AFOSR - TR - 72 - 1637

AD 748277

TECHNICAL REPORT NO. AFOSR-70-6

Prepared under subcontract with the Aviation Research Laboratory, Institute of Aviation, University of Illinois at Urbana-Champaigne, for the Life Sciences Program, office of Scientific Research, AFSC, United States Air Force.

THE USE OF CENTRAL-COMPOSITE DESIGNS IN HUMAN FACTORS ENGINEERING EXPERIMENTS

Charles W. Simon

AEROSPACE GROUP
HUGHES
DISPLAY SYSTEMS AND HUMAN FACTORS DEPT.
HUGHES AIRCRAFT COMPANY
CULVER CITY, CALIFORNIA

DECEMBER 1970

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

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

1. ORIGINATING ACTIVITY (Corporate author) Aviation Research Laboratory, Institute of Aviation University of Illinois Urbana, Illinois	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

THE USE OF CENTRAL-COMPOSITE DESIGNS IN HUMAN FACTORS ENGINEERING EXPERIMENTS

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

Scientific Interim

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

Charles W. Simon

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1970	57	13

8a. CONTRACT OR GRANT NO. F44620-70-C-0105 b. PROJECT NO. 9778 c. 61102F d. 681313	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR - TR - 70 - 1637

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
TECH,OTHER	Air Force Office of Scientific Research 1400 Wilson Boulevard (NL) Arlington, Virginia 22209

13. ABSTRACT

Central-composite experimental designs for exploring and fitting response surfaces were developed nearly twenty years ago. In spite of their successful applications in chemical and engineering research, these designs have been virtually ignored in human factors engineering experimentation. This is a serious oversight since these designs, as well as the whole concept of response surface methodology, are particularly suited for research relating human performance to equipment parameters. A study of the effects of three sensor-display variables on the ability to recognize targets on a display is used to describe some of the valuable features of the central-composite design and to illustrate some of its advantages and disadvantages for human factors engineering research.

DD FORM 1 NOV 65 1473

UNCLASSIFIED
Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
HUMAN FACTORS ENGINEERING						
RESPONSE SURFACE METHODOLOGY						
CENTRAL-COMPOSITE DESIGN						

UNCLASSIFIED

Security Classification

THE USE OF CENTRAL-COMPOSITE DESIGNS IN HUMAN FACTORS ENGINEERING EXPERIMENTS

Charles W. Simon
Hughes Aircraft Company

Technical Report No. AFOSR-70-6

Prepared under subcontract with the Aviation Research Laboratory, Institute of Aviation, University of Illinois at Urbana-Champaigne, for the Life Sciences Program, Office of Scientific Research, AFSC, United States Air Force.

December 1970

Equipment Engineering Divisions
AEROSPACE GROUP
Hughes Aircraft Company • Culver City, California

CONTENTS

CONTENTS (Continued)

ILLUSTRATIONS

TABLES

FOREWORD

Central-composite experimental designs for exploring and fitting response surfaces were developed nearly twenty years ago. In spite of their successful applications in chemical and engineering research, these designs have been virtually ignored in human factors engineering experimentation. This is a serious oversight since these designs, as well as the whole concept of response surface methodology, are particularly suited for research relating human performance to equipment parameters. A study of the effects of three sensor-display variables on the ability to recognize targets on a display is used to describe some of the valuable features of the central-composite design and to illustrate some of its advantages and disadvantages for human factors engineering research.

This paper was prepared in the Display Systems and Human Factors Department of Hughes Aircraft Company under Subcontract 2 with the Aviation Research Laboratory, Institute of Aviation, University of Illinois at Urbana-Champaigne. The research is being supported by the Life Science Program, Air Force Office of Scientific Research, Air Force Systems Command, United States Air Force, under prime contract No. F44620-70-C-105 with the University of Illinois. Dr. Glen Finch of AFOSR is technical monitor of the program.

INTRODUCTION

Response surface methodology is a procedure and a philosophy for the design, the conduct, the analysis, and the interpretation of experiments performed to determine the quantitative relationship between a dependent variable (the response) and one or more quantitative, continuous independent variables.

The basic approach, first suggested by Box and Wilson in 1951, ingeniously combined elements of multiple regression theory and its specialized form in analysis of variance with special features of the factorial designs, including principles of partitioning, confounding, and fractional replicates.

The central-composite design is one of a number of experimental designs developed specifically for use in response surface exploration in order that the data collection phase be performed as completely, as cheaply, and as efficiently as possible.

Traditionally, most human factors engineers have employed the factorial, analysis of variance models in the design of their experiments. Results from such studies are reported in terms of the mean performance for the experimental conditions and the reliability of differences among these means. When the evaluation of differences between existing equipments or systems is desired, this approach is useful. However, when one wishes to determine quantitative relationships between human operator performance and a multitude of equipment parameters, these analyses of variance models are inadequate. At best, they result in expensive and wasteful research and fail to yield the information desired. Response surface methodology and central composite designs are more suited for most applied human factors engineering research today.

The human factors engineer who is preparing a research program should ask himself these questions to determine whether the RSM approach is suitable for his problem:

1. Are the critical variables quantitative and continuous?

2. Is the real purpose of this program to discover the quantitative relationship among performance and equipment variables?

3. Am I more interested in understanding the broad, less precise relationships across a large multi-variate space than in obtaining highly reliable information about a few points in a small segment of the experimental region?

4. Do I believe that the higher-order interactions, three-factor and above, exert relatively little influence on the performance in which I am interested?

5. Am I under some obligation to do the study as quickly and cheaply as possible.

6. If I handle all of the variables which are considered critical, must I become concerned about the size of the study?

7. Will many observers be unable to run all of the experimental conditions during a single session?

8. Is the number of available observers and experimental materials limited?

9. Does the experimental equipment tend to vary and make constant settings difficult?

10. Am I more concerned with obtaining answers than performing a well-defined formal experiment?

The more "yes" answers that are given to the above questions, the more likely the experimenter could find the response-surface methodology and a central-composite experimental design useful.

CENTRAL-COMPOSITE DESIGN

Box and Hunter suggested the characteristics of experimental designs for fitting response surfaces. They felt that a good design should:

1. Utilize a grid of data points of minimum density over a multi-variate space of greatest practical interest.
2. Allow for approximating a polynomial of an order tentatively assumed to be representationally adequate to fit the response surface; when no assumption is made of the form of the function initially, one starts with a first-order polynomial model.
3. Allow a check on the adequacy of the function by allowing certain combinations of higher order terms to be examined.
4. Permit the already completed design of order d to form the nucleus from which a design of order d + 1 may be built, if the assumed polynomial proves inadequate.
5. Lend itself to blocking which
    a. helps maintain a steadier experimental environment when an experimental program is extended over many data points and time, and
    b. permits an experiment to be carried out sequentially, so that certain changes can be made in the experimental plan based on information obtained from the previous data collection period.
6. Be "rotatable" so that the orthogonal axes of the experimental design can take any orientation without changing the confidence in the prediction made at any given point.

The original central-composite designs, when completed, satisfy these criteria.

Construction

Central-composite designs capable of handling any number of factors are composed of three parts. They can be built by combining the vertices of a hypercube (which is the k-dimensional analogue of a cube having 2^k vertices) with those of a measure polytope (which is the k-dimensional analogue of an octahedron having 2k vertices) and with a specified number of center points. The three-dimensional model shown in Figure 1 illustrates the cubic factorial portion, the octahedron (or star), and the center portions of the design. Examining the construction of the design reveals a number of their properties and advantages.

