_{i=1}^{N} w_i \frac{r_i(y_o-y_i)(z_o-z_i)}{d_i^3} \Delta y$$

$$+ \sum_{i=1}^{N} w_i \left(\frac{r_i(z_o-z_i)^2}{d_i^3} + \frac{d_i-r_i}{d_i} \right) \Delta z + \left. \frac{\partial E}{\partial z} \right|_p = 0.$$

$$(6)$$

Note that solving this system necessitates the calculation of

$d_i = \sqrt{(x_0-x_i)^2 + (y_0-y_i)^2 + (z_0-z_i)^2}$ for each locator i. Since the taking

of a square root on our computer consumes the same amount of CPU time as

about 12 multiplications, this algorithm was replaced by the Least Squares

Squared (LSS) algorithm, which minimizes

$$E = \sum_{i=1}^{N} \frac{w_i(d_i^2-r_i^2)^2}{4r_i^2}$$

-558-

Factoring, we see that

$$\frac{w_i(d_i^2 - r_i^2)^2}{4r_i^2} = \frac{w_i(d_i - r_i)^2(d_i + r_i)^2}{4r_i^2} = w_i(d_i - r_i)^2 \left(1 + \frac{d_i - r_i}{2r_i}\right)^2 .$$

Now, when the starting point for the Newton-Raphson method is "close to" the true location (i.e., units have been successfully tracked), then $d_i \approx r_i$ and the factor above, $1 + (d_i - r_i)/2r_i$, is approximately 1. Hence

$$E \approx \sum_{i=1}^{N} w_i(d_i - r_i)^2 ,$$

so one would expect that the least squares squared and the least squares algorithms should produce almost identical answers. This was borne out by two one-shot accuracy tests of 72 trials each; the errors produced by the two algorithms agreed to several figures. In the test with smaller range errors (Gaussian, with standard deviation of 1 m), the position-location errors ranged up to 3.32 m, but the LS and LSS estimates (for the same x_i, y_i and r_i) never differed by more than .0004 m. In the other test, using a standard deviation of 10 m, position errors reached 31.9 m, but the LS and LSS estimates never differed by more than .045 m.

The equations for LSS are

$$\sum_{i=1}^{N} \frac{w_i \left(2(x_0-x_i)^2 + d_i^2 - r_i^2\right)}{r_i^2} \Delta x + \sum_{i=1}^{N} \frac{2w_i(x_0-x_i)(y_0-y_i)}{r_i^2} \Delta y$$

$$+ \sum_{i=1}^{N} \frac{2w_i(x_0-x_i)(z_0-z_i)}{r_i^2} \Delta z + \frac{\partial E}{\partial x} = 0$$

$$\sum_{i=1}^{N} \frac{2w_i(x_0-x_i)(y_0-y_i)}{r_i^2} \Delta x + \sum_{i=1}^{N} \frac{w_i\left(2(y_0-y_i)^2 + d_i^2 - r_i^2\right)}{r_i^2} \Delta y$$

$$+ \sum_{i=1}^{N} \frac{2w_i(y_0-y_i)(z_0-z_i)}{r_i^2} \Delta z + \frac{\partial E}{\partial y} = 0$$

$$\sum_{i=1}^{N} \frac{2w_i(x_0-x_i)(z_0-z_i)}{r_i^2} \Delta x + \sum_{i=1}^{N} \frac{2w_i(y_0-y_i)(z_0-z_i)}{r_i^2} \Delta y$$

$$+ \sum_{i=1}^{N} \frac{w_i\left(2(z_0-z_i)^2 + d_i^2 - r_i^2\right)}{r_i^2} \Delta z + \frac{\partial E}{\partial z} = 0.$$

(6')

Tests showed that LSS converged after 2 to 4 iterations in nearly all cases. The solution was in the neighborhood of the correct local minimum even when the starting point was 200 meters away from the true position.

One-shot tests showed that LSS and LS were more accurate than LSL, which in turn was more accurate than MINMAX, STC, and LM. LSS was chosen over LS because of slightly lower computer-time (and nearly exactly the same accuracy; see above), and LSL was also kept because it was the best of the algorithms not needing a starting point. It serves to provide a starting estimate for LSS when a unit first joins (or rejoins) the system. Otherwise, such a starting point is found by extrapolating on two recent position estimates of the locatee.

3. RESULTS OF SIMULATION TESTS

Since there were well over 150 computer runs made (in addition to debugging runs), only summaries of the results will be given. Several classes of runs can be distinguished:

(A) The early runs designed to settle on the best general algorithm;
(B) The large group of runs designed to aid in selection of values for the many parameters and options;
(C) Controlled tests on scenarios involving aircraft, to decide when to use a 3-dimensional method;
(D) "Proof" runs on actual (digitized) terrain, with realistic movements and calculated intervisibilities;
(E) "Tracking" runs, designed to evaluate the potential for improvement offered by tracking and also to furnish the data for a controlled comparison between the algorithms reported here and a Kalman filter implementation due to David Plutchak at General Dynamics in San Diego.

These different groups of runs were intermixed during the study, of course, but they will be described in turn below.

A. Preliminary tests. Most of the early simulation runs involved two-dimensional situations, with some or all units moving, with random changes in direction at random times, but with velocity vectors always restricted to the first quadrant. These runs had two aims: to verify the indications of the one-shot tests already described, and to shed some light on whether satisfactory system behavior required some units to be stationary. The one-shot test results were verified. With respect to simultaneous movement, interesting results were obtained. It quickly became obvious that when all units were moving, no algorithm could avoid a fatal accumulation of errors corresponding to a shift in the coordinate system, and a concomitant growth in the relative errors as well. In many runs, when the errors had grown sufficiently, convergence became a problem: the algorithm could not find a solution within the allotted number of iterations. For example, with 10 stopped and 10 moving units reporting ranges (and with perfect intervisibility), velocities up to 10 km/hr, a cycle time of 30 s, and a simulated operating time of 2 hrs, three runs were done, with the following results. In the first two runs, range errors were uniformly distributed between -3m and +3m (giving a standard deviation of 1.73 m). The first run had none of the units known to be fixed; position errors reached 10 km before the 2 hrs. were up. The second run had 4 of the units known fixed; the average error at 2 hrs was 14 m, and the maximum error throughout the 2 hr was

56m. For comparison, the third run used Gaussian range errors, with a standard deviation of 6 m, and again had 4 known fixed units. The average error at 2 hr was 9.6 m, considerably less than the 14 m on the second run, in spite of the more than threefold increase in range-error standard deviation. The improvement is due, of course, to the fact that LS (or LSS is the optimal algorithm for a Gaussian error distribution.

B. Parameter and Option Studies. The choice of parameters and options is endless. Thus, the selections reached in this study are in no general sense optimum. The broad decisions to be made include:

(1) Choice of weights for the LSS procedure;
(2) Choice between a 2-dimensional and a 3-dimensional algorithm;
(3) Whether to track (or filter), and how;
(4) Algorithm refinement to enhance convergence and/or decrease running time; and
(5) Whether to use only locators known to be fixed (briefly,known-fixed).

Choice of weights. The ranges considered in this study are generally several orders of magnitude larger than their measurement errors. Consider the locator circles $C_i = 0$ representing points whose distance form the estimated position of locator i is exactly the measured range r_i. Within the area of uncertainty of the position being estimated,each locator circle is represented by a short arc that is almost a straight line. For the LS estimate, we wish to find that point which minimizes the sum of squared errors, i.e., the sum of squared distances to the nearly-straight lines. But the positions of the lines are not equally well determined, so we should use weights. These weights should be inversely proportional to the uncertainties in the positions of the lines. There are two sources of uncertainty: the range measurement error, which we take to be Gaussian with mean zero and variance σ_R^2, and the error in the estimated position of the locator (measured in the direction of the point being located), which one can assume to be Gaussian with mean zero and variance σ_P^2. (Only the component of error in the direction of the point being located matters, because with the small errors involved, the error in the orthogonal direction does not appreciably affect the position of the line.) Now σ_R^2 has been taken as an input parameter, but, in practice, it could be estimated while the system is in operation. On the other hand, σ_P^2 must be estimated. For a fixed locator, whose position has been surveyed (either independently or by a modification of the system under discussion), an estimate of the variance should be in hand - and it may well be different for different directions. But a moving locator's position will be known to varying degrees of accuracy, depending on how many range measurements have been available, which users made the measurements, the geometry in effect,and the smoothness of the locator's path (which determines how useful tracking or smoothing will be). Again the uncertainty may be different in different directions. Ways can be devised to keep track of the uncertainty in three (orthogonal directions, but that was not done for this study. Instead, an estimate of the position uncertainty is obtained each time a new estimate of the locator's position is found. The method used is analogous to

estimating the residual (observational) variance from a linear regression. The newly estimated position plays the role of the regression line, and the discrepancies between the measured ranges and the distances to this estimated position play the role of deviations of the observations from the regression line. Proceeding as if the range measurements all had the same variance, one obtains an estimate of that variance by summing the squared deviations and dividing by (N-2) in two dimensions or (N-3) in three dimensions, where N is the number of observations (ranges) available. Now one uses the fact that for N ranges, from directions properly spread around the compass, the resulting position estimate (using LS or LSS) will have a circular distribution with variance $2s^2/N$, where s^2 is the variance of each measurement (which has just been estimated). The phrase "properly spread" means that for each direction from which a range measurement is taken, another measurement is taken in the perpendicular direction; in three dimensions, the measurements must be taken in sets of three mutually perpendicular directions to preserve circularity (i.e., sphericity) of the distribution.

The weights used for each range measurement are then taken to be the reciprocals of the quantities ($\sigma_R^2 + \sigma_P^2$) for the fixed locators, and ($\sigma_R^2 + 2s^2/N$) for the moving location where each value of s^2 was calculated the last time that locator was a locatee.

There are two considerations which should be mentioned here. First, locators are not in general split so nicely along perpendicular lines. Does this matter? Of course it does, in the sense that the uncertainty will no longer be the same in all directions. On the average, however, things will even out: if azimuths (from locatee to locators) are uniformly distributed from $0°$ to $180°$, the average (or expected) variance in a given direction turns out to be the value derived above. The second consideration is: What if the estimated variances for the locators are wrong? Two consequences follow: (a) to the extent that the true variances are different multiples of the assumed values, the estimate of position is less than optimum (because the true relative variances should be used to get the estimate); and (b) to the extent that the variances are (as a group) larger (or smaller) than assumed, the estimated variance of the locatee's position will be too small (or large, respectively). Point (a) is not likely to be important, since it would take very large discrepancies to affect the estimate noticeably. Point (b), on the other hand, should at least be investigated. One technique is to act as though the true variances are proportional to the estimated values, with an unknown constant of proportionality (say c). Then c can be estimated, and used to produce a fair variance estimate. Specifically, if each range observation has variance $c\sigma_i^2$ (σ_i^2 = location variance + range measurement variance), then the best estimate for the locatee can be found without considering c. If ϵ_i represents the discrepancy between the calculated range and the corresponding measured range from this position to the assumed position

of the i-th locator, the $\frac{1}{N-2} \Sigma \epsilon_i^2/\sigma_i^2$ is the (multiplicative) correction to

be applied to the variance estimate already given. (If $c \sim 1$, this value should also be ~ 1 .) This modification has not yet been made to the algorithms.

Random stopping and starting was implemented, to investigate the error
behavior when the composition of the set of locators changes. For this
purpose, it was decided that when a user stops, ten estimate of his position
would be averaged to get a good estimate before using his as a locator.
The accuracy of this estimate can of course be estimated from the consistency
of the individual estimates.

A series of 2-dimensional runs was done, with 30 units. Initially, the
fixed units were locators. During the runs units became locators as
soon as they had been fixed for ten location-operations, and ceased to
be locators as soon as they began to move. Average times moving and
fixed were set at 15 min., so that about 15 (half the units) were stopped
at any one time. The range errors were Gaussian with mean zero and variance
36 m^2. With a 30-second cycle time, 2 and 1/2 hrs. of operating time was
simulated. The overall average error was 3.2m, and the maximum error was
3.5 m; in the last set of 5 cycles, the average was 3.5m, and the maximum
was 11 m. Next, a run using longer cycles (90 s instead of 30 s) produced
errors of 3.8 m and 50 m overall, 5.7 m and 33.8 m for the last 5 cycles.
With 15 minute average "fixed r moving" times and 1 and 1/2 min. cycles,
many units didn't stay stopped long enough to be much use as locators; this
run was therefore repeated with the average times fixed and moving set at
45 min. The error figures then came back down toward those of the first
run: 3.4 m and 33.3 m overall, 5.9 m and 22.5 m for the last 5 cycles.
These values represent some improvement over corresponding values obtained
using equal weights, but the differences were not striking.

Choice between 2- and 3- dimensional algorithms. Clearly, one would not
use a three-dimensional algorithm if all units were known to be on a given
plane. Just as clearly, one needs to use a three dimensional algorithm when
the geometry is far from planar, at least when the measured ranges are the
only data available. A test had to be devised, to determine which algorithm
to use for each location operation. The problem was complicated further by
the availability of independent estimates of altitude differences between
units, which could be combined with the measured ranges to estimate.

(as $[r^2 - (\Delta z)^2]^{1/2}$) the planar ranges - i.e., the distances between the
projections of the units' positions onto a horizontal plane - which could
then be used with a two-dimensional algorithm to find (x,y) coordinates of
the locatee; the altitude differences can also be used separately, of course,
to estimate the locatee's altitude (with respect to a reference plane.)

Thus there were three choices, which will be referred to as 3D, 2D and
2 DSH (for 2D after slant-height reduction). The choice does matter: if
all the locators are in a plane (more or less), and the locatee is off the
plane, then the least squares problem in 3D will have two local minima -
one near the true position, and one located symmetrically on the other
side of the plane. There will also be a saddle point in some cases, located
roughly between the to minima, at which the derivatives are also zero.

-564-

It is not likely that the algorithm will converge to the saddle point if enough iterations are done. Also, if the starting point for the algorithm is chosen close to the proper solution, it should converge to that solution. However, the problem can become rather illbehaved: all along the line connecting the two minima, the penalty-function can be nearly constant, and then the accuracy of the solution suffers. Furthermore, for low-flying aircraft (up to several hundred meters), the algorithms did occasionally converge to the saddle point or to the below-ground solution, because only 4 to 8 iterations were allowed.

Certain intuitive expectations were verified by runs using all three algorithms (2D, 3D, 2DSH). It was found that when differences in terrain altitudes were less than the typical errors in the independent height measurements (and no aircraft were involved), it was best to treat the ranges as 2D ranges directly - i.e., assume all the users were coplanar. It was also found that for 2DSH, height errors of 15 m (standard deviation) did not significantly affect the accuracy of the (x,y) position determination - i.e., the position errors were comparable to those obtained with all true heights set to zero, the same range errors applied, and a 2D algorithm used.

A systematic study was done specifically to decide when to use 3D; this study is described in Sec. C below.

Smoothing and filtering. There are three levels at which filtering or smoothing might be used: (1) fitting a smooth path through the last two or more estimated positions, to obtain a first estimate of current position, which will serve as the starting point for an iterative method of position estimation; (2) fitting a smooth path as in (1), in order to provide an estimate of current position when there are not enough ranges to use the algorithms; and (3) using some sort of "filtering" rule which combines current range information with previous position values internally to produce its position estimate. The first level is used in these algorithms, but since the final estimate is insensitive to the quality of the starting point, only the last two position estimates of the locatee are used; they are extrapolated linearly to obtain the starting point for estimating the current position. The second level is also used, with refinements; when there are no range measurements, one has only the extrapolated position, but when there are one or two range measurements, one can modify the extrapolation accordingly. Many rules can be devised. The ones used in these algorithms are rather simple: with 2 ranges, reduce to planar ranges and use 2D; with one range, modify the extrapolation to reduce the discrepancy between the measured range and the calcultaed distance (from locatee's extrapolated position to locator), the amount of modification to be based on the relative sizes of the extrapolation estimate variance and the range-plus-position-error variance. This modification is in the spirit of the relocation described in the next paragraph. The third level of filtering has not yet been implemented, except for the z-coordinate when using 2DSH; there, an exponential filter is used, with constant 0.6. (I. e., the current smoothed value of z is taken to be $(0.6)z_{old} + (0.4)z_{est}$

where z_{est} is the new reading, and z_{old} is the last smoothed value.) This

reduced the variance of the z-estimate by about 30 percent. Various
polynomial smoothing techniques were tried on the x and y estimates,
but the best of these (which involved a quadratic curve fit through the
last five points, in x and y separately) made only slight improvement
in accuracy, and then only when the extrapolation was weighted one-tenth
and the position estimate from current range values was weighted nine-
teenths in obtaining the final estimate. No other filters were tried,
but through the cooperation of David Plutchak of General Dynamics, a set
of range data was run through these algorithms and through his filter pro-
gram, and the results compared. The results were predictable: when the
path is smooth enough and the data are obtained frequently enough, the
Kalman approach does better, while for sufficiently erratic paths or sufficiently
infrequent data, the algorithm presented here is more accurate. Details
on these runs will be found in Sec. E.

REFINEMENTS OF THE ALGORITHMS. Refinements consisted of two basic types:
miscellaneous aids to convergence, allowing more accurate estimates for a
given number of iterations; and a scheme called "relocation". The miscellaneous
aids included halving the step size when an iteration would result in an
increase rather than a decrease in the objective function, decreasing the
step size when a step would tend to return to the previous point (as when
the iteration is bouncing from one side of a valley to the other, working
down the slope); and increasing the step size substantially when several
steps are basically in the same direction and are not decreasing in size
fast enough. Since there are many ways to do these things, it does not
seem appropriate to go into detail.

Relocation refers to an attempt to adjust the estimated position of the
locators slightly, to decrease the discrepancies between measured and
calculated ranges. The impetus comes from the realization that such
discrepancies come not just from errors in the locatee's estimated position,
but also from errors in the estimated positions of the locators. The
amount of adjustment was calculated from the relative uncertainties of the
locatee and locator, in such a way that if a locator's position is much
better known than that of the locatee, little adjustment is done to that
locator. Specifically, the amount of adjustment is a fraction of the ratio:
(locator's variance)/(locator's variance plus locatee's variance). Several
values of the fraction were tried. The optimum value was zero unless the
algorithm was run for many cycles, in which case very small amounts of
relocation (fraction = 0.1, the smallest value tried) seemed to help slightly.
(But by this time, the errors had already grown to unacceptable levels.) A
better alternative, not evaluated in this study, would seem to be to adjust
the positions of all (or all but two or three) of the fixed locators simul-
taneously, using several successive measurements on each range, using a
criterion like minimum (weighted) mean square range-discrepancy, possibly
combined with a criterion of minimum total displacement of positions; this
latter device would tend to prevent translations or rotations of the set
of all unit positions without changes in relative position.

CHOICE OF LOCATORS. As mentioned in section A above, not all locators could
be allowed to move. The difficulty can be easily seen in the following
examples: suppose the velocity of every unit was suddenly changed by adding
a certain vector, say 5 meters per second in a given direction. Since the

relative positions (i.e., the ranges between units) are unchanged, the algorithm would never notice the change in velocities. Similarly, if a rotation were applied to the entire cluster of units, this also would go unnoticed. These are special cases, of course, but the same principle applies to the average velocity change; these errors may (and do) accumulate into a large translation velocity which is entirely spurious. It does not help to have units fixed, unless the algorithm is allowed to use the fact that they are fixed: i.e., they must be known fixed. Thus it is necessary to have some locators whose positions are not routinely updated. Furthermore, if the fraction of locators so fixed is too small, the algorithm will tend to let the corresponding ranges deviate in order to fit the other ranges, with the result that the moving cluster will acquire a spurious velocity anyway, with the ranges to the fixed units eventually being rejected as "outliers". Thus a reasonable fraction of the locators should be fixed.

C. AIRCRAFT TESTS. A systematic study was done to determine when to use 3D , and when to use 2DSH . For this study, an aircraft was to be located, using ground stations as locators. Four, eight and twelve locators were used, and many location operations were performed on aircraft at various elevations above various ground positions. The range-error sigma was set at 4 m and 8 m; the height error sigma was set at 15 m and 25 m. (The true value of the range-error sigma was thought to be about 6 m; that of the height difference was 20 m , corresponding to a height error sigma of 14 m .) Finally, runs were duplicated with the locator's position perturbed by random errors. Twelve positions were randomly chosen on a 15km x 15km square; eight of these were randomly chosen for the eight-locator trials; and four of these for the four-locator trials. The set of eight covered the 15km by 15km region; the set of four were contained within an 11km by 11km square. Eleven aircraft positions were chosen eight within the 15 km square, to which unit numbers 1 to 8 were assigned; one about 5 km outside, one 20 km outside, and one 50 km outside, numbered 9, 10, and 11 respectively. Eight altitudes were used in each position: from 300m to 2400m in steps of 300 m. Fifty trials were done for each position, each altitude, and each algorithm variation, with trials matched across algorithms for range and height errors. Errors were averaged across each set of 50 trials, and these averages compared for consistency and analyzed for meaning.

Now the height above which it pays to use 3D depends on several factors: number of locators, error magnitudes (both range errors and height errors), and geometry. But geometry, unlike the other factors mentioned, is a complex factor not amenable to summary by a single number: it involves the locations of a whole set of locators relative to the aircraft location, the error magnitudes and the elevations of the lines of sight above horizontal, which are hereafter called the look angles. Thus it was not surprising that the results do not clearly specify exact rules for which algorithm to use. The expected general trends were clearly in evidence: an aircraft having only low look angles, a small number of locators, a large range error sigma, a small height error sigma, all tended to favor the use of 2D methods. But the crossover altitude was not sharply defined - i.e., in a neighborhood around the crossover point, both methods did almost equally well. (This allows simplification of the rules of thumb with which one must eventually work.)

-567-

A reasonable set of rules seems to be the following: with as many as 12 locators, use 3D only when the aircraft is inside (or near) the area covered by the locators, and above 1000 m . With as few as 4 locators, use the same altitude cutoff, but be more stringent on geometry: the locators ought to span more than 90° of azimuth at the (ground projection of the) locatee. Examples of the results upon which this set of rules is based will be presented in the next paragraph, in lieu of the nearly undigestible mass of data which was obtained.

SPECIFIC RESULTS (SELECTED). The best conditions tested are of course $\sigma_R = 4$ and 12 locators. For this situation (with perturbed locator coordinates) the results were essentially the same for $\sigma_h = 10$ as for $\sigma_h = 25$. They are summarized below

	Unit Numbers			
	1-8	9	10	11
x, y error, 3D	< 4	5	15	35
x, y error, 2D	< 4-5 (4-10 for $\sigma_h = 25$)	5	7	13
z error, 3D:				
Altitude 300m.	15-56	360	852	633
600m.	7-21	55	308	1400
1200m.	4-10	24	459	625
1800m.	4-6	17	550	638
2400m.	2-6	13	117	600

Table 1. Average errors in meters, for 12 locators, $\sigma_R = 4m$.

At the other extreme is the case with 4 locators and $\sigma_R = 8$. Again, the height σ makes a little difference. The errors are given in the following table.

	Unit Numbers				
	4-8 (inside)	1-3 (nearby)	9	10	11
x,y error, 3D	8-13	13-20	17	40	75
x,y error, 2D	7-11	7-11	12	27	50
z error, 3D:					
Altitude 300m	43-117	224-612	654	618	393
600m	17-51	90-258	214	1229	1045
1200m	11-25	44-99	117	468	1833
1800m	8-19	25-62	61	617	1173
2400m	6-15	21-48	53	312	988

Table 2. Average errors in meters, for 4 locators, $\sigma_R = 8m$.

Note: The x, y error figures for 2D, for units 1 - 8 only, grow slowly with altitude, and (at the higher altitudes) grow with σ_h; at 2400 m altitude, with $\sigma_h = 25$, they are 10 - 19 instead of 7 - 11.

The results for 8 locators fell approximately between these results, except for units 10 and 11. For these, they provided further evidence that the great variability in average z-error across the different altitudes is indeed simply a reflection of the uncertainty of z-estimates under poor geometry.

D. TRIALS ON DIGITIZED ACTUAL TERRAIN . A portion of the area near Boston, consisting of rolling terrain, was chosen for testing the algorithms and was entered into the computer in digitized form. Eighty-three units were placed on the terrain: 44 moving, more or less in the same general direction, and 39 fixed (scattered among 13 different locations). Many trials were made, with various sets of units reporting. In each case, the trial consisted of 240 cycles (two hours); ground units were located every 30 sec. except that ground locators were located every 5 seconds. Aircraft were located every 5 sec., except that aircraft locators were located every second.

Several ground-units-only results will be presented first, then several runs with aircraft. (Seven aircraft were used: however, it was necessary to put them all at one altitude, which was chosen to be 300 m.) Errors are reported as x, y-errors - i.e., errors in the horizontal plane - and

z-errors. The range errors again followed a normal (Gaussian) distribution with mean zero and standard deviation 6 m; in addition, a random 1% of the range values were inflated by 7 m to reflect a positive bias due to occasional masking of the direct signal and acceptance of a reflected signal.

With all 39 fixed units reporting, the average x,y-error was 4.3 m, the maximum was 33.4 m; but quite a few units were lost (i.e., position estimates were not obtained by the algorithms) fairly often.

With all 83 units reporting, the average and maximum errors were 4.8 and 101 m; both values were due to large reported errors for units that were not even located when only fixed units reported.

When the number of locators was reduced to 21, care had to be exercised in choosing them. With one representative of each of the 13 sets of colocated fixed units reporting, and 8 moving units (chosen to be those which could "see" many of the other units), the average and maximum were 5.9 and 212; when one of these fixed units was replaced by a different moving one, the average was about the same but the maximum decreased to 140.

The 3 best aircraft runs were done incorporating a set of rules for deciding which algorithm to use, and with exponential smoothing of the z-coordinate. The rules were:
(a) For ground units, use LSS3D if there are at least 6 ranges of which at most 4 are at look angles below 30°; otherwise use LSS2D, using only the ranges below 30°.
(b) For aircraft, use LSS3D if there are at least 6 ranges, and either there are at least 3 ranges above 20°, or the aircraft is high enough (height greater than 16 km/(4 + number of ranges)); otherwise use LSS2D on all ranges.
(c) Before either of these rules is implemented, check whether there are at least 4 ranges from fixed locators. If so, use only the ranges from fixed locators.

Under these rules, incidentally, LSS3D was never used.

The table below presents the average and maximum errors for both ground units and aircraft, for x, y plane and for z, for each of the three runs. Each run represents 7080 location operations on ground units, and 22,320 location operations on aircraft.

	Ground units				Aircraft			
	Averages		Maxima		Averages		Maxima	
	x,y	z	x,y	z	x , y	z	x,y	z
6 air, 7 fixed ground, 3 moving ground locators; smooth z for aircraft only	7.3	11.0	64	51	7.6	8.0	93	42
Same locators; smooth z for all moving units	7.1	7.7	58	38	7.5	8.1	88	42
11 fixed ground, 5 aircraft locators; smooth z for aircraft.	5.2	11.1	73	57	6.2	8.0	84	41

Table 3. Average and maximum errors in meters for several
runs involving aircraft.

E. TRACKING AND FILTERING. As mentioned above, exponential smoothing was
applied to the z-coordinate; also, polynomial smoothing was implemented for
each of the other two coordinates. These techniques were compared with a
9-state (position, velocity, acceleration) Kalman filter technique
developed by General Dynamics. The comparisons are described below.

Two scenarios were used. Each had six fixed locators on the ground,
tracking an aircraft at 400 m altitude. One had good geometry; the other
had relatively bad geometry. True paths were calculated, and true ranges
to each locator; these ranges were then perturbed with independent
Gaussian errors having mean zero and standard deviation 6 m. Finally,
these perturbed values were treated as data, the estimated positions were
calculated by the various methods, these estimated positions were compared
with the true positions to get the true errors, and the errors compared among
the different methods to see which was most accurate. (Another step, taken
for the Kalman estimates, is described below.)

The first path, hereafter known as the circle path, consisted of eight turns
around a circular path of 1 km diameter, with a "drift" of 100 m per revolution,
at a speed of 225 km/hr . The geometry was good: 80 percent of the time
was spent within the convex hull of the set of six locators, which were
arranged in a more or less elliptical configuration (see Fig. 1); there was
always at least one locator close enough that the look angle was greater
than 30° , and occasionally there were four such, with an average of 2.14;
the aircraft was never more than 1.8 km from any locator. The other path,

-571-

hereafter called the back-and-forth path, consisted of starting from
stationary, traveling 12.7 km in a straight line at 212 km/hr , hovering
for 72 s., returning, and again hovering for 72 s.; then repeating. Almost
2 cycles of this path were done. The starts and stops were kept gradual,
so as to not exceed the value of acceleration for which the Kalman filter
was designed. The locators were scattered along the path, but not near the
extremes: at the ends, the look angles were all less than 6° , and the
ranges varied from 2.7 to 10.2 km. (See Fig 2.) Only 7 percent of the
time was there one look angle above 30°; there was never more than one.
(It should be remembered that 3D is not recommended for situations like this.
However, for this test, no auxiliary source of height information was assumed.)
The results will be presented first for smoothing versus unsmoothed LSS,
then for Kalman filter vs. LSS.

For many of the runs, the smooth paths were "roughed up" a bit, by perturbing
each coordinate of each point with independent Gaussian errors; the standard
deviation of the errors was 3 m. This was then treated as the true path.

Runs were made with the smoothing parameter set at 0.5, 0.8, and 0.9, at
one-second data interval. The best value was 0.9. For this run, the
smoothed track was clearly better than the LSS solution: the numbers of
times the smoothed LSS had smaller errors than the simple LSS were 269 out
of 400 and 669 out of 1000, which are significant at 7 and 11 sigmas, for
the circle and the back-and-forth run respectively. When in addition no
elimination of apparently bad ranges was allowed, the levels were 281 out
of 400 (8 σ's) and 672 out of 100 (11 σ's). (These are with reference to
the x,y error. For x, y, z error, the figures were 286/400 and 614/1000,
both at the 8σ level, with elimination allowed; the run without elimination
was not done for x, y, z errors. In addition, for these runs, the path
perturbations were not made, so the conclusions are tentative at best.)
Other trials were done, with different random errors, with essentially the
same results: the most different showed differences significant at 6σ's
and 4σ's respectively. One comparison was made for the perturbed path, with
no elimination of ranges, and smoothing parameter set at 0.9. The smoothed
track had smaller errors in 261 of 400 tries (6σ's) for the circle, and in
651 of 1000 tries (10σ's) for the back-and-forth path.

The more interesting runs were those comparing Kalman filtering with LSS.
(Due to lack of time, it was not possible to make as many runs as desired,
so that the filter was not compared with smoothed LSS.) These were run
with data intervals of one s. and five s. The differences were marked, as
one might expect. At the five s. interval, the Kalman x,y errors were larger
in 33 of 40 trials for the circle, and in 32 of 58 trials for the back-and-
forth path. (These figures are significant at 4σ and 0.8σ respectively.)
At one s. interval, results were reversed: LSS errors were larger in 25 of
40 trials for the circle, and in 44 of 58 for back-and-forth. (Significant
at 1.8σ and 3.8σ .)

The maximum errors, both (x,y) and z, are given in the accompanying Tables.

	Circle path (40 trials).	Back-and-forth (58 trials)
LSS	12.6	25.0
Kalman, 1 s.	11.9	24.4
Kalman, 5 s.	80.5	40.1

Table 4. Maximum (x, y) errors in meters.

	Circle path	Back-and-forth
LSS	15.6	339
Kalman, 1 s.	13.7	100
Kalman, 5 s.	154.1	142

Table 5. Maximum z-errors in meters.

On the basis of this admittedly thin evidence, one might suggest that Kalman filtering does not help unless the time interval between data points is 1 s. or less; and even at 1 s., the difference is not over-whelming. (Remember too that with independent height information, the reduction to a 2D method will give considerably better figures for aircraft below 500 - 1000 m altitude.)

The Kalman filtering program carries along an internal estimate of error, in the form of an estimate of the covariance matrix of the state variables (which include the position coordinates). It is to be hoped that the variances of the coordinates, as carried in this matrix, would accurately reflect the uncertainty of the corresponding positions, since these variances are used to weight the measurements. If so, the actual errors ought to correspond to those variances. Since the x- and y-variances are printed out, and since for these tests the true errors are known, one can check whether the internal estimates are accurate. This has been done. For both paths, the actual errors were compared with the printed variances. The numbers of errors within k times the corresponding sigma were tabulated for k = 1,2,... .

It is not proper simply to compare these frequencies with the expected frequencies for a Gaussian distribution with known sigma, since all we have is an estimate of sigma. One must use the t-distribution. And even so, there is a difficulty: one does not know which t-distributions to use, since the "degrees of freedom" (a parameter of the t-distribution) varies from trial to trial in an unknown way. An approximate treatment was done, by comparing the frequencies with various t-distributions at various scale factors, to find that distribution which best fit the tabulations. (There will be at least 3 degrees of freedom, since the six basic range measurements are used to fit three parameters, the coordinates, leaving three degrees of freedom. How many more there will be depends on the accuracy of the extrapolation from past positions, which in turn varies with the geometry and the data frequency; there might be as many as 6 to 9 degrees of freedom with good geometry and frequent data.) Since the Kalman filter was less accurate than LSS for 5-second data intervals, no tabulations were done for those tracks. The results were as follows:

There were 28 errors greater than 1σ, of which 6 were greater than 2σ, and none greater than 3σ, for 80 points on the circle path; the expected numbers for a t-distribution with 8 degrees of freedom are 28, 6, and 1. There are 39 above 1σ, 9 above 2σ, and 2 above 3σ, out of 116 points on the back-and-forth path; the expected numbers are also 39, 9, and 2. This is indeed a surprisingly good fit.

One more point should be brought out regarding the comparison of objective error from simulation trials with estimated errors based on real trials. The errors quoted in the simulation studies done for this project are errors in the x,y-plane. If, as seems reasonable, x- and y-error are independent, approximately Gaussian, and equally variable, then the mean absolute (x,y) -error will be 1.25 times the standard deviation of the x-error. Therefore, the estimated standard deviation as given by the Kalman program should be inflated by a factor of (1.25) before comparison with the average (x,y) -error as given by the simulation trials.

SUMMARY OF RESULTS. This section can be reasonably summed up as follows:

(1) The algorithm of choice is Least Squares Squared; it is at least as accurate at Least Squares, and takes less computation. (This assumes that errors are approximately Gaussian-distributed, and that outlier rejection techniques are used to eliminate bad measurements.)

(2) The distribution of errors, not just its standard deviation, affects the accuracy of the algorithms (and the proper choice of algorithm).

(3) At least a reasonable fraction of the locators must be known to be fixed. A reasonable fraction may be four out of twenty, but it helps to have a larger fraction.

(4) Using weights which reflect the accuracy of the locators' position estimates does increase the accuracy of the algorithm, but not very much

- at least for the runs made, for which the variance of the locators' positions varied over a two to one range. This is because the variance of the range measurement is added to the position variance before calculating weights. If the range variance were much smaller (than 36 m^2), weights would make a larger improvement.

(5) An example of the accuracy attainable: With 30 units spread over an area 20 km by 20 km, in 2 dimensions, with units starting and stopping, and only stopped units being used as locators (an average of 15 stopped at any time), the average position error throughout a run simulating 2 and 1/2 hrs. of real time, consisting of 300 cycles, was about half the range error standard deviation.

(6) For situations which are not far from planar (e.g., where many of the locators are at look angles less than 30°), an independent height measurement of reasonable accuracy can increase considerably the accuracy of the (x,y)-position estimate, by allowing the calculation of projected ranges in the (x,ȳ)-plane and the use of a 2-dimensional algorithm.

(7) The starting point for the Least Squares Squared algorithm is noncritical.

(8) There are indications, which could be verified by further analysis, that the technique of reducing slant ranges to horizontal (projected) ranges and applying a two-dimensional algorithm is very useful when locating low-flying aircraft with locators on the ground. (It requires a separate estimate of the aircraft height, of course.) Two factors are important: when the height is small compared to the ranges, then (1) the projected ranges are relatively insensitive to height measurement errors, and (2) any 3-dimensional method will have relatively low accuracy for the (x,y)-coordinates, and even lower accuracy for the z-coordinate. Very roughly, the 3D version is about as accurate as the 2D version when several of the look angles (angles between the horizontal plane and lines connecting the aircraft with the locators) are about 20° or more, when there are 12 locators and the range errors have a standard deviation of 4 m ; on the other hand, with only 4 locators and a standard deviation of 8 m, the aircraft had to be surrounded by the locators, and the angles had to be on the order of 30° before the 3D method approached the accuracy of the 2D method.

(9) Smoothing (using past positions as well as current range data to estimate current position) seems to help significantly, only on the z-coordinate of aircraft; even this conclusion depends, of course, on how stable the altitude of the aircraft really is.

(10) Use of a 9-state (position, velocity, acceleration) Kalman filter seems to give more accurate position estimates, when data are available at least every second (that is, several range measurements every second) and the path of the unit being located is quite smooth - i.e., variations from a smooth curve which are on the order of half the magnitude of the range errors. Otherwise the LSS3D algorithm will be more accurate. (The LSS2D algorithm on projected ranges was not compared with the Kalman filter approach - but it is felt that it might well produce more accurate results, even if the height information were also fed to the Kalman filter.)

-575-
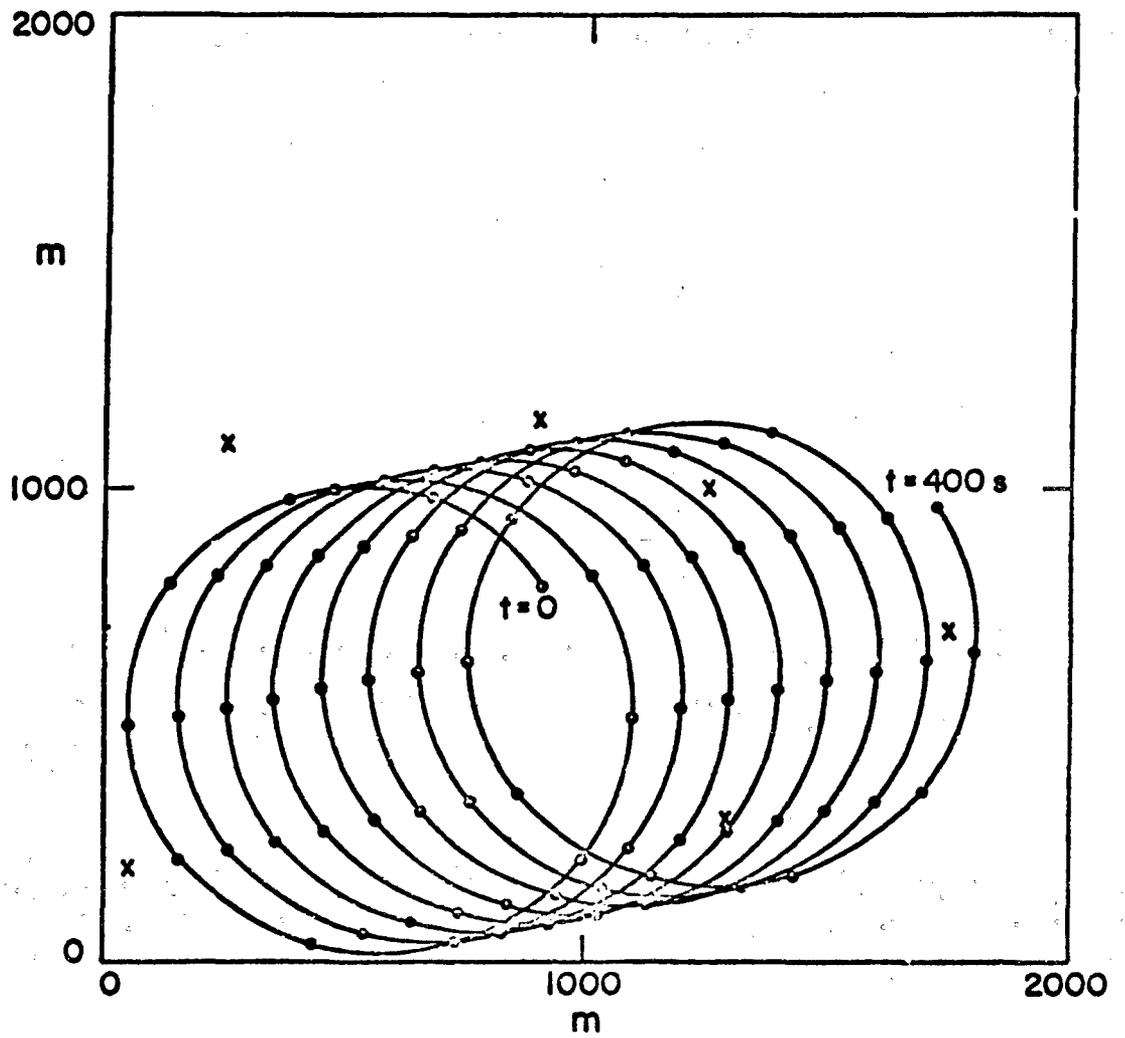
Figure 1. The circle path. (x indicates the positions of the locators.)
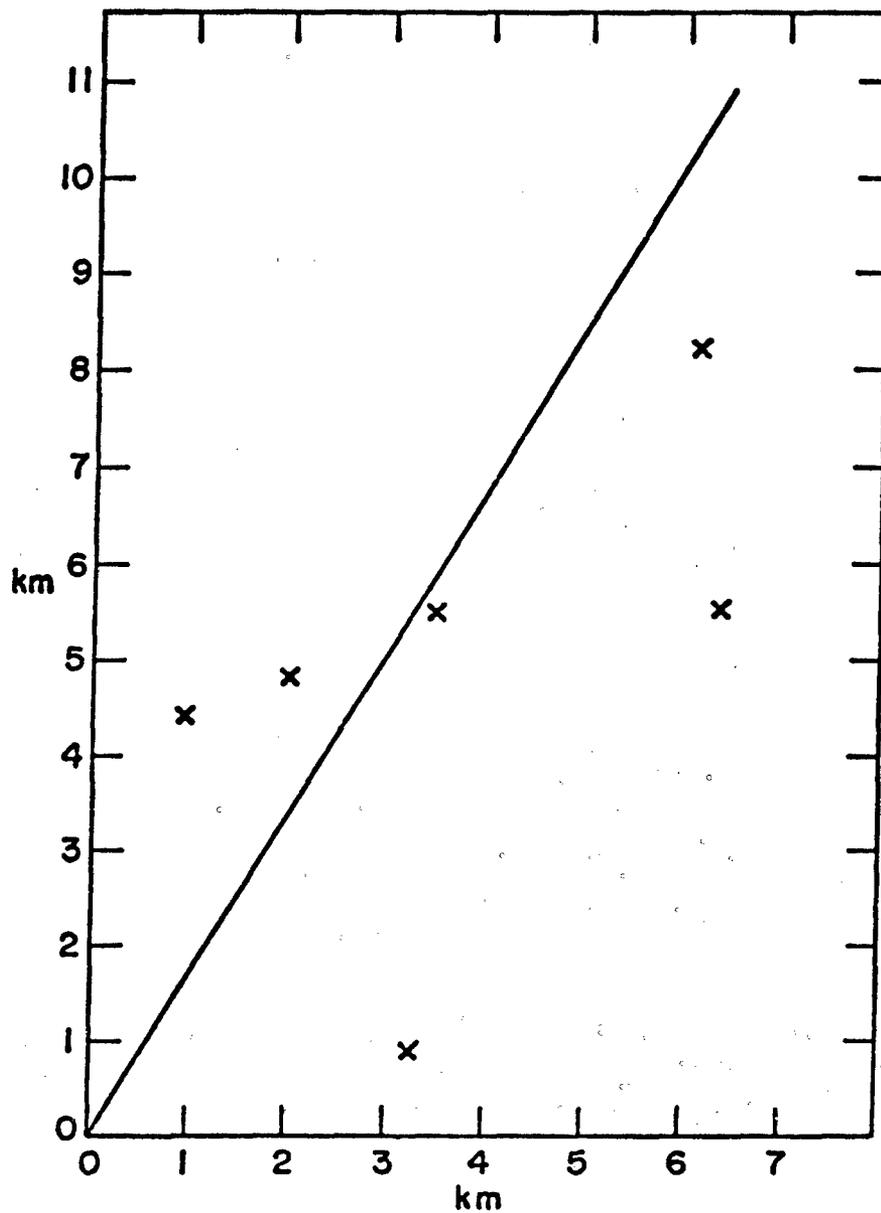
Figure 2. The back-and-forth path. (x indicates the position of the locators.)

APPROXIMATE LOWER CONFIDENCE LIMIT ON THE CIRCULAR ERROR PROBABILITY (CEP) IN THE CASE OF UNEQUAL VARIANCES

EDMUND H. INSELMANN

HEADQUARTERS, U. S. ARMY MATERIEL COMMAND
ALEXANDRIA, VIRGINIA

INTRODUCTION AND SUMMARY

This paper is concerned with the interval estimation of the Circular Error Probability (CEP) which is used as a measure of dispersion of rounds fired at a target. Most of the literature on this subject deals with circular normal distribution. In this case, the CEP problem works out easily and the lower confidence limit has been considered in other papers, e.g., Inselmann and Granville [1]. The attention here will be restricted to the case where target errors (in azimuth and range) are independent Gaussian variates with zero means and unequal variances. The main result herein is limited to the case where the ratio of the standard deviations is between .5 and 2. Probabilistic questions concerning the CEP have been discussed by Grubbs [2]; this work was based on a paper by Patnaik [3]. Other related work was done by Moore [4]. Both Grubbs and Moore's papers involve CEP's which are linear combinations of the variances. These results can be used to obtain point estimates but have not been used in the construction of the confidence limits. The statistic suggested for estimation, testing hypothesis, and calculating the confidence interval will be the geometric mean of the sample standard deviations.

Preceding page blank

PROBLEM STATEMENT

The object is to determine the lower confidence limit of the CEP, where CEP is defined as the radius of the circle about the origin that includes half of the probability in the plane. This of course depends on the distribution of the probability mass, which in this case is bivariate normal with zero means and no correlation. There are now two cases to be considered, equal and unequal variances. The solution for the equal variance case is well known [2]. This paper forwards an approximate solution for the lower confidence limit for the case of unequal variances. To make things more mathematical, consider the bivariate random variable (X, Y) with $EX = EY = EXY = 0$, $EX^2 = \sigma_1^2$ and $EY^2 = \sigma_2^2$. The joint density of (X, Y) is taken as

$$(1) \qquad f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{x^2}{2\sigma_1^2} - \frac{y^2}{2\sigma_2^2} \right]$$

The x and y are taken to be the azimuth and range components of the projectile impact on the target. Recall that the object is to find the radius of the circle that has a probability .5. So the interest is in the distribution of $R = \sqrt{X^2 + Y^2}$. The density of this is well known and is given by

$$(2) \qquad g(r) = \frac{r}{\sigma_1\sigma_2} I_0\left[\frac{r^2}{4\sigma_1^2}\left(\frac{\sigma_1^2}{\sigma_2^2} - 1\right) \right] \exp\left[\frac{-r^2}{4\sigma_1^2}\left(\frac{\sigma_1^2}{\sigma_2^2} + 1\right) \right]$$

where $I_0(x)$ is the Bessel function of the first kind with an imaginary argument. Then the first step in our analysis is to find R_0 such that

$$(3) \qquad .5 = \int_0^{R_o} g(r) \, dr.$$

The integrand here is such that R_0 can't be explicitly determined. Moore [4] has made some numerical calculations for R_0 which give an indication over what range the integrand must be approximated. It can be readily verified that $I_0(x) \approx 1$, for $0 \le x \le 1$. Moore's work [4] demonstrates that argument of the Bessel function in the integrand is less than one provided that the ratio of standard deviations is between .5 and 2. With this restriction one can readily solve the following equation for R'

$$.5 = \frac{1}{\sigma_1 \sigma_2} \int_0^{R'} r \exp \left[\frac{-r^2}{4\sigma_1^2} \left(\frac{\sigma_1^2}{\sigma_2^2} + 1 \right) \right] dr$$

$$= \frac{2\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2} \left[1 - \exp \left(-R_0'^2 \frac{\sigma_1^2 + \sigma_2^2}{4\sigma_1^2 \sigma_2^2} \right) \right]$$

The solution for $R_0'^2$ then is

$$(4) \qquad \left(R_0' \right)^2 = \frac{-4\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \ln \left(1 - \frac{\sigma_1^2 + \sigma_2^2}{4\sigma_1 \sigma_2} \right)$$

Expanding the log term in power series and reducing, Eq. (4) can be written

(5) $$\left(R_0^{\prime}\right)^2 = \sigma_1 \sigma_2 \left[\frac{-1}{t(\alpha)} \ln (1-t(\alpha))\right] = \sigma_1 \sigma_2 \Phi(\alpha)$$

where

$$t(\alpha) = \frac{1+\alpha^2}{4\alpha^2} , \quad \alpha = \sigma_1/\sigma_2, \quad 1/2 \leq \alpha \leq 2 \text{ and } .5 \leq t(\alpha) \leq .625.$$

Now taking the derivative of Φ with respect to α one has

$$\Phi^{\prime}(\alpha) = \frac{t^{\prime}}{t} \left(\frac{\ln(1-t)}{t} + \frac{1}{1-t}\right) = \frac{t^{\prime}}{t} \left[\frac{t}{2} + \frac{2t^2}{3} + \frac{3t^3}{4}\ldots\right] .$$

Note that the second factor is positive, while the first factor $t^{\prime}(\alpha) = 1/4 \left(1 - \frac{1}{\alpha^2}\right)$ is negative in [1/2, 1), zero at $\alpha = 1$ and positive in (1, 2). Hence,

$$\Phi^{\prime}(\alpha) < 0, \quad \alpha \in [1/2, 1)$$
$$= 0, \quad \alpha = 1$$
$$> 0, \quad \alpha \in (1, 2].$$

Simple calculation shows that

$$1.177 \sqrt{\sigma_1\sigma_2} \leq R^{\prime} \leq 1.2527 \sqrt{\sigma_1\sigma_2}.$$

-582-

The graph of $\sqrt{\phi}$ is
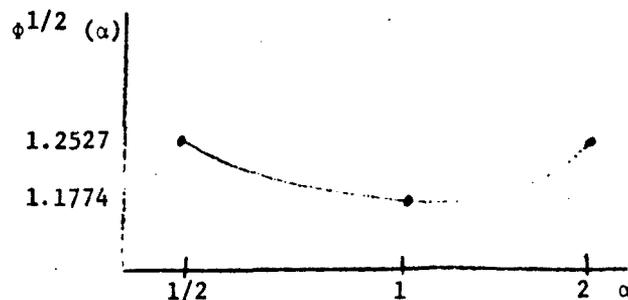
The R_0' that are computed at $\alpha = .2$, 1 and 2 agree with the results computed by Moore [4]. This suggests $\bar{R}_0 = R_0'$ for practical purposes in $(1/2, 2)$. Since the range of ϕ is small it can be replaced by a constant which is taken at the midpoint of the range, namely 1.214. This now gives the final approximation $R_0^* = 1.214 \sqrt{\sigma_1 \sigma_2}$.