Regression Model

The tendency to rely on factorial designs has limited considerably the nature of research performed by human factors engineers. Because of the horrendous size of an experiment after only a relatively few factors have been included, many investigators are forced by practical considerations to limit the number of factors studied to fewer than they really believe have a critical effect on performance. When a factorial study is completed, they seldom try to interpret interactions of three-factors or higher, generally because they are unable to and often because they recognize that the effects, though statistically significant, are of little practical importance. Box recognized these facts in the construction of his central composite designs. He chose the pattern of data collection points in the designs so that the complete design would permit an approximation of the response surface with a second-order polynomial of the form:

$$Y = \beta_o + \beta_i X_i + \beta_{ii} X_{ii}^2 + \beta_{ij} X_i X_j$$

where the β coefficients are parameters to be estimate from the experimental data. Graduating models such as these are referred to as empirical models to distinguish them from theoretical models since they do not seek to explain underlying fundamental mechanisms but merely to describe a relationship which exists. Engineers will find the regression model more useful than the ANOVA models more frequently used in human factor study. Regression equations can be used to (1) estimate performance when equipment variables are specified; (2) estimate values of equipment variables needed to obtain required performance levels; (3) determine how equipment trade-offs should be made in order to optimize performance when one or more system parameters must be constrained; and (4) obtain information on the relative importance of equipment parameters in order to plan future research efforts.

Economy

A major feature of the central composite design lies in its emphasis on economy of data collection. No other consideration has so limited the quality of human factors research as the inability to look at large enough pieces of problems. While on an absolute scale, the number of factors which could be considered critical in a single experiment would probably be less than ten, the traditional human factors approach to research and the experimental designs have conspired to prevent studies of even more modest size from being conducted. The central composite designs were planned to overcome such limitations by minimizing redundancy and limiting the data collection only to that which was really necessary.

Theoretically a minimum of N data collection points are required to write a polynomial of N coefficients. Thus to write a second order polynomial (Taylor series expansion) for five factors, at least 21 observations are required; this number is considerably less than the 243 observations required to complete a 3^5 factorial design. While more than the minimum are used in central composite designs in order to make other estimates, the number still is relatively small compared to the requirements of a factorial design.

Obviously, the unequal amount of data collected in the two designs means that unequal amounts of information will be obtained from them. The central composite design was planned to provide the most essential information first and to allow an experimenter to decide whether he must collect more data, rather than making plans to collect large amounts of data from the beginning. In the five factor case, the data which is not available from analysis of the 21 data collection points, but would be available from analysis of the 243 data collection points, are all interactions and non-linear terms of greater than second order. As mentioned earlier, these seldom have much effect on performance and, if they were found statistically significant, are seldom ever interpreted. Box suggested that one collect enough data to examine lower order relationships first, and only if these do not explain the data should more data be collected to estimate the higher order terms. Therefore, many of the data points eliminated from the central composite designs reflect this point of view.

Central composite designs reduce the size of the experiment by eliminating data collection in those parts of the experimental region which are least interesting. In some cases, this is done completely; in others, it is accomplished by reducing the precision of that information which is obtained. Box reasoned that normally an experimenter will know enough about his problem to localize his experiment within the region of greatest interest. Therefore the central composite design is planned to collect the most information at the center of the region and to take less and less data the further one moves from center. The experimental region therefore is in the form of a hypersphere around the center point. Many human factors experiments, in order to fill every cell of the factorial design, expend considerable time and effort collecting data for corner cells of the design composed of experimental conditions where the factors are at their extreme levels and where performance is either the poorest or the best.

In either case, the experimenter knows full well what the results will be but must run the cells in order to complete the factorial. The spherical space (Figure 1) covered by the central composite design reduces the problem by eliminating corner cells from the experimental region, although these data points could be added later if they, indeed, prove to be of interest.

When the number of factors reach five or more, not all of the 2^k vertices of the cube need be included. Instead a fractional factorial with enough points to keep all main effects and two-factor interactions unconfounded with one another can be used. The fractional replicates, $(1/2)^p$, of the 2^k cubic portion of the central composite designs which meet the criteria of unconfounded main and two-factor effects are: $k \geq 5$, $p = 1$; $k \geq 8$, $p = 2$; $k \geq 10$, $p = 3$, etc.

Considering the above, the number of data points required for an unreplicated central-composite design are: 3 factors, 20 points; 4 factors, 30; 5 factors, 32*; 6 factors 53*; and 7 factors, 90*. Those marked with an asterisk involve a fractional replicate of the cubic portion of the design. If, for example, the complete replicate had been used with the design for 6 factors, the total number of data collection points would have increased from the 53 to 90.

Information Distribution

Box defines the "information" at any point on the response surface as the reciprocal of the variance at that point. This measure relates to the reliability of values estimated at any point in the experimental space. The central composite designs were planned with two information qualities in mind: 1) rotatability; 2) uniformity.

Rotatability. A rotatable design is one in which the "information" is equal for all points equadistant from the center. This quality permits the orthogonal axes of the experimental design to be rotated to any orientation without changing the confidence in a prediction made at any given point. The value selected for the length of the axial arm of the star portion of the central composite design determines whether

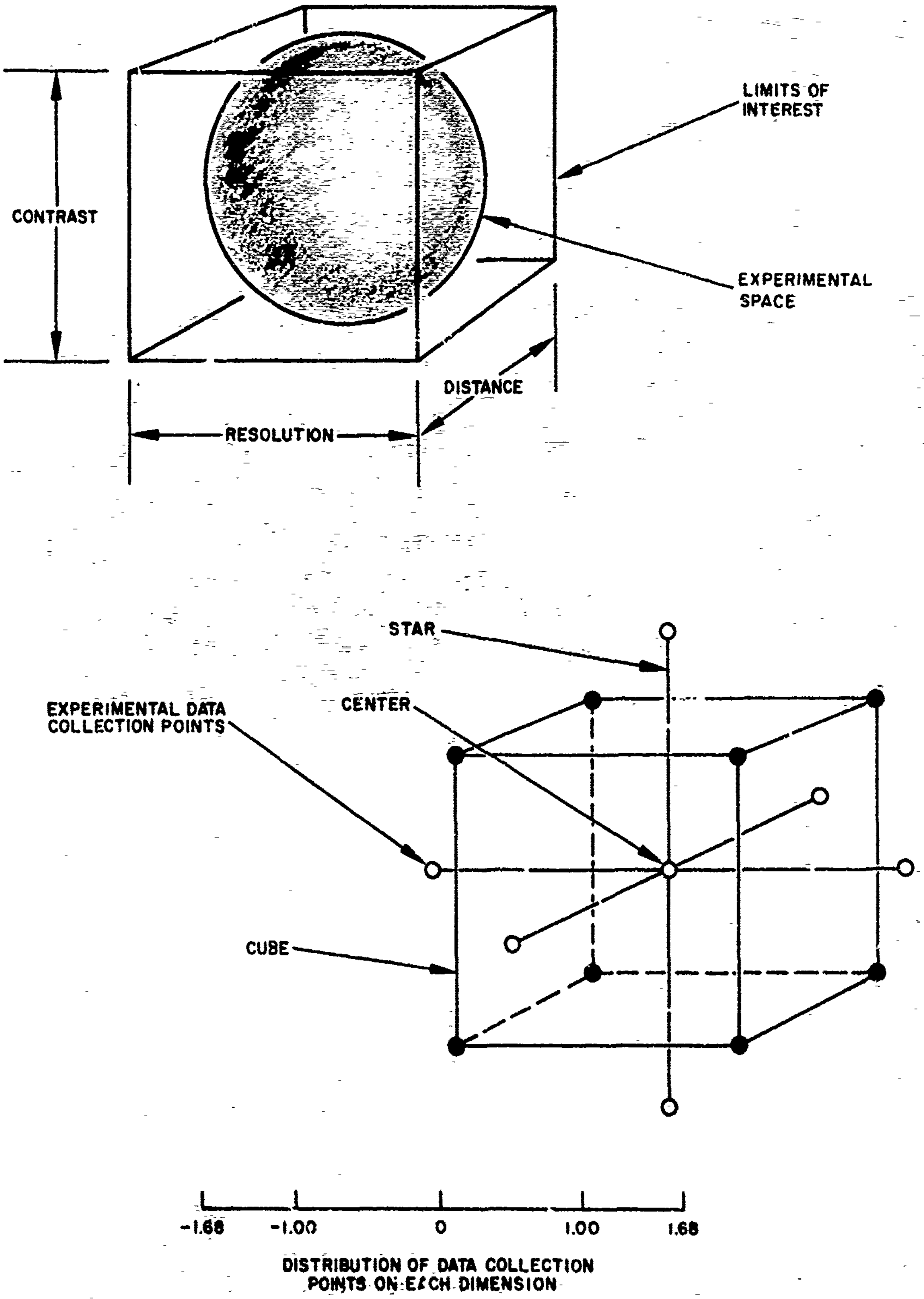

Figure 1. Geometric relationships in the central-composite design.

the quality of rotatability will exist. For rotatability in a k-factor design, the arm from center should equal $2^{k/4}$, except when fractional factorial designs of $(1/2)^p$ are used in place of the hypercube. In those cases, it should equal $2^{(k-p)/4}$.