THE STATISTICAL PROBLEM

The natural estimate for CEP is then $1.214 \sqrt{S_1 S_2}$ where $S_1^2 = \sum_{i=1}^{n} X_i^2 / n$ and $S_2^2 = \sum_{i=1}^{n} Y_i^2 / n$. The last step needed before constructing the lower confidence limit is to find the distribution of $V = S_1^2 S_2^2$. The density of V is

$$(6). \qquad h(v) = \frac{1}{\Gamma(n/2)^2} \left(\frac{n^2}{4\sigma_1^2 \sigma_2^2}\right)^{n/2} v^{n/2-1} \int_0^\infty \frac{1}{x} \exp\left[\frac{-xn}{2\sigma_1^2} - \frac{vn}{2\sigma_2^2 x}\right] dx$$

$$= \frac{2}{\Gamma(n/2)^2} \left(\frac{n^2}{4\sigma_1^2 \sigma_2^2}\right)^{n/2} v^{n/2-1} K_0\left(\frac{n\sqrt{v}}{\sigma_1 \sigma_2}\right),$$

The integral in (6) is found in the Bateman Tables [5] p146 formula (29), and K_0 is the Bessel function of the second kind. One is now able by numerically integrating this density to find the confidence interval. This is done as follows. Define V_β such that

$$\beta = \int_{V_\beta}^\infty h(x) \, dx$$

Then to find the lower confidence limit on R_0^* the following calculations are made.

$$1 - \beta = \Pr\left\{ \frac{n^2 S_1^2 S_2^2}{\sigma_1^2 \sigma_2^2} < V_\beta \right\}$$

$$= \Pr\left\{ \frac{n S_1 S_2}{\sqrt{V_\beta}} < \sigma_1 \sigma_2 \right\}$$

$$= \Pr\left\{ \frac{1.214 \sqrt{n S_1 S_2}}{\sqrt[4]{V_\beta}} < 1.214 \sqrt{\sigma_1 \sigma_2} \right\}$$

$$= \Pr\left\{ \frac{1.214 \sqrt{n S_1 S_2}}{\sqrt[4]{V_\beta}} < R_0^* \right\}$$

This determines the lower confidence limit for the approximation, R_0^*, of the CEP.

REFERENCES

1. E. H. Inselmann, W. Granville "Circular Distribution Estimation"
   JORSA Vol 15 No 1 (1967) pp 160-165.

2. F. E. Grubbs "Approximate Circular and Noncircular Offset
   Probabilities of Hitting" JORSA Vol 12 (1964) pp 51-62.

3. P. B. Patnaik "The Non-Central χ^2- and F- Distributions and
   Their Application BIOM Vol 36 (1949) pp 202-232.

4. R. A. Moore, "CEP for Correlated Errors With Unequal Standard
   Deviations" The Ramo-Wooldridge Corporation Memorandum GM 41.2-108.

5. H. Bateman "Tables of Integral Transforms" Vol 1 McGraw-Hill.

ON THE VARIATION IN MECHANICAL PROPERTIES
OF LARGE CALIBER GUN TUBE FORGINGS

Peter A. Thornton
Watervliet Arsenal
Watervliet, New York

ABSTRACT. The variations in tensile and Charpy impact properties of nine (9) large caliber gun tube forgings were evaluated by an analysis of variance technique. Utilizing the variance within individual disks as an estimate of error variance, the variation between disks was determined for each forging. In addition, a two factor analysis (cross-classification) was conducted on tubes of similar configuration to define significant variations in mechanical properties between like tubes. At the 5% significance level, real disk to disk variations were determined in yield strength and ultimate tensile strength for all forgings evaluated. However, no significant variation between disks was exhibited by percent elongation or reduction of area for the same forgings. Charpy impact data displayed real variation in five of the nine forgings analyzed. The cross classification revealed that variation in yield strength, %RA and Charpy impact strength was highly significant when the forgings resulted from different positions in the ingot. Conversely, %RA showed insignificant variation, while yield strength and Charpy impact (R.T.) exhibited slightly significant variation when the forgings resulted from similar ingot positions. The latter variation occurred in identical forgings produced from identical size ingots but different heats of steel.

GLOSSARY.

ANOVA	Analysis of Variance
σ^2	Variance (universe)
S^2	Estimate of variance
F	Ratio of mean square in question to error mean square
S.S.	Sum of Squares
D.F.	Degrees of Freedom
M.S.	Mean Square
SN	Serial Number

Preceding page blank

YS	Yield Strength (ksi)
UTS	Ultimate Tensile Strength (ksi)
El	Elongation (%)
RA	Reduction of Area (%)
Cv	Charpy Impact (ft-lb)

INTRODUCTION. During the late 1940's the problem of mechanical property variation was recognized in large gun tube forgings [5]. Since that time better metallurgical controls and advances in thermal-mechanical processing have been utilized in an attempt to correct this condition. However, simultaneous advances in weaponry, such as complex shapes and larger component size have tended to offset these improvements. Consequently, the problem of property variation still exists in large forgings. This fact was recently brought to light during the mechanical property investigation of a 175mm M113 gun tube in 1966 [2]. As a result of the investigation of this particular tube (SN 733), a study of the mechanical property variation in additional 175mm tubes was conducted [4]. The significant findings (conducted on 38 gun tubes) can be briefly summarized as follows: tensile ductility, Charpy impact and fracture toughness (pre-crack Charpy) varied considerably within a tube, within a disc, within a vendor's practice and from vendor to vendor.

A following study then attempted to determine the level and repro-ducibility of mechanical properties in present gun tube material, quenched to a uniform microstructure of 100% martensite and tempered to yield strength ranges to 140-160 ksi and 160-180 ksi [1]. Two important facts resulted from this work: (1) the optimum microstructure obtainable in this material was established by reheat treating small specimen blanks and (2) the variation in mechanical properties (excluding yield strength), found in the reheat treated test specimens was controlled by some factor in the manufacturing process other than heat treatment.

Accordingly, the object of this present examination is to define the variations that exist in mechanical properties of larger size specimens, viz., full size gun tubes. Although similar studies have been conducted, it is felt that the present evaluation is necessary because vendors periodically change their practice. Also, dealing with full size components will permit the appraisal of certain statistical analyses on the data and determine their applicability in future planned studies of this problem.

THEORY. When two or more independent sources of variation operate, the resulting variance is the sum of the separate variances [3]. The two types of errors which arise, when estimating the property of a bulk material are:

1. Errors of sampling (variance denoted by σ_1^2)

2. Errors of analysis (variance denoted by σ_0^2)

These sources of error operate independently and the total variation may be obtained by the addition of the two.

In order to separate and estimate the variances due to testing and sampling an Analysis of Variance (ANOVA) can be conducted with the experimental data. The ANOVA is essentially a method of analyzing the variance to which a response (test measurement) is subject, into its various components corresponding to the sources of variation which can be identified. The details of this method can be briefly summarized as follows:

Suppose there are k samples (disks) and n repeat analyses on each, giving a total number of analyses N = kn. The analytical error is responsible for the variation in the repeat analyses on each sample, and its variance is denoted by σ_0^2. This variance is estimated by:

$$\frac{\text{TOTAL of the sums of squares about the sample means}}{\text{TOTAL of the degrees of freedom}} = \sum_{i=1}^{k} \sum_{j=1}^{n} \frac{(x_{ij}-\bar{x}_i)^2}{k(n-1)}$$

where x_{ij} - individual responses (within disks)

$\bar{x}_i$ - disk mean

Similarly, the sampling error variance denoted by σ_1^2 is estimated by:

$$n \sum_{i=1}^{k} (\bar{x}_i-\bar{x})^2/(k-1)$$

where x_i - disk mean

$\bar{x}$ - grand mean

The sums of squares and degrees of freedom "between disks", "within disks" and "total" may be set out in tabular form called the ANOVA table as below:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Quantity Estimated by Mean Square
Between disks	$n \sum\limits_{i=1}^{k} (\bar{x}_i - \bar{x})^2 = S_1$	$k-1$	$S_1/(k-1)$	$\sigma_0^2 + n\sigma_1^2$
Within disks	$\sum\limits_{i=1}^{k} \sum\limits_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 = S_0$	$k(n-1)$	$S_0/k(n-1)$	σ_0^2
Total	$\sum\limits_{i=1}^{k} \sum\limits_{j=1}^{n} (x_{ij} - \bar{x})^2$	$nk-1$		

The results of the Analysis of Variance can then be tested for significance. This is accomplished by setting up the Null Hypothesis that there is no disk to disk variation ($S_1^2 - 0$). Consequently, two independent estimates of S_0^2 are realized; one from the mean square within disks, and the other from the mean square between disks. To test whether these two estimates differ significantly, i.e., whether they differ by more than can be reasonably explained on the grounds of errors in the estimates, the ratio of the mean square between disks to the mean square within disks is calculated. This ratio (F) is the measure of the variation caused by the effect divided by the variation due to repeat tests. The resultant F value is then compared with a table of variance ratio for the respective degrees of freedom, and a particular significance level. A significant value of F_{calc} ($F_{calc} > F_{table}$) discredits the Null hypothesis and it can be concluded that real variations exist in the property under consideration, from disk to disk.

APPROACH. The mechanical properties of nine (9) full size gun tube forgings from a single vendor, were evaluated transverse to the forging direction. These tubes are identified in Appendix A, along with their respective chemical analyses and heat treatments. Sampling of the tubes was accomplished by cutting 1-1/4" thick disks at approximately 12" intervals along the tube length. This procedure is illustrated on a finished gun tube in Figure 1. Tensile (o.357" dia) and standard Charpy impact specimens were machined from the midwall region of these disks as shown in Figure 2. The tensile specimens were tested in a commercial tensile testing machine and the strain measured during the tests with an extensometer. After testing, % reduction of area and % elongation were determined from the deformed specimens. Impact testing was conducted on standard Charpy V-notch specimens at -40°F and in some instances, at room temperature also. Then, in order to determine if real variations exist in the material along the tube length (disk to disk), an analysis of variance (one-way classification) was conducted on the data from each tube. Also, a two
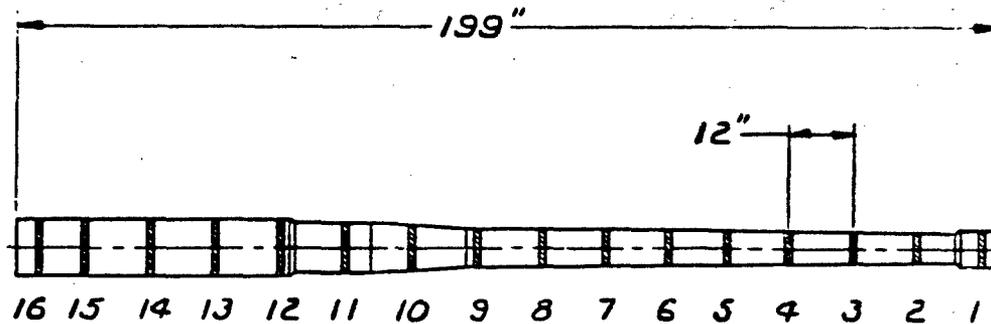
Figure 1. Schematic Showing Test Specimen Sampling
Plan Along Tube Length
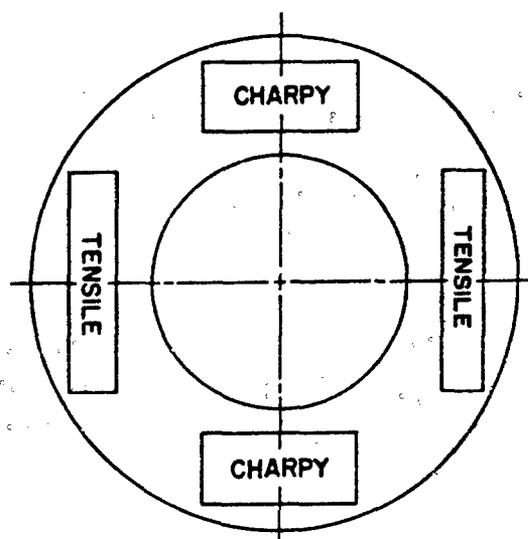
Figure 2. Schematic of Test Specimen Layout Within
A Disk

TABLE I ANOVA 155MM M126 - TUBE 8890

RESPONSE	SOURCE OF VARIATION	S.S.	D.F.	M.S.	Fcalc.	F(5%)	VARIATION
YS	B	543.5	11	49.4	9.0	2.7	Significant
	W	66.1	12	5.5			
UTS	B	250.0	11	22.7	37.7	2.7	Significant
	W	7.2	12	0.6			
%El	B	11.4	11	1.0	1.4	2.7	Insignificant
	W	9.0	12	0.8			
%RA	B	290.9	11	26.4	1.5	2.7	Insignificant
	W	207.1	12	17.2			
Cv(R.T.)	B	24.7	10	2.5	5.1	2.9	Significant
	W	5.3	11	0.5			
Cv(-40°F)	B	10.2	10	1.0	2.3	2.9	Insignificant
	W	4.8	11	0.4			

TABLE II ANOVA 155MM M126 - TUBE 8913

RESPONSE	SOURCE OF VARIATION	S.S.	D.F.	M.S.	Fcalc.	F(5%)	VARIATION
YS	B	295.9	11	26.9	14.9	2.7	Signif cant
	W	21.8	12	1.8			
UTS	B	184.5	11	16.8	49.9	2.7	Signif cant
	W	4.0	12	0.3			
%El	B	15.4	11	1.4	2.6	2.7	Insignificant
	W	6.3	12	0.5			
%RA	B	267.8	11	24.3	1.7	2.7	Insignificant
		172.6	12	14.4			
Cv(R.T.)	B	24.2	10	2.4	2.2	2.9	Insignificant
	W	11.8	11	1.1			
Cv(-40°F)	B	9.4	10	0.9	0.5	2.9	Insignificant
	W	19.4	11	1.8			

B - Between disks

W - Within disks

TABLE III ANOVA 155MM M46 - TUBE 16

RESPONSE	SOURCE OF VARIATION	S.S.	D.F.	M.S.	Fcalc	F(5%)	VARIATION
YS	B	708.9	27	26.2	25.4	1.9	Significant
	W	28.9	28	1.0			
UTS	B	688.5	27	25.5	42.9	1.9	Significant
	W	16.6	28	0.6			
%E1	B	23.1	27	0.8	1.5	1.9	Insignificant
	W	15.5	28	0.6			
%RA	B	257.1	27	9.5	1.6	1.9	Insignificant
	W	170.0	28	6.1			
C_V(R.T.)	B	367.0	22	16.7	9.1	2.0	Significant
	W	41.7	23	1.8			
C_V(-40°F)	B	434.7	25	17.4	6.4	2.0	Significant
	W	70.7	26	2.7			

TABLE IV ANOVA 152MM M81 - TUBE 1151

RESPONSE	SOURCE OF VARIATION	S.S.	D.F.	M.S.	Fcalc.	F(5%)	VARIATION
YS	B	220.5	16	13.8	12.0	2.4	Significant
	W	19.4	17	1.1			
UTS	B	129.2	16	8.1	9.2	2.4	Significant
	W	14.9	17	0.9			
%E1	B	17.9	16	1.1	2.2	2.4	Insignificant
	W	9.5	17	0.5			
%RA	B	145.8	16	9.1	2.6	2.4	Marginal
	W	59.4	17	3.5			
C_V(-40°F)	B	71.2	16	4.4	1.4	1.8	Insignificant
	W	162.4	51	3.2			

B - Between disks

W - Within disks

TABLE V   ANOVA   152MM INGOT 8178-5 GUN XP3

RESPONSE	SOURCE OF VARIATION	S.S.	D.F.	M.S.	Fcalc.	F(5%)	VARIATION
YS	B	97.3	27	3.6	7.6	1.9	Significant
	W	13.3	28	0.5			
UTS	b	181.0	27	6.7	19.7	1.9	Significant
	W	9.5	28	0.3			
%E1	B	29.8	27	1.1	1.1	1.9	Insignificant
	W	27.4	28	1.0			
%RA	B	559.6	27	20.7	0.9	1.9	Insignificant
	W	673.6	28	24.1			
Cv(-40°F)	B	71.5	27	2.6	2.2	1.9	Marginal
	W	33.5	28	1.2			

TABLE VI   ANOVA   152MM INGOT 8178-5 GUN XP-4

RESPONSE	SOURCE OF VARIATION	S.S.	D.F.	M.S.	Fcalc.	F(5%)	VARIATION
YS	B	67.4	27	2.5	5.7	1.9	Significant
	W	12.2	28	0.4			
UTS	B	413.3	27	15.3	36.5	1.9	Significant
	W	11.7	28	0.4			
%E1	B	36.3	27	1.3	1.9	1.9	Insignificant
	W	19.9	28	0.7			
%RA	B	557.0	27	20.6	1.5	1.9	Insignificant
	W	381.8	28	13.6			
Cv(-40°F)	B	212.0	27	7.8	6.4	1.9	Significant
	W	34.5	28	1.2			

B - Between disks

W - Within Disks

-594-

TABLE VII   ANOVA   90MM M41

RESPONSE	SOURCE OF VARIATION	S.S.	D.F.	M.S.	Fcalc.	F(5%)	VARIATION
YS	B	184.4	13	14.2	9.6	2.5	Significant
	W	20.6	14	1.5			
UTS	B	72.3	13	5.6	11.9	2.5	Significant
	W	6.5	14	0.5			
%RA	B	215.4	13	16.6	1.1	2.5	Insignificant
	W	202.6	14	14.5			
C_v(-40°F)	B	15.2	13	1.2	1.4	2.5	Insigniticant
	W	11.5	14	0.8			

TABLE VIII   ANOVA   105MM M137 HOWITZER GUN XP1

RESPONSE	SOURCE OF VARIATION	S.S.	D.F.	M.S.	Fcalc.	F(5%)	VARIATION
YS	B	209.7	13	16.1	13.1	2.0	Significant
	W	51.6	42	1.2			
UTS	B	93.6	13	7.2	28.2	2.0	Significant
	W	10.7	42	0.3			
%E1	B	5.8	13	0.4	1.9	2.0	Insignificant
	W	9.9	42	0.2			
%RA	B	107.9	13	8.3	2.2	2.0	Marginal
	W	159.7	42	3.8			
C_v(-40°F)	B	46.0	13	3.5	3.3	2.0	Significant
	W	44.4	42	1.1			

B - Between disks

W - Within disks

factor analysis of variance (cross-classification) was conducted on forgings with identical configurations (155mm, 8890-8913 and 152mm, XP3-XP4) to determine if significant variations existed in mechanical properties between similar forgings.

RESULTS AND DISCUSSION. The tensile and Charpy impact data from the nine gun tube forgings are compiled in Appendix B. The results of the one-way classification program (Appendix C) are presented for each tube in Tables I-IX. The experimental F ratio is compared with the F ratio at the 5% signigicance level (Prooability = 0.05) and significant differences noted. The column F (5%) indicates how large the F ratio must be so that there is only a one in twenty chance that an effect is just a random occurrence. Therefore, calculated values of F, larger than F(5%) have less than a one in twenty chance of being caused by random fluctuations. These variation determinations are summarized by response and forging in Table X.

Examination of the individual responses in this summary discloses that yield strength and ultimate tensile strength exhibited real disk to disk variations for all tubes analyzed. This finding can be interpreted by examining the yield strength plots in Figures 3-11. These graphs indicate the average value and spread within a disk. If the maximum and minimum values tend to approach the average, no spread is denoted. It is evident that in general little difference in yield strength was measured within disks and, therefore, the estimate of analytical error variance (S_o^2) is small. Consequently, deviations from disk to disk may show up as significant variations even though their magnitude is small relative to the grand mean. This point is exemplified in Figure 8 where the yield strength range deviates only 4.5 ksi, for the entire tube, yet statistically this amount of deviation constitutes a significant variation between disks. However, it is not highly unusual to obtain yield strength ranges on the order of 20 ksi in cannon tube forgings. Therefore, from a practical standpoint, "real" variations must also be considered in regard to the particular components, properties and previously established ranges. The same interpretation applies to ultimate tensile strength in these nine (9) forgings investigated.

On the other hand, percent elongation and percent reduction of area exhibited no significant variation from disk to disk for the forgings analyzed. Examination of the % RA plot in Figures 3-11 reveals that, in general, considerable deviations in this parameter were measured within disks. Referring to Figure 7, considerable spread in % RA between disks is apparent (20%). However, the scatter within disks generally appears considerable. Accordingly no significant variation in this parameter was realized between disks, although two forgings were marginal. As in the case of yield strength (statistically significant variation) this

-596-

Figure 4. Mechanical Property Variations
in 155mm M126 Tube SN8913

Figure 3. Mechanical Property Variations
on 155mm M126 Tube SN8890

-597-
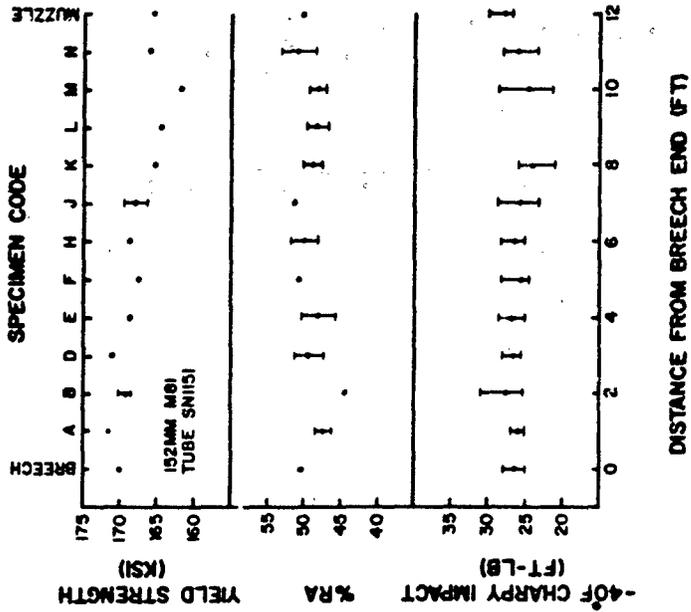
Figure 6. Mechanical Property
Variations in 152mm M81
Tube SN1151
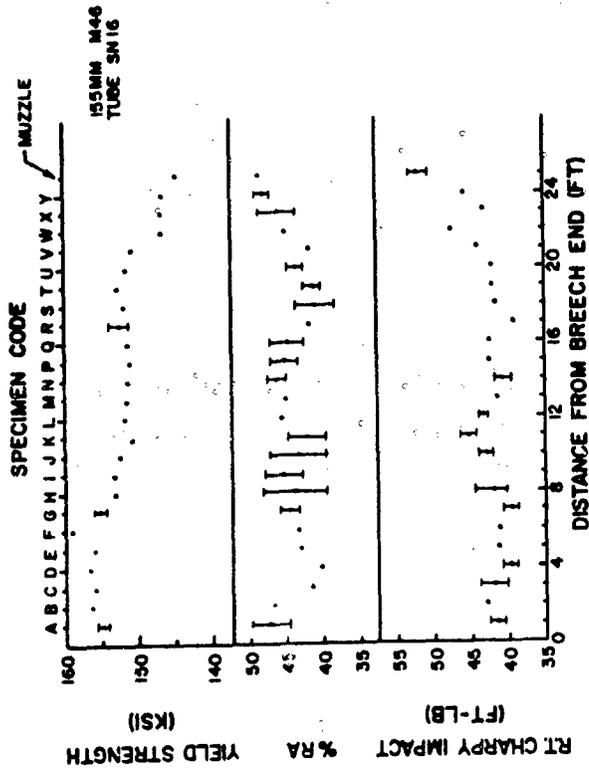
Figure 5. Mechanical Property Variations
in 155mm M46 Tube SN16

Figure 8. Mechanical Property Variations
in 152mm M81 Gun XP4

Figure 7. Mechanical Property Variations
in 152mm M81 Gun XP3

margination results from the relatively uniform spread in % RA for these two forgings (Figures 6 & 10). Also the type of test involved should be considered. It is likely that these determinations were affected by the precision of the individual measurements, yield and ultimate tensile strengths being more precisely measurable than elongation and reduction of area. Another important factor influencing the outcome of this type analysis is the number of tests taken within a disk (replicates). As the number of replicates increases, errors of estimation can be reduced considerably, resulting in more exact variance estimates and consequently accurate variation determinations. This point is expressed by the plots of Figure 10, where the average of four replicates are shown along with the high and low values. These "curves" graphically express the relative uniformity of spread in the data within disks. This consistency could be attributable to the number of observations recorded for this particular tube.

Lastly, Charpy impact data, both room temperature and −40°F, showed real variations in five of the tube forgings analyzed. Examination of Table X reveals that in one particular tube (No. 16), this parameter varied significantly between disks, at both test temperatures, while just the opposite was recorded in Tube #8913. Also, instances were noted where the Charpy impact values of a specific tube varied significantly for one temperature but not the other. Yet the limited data concerning room temperature tests restricts the analysis of these contrasting variation determinations.

Furthermore, it is noteworthy that the two forgings with largest ingot size and wall thickness (155mm M46 and 8" M2A1) exhibited significant variation in −40°F Charpy impact strength. Likewise, the forging with the smallest ingot size and wall thickness (105mm M137) showed real variation in the same parameter. Although the latter observation is most likely due to the small spread in impact strength within disks, these conflicting effects preclude any meaningful conclusions concerning the effect of ingot size and percent reduction on the variation of charpy impact strength for these particular forgings.

The results of the two factor analyses of variance conducted on the 155mm M126 forgings and the 152mm XP forgings are presented in Tables XI and XII. The mechanical responses analyzed were: yield strength, reduction of area and Charpy impact energy absorption. Since no interaction between the factors A and B (A - similar caliber forgings, B - similar test disk location) was anticipated, the sum of squares AB was added to the sum of squares error and this value used to estimate the experimental error variance. Examination of these tabulations shows that for the 155mm forgings no significant variation exists in % RA between the

Figure 10. Mechanical Property Variations in 105mm M137 Howitzer XP1

Figure 9. Mechanical Property Variations in 90mm M41

-109-
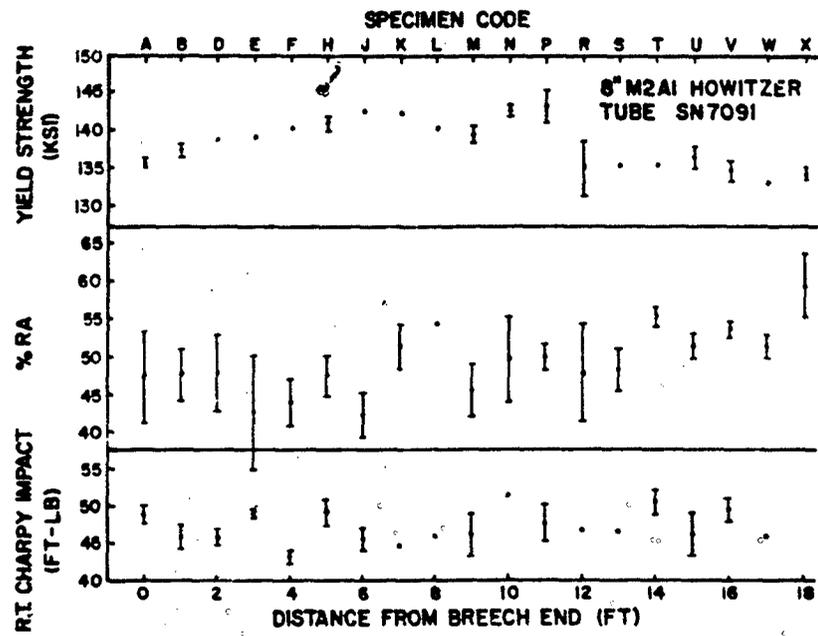
Figure 11. Mechanical Property Variations in 8" M2A1
Howitzer Tube SN7091

TABLE IX   ANOVA   8" M2A1 HOWITZER - TUBE 7091

RESPONSE	SOURCE OF VARIATION	S.S.	D.F.	M.S.	Fcalc.	F(5%)	VARIATION
YS	B	387.9	18	21.5	7.7	2.2	Significant
	W	53.4	19	2.8			
UTS	B	545.4	18	30.3	20.5	2.2	Significant
	W	28.1	19	1.5			
%El	B	42.2	18	2.3	1.9	2.2	Insignificant
	W	23.8	19	1.2			
%RA	B	699.1	18	38.8	1.3	2.2	Insignificant
	W	581.2	19	30.6			
C_v(R.T.)	B	206.9	18	11.5	2.6	2.2	Marginal
	W	83.6	19	4.4			
C_v(-40°F)	B	4110.8	18	228.4	24.0	2.2	Significant
	W	180.8	19	9.5			

B - Between disks

W - Within disks

TABLE X   STATISTICAL VARIATION SUMMARY BETWEEN DISKS
(5% SIGNIFICANCE LEVEL)

FORGING	No.	YS	UTS	%El	%RA	C_{v}R.T.	C_v -40
155mm M126	8590	S	S	I	I	S	I
155mm M126	8913	S	S	I	I	I	I
155mm M46	16	S	S	I	I	S	S
152mm M81	1151	S	S	I	M	-	I
152mm	XP3	S	S	I	I	-	M
152mm	XP4	S	S	I	I	-	S
90mm M41		S	S	-	I	-	I
105mm M137	XP1	S	S	I	M	-	S
8"M2A1	1091	S	S	I	I	M	S

S - Significant variation between disks

I - Insignificant

M - Marginal

TABLE XI ANALYSIS OF VARIANCE BETWEEN TUBES
155MM M126 FORGINGS - 8890, 8913

RESPONSE	SOURCE OF VARIATION	S.S.	D.F.	M.S.	Fcalc.	F(5%)	VARIATION
YS	A	11,532.0	1	11,532.0	5.5	4.1	Significant
	B	61,210.0	11	5,564.5	2.7	2.1	Significant
	SSE	72,742.0	35	2,078.3	-		
%RA	A	130.1	1	130.1	0.07	4.1	Insignificant
	B	30,621.2	11	2,783.7	1.5	2.1	"
	SSE	63,219.2	35	1,806.3	-		
Cv(R.I.)	A	769.5	1	769.5	8.6	4.2	Significant
	B	3,793.7	10	379.4	4.2	2.2	"
	SSE	2,876.0	32	89.9	-		

TABLE XII ANALYSIS OF VARIANCE BETWEEN TUBES
152MM INGOT 8178-5 FORGINGS XP3, XP4

RESPONSE	SOURCE OF VARIATION	S.S.	D.F.	M.S.	Fcalc.	F(5%)	VARIATION
YS	A	18,204.6	1	18,204.6	142.4	4.0	Significant
	B	7,644.0	25	305.8	2.4	1.7	"
	SSE	9,839.4	77	127.8	-		
%RA	A	195,231.3	1	195,231.3	101.8	4.0	Significant
	B	36,904.2	25	1,476.2	0.8	1.7	Insignificant
	SSE	147,708.7	77	1,918.3	-		
Cv(-40°F)	A	15,753.8	1	15,753.8	63.2	4.0	Significant
	B	10,588.4	25	423.5	1.7	1.7	Insignificant
	SSE	19,206.2	77	249.4	-		

A - Between similar caliber forgings

B - Between similar test disk locations

SSE - Experimental error (includes AB)

tubes, but does exist in the yield strength and Charpy impact data (R.T.) at the 5% significance level. Conversely, the 152mm forgings exhibited real variation between tubes for all three responses. Unfortunately, the Charpy test temperature for the latter forgings was -40°F, therefore qualitative comparison between different forgings is not feasible. It should be noted however, that the 152mm forgings were produced from the same ingot as a double forging. Consequently, the two tubes represent different positions in the original ingot with XP3 coming from the top portion and, XP4 from the bottom portion. It has been reported in connection with transverse reduction of area, that comparable tubes from a similar position in ingots cast from a single heat usually have about the same properties [5]. However, the quality of tubes coming from the bottom thirds of ingots is generally slightly lower and occasionally much lower than that of tubes produced from the middle or top thirds of ingots. The results of the 152mm cross classification appear to agree with this information. Conversely, the 155mm M126 forgings were produced from different ingots. Although these ingots came from different heats, they were identical in size. Subsequently, no significant variation was found in reduction of area, between tubes. In addition, the variation observed for yield strength and Charpy impact strength is just barely significant compared to that of the 152mm analysis. For example, the yield strength F ratio of 142 for 152mm XP tubes shows that there is 142 times as much variance arising from different tubes as compared to the repeat tests within the tubes. Similarly, an F ratio of about 6 exists for the yield strength variance between different 155mm M126 tubes.

Since the forgings studied herein were of equivalent chemistry and had heat treatments intended to produce identical microstructures, the observed variations in mechanical properties within forgings and between like forgings indicates that other factors contribute to the variance in final properties of the material. Elements such as melting, solidification and forging procedures can vary widely even within one vendor's practice. For instance, the effect of solidification was implied by the real variations noted between the two 152mm XP forgings produced from different positions of the same ingot. Likewise, the effect of melting and forging procedure was implied by variations between the two 155mm M126 tubes produced from identical size ingots poured from different heats. Thus, it is clearly obvious that in addition to determining the principal causes of mechanical property variation in forged materials, the problem of defining significant variations themselves, exists. Once a method is established (such as ANOVA) it should be combined with subsequent mechanical and metallurgical investigations thereby allowing accrual of statistical information in conjunction with specific material conditions. An accumulation of these data will then provide a basis for effective evaluation of mechanical property variation in large forgings. Therefore, investigation of the

basic issues, viz. melting and deoxidation practice, chemical segregation, fibering or alignment of second phases, forging reduction, ingot size, etc. will be continued. These efforts will be combined with statistical analysis of the mechanical property data in an attempt to optimize the properties and reduce variation in large forgings.

CONCLUSIONS. The mechanical property data and attendant statistical analyses permit the following conclusions:

1. Real disk to disk variations exist in yield strength and ultimate tensile strength for all nine (9) forgings evaluated.

2. No significant variation between disks existed in % elongation and % reduction of area for the forgings, although two (2) tubes were marginal in this respect.

3. Charpy impact data exhibited real disk to disk variation in five of the nine (9) forgings analyzed.

4. Percent reduction of area in forgings of equivalent configuration showed significant variation when the forgings resulted from different ingot positions. Conversely, the same response between similar forgings exhibited insignificant variation when the forgings came from similar ingot positions. This demonstrates the effect of solidification parameters on variation of mechanical properties in large forgings.

5. Significant variation was found in yield strength and room temperature Charpy impact strength for similar forgings produced from identical size ingots but different heats of steel. This variance indicates the effect of melting variables upon mechanical properties.

6. The Analysis of Variance Technique is useful in establishing variation validity but must be interpreted from a practical as well as statistical stand point.

REFERENCES

[1] Baldrey, D. and Lyons, T., "Variation in Mechanical Properties of Tempered Martensite Gun Steel", Watervliet Arsenal Report WVT-7020, March 1970.

[2] Bechtol, H.A., Bernstein, S.B., and Davidson, T.E., "Investigation of 175mm M113 Tube #733, Failure", AWC, Rock Island, Illinois, June 1966.

[3] Davies, O.L., "Statistical Methods in Research and Production", Oliver and Boyd, London, 1961, p. 100.

[4] Slawsky, M.L., Heiser, F.A. and Liuzzi, L., "The Variation of Mechanical Properties in 175mm M113 Gun Tubes", Watervliet Arsenal Technical Report, WVT-6734, July 1967.

[5] Wells, C. and Mehl, R.F., Trans. ASM, vol. 41, p. 715, 1949.

APPENDIX A

GUN TUBE MATERIAL HISTORY

Chemical Analysis

TUBE		C	Mn	P	S	Si	Ni	Cr	Mo	V
155mm M126	8890	.32	.59	.010	.010	.19	2.30	1.01	.52	.11
155mm M126	8913	.32	.58	.010	.009	.20	2.31	.98	.55	.11
155mm M46	16	.32	.65	.010	.010	.23	2.11	.97	.47	.09
152mm M81	1151	.31	.60	.011	.008	.20	2.08	.99	.46	.10
152mm	XP3	.33	.61	.011	.009	.20	2.09	1.00	.48	.12
152mm	XP4	.33	.61	.011	.009	.20	2.09	1.00	.48	.12
90mm M41		.33	.58	.009	.009	.24	2.05	.95	.51	.14
105mm M137	XP1	.33	.60	.010	.008	.20	2.09	.97	.48	.11
8" M2A1	7091	.32	.62	.010	.009	.21	2.24	.99	.51	.12

Thermal Treatment

155mm M126 - 8890, 8913

Heated to 1750°F in 10 hrs., held 12 hrs., air cooled 8 hrs. Reheated to 1550°F in 10 hrs., held 12 hrs., water bore quenched 15 minutes. Charged in draw furnace holding at 500°F. Raised temperature to 1030°F on muzzle and 1045°F on breech in 10 hrs., held 16 hrs., water cooled to 200°F.

155mm M46-16

Heated to 1750°F in 10 hrs., held 12 hrs., air cooled 8 hrs.
Reheated to 1550°F in 10 hrs., held 12 hrs. Water bore quenched
20 minutes. Charged in draw furnace holding at 500°F. Raised
temperature to 1110°F on muzzle and 1110°F on breech in 10 hrs.,
held 16 hrs., water cooled to 200°F.

152mm M81 - 1151

Heated to 1750°F in 10 hrs., held 12 hrs., air cooled 8 hrs.
Reheated 1525°F in 10 hrs., held 12 hrs., water bore quenched 10
minutes. Charged in the draw furnace holding at 500°F. Raised
temperature to 1040°F on muzzle and 1040°F on breech in 10 hrs.,
held 14 hrs., water cooled to 200°F.

152mm - XP3, XP4

Heat treated as a double tube to 1750°F in 10 hrs., held 12
hrs., air cooled 8 hrs. Reheated to 1525°F in 10 hrs., held 12
hrs., water bore quenched 15 minutes. Charged in draw furnace
holding at 500°F. Raised temperature to 1045°F on muzzle and
1045°F on breech in 10 hrs., held 16 hrs., water cooled to 200°F.

90mm M41

Same heat treatment as 152mm XP3, XP4, with the exception of
1030°F on muzzle, 1040°F on breech in draw furnace.

105mm M137 XP1

Same heat treatment as 152mm XP3, XP4, with the exception of
1025°F on muzzle and 1045°F on breech, held 14 hrs. in draw furnace.

8"M2A1 7091

Heated to 1750°F in 6 hours, held 8 hours, air cooled 6 hours. Re-
heated to 150°F in 6 hours, held 8 hours. Water bore quenched 17
minutes. Charged in draw furnace holding at 500°F. Raised temper-
ature to 1070°F on muzzle and 1070°F on breech in 6 hours, held 10
hours, water cooled to 200°F.

Starting Ingot Data

Type of Ingot: 12 Side Round Fluted

Ingot Size:

Tube Forging No.	Maximum Diameter (in)	Minimum Diameter (in)	Length (in)
155mm M126-8890	27-7/8	19-9/16	87-3/8
155mm M126-8913	27-7/8	19-9/16	87-3/8
155mm M46 - 16	35	29-5/8	101
152mm M81 -1151	22-3/4	16-5/8	75
152mm XP3 152mm XP4	27-7/8	19-9/16	87-3/8
90mm M41	24-1/2	16	78
105mm M137 XP1	22-3/4	16-5/8	67-3/4
8" M2A1-7091	34-3/4	29-1/2	104

Rough Forging Data

Tube Forging No.	Breech O.D. (in)	Muzzle O.D. (in)	Tube I.D. (in)
155mm M126	12-1/2	8-3/4	5-7/16
155mm M46	15-1/8	9-5/8	5-7/16
152mm M81	10-11/16	8-1/4	5-3/8
90mm M41	10-1/4	7-1/4	2-15/16
105mm M137	8-1/4	6-3/4	3-3/16
8" M2A1	17-7/8	13-1/4	7-1/8

APPENDIX B

Mechanical Property Data

155MM M126 Tube 8890

CODE	YS(ksi)	UTS(ksi)	%El	%RA	Cv(R.T.) ft-lb	Cv(-40) ft-lb
CFB	175.8	192.2	13.2	35.0	20.2	15.8
	174.9	191.2	11.1	26.6	20.2	16.5
CFD	174.0	192.2	11.9	31.8	21.2	18.2
	173.7	193.0	12.8	36.3	21.2	18.2
CFE	176.3	193.1	14.0	41.2	21.5	17.0
	176.3	192.8	11.9	33.2	21.2	16.5
CFF	174.9	192.1	12.8	35.0	20.5	17.5
	176.7	192.7	13.6	37.7	20.5	17.5
CFH	176.9	193.4	13.2	39.8	21.7	16.5
	176.6	192.5	14.0	39.4	21.2	16.2
CFJ	175.5	192.7	14.3	42.8	24.8	17.5
	174.6	191.7	13.2	39.8	22.2	17.5
CFK	175.7	192.6	13.6	40.7	23.0	17.0
	174.3	191.5	14.0	39.0	23.0	17.5
CFL	171.9	190.0	14.7	38.5	21.2	17.0
	170.1	189.0	13.6	38.5	21.0	19.5
CFM	156.6	185.2	14.7	39.8	23.2	16.2
	167.4	185.7	14.5	42.0	23.2	16.5
CFN	168.4	186.6	13.6	39.4	21.2	18.0
	168.0	185.6	13.2	37.7	23.2	16.5
CFP	167.1	184.9	11.9	27.1	-	-
	165.0	182.7	14.0	41.6	-	-
CFO	178.8	193.0	12.8	33.2	-	-
	177.9	192.5	12.8	28.5	-	-

155MM M126 Tube 8913

Code	YS(ksi)	UTS(ksi)	%El	%RA	Cv (R.T.) ft-lb	Cv (-40) ft-lb
CFB-1	172.5	190.2	12.8	37.0	20.2	16.2
	173.0	189.9	11.9	29.0	21.8	16.2
CFD-1	171.9	189.5	13.2	34.5	21.0	16.2
	168.0	190.6	13.5	38.1	19.2	18.2
CFE-1	173.3	191.0	14.0	40.6	18.8	15.0
	170.0	190.1	13.2	39.4	20.2	17.0
CFF-1	174.2	190.8	12.5	36.3	21.0	15.8
	174.0	189.5	13.2	37.2	19.0	18.2
CFH-1	173.0	190.4	14.1	39.4	21.5	17.5
	173.6	190.6	14.3	39.2	20.2	17.2
CFJ-1	174.0	190.4	14.3	41.8	22.5	18.0
	173.1	189.8	12.8	33.4	21.2	15.8
CFK-1	174.6	190.3	14.3	38.1	22.8	18.5
	174.8	190.3	14.7	42.8	22.5	16.2
CFL-1	169.2	186.8	14.7	40.3	22.5	16.8
	168.9	187.0	14.3	40.7	21.0	18.0
CFM-1	164.2	184.5	14.3	39.0	21.0	16.0
	165.4	184.3	13.2	30.8	23.2	18.2
CFN-1	165.0	183.7	11.1	28.5	19.5	14.0
	164.8	183.3	12.5	29.5	19.2	16.5
CFP-1	165.7	183.5	14.7	42.8	-	-
	164.8	183.4	13.6	36.8	-	-
CFO-1	169.0	188.7	12.8	34.5	-	-
	172.7	190.9	14.7	42.8	-	-

155MM M46 Tube 16

Code	YS(Ksi)	UTS(ksi)	%El	%RA	Cv(R.T.) ft-lb	Cv(-40) ft-lb
Vendor	155.0	168.8	13.5	46.1	-	45.0
(Breech)	155.0	168.4	14.2	46.5	-	44.0
	155.0	167.8	14.2	49.0	-	41.0
	153.5	167.6	14.2	49.4	-	41.0
CE-1	156.3	169.7	15.7	43.3	-	-
	156.0	168.9	15.5	43.3	-	-
CEA	154.2	167.5	17.0	49.8	40.7	39.5
	155.7	167.8	15.4	44.6	42.5	38.5
CEB	156.9	169.2	15.4	44.1	43.2	37.0
	156.6	169.0	15.0	44.6	-	38.2
CEC	156.6	170.1	15.0	41.0	40.2	39.0
	155.4	169.5	16.2	42.0	43.7	36.2
CED	157.2	170.7	14.3	40.3	39.0	38.7
	156.3	170.0	14.0	40.1	40.7	38.7
CEO	154.2	168.0	15.0	41.2	38.0	38.0
	155.7	168.0	15.4	44.1	40.2	34.0
CEE	156.3	169.7	15.7	43.3	41.7	39.0
	156.0	168.9	15.5	43.3	41.2	37.0
CEF	156.9	169.8	14.7	44.1	41.5	41.0
	157.2	170.0	14.3	42.6	41.0	39.5
CFG	156.0	169.9	16.4	46.0	38.7	38.0
	154.5	168.9	15.0	43.5	40.7	39.0
CEH	153	167.3	16.1	48.3	44.5	37
	153.3	168.7	15	39.6	40.3	39.5
CEI	152.1	165.5	14.7	42.8	-	42.7
	154.2	167.7	16.4	47.8	-	43.2
CEJ	152.1	165.7	17.1	47	44	39
	153.3	167.7	14.7	39.4	42.5	39

155MM M46 Tube 16 (cont'd)

Code	YS(ksi)	UTS(ksi)	%El	%RA	Cv(R.T.) ft-lb	Cv(-40) ft-lb
CEK	150	164.7	16.1	44.9	46.5	39.5
	151.8	165.3	14.3	39.4	44.5	39.5
CEL	152.1	166.1	15.7	45.4	44	40.2
	151.8	165.6	16.1	45.8	43	42
CEM	152.1	165.7	16.1	44.9	41.7	41.2
	151.5	165.4	16.1	45.4	41.2	37
CEN	152.1	166.1	15	44.9	39.7	38
	151.2	166.2	16.1	47.4	41.5	41
CEP	151.2	166.2	16.1	47	42	41.7
	151.2	165.6	15.7	43.5	43.5	37.5
CEQ	151.5	165.4	16.1	47	42.5	39
	151.5	166.4	15.4	42.4	42.5	36.5
CER	154.2	168	15	41.6	38.2	39
	151.5	165.7	14.3	41.6	40	34.5
CES	152.4	165.5	16.4	43.6	42.5	38.2
	152.1	166.7	15.5	38.5	41.2	38
CET	153	167.5	15	40.3	42	37.7
	153.6	168.6	15.7	42.4	42.5	39
CEU	151.8	166.3	14.7	42.8	43.2	40.5
	151.2	166.1	15.4	44.9	40	39.2
CEV	150.3	164.7	14.7	42.4	44.5	39.2
	151.8	165.5	15.4	41.2	44.2	39.5
CEW	149.4	164.4	16.1	45.8	46.7	37.5
	144.3	161.8	15	44.5	48.5	41.2
CEX	146.7	161.2	17	48.6	43.2	44.5
	147	160.9	15.7	43.6	43.5	43.2
CEY	147.3	160.2	16.4	49	46	48
	146.1	159.3	17.1	47.4	-	46.5
CEZX						

-614-

155MM M46 Tube 16 (cont'd)

Code		YS(ksi)	UTS(ksi)	%El	%RA	Cv(R.T.) ft-lb	Cv(-40) ft-lb
CEZ		144	157.7	16.1	48.3	53.5	50
		144	156.8	17.5	49	51	47.5
CE3		144.6	156.7	17.5	46.6	51.7	49.5
		144.9	156.8	16.4	46.6	-	-
Vendors	12	147.2	159	15	51	-	47
Muzzle	3	145.5	158.6	15	51		51
Results							
	6	144.2	157.4	14.29	49.8	-	54
	9	145.2	158	15	51	-	52

152mm M81 Tube 1151

CODE	YS(ksi)	UTS(ksi)	%E1	%RA	Cv(-40) Ft-lb	
Vendors	170	187.4	14.2	50.6	27.	
Breech	170	187.2	14.2	50.6	28.	
	170.8	187.8	14.2	50.6	26.	
	170.5	187.6	14.2	50.6	25.	
CD-1	171	186.7	15.7	45.8	26.5	25
	169.8	186.2	15.7	45.8	26	26
CD-A	171.6	186.2	16.1	46.2	25.7	26.5
	172.2	187.2	16.1	48.2	26.5	25
CDAX						
CDB	168.6	185.2	14	44.1	26.5	27.7
	170.1	186.7	15.4	44.6	25.5	31
CD2	171.3	186.7	15	45.4	27	25.2
	167.7	184	15.7	49.8	26	25
CDO	169.2	185.2	16.1	49.2	28.2	26
	171	186.5	15.7	51	28	-
CDOX						
CDD	171.3	187.1	15.	47.6	26.7	28
	171	187	17.1	51.7	27	25.5
CDE	168.9	187	14.7	45.8	27	28.5
	168.6	185.6	16.4	50.2	26.7	25
CDF	166.5	183.4	16.4	50.6	27.7	25.5
	168.6	186.1	16.4	50.6	25.2	24.5
CDH	168.6	185.7	17.1	51.7	26.2	25
	169.8	186.5	15.7	48.2	28.2	26.5
CDJ	169.5	183.2	16.1	51.7	26	22.7
	166.5	183.7	16.8	51.	25	28.7
CDK	165.9	182.7	15.	47.8	26	21
	165.	181.7	16.4	50	26	23.5