Uniformity. Box proposed that since an experimenter may not initially have a clear idea of where the most interesting portion of the response will lie in the experimental region, the quality of information obtained should be relatively equal throughout the space. Information is considered to be uniform when the reciprocal of the variances at any point from the center of the design to the vertices of the cube are approximately equal. The number of points at the center, thus, can considerably affect the "information" profile, and must be taken into consideration in planning central-composite designs.

Orthogonal Blocking

A very useful feature of the central-composite design when used for research in human factors engineering is that of orthogonal blocking. Blocking is achieved by dividing the total data collection points into subsets or blocks of conditions which are studied together. Blocking is orthogonal when any differences in mean performances among blocks will not affect the second order regression equation. The cube and the star portions of a design each represent a natural block. If the design is large enough, the cube portion can also be fractioned so that no main effect is confounded with any other. Blocking is a particularly useful tool in human factors engineering studies where unwanted changes often occur in the human subjects, the equipment, and other environmental conditions. It is also helpful when subject time and experimental materials are limited. Examples of how blocking can be employed to improve the precision of experimental data from human factors studies are given in a paper by Simon (1970).

Meeting the criterion for orthogonal blocking affects the selected length, α, of the arms of the star, and the number of center points in the central compocite design. To guarantee orthogonal blocking in the central composite designs, it is necessary that

$$2^k/2\alpha^2 = (N_c + N_{co})(N_s + N_{so})$$

where N_{co} and N_{so} are the number of center points to be added to the cube and the measure polytope respectively. When additional blocking occurs within the hypercube, the center points should be divided equally among the sub-blocks.

In certain cases, these relationships can only be approximated. Furthermore, it is not always possible to simultaneously provide for rotatability and orthogonal blocking. For human factors studies of any size, if a choice must be made, it would appear at this time that preference should be given to orthogonal blocking.

Sequential Designs

In addition to using blocking to reduce the distortion of experimental results, Box employed it to facilitate response surface exploration. He correctly pointed out the difficulty of planning a good experiment beforehand and recommended a plan-look-replan iterative approach. He achieved this sequential plan by breaking his co iplete second order central-composite designs into blocks which were first order rotatable designs. This meant that all main effects were unconfounded with one another. He recommended beginning an experiment by completing one of the first order blocks and reviewing the data before going further. Based on these initial results, the experimenter could compare the magnitudes of th fitted coefficients in the first order model and decide whether or not one or more independent variables should be dropped from further consideration. He could re-evaluate whether the range of values being investigated should be extended or reduced. He could evaluate whether a first order model was alone sufficient to represent the unknown function by testing for lack of fit and thereby, decide whether or not to continue the study.

The iterative procedure of examination and decision could continue until the total study was completed.

The ability to test the adequacy of an equation to fit experimental data (i.e. the "lack of fit" test) is provided in central-composite designs by adding data collection points beyond the minimum required to fit a second order polynomial. Extra data points at the center of a design not only help create a uniform information surface, but can also supply an estimate of experimental error. These are the only replicated points in the basic central-composite design, although later for human factors studies, other reasons will be suggested for replicating an entire design. Second, the distribution of data collection points in the basic central-composite design allows enough degrees of freedom to write a second order polynomial, to obtain an estimate of the error, and to have enough left over to estimate the effects of higher order factors which cannot be individually isolated. If the variance associated with these higher order effects are significantly greater than the variance associated with the error, one must reject the hypothesis that the equation fits the data and assume that higher order effects are present. To identify these effects, more data points must be added.

A TARGET RECOGNITION EXPERIMENT USING A CENTRAL-COMPOSITE DESIGN

The experiment presented below illustrates some of the features of response surface methodology and the central-composite designs as they might be applied to human factors engineering research. A target recognition study* aimed at specifying requirements for the design of a sensor-display system was used to exemplify the special considerations which must be given when applying the technique to problems where human performance is investigated. Since numerous references describe the three-factor central-composite design and the rationale for its construction, the emphasis here will be upon its application to human factors engineering problems and less on its mathematical basis.

The Problem

Forward looking infra-red (FLIR) systems are thermal imaging systems in which a detector array is mechanically scanned across an infra-red telescope field of view. The detector elements are sampled and multiplexed, then fed to a CRT for display. Whereas increasing the multiplexing rate improves system performance, the corresponding video frequencies become difficult and costly to display. A mathematical model has been developed to effect the trade-off between multiplexing rates and other display parameters, as well as operator performance. A laboratory experiment was carried out to supply empirical data on human performance to support the development of the mathematical model. While several studies were performed, only one will be described here.

The prime purpose of the experiment was to determine the functional relationship between the ability of human observers to

*B. Mueller and C. W. Simon, Evaluation of Infrared Video High Speed Commutation. Wright-Paterson AFB, Ohio. Report No. AFAL-TR-69-48, 13 March 1969.

Preceding page blank

recognize armored vehicles on the display as a function of the vertical spacing of the FLIR sensor elements and the frequency of multiplexing. Electrical noise was added as a third variable.*

Experimental Procedure

The observer was shown a display on which the picture of an armored vehicle, a tank, was barely visible. His task was to identify which of ten HO-scale model tanks on a shelf located below his display represented the displayed vehicle. The image on the display could be made progressively larger by the observer, stopping intermittently to study the image and relate it to the model tanks before him. The process was continued until enough similarities between image and model could be observed to permit a positive recognition at the greatest possible range; i.e. smallest image.

Simulation of Display and Range Closure

A closed-circuit television system was used to simulate the display subsystem of the FLIR. The TV camera was pointed toward a positive, transparent image of an armored vehicle (tank) mounted before a light box illuminated from the rear. A pulley and gear mechanism allowed the light box to be moved by a drive motor toward the camera, simulating range closure. The observer had a button which allowed him to stop the closing process at any point.

---

* Originally, it was planned to study five variables, the above three plus amplifier bandwidth and CRT spot size. The final study was limited to three variables because of equipment difficulties and not because of the possible size of the study. The three factors study would have required 20 data collection points to complete a single replicate of a central composite design; a five-factors study would have required by 33. This would have been enough to estimate all main effects and all two-factor, linear x linear interactions.

Imagery

Photographs of 20 tanks were used in the study. The tanks were highly accurate HO models, reproduced from authentic blueprints and included military equipment of World War II vintage up to more recent models (Figure 2A). Miniature foliage was placed behind each tank to obscure its gross outline when viewed from a distance; a combination of both gross features and finer detail had to be visible before a tank could be recognized. Large variations in overall tank size were removed as an identifying feature in the pictures by photographing them from different distances so that the vertical dimension of each tank on the film was approximately 2.5 inches. The angular direction from which these photographs were taken provided a view of two sides and the top of the vehicle. A typical scene is shown in Figure 2B. No effort was made to authentically simulate the lights and shadows of an infra-red scene.

Equipment (Independent) Variables

The FLIR sensor consists of an array of vertically spaced elements which are scanned horizontally across the field of view of an IR telescope. During scanning, multiplexing occurs at a rapid rate down through the elements of the vertical array. As seen by a viewer, this creates an image composed of parallel horizontal lines which are being sampled intermittently. Thus, the CRT of the TV display, while physically different from the FLIR display, provides an adequate simulation from an observer's viewpoint.