152mm M81 Tube 1151   (Continued)

CODE		YS(ksi)	UTS(ksi)	%E1	%RA	Cv(-40) Ft-lb	
CDL		164.1	181.4	16.1	49.8	25.5	22.2
		165.3	183.3	14.7	46.8	27.7	-
CDM		162	180.5	15.7	49.4	24.5	21.5
		162	180.1	15.7	47.4	28.7	27
CDNX							
CDN		165.7	183.5	16.4	48.9	26	23.5
		166.5	182.8	16.8	53.6	28	27
CD3		166.5	183.8	16.1	53.6	24	22.5
		166	182.5	16.8	51.9	26	27
Vendors	12	166.2	185	14.2	50.6	29	
Muzzle	3	165.5	184.6	14.2	50.2	27	
	6	166.5	185.2	14.2	50.6	27	
	9	165.8	183.2	14.2	49.0	30	

152MM Gun XP3

Code	YS(ksi)	UTS(ksi)	%RA	%El	Cv(-40) ft-lb
	165,500	181,600	41.9	12.14	22.0
	163,500	181,200	30.3	9.29	23.0
XP3-1	166,000	183,600	30.3	10.71	22.0
	165,800	182,200	40.7	12.14	23.0
XP3-2	164,500	181,800	34.9	11.43	25.0
	165,500	183,000	43.6	12.86	24.0
XP3-3	165,800	183,600	41.1	12.14	23.0
	165,000	183,000	41.1	12.14	23.0
XP3-4	164,500	182,000	40.2	12.14	24.0
	165,000	182,600	28.9	10.00	23.0
XP3-5	164,500	182,600	34.5	11.43	25.0
	165,500	181,600	35.8	11.43	22.0
XP3-6	165,000	182,400	33.5	10.71	23.0
	164,500	181,600	41.9	12.86	22.0
XP3-	164,000	181,800	33.1	10.71	22.0
	163,500	181,200	25.1	10.00	23.0
XP3-	163,800	181,600	36.7	11.43	23.0
	162,500	180,600	41.1	12.86	20.0
XP3-9	163,500	181,400	34.9	11.43	23.0
	163,500	181,200	39.4	12.14	22.0
XP3-10	162,500	180,800	33.5	10.71	22.0
	162,500	180,800	36.3	11.43	23.0
XP3-11	163,500	181,400	40.2	12.86	22.0
	163,500	180,600	36.7	11.43	23.0
XP3-12	163,500	180,800	41.9	12.86	23.0
	163,000	180,200	35.8	11.43	22.0
XP3-13	163,800	180,600	31.7	10.71	21.0
	163,500	180,800	32.6	10.71	23.0
XP3-14	163,000	180,400	37.6	11.43	23.0
	163,000	180,600	38.5	11.43	22.0

152MM Gun XP3 (cont'd)

Code	YS(ksi)	UTS(ksi)	%RA	%E1	Cv(-40) ft-lb
XP3-15	161,500	180,200	34.0	10.71	22.0
	161,500	180,400	32.6	10.00	21.0
XP3-16	163,000	180,000	42.4	12.86	23.0
	162,500	179,600	38.9	12.14	22.0
XP3-17	162,500	179,600	31.2	10.71	21.0
	163,500	180,200	38.9	12.14	22.0
XP3-18	162,500	179,200	40.2	12.14	19.0
	163,000	180,000	24.6	10.00	18.0
XP3-19	162,500	179,600	29.4	10.71	23.0
	160,500	176,800	41.9	12.86	24.0
XP3-20	161,800	178,600	35.8	11.43	22.0
	161,000	177,800	31.7	10.71	22.0
XP3-21	162,500	182,000	41.5	12.86	24.0
	163,500	182,600	41.1	12.86	21.0
XP3-22	162,000	181,600	34.5	10.00	23.0
	164,200	181,600	41.9	11.43	22.0
XP3-23	164,200	182,200	38.9	11.43	22.0
	163,000	182,000	37.6	10.00	20.0
XP3-24	164,000	185,200	42.4	12.86	21.0
	164,500	184,800	43.2	12.86	22.0
XP3-25	165,500	184,400	44.5	12.86	25.0
	167,000	184,600	39.4	12.14	22.0
XP3-26	165,500	184,000	36.3	11.43	23.0
	166,000	184,600	38.5	11.43	22.0
	165,000	185,800	31.7	9.29	19.0
	166,000	185,600	41.1	12.14	21.0

152MM Gun XP4

Code	YS(ksi)	UTS(ksi)	% RA	% El	Cv(-40) ft-lb
	165,000	185,800	31.7	9.29	19.0
	166,000	185,600	41.1	12.14	21.0
XP4-26	168,000	186,600	44.5	12.86	22.0
	167,000	186,400	46.1	12.86	22.0
XP4-25	165,500]	187,400	47.8	14.29	25.0
	166,500	187,200	38.9	12.86	22.0
XP4-24	168,000	186,800	39.8	12.86	23.0
	168,000	186,800	43.6	12.86	23.0
XP4-23	168,800	190,200	43.2	12.86	22.0
	168,500	189,600	42.4	12.86	22.0
XP4-22	169,500	189,000	40.2	11.43	21.0
	168,800	188,000	48.6	13.57	23.0
XP4-21	168,500	189,000	38.9	12.14	22.0
	167,800	188,600	45.3	13.57	20.0
XP4-20	167,500	186,400	44.9	12.86	23.0
	166,000	185,000	38.0	11.43	23.0
XP4-19	166,000	184,200	48.6	14.29	26.0
	166,500	186,200	44.5	13.57	26.0
XP4-18	164,800	182,400	44.1	12.86	26.0
	166,000	183,000	47.0	13.57	27.0
XP4-17	166,000	182,800	44.5	12.86	26.0
	165,500	184,000	45.7	13.57	26.0
XP4-16	166,200	182,800	46.5	12.86	23.0
	166,500	183,000	47.4	13.57	25.0
XP4-15	166,000	182,800	45.7	13.57	25.0
	166,000	182,400	49.0	14.29	25.0
XP4-14	164,000	180,600	47.4	13.57	25.0
	166,000	182,400	51.0	14.29	26.0
XP4-13	165,500	182,600	48.6	14.29	28.0
	165,000	181,600	50.2	14.29	28.0

152mm Gun XP4 (cont'd)

Code	YS(ksi)	UTS(ksi)	% RA	% El	Cv(-40) ft-lb
XP4-12	165,800	183,000	47.4	13.57	27.0
	166,000	183,400	51.0	14.29	26.0
XP4-11	166,000	181,800	44.1	13.57	25.0
	166,500	182,000	48.6	14.29	26.0
XP4-10	166,000	182,800	43.6	12.86	25.0
	165,500	181,800	34.0	11.43	25.0
XP4-9	165,000	180,800	46.5	12.86	24.0
	166,500	180,600	41.5	11.43	26.0
XP4-8	164,500	181,800	44.5	13.57	27.0
	165,800	182,000	46.5	13.57	26.0
XP4-7	165,500	181,000	49.0	13.57	26.0
	166,000	182,400	43.6	12.14	27.0
XP4-6	166,000	182,200	50.2	14.29	28.0
	166,200	182,600	41.1	12.14	27.0
XP4-5	165,000	181,000	47.4	13.57	25.0
	166,200	182,400	51.7	14.29	26.0
XP4-4	164,800	181,000	50.2	14.29	23.0
	165,800	181,600	50.2	14.29	27.0
XP4-3	166,000	180,400	44.5	13.57	24.0
	167,500	180,800	48.6	13.57	26.0
XP4-2	167,500	180,400	44.1	12.86	23.0
	166,500	181,000	50.2	13.57	26.0
XP4-1	165,500	182,200	44.1	12.86	25.0
	165,000	181,600	48.6	14.29	26.0
	165,500	180,400	43.6	12.86	26.0
	164,500	180,800	46.1	12.86	25.0

90MM M41

Code	YS(ksi)	UTS(ksi)	%RA	Cv(-40)ft-lb
1-12	170,000	187,000	40.7	16.0
2-12	170,000	187,200	43.6	18.0
2-3	169,500	186,600	39.4	17.0
3-12	170,000	187,600	27.5	17.0
3-3	170,000	187,800	41.9	16.0
4-12	170,500	187,200	42.8	17.0
4-3	167,500	186,000	39.4	17.0
5-12	170,500	187,800	41.9	17.0
5-3	168,000	186,000	34.9	17.0
6-12	169,500	187,400	38.0	17.0
6-3	170,500	188,200	37.1	16.0
7-12	170,500	188,000	37.6	16.0
7-3	170,000	188,000	37.6	17.0
8-12	169,500	187,400	35.8	17.0
8-3	170,000	187,800	30.8	17.0
9-12	174,000	188,800	40.7	18.0
9-3	170,000	188,400	44.1	20.0
10-12	176,000	191,200	40.7	19.0
10-3	175,800	191,400	37.6	17.0
11-12	176,500	192,200	37.1	18.0
11-3	176,500	191,400	34.9	17.0
12-12	175,000	191,200	38.9	19.0
12-3	176,500	190,600	42.4	17.0
13-12	174,500	189,800	35.4	18.0
13-3	173,500	188,800	34.0	16.0
14-12	171,000	187,200	38.0	17.0
14-3	173,000	189,400	38.5	18.0
15-12	172,500	188,000	31.2	19.0
15-3	173,500	188,400	38.5	18.0
16-12	173,000	188,200	41.1	19.0

105mm M137 Howitzer Gun XP1

CODE	YS(ksi)	UTS(ksi)	% El	% RA	Cv(-40) ft-lb
1 (muzzle)	172.0	190.6	13.6	46.1	
	171.8	190.4	13.6	47.4	21.5
	171.5	190.0	12.9	43.2	22.0
	171.8	190.6	12.9	44.5	21.0
					22.0
3	172.5	190.2	13.6	48.2	
	172.0	189.8	12.1	41.5	23.0
	172.5	189.8	12.9	45.7	22.0
	170.5	188.6	12.1	39.8	23.5
					20.0
5	173.8	190.8	13.6	46.5	
	173.8	191.6	12.9	42.8	22.5
	173.0	190.8	12.1	41.9	21.5
	173.0	191.0	12.1	42.4	20.0
					20.0
7	169.0	191.8	12.1	42.4	
	175.5	191.8	12.9	43.6	20.0
	170.2	193.0	12.9	44.9	20.5
	174.5	192.4	12.9	42.4	19.5
					21.0
9	176.8	193.4	12.9	44.9	
	176.5	193.2	12.1	41.5	22.0
	175.0	191.6	12.1	40.2	21.5
	175.5	191.6	12.1	40.2	21.0
					20.0
11	176.0	192.8	12.1	41.9	
	176.0	192.2	12.9	43.2	20.0
	174.5	192.6	12.1	40.7	21.0
	175.5	193.0	12.1	41.9	20.0
					22.5
13	175.5	192.0	12.9	44.9	
	175.5	192.4	12.9	43.6	19.0
	173.0	191.4	12.9	44.5	20.5
	173.8	192.4	12.1	40.2	21.0
					20.0
15	175.0	192.8	13.6	44.9	
	175.0	192.6	12.9	44.1	22.0
	173.8	192.0	12.9	44.9	23.5
	175.5	193.2	12.1	40.2	22.0
					21.5
17	176.0	192.4	12.9	44.1	
	176.0	192.6	12.1	41.9	21.5
	175.0	192.2	12.9	43.2	22.0
	175.5	193.0	12.9	43.6	21.0
					22.0

105mm M137 Howitzer Gun XP1 (continued)

CODE	YS(ksi)	UTS(ksi)	%E1	% RA	Cv(-40) ft-lb
19	177.0	193.2	12.1	41.5	22.0
	176.2	193.6	12.1	40.2	19.5
	175.5	192.8	12.1	41.1	22.5
	177.8	194.0	12.1	40.2	20.0
21	177.5	193.6	12.1	39.4	21.0
	177.0	194.4	12.9	42.4	18.5
	176.2	194.0	12.1	41.1	21.0
	178.0	194.0	12.1	40.2	19.0
23	178.2	194.0	12.1	41.5	21.0
	177.8	193.4	12.9	42.8	22.0
	177.0	194.0	11.4	36.7	19.5
	178.0	193.0	12.1	41.5	20.0
25	177.8	194.6	12.9	44.5	20.5
	177.5	194.4	12.9	44.1	21.0
	176.5	194.4	12.9	42.8	20.5
	176.5	194.6	12.9	43.2	19.5
27	174.0	192.2	12.9	45.3	24.0
	173.5	192.2	12.9	44.5	23.0
	173.5	191.6	13.6	46.1	24.0
	174.0	192.8	12.1	41.9	22.0

8" M2A1 Howitzer Tube 7091

CODE	YS(ksi)	UTS(ksi)	%El	%RA	Cv(R.T.) ft-lb	Cv(-40) ft-lb
CGA	136.2	152.6	18.3	52.7	50.0	27.2
	135.0	151.0	17.0	40.3	47.2	24.5
CGB	138.0	154.2	17.0	44.1	47.5	21.5
	136.8	151.8	17.0	51.0	44.2	18.2
CGD	138.9	154.5	15.7	42.8	44.8	24.0
	138.6	154.1	17.5	52.9	47.0	22.0
CGE	139.8	156.1	18.3	50.2	49.5	25.0
	138.8	155.5	15.4	35.0	48.5	17.0
CGF	140.6	156.4	17.0	41.0	42.2	19.0
	139.8	155.1	16.1	47.0	44.0	22.0
CGH	142.0	157.4	17.5	50.3	51.0	18.5
	140.1	155.8	17.5	44.9	47.8	17.0
CGJ	143.1	157.2	15.4	45.4	47.0	37.2
	142.5	156.5	14.7	39.4	44.0	35.0
CGK	142.9	157.3	16.4	48.5	44.5	30.8
	141.3	156.0	17.9	54.4	45.0	30.0
CGL	140.1	154.6	17.9	53.8	45.2	40.5
	139.8	156.8	17.9	54.8	46.8	44.5
CGN	143.4	155.0	19.7	55.3	52.0	44.0
	141.6	154.2	16.4	44.1	51.0	48.0
CGM	140.4	155.1	17.0	49.0	43.2	37.0
	138.3	153.3	14.3	42.0	49.0	37.5
CGP	141.0	152.5	17.0	48.3	45.2	36.5
	145.2	152.2	17.9	51.7	50.0	39.0
CGR	138.6	149.5	18.3	54.4	47.0	45.0
	131.4	149.8	15.4	41.2	46.5	46.5
CGS	135.6	148.2	17.0	45.4	46.0	48.2
	134.7	148.0	17.0	51.0	47.0	47.5

8" M2A1 Howitzer Tube 7091   (continued)

CODE	YS(ksi)	UTS(ksi)	%El	%RA	Cv(R.T.) ft-lb	Cv(-40) ft-lb
CGT	135.0 135.3	143.0 144.6	18.6 19.4	54.0 56.3	49.0 52.0	46.5 45.0
CGU	138.0 135.3	149.5 148.7	18.6 17.9	53.0 49.8	43.2 49.0	39.5 44.2
CGV	135.9 133.5	147.8 147.2	19.7 18.3	54.8 52.7	48.0 51.2	46.0 35.2
CGW	132.9 132.9	147.4 146.2	18.3 17.9	52.5 49.8	45.0 46.8	41.5 49.0
CGX	135.0 132.9	152.7 147.5	19.8 19.3	63.7 55.2		

APPENDIX C

COMPUTER PROGRAM FOR ANALYSIS OF VARIANCE

(ONE WAY CLASSIFICATION)

```
BPS FORTRAN D COMPILER
            C
            C       GUN TUBE ANALYSIS OF VARIANCE
            C
S.0001              DIMENSION TITLE(20),X(50,50),XBAR(50)
S.0002              READ (5,10) NSET
S.0003              DD 1000 L=1,NSET
S.0004              READ (5,5) TITLE
S.0005      5       FORMAT (20A4)
S.0006              READ (5,10) N,K
S.0007      10      FORMAT (212)
S.0008              WRITE (6,15) TITLE
S.0009      15      FORMAT (1H1,1X,20A4,/)
S.0010              DD 100 I=1,N
S.0011              READ (5,20) (X(I,J),J=1,K)
S.0012      20      FORMAT (7F10.0)
S.0013              WRITE (6,25) (X(I,J),J=1,K)
S.0014      25      FORMAT (10F10.2)
S.0015      100     CONTINUE
S.0016              SUM XM = 0
S.0017              DD 300 J=1,K
S.0018              SUM X = 0
S.0019              DD 200 I=1,N
S.0020              SUM X = SUM X + X(I,J)
S.0021      200     CONTINUE
X.0022              XBAR(J) = SUM X/N
S.0023              SUM XM = SUM XM + XBAR(J)
S.0024              WRITE (6,30) XBAR(J),N
S.0025      30      FORMAT (F8.3, 15)
S.0026              WRITE (6,32) SUM XM
S.0027      32      FORMAT (F10.3)
S.0028      300     CONTINUE
            C
            C       GM EQUALS GRAND MEAN (XBAR)
            C
S.0029              GM = SUM XM/K
S.0030              WRITE (6,35) GM,K
S.0031      35      FORMAT (' GM=',F8.2,' K=',15)
S.0032              SUM XB = 0
S.0033              DD 400 J=1,K
```

```
S.0034              XDEV = XBAR(J) - GM
S.0035              SUM XB = SUM XB + XDEV**2
S.0036       400    CONTINUE
             C
             C      XSB EQUALS SUM OF SQUARES BETWEEN DISKS
             C
S.0037              XSB = N*SUM XB
S.0038              WRITE (6,40) XSB
S.0039       40     FORMAT (' XSB=',F10.3)
S.0040              SUM XT = 0
S.0041              DD 600 J=1,K
S.0042              DD 500 I=1,N
S.0043              XT DEV = X(I,J) - GM
             C
             C      SUM XT EQUALS SUM OF SQUARES TOTAL
             C
S.0044              SUM XT = SUM XT + XTDEV**2
S.0045       500    CONTINUE
S.0046       600    CONTINUE
S.0047              WRITE (6,45) SUM XT
S.0048       45     FORMAT (' SUM XT=',F10.3)
             C
             C      XSW EQUALS SUM OF SQUARES WITHIN DISKS
             C
S.0049              XSW = SUM XT - XSB
S.0050              WRITE (6,50) XSW
S.0051       50     FORMAT (' XSW=',F10.3)
S.0052              SUM XW = 0
S.0053              DD 800 J=1,K
S.0054              DD 700 I=1,N
S.0055              XW DEV = X(I,J) - XBAR(J)
             C
             C      SUM XW IS A CHECK ON XSW FOR SUM OF SQUARES WITHIN DISKS
             C
S.0056              SUM XW = SUM XW + XWDEV**2
S.0057       700    CONTINUE
S.0058       800    CONTINUE
S.0059              WRITE (6,55) SUM XW
S.0060       55     FORMAT (' SUM XW=',F10.3)
             C
             C      YSB EQUALS MEAN SQUARE BETWEEN DISKS
             C
             C      YSW EQUALS MEAN SQUARE WITHIN DISKS
             C
             C      YST EQUALS TOTAL MEAN SQUARE
             C
S.0061              YSB = XSB/(K-1)
S.0062              YSW = XSW/(K*(N-1))
```

```
S.0063          YST = SUM XT/((N*K)-1)
S.0064          WRITE (6,60) YSB,YSW,YST
S.0065      60  FORMAT (' YSB=',F10.3,' YSW=',F10.3,' YST=',F10.3)
            C
            C
            C   DFI EQUALS DEGREES OF FREEDOM BETWEEN DISKS
            C
            C   DF2 EQUALS DEGREES OF FREEDOM WITHIN DISKS
            C
S.0066          DF1 = K-1
S.0067          DF2 = K*(N-1)
S.0068          WRITE (6,65) DF1,DF2
S.0069      65  FORMAT (' DF1=',F5.0,' DF2=',F5.0)
            C
            C   S1 EQUALS STANDARD DEVIATION WITHIN DISKS
            C
            C   S2 EQUALS STANDARD DEVIATION BETWEEN DISKS
            C
S.0070          S1 = SQRT(YSW)
S.0071          V2 = (YSB - YSW) / N
S.0072          S2 = SQRT ( V2 )
S.0073          WRITE (6,70) S1,S2
S.0074      70  FORMAT (' S1=',F5.2,' S2=',F5.2)
            C
            C
            C
            C   F RATIO FOR SIGNIFICANT DIFFERENCE
            C
            C
S.0075          R = YSB/YSW
S.0076          WRITE (6,75) R
S.0077      75  FORMAT ('R=',F8.3)
S.0078     1000 CONTINUE
S.0079          END
```

SEQUENTIAL AND PRIOR ANALYSIS FOR 2^k FACTORIALS

J.S. Hunter
Princeton University
Princeton, New Jersey

INTRODUCTION. The analysis of the observations obtained from a 2^k factorial design program usually awaits the completion of the entire design, in one or more replicates. However, the individual trials comprising the design are often run in sequence (hopefully random) and experimenters are moved to infer from observations already in hand without waiting for the completion of a replicate of the full design. Such inferences are often performed before the first replicate of the design is completed, the prior knowledge of the experimenter replacing informally, the yet-to-be-run observations. This paper reviews, with examples, a method whereby every factorial effect may be re-estimated at the conclusion of each completed run, both without and with the use of prior estimates for the observations.

REVIEW. Factorial designs are frequently used to explore a response η as a function of k variables under the control of an experimenter. A 2^k factorial design consists of the $N = 2^k$ experiments formed by taking all possible combinations of two levels (or versions) of k controlled variables $x_1, x_2, ..., x_k$. For k=3, the $N = 2^3 = 8$ runs are displayed as the design matrix given in Table 1:

TABLE 1

Design Matrix			Obs.	Yates Algorithm			Coefficients
x_1	x_2	x_3	y				
−	−	−	59	116	248	560	$b_0 = 70.0$
+	−	−	57	132	312	−8	$b_1 = -1.0$
−	+	−	71	144	−12	40	$b_2 = 5.0$
+	+	−	61	168	4	−16	$b_{12} = -2.0$
−	−	+	69	−2	16	64	$b_3 = 8.0$
+	−	+	75	−10	24	16	$b_{13} = 2.0$
−	+	+	85	6	−8	8	$b_{23} = 1.0$
+	+	+	83	−2	−8	0	$b_{123} = $ zero

Preceding page blank

The + and - signs are used to define the high and low versions of each of the controlled variables. Also recorded in Table 1 is an example set of observations recorded for each run. When a 2^k factorial design is performed it is usual to fit the "factorial" model to the observations. The factorial model contains, in addition to the constant term β_o, k first order terms β_i, (i=1,2,...,k): k(k-1)/2 two-factor interaction terms β_{ij} (i≠j); k(k-1)(k-2)/6 three factor interaction terms β_{ijm} (i≠j≠m); etc. ending with a single k factor interaction term $\beta_{ijm...k}$. There are $N=2^k$ experimental environments established by the 2^k design, and there are N constants or parameters in the factorial model. For example, when k=3 we have

$$\eta = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3.$$

Of course, the experimenter does not see the true response η but rather an observation y where $y = \eta + \varepsilon$. We assume the "errors" ε to be $NID(0,\sigma^2)$, that is, normally and independently distributed with a zero mean and fixed variance σ^2. Given these assumption and design we find the least squares estimates of all the coefficients β_i in the model to be mutually orthogonal with minimum variance, i.e. $V(b_i) = \sigma^2/mN$ where m = number of times the 2^k design has been performed. When one or more of the controlled variables are qualitative the + and - signs in the design matrix are taken to indicate the presence (+) and absence (-), or the two versions, of the qualitative variable. When the variables are qualitative the main "effects", and various interaction "effects" are equal to twice the corresponding estimated coefficients, i.e.

Estimated factorial effects = 2 (Estimated Factorial coefficients)
The variance of any effect is simply $4\sigma^2/mN$.

The least squares estimates of the coefficients in the factorial model may be quickly obtained using Yates' algorithm. The algorithm, and the derived estimated coefficients, are displayed in Table 1. (Almost all standard texts on the design of experiments describe Yates' algorithm). Given the data in Table 1, the fitted factorial model is

$$\hat{y} = 70.0 - 1.0x_1 + 5.0x_2 + 8.0x_3 - 2.0x_1 x_2 + 2.0x_1 x_3 + 1.0x_2 x_3 + (zero)x_1 x_2 x_3.$$

Given we define a "valid" estimate of σ^2 as one determined from repeated experimental trials only, no estimate of σ^2 is possible with the present data and factorial model. To obtain an estimate σ^2 one or more experimental trials must be repeated.

Suppose now, adding replicate runs sequentially, the experimenter repeats run (-,-,-). The recorded data are displayed in Table 2. With this additional information, an estimate of σ^2 is now possible. However, each of the coefficients in the factorial model must also be re-estimated to reflect the information provided by the additional observation. Re-estimation of the b's is easily accomplished using the "predicter-corrector" equation.

-632-

$$d = \frac{1}{N(m+1)} \, (y_i - \hat{y}_i)$$

Where N = number of observations in a single completed replicate. Here $N = 2^3 = 8$.

m = number of replicates on N runs completed.

Here $m = 1$, y_i = newest observation for run i,

$\hat{y}_i$ = predicted response for run i made at the conclusion of the last completed replicate of N runs

Note, that at the conclusion of the first replicate $\hat{y}_i$ equals the previous observation at run i, and that $\hat{y}_i$ equals the previous average of the observations made at run i, should there be more than a single completed replicate.

TABLE 2

x_1	x_2	x_3	Obs	Old b's	Correction	New b's
-	-	-	59,60	70	+0.0625	70.0625
+	-	-	57	-1.0	-0.0625	-1.0625
-	+	-	71	5.0	-0.0625	4.9375
+	+	-	61	-2.0	-0.0625	7.9375
-	-	+	69	8.0	+0.0625	-1.9375
+	-	+	75	2.0	+0.0625	2.0625
-	+	+	85	1.0	+0.0625	1.0625
+	+	+	83	zero	-0.0625	-0.0625

Thus, for our example

$$d = \frac{1}{8(2)} \, (60-59) = 0.0625.$$

Each of the estimated coefficients must now be corrected by either adding or subtracting d. The correction d is always added to the constant term b_0. The correction is added or subtracted to the remaining coefficients in the fitted model as determined by the signs of the associated factorial run. Since the run here is $(x_1 = -1, x_2 = -1, x_3 = -1)$, the correction is subtracted from b_1, b_2 and b_3, added to b_{12}, b_{13}, b_{23} (e.g. $x_1 x_2 = (-1)(-1)=1$) and subtracted from b_{123} (e.g. $x_1 x_2 x_3 = -1$). The new b cofficients are given in Table 2.

As each additional run is performed, the associated correction d can be determined and the estimates up-dated. Thus, at the end of the four additional runs $(-,-,-)$; $(-,+,-)$; $(+,+,-)$ and $(-,+,+)$ we have the results given in Table 3.

TABLE 3

x_1	x_2	x_3	Obs	Original b's	(- - -) .0625	(- + -) .1875	(+ + -) -.3125	(- + +) 0.2500	Newest b's
-	-	-	59, 60	$b_0 = 70.0$	+	+	+	+	70.1875
+	-	-	57	$b_1 = -1.0$	-	-	+	-	-1.8125
-	+	-	71, 74	$b_2 = 5.0$	-	+	+	+	5.0625
+	+	-	61, 66	$b_3 = -2.0$	-	-	-	+	-1.6875
-	-	+	69	$b_{12} = 8.0$	+	-	+	-	7.3125
+	-	+	75	$b_{13} = 2.0$	+	+	-	-	2.3125
-	+	+	85, 89	$b_{23} = 1.0$	+	-	-	+	1.4375
+	+	+	83	$b_{123} = $ zero	-	+	-	-	0.1875

The experimenter determines which experiments to repeat. Any sequence within a replicate is permissable up to and including the last run performed to complete a full replicate of the original design. Estimates from partially replicated factorial and fractional factorial designs are thus readily obtained using the predictor-corrector equation. The derivation of the equation may be found in (1) along with examples illustrating its application to the 2^{k-p} fractional factorial designs, and to factorial designs run in blocks. Another exposition, with an example illustrating the application of the predictor-corrector equation in Evolutionary Operation, appears in (2).

PRIOR ANALYSIS. An important consideration in planning a 2^k factorial design, or any experimental design, should be an a-priori analysis. To perform a prior analysis, the experimenter should predict, prior to taking any data, the outcome of each experiment in the proposed desing. These "observations" $\tilde{y}$ (the ˜ is employed to distinguish such data from real observations y) should then be analyzed as though they were actual data. Many valuable consequences derive from such an analysis:

1) one or more of the estimated main effects, or equivalently the first order coefficients b_i, may be found to be small. In such circumstances a larger change in the associated controlled variable x_i would be in order. In a good design, the "step" to take for each x_i should produce, a priori, equal changes (ignoring sign) in the response. Said another way, the magnitude of the first derivative of the response η with respect to each of the controlled variables x_i should be equal. The design is then appropriately scaled. (such scaling of the x_i is usually not possible when the x_i are qualitative).

ii) the metric for either, or both, the response and controlled variables may be found to be inappropriate. On seeing a data analysis it is common for the experimenter to wish that the response had been measured

another way (in terms of original units or logs rather than as a ratio); or that one or more of the controlled variables had been investigated over a different metric (instead of settings of 20,60, 100, to have used the logarithmic spacing of 20,44.7,100 instead). On other occasions the experimenter may find that a change in the response metric may lead to the elimination of interaction effects in the factorial model.

iii) the statistical description of the experimenter's prior knowledge might lead to a complete re-expression of the original problem. For example, the experimenter may discover he does not want a factorial model, but rather a response surface model to describe his response function. Such a decision can materially alter the experimental design.

iv) the computations can, in themselves, be important. The time to discover the inappropriateness of some computer program, or the inadequacies or unavailability of some desired statistical test, is before the real data become available.

It is not difficult to imagine many other signals arising from a prior analysis. Faced clearly with the statistical aspects of his prior information, the experimenter is forced to review the entire gamut of design considerations; replication, blocking, co-variables, size of experimental region, etc. It seems safe to say that only rarely would an experimental design remain unchanged after a prior analysis. George E. P. Box once commented that the only time one could properly design an experimental program was after the experiments were completed. A careful prior analysis comes as close to this desire as is possible.

COMBINING PRIOR AND REAL OBSERVATIONS. We are now moved to consider what happens as the actual observations y are compared with the prior observations $\tilde{y}$. Whenever a difference $y-\tilde{y}$ is small, the experimenter confirms his prior knowledge. Whenever $y-\tilde{y}$ is large the experimenter questions his prior knowledge (or on some occasions, the observation). In either instance the experimenter's learning experience is heightened and made far more explicit than is possible without the use of the prior observations $\tilde{y}$.

As an example of combining both prior and real data, let the prior distribution of belief for the response η at any point p in the experimental region be Normally distributed with mean $\tilde{y}$ and variance σ_p^2. (The variance σ_p^2 may be roughly determined by sketching the Normal prior for the response and then determining σ_p from the sketched distribution.) Let the distribution of belief of the response η, determined from a recorded observation y at any point in the experimental region, also be Normally distributed with mean y and variance σ^2. (The variance σ^2 should be estimated from repeated observations.) The best estimate of η at any point in the design is then simply

$$\hat{y} = [\ \frac{1}{\sigma_p^2}\ \tilde{y}_p + \frac{1}{\sigma^2}y\ ] \Big/ [\ \frac{1}{\sigma_p^2} + \frac{1}{\sigma^2}\ ]$$

The quantity $\hat{y}$ may then be used to estimate the coefficients in the factorial model. Alternatively, the predicter-corrector equation can be employed, modified for the present situation in which the observations have different weight. Let $w = \sigma_p^2/\sigma^2$, then it can be shown that

-635-

$$d = \frac{w}{N} \left[\frac{1}{m+w} \right] (y-\tilde{y})$$

where

$$w = \sigma_p^2/\sigma^2$$

y = actual response

$\tilde{y}$ = prior response

σ_p^2 = variance of the prior distribution of belief for η

σ^2 = experimental error variance

For example, suppose the data given in Table 1 consisted of the <u>prior judg-ments</u> of the experimenter. Suppose now that the single, new, <u>real</u> observation $y = 60$ was obtained from run $(-,-,-)$. Further, let $\sigma_p^2/\sigma^2 = 4$. Then all the prior coefficients in the factorial model would be changed as illustrated in Table 4.

TABLE 4

$$d = \frac{4}{8} \left[\frac{1}{1+4} \right] (60-59) = 0.1$$

Prior obs $\tilde{y}$	New Obs y	Prior Coefficients	$(- - -)$ d=0.1	New Coefficients
59	60	$b_0 = 70$	+	70.1
57		$b_1 = -1.0$	-	-1.1
71		$b_2 = 5.0$	-	4.9
61		$b_3 = -2.0$	-	-2.1
69		$b_{12} = 8.0$	+	8.1
75		$b_{13} = 2.0$	+	2.1
85		$b_{23} = 1.0$	+	1.1
83		$b_{123} = 0$	-	0.1

As more observations become available, the prior coefficients would be successively modified by the actual data. At the end of all N runs the new coefficients are the weighted average of the prior coefficients, and those coefficients separately determined from the real data, the weights being determined by the inverse of the variances of the prior, and the "real" observations.

REFERENCES

1. Hunter, J. S., Sequential Factorial Estimations, Technometrics VI, No. 1, p 41-55 (1964).

2. Hunter, J. S. Factorial, One Run at a Time Experiments, Indus. Qual. Cont., XXII, pp 391-6, (1966).

Some Recent Advances in
Forecasting and Control†

Part I

G. E. P. Box
University of Wisconsin, U.S.A.
and
G. M. Jenkins
University of Lancaster, U.K.

1. Introduction

1.1. *Nature of Forecasting and Control Problems*

UNTIL fairly recently the word "control" has been principally associated in the statistician's mind with quality control, and especially with the quality control chart techniques developed originally by Shewhart in the United States and by Dudding and Jennett in Great Britain.

During the war the development of sequential inspection methods by Wald and Barnard gave new impetus to techniques in which sequential aspects were emphasised and led eventually to the introduction of cumulative sum charts by Page (1954, 1957) and by Barnard (1959).

The need for control implies the existence of an inherent disturbance in the process of one kind or another such as might be described by a time series. Thus, in recent years we find contributions to control problems from workers in stochastic processes such as Whittle (1963) and Bather (1963). Because one approach to control would be to forecast the deviation from target which would occur if no action were taken and then to act so as to cancel out that deviation, *forecasting* and *control* problems are closely linked together. However, we can forecast a time series in an optimal manner only if we have an adequate stochastic model for that series.

In the past a great deal of attention has been given to stationary time series models which have the property of remaining in equilibrium about a constant mean. However, forecasting has been of particular importance in business and economics where many series (for example, the monthly sales of an industrial product) are *non-stationary* and have no natural mean. It is not surprising, therefore, that the economic forecasting methods which have been proposed by such workers as Holt (1957, 1960), Winters (1960) and Brown (1962) and the control chart techniques proposed by Roberts (1959), all using the exponentially weighted moving averages, are appropriate for a particular type of *non-stationary* process. The fact that such methods have been successful supplies a clue to the *kind* of non-stationary model which might be useful in these problems.

† Originally presented at the European Meeting of Statisticians, held at Imperial College, London, in September 1966.

To a control engineer the word "control" has had a different connotation. He usually thinks in terms of feedback and feed forward control loops, the dynamics and stability of the system, and often of particular types of hardware to carry out the control action. In this paper we outline a statistical approach to optimal forecasting and to the optimal design of feedback and feedforward control schemes that we have developed in previous papers (Box and Jenkins, 1962, 1963, 1965; Box et al. 1967) and which will be described in a forthcoming book (Box and Jenkins, 1968).

The control techniques we discuss are different from those of standard quality control procedures, but this is because they have a different purpose. We certainly do not believe that the traditional quality control chart is unimportant or outmoded. Appropriate *display* of data on such a chart (rather than the burying of it in a process record book) ensures that changes that occur are regularly brought to the attention of those in charge of the process. They are thus nudged into seeking "assignable causes" for the changes and a continuous incentive for process improvement is achieved. This device is of enormous importance because it can stimulate new thinking about the process. However, in many situations a control scheme is required which adjusts some variable, whose precise effect on the quality characteristic is *known*, so as to minimize the variation of this quality characteristic about a target value. It is with such control problems that we are concerned here.

1.2. *An Outline of the Approach*

We suppose throughout that observations are available at *discrete* equispaced intervals of time. For example, in a sales forecasting problem, figures might be available every month and we might wish to forecast sales for 1,2,3,...,12 months ahead. Again, in a chemical process, observations and the opportunity to make control changes might occur every 5 minutes, every hour, or every shift depending on the rate at which the state of the system could change. In the case of a chemical process discrete observations might arise from a discrete or batch process, or a continuous record of the process characteristic might be "sampled" at equally spaced intervals. In practice, if the sampling interval is suitably chosen almost nothing will be lost by employing the discrete rather than the continuous record and there may be considerable gain in the simplicity of the analysis.

The optimal forecasts of future values of a time series are determined by the stochastic model that describes that series. Therefore the main object in statistical analysis directed to forecasting must be in obtaining a suitable stochastic model for the series in question. Therefore, we first develop a class of *stochastic models* which are capable of representing not only stationary behaviour but also non-stationary behaviour of the kind that we have encountered in practice. We show how models which satisfactorily describe a particular series may be derived and how they can be used to forecast seasonal as well as non-seasonal series. The same kind of stochastic model used in the forecasting problem may also be used to represent the disturbances which infect a system and which make control action necessary.

Now any control action which is taken will not be felt immediately but usually its effect will build up gradually because of the inertia of the system. Therefore we next describe *dynamic models* capable of representing the dynamic relationship between a controlling variable X and a controlled variable Y and we show how these dynamic models may be fitted to data obtained from the system.

An important principle in the choice of our models is that they should, whilst

adequately representing the data, contain as few parameters as possible. Following Tukey we call this the *principle of parsimony.*

In Part II of this paper we shall describe how the stochastic and dynamic models may be brought together to design optimal feed-forward and feedback control schemes and also how the parameters in the stochastic and dynamic models may be simultaneously estimated from measurements made on the operating system.

2. TIME SERIES MODELS

A criterion of great importance in discussing time series is *stationarity.* A series is strictly stationary if its properties are completely unaffected by a shift in the time origin. In particular, a stationary series varies about some *fixed* mean μ. It exhibits no change in mean and no drift.

2.1. *Autoregressive and Moving Average Models for Stationary Time Series*

Suppose we denote the values of a stationary series at equally spaced times $t, t-1, t-2, \ldots$ by $w_t, w_{t-1}, w_{t-2}, \ldots$ Let $a_t, a_{t-1}, a_{t-2}, \ldots$ be a "white noise" series consisting of uncorrelated random Normal deviates all having mean zero and variance σ_a^2. It is helpful to think of these a's as a series of random "shocks".

The time series model we employ, originally developed by Yule, is essentially a device for transforming the original series w_t, the observations of which are often highly correlated, into a series of uncorrelated component shocks a_t which can be thought of as generating the series. There are basically two different ways in which this is done.

The deviation $\dot{w}_t = w_t - \mu$ from the mean μ can be made linearly dependent on previous deviations $\dot{w}_{t-1} = w_{t-1} - \mu$, $\dot{w}_{t-2} = w_{t-2} - \mu$, etc., and on a_t. We then have what is called an *autoregressive* model. Thus

$$\dot{w}_t = \phi_1 \dot{w}_{t-1} + a_t, \tag{1}$$

$$\dot{w}_t = \phi_1 \dot{w}_{t-1} + \phi_2 \dot{w}_{t-2} + a_t, \tag{2}$$

are autoregressive models of orders 1 and 2, respectively.

Alternatively, we can make $\dot{w}_t$ linearly dependent on a_t and on one or more previous a's. We then have what is called a *finite moving average model.* Thus

$$\dot{w}_t = a_t - \theta_1 a_{t-1}, \tag{3}$$

$$\dot{w}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}, \tag{4}$$

are moving average models of orders 1 and 2, respectively. One might ask: can an autoregressive model be used to represent moving average behaviour? The answer is that this can be done but an infinite number of autoregressive terms are needed to represent a finite moving average model and vice versa.

To ensure parsimony we may need terms of both kinds and we are thus led to the *general mixed autoregressive-moving average model of order (p, q)*, which may be written

$$\dot{w}_t - \phi_1 \dot{w}_{t-1} - \ldots - \phi_p \dot{w}_{t-p} = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}, \tag{5}$$

where p and q would by 0, 1 or 2 in most applications. To manipulate models of this kind it is convenient to define a backward shift operator B such that

$$Bw_t = w_{t-1}. \tag{6}$$

Using the operator B, (5) can be written

$$\phi_p(B)\dot{w}_t = \theta_q(B)a_t, \tag{7}$$

where

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 \ldots - \phi_p B^p,$$
$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 \ldots - \theta_q B^q,$$

are polynomials in B of degree p and q respectively and $\phi_p(B)$ is called the *auto-regressive* operator and $\theta_q(B)$ the *moving average* operator.

For example, the models of equations (1), (2), (3), and (4) could be written

A.R. 1: $\qquad (1 - \phi_1 B)\dot{w}_t = a_t,$

A.R. 2: $\qquad (1 - \phi_1 B - \phi_2 B^2)\dot{w}_t = a_t,$

M.A. 1: $\qquad \dot{w}_t = (1 - \theta_1 B)a_t,$

M.A. 2: $\qquad \dot{w}_t = (1 - \theta_1 B - \theta_2 B^2)a_t.$

Now consider the first order autoregressive model (1). The values of the series may be built up recursively as follows:

$$\dot{w}_1 = \phi_1 \dot{w}_0 + a_1,$$

$$\dot{w}_2 = \phi_1 \dot{w}_1 + a_2 = \phi_1^2 \dot{w}_0 + \phi_1 a_1 + a_2,$$

$$\dot{w}_3 = \phi_1 \dot{w}_2 + a_3 = \phi_1^3 \dot{w}_0 + \phi_1^2 a_1 + \phi_1 a_2 + a_3,$$

$$\dot{w}_t = \phi_1 \dot{w}_{t-1} + a_t = \phi_1^t \dot{w}_0 + \phi_1^{t-1} a_1 + \phi_1^{t-2} a_2 + \ldots + a_t. \tag{8}$$

We can ensure stationarity for this series by requiring that ϕ_1 lies between the values -1 and $+1$. If ϕ_1 lay outside these limits (if for example, ϕ_1 were equal to 2) then we can readily see from equation (8) that the deviation $\dot{w}_t$ would be dominated by remote events led by $\dot{w}_0$ and a_1 which would become more and more important as t became larger. On the other hand, if ϕ_1 lay between -1 and $+1$, as we require, the behaviour of $\dot{w}_t$ would be dominated by the most recent shock a_t, as is sensible.

A similar argument applied to the first order moving average model (3) leads to the conclusion that θ_1 must lie between -1 and $+1$ if a_t is not to be dominated by remote events. If this condition is satisfied the moving average model is said to be *invertible*.

Now one way of expressing the condition that ϕ_1 in the autoregressive operator $1 - \phi_1 B$ lies between -1 and $+1$ is to say that the roots of the equation $1 - \phi_1 B = 0$ (where B is regarded as a variable) lie *outside* the interval -1 to $+1$.

The corresponding condition for stationarity and invertibility of the general mixed autoregressive moving average model (5) is that the roots (which may be complex) of $\phi(B) = 0$ and $\theta(B) = 0$ must lie outside the unit circle and we shall suppose in all that follows that this condition is imposed.

With these conditions satisfied the model (5) turns out to be a valuable device for representing stationary time series. If the model is expressed in terms of the w_t's themselves, instead of deviations from the mean, the general form of the model may be written

$$\phi_p(B)w_t = \theta_0 + \theta_q(B)a_t, \tag{9}$$

where

$$\theta_0 = (1 - \phi_1 - \phi_2 - \ldots - \phi_p)\mu.$$

2.2. A General Model which can represent Stationary and Homogeneous Non-stationary Time Series

Time series representing economic phenomena and disturbances in processes to be controlled are often best represented by non-stationary models. There is an unlimited number of ways in which a time series may be non-stationary. We now adapt our models to take account of the kinds of non-stationarity which we have frequently met in practice. Figure 1(a) shows one type of non-stationary series of common occurrence. This series is homogeneous except in its level. By this is meant that apart from a vertical translation, one part of the series looks much like another. A series z_t which is stationary in its *first difference*

$$\nabla z_t = z_t - z_{t-1} = (1-B)z_t$$

exhibits precisely this kind of behaviour. Again Figure 1(b) shows a second kind of non-stationarity which is frequently met. This series has neither a fixed level nor a fixed slope but is homogeneous if one allows for differences in these characteristics. We can reproduce such behaviour in a series z_t by a representation in which the *second difference*

$$\nabla^2 z_t = z_t - 2z_{t-1} + z_{t-2} = (1-B)^2 z_t$$

follows a stationary model.

Finally then, if z_t is the variable whose behaviour we wish to represent, it is assumed that its dth difference $\nabla^d z_t = w_t$ can be represented by the stationary and invertible model of equation (9). Since $\nabla^d = (1-B)^d$, the model for z_t becomes

$$\phi_p(B)(1-B)^d z_t = \theta_0 + \theta_q(B)a_t, \tag{10}$$

which will be non-stationary unless $d = 0$. The model is said to be of order (p,d,q) where p, d, and q are usually 0, 1, or 2.


(a)

(b)
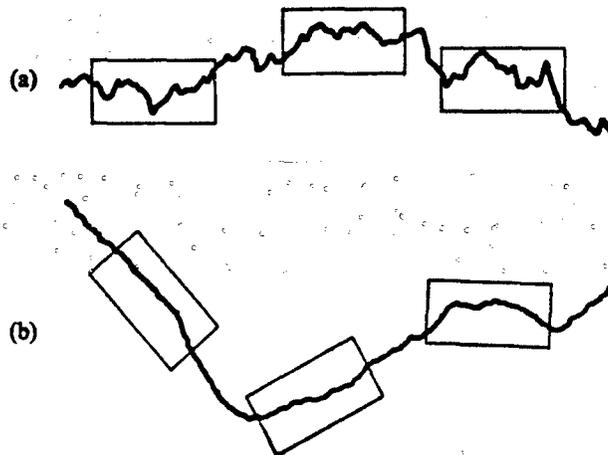
FIG. 1(a). A series showing non-stationarity in level such as can be represented by the model $\phi(B)\nabla z_t = \theta(B)a_t$.

FIG. 1(b). A series showing non-stationarity in level and in slope such as can be represented by the model $\phi(B)\nabla^2 z_t = \theta(B)a_t$.

The operator $\Phi_{p+d}(B) = \phi_p(B)(1-B)^d$ is called the *general* autoregressive operator. Since d of the roots of $\Phi_{p+d}(B) = 0$ are unity, this non-stationary operator will, of course, not satisfy the stationarity condition that all roots lie outside the unit circle. In many practical cases where differencing is needed to obtain stationarity (that is where $d \geqslant 1$), $\nabla^d z_t = w_t$ can be assumed to have a zero mean so that θ_0 in (10) can be set equal to zero.

Suppose we wish to determine a suitable model for a series for which observations $z_1, z_2, z_3, \dots$ are available (where if possible there should be at least 50 and preferably more than 100 observations). In practice such model determination has to be done iteratively using a process of *identification, estimation, diagnostic checking, refitting and rechecking* until a satisfactory representation is found.

2.3. *Identification*

Equation (10) supplies too rich a class of models to permit immediate estimation. Therefore, using experience and the data we first identify a sub-class of models *worthy to be entertained*.

The primary data-analysis tool at this stage is the *sample autocorrelation function* of the original series and its differences. Suppose that n differences $w_1, w_2, \dots, w_n$ are available. The sample autocorrelation coefficient at lag k for $w_t = \nabla^d z_t$ is

$$r_k(w) = c_k(w)/c_0(w),$$

where

$$c_k = \frac{1}{n}\sum_{t=1}^{n-k}(w_t - \bar{w})(w_{t+k} - \bar{w}) \quad \text{and} \quad \bar{w} = \frac{1}{n}\sum_{t=1}^{n}w_t.$$

We shall use $\rho_k(w)$ for the corresponding theoretical autocorrelation.

A suitable value for d may be inferred by finding the degree of differencing necessary to induce the sample autocorrelation function to damp out fairly quickly. For example, Table 1 shows the sample autocorrelation function of z, ∇z, and $\nabla^2 z$ for a series of IBM Common Stock Daily Closing Prices given by Brown (1962). While the sample autocorrelations for the original series are very slow to die out, indicating non-stationarity, its first and higher differences behave like those of a stationary series suggesting that we set $d = 1$.

Values to be entertained for p and q may usually be deduced by inspecting the sample autocorrelations using knowledge of the behaviour of the theoretical autocorrelation function ρ_k for various types of models. The characteristics of $\rho_k(w)$ for

TABLE 1

Sample autocorrelations for various differences of the IBM Common Stock Daily Closing Prices

Source: New York Stock Exchange, May 1961–November 1962 (369 observations)

		1	2	3	4	5	6	7	8	9	10
z	Lags 1–10	·99	·99	·98	·97	·96	·96	·95	·94	·93	·92
	11–20	·91	·91	·90	·89	·88	·87	·86	·85	·84	·83
∇z	Lags 1–10	·09	·00	−·05	−·04	−·02	·12	·07	·04	−·07	·02
	11–20	·08	·05	−·05	·07	−·07	·12	·12	·05	·05	·07
$\nabla^2 z$	Lags 1–10	−·45	−·02	−·04	·00	−·07	·11	−·01	·04	−·10	·02
	11–20	·04	·04	−·12	·13	−·17	·19	·05	−·04	−·01	·09

SOME RECENT ADVANCES IN FORECASTING AND CONTROL

models of order $(1,d,0)$, $(2,d,0)$, $(0,d.1)$, $(0,d,2)$ and $(1,d,1)$ are shown in Table 2. The boundaries of the admissible parameter space are indicated by the inequalities. We see from Table 1 that the autocorrelations of ∇z are all small and appear consistent with a model of order $(0,1,0)$ or perhaps $(0,1,1)$.

Of considerable help in judging the reality of sample autocorrelations is the following approximate formula due to Bartlett for the standard error (S.E.) of r_k, namely

$$\text{S.E. } [r_k] \simeq \sqrt{\left\{\frac{1}{n}(1+2\rho_1^2+2\rho_2^2+...)\right\}}. \tag{11}$$

Since we do not know the theoretical autocorrelations ρ_k, they have to be replaced by their sample estimates r_k.

Thus, under the assumption that the first difference of the IBM series is a moving average of order 1 (that is, the series is of order $(0,1,1)$)

$$\text{S.E. } [r_k] \simeq \sqrt{[\tfrac{1}{368}\{1+2(0.09)^2\}]} = 0.05.$$

Referring to Table 1, we see that only 3, that is 6 per cent of the sample autocorrelations of ∇z from the second onwards are greater than two standard deviations, confirming that a model of order $(0,1,1)$ is worthy to be entertained.

TABLE 2

Behaviour of theoretical autocorrelation function of dth difference of series for various simple (p.d,q) models

Order	$(1, d, 0)$	$(0, d, 1)$
Behaviour of ρ_k	$\rho_k = \phi^k$ decays exponentially	only ρ_1 non-zero
Preliminary estimates from	$\phi_1 = \rho_1$	$\rho_1 = \dfrac{-\theta_1}{1+\theta_1^2}$
Admissible region	$-1 < \phi_1 < 1$	$-1 < \theta_1 < 1$

Order	$(2, d, 0)$	$(0, d, 2)$
Behaviour of ρ_k	mixture of exponentials or damped sine wave	only ρ_1 and ρ_2 non-zero
Preliminary estimates from	$\phi_1 = \dfrac{\rho_1(1-\rho_2)}{1-\rho_1^2}\quad \phi_2 = \dfrac{\rho_2-\rho_1^2}{1-\rho_1^2}$	$\rho_1 = \dfrac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}\quad \rho_2 = \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2}$
Admissible region	$-1 < \phi_2 < 1$ $\phi_2+\phi_1 < 1$ $\phi_2-\phi_1 < 1$	$-1 < \theta_2 < 1$ $\theta_2+\theta_1 < 1$ $\theta_2-\theta_1 < 1$

Order	$(1, d, 1)$	
Behaviour of ρ_k	decays exponentially after first lag, $\rho_k = \phi\rho_{k-1}$ $(k \geqslant 2)$	
Preliminary estimates from	$\rho_1 = \dfrac{(1-\theta_1\phi_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1}$	$\rho_2 = \rho_1\phi_1$
Admissible region	$-1 < \phi_1 < 1$	$-1 < \theta_1 < 1$

By substituting sample estimates for ρ_k in Table 2, preliminary values for the model parameters (which, however, are in general not efficient estimates) may be obtained. For instance, in the case of the IBM Stock Price series suppose that we tentatively entertain the model $\nabla z_t = (1-\theta B)a_t$ of order $(0,1,1)$. Then, because r_1 of ∇z_t is $0\cdot 09$, a first guess for the parameter θ_1 is $-0\cdot 09$ since this is the root of the equation $0\cdot 09 = -\theta_1/1 + \theta_1^2$ which lies within the admissible region $-1 < \theta_1 < 1$.

A complementary tool for identification called the *sample partial autocorrelation function* may also be used (see for example Box and Jenkins (1968)).

2.4. Fitting

Using efficient statistical methods we may now fit the tentatively identified model, or to be on the safe side, a slightly over-parameterized version of it.

On the assumption that the a's are Normally distributed, a close approximation to the maximum likelihood estimates of $\phi = (\phi_1, \phi_2, ... \phi_p)$ and $\theta = (\theta_1, \theta_2, ... \theta_q)$ will be obtained by minimizing the sum of squares

$$S(\phi, \theta) = \sum a_t^2 (\phi, \theta).$$

The values $a_t (\phi, \theta)$ for any ϕ and θ may readily be calculated recursively using

$$a_t = \theta_0 + \theta_1 a_{t-1} + ... + \theta_q a_{t-q} + w_t - \phi_1 w_{t-1} - ... - \phi_p w_{t-p},$$

with $w_t = \nabla^d z_t$ and when $d \neq 0$, θ_0 would often be set equal to zero. The process can be started off by commencing with a_{p+1} and setting $a_p, a_{p-1}, ... a_{p-q+1}$ equal to their expected values of zero. This procedure is adequate for most purposes but a more exact calculation of the likelihood function will be described in Box and Jenkins (1968).
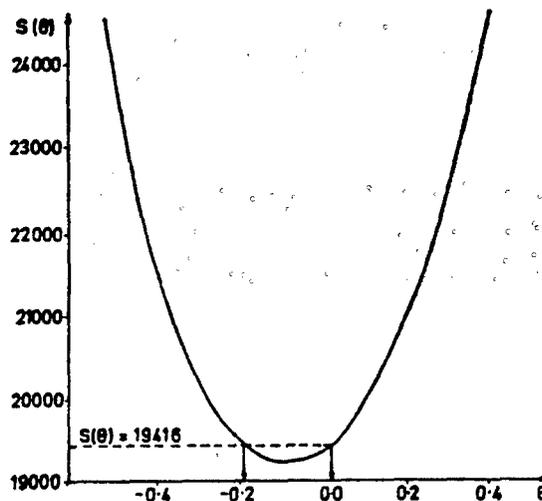

Fig. 2. Sum of squares function for I.B.M. data with approximate 95% confidence region for θ.

An approximate $1-\varepsilon$ confidence region for ϕ and θ is enclosed by the contour

$$S_{1-\varepsilon}(\phi,\theta) = S(\hat\phi,\hat\theta)[1 + \{\chi^2_{1-\varepsilon}(p+q)/v\}] \tag{12}$$

where $\chi^2_{1-\varepsilon}(m)$ is the upper $1-\varepsilon$ significance point of the chi-square distribution having m degrees of freedom and v is equal to the number of a's in the sum $S(\phi,\theta)$ less the number of parameters fitted.

We illustrate again with the IBM data, using the tentatively entertained model $\nabla z_t = a_t - \theta a_{t-1}$. Figure 2 shows a plot of $S(\theta)$ against θ with a minimum at $\hat\theta = -0 \cdot 09$ and $S(\hat\theta) = 19,216$. The approximate 95 per cent confidence limits for θ of $-0 \cdot 19$ and $0 \cdot 03$ are those values for which

$$S(\theta) = 19,216\ (1+3 \cdot 84/367) = 19,417.$$

Complicating the model by adding an extra term on either side produced no appreciable reduction in the residual sum of squares. Hence the form of the model which was finally accepted is $\nabla z_t = a_t + 0 \cdot 1 a_{t-1}$. Least squares estimates and approximate confidence limits may be obtained without the use of graphical methods using iterative non-linear least squares procedures described in Box et al. (1967). However, in general graphs and contour plots of the sum of squares function $S(\phi,\theta)$, or of sections of it, are of great value in illuminating the estimation situation.

2.5. Diagnostic Checks

If the form of the model is correct and if $\hat\phi$ and $\hat\theta$ are close to their "true" values, then the estimated residuals $\hat a_t = a_t(\hat\phi,\hat\theta)$ will be (very nearly) uncorrelated random deviates. Inadequacies of the model may be shown up for example by examining the autocorrelation function of the residuals. A fuller discussion is given in Box and Jenkins (1968).

2.6. Seasonal Models

One often has to analyse time series in which recurrent patterns with *known period* s occur, for example, yearly patterns in monthly sales data ($s = 12$). Here parsimony can often be achieved using multiplicative models of the type

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D z_t = \theta_q(B)\Theta_Q(B^s)a_t. \tag{13}$$

To see how this model is arrived at, suppose we are analysing a series of monthly sales data so that $s = 12$. Suppose we consider all the data at a fixed point in the period s. For example, suppose we consider the sequence of January sales figures. This series would be free of seasonality and might be described by a suitably chosen model of the general form given in equation (10). Bearing in mind that successive Januarys are $s = 12$ months apart and assuming that $\theta_0 = 0$, we would have

$$\Phi_P(B^s)\nabla_s^D z_t = \Theta_Q(B^s)e_t, \tag{14}$$

where

$$\nabla_s z_t = z_t - z_{t-s} \qquad \text{and} \qquad B^s z_t = z_{t-s}.$$

It could reasonably be assumed that February sales, March sales, etc. would follow precisely similar models *with the same parameters*. However, it could not be expected that the residuals e_{t+1} from February sales would be independent of the residuals e_t from January sales. To allow for this dependence a second model may be fitted to the "seasonal free" residuals e_t in the form

$$\phi_p(B)\nabla^d e_t = \theta_q(B)a_t. \tag{15}$$

On eliminating e_t between (14) and (15), we obtain (13). When $s = 12$ the model embodies parameters which describe month-to-month variation (little letters) and parameters which describe year-to-year variation (capital letters).

Procedures for identifying, fitting and checking such models closely follow those described above. For instance, it was shown in Box *et al.* (1967) that the airline passenger data of Fig. 4 was closely fitted by the model

$$(1-B)(1-B^{12})z_t = (1-0\cdot4B)(1-0\cdot6B^{12})a_t \tag{16}$$

corresponding to $p = 0$, $P = 0$, $s = 12$, $d = 1$, $D = 1$, $q = 1$, $Q = 1$, $\theta_1 = 0\cdot4$, and $\Theta_1 = 0\cdot6$. The sum of squares plot for this example is shown in Fig. 3.

2.7. Forecasting

Suppose now that we have determined an adequate model for a given series and we have new data $z_t, z_{t-1}, \ldots$ from the same series extending up to the present time t from which we wish to make a forecast l steps ahead. We call this an *origin* t forecast for *lead time* l.

It may be shown that the minimum mean square error forecast for any lead time is given by

$$\hat{z}_t(l) = \underset{t}{E}[z_{t+l}],$$
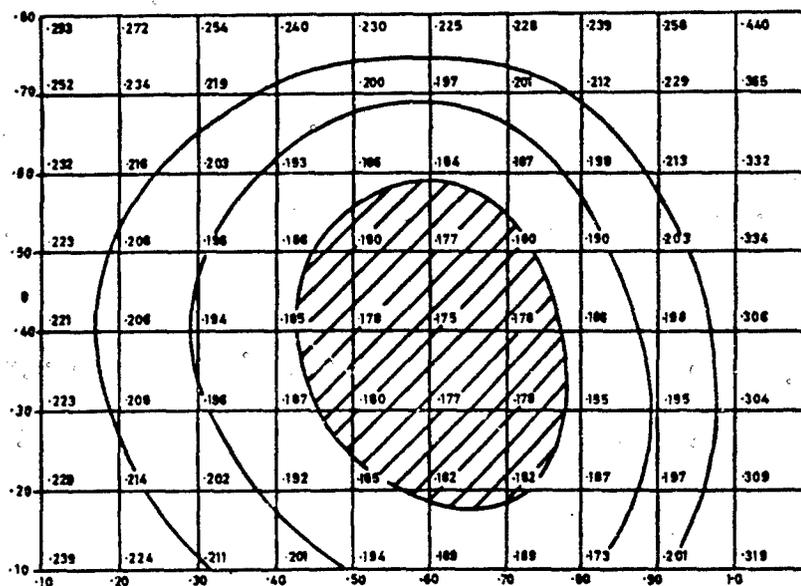
Fig. 3. Sum of squares grid and contours of $S(\theta, \Theta)$ for airline data with approximate 95% confidence region shaded

where $\underset{t}{E}$ is the conditional expectation given the z's up to time t. It follows, in particular, that

$$a_t = z_t - \hat{z}_{t-1}(1). \tag{17}$$

-646-

Thus the "shocks" a_t in the models (10) and (13) are in fact the forecast errors for unit lead time. That for an optimal forecast these "one step ahead" forecast errors ought to form an uncorrelated series is otherwise obvious. For suppose these forecast errors were autocorrelated; then it would be possible to forecast the next forecast error in which case the forecast could not be optimal.

The required expectations are easily found because

$$E_t[z_{t+j}] = \hat{z}_t(j), \qquad E_t[a_{t+j}] = 0, \qquad\qquad j = 1,2,3,\dots$$

$$E_t[z_{t-j}] = z_{t-j}, \qquad E_t[a_{t-j}] = a_{t-j} = z_{t-j} - \hat{z}_{t-j-1}(1), \qquad j = 0,1,2,\dots \quad (18)$$

For instance, to determine the 3-month ahead forecast for the airline series, we first use (16) to write down

$$z_{t+3} = z_{t+2} + z_{t-9} - z_{t-10} + a_{t+3} - 0{\cdot}4a_{t+2} - 0{\cdot}6a_{t-9} + 0{\cdot}24a_{t-10}.$$

Taking conditional expectations at time t,

$$\hat{z}_t(3) = \hat{z}_t(2) + z_{t-9} - z_{t-10} - 0{\cdot}6a_{t-9} + 0{\cdot}24a_{t-10},$$

and using (17),

$$\hat{z}_t(3) = \hat{z}_t(2) + z_{t-9} - z_{t-10} - 0{\cdot}6\{z_{t-9} - \hat{z}_{t-10}(1)\} + 0{\cdot}24\{z_{t-10} - \hat{z}_{t-11}(1)\},$$

that is

$$\hat{z}_t(3) = \hat{z}_t(2) + 0{\cdot}4z_{t-9} - 0{\cdot}76z_{t-10} + 0{\cdot}6\hat{z}_{t-10}(1) - 0{\cdot}24\hat{z}_{t-11}(1).$$

The forecast $\hat{z}_t(2)$ can be obtained in a similar way in terms of $\hat{z}_t(1)$ from $E_t[z_{t+2}]$.
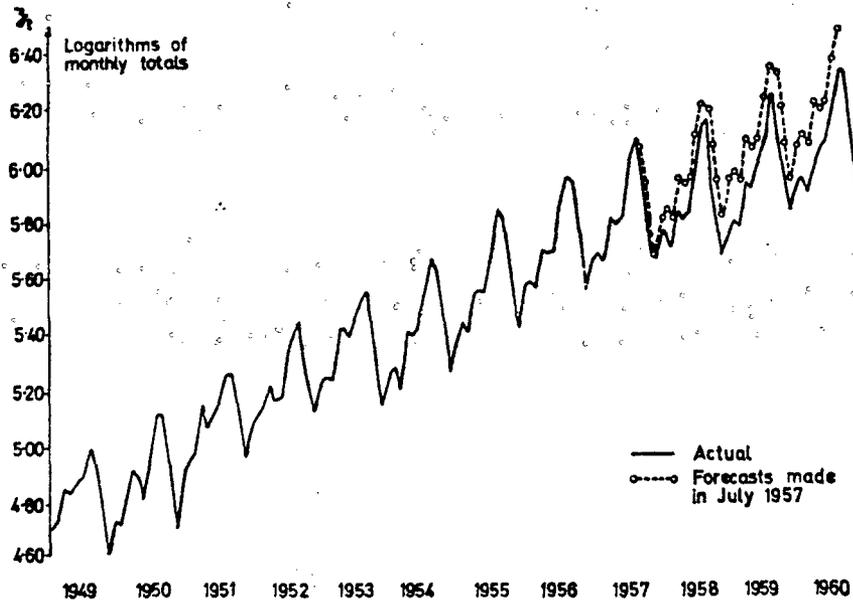
Fio. 4. Logarithms of monthly totals of international airline passengers with forecast made origin July 1957 for 1, 2, 3, ..., 36 months ahead.

Similarly $\hat{z}_t(1)$ can be obtained from $E[z_{t+1}]$ which employs only values of the per-

viously occurring z's and a's. In practice then it is a very simple matter to compute the forecasts $\hat{z}_t(1)$, $\hat{z}_t(2)$, $\hat{z}_t(3)$, etc. recursively, using the forecast function

$$E[z_{t+l}] = E[z_{t+l-1}+z_{t+l-12}-z_{t+l-13}-0{\cdot}4a_{t+l-1}-0{\cdot}6a_{t+l-12}+0{\cdot}24a_{t+l-13}]$$

and (18). Note that this form of computation is ideally suited for use on an automatic computer. Using these methods, forecasts made at origin July 1957 for lead times 1,2,3,...,36 months ahead are shown in Fig. 4 where they may be compared with the values actually realised.

The procedure provides a very convenient and efficient method for industrial forecasting. In particular, it is ideally suited for forecasting sales or inventory on a large variety of products. Since only a very small amount of previous information need be stored for each product a computer with only modest storage capacity may be employed. In those cases where a past history of 50 or so observations is not available one can proceed by using experience and whatever past information *is* available to yield a preliminary model which may then be updated from time to time as more information becomes available.

3. DYNAMIC MODELS

In this section we consider the estimation of dynamic models which describe the relationship between a manipulated variable X and a controlled variable Y. Since the dynamic model describes how changes in X are transmitted into Y, it may be said to describe the *transfer function* between X and Y. Knowledge of the appropriate transfer function is essential for the design of control schemes. However, dynamic models of the type we now describe are also useful in forecasting a time series Y from past values of another time series X as well as from past values of Y.

3.1. *Linear Dynamic Models*

Suppose that in the study of the dynamic characteristics of some system, such as a chemical reactor, pairs of observations $(X_1, Y_1), (X_2, Y_2),\ldots$ are available of an input X, such as gas feed rate and an output Y, such as product viscosity. Suppose further that over the operating ranges of variation of Y and X there exists an approximately linear steady-state relationship

$$\dot{Y} = g\dot{X},$$

where $\dot{Y}, \dot{X}$ denote deviations from some average levels, and g is called the *steady state gain* of the system (or the linear regression coefficient between Y and X).

The *dynamic* characteristics of such systems can usually be represented parsimoniously by linear difference equations of the form

$$\zeta(\nabla)\, \dot{Y}_{t+1} = g\eta(\nabla)\dot{X}_{t-b} \qquad (19)$$

with

$$\zeta(\nabla) = 1+\zeta_1\nabla+\ldots+\zeta_u\nabla^u,$$

$$\eta(\nabla) = 1+\eta_1\nabla+\ldots+\eta_v\nabla^v,$$

where b represents the number of whole intervals of pure dead time (delay) in the

system. Most systems occurring in practice can be represented parsimoniously with u and v at most 2. For instance, the simple model

$$(1+\zeta\nabla)\,\hat{Y}_{t+1} = g(1+\eta\nabla)\hat{X}_t \qquad (20)$$

or

$$\hat{Y}_{t+1} = \frac{\zeta}{1+\zeta}\,\hat{Y}_t + g\frac{(1+\eta)}{1+\zeta}\,\hat{X}_t - \frac{g\eta}{1+\zeta}\,\hat{X}_{t-1} \qquad (21)$$

can represent a system whose response to a step change of X_0 in the input is to produce an eventual change gX_0 in the output which is approached exponentially at a rate depending on ζ and delayed by an amount depending on η. Fig. 5 illustrates the model (20) with $\zeta = 1$, $g = 4$ and $\eta = -0.5$.

By solving the difference equation (19) the dynamic model can be written in the alternative form

$$\hat{Y}_{t+1} = v_0\hat{X}_t + v_1\hat{X}_{t-1} + \dots$$
$$= V(B)\hat{X}_t, \qquad (22)$$

where the weights v_j applied to past inputs are called the *impulse response function* of the discrete system. The form (22) is not a parsimonious way of representing the dynamic model, but is useful in identifying the model (19) as will be shown in section 3.2.
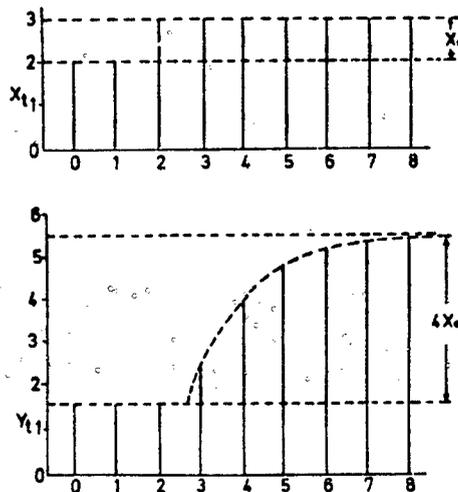
FIG. 5. Delayed exponential response to a step change produced by
$(1+\nabla)(Y_{t+1} - 1.5) = 4(1 - 0.5\nabla)(X_t - 2)$ or $(1+\nabla)Y_{t+1} = -6.5 + 4(1 - 0.5\nabla)X_t$

Dynamic models with added noise

The relationship (19) between the input and output will usually be obscured by

noise due to measurement error and variation in other variables not under one's control. In this case, we can write (19) as

$$\zeta(\nabla)\, \dot{Y}_{t+1} = g\eta(\nabla)\dot{X}_{t-b} + E_{t+1} \qquad (23)$$

and (22) as

$$\dot{Y}_{t+1} = V(B)\dot{X}_t + F_{t+1}, \qquad (24)$$

where E_{t+1} and F_{t+1} are supposed not correlated with the input process $\dot{X}_t$.

3.2. Identification of Dynamic Models

In the same way that the sample autocorrelation function can be used to identify univariate time series models, the basic tool in the identification of dynamic models is the *sample cross correlation function*. To describe a pair of time series by their cross correlation function it is necessary to assume that both series are stationary. Hence it is first necessary to difference both input and output d times until the resulting input and output series are stationary.

If this differencing operation is applied to both sides of (23) and (24) the differenced dynamic models become

$$\zeta(\nabla)y_{t+1} = g\eta(\nabla)x_{t-b} + e_t \qquad (25)$$

and

$$y_{t+1} = V(B)x_t + f_t, \qquad (26)$$

where $\qquad y_t = \nabla^d \dot{Y}_t, \qquad x_t = \nabla^d \dot{X}_t, \qquad e_t = \nabla^d E_t, \qquad f_t = \nabla^d F_t.$

Suppose that after differencing, n pairs of differences $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$ are available. Then the sample cross correlation function at lag $+k$ is defined by

$$r_{xy}(k) = \frac{c_{xy}(k)}{\sqrt{\{c_{xx}(0)\,c_{yy}(0)\}}}, \qquad k = 0, +1, +2, \dots \qquad (27)$$

where

$$c_{xy}(k) = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(y_{t+k} - \bar{y}). \qquad (28)$$

and at lag $-k$ by

$$c_{xy}(-k) = c_{yx}(k),$$

where $\bar{x}, \bar{y}$ are the means of the x and y series.

Prewhitening of the input series

Suppose that it is assumed that the input x_t is uncorrelated with the noise in (26). Then, on multiplying throughout in (26) by x_{t-k+1} and taking expectations,

$$\gamma_{xy}(k) = V(B)\gamma_{xx}(k-1). \qquad (29)$$

where $\gamma_{xy}(k), \gamma_{xx}(k)$ are the theoretical cross covariance function and input autocovariance function respectively, and B now operates on k.

Suppose now that we carry out the usual identification and estimation methods as described in section 2 to obtain a model

$$\phi(B)\theta^{-1}(B)x_t = x_t' \qquad (30)$$

which transforms the correlated input series x_t to a white noise series x'_t. Suppose also that this transformation is now applied to both sides of (26), yielding,

$$y'_{t+1} = V(B)x'_t + f'_t, \qquad (31)$$

where x'_t is white noise uncorrelated with f'_t. On multiplying throughout in (31) by x'_{t-k+1} and taking expectations, we obtain

$$\gamma_{x'y'}(k) = v_k \sigma_{x'}^2. \qquad (32)$$

In terms of the cross correlation function, (32) may be rewritten

$$v_k = \rho_{x'y'}(k) \frac{\sigma_{y'}}{\sigma_{x'}}. \qquad (33)$$

Hence after "prewhitening", the cross correlation function is directly proportional to the impulse function.

The presence of small initial values of v_k is indicative of pure delay or dead time. Thereafter the presence of values of v_k not following a pattern indicates that terms should be introduced on the *right hand side* of the model (19) and the presence of exponential decay or damped sine wave behaviour in v_k indicates that terms should be introduced on the *left hand side* of the model (19).

3.3. *An Example of identifying a Dynamic Model*

Fig. 6 shows continuous records of the input airfeed (X) and the output carbon dioxide concentration (Y) from a gas furnace. The input airfeed was deliberately varied so as to follow an autoregressive process and the input and output records read at 9-sec intervals resulting in 226 pairs of observations.
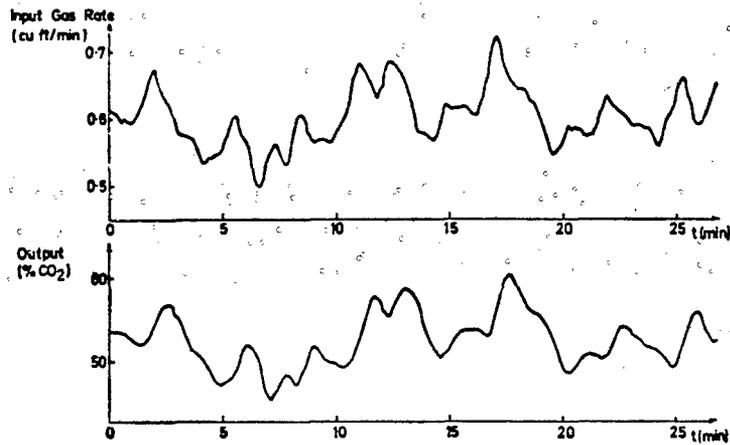
FIG. 6. Input and output records to a gas furnace.

The sample auto- and cross-correlation functions damped out fairly quickly indicating that no differencing was necessary. Hence $x_t = \dot{X}_t$, $y_t = \dot{Y}_t$. The usual identification and fitting procedure applied to the input indicates that it is a third order autoregressive process

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)x_t = x'_t$$

with $\hat{\phi}_1 = 1\cdot97$, $\hat{\phi}_2 = -1\cdot37$, $\hat{\phi}_3 = 0\cdot34$ and $s^2_{x'} = 0\cdot0353$.

Hence the transformations

$$x'_t = (1 - 1\cdot97B + 1\cdot37B^2 - 0\cdot34B^3)x_t$$

$$y'_t = (1 - 1\cdot97B + 1\cdot37B^2 - 0\cdot34B^3)y_t$$

were applied to the input and output series to yield the series x'_t and y'_t with $s_{x'} = 0\cdot188$, $s_{y'} = 0\cdot358$. The sample cross-correlation function between x' and y' is shown in Table 3 together with the estimate of the impulse response function obtained from (33), that is

$$v_k = \frac{0\cdot358}{0\cdot188} r_{x'y'}(k).$$

Fig. 7 shows the plot of v_k versus k and indicates that there are two whole periods of delay, then one or two preliminary values v_3 and v_4 which do not correspond to a pattern, followed by a decay pattern which could be first or second order.
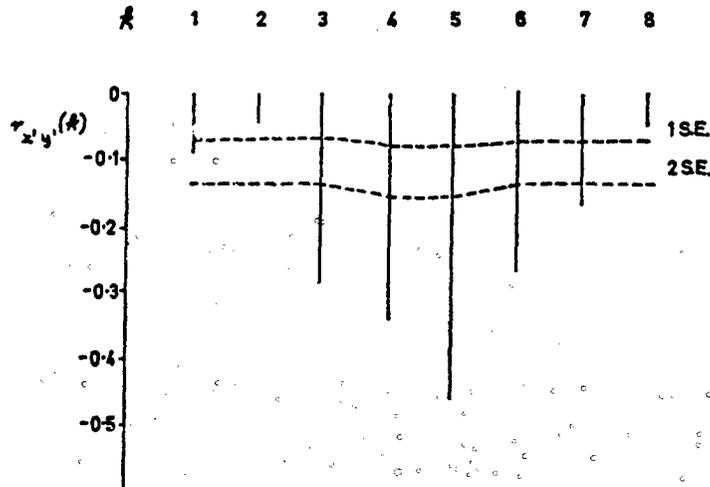
FIG. 7. Gas furnace data sample cross correlations after prewhitening.

TABLE 3

Cross correlation function and approximate impulse response function for gas furnace data

k	1	2	3	4	5	6	7	8
$r_{x'y'}(k)$	-0·05	-0·03	-0·28	-0·33	-0·46	-0·27	-0·17	-0·03
S.E. $[r]$	0·07	0·07	0·07	0·08	0·08	0·07	0·07	0·07
v_k	-0·09	-0·04	-0·53	-0·63	-0·88	-0·52	-0·32	-0·04

To help in the identification procedure, approximate standard errors for the sample cross correlations were computed using Bartlett's approximate formula

$$\text{cov}\{r_{x'y}(k), r_{x'y}(l)\} \approx \frac{1}{N} \sum_{j=-\infty}^{+\infty} \{\rho_{x'x'}(j)\rho_{y'y'}(j+l-k) + \rho_{x'y'}(j+l)\rho_{y'x'}(j-k)\} \quad (34)$$

for the covariance between two values of the sample cross correlation at different lags k and l. On writing $k = l$ in (34) and making use of the fact that the x' series is white noise, the variance of a single cross correlation coefficient is given approximately by

$$\text{var}\{r_{x'y}(k)\} \approx \frac{1}{N}\left\{1 + \sum_{j=-\infty}^{+\infty} \rho_{x'y}(j+k)\rho_{y'x'}(-k)\right\}. \quad (35)$$

The standard errors given in Table 3 are based on the assumption that the cross correlations up to lag+2 and from lag+8 onwards are effectively zero. The one standard error and two standard error limits are plotted on Fig. 7 and confirm the identification of a dynamic model

$$(1 + \zeta_1\nabla + \zeta_2\nabla^2)\, \hat{Y}_{t+1} = (1 + \eta_1\nabla + \eta_2\nabla^2)\dot{X}_{t-b} \quad (36)$$

(probably with $b = 2$) or some simplification of it.

3.4. *Estimation of the Transfer Function*

In the first instance η_2 was set equal to zero in (36) and the model with added noise written as

$$\dot{Y}_{t+1} = \beta_1 \dot{Y}_t + \beta_2 \dot{Y}_{t-1} + \beta_3 \dot{X}_{t-b} + \beta_4 \dot{X}_{t-b-1} + E_{t+1}. \quad (37)$$

If the errors E_t were uncorrelated, then the parameters in (37) could be estimated by linear least squares. Under the added assumption that the E_t are Normal, these would also be maximum likelihood estimates. In practice the E_t's would rarely be uncorrelated, and hence it would be necessary to arrive at a model by iteration as is now illustrated for the gas furnace example.

Initially the model (37) was fitted by linear least squares for different values of the delay parameter b assuming that the errors E_t were uncorrelated. The minimum sum of squares was attained when $b = 1$, yielding the preliminary fitted model

$$\dot{Y}_{t+1} = 1\cdot39\,\dot{Y}_t - 0\cdot55\,\dot{Y}_{t-1} - 0\cdot14\dot{X}_{t-1} - 0\cdot34\dot{X}_{t-2} + E_t,$$

where the dot notation is used to denote deviations of Y and X from their average values. The first ten autocorrelations of the residuals E_t from this model are given in Table 4.

TABLE 4
Autocorrelations of residuals from fitted dynamic model

k	1	2	3	4	5	6	7	8	9	10
r_k	·24	·18	·00	−·02	·01	·16	·08	·06	−·07	·00

These residuals might be explained by a first order autoregressive noise model

$$E_{t+1} = \phi E_t + a_{t+1}.$$

If so, then for an appropriate choice of ϕ we can rewrite the model (37) as

$$\tilde{Y}_{t+1} = \beta_1 \tilde{Y}_t + \beta_2 \tilde{Y}_{t-1} + \beta_3 \tilde{X}_{t-b} + \beta_4 \tilde{X}_{t-b-1} + a_{t+1} \tag{38}$$

where $\tilde{Y}_t = \dot{Y}_t - \phi \dot{Y}_{t-1}$, $\tilde{X}_t = \dot{X}_t - \phi \dot{X}_{t-1}$, and a_t is now white noise.

The model (38) was fitted to the transformed data $\tilde{Y}_{t+1} = \dot{Y}_{t+1} - \phi \dot{Y}_t$ and $\tilde{X}_{t+1} = \dot{X}_{t+1} - \phi \dot{X}_t$ for a grid of values of ϕ and b. The minimum sum of squares occurred at $b = 2$, $\phi = 0\cdot7$ yielding the model

$$\tilde{Y}_{t+1} = \underset{(\pm\cdot05)}{0\cdot90} \tilde{Y}_t - \underset{(\pm\cdot04)}{0\cdot19} \tilde{Y}_{t-1} - \underset{(\pm\cdot07)}{0\cdot48} \tilde{X}_{t-2} - \underset{(\pm\cdot10)}{0\cdot44} \tilde{X}_{t-3},$$

the figures in parentheses under the estimated parameters being their standard errors obtained from the usual least squares formula. The autocorrelations of the residuals a_{t+1} from this model were all small, confirming that the model is adequate.

Hence the final model is

$$(1 - 0\cdot7B)(1 - 0\cdot90B + 0\cdot19B^2) \dot{Y}_{t+1} = -(1 - 0\cdot7B)(0\cdot48B^2 + 0\cdot44B^3)\dot{X}_t + a_{t+1}. \tag{39}$$

Rewriting (39) as

$$\dot{Y}_{t+1} = \frac{-(0\cdot48B^2 + 0\cdot44B^3)\dot{X}_t}{1 - 0\cdot90B + 0\cdot19B^2} + \frac{a_{t+1}}{(1 - 0\cdot7B)(1 - 0\cdot90B + 0\cdot19B^2)}, \tag{40}$$

we see that the fitted dynamic model is

$$(1 - 0\cdot90B + 0\cdot19B^2) \dot{Y}_{t+1} = -(0\cdot48B^2 + 0\cdot44B^3)\dot{X}_t. \tag{41}$$

This model implies transfer function characteristics which agree very closely with those estimated in Jenkins and Watts (1968) using cross spectral analysis. In the control engineer's language it corresponds to a second order system with time constants $T_1 = 15\cdot8$ seconds and $T_2 = 8\cdot2$ seconds, and a pure delay or dead time of $22\cdot8$ seconds.

The model (40) also implies that the noise n_{t+1} at the output of the system is a third order autoregressive process

$$(1 - 0\cdot70B)(1 - 0\cdot90B + 0\cdot19B^2)n_{t+1} = a_{t+1}. \tag{42}$$

A more direct fitting procedure which employs iterative non-linear least squares, and which is readily adapted to the analysis of multiple input data, is described in Box and Jenkins (1968).

REFERENCES

BARNARD, G. A. (1959). Control charts and stochastic processes. *J. R. Statist. Soc.* B, 21, 239–271.
BATHER, J. A. (1963). Control charts and the minimisation of costs. *J. R. Statist. Soc.* B, 25, 49–80.
Box, G. E. P. and JENKINS, G. M. (1962). Some statistical aspects of adaptive optimisation and control. *J. R. Statist. Soc.* B, 24, 297–343.
Box, G. E. P. and JENKINS, G. M. (1963). Further contributions to adaptive quality control: simultaneous estimation of dynamics: non-zero costs. *I.S.I. Bulletin*, 34th Session, Ottawa, Canada, 943–974.
Box, G. E. P. and JENKINS, G. M. (1965). Mathematical models for adaptive control and optimisation. *A.I.Ch.E.–I.Chem.E. Symposium* Series No. 4, 61–68.
Box, G. E. P., JENKINS, G. M., and BACON, D. W. (1967). Models for forecasting seasonal and non-seasonal time series, pp. 271–311 of *Advanced Seminar on Spectral Analysis of Time Series*. Edited by B. Harris, New York: John Wiley.

SOME RECENT ADVANCES IN FORECASTING AND CONTROL

Box, G. E. P. and Jenkins, G. M. (1968). *Time Series Forecasting and Control*, San Francisco: Holden-Day. The original draft of this book has been issued in 1966 and 1967 as Technical Reports Nos. 72, 77, 79, 94, 95, 99, 103, 104, 116, 121 and 122 of the Department of Statistics, University of Wisconsin, Madison and simultaneously as Technical Reports Nos. 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 13 of the Department of Systems Engineering, University of Lancaster.

Brown, R. G. (1962). *Smoothing, Forecasting and Prediction of Discrete Time Series*. New Jersey: Prentice-Hall.

Holt, C. C. (1957). Forecasting trends and seasonals by exponentially weighted moving averages. O.N.R. Memorandum No. 52, Carnegie Institute of Technology.

Holt, C. C., Modigliani, F., Muth, J. F., and Simon, H. A. (1960). *Planning Production, Inventories and Work Force*, New Jersey: Prentice-Hall.

Jenkins, G. M. and Watts, D. G. (1968). *Spectral Analysis and its Applications*. San Francisco: Holden-Day.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41, 100–114.

Page, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, 44, 248–252.

Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, 1, 239–250.

Whittle, P. (1963). *Prediction and Regulation*, London: English Universities' Press.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6, 324–342.

CSP-V: A CONTINOUS SAMPLING PLAN WITH A PROVISION FOR A REDUCED CLEARANCE NUMBER

Gary L. Aasheim

U. S. Army Ammunition Procurement and Supply Agency
Joliet, Illinois

1. INTRODUCTION

The purpose of this paper is to introduce CSP-V, a continous sampling plan which provides a means for reduced screening inspection. The discussion will begin with introductory remarks about continuous sampling plans in general and about some recent developments that have simplified the formulation of mathematical expressions for continuous sampling plans.

2. CONTINUOUS SAMPLING PLANS

A continuous sampling plan is a sampling plan in which inspection is carried out as the product flows along the production or assembly line without any grouping of the units of product into lots for inspection purposes. The procedure alternates between sequences of screening (100% inspection) and sampling inspection, where the severity of inspection is dependent upon the discovery and spacing of defective units of product. Units of product that have passed an inspection station, whether inspected or not, are considered acceptable.

A continuous sampling plan offers certain advantages to a user. The plan can be used as a process control device, providing a signal of possible process difficulty immediately upon the finding of a defective unit. When dealing with the manufacture of dangerous materials such as ammunition, the use of a continuous sampling plan makes possible the avoidance of the hazards of grouping large quantities of product in a storage area until a lot is formed.

A continuous sampling plan levies certain requirements upon the manufacturing process to which it is to be applied. The process must present a moving product to the inspector. There must be ample physical facilities at the manufacturing site to permit 100% inspection when necessary. The inspection procedure must be relatively easy to carry out and the manufacturing process must be one which is capable of producing a homogeneous product.

3. TERMINOLOGY

Certain of the terminology peculiar to sampling plans is reviewed here:

a. The Average Fraction Inspected (AFI) is the fraction of units inspected over an indefinitely long period of time when the process average is some constant value p.

b. The Average Outgoing Quality (AOQ) is the percent of units in the passed product which are defective when the process average is some constant value p.

c. The Average Outgoing Quality Limit (AOQL) is the maximum value that the AOQ assumes over the range of p.

d. Responsiveness is the reaction of a continuous sampling plan to a sudden breakdown in product quality. The particular measure that is chosen for this property is somewhat arbitrary. One of these measures will be described later in the paper.

4. A BRIEF HISTORY OF CONTINUOUS SAMPLING PLANS

Harold F. Dodge developed the first continuous sampling plan in 1943. The plan, which has since been designated CSP-1, simply provides for screening inspection until a predesignated number, i, consecutive items passing the inspection station are found defect-free. Clearing i in the screening phase signals initiation of a sampling phase in which only a fraction, f, randomly selected items passing the inspection station are examined. A procedural diagram for CSP-1 is shown in Figure 1.

Dodge and others later developed more complex continuous sampling plans, the main object of which was to reduce the amount of inspection without reducing product quality. Another development, this one by Lieberman and Solomon, was the use of Markov chains in constructing the formulae to describe characteristics of continuous sampling plans. Brugger and others subsequently developed a simplified Markov chain approach for use in the problem of developing continuous sampling plan formulae.

5. THE SIMPLIFIED MARKOV CHAIN APPROACH

As continuous sampling plans grew in complexity beyond CSP-1, the difficulty in applying the Lieberman and Solomon Markov chain analysis also grew. This method dealt with each unit of product individually as a state of the Markov chain. The simplified method groups together all units of product inspected under a particular phase of the sampling plan and then treats each phase of the sampling plan as a state of the Markov chain.

The simplified Markov chain method can be used for continuous sampling plans which:

a. can be described by finite, ergodic Markov chains. For purposes of the simplified Markov chain method, the condition of ergodicity will be satisfied as long as there is a probabalistic path leading from each jth phase of the sampling plan to every other phase of the plan and back again to phase j.

```
  START
         ┌──────────────────────────────────────────────────┐
         │  The screening crew inspects 100% of the units.   │
         ├──────────────────────────────────────────────────┤
         │  When i consecutive units are found free of the   │
         │  defects concerned,                                │
         └──────────────────────────────────────────────────┘
                                    │
                                    │
                                    ▼
         ┌──────────────────────────────────────────────────┐
         │  The screening crew is released from 100% inspec- │
         │  tion and the sampling inspector inspects a frac- │
         │  tion, f, of the units, where the sample units    │
         │  are selected in a random manner.                  │
         ├──────────────────────────────────────────────────┤
         │  When the sampling inspector finds one of the     │
         │  defects concerned,                                │
         └──────────────────────────────────────────────────┘
```
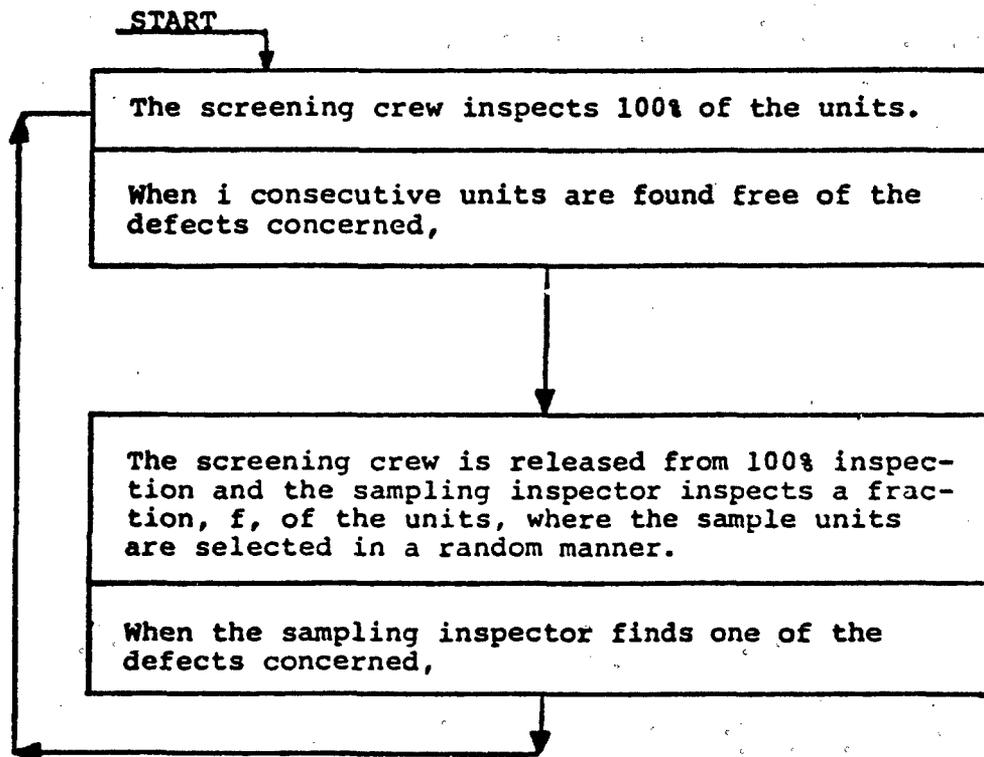
FIGURE 1.   PROCEDURE FOR CSP-1 PLANS

b. have no inspection phases other than those of the following types:

Screening Phases: Units of product are inspected until a pre-designated number, i, of them are found defect-free. Inspection is subsequently begun in a new inspection phase.

Checking Phases: A predesignated number of units of product are inspected. When this inspection is completed, one of two inspection phases, which shall be designated j_1 and j_2, will be entered. Phase j_1 will be entered if no defects were found during the checking phase; phase j_2 will be entered if a defect is found during the checking phase.

Limited Sampling Phases: Sampling inspection is conducted at some predesignated rate f until (i) a defective unit is found or (ii) some predesignated number k of units are found nondefective. The next inspection phase entered depends upon the occurrence of (i) or (ii). A phase of this type is limited in length, of course, to k inspected units.

Unlimited Sampling Phases: Sampling is conducted at rate f until a defect is found, at which time a new phase is entered. This type of phase is unlimited in length since it terminates only when a defective unit is found.

It can be shown by a relatively simple derivation that continuous sampling plans employing at most a finite number of inspection phases selected only from the above list can be described by a finite Markov chain.

The simplified Markov chain method then allows us to express the <u>steady phase</u> probability P_j, of each of the j inspection phases of the inspection plan in terms of each of the <u>steady state</u> probabilities P_j'' of the Markov chain describing the sampling plan and the expected lengths $E(j)$ of the inspection phases of the plan:

(5.1) $P_j = P_j'' E(j) / \sum_i P_i'' E(i)$

We define the steady phase probability, P_j, as the long run proportion of time, expressed in terms of all units reaching the inspection station, inspection is conducted in the jth inspection phase. The steady state probability, P_j'', is the long run relative frequency with which the jth inspection phase occurs.

We then express each of the steady state probabilities as a function of a single one of the steady state probabilities, call it P_0'', so that

(5.2) $P_j'' = h_j P_0''$

Here, h_j is simply the coefficient of P_0'' that arises when expressing P_j'' in terms of P_0''. (5.1) can then be rewritten:

(5.3) $P_j = P_0'' h_j E(j) / \sum_i P_0'' h_i E(i)$

$= h_j E(j) / \sum_i h_i E(i)$

We are also generally interested in finding an expression for the average fraction inspected (AFI). By definition AFI $= \sum_i f_i P_i$. Substituting from the final result of (5.3) we get

(5.4) AFI $= \sum_i f_i h_i E(i) / \sum_i h_i E(i)$

where f_i is the sampling frequency of the ith inspection phase. The AOQ may also be written in terms of the AFI and the population percent defective, p:

(5.5) AOQ $= p[1 - \text{AFI}]$ when defective items are removed, and then replaced

(5.6) AOQ $= p[1 - \text{AFI}]/[1 - p \, \text{AFI}]$ when no replacement of defective units of produce is made.

6.   CSP-V

CSP-V is a continuous sampling plan for use with a product with a good quality history where reduction in inspection has economic merit, but reduction in sampling inspection does not. As was mentioned earlier, most continuous sampling plans developed since CSP-1 had the objective of reducing the AFI without sacrificing product quality. This reduction has been accomplished primarily by including in the later plans sampling phases with reduced sampling frequencies.

But consider the case of the isolated sampling inspector who has no tasks to perform other than his inspection duties, whose idle time only increases with reduced sampling. Introducing into this situation a sampling plan which reduces the amount of sampling inspection does nothing to improve the efficiency with which manpower is utilized. If we instead reduce screening inspection, we free screening inspectors, who are normally also production line workers, for regular production line duties for a larger proportion of the time. To provide a continuous sampling plan which would reduce the proportionate amount of screening inspection in a situation such as the one just described was the objective in designing CSP-V.

-661-

```
        START
SC1 ┌────────────────────────────────────────────────────────────┐
    │  The screening crew inspects 100% of the units.             │
    ├────────────────────────────────────────────────────────────┤
    │  When i consecutive units are found free of the             │
    │  defects concerned,                                         │
    └────────────────────────────────────────────────────────────┘

SA1 ┌────────────────────────────────────────────────────────────┐
    │  The screening crew is released from 100% inspec-           │
    │  tion and the sampling inspector inspects a frac-           │
    │  tion, f, of the units, where the units are select-         │
    │  ed in a random manner.                                     │
    ├───────────────────────────────┬──────────────────────────────┤
    │  If a defect is found         │  If i consecutive sample      │
    │                               │  units are free of the        │
    │                               │  defects concerned,           │
    └───────────────────────────────┴──────────────────────────────┘

SA2 ┌────────────────────────────────────────────────────────────┐
    │  The sampling inspector continues to inspect a frac-        │
    │  tion, f, of the randomly selected units.                   │
    ├────────────────────────────────────────────────────────────┤
    │  When the sampling inspector finds one of the de-           │
    │  fects concerned,                                           │
    └────────────────────────────────────────────────────────────┘

SC2 ┌────────────────────────────────────────────────────────────┐
    │  The screening crew begins inspection, 100% of the          │
    │  units.                                                     │
    ├───────────────────────────────┬──────────────────────────────┤
    │  If a defect is found         │  If x consecutive units       │
    │  in the next x con-           │  are free of the defects      │
    │  secutive units.              │  concerned,                   │
    └───────────────────────────────┴──────────────────────────────┘
```

FIGURE 2.   PROCEDURE FOR CSP-V PLAN

Figure 2 shows the procedural diagram for CSP-V. Inspection begins with a screening phase, SC1, and then moves into a limited sampling phase, SA1. The inspection procedure, in a good quality situation, can lead from SA1 into SA2 and SC2. The reduced screening for which the plan aims is made possible through the introduction of phase SC2 in which is employed a clearance number x that is less than i of SC1. The formulae (AFI, AOQ, etc.) that are developed for CSP-V in the following paragraphs will reduce to those for CSP-1 for x=i.

6.1   Derivation of the Formulae

The transitional matrix for the Markov chain describing CSP-V is shown in Figure 3 and gives the probabilities of moving from one phase of the inspection plan to another. On these pages p is the fraction defective of the units reaching the inspection station and q = 1-p.