Three variables of the FLIR system were simulated in the experiment:

1. Vertical Spacing of the TV Lines (V)
   To simulate the effect of an expanded field of view which could be obtained by separating the FLIR detector array elements vertically, the vertical deflection of the camera subsystem was modified to allow an adjustment of the line-to-space ratio of the display tube without changing the

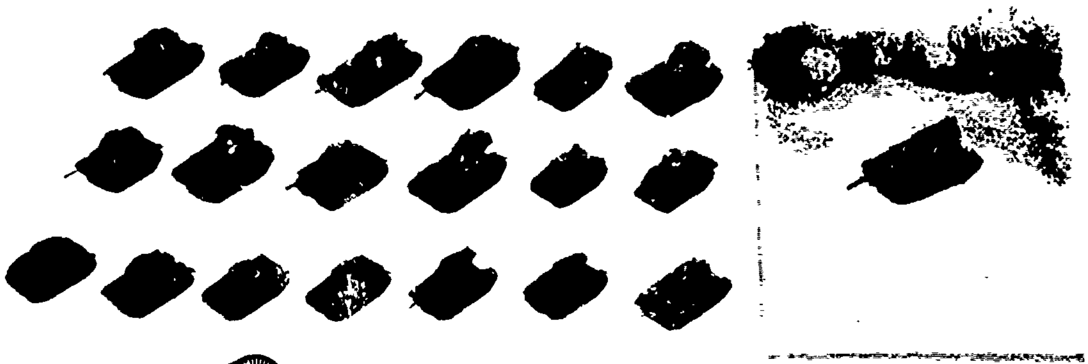

Figure 2A. Battery of tank models.

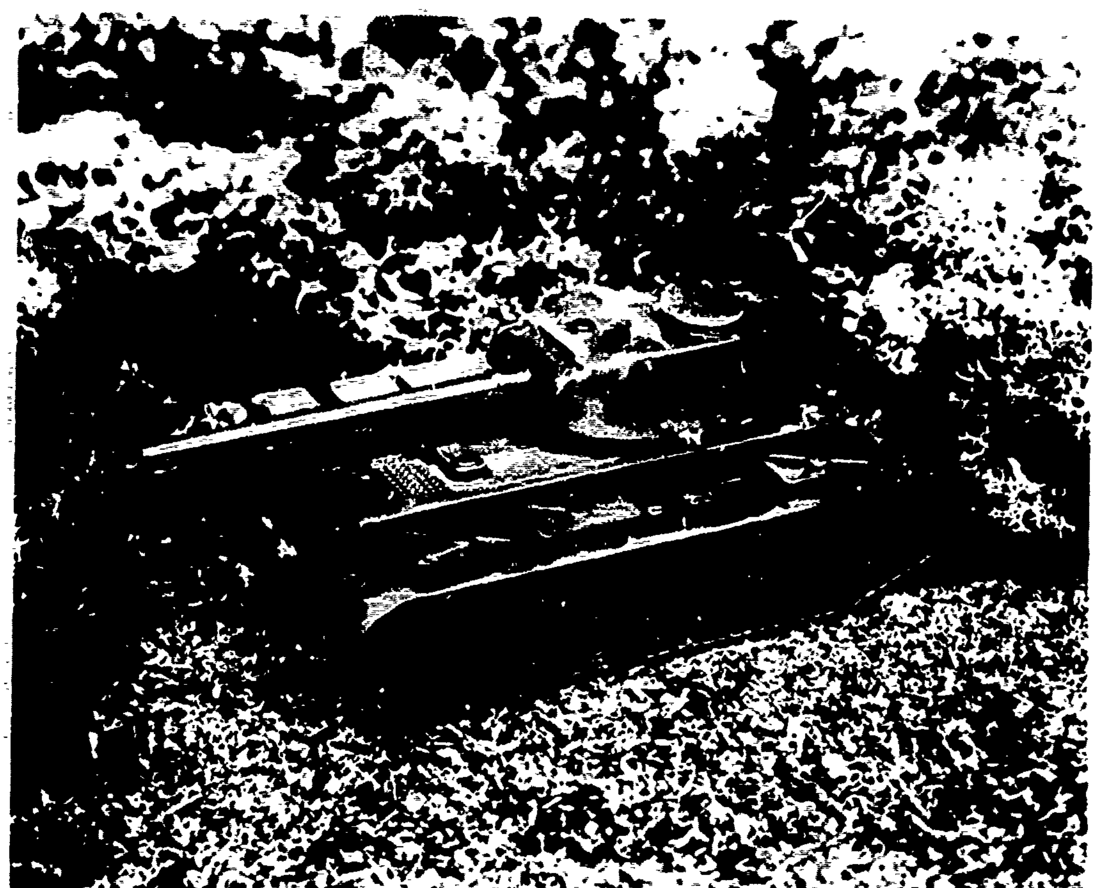

Figure 2B. Close-up of a tank model.

size of the vertical image. The vertical spacings of the horizontal raster lines were measured in lines per inch. Those selected for the study were 30, 48, 75, 102, and 120 lines per inch.

2. Display Multiplexing (H)

   To simulate the horizontal characteristics of FLIR display multiplexing, a pulse was mixed with the video in the display video amplifier; this pulse provided a variable frequency sampling that blanked the horizontal raster line of the CRT. The pulse repetition rates selected for the experiment were 0.5, 0.42, 0.3, 0.18, and 0.1 microseconds.

3. Random Noise (N)

   Random noise was mixed with the video in the CRT video amplifier. Noise components up to 5 Mc with varying amplitudes were injected. The amplitudes selected for the experiments provided peak-to-peak rms noise levels of 4.6, 6, 8, 10 and 11.4 volts.

Considerations in Selecting Factor Levels

The selection of experimental factor levels depends on several things. First, it depends on applied interests. The ranges to be considered should cover not only the conditions of immediate interest, but be broad enough to prevent having to do a new study as soon as requirements change slightly. Whenever possible, it is desirable to use a range of values which will include on one end that value at which the human will barely be able to do the task and on the other end, to include a value where the human performs about as well as possible. These points can generally be determined by a small preliminary study. Second, the use of a central-composite design itself determines the selection of the other levels. This is one disadvantage of the central-composite design: all factors must have five levels i.e., 0, ± 1, and $\pm \alpha$. There are times when this number is not practical. For example, performance may not change radically enough to justify five levels. Also there will be certain experimental factors which can not be simulated at five levels. For example, if an experimenter must use the imagery already collected on previous flight missions for a study of radar image quality, he might find that no radar maps were ever collected at five altitude levels or at the particular altitudes called for by the experimental design. Third, not only is it necessary to decide what the range of values should be, but also what the scale should be. In many human factors studies, classical psychophysical relations exist between equipment variables and human performance. Under those conditions, if the levels of the independent variable are expressed on a log scale before selecting the levels required for the central-composite design, the subsequent analysis and interpretation will be simpler than if the log transformation is made after the levels are selected and data have been collected. The importance of preliminary trial runs in planning human factors experiments cannot be underestimated.

Coding

One advantageous feature of the central-composite design is its use of coding to simplify the analysis. The real world levels of the independent variables are converted into a new coordinate system which materially reduces the calculations required for the analysis. After the calculations are made with the coded values, the results can then be translated back to real world values. As an example of coding, the conversion equation for V, lines per inch, in this study, would be:

$$V\,(\text{coded}) = \frac{V\,(\text{real world}) - 75}{27}$$

which yields the coded values shown in Table 1. The other two conversion equations for H and N in this study are: $H_c = \frac{H - 0.3}{.12}$ and $N_c = \frac{N - 8}{2}$. The numbers in the conversion equations are selected so that the center level will be zero and the levels on either side become

Table 1. Coded Values . Levels of the Experimental Variables

Variable Symbol	Coded Values					Variable Name
	-1.63	-1.0	0	+1.0	+1.63	
V	30	48	75	102	120	Lines per inch
H	0.1	0.18	0.3	0.42	0.5	microseconds
N	4.6	6.0	8.0	10.0	11.4	volts rms

±1. In practice, one works backwards by first selecting the extreme values of interest in real world terms and setting them equal to $\pm\alpha$. Plus or minus 1.63 is the appropriate α for a three factor design with orthogonal blocking. It differs slightly from the 1.68 required for rotatability, a difference of no practical importance in most studies involving human performance.