<div align="center">TO</div>

		SC1	SA1	SA2	SC2
	SC1	0	1	0	0
	SA1	$1-q^i$	0	q^i	0
FROM	SA2	0	0	0	1
	SC2	$1-q^x$	q^x	0	0

<div align="center">FIGURE 3</div>

If, from the transitional probability matrix, we express each of the steady state probabilities as linear combinations of the other steady state probabilities, we get:

(6.1.1) $P''_{SC1} = (1-q^i) \, P''_{SA1} + (1-q^x) \, P''_{SC2}$

(6.1.2) $P''_{SA1} = P''_{SC1} + q^x \, P''_{SC2}$

(6.1.3) $P''_{SA2} = q^i \, P''_{SA1}$

(6.1.4) $P''_{SC2} = P''_{SA2}$

If we now express these steady state probabilities in terms of P''_{SA2}, we get:

(6.1.1)' $P''_{SC1} = [(1-q^i q^x)/q^i] \, P''_{SA2}$

$(6.1.2)'\ P''_{SA1} = q^{-1} P''_{SA2}$

$(6.1.3)'\ P''_{SA2} = P''_{SA2}$

$(6.1.4)'\ P''_{SC2} = P''_{SA2}$

Equations $(6.1.1)'-(6.1.4)'$ will be recognized as having the form of (5.2) where $P''_o = P''_{SA2}$ in this case.

The next task will be to determine the steady phase probabilities (5.3), the AFI (5.4) and the AOQ for CSP-V. In order to expedite the necessary computations, a working table is constructed:

TABLE 1

WORKING TABLE FOR FINDING
PHASE OCCUPANCY PROBABILITIES

	(1)	(2)	(3)	(4)	(5)
			Expected Number of Units in	Multiplication of Column 3 by	Product of Columns
Phase	Coefficient	Multiplication of Column 1 by q^1	Phase	f_{pq}^1	2 & 4
SC1	$(1-q^1 q^x)/q^1$	$1-q^1 q^x$	$(1-q^1)/pq^1$	$f(1-q^1)$	$f(1-q^1)(1-q^1 q^x)$
SA1	$1-q^1$	1	$(1-q^1)/fp$	$q^1(1-q^1)$	$q^1(1-q^1)$
SA2	1	q^1	$1/fp$	q^1	q^{21}
SC2	1	q^1	$(1-q^x)/p$	$fq^1(1-q^x)$	$fq^{21}(1-q^x)$

Column (1) lists the coefficients, h_j, of the indicated inspection phases as their steady state probabilities P''_j were given in terms of P''_{SA2}. Column (2) is obtained from (1) by multiplying each term in (1) by the least common denominator of all terms in (1).

Column (3) is the expected number, $E(j)$, of units that reach the inspection station during an occurrence of the indicated phase. Column (4) is obtained from (3) as (2) is from (1). Column (5) is the product of (2) & (4).

P_{SA1}, for example, may now be taken from column (5) according to (5.3):

$$P_{SA1} = h_{SA1} E(SA1) / \sum_j h_j E(j)$$

$$= q^1(1-q^1)/[f(1-q^1)(1-q^1q^x)+q^1(1-q^1)+q^{21}+fq^{21}(1-q^x)]$$

and

$$AFI = \sum_i f_i h_i E(i) / \sum_i h_i E(i)$$

$$= \frac{[f(1-q^1)(1-q^1q^x)+fq^1(1-q^1)+fq^{21}+fq^{21}(1-q^x)]}{[f(1-q^1)(1-q^1q^x)+q^1(1-q^1)+q^{21}+fq^{21}(1-q^x)]}$$

$$= f[1-q^1(q^x-q^1)]/[f(1-q^1\{1+q^x-q^1\})+q^1]$$

The AOQ formula may now be easily gotten from (5.5) or (5.6).

6.2   Responsiveness

In an earlier paragraph, the term responsiveness was mentioned, a measure of the plan's ability to detect a drop in product quality to an undesirable level. The particular measure used by the Army Ammunition Procurement and Supply Agency is the expected time for the plan to return from its most liberal sampling phase to its most restrictive screening phase when the percent defective of the product reaching the inspection station has some fixed value p. In CSP-V we have somewhat arbitrarily called SA2 a more liberal sampling phase than SA1 and have determined the expected length of time to return from SA2 to SC1 as the desired measure of responsiveness.

We once again employ the simplified Markov chain approach to determine the plan's responsiveness. A procedural diagram for CSP-V is shown in Figure 4 to illustrate the Markov chain. Notice that the diagram is changed somewhat from that shown in Figure 2:

in Figure 4, $P(SC1 \to SA1) = 0$ and $P(SC1 \to SA2) = 1$;

in Figure 2, $P(SC1 \to SA1) = 1$ and $P(SC1 \to SA2) = 0$.

The change excludes all paths of the original model that do not lead from SA2, as a starting point, directly or indirectly to SC1 as a terminal point. The change introduces one path, a recycling path in effect, which simply leads from the terminal phase, SC1, to the initial phase, SA2.

START

SC1

The screening crew inspects 100% of the units.

When i consecutive units are found free of the defects concerned,

SA1

The screening crew is released from 100% inspection and the sampling inspector inspects a fraction, f, of the units, where the units are selected in a random manner.

| If a defect is found | If i consecutive sample units are free of the defects concerned, |

SA2

The sampling inspector continues to inspect a fraction, f, of the randomly selected units.

When the sampling inspector finds one of the defects concerned,

SC2

The screening crew begins inspection, 100% of the units.

| If a defect is found in the next x consecutive units. | If x consecutive units are free of the defects concerned, |

FIGURE 4.   PROCEDURAL DIAGRAM OF CSP-V PLANS ALTERED TO ALLOW CONSTRUCTION OF RESPONSIVENESS MODEL

-666-

The expression for responsiveness is taken from the identity

(6.2.1) $P_{SC1} = E(SC1)/[E(SC1) + E(REST)]$

which is just the proportion of time inspection is in SC1 over the long run in terms of the responsiveness procedural diagram. In this expression for P_{SC1}, E(REST) is the expected length of time to get from SA2 to SC1 and is the responsiveness term we want to evaluate. (6.2.1) can then be rewritten:

(6.2.2) $E(REST) = E(SC1)[1-P_{SC1}]/P_{SC1}$

The transitional probability matrix for the Markov chain associated with the responsiveness problem follows:

		TO			
		SC1	SA1	SA2	SC2
FROM	SC1	0	0	1	0
	SA1	$1-q^i$	0	q^i	0
	SA2	0	0	0	1
	SC2	$1-q^x$	q^x	0	0

Each steady state probability is expressed, as before, in terms of transitional probabilities and other steady state probabilities. The following working table is constructed in the same manner as was the table of 6.1, this time taking $P_0'' = P_{SC2}''$.

TABLE II

WORKING TABLE FOR FINDING PHASE OCCUPANCY PROBABILITIES FOR THE RESPONSIVENESS MODEL

	(1)	(2) Expected Number of	(3) Multiplication of	(4) Product of
Phase	Coefficient	Units in Phase	Column (2) by fpq^i	Columns (1) & (3)
SC1	$1-q^i q^x$	$(1-q^i)/pq^i$	$f(1-q^i)$	$f(1-q^i)(1-q^i q^x)$
SA1	q^x	$(1-q^i)/fp$	$q^i(1-q^i)$	$q^i q^x(1-q^i)$
SA2	1	$1/fp$	q^i	q^i
SC2	1	$(1-q^x)/p$	$fp^i(1-q^x)$	$fq^i(1-q^x)$

-667-

E(REST) can now be determined. First, from the working table

$$P(SC1) = [f(1-q^1)(1-q^1q^x)]/[f(1-q^1)(1-q^1q^x)+q^1q^x(1-q^1)+q^1+fq^1(1-q^x)]$$

Then, substituting into (6.2.2) and simplifying

$$E(REST) = [f(1-q^x)+q^x(1-q^1)+1]/fp(1-q^1q^x)$$

Figure 5 is a display of responsiveness curves. Since CSP-1 generally requires a greater amount of sampling than other continuous sampling plans and is generally the most responsive of all continuous sampling plans, it is usually the standard against which other sampling plans are compared. Hence, for comparison purposes, we have included in Figure 5 the responsiveness curve for CSP-1 along with the responsiveness curves for CSP-V at i=2x and i=3x. Spacing between the curves will vary with AOQL and sampling frequency; however, the position of the three curves relative to each other remains unchanged.

6.3   Summary

A complete set of i values has been computed for CSP-V to be included in a future revision of MIL-STD-1235. This "complete set" includes computations for x=1/2 and x=1/3. It was decided not to go below x=1/3, in choosing values for x, since some of the x values for the small i-value plans would get, we somewhat arbitrarily decided, too small. We restricted our attention to integer multiples 2 and 3 for ease in remembering by the user.

CSP-V is currently undergoing plant tests at Pine Bluff Arsenal, Arkansas, where early responses to the plan have been favorable. Specifically, Arsenal representatives state that the amount of time production line workers are needed to serve on screening inspection teams has been reduced.

7.   BIBLIOGRAPHY

7.1   Aasheim, Gary L., "CSP-V: A Continuous Sampling Plan with a Provision for a Reduced Clearance Number", QEM 21-230-9, U.S. Army Ammunition Procurement and Supply Agency, Joliet, Illinois, April 1972.

7.2   Brugger, R.M., "A Simplification of the Markov Chain Approach to Continuous Sampling Plan Formulation", QEM 21-230-12, U.S. Army Ammunition Procurement and Supply Agency, Joliet, Illinois, March 1972.

7.3   Brugger, R.M., "An Introduction to Continuous Sampling", QEM 21-230-16, U.S. Army Ammunition Procurement and Supply Agency, Joliet, Illinois, March 1972.

FIGURE 5

CSP-V Responsiveness Curves

AOQL = .79%
f = 1/100
i = 330 for CSP-1 Sampling Plan
i = 333 for CSP-V Sampling Plan, x = 1/2
i = 333 for CSP-V Sampling Plan, x = 1/3
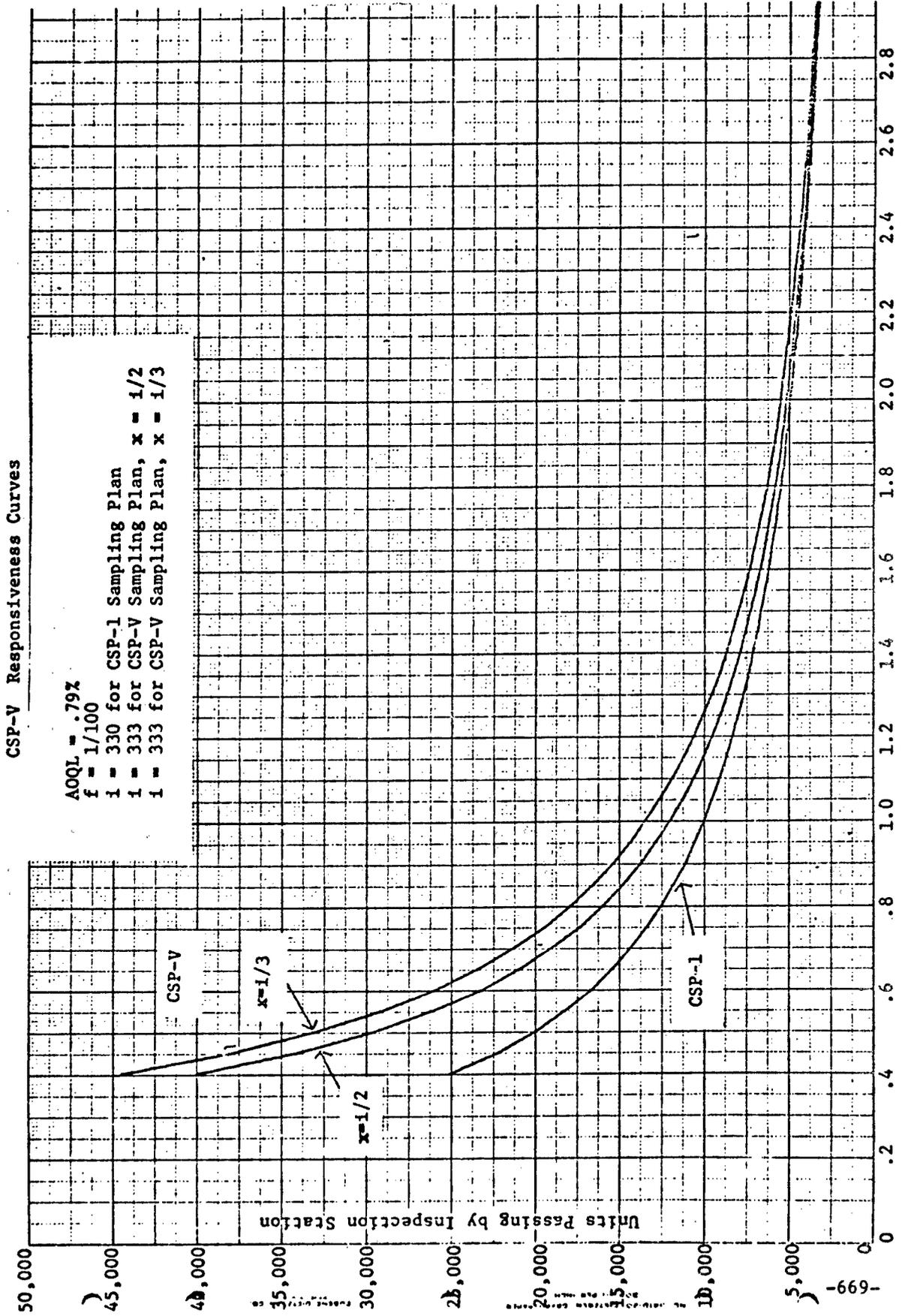
THE SELECTION OF THE MOST MEANINGFUL SUBSET OF RESPONSES

IN A MULTIPLE RESPONSE EXPERIMENT

Walter D. Foster
Fort Detrick, Frederick, Maryland

ABSTRACT

A sensitive plant extensively used in experimental agriculture is the black valentine bean seedling which has been used with a wide variety of growth regulators in controlled environment experiments. Some of the responses to treatment that have been used are fresh and dry weights of plant tops, roots, total plant, plus a variety of height observations. The problem has been to select that subgroup of responses most informative in terms of statistical analysis.

Univariate approaches have included the magnitude of the F ratio as a measure of relative sensitivity to treatment and the magnitude of the coefficient of variation as a measure of efficiency. In factorial experiments, the F ratio was examined separately for each main effect.

A multivariate approach consisted of following Smith, Gnanadesikan and Hughes[1] multivariate analysis of variance and isolating the characteristic vector corresponding to the maximum root in the characteristic equation where the size of the elements of the vector indicate relative contribution to analysis after removal of the correlations.

[1] Multivariate Analysis of Variance. Biometrics, Vol 18 #1, p 22ff. March 1962.

Preceding page blank

INTRODUCTION: At the beginning of an extensive series of greenhouse experiments with growth regulators, the question was raised concerning the most appropriate response measurements to make. Using the black valentine bean plant, the following responses to environment and treatment factors were considered: fresh and dry weights of tops, roots, plant, second and fifth trifoliate leaves, and height of plant measured at 3, 7, 12, and 14 days after emergence. It was the objective of this study to identify that subset of these responses showing both the greatest sensitivity to the treatment factors and the most efficiency in the sense of needing least replication for a given effect.

Environment factors consisted of two temperatures, 25° and 30° C. in the growth chambers; treatment factors were four levels of a growth inhibitor applied as a liquid to the nutrient solution. Five separate pots per treatment provided the replication in a completely randomized design.

METHODS OF ANALYSIS:

A. Univariate Approach.

Analysis of each response separately was proposed as the first approach. The F-ratio in the analysis of variance was used as a measure of relative sensitivity of response to treatment on the basis of the following reasoning. Each response, being part of the same bean plant, had been subjected to the same treatments as every other response. A response not affected by treatment was considered to be insensitive and would be expected under the null hypothesis to have an F ratio of unity. Conversely, the greater the effect of treatment on a particular response, the higher would be the F-ratio. This is illustrated in the following table of expected means squares for the design used in this experiment.

Source	d.f.	Expected Mean Square
Temperature	$t-1$	$\sigma^2 + dw\,\theta_T^2$
Dose	$d-1$	$\sigma^2 + tw\,\theta_D^2$
T x D	$(t-1)(d-1)$	$\sigma^2 + w\,\theta_{DT}^2$
Within	$td(w-1)$	σ^2

F-test for Temperature: $(\sigma^2 + dw\,\theta_T^2)/\sigma^2$ under the null hypothesis that $\theta_T^2 = 0$. Thus the magnitude of F depends upon the size of θ_T^2 and similarly for the effect of Dose.

The next step was to select that subset of responses with the highest F-ratios. The following statistical analysis was proposed for this selection. A transformation of F to Fisher's original Z-statistic,

$$Z = (1/2)\ln F$$

results in a quantity approximately normally distributed[2/] with mean

$$E(Z) = -\tfrac{1}{2}(1/n_1 - 1/n_2)$$

and variance

$$V(Z) = \tfrac{1}{2}(1/n_1 + 1/n_2)$$

provided n_1 and n_2, the degrees of freedom for numerator and denominator of F, are not small. Thus a test of two F-ratios as a device to pick the larger could be accomplished by a t-test of the corresponding Z's:

$$t = (Z_A - Z_B)/\left[V(Z_A) + V(Z_B) - 2\text{Cov}\,(Z_A, Z_B)\right]^{1/2}$$

Two problems arise with this use of t. Granted that the F-values are not likely to be independent because the different responses were measured on the same plants, the t-test allows for a covariance term as indicated above. The problem arises in estimating the covariance. On the assumption that any

meaningful correlation is likely to be positive, it is noticed that its evaluation would serve to diminish the denominator of t. Therefore, ignoring the covariance term would be a conservative step.

The second problem is the degrees of freedom to be used with the t-test. If the F-ratios are considered as a sample statistic, then it might be intuitively appropriate to take df = $n_{2_A} + n_{2_B}$. Alternatively, it might be assumed that the experimental design, once designated, fixed the first and second moments of F so that the variance of the Z transform can be considered to be known. Thus df = ∞ would be appropriate. My own observation is as follows: if the df for F are not small as required by the Z-transform of normality, then it should make little difference and the user could take df = ∞ and remain happily apathetic to the potential troubles of exactness of theory.

The rationale for a t-test between two Z-values can be extended to a multiple range test for the ordering of all of the observed F-ratios. Thus, a multiple range ordering of the Z's and hence the F's should serve to order the relative sensitivities of the various responses to the treatments.

The second objective of the univariate approach was to identify the most efficient responses in the sense of distinguishing a treatment effect with minimum replication. This approach is identified with that of Neyman-Pearson and the concept of power of the test. The power of the familiar t-test is written as 1 - B where B, the probability of accepting a false hypothesis when a specified alternative hypothesis is true, is estimated by

$$ t = (\mu_A - D)/ s\ n^{-1/2}\ , \quad \text{where} $$

μ_A = alternative hypothesis, and

D = value in X-scale corresponding to the decision value of t under the null hypothesis and a given level of alpha.

-674-

To relate the concept of the power of the t-test to the coefficient of variation, the latter is rewritten as

$$s = \bar{X} \ (CV) \quad \text{and}$$

s is substituted in the expression above for t:

$$t = k/ \ CV \quad \text{where}$$

$$k = N^{1/2} \ (\mu_A - D)/ \ \bar{X} \ .$$

Thus, the efficiency of detecting a difference of a given size is inversely related to the coefficient of variation when the experimental design has been stated and the parameters experimentally estimated: namely, N, the alternative hypothesis, level of alpha, and the experimental mean response.

B. Multivariate Approach.

Smith, Gnanadesikan and Hughes [1] multivariate analysis of variance (MAV) is a technique designed to test for treatment effects where multiple responses have been observed. It also gives a procedure by which the most effective subset of the responses may be identified--which is the objective of this paper. This procedure consists of computing the characteristic vector associated with the maximum characteristic root. The largest absolute values in the characteristic vector are associated with the responses suggested to be the most effective subset for detecting treatment differences, analogous to the way that the magnitude of the standard partial regression coefficients are indicative of the most effective predictors in muliple regression.

While the flow chart for their MAV is very effective, it did not include the computation of the characteristic vectors. Adding this as a printout to the MAV program makes this portion of their algorithm applicable

to the analysis here. Because multivariate techniques in general and this one specifically allows for linearly correlated responses, they are especially applicable to problems like these.

APPLICATION OF METHODOLOGY:

These methods were computed for the growth chamber experiment on black valentine beans with two temperature levels and four dose levels in factorial arrangement. The seven responses observed in the experiment reported here were coded as follows:

	Fresh Weight	Dry Weight
Tops	FWT	DWT
Roots	FWR	DWR
Plant	FWP	DWP
Growth	Δ Ht-14	

The results of the univariate analyses of variance have been condensed as F-ratios for Temperature and Dose in the following table which also gives the CV for each response, their rankings, and the rank of the absolute magnitude of the values of the characteristic vector in the MAV.

	FWT	FWR	FWP	DWT	DWR	DWP	ΔHT
Temp	25.4	.019	14.8	6.7	1.0	6.0	4.2
Dose	4.0	36.7	17.0	2.7	17.8	5.1	2.2
CV	13.2	18.6	12.4	15.4	15.4	14.4	15.3

Rank of F:

	FWT	FWR	FWP	DWT	DWR	DWP	ΔHT
Temp	1	7	2	3	6	4	5
Dose	5	1	3	6	2	4	7

Rank of CV:

	FWT	FWR	FWP	DWT	DWR	DWP	ΔHT
	2	7	1	5.5	5.5	3	4

Rank of MAV:

	FWT	FWR	FWP	DWT	DWR	DWP	ΔHT
	2	1	-	4	5	-	3

The ranks of F ratios for Temperature were highly consistent with those of other similar experiments, confirming these results. The use of Fisher's Z as suggested in the methodology was not followed because of the single degree of freedom in the numerator of F. Results very similar to these for the ranks of F for Dose were also found in other experiments. Again the Z transform was not computed. What was of primary interest was the almost diametric failure of the temperature and dose effects to agree for FWT and FWR. When I pointed this out to the plant physiologist, he set out to construct a functional model based on the tops being more responsive to the temperature effects and the roots more responsive to doses which had been applied directly to the roots in the nutrient solution. Moreover, these results confirmed a long established practice of analyzing top and root tissues separately. Whether or not to choose Fresh or Dry was left to the experimenter's judgment depending upon the importance of water content in the plant tissues. Little is said of the results for the coefficients of variation; they seemed to follow closely the results for temperature.

The MAV rankings confirmed that the primary information lay in fresh weight of tops and roots, that for the dry weight apparently being rather highly correlated with that for the fresh weights--an expected result. The addition of growth as a third response as suggested by the MAV ranks was not strongly backed by the univariate analyses. Whether or not to include it as a response may be better answered by its importance experimentally, e.g. does tallness matter, or by the cost of measuring it.

REFERENCES

1. Smith, Gnanadeskian, and Hughes. Multivariate Analysis of Variance. Biometrics Vol #18 No.1,p 22 ff. March 1962

2. Kendall, M. G.; Advanced Theory of Statistics, Vol II, p 116 ff. 1946

RESPONSE SURFACE ANALYSIS FOR DUAL RESPONSE SYSTEMS

Raymond H. Myers

Department of Statistics
Virginia Polytechnic Institute
and State University

ABSTRACT

The purpose of this work is to develop the theory associated with a dual response surface model. A dual response surface system is assumed and the theoretical framework is developed for arriving at "optimum" conditions on a set of independent variables.

The approach is to find conditions which maximize a "primary response" subject to the constraint that a "secondary response" takes on some specified or desirable value. An algorithm is outlined whereby a user can generate simple two dimensional plots to determine the conditions of constrained maximum primary response subject to the secondary response taking on any value he wishes. He, thus, is able to reduce to simple plotting the complex task of exploring the dual response system.

In certain situations it becomes necessary to apply a double constraint, the second being that the located operating conditions be a certain "distance" from the origin of the independent variables, (or the center of the experimental design).

The procedure applies to optimizing in cases where it is desirable to employ two measures of effectiveness, cost often being the prime candidate for the secondary response and a single measure of performance as the primary response.

Preceding page blank

RESPONSE SURFACE EXPLORATION IN PROBLEMS
INVOLVING TWO RESPONSES

1. **Introduction**

Much has been written concerning the exploration of an experimental region using response surface methods. Basically, a polynomial type response function is used to graduate a mechanism given by

$$\eta = g(x_1, x_2, \ldots, x_k)$$

in some experimental region. The most frequently fitted response function and the one to be used here is the quadratic model which gives rise to a fitted response function of the form

$$\hat{y} = b_0 + \underline{x}'\underline{b} + \underline{x}'B\underline{x}, \tag{1.1}$$

where $\underline{x}$ is a vector of independent or design variables and $\hat{y}$ is the estimated response. The elements in $\underline{b}$ and B represent least squares estimators, the latter being a k x k matrix

$$B = \begin{bmatrix} b_{11} & b_{12}/2 & \cdots & b_{1k}/2 \\ & b_{22} & \cdots & b_{2k}/2 \\ & & & \vdots \\ & & & b_{kk} \end{bmatrix}$$

where the b_{ij} are <u>second order coefficients</u>. The total exploration following the estimation of (1.1) involves finding the stationary point

$$\underline{x}_0 = - B^{-1} \underline{b}/2$$

and conducting a canonical analysis to determine the nature of the stationary point. Discussions of these procedures are given in [1], [2], and [5].

Quite often the researcher is confronted with the problem of simultaneous optimization of two or more response variates. It is not unusual in this

situation to obtain a solution, $\underline{x}$, which is optimal for one response and far from optimal or even physically impractical for the other(s). The task is then to arrive at some compromise conditions involving the two responses. The problem is a natural one but only a few papers dedicated to it have appeared in the statistical literature. See for example [4] and [6].

2. The Dual Response Problem

Let us suppose that the experimenter has a primary response, with fitted response function given by

$$\hat{y}_p = b_0^{(1)} + \underline{x}'\underline{b}^{(1)} + \underline{x}'B^{(1)}\underline{x} \qquad (2.1)$$

and what we shall refer to as a secondary response (although indeed the two responses may be equally important) with response function given by

$$\hat{y}_s = b_0^{(2)} + \underline{x}'\underline{b}^{(2)} + \underline{x}'B^{(2)}\underline{x} \qquad (2.2)$$

The expression in (2.2) may have been obtained from the same experiment through the use of multivariate multiple regression or perhaps externally. The latter may be the case when the secondary response is the cost variable in say a yield-cost study. Indeed, the coefficients in (2.2) may possibly not be random variables

The solution proposed and discussed in the sequel is to find the conditions on $\underline{x}$ which optimize $\hat{y}_p$ subject to $\hat{y}_s = k$, where k is some desirable or acceptable value of the secondary response. (Actually, there are situations in which it is necessary to consider a double constraint. This will be discussed in a later section). To arrive at the solution mentioned above, Lagrangian multipliers are needed. Thus, we consider

$$L = b_0^{(1)} + \underline{b}^{(1)}{}'\underline{x} + \underline{x}'B^{(1)}\underline{x} - \mu(b_0^{(2)} + \underline{b}^{(2)}{}'\underline{x} + \underline{x}'B^{(2)}\underline{x} - k)$$

and require solutions for $\underline{x}$ to the set of equations

$$\frac{\partial L}{\partial \underline{x}} = \underline{0}$$

which results in the following:

$$(B^{(1)} - \mu B^{(2)})\underline{x} = \tfrac{1}{2}(\mu \underline{b}^{(2)} - \underline{b}^{(1)}). \qquad (2.3)$$

It is important at this point to study the nature of the "stationary point" generated by equation (2.3). We begin by considering the matrix of second partial derivatives, the (i, j) element of which is

$$\frac{\partial^2 L}{\partial x_i \, \partial x_j} \quad (i, j = 1,2,\ldots,k).$$

It follows immediately that

$$M(\underline{x}) = 2(B^{(1)} - \mu B^{(2)}). \qquad (2.4)$$

Much of the development that follows is somewhat similar to the approach taken by Draper [3] in Ridge Analysis. In fact, one can consider Ridge Analysis in which it is desired to maximize an estimated response, $\hat{y}$, subject to the constraint $\underline{x}'\underline{x} = R^2$, as a special case of the dual response problem. However, in the dual response problem, the solution must depend on the nature of the matrices $B^{(1)}$ and $B^{(2)}$.

It is well known that if the matrix of second partial derivatives given by equation (2.4) is negative definite, the value of $\underline{x}$ generated by equation (2.3) will give rise to a <u>local maximum</u> on $\hat{y}_p$ (local minimum if the matrix of second partials is positive definite). Therefore, rather than fixing $\hat{y}_s = k$, an appropriate procedure would be to select directly values of the Lagrange multiplier, μ, in the region which gives rise to operating conditions on $\underline{x}$ from (2.3) that result in <u>absolute maxima</u> on y_p, conditional on being on a surface of the secondary response given by (2.2). In what follows, we make use of the following theorem.

<u>Theorem 2.1</u>: Let $\underline{x}_1$ and $\underline{x}_2$ be solutions to equation (2.3), using μ_1 and μ_2 respectively and let $\hat{y}_{s,1} = \hat{y}_{s,2}$. If the matrix $(B^{(1)} - \mu_1 B^{(2)}$ is negative definite then $\hat{y}_{p,1} > \hat{y}_{p,2}$. It also follows that if $(B^{(1)} - \mu_1 B^{(2)})$ is positive

definite, then $\dot{y}_{p,1} < \dot{y}_{p,2}$.

Proof

If $\underline{x}_1$ and $\underline{x}_2$ give rise to the same value of the secondary response, then

$$b_0^{(2)} \underline{x}_1' \; B^{(2)} \underline{x}_1 + \underline{x}_1' \; \underline{b}^{(2)} = \underline{x}_2' \; B^{(2)} \; \underline{x}_2 + \underline{x}_2' \; \underline{b}^{(2)} + b_0^{(2)} \; . \tag{2.5}$$

Consider now $\dot{y}_{p,1} - \dot{y}_{p,2}$. We can write

$$\dot{y}_{p,1} - \dot{y}_{p,2} = \underline{x}_1' \; B^{(1)} \; \underline{x}_1 - \underline{x}_2' \; B^{(1)} \; \underline{x}_2 + (\underline{x}_1' - \underline{x}_2') \underline{b}^{(1)} \; .$$

By adding and subtracting $\mu_1 \underline{x}_2' B^{(2)} \underline{x}_2$, we obtain

$$\dot{y}_{p,1} - \dot{y}_{p,2} = \underline{x}_1' \; B^{(1)} \underline{x}_1 - \underline{x}_2' \; (B^{(1)} - \mu_1 \; B^{(2)}) \underline{x}_2 - \mu_1 \underline{x}_2' \; B^{(2)} \underline{x}_2$$
$$+ (\underline{x}_1' - \underline{x}_2') \; \underline{b}^{(1)} \; . \tag{2.6}$$

From equation (2.3) with $\mu = \mu_1$ and $\underline{x} = \underline{x}_1$, we have

$$\underline{x}_1' \; B^{(1)} \; \underline{x}_1 = \mu_1 \underline{x}_1' \; B^{(2)} \underline{x}_1 + \tfrac{1}{2} \mu_1 \underline{x}_1' \; \underline{b}^{(2)} - \tfrac{1}{2} \underline{x}_1' \; \underline{b}^{(1)}$$

which from (2.5) becomes

$$\underline{x}_1' \; B^{(1)} \; \underline{x}_1 = \mu_1 \dot{y}_s - \mu_1 b_0^{(2)} - \tfrac{1}{2} \mu_1 \underline{x}_1' \; \underline{b}^{(2)} - \tfrac{1}{2} \underline{x}_1' \; \underline{b}^{(1)} \; . \tag{2.7}$$

From (2.5) we also have

$$\underline{x}_2' \; B^{(2)} \; \underline{x}_2 = \dot{y}_s - \underline{x}_2' \; \underline{b}^{(2)} - b_0^{(2)} \; . \tag{2.8}$$

Hence, from (2.7) and (2.8) it follows that

$$\underline{x}_1' \; B^{(1)} \; \underline{x}_1 - \mu_1 \underline{x}_2' \; B^{(2)} \; \underline{x}_2 = - \tfrac{1}{2} \mu_1 \underline{x}_1' \; \underline{b}^{(2)} + \mu_1 \underline{x}_2' \; \underline{b}^{(2)} - \tfrac{1}{2} \underline{x}_1' \; \underline{b}^{(1)} \; .$$

Thus, (2.6) becomes

$$\dot{y}_{p,1} - \dot{y}_{p,2} = (\underline{x}_2' - \tfrac{1}{2} \underline{x}_1') \; (\mu_1 \underline{b}^{(2)} - \underline{b}^{(1)}) - \underline{x}_2' (B^{(1)} - \mu_1 \; B^{(2)}) \underline{x}_2 \; .$$

From equation (2.3), we have

$$(\underline{x}_2{}' - \tfrac{1}{4}\,\underline{x}_1{}')\,(\mu_1\,\underline{b}^{(2)} - \underline{b}^{(1)}) = 2\underline{x}_2{}'\,(B^{(1)} - \mu_1\,B^{(2)})\,\underline{x}_1 - \underline{x}_1{}'\,(B^{(1)} - \mu_1\,B^{(2)})\,\underline{x}_1$$

and, as a result $\hat{y}_{p,1} - \hat{y}_{p,2} = (\underline{x}_2 - \underline{x}_1)'\,(B^{(1)} - \mu_1\,B^{(2)})\,(\underline{x}_1 - \underline{x}_2)$

$$= -\,(\underline{x}_1 - \underline{x}_2)'\,(B^{(1)} - \mu_1 B^{(2)})\,(\underline{x}_1 - \underline{x}_2). \qquad (2.9)$$

which is positive if $B^{(1)} - \mu_1\,B^{(2)}$ is a negative definite matrix and negative if $B^{(1)} - \mu_1\,B^{(2)}$ is a positive definite matrix.

Theorem (2.1) indicates that in the quest for values of $\underline{x}$ which yield constrained maxima (minima) we can limit ourselves to values of μ which make $B^{(1)} - \mu\,B^{(2)}$ negative definite (positive definite) assuming that such values exist. It shall be demonstrated that this "working region" in μ does often exist and that its location depends on the nature of the matrices $B^{(1)}$ and $B^{(2)}$. Equation (2.9) also indicates

$$\hat{y}_{p,1} - \hat{y}_{p,2} = (\underline{x}_1 - \underline{x}_2)'(B^{(1)} - \mu_2\,B^{(1)})\,(\underline{x}_1 - \underline{x}_2)$$

which implies that while $B^{(1)} - \mu_1\,B^{(2)}$ is negative definite, $B^{(1)} - \mu_2\,B^{(2)}$ cannot be negative definite unless both give rise to the same solution for $\underline{x}$. It will become apparent later that the latter cannot occur.

2.1 $B^{(2)}$ Positive Definite

Suppose that the stationary point of the secondary response results in a minimum, implying that $B^{(2)}$ is positive definite. Consider the quadratic form with matrix given by $M(\underline{x})$, i.e.,

$$q = \underline{u}'(B^{(1)} - \mu\,B^{(2)})\,\underline{u}$$

Since $B^{(2)}$ is symmetric positive definite, there exists a nonsingular matrix R (Rao [7]) such that

$$R'\,B^{(1)}R = \mathrm{diag}(\lambda_1,\ \lambda_2,\ \ldots\ \lambda_k)$$

-684-

and

$$R' B^{(2)} R = I_k.$$

Performing the transformation

$$\underline{u}' = \underline{v}' R'$$

we have

$$q = \underline{v}' \text{ diag } (\lambda_1 - \mu, \ \lambda_2 - \mu, \ \ldots, \ \lambda_k - \mu) \ \underline{v}. \tag{2.10}$$

The λ's are merely the eigenvalues of the real, symmetric matrix

$$D_2^{(-\frac{1}{2})} Q' B^{(1)} Q D_2^{(-\frac{1}{2})} = S. \tag{2.11}$$

Here Q is the orthogonal matrix for which

$$Q' B^{(2)} Q = D_2 \tag{2.12}$$

and D_2 is the diagonal matrix containing the eigenvalues of $B^{(2)}$. We use the notation $D_2^{(-\frac{1}{2})}$ to denote a diagonal matrix containing the reciprocals of the square roots of the eigenvalues of $B^{(2)}$. From equation (2.10), it is clear that we can insure a negative definite $M(\underline{x})$ if $\mu > \lambda_k$ (positive definite if $\mu < \lambda_1$) where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the eigenvalues of the matrix S arranged in ascending order.

In what follows, it becomes apparent that this indeed defines the working region for μ and, in fact, any $\mu_i > \lambda_k$ yields $\underline{x}_i$ which gives rise to an absolute maximum $\hat{y}_{p,i}$ (absolute minimum for $\mu_i < \lambda_1$) conditional on being on a surface of secondary response given by

$$\hat{y}_{s,i} = b_0 + \underline{x}_i' \ \underline{b}^{(2)} + \underline{x}_i' \ B^{(2)} \ \underline{x}_i.$$

It turns out that by choosing μ values in this region one generates $\underline{x}$'s which give all possible values of $\hat{y}_s$.

The following theorem will be useful in obtaining an understanding of the relationship between the Lagrangian multiplier, μ, and the resulting estimated

value of the secondary response function.

Theorem 2.2 : Let $\underline{x}$ be a solution to (2.3) where $B^{(2)}$ is positive definite.

Then $\dfrac{\partial^2 \dot{y}_s}{\partial \mu^2} \geqslant 0$ with the equality holding only in the limit as μ approaches $\pm \infty$.

<u>Proof</u>:

Differentiating both sides of equations (2.2) and (2.3) with respect to μ yields

$$\frac{\partial \dot{y}_s}{d \mu} = \underline{b}_2' \frac{\partial \underline{x}}{\partial \mu} + 2 \underline{x}' B^{(2)} \frac{d \underline{x}}{\partial \mu} \tag{2.13}$$

and

$$(B^{(1)} - \mu B^{(2)}) \frac{d \underline{x}}{\partial \mu} = \tfrac{1}{2} \underline{b}_2 + B^{(2)} \underline{x}. \tag{2.14}$$

Upon taking the second partial in (2.13) and (2.14) with respect to μ, one can write

$$\frac{\partial^2 \dot{y}_s}{\partial \mu^2} = \underline{b}_2' \frac{\partial^2 \underline{x}}{\partial \mu^2} + 2 \left[\underline{x}' B^{(2)} \frac{\partial^2 \underline{x}}{\partial \mu^2} + \frac{\partial \underline{x}'}{\partial \mu} B^{(2)} \frac{\partial \underline{x}}{\partial \mu} \right] \tag{2.15}$$

$$(B^{(1)} - \mu B^{(2)}) \frac{\partial^2 \underline{x}}{\partial \mu^2} = 2 B^{(2)} \frac{\partial \underline{x}}{\partial \mu} \tag{2.16}$$

Upon premultiplying (2.14) by $\dfrac{\partial^2 \underline{x}'}{\partial \mu^2}$ and (2.16) by $\dfrac{\partial \underline{x}'}{\partial \mu}$ and subtracting the result-

ing equations we find that

$$\tfrac{1}{2} \underline{b}_2' \frac{\partial^2 \underline{x}}{\partial \mu^2} = 2 \frac{\partial \underline{x}'}{\partial \mu} B^{(2)} \frac{\partial \underline{x}}{\partial \mu} - \underline{x}' B^{(2)} \frac{\partial^2 \underline{x}}{\partial \mu^2} \tag{2.17}$$

Substituting the expression for $\underline{b}_2' \dfrac{\partial^2 \underline{x}}{\partial \mu^2}$ from (2.17) into (2.15) results in
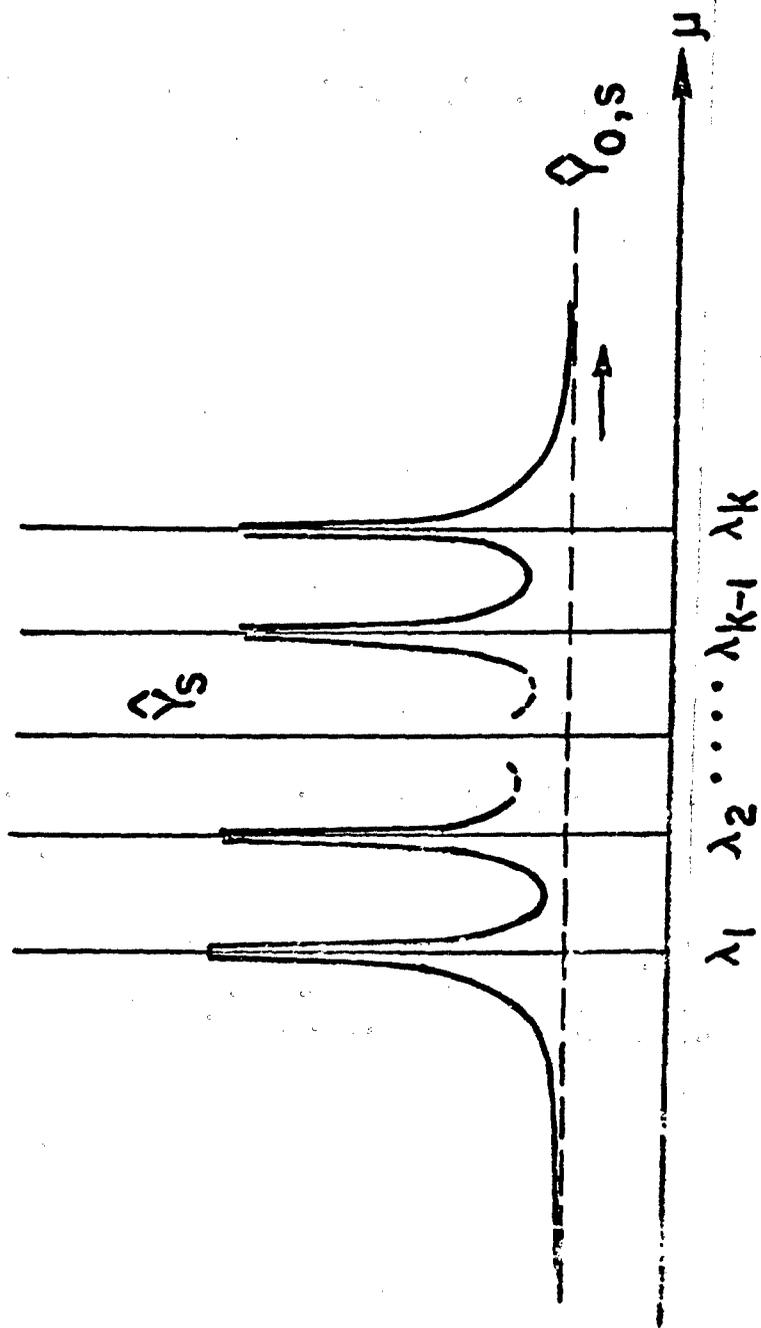
Fig. 2.1 PLOT OF $\hat{Y}_s$ AGAINST μ.

$$\frac{\partial^2 \hat{y}_s}{\partial \mu^2} = 6 \frac{\partial \underline{x}'}{\partial \mu} B^{(2)} \frac{\partial \underline{x}'}{\partial \mu}$$

which, of course, is positive except when $\frac{\partial \underline{x}}{\partial \mu} = \underline{0}$. From (2.3) and (2.14)

$\frac{d \underline{x}}{\partial \mu} = 0$ only in the limit as μ approaches either plus or minus infinity.

It is important to note that the relationship between y_s and μ is of the form illustrated in Figure (2.1). In the figure, $\hat{y}_{s,0}$ is the value of the estimated secondary response function at its stationary point, the latter being a point of minimum response. The existence of the asymptotes is easily seen since from (2.3)

$$\lim_{\mu \to \infty} \underline{x} = -\frac{B^{(2)^{-1}} \underline{b}^{(2)}}{2} = \underline{x}_{s,0}$$

which is the center or stationary point for the secondary system. As μ approaches λ_i ($i = 1,2,\ldots,k$), $\hat{y}_s$ approaches infinity since

$$|B^{(1)} - \lambda_i B^{(2)}| = 0. \qquad\qquad (i = 1,2,\ldots,k)$$

Hence, the asymptotes at the λ_i.

Theorems (2.1) and (2.2) indicate that the "working region" for μ resulting in a maximization of $\hat{y}_p$, subject to specific values of $\hat{y}_s$ is $\mu > \lambda_k$ and $\mu < \lambda_1$ for minimization. In a practical situation, interest would only be centered upon that part of the working region that generates values of $\hat{y}_s$ and thus $\underline{x}$ in the region of the experiment which generated either or both response functions. The procedure of determining operating conditions can be reduced to one of constructing a few simple graphs. Numerical examples of this procedure are given in a later section following the discussion of the problem for the case where $B^{(2)}$ is negative definite.

2.2 $B^{(2)}$ Negative Definite

When $B^{(2)}$ is negative definite, the stationary point for the secondary response function is a point of maximum response. Much of the development given in the previous section carries over, with a few modifications that deserve some attention. Consider again the matrix $M(x)$ given in equation (2.4) and the associated quadratic form $q = \underline{u}' (B^{(1)} - \mu B^{(2)})\underline{u}$. Again, there exists an orthogonal matrix Q for which

$$Q' B^{(2)} Q = D_2 \tag{2.18}$$

where D_2 is a diagonal matrix containing negative values. Let the matrix $D_2^* = - D_2$ and make the transformation

$$\underline{u} = Q D_2^{*(-\frac{1}{2})} \underline{y},$$

where $D_2^{*(-\frac{1}{2})}$ is diagonal containing reciprocals of square roots of the diagonal elements of D_2^*. Therefore, the quadratic form q can be written as

$$q = \underline{y}' \left[P' B^{(1)} P + \mu I \right] \underline{y} \tag{2.19}$$

where $P = Q D_2^{*(-\frac{1}{2})}$. The matrix $P' B^{(1)} P$ is real symmetric and thus there exists an orthogonal O for which

$$O' \left[P' B^{(1)} P \right] O = \Lambda^* \tag{2.20}$$

where Λ^* is a diagonal matrix of eigenvalues of $P' B^{(1)} P$. We can then make the orthogonal transformation

$$\underline{y} = O \underline{z} \tag{2.21}$$

and as a result

$$q = \underline{z}' [\Lambda^* + \mu I] \underline{z} \tag{2.22}$$

If we call Λ the diagonal matrix containing the eigenvalues of the symmetric matrix

$$S = D_2^{(-\frac{1}{2})} \, Q' \, B^{(1)} \, Q \, D_2^{(-\frac{1}{2})},$$

which is real in spite of the fact that $D_2^{(-\frac{1}{2})}$ contains purely imaginary values,

$$\Lambda^* = - \Lambda, \tag{2.23}$$

thus

$$q = \underline{z}' \, [\mu \, I - \Lambda] \, \underline{z}. \tag{2.24}$$

So in order to render q negative definite, and thus find $\underline{x}$ from (2.3) which maximize $\dot{y}_p$ subject to a constraint on $\dot{y}_s$, we are led to choosing values of μ which are smaller than the smallest eigenvalue of S. On the other hand, if our desire is to minimize $\dot{y}_s$, we find conditions by choosing μ larger than the largest eigenvalue of S.

A theorem analogous to Theorem 2.2 is again helpful in showing that constrained absolute maxima (minima) are obtained by choosing $\mu < \lambda_1$ ($\mu > \lambda_k$).

Theorem 2.3: Let $\underline{x}$ be a solution to (2.3) with $B^{(2)}$ negative definite. Then $\dfrac{\partial^2 \hat{y}_s}{\partial \mu^2} \leq 0$, with the equality holding in the limit as μ approaches $\pm \infty$.

Proof: The proof is similar to that of Theorem 2.2

As a result, the nature of the plot of $\hat{y}_s$ against μ is an inverted version of that given in Figure 2.1. That is, $\hat{y}_s$ will approach $- \infty$ as μ approaches an eigenvalue of S. For values of μ smaller than λ_1, $\hat{y}_s$ will increase with decreasing μ and asymptote to $\hat{y}_{s,0}$ which is its maximum value. Hence, the "working region" is $\mu < \lambda_1$ for constrained maximization of $\hat{y}_p$ and $\mu > \lambda_k$ for constrained minimization of $\hat{y}_p$.

3. Summary and Example for Case where $B^{(2)}$ is Definite

Perhaps the best way to summarize the results obtained when $B^{(2)}$ is definite is to outline the procedure which would be followed when it is of interest to obtain operating conditions resulting in a constrained optimum primary response variable. Following this outline will be a numerical example.

Once the parameters of the two response functions have been obtained, the eigenvalues of the matrix S should be determined. If one is interested in the constrained maximization of $\hat{y}_p$ and $B^{(2)}$ is positive definite, then values of $\mu > \lambda_k$ should be substituted into equation (2.3) and stationary values of $\underline{x}$ generated. These values of $\underline{x}$ represent points of absolute maximum response conditional on the estimated secondary response being given by equation (2.2). If minimization is desired, then values of $\mu < \lambda_1$ should be chosen. If $B^{(2)}$ is negative definite, values of $\mu < \lambda_1$ provide constrained maxima and values of $\mu > \lambda_k$ provide constrained minima.

Exploration of the dual response system can be carried out simply and concisely by constructing plots of x_1 vs. $\hat{y}_s$, x_2 vs. $\hat{y}_s$, ..., x_k vs. $\hat{y}_s$, and $\hat{y}_s$ vs. $\hat{y}_p$. When these simple two dimensional graphs are available to the experimenter, it will be possible for him to make a decision regarding what operating conditions should be used. In particular, for any value of the secondary response chosen, values of the x's are found which give rise to the maximum (or minimum) primary response.

One must be careful, of course, to consider as reliable only those results corresponding to values within or on the periphery of the experimental region. In addition, caution must be exercised in placing heavy reliance on results where either or both response functions are derived from empirical data that may have large random errors associated with them. (This too is a hazard with Ridge Analysis as pointed out by Draper.)

3.1 A Numerical Example

Consider a dual response surface problem where y_p and y_s depend on three independent variables x_1, x_2, and x_3. The following two response functions were fit to a set of experimental data

$$\hat{y}_p = 65.39 + 9.24x_1 + 6.36x_2 + 5.22x_3 - 7.23x_1^2 - 7.76x_2^2$$
$$- 13.11x_3^2 - 13.68x_1 x_2 - 18.92x_1 x_3 - 14.68x_2 x_3.$$

$$\hat{y}_s = 56.42 + 4.65x_1 + 8.39x_2 + 2.56x_3 + 5.25x_1^2 + 5.62x_2^2$$
$$+ 4.22x_3^2 + 8.74x_1 x_2 + 2.32x_1 x_3 + 3.78x_2 x_3$$

giving

$$B^{(2)} = \begin{bmatrix} 5.25 & 4.37 & 1.16 \\ & 5.62 & 1.89 \\ \text{sym} & & 4.22 \end{bmatrix}$$

with eigenvalues of $B^{(2)}$ being (10.553, 3.557, 0.979). Thus, the secondary response function yields a stationary point which is a point of <u>minimum response</u>, with the stationary point and the estimated response at the stationary point being

$$\underline{x}_{s,0} = - B^{(2)^{-1}} \frac{\underline{b}^{(2)}}{2} = \begin{bmatrix} 0.5194 \\ -1.178 \\ 0.0814 \end{bmatrix}, \quad \hat{y}_{s,0} = 52.79$$

For the primary response function

$$B^{(1)} = \begin{bmatrix} -7.23 & -6.84 & -9.46 \\ & -7.76 & -7.34 \\ \text{sym} & & -13.11 \end{bmatrix} \text{(eigenvalues are 0.1765, -2.6304 -25.6460)}$$

with $\underline{x}_{p,0} = - \frac{B^{(1)^{-1}} \underline{b}^{(1)}}{2} = \begin{bmatrix} -8.077 \\ 3.8862 \\ 3.8516 \end{bmatrix}$, $\hat{y}_{p,0} = 50.4849$
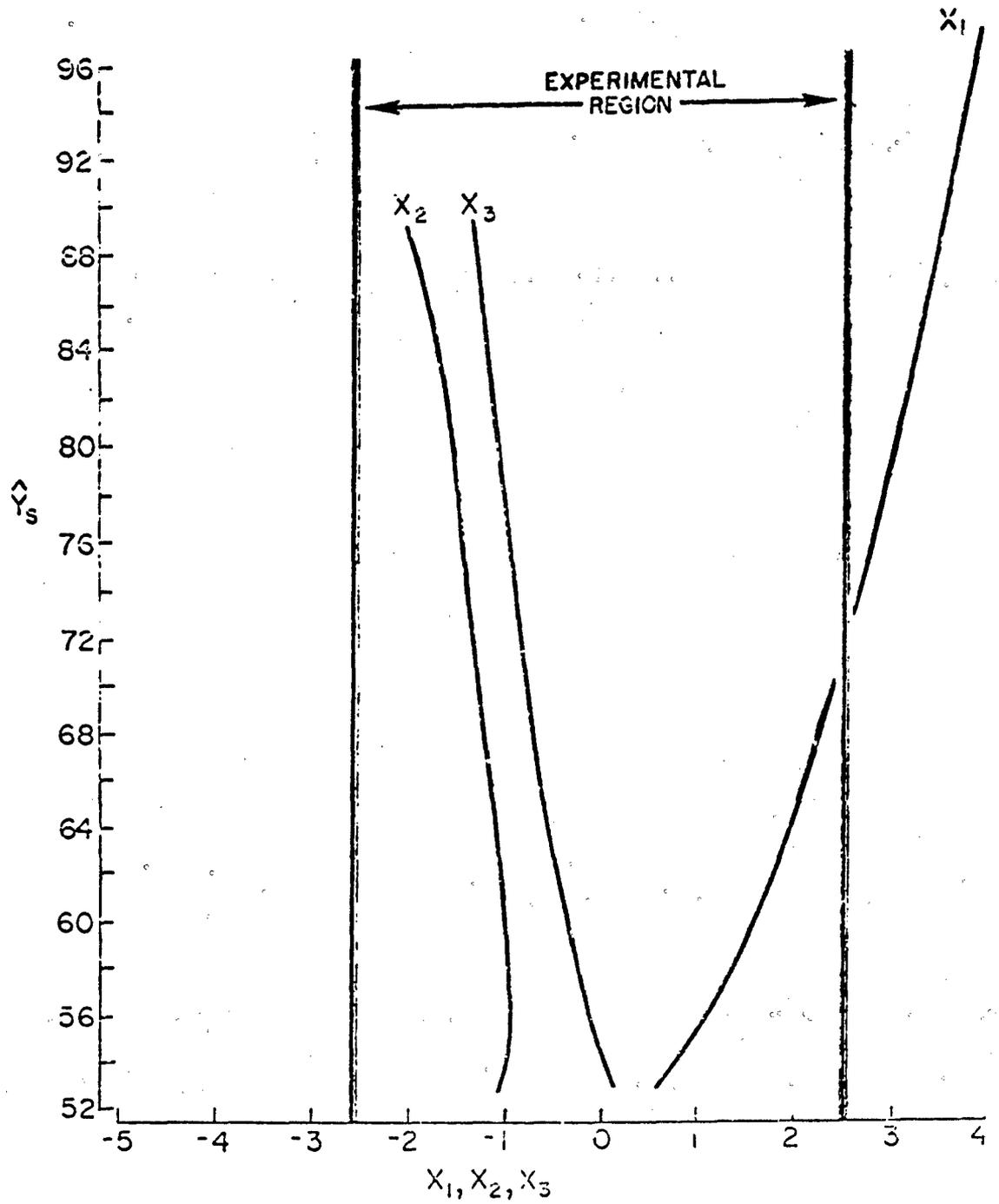
Fig. 3.1 CONDITIONS OF CONSTRAINED MAXIMA ON PRIMARY RESPONSE FOR FIXED VALUES OF $\hat{Y}_s$.
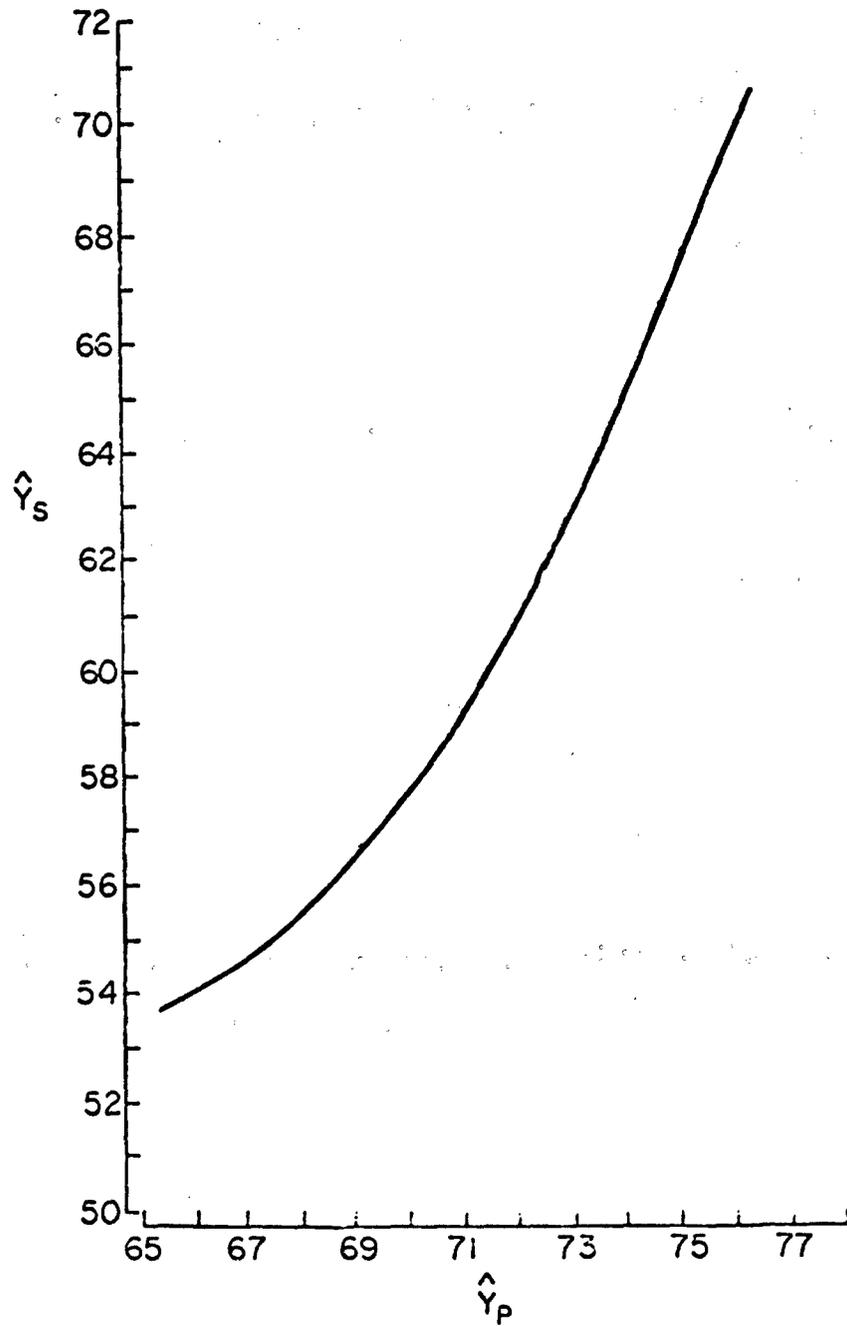
Fig. 3.2 MAXIMUM ESTIMATED PRIMARY RESPONSE AT SPECIFIC VALUES OF THE SECONDARY RESPONSE.

From these results, it follows that the primary response system is a "saddle system" with center at $x_{p,0}$ which is outside the experimental region. The goal of the investigation was to determine operating conditions which __maximize__ $\hat{y}_p$ but do not allow $\hat{y}_s$ to become too large. It was felt that values of the secondary response larger than about 65 would probably be excessive. Recall that the matrix S is given by

$$S = D_2^{(-\frac{1}{2})} \, Q' \, B^{(1)} \, Q \, D_2^{(-\frac{1}{2})}$$

For this example, we have

$$S = \begin{bmatrix} 0.3078 & 0 & 0 \\ 0 & 0.5302 & 0 \\ 0 & 0 & 1.0105 \end{bmatrix} \begin{bmatrix} 0.64276 & 0.69381 & 0.32478 \\ -0.34969 & -0.11148 & 0.93021 \\ 0.68159 & -0.71147 & 0.17097 \end{bmatrix} B^{(1)} \, Q \, D_2^{(-\frac{1}{2})}$$

$$= \begin{bmatrix} -2.0338 & -1.4100 & -0.7715 \\ -1.4100 & -1.4566 & -1.3899 \\ -0.7715 & -1.3899 & -1.4861 \end{bmatrix}$$

The eigenvalues of S are

$$\lambda = -4.0617 \qquad \lambda_2 = -0.9945 \qquad \lambda_3 = 0.08017$$

Equation (2.3) was used with $\mu > 0.08017$ to generate values of $\underline{x}$ representing points of constrained maximum primary response. Corresponding values of $\hat{y}_p$ and $\hat{y}_s$ were computed and the plots given in Figures 3.1 and 3.2 were constructed. Figure 3.1 indicates the locus of operating conditions giving absolute maxima on the primary response for various fixed values of the estimated secondary response. Figure 3.2 gives the value of the maximum estimated primary response for values of the estimated secondary response. In this example, the operating conditions which maximize $\hat{y}_p$, conditional on $\hat{y}_s = 65.0$ are found in Figure 3.1 to be

$$x_1 = 2.07; \ x_2 = -1.15; \ x_3 = -0.6$$

with an estimated primary response found in Figure 3.2 to be approximately 74.

These two dimensional plots can be very revealing in exploring dual response systems and the method of course can be used for any number of independent variables. A computer algorithm can be easily altered to handle the case where $B^{(2)}$ is negative definite. One merely needs to change all the signs from negative to positive of the eigenvalues of $B^{(2)}$ in forming D to avoid imaginary values. Then all of the signs of the elements in the resulting S matrix should be changed and the λ's obtained as the eigenvalues of this matrix are appropriate. Points of maximum $\hat{y}_p$ are then obtained by choosing $\mu < \lambda_1$ and points of minimum $\hat{y}_p$ are obtained by choosing $\mu > \lambda_k$.

4.0 $B^{(2)}$ Indefinite

When $B^{(2)}$ is indefinite, situations exist for which it is impossible to obtain a solution to the dual response optimization problem as it is currently stated. This will become obvious to the reader who attempts to maximize a two dimensional primary response system that is ellipsoidal in nature with a minimum at the center, given some specific value of a secondary variable with the latter having a saddle point system, i.e., $B^{(2)}$ is indefinite. No solution is found without further constraints.

In the case of a $B^{(2)}$ which is indefinite and the desire is a constrained maximization of $\hat{y}_p$ (constrained minimization is discussed in section 5.0), a solution exists if the primary response system yields a maximum at the center, i.e., $B^{(1)}$ is negative definite. Likewise, a constrained minimization is possible if the primary system yields a minimum at the center (constrained maximization is discussed in section 5.0). For the former case, consider the matrix of second partial derivative $M(\underline{x}) = 2(B^{(1)} - \mu B^{(2)})$. To make $M(\underline{x})$ negative definite, we require that the quadratic form

$$q = \underline{u}' [\mu B^{(2)} + (- B^{(1)})] \underline{u}$$

be positive definite. Again, we make use of the fact that there exists a non-

-696-

singular matrix R for which

$$R' \, B^{(2)} \, R = \text{diag} \, (\lambda_1, \, \ldots, \, \lambda_k)$$

$$R' \, (- \, B^{(1)}) \, R = I_k,$$

the roles of $B^{(1)}$ and $B^{(2)}$ having been reversed. The λ's are the eigenvalues of the matrix

$$S^* = D_1^{(-\frac{1}{2})} \, P' \, B^{(2)} \, P \, D_1^{(-\frac{1}{2})}, \tag{4.1}$$

where P is the orthogonal matrix for which

$$P' \, [- \, B^{(1)}] \, P = D_1$$

and $D_1^{(-\frac{1}{2})}$ contains reciprocals of the square roots of eigenvalues of $(- \, B^{(1)})$.

Letting $\underline{u} = R \, \underline{v}$, we have

$$q = \underline{v}' \, [\mu \, \text{diag} \, (\lambda_1, \, \ldots, \, \lambda_k) + I_k] \, \underline{v}.$$

The values of μ required to insure a local maximum on $\hat{y}_p$ are those for which $\mu \, \lambda_i > -1$ ($i = 1, 2, \ldots, k$) and, as a result, are the values of μ to employ in (2.3). From the definition of S^* and since $B^{(2)}$ is indefinite, the signs of the λ's will be mixed. Thus, the appropriate values of μ to use are given by the inequality

$$-\frac{1}{\lambda_1} > \mu > -\frac{1}{\lambda_k} \tag{4.2}$$

Again, a plot of $\hat{y}_a$ against μ in the working region of μ, is very revealing. Figure 4.1 indicates the appearance of this plot. In this case, the asymptotes will not be at $\mu = \lambda_i$ but rather at $\mu = -\frac{1}{\lambda_i}$ ($i = 1, 2, \ldots, k$), where a solution to equation (2.3) does not exist. To show this, we first consider

$$|B^{(1)} - \mu \, B^{(2)}| = (- \, \mu)^k \, |B^{(2)} + \frac{1}{\mu} \, (- \, B^{(1)})|.$$

Thus, we have

-697-
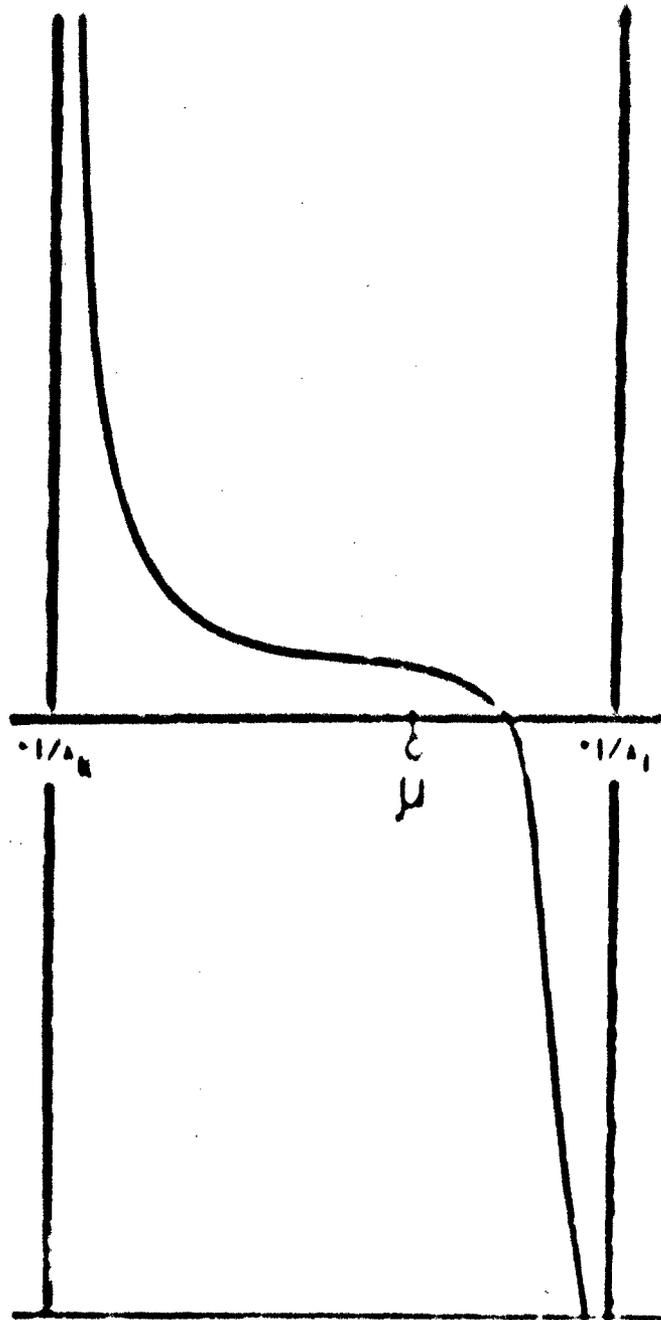
Fig. 6.1 TYPICAL PLOT OF $\hat{\theta}_i$ AGAINST μ FOR ϕ^2 INDEFINITE AND ϕ^{11} NEGATIVE DEFINITE IN THE WORKING REGION OF μ.

$$|D_1^{(-\frac{1}{2})}|\,|P'|\,|B^{(1)} - \mu\,B^{(2)}|\,|P|\,|D_1^{(-\frac{1}{2})}| = (-\mu)^k|S^* + \tfrac{1}{\mu}I|.$$

Thus, if $\mu = -\frac{1}{\lambda_1}$, $|S^* + \frac{1}{\mu}I| = 0$ and thus $|B^{(1)} - \mu\,B^{(2)}| = 0$.

The first derivative $\dfrac{\partial\, y_s}{\partial\,\mu}$ is negative in the working region of μ, i.e., in the region given by equation (4.2), for we can write, by combining equations (2.13) and (2.14)

$$\frac{\partial\, y_s}{\partial\,\mu} = \tfrac{1}{4}\,(\underline{b}^{(2)'} + 2\underline{x}'B^{(2)})(B^{(1)} - \mu\,B^{(2)})(\underline{b}^{(2)} + 2B^{(2)}\,\underline{x}).$$

The derivative cannot be other than negative in this region since $B^{(1)} - \mu\,B^{(2)}$ is negative definite and $(\underline{b}^{(2)} + 2B^{(2)})\underline{x}$ can only be zero when μ is infinite.

Again, the methodology involves choosing values of μ, this time in the region given by equation (4.2), generating $\underline{x}$ values from (2.3), computing $\dot{y}_s$ and $\dot{y}_p$ and plotting graphs similar to those in Figures (3.1) and (3.2) to describe the dual response system in the experimental region and using these plots to arrive at appropriate operating conditions.

If $B^{(1)}$ is positive definite, operating conditions can be found which <u>minimize</u> $\dot{y}_p$ for specific values of $\dot{y}_s$. The procedure involves using values of μ in equation (2.3) for which $\mu\,\lambda_i < 1$ ($i=1,2,\ldots,k$) where again the λ's are eigenvalues of the matrix S^* as given in (4.1). In this case, P is the orthogonal matrix for which

$$P'B^{(1)}P = D_1.$$

So the range of μ to use in this case is given by

$$\frac{1}{\lambda_1} < \mu < \frac{1}{\lambda_k} \tag{4.3}$$

5.0 Double Constraint Exploration

In the previous sections, we have considered the exploration of the dual response system with the goal of finding conditions that optimize the primary response with a simple constraint, namely, that the secondary response takes on a specific value. However, the experimenter will encounter many situations where mathematically the solution is valid but the recommended operating conditions $\underline{x}$ fall outside the region of the experiment that generated the estimated response functions and thus, would not be considered reliable. In the example given in section 3.1, the method would not have been successful if the operating conditions given in Figure 3.1 for $\dot{v}_s = 65$ had fallen outside the experimental region.

It seems that an appropriate procedure to follow would be to apply the additional constraint $\sum_{i=1}^{k} x_i^2 = R^2$ (using $\underline{x} = \underline{0}$ as the origin in the design variables) and employ the procedure where R is small enough to insure a solution inside the region of the designed experiment.

In fact, if $B^{(2)}$ is indefinite, there are cases when such a procedure is necessary. The solution to the problem as it has now been stated is obtained by employing, again, the method of Lagrangian multipliers. Hence, we consider the function

$$L = \dot{y}_p - \mu(\dot{y}_s - k) - \gamma(\underline{x}'\underline{x} - R^2).$$ (5.1)

The equation $\frac{\partial L}{\partial \underline{x}} = \underline{0}$ implies that

$$(B^{(1)} - \mu B^{(2)} - \gamma I) \underline{x} = \frac{1}{2} (\mu \underline{b}^{(2)} - \underline{b}^{(1)}.$$ (5.2)

Perhaps the most effective method for solving (5.2) is to choose values of μ and γ directly, making appropriate choices to insure that the values of $\underline{x}$ represent operating conditions where the maximum (or minimum) on $\dot{y}_p$ is achieved. For

a given value of μ, the matrix of second partials

$$M(\underline{x}) = 2(B^{(1)} - \mu B^{(2)} - \gamma I)$$

is made negative definite (and thus, a local maximum achieved) by selecting $\gamma > \lambda_k$, where λ_k is the largest eigenvalue of the matrix $B^{(1)} - \mu B^{(2)}$. [See Draper [3]]. Values of $\gamma < \lambda_1$ should be taken for local minima. In fact, for $\mu = 0$, the problem reduces to Ridge Analysis where the locus of coordinates generated by (5.2) represents points of absolute maxima on $\dot{v}_p$ without the constraint on $\dot{y}_s$.

The choice of μ essentially defines the direction taken as one moves away from $\underline{x} = \underline{0}$. Again, various two dimensional plots describe the dual response system. This will become apparent in the following section.

5.1 Example

Two responses were fit to a set of experimental data involving $k = 2$ independent variables. The two response functions were found to be the following:

$$\dot{y}_p = 53.69 + 7.26x_1 - 10.33x_2 + 7.22x_1^2 + 6.43x_2^2 + 11.36x_1\,x_2.$$

$$\dot{y}_s = 82.17 - 1.01x_1 - 8.61x_2 + 1.40x_1^2 - 8.76x_2^2 - 7.20x_1\,x_2.$$

The method can be used for any value of k, the number of independent variables. The above example was used so that the response contours can be drawn for illustrative purposes. The center of the primary and secondary systems are given by

$$\underline{x}_{p,0} = \begin{bmatrix} -3.7197 \\ 4.0891 \end{bmatrix} \qquad \underline{x}_{s,0} = \begin{bmatrix} -0.439 \\ -0.311 \end{bmatrix}$$

$B^{(1)}$ has eigenvalues given by 12.5187 and 1.1313 and thus, $x_{p,0}$ represents a point of minimum response. The eigenvalues of B(2) are -9.9063 and 2.5463 and so the secondary response system is hyperbolic in nature. Figure 5.1 shows the dual response system. Of course, in this example, it is impossible to find conditions

-701-

which maximize $\hat{y}_p$ subject to specific values of $\hat{y}_s$. However, by applying the additional restriction that $\underline{x}'\underline{x} = R^2$ equation (5.2) can be used with different values of ν and various values of γ exceeding the largest eigenvalue of $B^{(1)} - \nu B^{(2)}$ to generate values of $\underline{x}$ satisfying the constraints and giving rise to optimal operating conditions. The two dimensional plots in Figures (5.2), (5.3) and (5.4) are helpful in providing an exploration of the system and providing a recommendation for future operating conditions.

Suppose for example that we wish to find conditions in the experimental region which maximize $\hat{y}_p$ but we also require $84 < \hat{y}_s < 88$. Figures (5.2) and (5.3) indicate 'candidates' for operating conditions. The $\nu = 0$ line represents maximization of y_p subject only to $\underline{x}'\underline{x} = R^2$. The line $\nu = -2$ at $R = 1.0$ appears to be the proper choice. Figure 5.4 gives the values of the coordinates, $x_1 = 0.85$ and $x_2 = -0.6$, with the estimated responses at these conditions given from Figures (5.2) and (5.3) as

$$\hat{y}_p = 67, \qquad \hat{y}_s = 87.8$$

BIBLIOGRAPHY

[1] Box, G. E. P. and Wilson, K. B., 1951. On the Experimental Attainment
of Optimum Conditions. Journal of the Royal Statistical Society B 13
pp. 1-45.

[2] Davies, O. L. Design and Analysis of Industrial Experiments, Hafner
and Company, New York, (1956).

[3] Draper, N. R. "Ridge Analysis" of response surfaces. Technometrics 5
No. 4, 469.

[4] Michaels, S. F. and Pengilly, P. J. Maximum Yield for Specified Cost.
Applied Statistics, Volume 12, No. 3, pp. 189-193, 1963.

[5] Myers, Raymond H. Response Surface Methodology, Allyn and Bacon, Inc.,
Boston, 1971.

[6] Umland, A. W. and Smith, W. N. The Use of LaGrange Multipliers with
Response Surfaces. Technometrics, Volume 1, No. 3, pp. 289-292, 1959.

[7] Rao, C. R. Linear Statistical Inference and Its Applications. John
Wiley and Sons, Inc., New York, 1965.
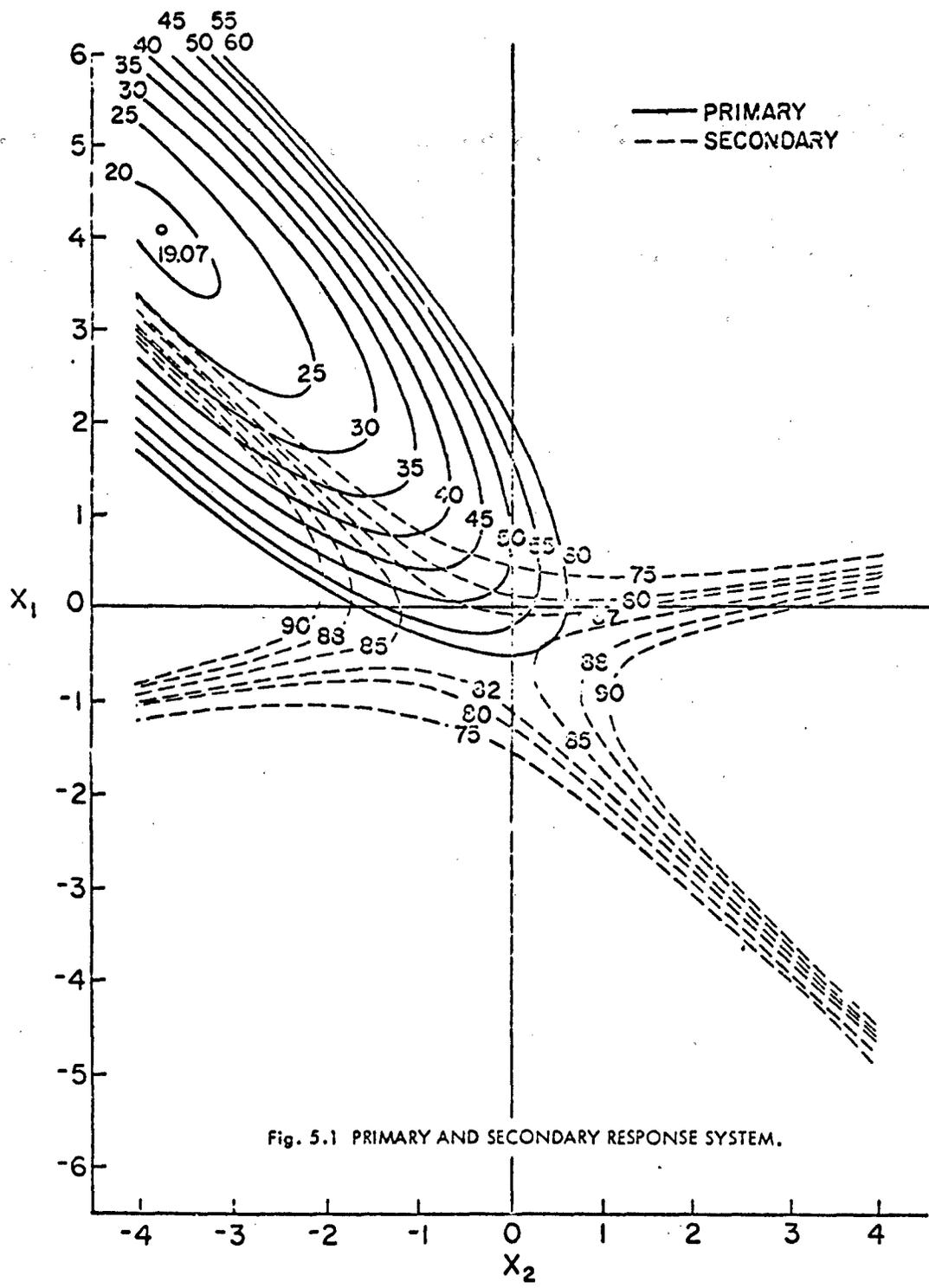
Fig. 5.1 PRIMARY AND SECONDARY RESPONSE SYSTEM.
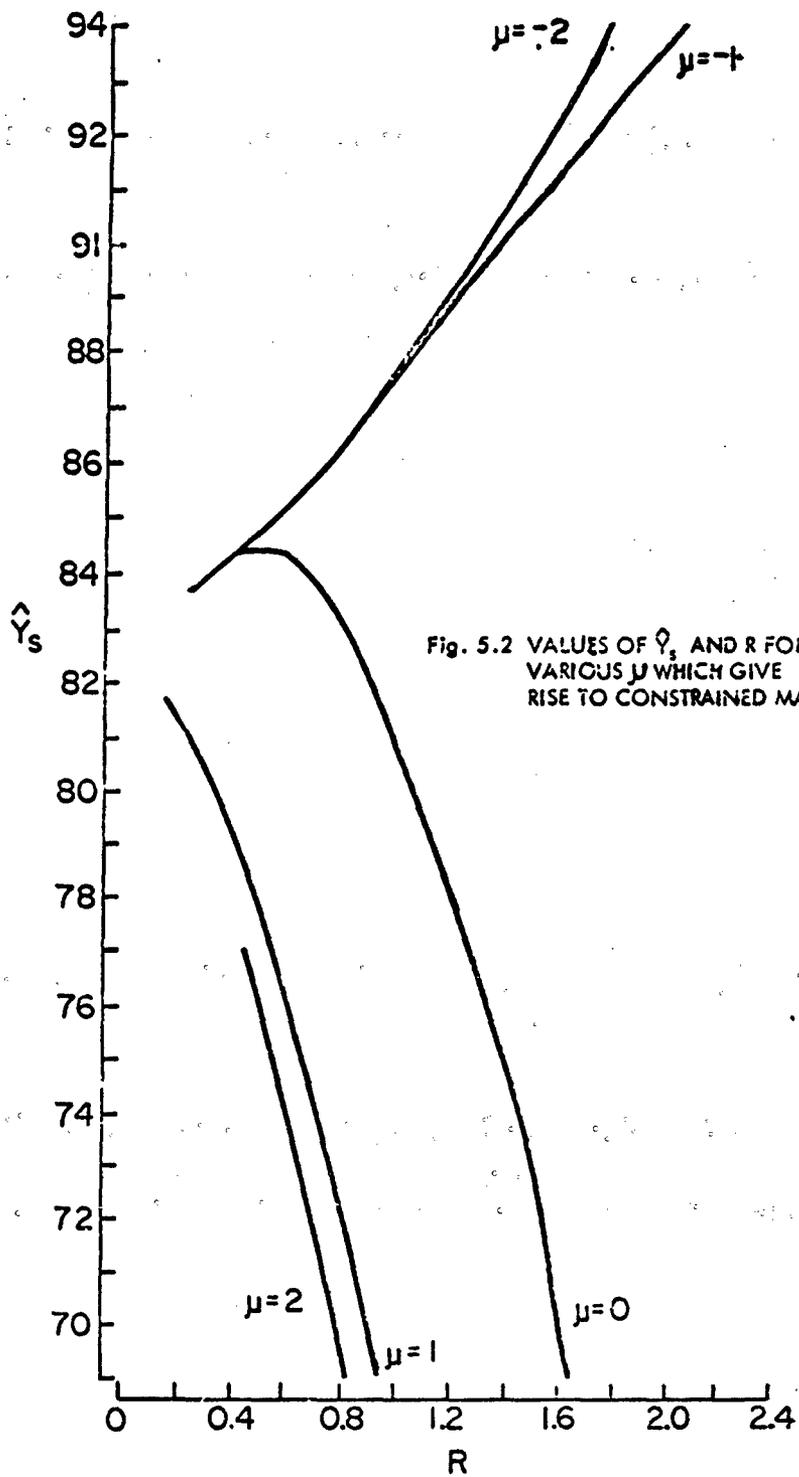
Fig. 5.2 VALUES OF $\hat{Y}_s$ AND R FOR VARIOUS μ WHICH GIVE RISE TO CONSTRAINED MAXIMA.
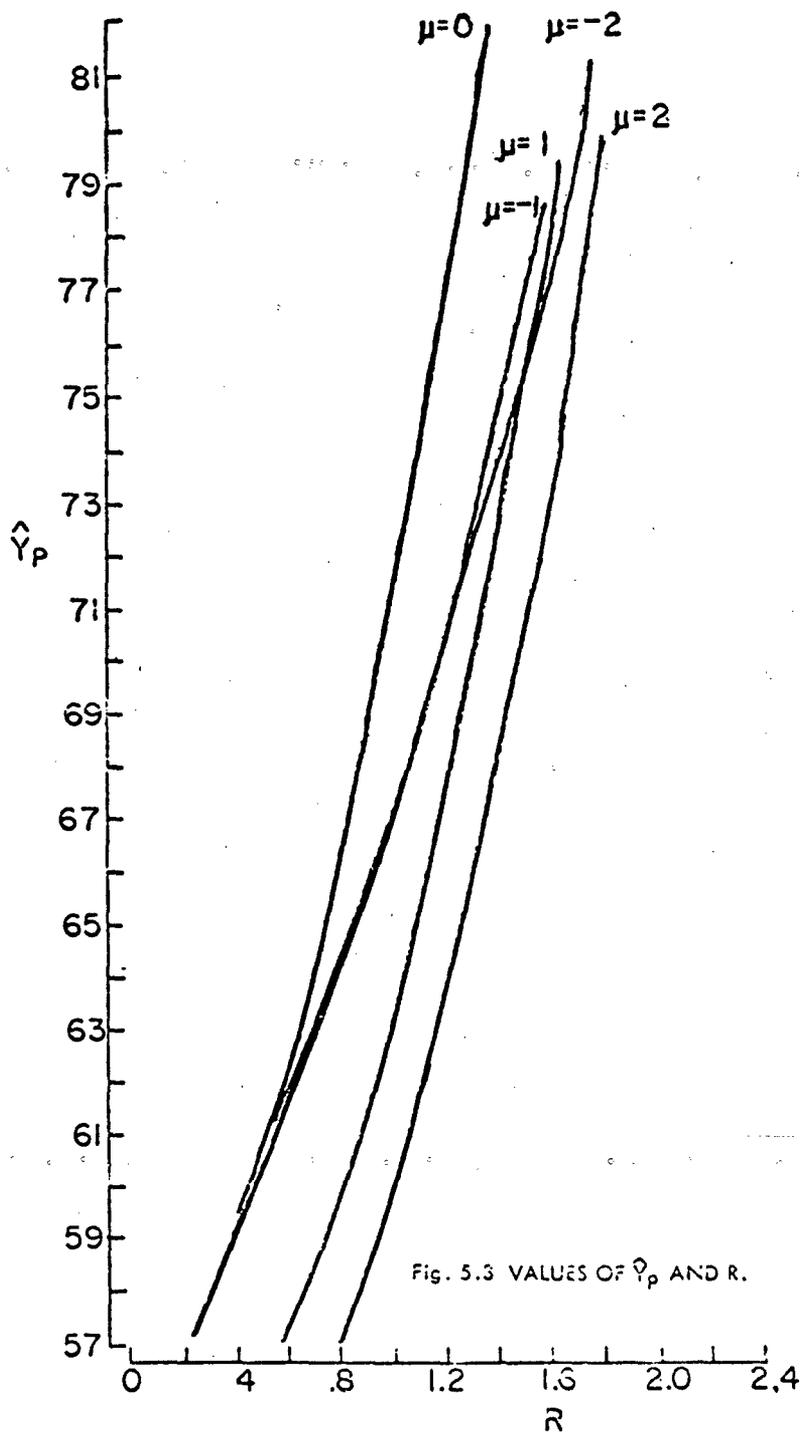
Fig. 5.3 VALUES OF $\hat{Y}_P$ AND R.
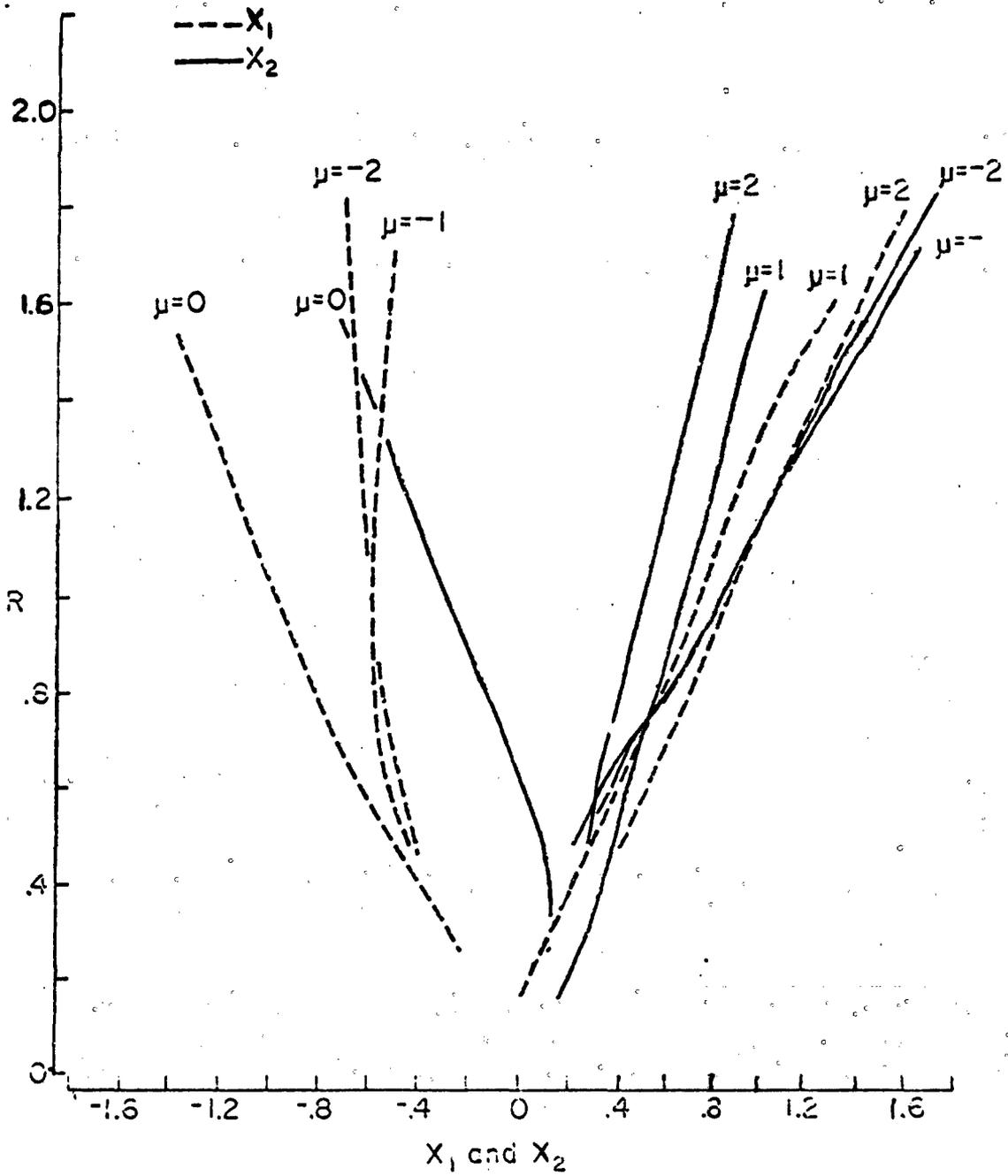
Fig. 5.4 OPERATING CONDITIONS FOR VARIOUS R.

OPTIMIZATION OF INPROCESS STORAGE DEVICES
USING A GERT III-Q SIMULATION

Ray L. Peterson[1]
US Army Munitions Command,
Frankford Arsenal,
Philadelphia, Pennsylvania

ABSTRACT. The purpose of this paper is to define storage devices, buffers, to insure continuous operation of the subsystem and optimize output efficiency for the entire system.

The optimization was achieved by using the simulation program GERT III-Q. A Delphi type method of gathering data was used to obtain imputs for this event type simulation.

The study shows that buffers contribute greatly to the total efficiency of the entire system by smoothing out the discontinuities in production due to failures. The optimum buffer size was approximately twenty minutes and the initial level should be full at the beginning of each day to maximize efficiency.

INTRODUCTION. During the Korean conflict there was little change in ammunition or manufacturing methods from those used in World War II and no real process changes of significance were made during the post-Korean period. New weapon systems gradually began to come into existance: the M-60 machine gun, the M-14 rifle, the M-16 rifle, and the minigun. These weapons required two new cartridges, the 7.62mm NATO and the 5.56mm rounds. Thus, when the Vietnam conflict began to involve more United States' forces, the World War II Small Caliber Ammunition Production Base was once again called upon to mass produce ammunition. The reactivating of existing production lines was complicated by the need to convert the equipment to manufacture a new family of small caliber ammunition. The conversion program was both costly and time-consuming. The age and condition of the equipment, coupled with the required modifications, dictated a major rebuild program. This required extensive retro-engineering and in-house manufacture of conversion and repair parts.

As a result of the increased requirement for Vietnam, it became apparent that improvement in production equipment was necessary. The slow production of cartridges was not compatible with the rapid-fire weapon concept currently gaining favor. Several studies were initiated to determine methods and means of improving and modernizing cartridge production. Results were promising. The search for modern technology of the small caliber ammunition production was initiated in mid 1968[2] at Frankford Arsenal when the United States Army Material Command implemented a program which established the Small Caliber Ammunition Modern-

---

ization Program (SCAMP). The present manufacturing techniques employed to produce small caliber ammunition were outdated due to the slow production, high costs, and the large storage areas required between operations. By providing a continuous process flow of raw materials and finished products, SCAMP intends to show its superiority over the batch process presently in use.

Due to the continuous flow nature of the SCAMP process, it readily lends itself toward an event type simulation. The following is a block diagram of the SCAMP process as applied to the 5.56mm production line. (Mod A)

The SCAMP process for the Mod A line is actually broken down into five discrete processes (submodules). Therefore, it can be easily seen that a breakdown in any one of the five submodules causes a disruption in the flow of the others. However, if an inprocess storage area was set aside for each of the submodules, a source to achieve continuous operation of the entire process could be provided. From this point on, the storage area will be referred to as the buffer level. Modifying the first flow chart, Figure 1, to include these buffer levels, it now takes on the following appearance (Figure 2).

The question that now presents itself is: What should be the optimal size and initial level of the buffer to insure continuous operation of the subsystem and optimize output efficiency?

The purpose of this paper is to develop a decision risk-analysis for the Mod A line to be utilized by the SCAMP management, and to show the effectiveness of GERT III-Q as applied to operations research type problems.

METHODOLOGY. In order to find the optimal buffer level in the Mod A line, an estimate efficiencies of the subsystems, a GERT III-Q simulation was chosen. (See Appendix A for a description of GERT III-Q).

Like any simulation, GERT III-Q needs input data from each submodule. The types of data needed are: the estimated availabilities for each submodule, the estimated repair times for each submodule, and the operating speed of the Mod A line. Since Mod A is a new process, there is no existing data available on the reliability of the submodule, so the data was gathered subjectively using a Delphi type[3] method.

This judgemental data, subjective ratings, were obtained from a series of Delphi type questionnaires administered to a group of ten experts from Frankford Arsenal and Government contractors. Three rounds of questionnaires requiring the same ratings were administered. In the second and third rounds, the experts were asked to reconsider their own ratings in the previous round in the light of the feedback information summarizing the groups ratings in the previous round. This mean was used to represent the concensus among the experts. (Appendix B is a sample questionnaire).

5.56mm PRODUCTION LINE
(Mod-A)

Input-Primers

Input-Brass Cups

5.56mm CASE → PRIMER INSERT

Input-Propellant

Input-Lead Wire
Input-Jackets Cups

BULLET

LOAD & ASSEMBLE

Input-Cardboard Boxes
Bandoliers
Crates
Ammo cans
Pallets

PACKAGING

PALLETIZED 5.56mm AMMUNITION

OUTPUT
- - - -
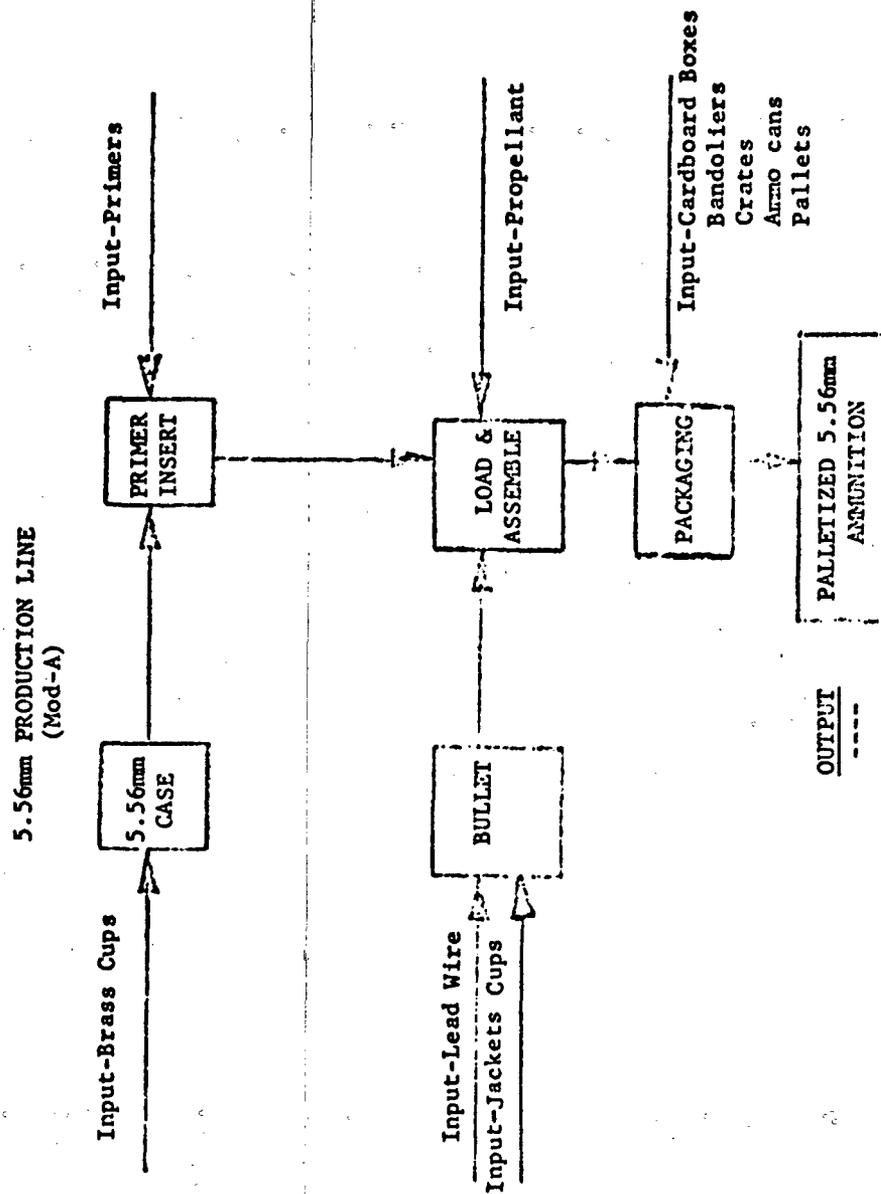
FIGURE 1.