Performance (Dependent) Variable

The performance score on each trial run is the distance d that the target image was from the camera lens at the time of recognition. For the analysis in the paper, the d was determined by the numbers read from a digital counter at the time of recognition.

The score d can be converted into distance D in inches by means of the equation

$$D = 12 + 0.3d$$

and D can be expressed as spot size (SS) at the target by the following relationship:

$$SS = (2.8 \times 10^{-3})\ D \text{ inches.}$$

Experimental Design

With three independent variables, the coordinates of the basic central composite design are represented by the eight vertices of the cube, the six vertices of the star, and six center points. The geometric distribution of these 20 data collection points was shown earlier in Figure 1. The coded spatial coordinates of these 20 points are listed in Table 2, orthogonally blocked into three groups of 6, 6, and 8 conditions each. The blocked design is geometrically represented in Figure 3. Note that two of the six center points are in each block.

Data collected from any one of the blocks would permit an estimate of the linear effects of each of the three variables. Data collected from the first two blocks would complete the cube portion of the design and permit an estimate of all linear effects and two-factor interactions. Data collected from the total 20 points permits an estimate of all linear effects, all two-factor interactions, and all quadratic effects for the three variables. In addition, an estimate of experimental error and lack of fit can be made.

Observers were tested on all conditions in one block twice per day. After the sequence in a block was completed, it was repeated to provide two trials per condition. Within each block the order was "perfectly" counterbalanced among observers. This means that among observers each condition occurred only once at every ordered position within a block and was preceded or followed once by every other

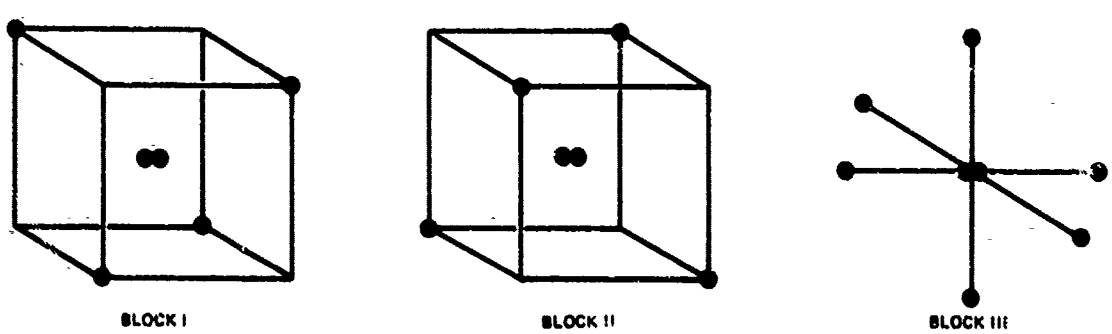

Figure 3. Blocking a central composite design

Table 2. Experimental Data Points Expressed in Coded Spatial Coordinates

Experimental Condition		V	H	N
BLOCK I	a.	1	-1	1
	b.	1	1	-1
	c.	-1	1	1
	d.	-1	-1	-1
	e.	0	0	0
	f.	0	0	0
BLOCK II	g.	-1	1	-1
	h.	-1	-1	1
	i.	1	-1	-1
	j.	1	1	1
	k.	0	0	0
	l.	0	0	0
BLOCK III	m.	-1.63	0	0
	n.	0	-1.63	0
	o.	0	0	-1.63
	p.	0	1.63	0
	q.	1.63	0	0
	r.	0	0	1.63
	s.	0	0	0
	t.	0	0	0

condition within the block. Figure 4 illustrates how the counterbalancing of observers, order, and conditions occurred within the three blocks. Effects of differences in the tank targets were removed by this counterbalancing, since every display condition within a block was tested with every target within the block.

Blocking

The value of blocking can be illustrated with this experimental design. The distribution of observers and targets among the blocks and the potential of unknown environmental changes from day to day are all likely to result in average performance differences from block to block which are not due to differences in the experimental conditions. For example, two more subjects were added in the third block to complete the counterbalancing procedure. Their performance could easily have shifted the average performance level for that block. In addition, there are different sets of targets used in each block. Since no effort had been made to equate the tank images for ease of recognition, this would be expected to cause differences in average performance levels among blocks. Finally, in any study, unspecified diurnal variations can be expected to occur which could result in unwanted shifts in performance among blocks. By using orthogonal blocking in this central composite design, average shifts in performance from block to block for any reason will not affect the estimates of the coefficients in the second order polynomial.

Several features were added in this study with human observers which might not have been used had the same design been employed in a chemical experiment. First of all, the counterbalancing and replication (with observers) of the design was introduced as a methodological rather than a statistical tool. Its purpose was not to increase data reliability (which it does do indirectly), but to improve data validity on the assumption that the counterbalancing will offset the failure to perform the time-consuming task of equating targets and to counteract any learning effects which might possibly occur. As a second precaution, the order in which the blocks were presented to

TARGETS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
ORDER	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	7	8
OBSERVERS 1	a	b	f	c	e	d	g	h	l	i	k	j	m	n	t	o	s	p	r	q
2	b	c	a	d	f	e	h	i	g	j	l	k	n	o	m	p	t	q	s	r
3	c	d	b	e	a	f	i	j	h	k	g	l	o	p	n	q	m	r	t	s
4	d	e	c	f	b	a	j	k	i	l	h	g	p	q	o	r	n	s	m	t
5	e	f	d	a	c	b	k	l	j	g	i	h	q	r	p	s	o	t	n	m
6	f	a	e	b	d	c	l	g	k	h	j	i	r	s	q	t	p	m	o	n
7	← BLOCK I →						← BLOCK II →					7	s	t	r	m	q	n	p	o
8												8	t	m	s	n	r	o	q	p
													← BLOCK III →							

Figure 4. Order in which each observer was tested on experimental conditions within each block.

each observer was counterbalanced among days (see Figure 5) to reduce possible block differences. Even though theoretically, block differences are orthogonal to the regression equation, non-linearities which are known to exist with human performance data warrants the added precaution of reducing block differences. Until more experience has been obtained with these designs in experiments with human subjects, the replications and counterbalancing techniques should probably be employed. However, the experimenter must eventually balance the advantages incurred by running enough subjects to perfectly counterbalance conditions within a block against the disadvantages of added time and costs. By counterbalancing the order that observers ran on the different blocks, it was not possible to complete only one block and examine the data to decide on how to run the remainder of the experiment. The advantages of this procedure would be considerably greater as the number of factors increased. With only three factors

OBSERVERS	ORDER (DAYS)		
	1	2	3
1	I	II	III
2	II	III	I
3	III	I	II
4	I	III	II
5	II	I	III
6	III	II	I
7	III		
8	III		

Figure 5. Order in which observers were tested by block.

the degrees of freedom available for tests within blocks are too small to be meaningful.

If counterbalancing is considered to be a prime requirement, another advantage of blocking can be shown. A "perfect" counterbalance (meaning each condition appearing once in every column, in every row, and preceding and following each condition once) of the twenty experimental conditions unblocked would have required a 20 x 20 design, or 400 data collection points. By blocking, the total number of data points are reduced to 136, by first counterbalancing within the three blocks, 6 x 6, 6 x 6, and 8 x 8, and then counterbalancing the block order at no additional cost. Therefore, in comparing experimental designs for experiments in which humans will be employed as subjects and where counterbalancing is to be used, it is not enough to merely compare the total number of data collection points for a single replicate. Instead, the effects of blocking on the total number of data collection points for the replicated design must be taken into consideration.

An interesting illustration of this point can be made by comparing two central-composite designs for a five-factor study. A full central-composite design would require 54 data collection points for a single replication. A design in which a fractional half of the cube portion is used (which would still keep main effects and two factor interactions clear) would require only 33 data collection points. However if replication and counterbalancing are employed, the 33 point design is not the more economical. The difference lies in the blocking which is possible with the two designs. The 33 point design can be divided into two blocks of 22 and 11 points each. The 54 point design can be divided into five blocks of 10, 10, 10, 10, and 14 points each. Thus, a perfect counterbalance of the 33 point design would require 22x22 plus 11x11 or 605 data collection points and a minimum of 22 subjects. A perfect counterbalance of the 54 point design would require 4 x (10x10) plus 14 x 14, or 596 data collection points and a minimum of only fourteen subjects. Furthermore, the additional three blocks in the 54 point design provide a greater opportunity for controlling unwanted environmental variations. Given the requirement for perfect counterbalancing, the larger basic design would actually be better. The experimenter working with human observers will have to decide whether the extra replications required for counterbalancing are desirable or necessary.

Observers. Eight observers were used in this study. Each were allowed three practice trials before beginning a block of trials. Six of the observers were used on all conditions in all blocks. Two of the observers were used only on the third block of conditions for reasons indicated previously.

Performance Measure. The d on the two trials per condition correlated 0.85. The two trials were averaged to obtain a single performance score per condition per observer. The median d score among the observers for each condition was then obtained to represent the average distance, d, at which all targets in each block were recognized by the observers in that block on each experimental condition. These twenty d scores, each representing performance on one of the twenty display conditions, were used in the data analysis.

By using the median performance scores for each display, any variability due to observers was essentially removed from the regression analysis. The procedure is justified on the grounds that:

1. The study was performed to determine the relationship between equipment variables and performance, i.e., the response surface. By separating equipment effects from observer effects, a clearer relationship is established facilitating the interpretation of results.

2. Tests of significance should be based on the error term of the replicated center points, rather than on the variability among individuals or the interaction of individuals with experimental conditions.

3. If there is an interest in the variance among individuals, it can be calculated separately. However, in this study, since the observers could not be considered representative of any particular group, knowledge of how their performance varied would have had little generality.

4. If one were willing to assume a linear relationship, the subject variability could be combined with the original analysis of variance.

RESULTS OF TARGET RECOGNITION EXPERIMENT

The purpose of this section is primarily to show the types of questions which can be asked of the experimental data, to illustrate some of the features available when using the central-composite designs, and to indicate considerations necessary in the interpretation of results. While the results of the present study will be used to exemplify these, the calculations required to analyze the data will not be described; such information is explicitly provided in a number of other publications.

The Raw Data

The median performance Y, expressed in terms of d, for each of the twenty experimental conditions is shown in Table 3. The first three columns of X variables, V, H, and N, replicate the coded values of the original design. The additional columns represent the remaining terms of the second order polynomial. The values for these columns are derived by performing the indicated operation on the values of the first three columns. For example, if for an experimental condition, V equals +1 and H equals -1, then for that same condition, VH would equal (+1)(-1) or -1 and H^2 would equal (-1)(-1) or +1.

Regression Analysis

A least square fit performed on the coded data matrix yielded the following multiple regression equation:

$$\hat{Y}_c = 116.14 + 10.54V - 14.95H - 15.34N - 6.62VH - 1.31VN + 1.36HN - 7.75V^2 + 1.53H^2 - 0.20N^2 \quad \text{(Equation 1)}$$

where $\hat{Y}_c$ = d and all values of V, H, and N are coded.

Table 3. Experimental Data

	Dependent(Y)	Independent(X)								
	d	V	H	N	VH	VN	HN	V^2	H^2	N^2
1)	113.25	1	-1	1	-1	1	-1	1	1	1
2)	94.00	1	1	-1	1	-1	-1	1	1	1
3)	115.50	0	0	0	0	0	0	0	0	0
4)	115.25	0	0	0	0	0	0	0	0	0
5)	81.50	-1	1	1	-1	-1	1	1	1	1
6)	142.00	-1	-1	-1	1	1	1	1	1	1
7)	125.25	0	0	0	0	0	0	0	0	0
8)	116.25	0	0	0	0	0	0	0	0	0
9)	107.00	-1	1	-1	-1	1	-1	1	1	1
10)	105.00	-1	-1	1	1	-1	-1	1	1	1
11)	198.25	1	-1	-1	-1	-1	1	1	1	1
12)	106.00	1	1	1	1	1	1	1	1	1
13)	107.75	0	0	0	0	0	0	0	0	0
14)	125.50	0	0	0	0	0	0	0	0	0
15)	62.75	-1.63	0	0	0	0	0	2.66	0	0
16)	116.25	0	-1.63	0	0	0	0	0	2.66	0
17)	123.75	0	0	-1.63	0	0	0	0	0	2.66
18)	102.25	1.63	0	0	0	0	0	2.66	0	0
19)	98.25	0	1.63	0	0	0	0	0	2.66	0
20)	81.50	0	0	1.63	0	0	0	0	0	2.66
		Experimental Design								

To express the relationship of Equation 1 in real world values instead of coded values, the following substitutions should be made in Equation 1:

$$V = \frac{V' - 75}{27} \qquad H = \frac{H' - 0.3}{0.12} \qquad N = \frac{N' - 8}{2}$$

where V, H, and N are the terms of the coded equations, and the primed terms are in real world measurement.

When this substitution is made and the equation simplified, the real world regression equation is:

$$\hat{Y}_r = 207.19 + 2.79V' - 487.20H' - 21.99N'$$

$$-2.04\, V'H' - 0.024\, V'N' + 56.51\, H'N' \qquad \text{(Equation 2)}$$

$$-0.011\, V'^2 \pm 106.4\, H'^2 - 0.05\, N'^2$$

where $\hat{Y}_r$ = d and all values of V', H', and N' are in terms of real world measurements.

Given the latter equation, an engineer can:

1. Estimate performance for values of V', H', and N' not included in the original study.
2. Estimate equipment design requirements for specified performance level.
3. Study the effects of trade-offs among two or more variables.
4. Determine the combination of variables which yield best performance.
5. Compare the effect of different factors on performance in order to better plan future research.

6. Determine the direction of slope of the response surface for planning the region in which subsequent experiments should be carried out.

As with any polynomial, it is dangerous to extrapolate beyond the region of the original experimental design. The curve which is obtained by a least square fit approximates the existing data; but beyond that point, the curve may be completely inaccurate.

Analysis of Variance

Before using the equation, the experimenter should ask:

1. How well does the equation estimate the performance in this study?

2. How well would this equation be expected to predict new data?

3. Does the second order polynomial adequately describe the empirical data?

4. Was the introduction of blocking into the experimental design justified?

5. What are the confidence limits for the predicted performance?

The first step toward understanding the data is to perform an analysis of variance. The results of the analysis of Table 3 are shown in Table 4.

Table 4 shows how the total variance was partitioned into that portion which can be accounted for by the regression equation and that which cannot (residual). The total variance is merely the variance of the performance obtained empirically from the experiment (i.e., the Y column of Table 3 and the A column in Table 5). If we had estimated performance for each of the 20 experimental conditions using the coded regression equation, #1, we would have obtained the values in Column B of Table 5 (i.e., $\hat{Y}$). The variance of this column is the

Table 4. Analysis of Variance of the Results from Coded Data

Source	Proportion	d. f.	Variance	F	ρ
Regression	.74	9	1143.	5.63	<.05
First Order Terms	.55	3	2533.	12.47	<.005
Second Order Terms	.19	6	448.	2.21	>.10
Residual	.26	10	363.		
Block	.14	2	1002.	4.94	<.05
Error	.12	8	203.		
Bias (Lack of Fit)	.10	5	286.	4.32*	>.10
Random (Center Points)	.02	3	66.		
TOTAL	1.0	19	732.		

(*Tested by Random error; all others tested by Error variance.)