5.56mm PRODUCTION LINE with BUFFERS
(Mod-A)

Input-Brass Cups → 5.56mm CASE → BUFFER → PRIMER INSERT ← Input-Primers → BUFFER → LOAD & ASSEMBLE ← Input-Propellant → BUFFER → PACKAGING ← Input-Cardboard Boxes / Bandoliers / Crates / Ammo Cans / Pallets

Input-Lead Wire / Input-Jacket Cups → BULLET → BUFFER → LOAD & ASSEMBLE

PACKAGING → PALLETIZED 5.56mm AMMUNITION

OUTPUT
- - - -
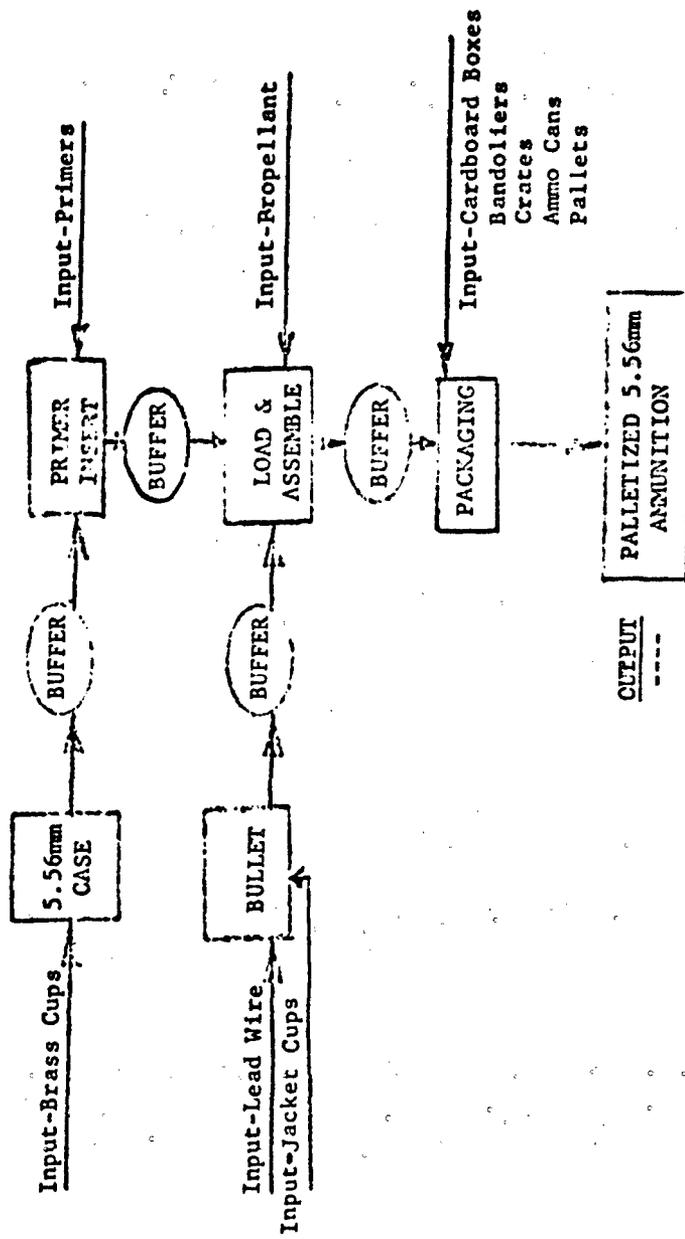
FIGURE 2

-712-

Figure 3 is the data in reduced form that was obtained from the questionnaires.[4]

Using the data from Figure 3, a series of 16 simulations were made using the mean values of the estimated availabilities with several different buffer sizes and levels. The size of the buffer is measured in minutes of production that it is capable of holding. The series was repeated for the three sigma limits of the availabilities. The upper bound on the availabilities became 100% for each submodule, which is a trivial case so the upper limit was changed to 95% available for a realistic performance.

Figure 4 shows the combinations of the simulations and their results which were carried out using the GERT III-Q program and exercised on the CDC-6400. Each simulation represents ten days production and the average efficiency for this period is calculated. Where average efficiency is defined as: the number of good pieces out divided by the number of good pieces possible. (Appendix C is a sample of the type output of this model).

Figures 4, 5, and 6 are graphs plotting average efficiency in percent versus the buffer sizes in minutes.

The differences between Figures 4, 5, and 6 are: Figure 4 represents the simulation with the mean values of the availabilities for the submodule, Figure 5 represents the simulation with the low values of the availabilities for the submodules, and Figure 6 represents the simulation for the 95% availabilities for the submodules.

The differences between the graphs appears to be that the curves tend not to bend as sharply the higher the availability. Therefore, that tends to show that the size of the buffer is not as critical to efficiency at the higher availabilities. For example, the change in efficiency from a ten-minute to a twenty-minute buffer, the low availability value is approximately 7%, while the change in the 95% availabilities is approximately 2%.

Another important observation made from the graphs is the full buffer, independent of the sizes or the availabilities of the submodules is more efficient. This can be observed in Figures 4, 5, and 6.

The third and most important observation made is the optimal buffer size. From the graphs it can be seen that the curves level off after approaching the twenty minute size. So the optimum buffer size is approximately twenty minutes, since at this point the percentage increase in efficiency decreases to almost zero, while the costs (time, space, and money) of making a larger buffer is going up.

Some of the assumptions of this study are: when the submodule is available it can be run at 1200 pieces per minute, no unacceptable pieces are formed by the submodules (which would decrease efficiency), the number of pieces that balk (overflow from a buffer) is not relevant to the optimum size, and that the repair times fit a Beta distribution around the mean for each submodule.

-713-

TABLE OF ESTIMATED AVAILABILITIES

TYPE OF SUBMODULE	$\bar{X}$ or AVERAGE ESTIMATED AVAILABILITIES in PER-CENT	$\bar{Y}$ or STANDARD DEVIATION of ESTIMATED AVAILABILITIES
CASE	78.0	8.1
BULLET	85.0	8.1
PRIMER INSERT	93.5	2.4
LOAD & ASSEMBLE	90.5	5.5
PACKAGING	94.0	2.1

TABLE OF ESTIMATED REPAIR TIMES

TYPE OF SUBMODULE	$\bar{M}$ or AVERAGE ESTIMATED REPAIR TIMES in PER-CENT	$\bar{N}$ or STANDARD DEVIATION of ESTIMATED REPAIR TIMES
CASE	15.0	5.3
BULLET	11.5	4.1
PRIMER INSERT	8.0	3.5
LOAD & ASSEMBLE	15.0	4.7
PACKAGING	10.0	4.7

FIGURE 3

TABLE LISTING SIMULATIONS AND RESULTS

BUFFER SIZE in MINUTES, and LEVEL at BEGINNING of EACH DAY	USING $\bar{X}$	USING $\bar{X}-3\bar{Y}$	USING 95%
	EFFICIENCIES IN PER-CENT		
10/full	70.7	46.2	88.7
10/half-full	69.5	45.4	87.6
10/empty	66.5	42.0	86.5
15/full	76.4	49.9	92.7
15/half-full	74.3	48.9	91.7
15/empty	70.3	45.1	88.2
20/full	78.7	51.7	94.3
20/half-full	76.2	50.4	93.4
20/empty	71.9	46.3	88.8
30/full	79.8	52.7	95.2
30/half-full	77.4	51.1	94.5
30/empty	72.9	46.7	89.2
40/full	80.0	53.0	95.3
40/half-full	77.6	51.5	94.7
40/empty	73.1	47.0	89.5
Infinite/empty	73.5	48.1	89.7

FIGURE 4.

GRAPH: Buffer Size vs Efficiency
Using Low Values Of The
Availabilities.

FIGURE 4

GRAPH: Buffer Size vs Efficiency
Using The Mean Values
Of The Availabilities

Average Total Efficiency (%)

Buffer Size (Minutes)

FULL

HALF-FULL

EMPTY

FIGURE 5

GRAPH: Buffer Size vs Efficiency
Using The 95% Value Of
The Availabilities

FULL

HALF-FULL

EMPTY

Average
Total
Efficiency
(%)

Buffer Size (Minutes)

FIGURE 6

-718-
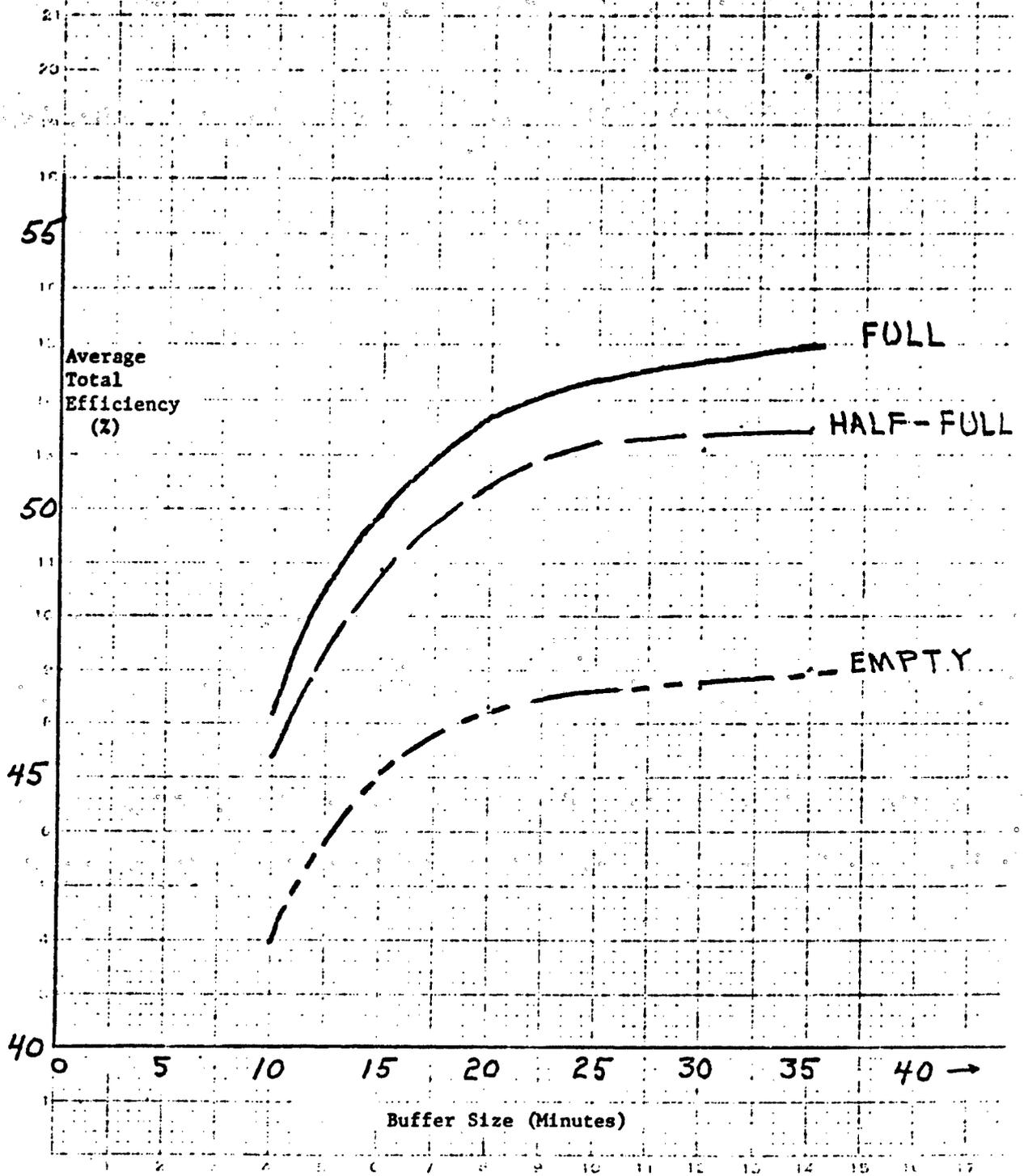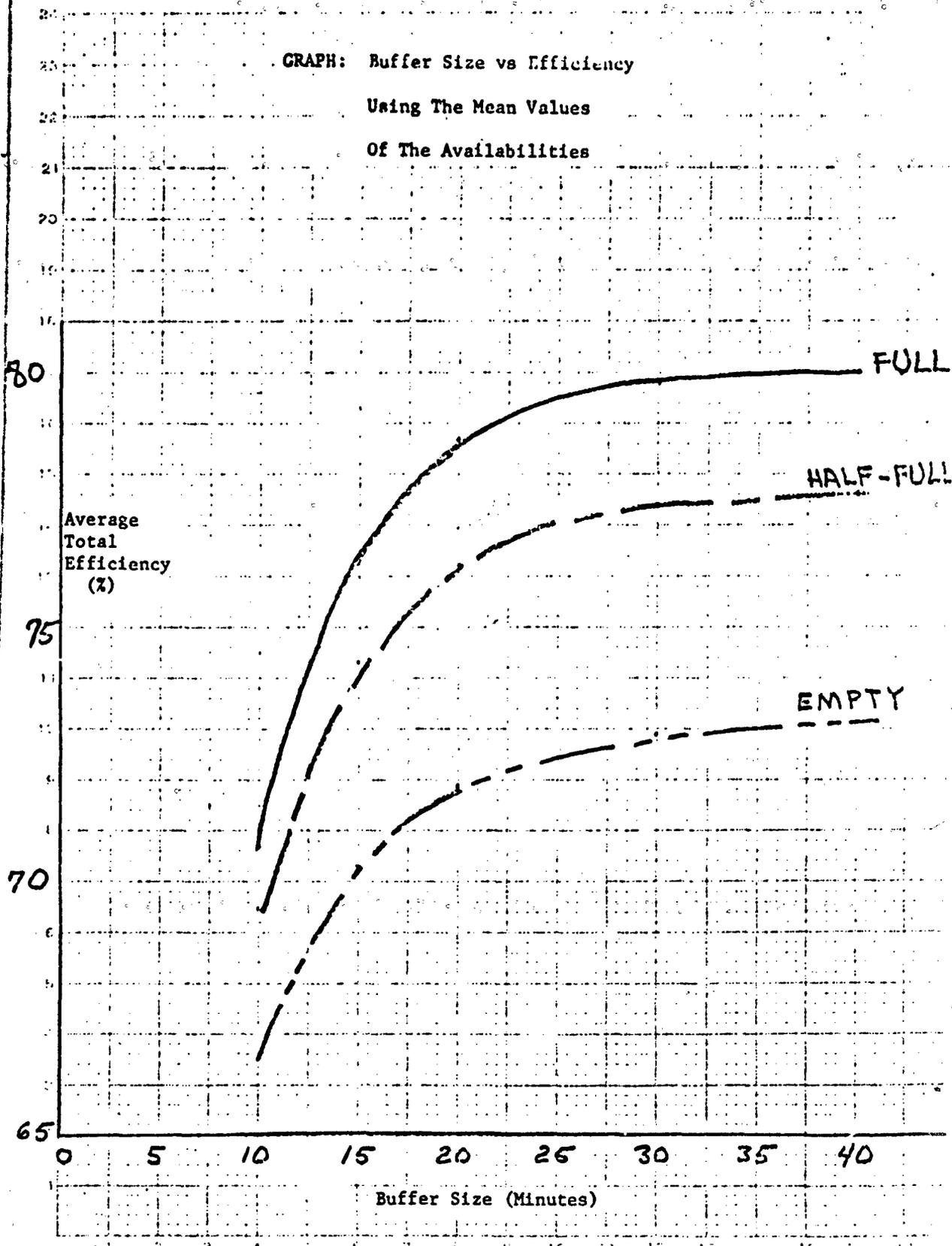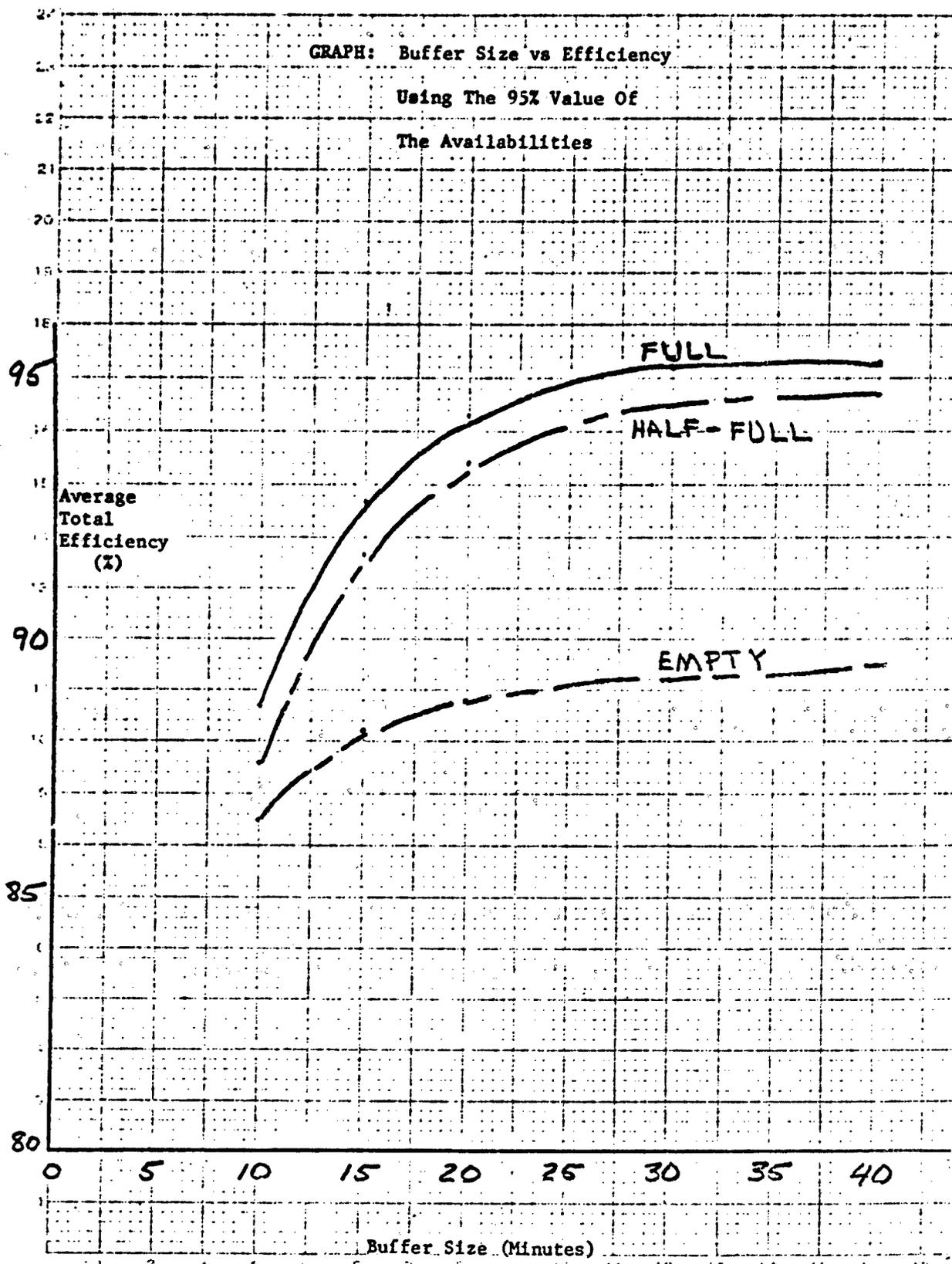
CONCLUSIONS. In summary, the conclusions of this paper are:

1. The optimal buffer level is approximately twenty minutes, independent of availabilities.

2. The optimal buffer level at the beginning of each day is full, independent of availabilities and the size of the buffer.

3. The higher the availabilities of the submodules, the less critical buffer size is to efficiency.
Other possible follow-up studies are: what effect does defective component parts have on the buffer; what effect would it have on efficiency to change the modular array of the submodules; the results should be updated when better repair times, availabilities and their distributions become available.

APPENDIX A
GERT III-Q

The GERT III-Q program performs a simulation of a network by advancing time from event to event (event simulation). The events and symbols associated with GERT III-Q network are:

1) Start of the simulation————————————————————————

2) End of an activity————————————————————————————

3) Completion of a simulation of the network——————————————

4) Storage capability or queue node——————————————————

The start event causes all source nodes to be realized and schedules the activities that emanate from the source nodes according to the output type of the source node. The output type for all nodes is either Deterministic or Probabilistic. In the former case, all activities starting from the node are scheduled. If an activity is completed that preceeds a Q-node, there are two things that can occur: the activity following the Q-node can be initiated; or the number in the queue can be increased by one. The Q-node can serve only one item at a time. If the node is in the process of serving an item, then the number in the queue increases by one.

Scheduling an activity means that an "end of activity" is caused to occur at some time in the future. The simulation proceeds from event to event until the simulation of the network is completed.

The above process is then repeated until the number of simulations of the network is completed.[5]

The following, Figure 7, is the nodal network of the Mod A production line used in the simulation. This figure also shows how it correlates to the block diagrams of the 5.56mm production line shown in the introduction.
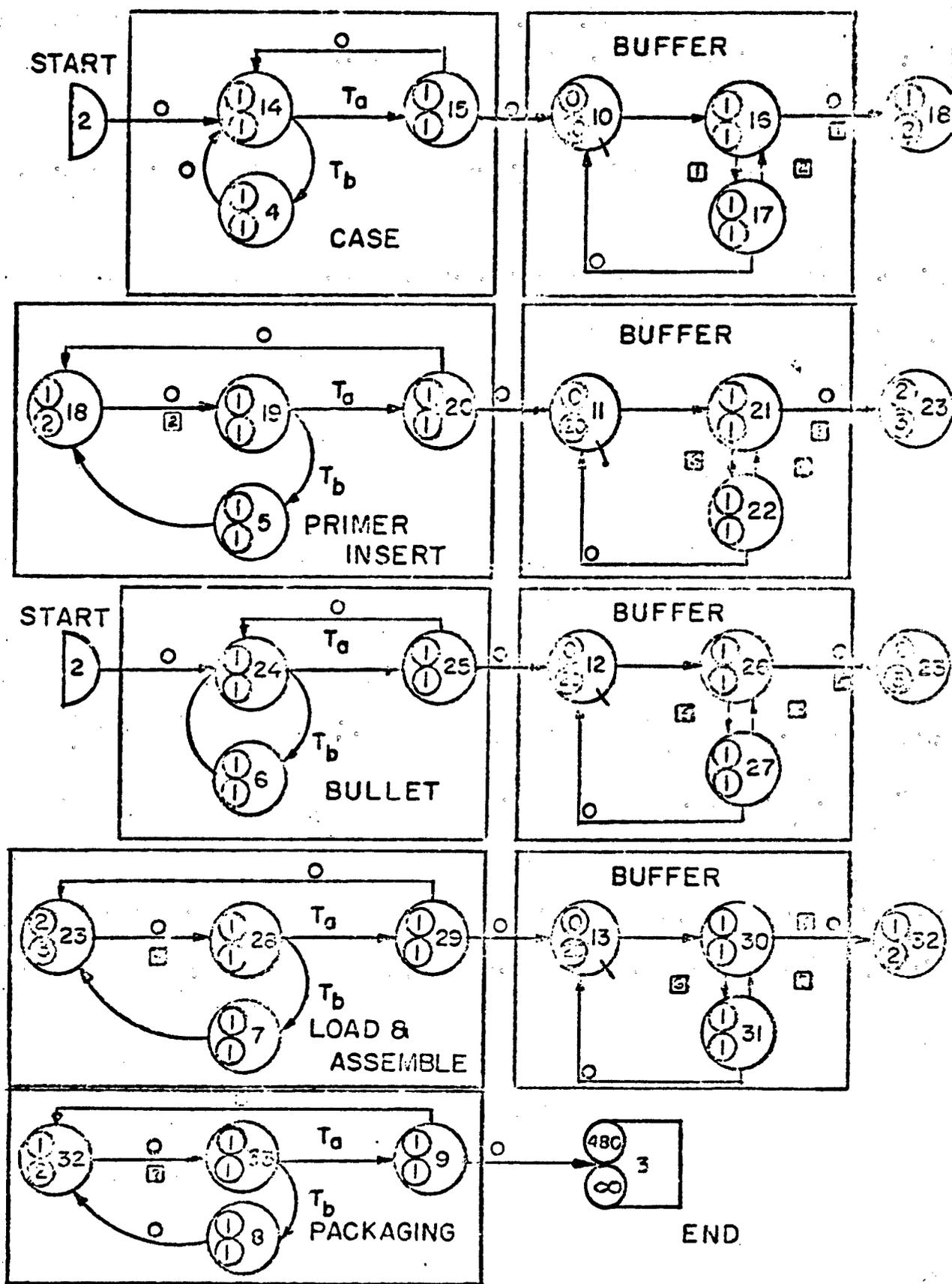
APPENDIX B
MOD A DECISION PROBLEM

NAME _____

ORGANIZATION AND/OR DEPARTMENT_____

NOTE: The supplementary material titled "Definition of Terms" has been
distributed for your easy reference in completing this questionnaire.

1. Frankford Arsenal has these objectives in modernizing its submodule
system:

    1) minimize system component scrap rate
    2) maximize system production rate

Rate the importance of these objectives relative to each other. Each rating
must lie between 0 and 100, and must be a multiple of 5. The sum of the
ratings must be 100. Note that a higher rating represents a more important
objective.

Objective	Priority Rating
Minimize Component Scrap	_____
Maximize Production Rate	_____
TOTAL	100

2. The effectiveness of Mod-A will be traceable to the quality of certain
submodules. Consider the five submodules.

1) CASE   _____

2) BULLET   _____

3) PRIMER INSERT   _____

4) LOAD & ASSEMBLE   _____

5) PACKAGING   _____

    Rate the submodules relative to how they meet Frankford's objectives.
The ratings must lie between 1 and 5. There is no other constraint.

3. Which one of the following would you rate the most important in
maximizing the production rate.

            1) in-process storage devices (buffers)
            2) batch system
            3) through process

    Rate the above relative to Frankford's objectives. The ratings
must lie between 0 and 7. The sum must add to 7.

4. If there is to be a buffer between each of the submodules, what size do you feel it should be to meet Frankford's objective?
Give your answer in minutes of producing 1200 pieces per minute. The answer should be in intervals of 5 minutes. Infinite is a valid answer.

_____minutes

5. Give estimates to the following questions based on the "2.5 million run" phase of Mod A.
1) Based on two shifts a day five days a week. How many minutes do you feel the submodule will be available in an eight hour day? There are 480 minutes possible.

CASE   _____minutes

BULLET   _____

PRIMER INSERT   _____

LOAD & ASSEMBLE   _____

PACKAGING   _____

2) In the 480, how many failures do you expect?

_____(number)

3) What type failures do you expect to occur most often? How many of each do you expect in two (2) eight hour shifts? (number)

_____ (a) Rotary failure (degrading type)

_____ (b) Serial failure (catastrophic type)

6. If the failure is rotary (degrading), how long do you think it will take to repair one such failure, to include purge and start up time. Convert time estimates to minutes.

_____minutes

7. What would you consider to be an estimate of the efficiency of Mod A during the "2.5 million" run?

_____%

Rating should be given in percent and be in multiples of 5.
Note efficiency is defined as good pieces out divided by good pieces in, or ?/1200 ppm.
where (?) represents good pieces out per minute.

-722-

8. How many pieces of component scrap do you expect in an eight hour shift given the submodule is producing 1200 per minute, when it is producing.

CASE	_____(scrap rounds)
BULLET	_____
PRIMER INSERT	_____
LOAD & ASSEMBLE	_____
PACKAGING	_____

9. How many minutes of an eight hour shift do you expect the submodule to be down. There are 480 minutes in an eight hour shift.

CASE	_____minutes
BULLET	_____
PRIMER INSERT	_____
LOAD & ASSEMBLE	_____
PACKAGING	_____

10. If the failure is serial (catastrophic), how long do you think it will take to repair one such failure, to include purge and start up time. Convert time estimates to minutes.

_____minutes

GERT SIMULATION PROJECT1111 BY PETERSON RAY
DATE 8/ 2/ 1972

FINAL RESULTS FOR 10 SIMULATIONS

NODE	PROB./COUNT	MEAN	STD.DEV.	# OF OBS.	MIN.	MAX.	NODE TYPE
3	1.0000	547.2795	31.2391	10.	497.4689	600.2850	F
9	1.0000	1.1404	1.2205	4800.	1.0000	18.7862	B
8	1.0000	131.5411	110.9766	24.	7.4484	443.0795	B
7	.8000	177.4120	137.8526	18.	22.8833	501.7055	B
6	1.0000	100.3216	88.2372	33.	11.9868	468.7500	B
5	1.0000	195.9835	152.4059	19.	20.2237	494.1824	R
4	.8000	176.8553	125.7665	17.	24.3216	448.4439	R

QUEUE NODES

NODE	MEAN	STD.DEV.	# OF OBS.	MIN.	MAX.	
10	5.6874	1.8933	10.	2.5991	9.6589	AVERAGE NUMBER IN THE QUEUE
10	.9710	.0256		.9328	.9991	AVERAGE BUSY TIME OF PROCESSOR
10	.0181	.0114		0.0000	.0353	AVERAGE BALKERS PER UNIT TIME
11	4.4404	2.2882	10.	1.1559	8.2508	AVERAGE NUMBER IN THE QUEUE
11	.9596	.0254		.9181	.9949	AVERAGE BUSY TIME OF PROCESSOR
11	.0323	.0373		0.0000	.1082	AVERAGE BALKERS PER UNIT TIME
12	5.9172	2.6208	10.	.7849	8.6592	AVERAGE NUMBER IN THE QUEUE
12	.9790	.0251		.9320	.9988	AVERAGE BUSY TIME OF PROCESSOR
12	.0531	.0385		.0060	.1299	AVERAGE BALKERS PER UNIT TIME
13	3.0915	2.0793	10.	.6374	7.0845	AVERAGE NUMBER IN THE QUEUE
13	.9265	.0541		.8313	.9877	AVERAGE BUSY TIME OF PROCESSOR
13	.0120	.0176	10			AVERAGE BUSY TIME OF PROCESSOR

FOOTNOTES

[1] Ray Peterson, SP/4, Project Engineer of Systems, SCAMP office, Frankford Arsenal, Philadelphia, Pa.

[2] USAMUCOM, The New Generation Small Caliber Ammunition Production Equipment Concept VOL I, 15 April 1969.

[3] Franklin Institute Research Labs, Applying the Delphi Method to the Collection of Risk Analysis Data, Dec. 1971

[4] Dr. Erwin Kreyszig, Advanced Engineering Mathematics (John Wiley and Sons, Inc., New York, 1968), 2nd Ed., Chapt 18, pp 732.

[5] Dr. Alan B. Pritzger, The GERT Simulation Program

Also

Dr. Alan B. Pritzger and Gary E. Whitehouse, GERT, Part II, Probabilistic and Industrial Engineering Applications, The Journal of Industrial Engineering, Volume IXL, No. 1, 1966, pp. 61-66.

BIASED SAMPLES FOR TESTING

John Fargher, Jr.

U. S. Army Munitions Command
Frankford Arsenal
Philadelphia, Pennsylvania

INTRODUCTION. The basic tools of the reliability and quality control engineer are attribute and variables sampling inspection plans. These tools, as useful as they may be, have presented a major obstacle in the automation of many complex processes. These obstacles are:
1. Parts are pulled for samples and cannot be returned to the process flow.
2. The sample size required to maintain consumer and producer risk is too large to be handled economically on off-line gaging.

Because of these obstacles, costs for inspection is not decreasing commensurate to the drop in production costs realized through automation.

AUTOMATED INSPECTION. The area of automatic quality assurance encompasses and is interfaced with all phases of the production process. With the advent of more sophisticated and complex production systems, on line collection of data is becoming commonplace. Inspections are performed 100% on a go-no go basis (attributes). Because of the high production rates, variables data is unobtainable. Current methods of on-line automated inspection are limited to simple feedback arrangements in which limited switches are activated when a production item enters the inspection station. The limit switch arrangement is usually a one sided test, either a dimension is too large or within tolerance, or too small or within tolerance. Two sided tests are accomplished on two separate Automated Inspection Stations. The general scheme for inspection by attributes (accept or reject) is shown in Figure 1 for one inspection station on line.
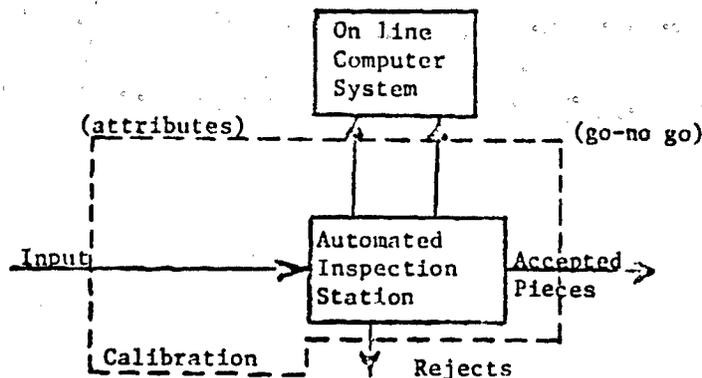

Figure 1. Diagram of Automated Inspection

Preceding page blank   -727-

As can be seen from Figure 1., there are really two basic elements.
1. An on-line computer system
2. An inspection station with calibration

The operating aspects of the calibration test set and calibration error are discussed in detail in Rusinoff's book[1] and the <u>Proceedings of the Instrument Society of America</u> (ISA): 1969[2]. There are two outputs from the automated inspection, accepted and rejected pieces. The accepted pieces continue through the production flow. Rejected pieces are normally not reworked but are discarded. Pieces from the process flow are pulled for samples somewhere after the inspection station for variables control charting and to check on the operation of the inspection station. Figure 2 illustrates the general scheme for 100% on-line inspection and feedback from off-line gaging.
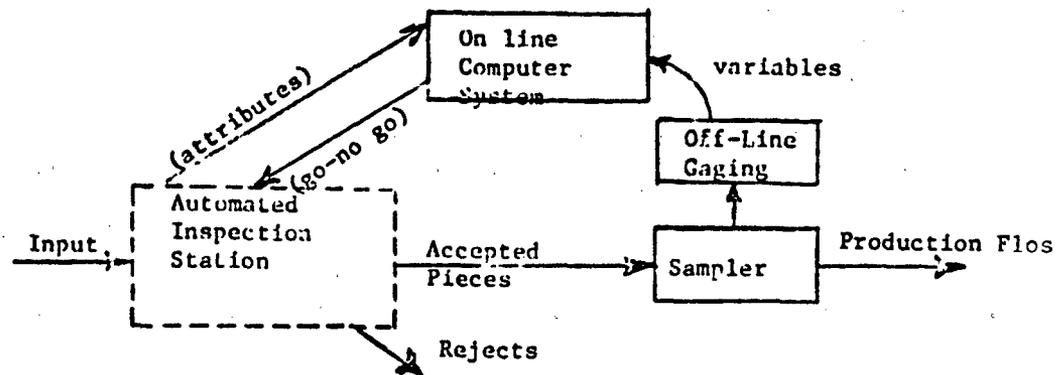
Figure 2. Diagram of Automated Inspection with off-line gaging of Production Flow.

The automated inspection station is considered to have several to many sensors for quality assurance measurement. Each inspection has a probability of accepting "out of tolerance" pieces, β, & rejecting "in tolerance" pieces, α. An operating characteristic curve for each inspection is illustrated in Figure 3.

Probability (accepting/out of tolerance) = β
Probability (rejecting/out of tolerance) = $1-\beta$

---

[1] Rusinoff, S. E., <u>Automation in Practice</u>, Chicago, Ill., American Technical Society, 1957.

[2] Williams, T.J., & Ryan, R.M., <u>Progress in Direct Digital Control</u>, Pittsburgh, Pa., Instrument Society of America, 1969.
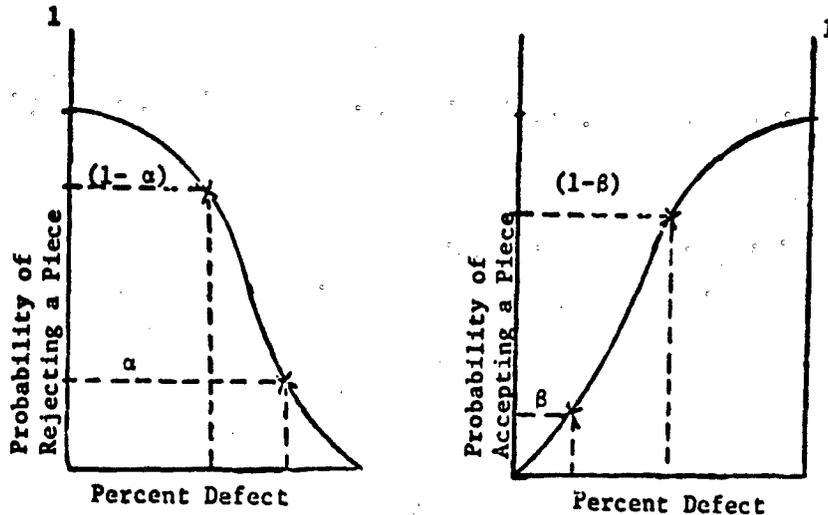
Figure 3. Operating Characteristic Curves for Individual Inspections

Probability (rejecting/in tolerance) = α
Probability (accepting/in tolerance) = $1-\alpha$

If several classes of defects are inspected for within an inspection station, each defect (or defect class) is judged independently of all other defects (or defect classes). This means that screening may be in effect for certain defects (classes) while at the same time sampling may be in effect for other defects (classes) at the same station.

Inspection may be set up so that the results are cumulative over all stations for a given defect class or so that results at each station are considered independently. It must be pointed out that the inspection results and judgments therefore must be individual defects for which the AQL or AOQL is specified. If inspection is stopped where the results are cumulative over the stations, all units which have not reached the last inspection station must be removed and/or screened and presented as resubmitted material.

From this, considering all inspections are independent, the probabilities for the automated inspection station become

$$\prod_{i-1}^{n} (1-\beta_i) = 1-\beta*$$

$$1 - \prod_{i=1}^{n} (1-\beta_i) = \beta*$$

where $\beta*$ is the probability of accepting an "out of tolerance" piece, and

$$\prod_{i=1}^{n} (1-\alpha_i) = 1-\alpha*$$

$$1 - \prod_{i=1}^{n} (1-\alpha_i) = \alpha*$$

-729-

where α^* is the probability of rejecting an "in tolerance" piece. The operating characteristic curves would be similar to Figure 3 but with less of a gradiant (slope).

Now, a new sampling technique is introduced, sampling from the production flow and using all of the rejects as samples. Figure 4 illustrates this technique.
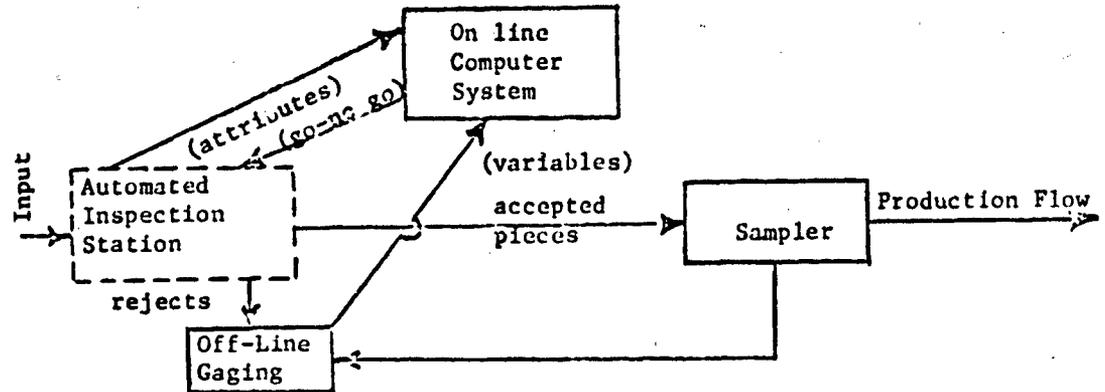

Figure 4. Diagram of Automated Inspection with off-line Gaging of accepted and rejected production flow.

The advantages offered by this sampling technique are:
1. Calibration of the Automated Inspection Station is checked without the use of masters.
2. Fewer "in tolerance" pieces needed for the sample.
Because of the smaller sample of "good" pieces, fewer pieces are lost from the production flow, fewer voids are introduced, and a greater production efficiency results. Also, a periodic check on calibration of the Automated Inspection Station is obtained without losing production time for the master gage. As this topic is further researched, more advantages should be found. This makes sense as "rejects" should tell more about the process than "good" pieces.

CONTROL OF PROCESS AND ON-LINE INSPECTION. An example of a typical characteristic (dimension) probability density function is given in the left hand side of Figure 5. The right hand side of Figure 5 illustrates a typical reject population. The normalized probability density function of the accepted population, Figure 5, has the area under the acceptance region of

$$acc_{acc} = \frac{1-x-p+\beta}{1-\alpha+2\beta-p}$$

-730-
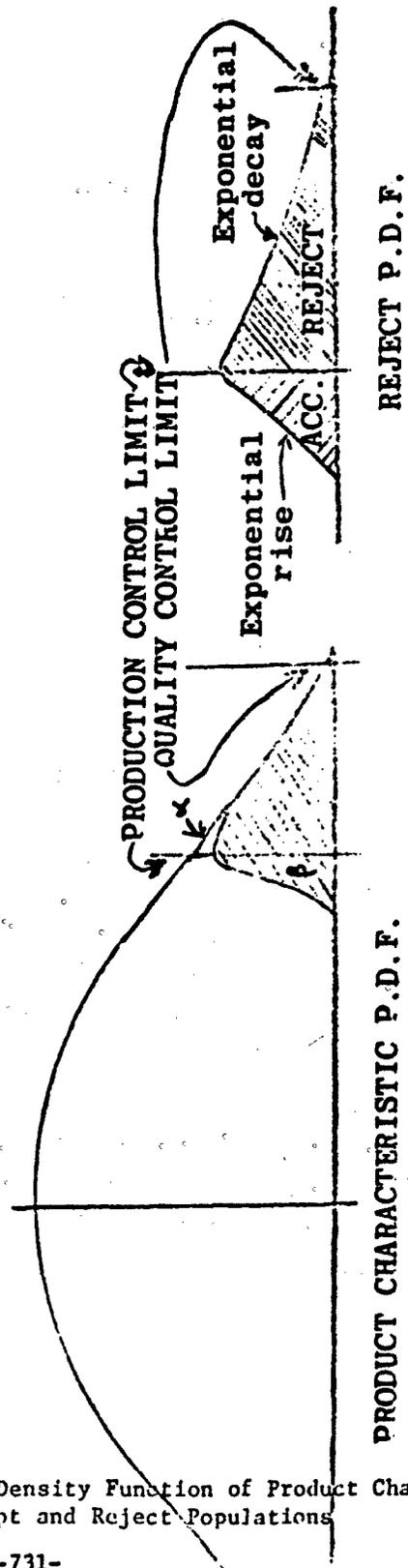
Figure 5. Probability Density Function of Product Characteristic, Showing Accept and Reject Populations

where p is the percent defective divided by 100, and the area under the rejection region of

$$rej_{acc} = \frac{\beta}{1-\alpha+2\beta-p}$$

The total area under the accepted population sums to 1.

The rejected normalized probability density function, Figure 5, has an area in the acceptance region of

$$acc_{rej} = \frac{\alpha}{\alpha+p-\beta}$$

where α is the percentage of good pieces rejected, β is the percentage of bad pieces that are accepted and p is the percent defective (area to the right of production limit).

The area in the reject region is

$$rej_{rej} = \frac{p-\beta}{\alpha+p-\beta}$$

The total area of the reject population also sums to 1.

Looking at the reject population, given $\alpha = 1/1,000 =$ rejected "good" pieces, $\beta = 1/10,000 =$ accepted "bad" pieces, p=1%.

Percentage of rejected pieces $= p+\alpha-\beta= 1\% + .1\%-0.01\%$

$$=1.09\%$$

The percentage of good pieces in the reject population is

$$rej_{acc} = \frac{.1}{1 + .1 - .01} = \frac{.1}{1.09} = 9.17\%$$

$$rej_{rej} = \frac{1 - .01}{1 + .1 - .01} = \frac{.99}{1.09} = \frac{90.83\%}{100.00\%}$$

The signal from the on-line inspection station to the reject station can be characterized by a binary code, Z,

$$Z = \begin{cases} 1, & y > X_p \\ 0 & \text{otherwise,} \end{cases}$$

where X_p is the production limit. The value for y is given by

$$y = X+\epsilon.$$

where X is the actual measurement and ϵ is the error term associated with gage noise. This error term is assumed to be an ensemble of error terms that is described by a Gaussian probability density function. A priori probability of the parameter y is given by

$$f(y) = \frac{1}{\sqrt{2\pi}\ \sigma_y} \exp\left(-\frac{(y-X)^2}{2\sigma_y^2}\right)$$

where X is the mean value of the ensemble and σ_y^2 is its variance.

The probability density function for the reject population is found, using Bayes formula,

$$f_{XZ}(X,1) = \frac{f_{XZ}(X,Z=1)}{f_Z(Z=1)}$$

$$= \frac{\displaystyle\int_{X_p}^{\infty} f_{y|X}(y|X)\, f_X(X)\, dy}{\displaystyle\int_{-\infty}^{\infty}\int_{X_p}^{\infty} f_{z|x}(y|x)\, f_X(x)\, dy\, dx}$$

$$= \frac{\displaystyle\int_{X_p}^{\infty} \frac{1}{\sqrt{2\pi}\ \sigma_y} \exp\left(-\frac{(y-X)^2}{2\sigma_y^2}\right) \frac{1}{\sqrt{2\pi}\ \sigma_X} \exp\left(-\frac{(X-\mu)^2}{2\sigma_X}\right) dy}{\displaystyle\int_{-\infty}^{\infty}\int_{X_p}^{\infty} \frac{1}{\sqrt{2\pi}\ \sigma_x} \exp\left(-\frac{(y-x)^2}{2\sigma_z^2}\right) \frac{1}{\sqrt{2\pi}\ \sigma_X} \exp\left(-\frac{(x-\mu)^2}{2\sigma_X^2}\right) dy\, dx}$$

$$= \frac{\dfrac{1}{\sqrt{2\pi}\,\sigma_X} \exp\left(-\dfrac{(X-\mu)^2}{2\sigma_X^2}\right)\left[1 - \phi\left(\dfrac{X_p-X}{\sigma_y}\right)\right]}{1 - \displaystyle\int_{-\infty}^{\infty} \phi\left(\dfrac{x-\mu}{\sigma_X}\right)\phi\left(\dfrac{X_p-x}{\sigma_z}\right)dx}$$

But the denominator of this final expression is one minus a function of the chosen values of σ_X^2, σ_z^2, μ, X_p but not on X. In the analysis of the probability density function, $1 - K(\sigma_X^2, \sigma_z^2, \mu, X_p)$ is only a scale factor and will be signified by K,

$$f(X|Z=1) = \begin{cases} \dfrac{1}{\sqrt{\pi}\,K\,\sigma_X}\exp\left(-\dfrac{(X-\mu)^2}{2\sigma_X^2}\right), & \text{for } X \to \infty, \\[2em] \dfrac{1}{2\sqrt{2\pi}\,K\,\sigma_X}\exp\left(-\dfrac{(X_p-\mu)^2}{2\sigma_X^2}\right), & \text{for } X = X_p, \\[2em] 0, & \text{for } X \to -\infty. \end{cases}$$

Therefore

$$f(X|Z=1) = \frac{1}{k}\exp\left(-\frac{(X-\mu)^2}{2\sigma_X^2}\right)\int_{(X_p-X)/\sigma_y}^{\infty}\exp\left(-\frac{t^2}{2}\right)dt.$$

The mode of the distribution can be found by setting the partial of $f(X|Z=1)$ with respect to X equal to zero and solving

$$\frac{\partial f(X|y \geq X_p)}{\partial X} = 0 = -\frac{(X-\mu)}{k\,\sigma_X^2}\exp\left(-\frac{(X-\mu)^2}{2\sigma_X^2}\right)\int_{(X_p-X)/\sigma_y}^{\infty}\exp\left(-\frac{t^2}{2}\right)dt$$

$$+\frac{1}{k\sigma_y}\exp\left(-\frac{(X-\mu)^2}{2\sigma_X^2} - \frac{(X_p-X)^2}{2\sigma_y^2}\right).$$

Now that we have the probability density function, the mean can be found from

$$E(X) = \int_{-\infty}^{\infty} X \, f(X) \, dX$$

$$= \int_{-\infty}^{\infty} \frac{X}{k} \exp\left(-\frac{(X-\mu)^2}{2\sigma_X^2}\right) \int_{(X_p-X)/\sigma_y}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt \, dx.$$

It can be shown that the function exists in closed form whenever σ_X, σ_y, μ, and X_p are chosen. Likewise, the tolerance can be found such that

$$\int_{-\infty}^{R_{LCL}} \frac{1}{k} \exp\left(-\frac{(X-\mu)^2}{2\sigma_X^2}\right) \int_{(X_p-X)/\sigma_y}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt \, dX = \frac{\alpha}{2},$$

$$\int_{-\infty}^{R_{UCL}} \frac{1}{k} \exp\left(-\frac{(X-\mu)^2}{2\sigma_X^2}\right) \int_{(X_p-X)/\sigma_y}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt \, dX = 1 - \frac{\alpha}{2}$$

and

$$P[R_{LCL} < X < R_{UCL}] = 1 - \alpha$$

This can be used for calculating the range on $\bar{X}$ (See Duncan, Table E2, Page 874) and the $\bar{R}$ value for the range chart, where $P(R_{LCL} < \bar{R} < R_{UCL}) = .50$.

To obtain a cell height for the χ^2 and Kolmogorov-Smirnov goodness of fit test, the function is integrated from X to X+ΔX for each cell interval, ΔX

$$\frac{1}{k} \int_{X_i}^{X_i + \Delta X} \exp\left(-\frac{(X-\mu)^2}{2\sigma_X^2}\right) \int_{(X_p-X)/\sigma_y}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt \, dX, \quad i=1,2,\ldots,n,$$

for n cells. ΔX should be chosen to assure that the interval is sufficiently large to allow at least 5 counts per cell whenever the statistical tests are run on actual data.

-735-

Once the "reject inspection plan" is in effect, ie. p, $\bar{X}$ and range charts are being kept, tests of trends on the recorded values should be implemented to test for trends in the measurement.

CONCLUSIONS. Application of this technique should be simulated before actually applying it to a process to check for sensitivity to the process and inspection parameters as opposed to the sampling plans now in effect.

METHODS TO EXTEND THE UTILITY OF LINEAR
DISCRIMINANT ANALYSIS*

CPT Lawrence E. Larsen
Walter Reed Army Institute of Research
Division of Neuropsychiatry
Dept. of Microwave Research
Walter Reed Army Medical Center
Washington, D. C.

ABSTRACT

When attempting to apply multiple group discriminant analysis in the setting of unequal population variance/covariance matrices, we propose a method which may improve classification success. The method is based on pre-clustering the groups according to similarity of their sample dispersion matrices. When such clustering is possible, a sequential discrimination of the groups in each cluster may allow substantially lower error rates than a straightforward discrimination of the groups in one shot, due to more appropriate pooling. This proposition is illustrated with data taken from an EEG (multiple group) pattern recognition problem where error rate estimates (unbiased) were impressively improved with the strategy of a layered decision process according to the above principles. Discussion emphasizes the role of components analysis of the separate, sample dispersions to direct the clustering between groups, and the effect of segmentation on the "common" variance/covariance matrix at each layer.

* * * ———— * * *

In previous studies (Larsen & Walter, 1969; Larsen & Walter, 1970) a method described as "the two stage machine" was employed to improve classification success in a testing data set (Larsen et al., 1971). It was based on the idea of a layered decision process aimed at first discriminating two aggregates of groups followed by a later discrimination for the groups within in each aggregate taken separately. The aggregates of groups were determined on the basis of cross-group clustering in a reduced discriminant space of two dimensions. This communication reports further studies on a possible theoretical basis for the efficacy of the two-stage machine as well as a means to determine aggregate structure and when the technique may be profitably employed.

---

*This paper was presented at the Seventeenth Conference on the Design of Experiments in Army Research, Development and Testing.

METHODS TO EXTEND THE UTILITY OF LINEAR
DISCRIMINANT ANALYSIS*

CPT Lawrence E. Larsen
Walter Reed Army Institute of Research
Division of Neuropsychiatry
Dept. of Microwave Research
Walter Reed Army Medical Center
Washington, D. C.

ABSTRACT

When attempting to apply multiple group discriminant analysis in the setting of unequal population variance/covariance matrices, we propose a method which may improve classification success. The method is based on pre-clustering the groups according to similarity of their sample dispersion matrices. When such clustering is possible, a sequential discrimination of the groups in each cluster may allow substantially lower error rates than a straightforward discrimination of the groups in one shot, due to more appropriate pooling. This proposition is illustrated with data taken from an EEG (multiple group) pattern recognition problem where error rate estimates (unbiased) were impressively improved with the strategy of a layered decision process according to the above principles. Discussion emphasizes the role of components analysis of the separate, sample dispersions to direct the clustering between groups, and the effect of segmentation on the "common" variance/covariance matrix at each layer.

* * * ——— * * *

In previous studies (Larsen & Walter, 1969; Larsen & Walter, 1970) a method described as "the two stage machine" was employed to improve classification success in a testing data set (Larsen et al., 1971). It was based on the idea of a layered decision process aimed at first discriminating two aggregates of groups followed by a later discrimination for the groups within each aggregate taken separately. The aggregates of groups were determined on the basis of cross-group clustering in a reduced discriminant space of two dimensions. This communication reports further studies on a possible theoretical basis for the efficacy of the two-stage machine as well as a means to determine aggregate structure and when the technique may be profitably employed.

---

*This paper was presented at the Seventeenth Conference on the Design of Experiments in Army Research, Development and Testing.

It should be noted that when all the assumptions underlying linear discriminant analysis are met, there cannot be any advantage to segmenting a multiple group discrimination problem. The relevant assumption underlying classical linear discriminant analysis is that the m groups being discriminated are samples from m Gaussian propulations with equal variance/covariance (dispersion) matrices. Any differences are limited to unequal mean vectors. (The mean vector is an ordered one dimensional array composed of within-group means for each variate on which the measurement space constructed. The dispersion matrix is a symmetrical two dimensional array of numbers containing the variances for each variate in the main diagonal and all possible pair-wise covariances in the off diagonal elements.) The dispersion matrix and mean vector occupy importance because these two parameters completely define a multivariate normal distribution. This statement is analogous to the fact that the distribution of a single Gaussian variate is determined by the population mean and variance. The assumption of equal within-groups covariance matrices allows for optimal discrimination (i.e., minimal losses when costs of misclassification are equal) with linear boundaries. When this assumption is violated, the optimal boundary is no longer linear; rather, the optimal boundary (in general) is quadratic. The linear boundary obtained by BMD 07M (Dixon, 1968) in such a setting is, therefore, not the optimal boundary nor is it, in general, even the best possible linear boundary. This results from the fact that the program pools across the individual groups to estimate the so-called common covariance matrix (ccm). To the extent that the assumption of equal dispersion matrices is true in the population sense, such a pooling (or "averaging" of within-group dispersions) makes for a better estimate of the population dispersion matrix than any one of the individual within-groups dispersion matrices taken alone. When this assumption is violated, the resulting ccm does a variable amount of violence to the separate population values. The extent of misrepresentation depends upon how much the separate population dispersions differ.

One way to find the best linear boundary, or at least a better boundary than the one determined by simple pooling, is to weigh the estimated within-groups dispersion matrices prior to pooling (Anderson & Bahadur, 1962). While this approach has generality, the determination of weights necessary to find the best linear boundary (in a multiple group discrimination with unequal within-group dispersions) is not always a simple exercise.

In certain settings, another approach proposed by Rao (1966) is possible. The required conditions are that the differences in dispersion are arranged in the test space such that dispersions are homogeneous within either of two subsets, but heterogeneous between subsets. In this case, the optimal boundary between subsets is quadratic, but the optimal boundary within subsets is linear; thus, two "common" covariance matrices result, each one common only to the groups within the subset. This framework is the one which we believe underlies the efficacy of the two stage machine.[1]

---

[1] Note that while a quadratic boundary is, in general, optimal for the separation of such subsets, it may not perform appreciably better than a linear one if the subsets do not extensively overlap in measurement space.

Rao's formulation is a tractable specialization of the more likely situation that the subsets have greater similarity of dispersion within subsets than across subsets. This presents additional measurement problems since testing equality of dispersion is simpler than indexing similarity due to the variety of ways different within-group dispersions may still be (colloquially) similar. At this point, it may be useful to note that one cannot reasonably use the failure of tests for homogeneity of dispersion to judge similarity of non-equality dispersions. Such a procedure would only indicate the violence with which we reject the hypothesis of equality. It is not reasonably extendable to the assertion that strength of rejection may be used to group similar dispersions by strength of rejection of equality to one or a set of "standard" dispersion matrices. Such a statement is analogous to the fact that the significance of Wilks' lambda cannot be used to quantitatively judge group separation. Rather, a statistic such as Mahalonbois' distance is required.

When covariance matrices are examined for equality, all parameters of description (the elements in the main diagonal and upper, or lower, triangle) resolve to a single, dichotomous, dimension—equality or nonequality. When dispersions differ, it seems reasonable to seek some summary indices rather than an exhaustive examination of elements in the main diagonal and upper triangle. Therefore, we propose to parameterize the dispersions in a way which capitalizes on geometric concepts (as an interpretive aid) and not to consolidate these measures, but rather to consider them separately with some empirical guides for determining their relative importance. Other than size (the volume contained in a given hyper-ellipse of equal likelihood), which may be measured by generalized variance (determinant of the dispersion matrix) for dispersions which are equal otherwise, we propose two indexes of dispersion: shape and orientation. We further propose that principal components analysis of the estimated within-group dispersion matrices may provide such a summarization.

Thus, we propose to cluster groups into aggregates on the basis of greater similarity of dispersion within clusters than between clusters. The technique may, therefore, be described as cluster seeking subset generation. The objective is to provide more appropriate pooling at the level of discrimination within the aggregates and to employ either linear or quadratic boundaries for discrimination between aggregates depending on earlier considerations.

In the setting of multivariate normal dispersions (with different centroids) which are equal except for rotations of their hyper-ellipses in measurement space, the most extreme dissimilarity would occur when the two major axes are orthogonal. The effect of pooling in this extreme situation would be to circularize the ccm. To the extent that the separate dispersions approached isotrophy, i.e. as the separate dispersions approximate circularity as evidenced by equal or near equal eigenvalues, the pooling would be of little harm. However, when the separate dispersions are strongly anisotropic, the potential error from pooling and consequent ccm circularization becomes more serious. In this way, anisotropy (i.e., shape) becomes an antecedent for characterizing similarity. That is, to the extent that the major axis of the hyper-ellipse is long relative to the minor axes, then the dispersions may be described, except for size, by the orientation (i.e., direction cosines) of the major axis. Consequently, similarity of dispersion may become similarity of orientation of major axes when the dispersions are strongly anisotropic.

If dispersions have been successfully clustered in this fashion, the effects of unequal size are easily dealt with by translation of the decision boundary for discrimination either within or between clusters. The optimal decision rule for discrimination when dispersions differ only in size is obtainable from the likelihood ratio.

Principal components analysis serves two roles: a determination of anisotropy by the inequality of the eigenvalues of the within group dispersion matrices; and description of the orientation of the major axis of the hyper-ellipse by its direction numbers, which are the elements of the eigenvector corresponding to the first eigenvalue.

Principal components analysis consists of a rotational transformation of co-ordinate axes with the object in mind that variance is maximized along the new axes. The set of new axes constitute a set of normalized linear combinations (sum of squares of coefficients equal to unity) of the old co-ordinates. That is, in the population sense we have a p-dimensional random vector[1] in the original co-ordinate system

$$\underline{X} = \{x_1, x_2 \ldots x_p\}'$$

where $x_1, x_2 \ldots x_p$ are random variates.

The random p-vector has a mean of zero,

$$E(\underline{X}) = 0$$

and a dispersion matrix

$$E(\underline{X}\underline{X}') = [\Sigma].$$

The linear combination is implemented by a p-dimensional column vector of coefficients

$$\underline{B} = \{b_1, b_2, \ldots b_p\}'$$

The linear combination is made by the following

$$U = b_1 x_1 + b_2 x_2 + \ldots + b_p x_p$$

$$U = \underline{B}'\underline{X}$$

The transformation $\underline{B}$ carries the random vector $\underline{X}$ into a random scalar variate U. The variance of U is

$$E(U)^2 = E(\underline{B}'\underline{X})^2$$

$$= E(\underline{B}'\underline{X}\underline{X}'\underline{B})$$

$$= \underline{B}'[\Sigma]\underline{B}$$

---

[1] A random vector is the multivariate analog of the random variate in univariate statistics.

Thus, the variance of U depends on the coefficients of the combination and the dispersion matrix in the original coordinate system.

It is these variances, one for each set of coefficients, which are maximized subject to the condition that

$$\sum_{k=1}^{p} b_k^2 = 1.0$$

and that the sets are mutually orthogonal. That is, the inner products are zero.

$$\underline{B}^{(j)} \cdot \underline{B}^{(i)} = 0 \text{ for } i,j = 1,2,\ldots p$$

$$\text{and } i \neq j$$

or equivalently,

$$\sum_{k=1}^{p} b_k^{(i)} b_k^{(j)} = 0 \text{ where } k \text{ indexes variates and } i, j \text{ index sets of coefficients}$$

The column vectors of coefficients which maximize the above variances are found as the vectors which are invariant under the linear transformation $[\Sigma]$. That is, invariant vectors are those which are changed only by a scalar multiple with the given linear transformation:

$$[\Sigma] \ \underline{B} = \lambda \underline{B}$$

where λ is a scalar constant. The invariant vector(s) are found by solving the following set of homogeneous equations

$$[\Sigma]\underline{B} = \lambda\underline{B}$$

$$[\Sigma]\underline{B} - \lambda\underline{B} = 0$$

$$([\Sigma] - \lambda[I])\underline{B} = 0$$

A non-trivial solution exists when the coefficient matrix is singular. That is, when

$$\left| \Sigma - \lambda I \right| = 0$$

The values of λ which make the determinant zero are found by expanding the above determinant to produce a polynominal in $\lambda, \phi(\lambda)$, of degree p. The roots of $\phi(\lambda)$ are substituted separately into the coefficient matrix in order to find the corresponding $\underline{B}$. Thus, there are p sets of $\underline{B}^{(i)}$ $i = 1,2,\ldots p$ corresponding to the p roots of $\phi(\lambda)$. Furthermore, under the above conditions, the variance of composite, scalar variates $U^{(i)} = \underline{B}^{(i)}\underline{X}$ can be shown to be equal to the value of the corresponding root. That is

$$E(U^{(i)})^2 = \lambda_i$$

If the roots of $\phi(\lambda)$ are ranked in decreasing magnitude, $\lambda_1 > \lambda_2 > ... > \lambda_p$. The first principal component, $\underline{B}^{(1)}$, makes the one linear combination with maximum possible variance. The variate values on this coordinate are covered by the random variate $U^{(1)}$ with variance of λ_1. Further, if $\underline{X}$ is from a multivariate normal population, the first principal component corresponds to the major axis at the hyperelipse of equal likelihood for the multivariate "scatter diagram".

Severa: additional features are of interest: 1) the roots of the characteristic polynominal $\phi(\lambda)$, are real and non-negative; 2) the p sets of coefficients $B^{(1)}$ i = 1,2,...p produce p random variates $U^{(1)}$, i = 1,2,...p which may constitute a random vector

$$\underline{U} = \{U^{(1)}, U^{(2)}, ...U^{(p)}\}$$

Similarly, the ordered set of column vectors $B^{(1)}$ may constitute a transformation matrix.

$$\underline{U} = [B]\underline{X}$$

where

$$[B] = \left[\underline{B}^{(1)} \; \underline{B}^{(2)} ... \underline{B}^{(p)} \right]$$

The transformation matrix is non-singular and orthogonal

$$|B| \neq 0, \quad [B][B'] = [I] = [B'][B]$$

It caries the random vector $\underline{X}$ into the random vector $\underline{U}$. This matrix of normalized eigenvectors is known as a factor pattern matrix when the eigenvectors are multiplied by the square root of their corresponding eigenvalues.

$$[A] = [B][\Lambda^{1/2}]$$

where

$$[\Lambda^{1/2}] = \begin{vmatrix} \sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ & & & \sqrt{\lambda_p} \end{vmatrix}$$

Later presentation of the components analysis will be in terms of the factor pattern matrix since the squared length of the vector will then equal the corresponding eigenvalue and thus be an analog of variance on the component.

The random vector $\underline{U}$ in the new p-space has a dispersion matrix given by

$$E(\underline{U}\underline{U}') = [\Lambda]$$

-742-

where is

$$[\Lambda] = \begin{vmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{vmatrix}$$

That is, the dispersion matrix in the transformed space is diagonal with the eigenvalues of the characteristic equation down the main diagonal. The generalized variance is invariant. That is, the determinant of the two dispersion matrices are equal:

$$|\Lambda| = |\Sigma|$$

Lastly, the number of non-zero roots to $\phi(\lambda)$ depends on the rank of $[\Sigma]$. If $[\Sigma]$ is of full rank (rank equal to order) there are p non-zero roots to $\phi(\lambda)$. If the rank, r is less than the order, p, then there are r<p non-zero roots to $\phi(\lambda)$; thus, there will be r co-ordinate axes to the new space. This is one possible route to dismensionality reduction. Another route is to select the m largest roots of $\phi(\lambda)$ according to some criterion which most often operates on the notion that the p-m smallest roots are either indistinct (i.e., equal or nearly equal) or constitute an empirically unimportant proportion of the trace.

Other approaches to development of the principal components solution include the following: 1) Pearson approached the problem from the point of view of fitting an orthogonal subspace of q-dimension to a swarm of n points in p-space where q<p. The criterion used was the subspace should minimize the sums of squares perpendicular distances from the points to the subspace. The solution is the first q eigenvectors of $[\Sigma]$. 2) Hotelling approached it from the random vector treatment where the object was to find a dimensionality reducing transformation matrix such that the transformed random q-vector is the best linear predictor of the original random p-vector where q<p. Again, the solution is the first q eigenvectors of $[\Sigma]$. (See Rao, 1963 for further development).

At this point, let us consider some 2 and 3 dimensional examples[1] with standardized variates:
Consider the following variance/covariance matrix $[S_1]$

$$[S_1] = [R_1] = \begin{bmatrix} 1.0 & 0.9 \\ 0.9 & 1.0 \end{bmatrix}$$

---

[1] In these examples, the variates are standardized (to have mean zero and variance unity) to simplify the figures and exposition. This has the consequence of making the covariance and correlation matrices equal. Furthermore, in the two dimensional examples, the principal component is constrained to make a 45° angle with abscissa, either in quadrant I or II. The fact that unities are in the main diagonal means that the total variance to be modelled is p where p is the order of the matrix. In general, the components analysis solution is not invariant with normalization of variates. Also, distributional properties of eigenvalues and eigenvectors of correlation matrices are not as well treated as for the dispersion matrix (see Anderson, 1965).

the eigenvalues are

$$\lambda_1 = 1.9$$

$$\lambda_2 = 0.1$$

Note that the sum of the eigenvalues equals the sum of the elements of the main diagonal, i.e., the trace of the matrix

$$\sum_{i=1}^{p} \lambda_i = \sum_{j=1}^{p} s_{jj} \quad \text{where p is the order of [S].}$$

The eigenvectors are

$$\underline{v}_1 = [+0.70711, +0.70711]'$$

$$\underline{v}_2 = [-0.70711, +0.70711]' \qquad [V] = [\underline{v}_1 \ \underline{v}_2]'$$

Note that these are normalized

$$\sum_{i=1}^{p} v_i^2 = 1.0$$

The factor pattern is

$$[A] = [V] [\Lambda^{1/2}]$$

$$[A] = \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix} \begin{bmatrix} 1.3748 & 0 \\ 0 & 0.3162 \end{bmatrix}$$

$$[A] = \begin{bmatrix} 0.9745 & -0.2234 \\ 0.9745 & 0.2234 \end{bmatrix}$$

If we take the columns of [A] as locating two vectors in a two dimensional test space, we have the following plot: (see Figure 1)
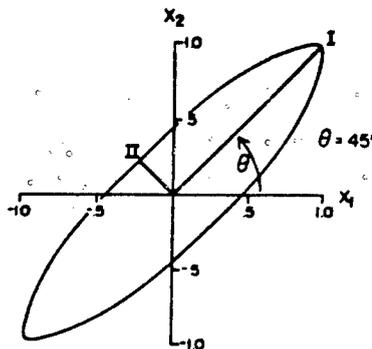

FIGURE 1
Major and minor axes of a bivariate gaussian swarm with dispersion matrix $[S_1]$.

Note that the sign of s_{12} determines the quadrant in which the principal component will be. When s_{12} is positive, the first component will be in quadrant I. When it is negative, the first component will be in quadrant II. Furthermore, the squared lengths of vectors are equal to the eigen-values. These, in turn, are proportional to the fraction of total variance explained on each axis. Given a bivariate normal scatter diagram, the ellipse encloses 1σ of variance on each axis as generated by the corresponding (new) composite variate. If the samples were from a bivariate gaussian population, the first and second components would be the major and minor axes of any two-dimensional ellipse of equal likelihood. The high positive correlation (and equal variances) puts the major axis in quadrant I, and implies a long major axis relative to the minor axis.

We may check the effect of sign on location of the first principal component by analyzing the following matrix:

$$[S_2] = \begin{vmatrix} 1.0 & -0.9 \\ -0.9 & 1.0 \end{vmatrix} = [R_2]$$

with $[S_1]$ for comparison

$$[S_1] = \begin{vmatrix} 1.0 & 0.9 \\ 0.9 & 1.0 \end{vmatrix} = [R_1]$$

The corresponding factor patterns are

$$A_1 = \begin{vmatrix} 0.975 & -0.223 \\ 0.975 & 0.223 \end{vmatrix}$$

$$A_2 = \begin{vmatrix} -0.975 & 0.223 \\ 0.975 & 0.223 \end{vmatrix}$$

The change in sign of covariance (correlation) moves the first principal component from quadrant I to quadrant II.

X_2

I

1.0

$.5$

II

θ

-1.0 $-.5$ $.5$ 1.0 X_1

$-.5$

-1.0
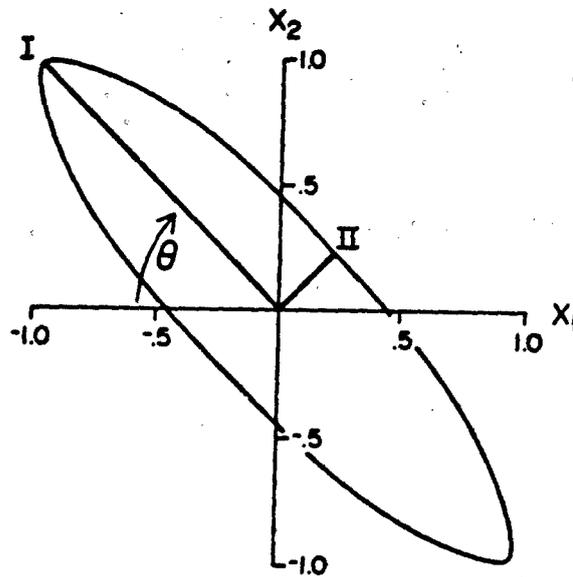
FIGURE 2
Major and minor axes of a bivariate gaussian swarm with dispersion matrix $[S_2]$.

The effect of change from a high positive covariance (correlation) to an equally high negative correlation was to rotate the principal components coordinate system 90°. The "shape" (ratio of lengths of major and minor axes) of the bivariate swarm is unchanged. Thus, the sign and values in the first column of [A] determine the direction of the first principal component in the test space.

Components analysis also provides information concerning shape of the test space hyper-ellipse by way of the relative sizes of the eigenvalues. We noted before that a high correlation, regardless of sign, implies that the test space has a long major axis relative to its minor axis. However, once out of the bivariate case, it is more informative to consider the proportion of total variance generated or explained by each component as an index of relative dispersion along that axis. Thus, we can determine something about shape of the multivariate swarm in measurement space. Consider a third two-dimensional example:

$$[S_3] = \begin{vmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{vmatrix} = [R_3]$$

The direction of its principal component is identical to the one for $[S_1]$. However, the shape of the ellipse is much rounder than before.
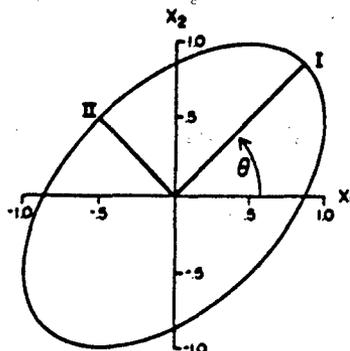

FIGURE 3

Major and minor axes of a bivariate gaussian swarm with dispersion matrix $[S_3]$.

This is in contrast to S_2 where shape was identical to S_1. We can summarize this information in a table or plot of eigenvalues as proportionate contribution to total dispersion as in fig. 4 and table 1.
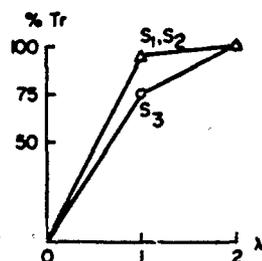

FIGURE 4

A plot of the eigenvalues of $[S_1]$, $[S_2]$, and $[S_3]$ as percentage of the traces.

TABLE 1

Eigenvalues for 3 Two-by-Two Dispersion Matrices

$$S_1: \begin{array}{l} \lambda_1 = 1.9 \ (95\%) \\ \lambda_2 = 0.1 \ (5\%) \end{array} \qquad S_2: \begin{array}{l} \lambda_1 = 1.9 \ (95\%) \\ \lambda_2 = 0.1 \ (5\%) \end{array}$$

$$S_3: \begin{array}{l} \lambda_1 = 1.5 \ (75\%) \\ \lambda_2 = 0.5 \ (25\%) \end{array}$$

Let us consider two additional examples in three dimensions. Given the variance/covariance matrix $[S_4]$ written in upper triangular form,

$$[S_4] = \begin{vmatrix} 1.0 & 0.7 & 0.5 \\ & 1.0 & 0.9 \\ & & 1.0 \end{vmatrix} = [R_4]$$

we have the following eigenvalues

$$\lambda_1 = 2.41 \ (80\%), \ \lambda_2 = 0.524 \ (18\%), \ \lambda_3 = 0.06 \ (2\%)$$

The factor pattern in

$$[A_4] = \begin{vmatrix} 0.80 & -0.59 & 0.06 \\ 0.97 & 0.12 & -0.19 \\ 0.90 & 0.40 & 0.15 \end{vmatrix}$$

Direction of principal component and shape of distribution are seen in the test space as in Figure 5.
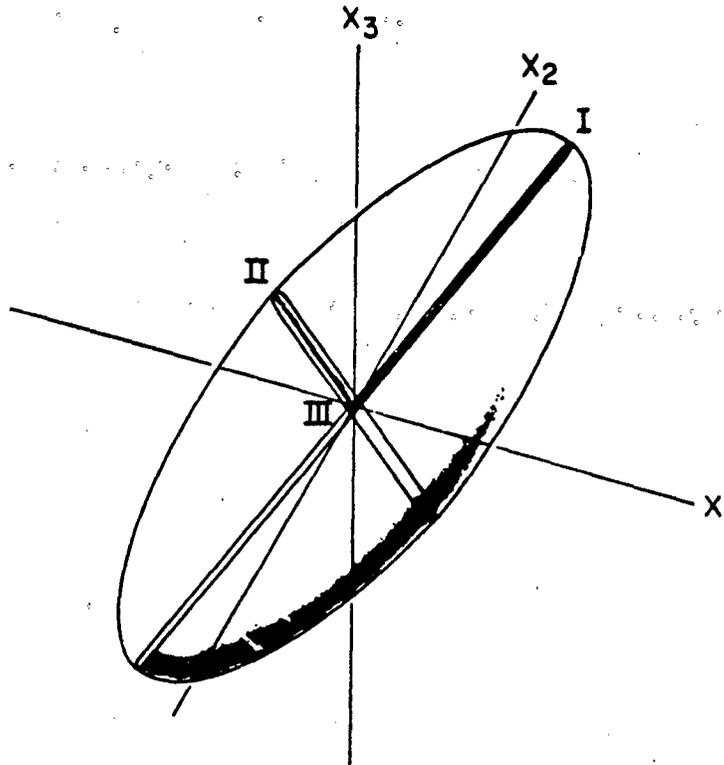
FIGURE 5
Axes of a trivariate normal swarm with dispersion matrix $[S_4]$.

The relative shape is further illustrated in Figure 6 by plotting the cumulative proportion of total dispersion against the index i.
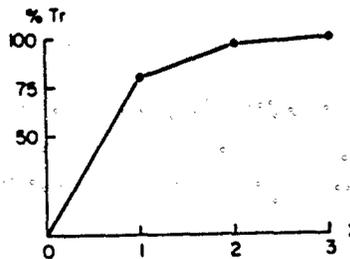

FIGURE 6
A plot of percentual eigenvalues for $[S_4]$.

We form a second variance/covariance matrix by interchanging the third element of the second row and the second element of the first row. Thus,

$$[S_5] = \begin{vmatrix} 1.0 & 0.9 & 0.5 \\ & 1.0 & 0.7 \\ & & 1.0 \end{vmatrix} = [R_5]$$

The eigenvalues are unchanged from before, but the orientation of the factor ellipse is slightly changed by rotation toward the X_1 axis (See Figure. 7). In the factor pattern, the first and third rows have been interchanged.
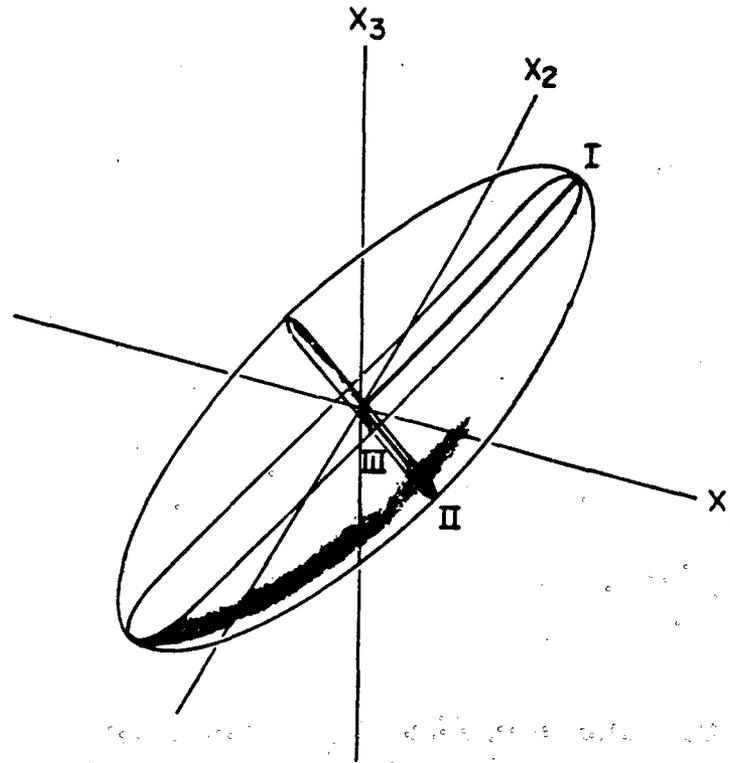

FIGURE 7

Axes of a trivariate normal swarm with dispersion Matrix $[S_5]$ which is related to $[S_4]$ by interchanging elements s_{23} and s_{12}.

$$[A_5] = \begin{vmatrix} 0.90 & 0.40 & 0.15 \\ 0.97 & 0.12 & -0.19 \\ 0.80 & -0.59 & 0.06 \end{vmatrix}$$

Thus, $[S_4]$ and $[S_5]$ do not differ at all in shape and differ only slightly in orientation. As a further index of similarity of orientation, we propose that the "coincidence" between the two principal components (as measured by the cosine of the angle between them) may be a useful guide when similarities of orientation are not obvious by inspection of the factor pattern.

Given the coordinates of each first principal component $[l_1, l_2, \ldots l_p]$ or specifically

$$S_4: \quad [0.80, 0.97, 0.90]'$$

$$S_5: \quad [0.90, 0.97, 0.80]'$$

we convert these direction numbers, l_i to direction cosines α_i and β_i by dividing each with the constant C defined below:

$$C = \sqrt{\sum_{i=1}^{p} l_i^2}$$

thus,

$$C = \sqrt{(0.80)^2 + (0.97)^2 + (0.90)^2}$$

$$C = \sqrt{.64 + .94 + .81}$$

$$C = \sqrt{2.39} = 1.546$$

The direction cosines are

$$S_4: \quad [\alpha_1, \alpha_2, \alpha_3]$$

$$S_4: \left[\frac{0.80}{1.53}, \frac{0.97}{1.53}, \frac{0.90}{1.53} \right] = [0.5174, 0.6274, 0.5821]'$$

$$S_5: \quad [\beta_1, \beta_2, \beta_3]$$

$$S_5: \left[\frac{0.90}{1.53}, \frac{0.97}{1.53}, \frac{0.80}{1.53} \right] = [0.5821, 0.6274, 0.5174]'.$$

The cosine of the angle between these two vectors is given by the sum of the products of the direction cosines

$$\cos \theta = \sum_{i=1}^{p} \alpha_i \beta_i$$

$$= (0.5174)(0.5821) + (0.6274) + (0.5821)(0.5174)$$

$$= 0.2677 + 0.3936 + 0.3011$$

$$\cos \theta = 0.9624$$

which confirms the impression that interchanging the two elements leaves the
test space orientation of the first principal component essentially unchanged.
    We see that principal components analysis provides information con-
cerning direction of the major axis of the hyper-elipse and shape of the
multivariate swarm. We propose that this information, in conjunction with the
presence of clustering among centroids in the test space, may provide a rea-
sonable basis for deciding the use of cluster seeking subset generation.
    Moving now to the 5 variate sleep-stage data (Larsen & Walter, 1970) and
a consideration of shape, we find the following plot of eigenvalues as cumu-
lative proportion of total dispersion.
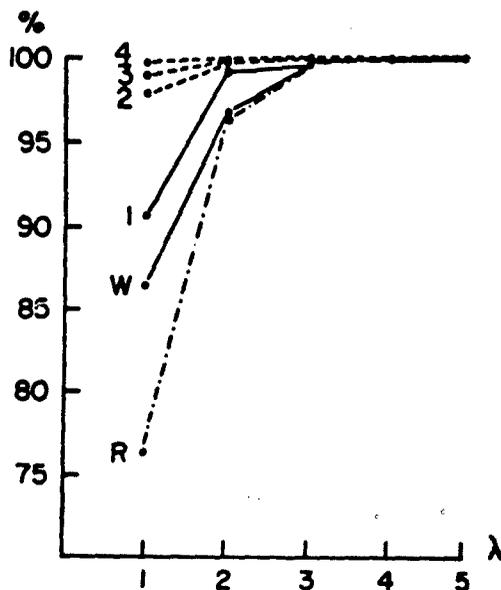
FIGURE 8
Plot of eigenvalues as percent of their trace for 6 sleep stages.

    Clearly, groups 2, 3, and 4 are strongly anisotropic with the first
principal component containing over 95% of the trace for any of the disper-
sions. Groups W, 1 and REM while less anisotropic than the others are still
dominated by the first principal component to extents ranging from 70% to
80% of the trace. In this setting, inappropriate pooling could introduce
substantial error if the dispersions are substantially different in orienta-
tion. Thus, the stage is set for characterizing similarity of dispersion by
similarity of orientation of the separate first principal components. The
object in mind being to aggretate groups in such a way that the dispersions
are similar within a subject and dissimilar between subsets.

The elements of the first principal component for each group are given below after applying the constant multiplier, $k = 100|\ell_i|/\Sigma_i|\ell_i|$, to the loadings. (This has the effect of removing absolute size.)[1,2]

TABLE 2

First Principal Component
Direction Numbers for 5 Sleep Stages on 5 Variates

	W	1	R	2	3	4
1	6.25	40.65	41.30	74.50	95.30	95.95
2	26.25	46.60	34.00	8.85	3.11	1.985
3	61.70	6.55	18.65	4.27	0.99	0.10
4	0.29	4.50	0.389	0.70	-0.24	0.06
5	0.25	0.16	0.231	0.66	0.29	0.04

Since these are also direction numbers, we may proceed to calculate the cosines between selected pairs of principal components. Simple inspection reveals that Stages 2, 3 and 4 are similarly oriented in terms of their first principal component. Among members of the DESYNC set, Stages 1 and REM are more similarly oriented than either is to Stage Wake. We will calculate the "coefficient of co-linearity" for selected pairs of principal components. The direction cosines are:

TABLE 3

First Principal Component
Direction Cosines for 5 Sleep Stages

	W	1	R	2	3	4
Δ 1	.093	.652	.728	.992	.999	.999
θ 2	.390	.747	.601	.118	.033	.021
σ 3	.916	.105	.329	.057	.010	.001
β_1 4	.004	.072	.007	.009	-.002	.001
β_2 5	.004	.003	.004	.009	.002	.001

---

[1]Note that we are not using the raw first eigenvector. Rather we are using the first component which is a normalized and weighted version of the first eigenvector.

[2]Removing absolute values obscures distributional properties of the loadings. However, we do not make use of these properties in present or later calculations. We do lose sight of the fact that eigenvectors associated with large eigenvalues will have more sampling variation in their loadings than vectors associated with smaller eigenvalues (Lawley, 1963).

The cosine of the angle between two p dimensional vectors is given by:

$$\cos \theta = \sum_{i-1}^{p} \alpha_i \beta_i$$

which leads to the following table of selected comparisons

TABLE 4

Cosines Between Selected Pairs of First Principal Components

$$\cos \theta_{W/1} = 0.448 \qquad \cos \theta_{W/2} = 0.191$$

$$\cos \theta_{W/R} = 0.603 \qquad \cos \theta_{1/2} = 0.741$$

$$\cos \theta_{1/R} = 0.959 \qquad \cos \theta_{2/4} = 0.994$$

Note that the orientation of the first principal component for stage Wake is rather unlike all other stages, but is most similar to stage REM.

On the basis of similarity of dispersion, it would appear that the groups 2, 3, and 4 should be aggregated into one cluster, named SYNC for the synchronous (slow, high voltage) appearance of the EEG; whereas groups W, 1 and REM should be aggregated in another cluster, named DESYNC for the desynchronous (fast, low voltage) appearance of the EEG. The groups in the SYNC subset have dispersions with principal components which are quite similar within (the lowest being 2/4 = 0.994) and dissimilar between subsets (the highest being 1/2 = 0.741). The picture in the DESYNC cluster is not as clear; nevertheless, for each group the highest "coefficient" is still within the DESYNC set. The highest within subset similarity being 1/R = 0.959, the lowest similarities are for those comparisons involving stage W. Yet the highest coefficient is for W/1 = 0.448, whereas the best comparison across subsets is still worse at 0.191 for W/2. Certainly, the average relationship is higher within a set rather than across sets.

It is interesting to compare the principal components for each group against the first component of the various pooled covariance matrices. We begin with calculation of direction cosines for the following poolings: 1) combining all 6 groups into one common covariance matrix; 2) combining W, 1 and R into a DESYNC "common" covariance matrix; and 3) combining 2, 3, and 4 into a SYNC "common" covariance matrix. The direction cosines of the principal components for these 3 ccm are as follows:

TABLE 5

Direction Cosines of the First Principal Component for Various Poolings of Covariance Matrices

	D+S	D	S
1	.999	.706	.999
2	.024	.684	.036
3	.006	.165	.001
4	-.004	.069	.008
5	.002	.028	.002

Selected "correlations" are

TABLE 6

Cosines Between Selected First Principal Components of
Groups and Pooled Covariance Matrices

$$2/(D + S) = 0.994 \qquad 2/S = 0.996$$
$$W/D = 0.484 \qquad 1/D = 0.994 \qquad R/D = 0.980$$
$$W/(D + S) = 0.099 \qquad 1/(D + S) = 0.665$$

It is plainly apparent that the groups which benefit most from seg-
mentation by cluster seeking subset generation are W, 1 and R. Their repre-
sentation by the DESYNC ccm is much more realistic than representation by
the [D+S] ccm. It is also these groups in which error rates improve the
most after segmentation.

When we consider both orientation and shape, it becomes clear that
clusters containing 2, 3, and 4 in SYNC; and W, 1, and R in DESYNC do pro-
vide a more reasonable basis for pooling than lumping both sets together.
The dispersions of SYNC and DESYNC subset members are more similar within
clusters than across clusters. We would expect pooling in the DESYNC
cluster to do somewhat more damage than pooling in the SYNC cluster because
R is less anisotropic and W is different in orientation.

ACKNOWLEDGEMENTS

REFERENCES

Anderson, T. W. (1963) Asymptotic theory for principal components analysis
Ann. Math. Stat. 34:122-148

Anderson, T. W. and Bahadur, R. R. (1962) Classification into two multi-
variate normal distributions with different covariance matrices. Ann.
Math. Stat., 33: 420-431

Dixon, W. J. (Ed.) (1968) BMD. Biomedical computer programs. Hlth, Sci.
Computing Facility, UCLA School of Medicine, Los Angeles.

Larsen, L. E., and Walter, D. O. (1969) On automatic methods of sleep staging by spectra of electroencephalograms. *Agressologie*, 10 (numero special): 1-14.

Larsen, L. E. and Walter, D. O. (1970) On automatic methods of sleep staging by EEG spectra. *Electroenceph. Clin. Neurophysiol.*, 28:459-467.

Larsen, L. E., Walter, D. O., McNew, J. J., Adey, W. R. On the problem of bias in error rate estimation for discriminant analysis. *Pattern Recognition*, 3:217-223.

Lawley, D. N. and Maxwell, A. E. (1963) *Factor analysis as a statistical method*. Butterworths, London.

Rao, C. R. (1963) The use and interpretation of principal components analysis in applied research. U. S. Office of Education contract report 2-10-065.

Rao, C. R. (1966). Discriminant function between composite hypothesis and related problems. *Biometrika*, 53:339-343.

LIST OF ATTENDEES

EIGHTEENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS
IN ARMY RESEARCH, DEVELOPMENT AND TESTING

Gary L. Aasheim	US Army Ammunition Procurement & Supply Agency, Joliet, Illinois
M. Acker	USAECOM, Fort Monmouth, New Jersey
Ronald C. Adamowicz	National Security Agency, Fort Meade, Md.
Claude Angers	Defense Research Establishment Valcartier, Courcelette, P. Q., Canada
Louis Avrami	Picatinny Arsenal, Dover, New Jersey
Paul G. Baer	US Army Ballistics Research Laboratories Aberdeen Proving Ground, Maryland
K. S. Banerjee	University of Delaware, Newark, Delaware
Norman E. Banks	US Army Ballistics Research Laboratories Aberdeen Proving Ground, Maryland
David R. Barton	National Security Agency, Ft. Meade,Md.
Walter J. Bauman	US Army Munitions Command, Edgewood Arsenal, Maryland
Ray Bell	US Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland
Alan W. Benton	US Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland
Erwin Biser	USAECOM, Fort Monmouth, New Jersey
G. E. P. Box	University of Wisconsin, Madison, Wisconsin
Barry H. Bramwell	US Army Munitions Command, Edgewood Arsenal, Maryland
Roger L. Brauer	US Army Construction Engineering Research Laboratory, Champaign, Illinois
O. P. Bruno	US Army Materiel Systems Analysis Agency Aberdeen Proving Ground, Maryland
Jagdish Chandra	ARO, Durham, North Carolina
L. D. Clements	Yuma Proving Ground, Yuma, Arizona
A. Clifford Cohen	University of Georgia, Athens, Georgia
Jacob L. Cohn	US Army Combat Development Command, Fort Belvoir, Virginia
Eugene A. Cogan	George Washington University, Alexandria, Virginia
Norman P. Coleman, Jr.	US Army Weapons Command, Rock Island, Ill.
E. Covert	Deseret Test Center, Fort Douglas, Utah
Paul C. Cox	White Sands Missile Range, New Mexico
Larry H. Crow	US Army Materiel Systems Analysis Agency Aberdeen Proving Ground, Maryland
Leo Davis	US Army Combat Development Maintance Agency, Aberdeen Proving Ground, Md.

Hany S. Dea	US Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland
Richard J. D'Accardi	USAECOM, Fort Monmouth, New Jersey
I. A. de Armon	Edgewood Arsenal, Maryland
Faust Denicola	US Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland
D. Dense	USAECOM, Fort Monmouth, New Jersey
Douglas Depriest	Office of Naval Research, Arlington, Va.
William de Rosset	US Army Ballistics Research Laboratories, Aberdeen Proving Ground, Maryland
Gerard Dobrindt	US Army Test and Evaluation Command, Aberdeen Proving Ground, Maryland
Reed E. Donnard	Frankford Arsenal, Philadelphia, Pa.
John J. Dougherty	Boeing-Vertol, Philadelphia, Pa.
Francis G. Dressel	Duke University, Durham, N. C.
John Dziobko, Jr.	US Army Munitions Command, Dover, N.J.
Churchill Eisenhart	National Bureau of Standards, Washington, D. C.
Henry Ellner	US Army Materiel Command, Washington, D.C.
Bernard F. Engebos	US Army Electronic Command, White Sands Missile Range, New Mexico
Oskar M. Essenwanger	US Army Missile Command, Redstone Arsenal, Alabama
Glenn Estell	Land Combat Materiel Test Division, White Sands Missile Range, New Mexico
Clarence F. Ey	US Army Small Arms Systems Agency, Aberdeen Proving Ground, Maryland
Harry Ferngold	US Naval Ships, R&D Center, Carderock, Md.
J. A. Feroli	US Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland
James J. Filliben	National Bureau of Standards, Washington, D. C.
Richard Fischer	Edgewood Arsenal, Maryland
R. A. Fisher	US Army Test and Evaluation Command, Aberdeen Proving Ground, Maryland
Edwards N. Fiske	Edgewood Arsenal, Maryland
Walter D. Foster	Fort Detrick, Frederick, Maryland
Fred Frishman	Office, Chief of Research and Development, Washington, D. C.
William F. Fulkerson	US Army Weapons Command, Rock Island, Ill.
A. S. Galbraith	ARO, Durham, North Carolina
SSG Bruce W. Geff	National Security Agency, Ft. Meade, Md.
Aaron H. Getzel	Frankford Arsenal, Philadelphia, Pa.
Henry Goldstein	Biomedical Laboratory, Edgewood Arsenal, Maryland
Benjamin S. Goodwin	US Army Test and Evaluation Command, Aberdeen Proving Ground, Maryland
Pamela C. Gosnell	National Security Agency, Ft. Meade, Md.
Harry Greveris	Frankford Arsenal, Philadelphia, Pa.
Maj. David S. Grieshop	Hq., Combat Development Command, Fort Belvoir, Virginia

Allan H. Layman	US Army Combat Development Maintance Agency, Aberdeen Proving Ground, Md.
Stuart Layman	Night Vision Laboratory, USAECOM, Fort Belvoir, Virginia
James Lechner	National Bureau of Standards, Washington, D. C.
Major Dick T. Leclere	US Army Infantry Board, Fort Benning,Ga.
Charles T. Lempke, Jr.	National Security Agency, Ft. Meade,Md.
William G. Lese, Jr.	US Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland
Raymond B. Loecke	US Army Management Engineering Training Agency, Rock Island, Illinois
Charles C. Lozar	US Army Construction Engineering Laboratory, Champaign, Illinois
Dale A. Madden	Food and Drug Administration, Rockville, Maryland
C. Maloney	Food and Drug Administration, Rockville, Maryland
James M. Mario	Picatinny Arsenal, Dover, N. J.
Joseph M. Marlin	Air Defense Materiel Test Division, White Sands Missile Range, N. M.
Jack Martin	US Army Combat Development Maintance Agency, Aberdeen Proving Ground, Maryland
Charles E. Mattes	Consuldyne, Inc., Bethesda, Maryland
Chester H. McCall, Jr.	Internal Careers Institute, Los Angeles, California
W. A. McCool	White Sands Missile Range, New Mexico
Bruce McDonald	Office of Naval Research, Arlington,Va.
Richard McGauley	US Army Test and Evaluation Command, Aberdeen Proving Ground, Maryland
William McIntosh	US Army Test and Evaluation Command, Aberdeen Proving Ground, Maryland
Georges J. McLaughlin	Defense Research Establishment Valcartier, Courcelette, P. Q. , Canada
David R. McVeigh	National Security Agency, Ft.Meade, Md.
Martin Messinger	Picatinny Arsenal, Dover, New Jersey
Robert Mioduski	US Army Combat Development Maintance Agency, Aberdeen Proving Ground, Md.
James N. Mitchell	Yuma Proving Ground, Yuma, Arizona
J. Richard Moore	US Army Ballistics Research Laboratories, Aberdeen Proving Ground, Md.
Paul More	Frankford Arsenal, Philadelphia, Pa.
Col. George T. Morris, Jr.	US Army Test & Evaluation Command Aberdeen Proving Ground, Maryland
Walter L. Mowchan	US Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland
Leslie Murray	US Army Weapons Command, Rock Island,Ill.
Raymond H. Myers	Virginia Polytechnic Institute & State University, Blacksburg, Virginia

Beatrice S. Orleans	Naval Ships Systems Command, Washington, D. C.
Maj. William J. Owen	US Army Infantry Board, Ft. Benning, Ga.
Rudolph J. Pabon	US Army Combat Development Command, Fort Belvoir, Virginia
Geza Pap	Picatinny Arsenal, Dover, New Jersey
Jack L. Perrin	US Army Topographic Laboratories, Fort Belvoir, Virginia
James E. Perry	Night Vision Laboratory, USAECOM, Fort Belvoir, Virginia
Elliot G. Peterson	Deseret Test Center, Fort Douglas, Utah
SP/4 Ray L. Peterson	Frankford Arsenal, Philadelphia, Pa.
Colonel Jack M. Pollin	US Military Academy, West Point, N.Y.
Colonel Lewington Ponder	US Army Test and Evaluation Command, Aberdeen Proving Ground, Maryland
Paul Randolph	New Mexico State University, Las Cruces, New Mexico
Robert C. Raybold	National Bureau of Standards, Washington, D. C.
Barry H. Rodin	US Army Ballistics Research Laboratories, Aberdeen Proving Ground, Maryland
Arnold J. Rosenthal	Celanese Research Corporation, Summit, New Jersey
Edward W. Ross, Jr.	US Army Natick Laboratories, Natick, Mass.
Alan Rowberg	US Army Medical Research Institute of Infectious Diseases, Ft. Detrick, Frederick, Maryland
George Schlenker	US Army Weapons Command, Rock Island, Ill.
MAJ. W. A. Schmidt	US Military Academy, West Point, N.Y.
Ralph E. Shear	US Army Ballistics Research Laboratories, Aberdeen Proving Ground, Maryland
Victor E. Shely	Night Vision Laboratories, USAECOM, Fort Belvoir, Virginia
J. Shea	Physics International Company
Gregory Sliwa	US Army Aviation Test Board, Fort Rucker, Alabama
Royce Soanes	Watervliet Arsenal, Watervliet, New York
Herbert Sólomon	George Washington University, Washington, D. C.
C. E. Sperry	Deseret Test Center, Fort Douglas, Utah
David J. Sookne	National Bureau of Standards, Washington, D. C.
Walter H. Squire	Frankford Arsenal, Philadelphia, Pa.
Todd Stevenson	US Army Test and Evaluation Command, Aberdeen Proving Ground, Maryland
Arthur Stuempfle	Edgewood Arsenal, Maryland
Douglas Tang	Walter Reed Army Institute of Research, Washington, D. C.
Malcolm S. Taylor	US Army Ballistics Research Laboratories, Aberdeen Proving Ground, Maryland

Jerry Thomas	US Army Combat Development Command, Fort Belvoir, Virginia
James W. Thompson	US Naval Weapons Laboratory, Dahlgren, Virginia
W. A. Thompson, Jr.	University of Missouri, Columbia, Mo.
Peter A. Thornton	Watervliet Arsenal, Watervliet, N. Y.
Robert M. Thrall	Rice University, Houston, Texas
Edward A. Timsans	Edgewood Arsenal, Maryland
Chris P. Tsokos	University of South Florida, Tampa, Florida
J. W. Tukey	Princeton University, Princeton, N.J.
Lt. Col. Luke Vavra	US Army Combat Development Maintance Agency, Aberdeen Proving Ground, Maryland
Edgar P. Virene	US Army Small Arms Systems Agency, Aberdeen Proving Ground, Maryland
V. V. Visnaw	US Army Materiel Testing Directorate, Aberdeen Proving Ground, Maryland
Mary Ann Wall	US Army Natick Laboratories, Natick, Massachusetts
G. S. Watson	Princeton University, Princeton, N. J.
Gertrude Weintraub	Picatinny Arsenal, Dover, New Jersey
Larry West	US Army Test and Evaluation Command, Aberdeen Proving Ground, Maryland
Edward F. Wilsey	US Army Ballistics Research Laboratories, Aberdeen Proving Ground, Maryland
Gary Witzling	Picatinny Arsenal, Dover, New Jersey