Table 5. Derivation of Residual Values

Experimental Condition	A Observed Performance (Y)	B Estimated Performance ($\hat{Y}$)	C Residual ($Y-\hat{Y}$)
1	113.25	111.62	1.62
2	94.00	101.76	- 7.76
3	115.50	116.14	- 0.64
4	115.25	116.14	- 0.89
5	81.50	90.39	- 8.89
6	142.00	135.10	6.89
7	125.25	116.14	9.10
8	116.25	116.14	0.108
9	107.00	91.31	15.68
10	105.00	79.92	25.07
11	198.25	172.05	26.19
12	106.00	95.59	10.40
13	107.75	116.14	- 8.39
14	125.50	116.14	9.35
15	62.75	78.26	-15.51
16	116.25	144.64	-28.39
17	123.75	140.64	-16.89
18	102.25	112.68	-10.43
19	98.25	95.80	2.44
20	81.50	90.55	- 9.05

variance associated with Regression in Table 3. If we calculated the differences between the obtained performance (Y) and the estimated performance ($\hat{Y}$), we would have the residual values shown in Column C of Table 5 (i.e., $Y-\hat{Y}$). The variance of these numbers provides the variance for the Residual in Table 4.

Equation Strength. Each value in the proportion column in Table 4 indicates that proportion of the total variance which can be accounted for by each of the sources of variance. It is obtained by dividing the sum of squares (i.e., variance multiplied by degrees of freedom) for the particular source by the total sum of squares. Thus, the regression equation in this study accounted for 0.74 of the total variance. This proportion, R^2, is referred to as the Coefficient of Multiple Determination. The square root of this value, 0.86, represents the Multiple Regression Coefficient, R, for the equation which is equivalent to the simple correlation between the observed (Y) and the estimated ($\hat{Y}$) performance scores. This relationship is plotted in Figure 6.

Equation Fit. Some explanation must be provided for the 0.26 of the variance not accounted for by the regression equation. In Table 4, we see that 0.14 of the 0.26 was due to different performance among the blocks. The remaining 0.12 is attributable to Error of which two possible sources can be determined. The Random error represents the variability in performance among the replicated center points within blocks. The Bias error is actually that which is left over after all other sources of variance have been accounted for. This latter source, not being a result of random variation, or block differences, or any term in the second order polynomial, must represent the presence of higher-than-second order effects which cannot be isolated with the amount of data collected in the present experiment. A comparison of the two error sources yield an F-ratio of 4.32, which for 5 and 3 degrees of freedom could happen by chance more than ten times in one hundred. With so few degrees of freedom, a conservative

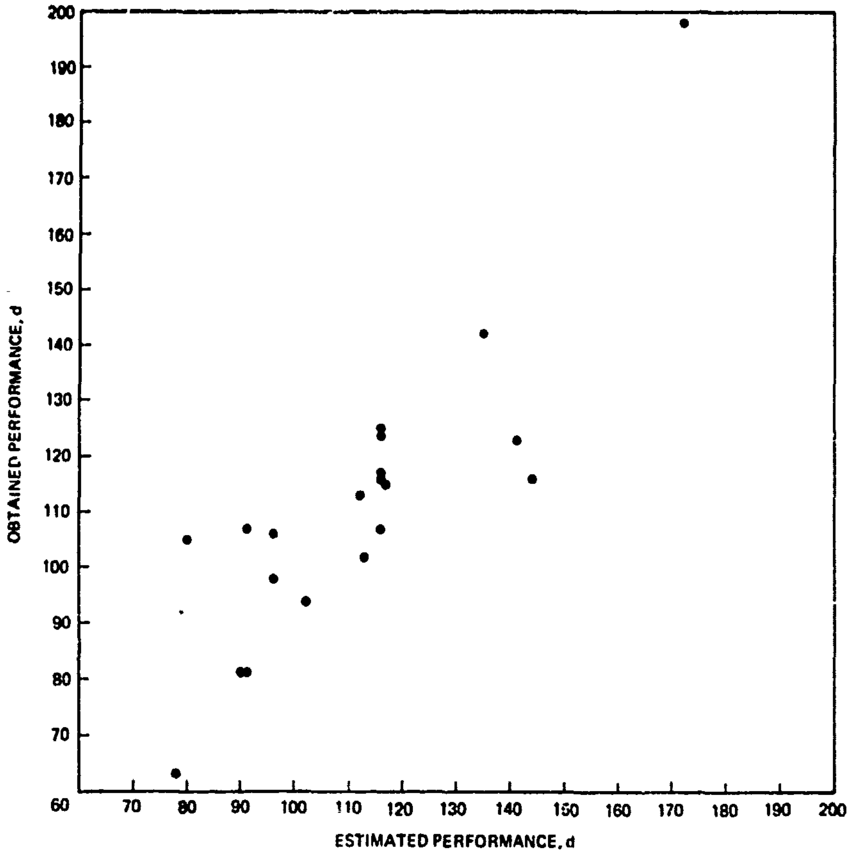

Figure 6. Scatter diagram showing the relationship between estimated and observed performance.

significance test (p = 0.10) is recommended. We therefore assumed that the B as variance was not reliably larger than the chance variance and that the second order polynomial is an adequate fit.

Block Effects. Combining the two "not significantly different" error sources into a single Error term provides more degrees of freedom for future tests of significance. Mean performances among blocks did vary significantly (Table 4) at the 0.05 probability level; however, with the central-composite design, these differences will not affect the coefficients of the regression equation. The use of blocking in this experiment, therefore, prevented unwanted sources of variance from distorting the results.

Equation Reliability. We can also test the reliability of the regression equation by calculating the ratio between the Regression and the Error variance. The F of 5.63 was statistically significant at the 0.05 probability level (Table 4). However, Box has suggested that this test is a relatively insensitive one and that to be of practical significance, the F-ratio should be four times greater than the F required for statistical significance. While this is an arbitrary value, the F test combined with the proportional contribution of the Regression equation to total variance together are the best indicators of the equations usefulness.

Equation Predictiveness. An equation which accounts for a high proportion of the variance of experimental date is not necessarily a good predictor of future data. Any set of data can be fitted by a polynomial with enough terms. Since the equation can be expected to account for some chance effects which are not likely to occur in a second data sample, the Coefficient of Determination will prove to be an overestimation when applied to a new sample. To estimate how well the equation might predict future data, corrections must be made for the number of terms in the equation relative to the number of observations from which the equation was derived. The following equation relates the two:

$$\bar{R}^2 = 1 - (1 - R^2)(n-1)/(n-t-1)$$

where n is the number of observations and t is the total number of terms in the equation. For the equation in this study, the estimated predictive strength would drop from 0.74 to 0.50.

Of course the value, 0.74, was obtained in an analysis in which 0.14 of the total variance was due to differences among blocks. We could have included blocks as still another linear term of the equation and raised the strength of the equation to 0.88. However, since the blocking effect is an artifact of the methodology, it should not be included in the regression equation. However, if we assume that the effects due to Regression and Error represent the total sources of variability, the Regression equation would then explain 0.86 of the variability of the present data (not due to blocking) and the predictive coefficient becomes 0.73.

Equation Order. Table 4 shows a partitioning of the Regression variance into that which can be accounted for by the First Order terms and by the Second Order terms. The effect of the Second Order terms in this analysis was not significantly greater than chance, implying that the response surface was essentially planar.

Let us digress at this point and remind the reader of one of the features of the central-composite design -- the sequential approach. This study was actually conducted without examining the results of first order effects after one block of data had been collected. There were several reasons why. First, the counterbalancing of blocks among days prevented a single block from being completed before the entire study was completed. Second, because of the number of degrees of freedom in a single block, any test of fit would have been relatively insensitive. Had a First Order Regression equation been written for a single block of data, only one degree of freedom would have been available each for the Bias and the Random Error. It would not have been possible to have made a meaningful test of Lack of Fit. (On the other hand, in fact, had the performance scores of each individual been used as replicates of the first block of the design, a suitable test might have been made.) Third, the use of the sequential approach is more appropriate when searching for an optimum or when the number of factors are greater than the three studied here. The inclusion of the second order terms do improve the fit of the present experimental data -- increasing the proportion of variance accounted for by 0.19.

Confidence Limits. The Error variance can be used to provide an estimate of the confidence limits for the equation as a whole. For the 8 degrees of freedom, 95 percent of the estimated responses will fall between ±3.65 (in terms of d). In practice, the confidence limits at any point in the space will vary slightly at different distances from the center of the experimental region.

Interpreting the Equation

Our analysis has shown that the equation does in fact describe the response surface. How can it be used?

By substituting values for the independent variables V', H', and N' in the equation, we can obtain performance estimates useful for evaluating capabilities of future systems or for judging the effects of trade-offs among the independent variables.

By examining the equation itself, a better understanding of the relationships among the dependent and independent factors can be gained.

Individual Terms. Mathematically, each coefficient of the equation represents how much change occurs in d for each unit of change in the particular term being studied. For example, in the real world regression equation, No. 2, the coefficient for the V term indicates that when a new line per inch is added to the display, the recognition range increases 2.79 d's. Unfortunately, to understand the effect of a particular variable is not that simple for two reasons.

First of all, this V term represents only the linear component of the effect of V. To estimate the total effect of changing lines per inch, all of the terms which include the V must be considered. Second, the terms of the real world regression equation (No. 2) are not independent. This was determined by examining the correlation matrix used to derive the equation. Therefore, for this equation it is not even possible to determine from the coefficient the effect of any single term. If one were to examine the table of intercorrelations among the 20 conditions of the nine terms of the equation, one would find, for example, that V correlates 0.65 with VH, 0.81 with VN, and 0.98 with V^2. Thus a change in performance due to the linear interaction between V and H cannot be determined in isolation from the effects of V separately because these terms are not independent of one another. Similar intercorrelations can be found among other terms of the equation. All this means is that while the real world regression equation as a whole represents an expression which will represent the response surface with the least average error, the effect of any term cannot be determined individually.

The table of intercorrelations for the coded independent variables, however, would show all but the quadratic terms independent of one another. The points of the central-composite design were selected with that goal in mind. The three quadratic terms were correlated -0.07. With the coded equation, the effect on $\underline{d}$ for unit changes in the isolated terms can be determined from the coefficients with only a slight error for the quadratic terms.

The significance of the coefficients of each of the terms in the equation can be tested. However, when the purpose of a study is to describe the response surface, Box and Hunter did not regard such a test with much favor. They wrote:

> "It should be noted here that the individual coefficients of the model have not been separately tested for significant departure from zero. If this has been done, and one coefficient was found to be not significantly different from zero, we would not be entitled to replace the given estimate with a zero, for regardless of its magnitude, it is still the best estimate of the unknown coefficient. To replace this estimate by a zero would in effect be replacing a best estimate by a biased one. The important test concerns the order of the model; i.e., whether a model of first order, or of second order, adequately represents the unknown function. Another test that could be run would be to determine whether a particular variable x_i contributed significantly to the response. In this case the sums of squares of all the coefficients bearing an i subscript would be pooled and then tested."

While an engineer might be interested in the relative effects of certain variables in order to decide where best to distribute time, money, and effort in improving a system, this might better be determined by a more direct approach in which changes in equipment factors are related to their cost, then seeing how much improvement in d is possible for differences in dollars.

Graphic Analysis. When an experimental region consists of only three dimensions or if an equation were reduced to only three factors (including their interactions and quadratic forms), it is possible to represent the response surface graphically. Figures 7-A, B, and C illustrate how this was done for the present study. The surface appears the same for either the Coded or the Real World regression equations provided the scales of the axes are equated. The solid contour lines represent equal performance levels (i.e., recognition ranges in terms of d) in the same way that lines on a contour map represent equal terrain altitudes. The three parts of Figure 7 represent three levels of the RMS noise; the size of the plotted area at each level characterizes the spherical shape of the experimental space.

An examination of these figures can provide some insight into the relative effects of variables and their interactions upon performance. These figures can be used to evaluate the effects of tradeoffs among variables, the shape of the response surface, the direction in which the optimum performance will be found and which combinations of the variables are required to optimize performance, if the optimum lies is within the experimental space.

To illustrate how Figures 7A, B, and C can be used, scan across the three figures and determine performance at the center. The d values are approximately 95, 110, and 135. This suggests that within the experimental region, the effect of RMS noise on performance was essentially linear, a fact supported by the very small coefficient for the N^2 term in the equation.

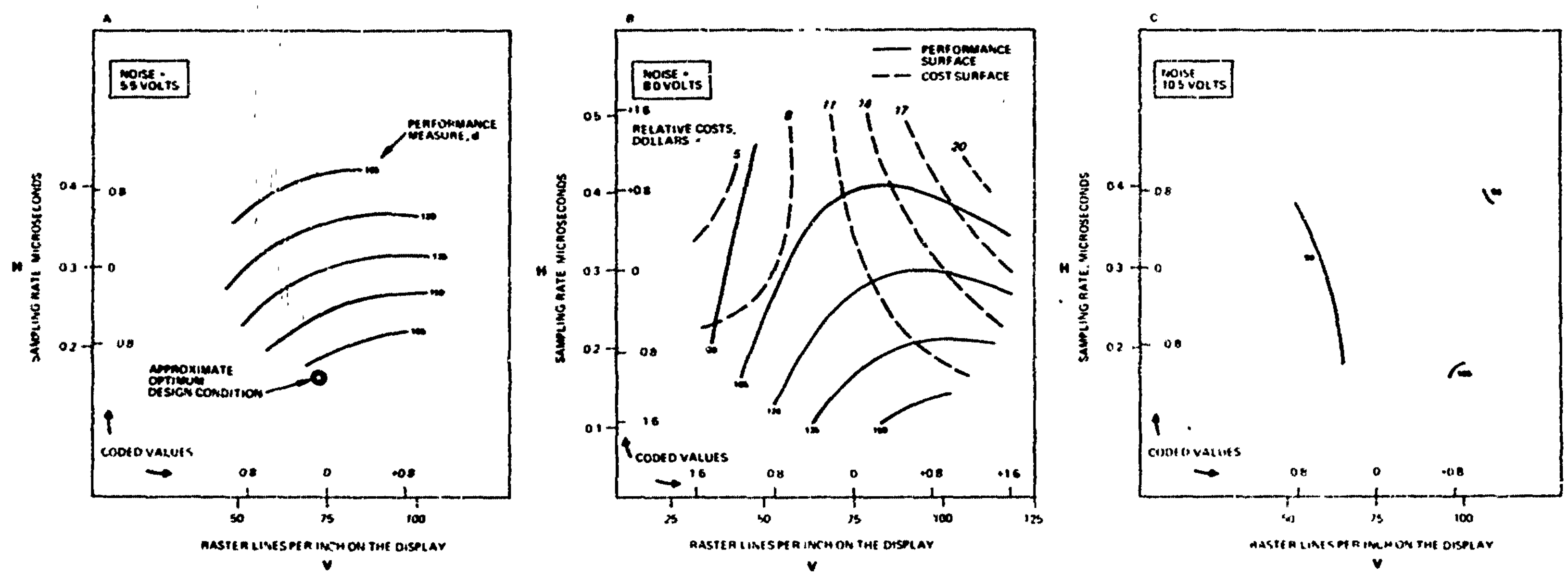

Figure 7. Graphic representation of experimental response surface.

What is the effect of sampling rate on performance? From the Figure 7, the strong interaction between H (sampling rate) and N is evident. When the noise level is high (Figure 7-C) changing the sampling rate (or for that matter changing lines per inch) has essentially no effect on performance. When the noise level is low (Figure 7-A), increasing the sampling rate results is a rather extensive reduction in recognition range. On the other hand, at this high noise level, the effect of changing the number of lines per inch on the display is practically insignificant.

At the center of the experimental space (Figure 7-B), both V and H affect performance. Performance is best (i.e., recognition occurs at the greatest distance) when the greatest number of lines per inch and the slowest sampling rate are used. That is not surprising; however, the graph also shows that if V and H are decreased together, recognition range will remain relatively constant.

Multiple Criteria. Plotting the data also facilitates the examination of multiple criteria. It is not enough for an engineer to know which combinations of V, H, and N would result in the greatest recognition range; it's equally important that he take into consideration the costs. To illustrate, the experimental conditions in Table 3 were related to dollars as well as to recognition distance. Estimates were made of the relative costs of the different combinations of sampling rates, lines per inch on the display, and noise levels for each of the 15 different experimental conditions. These relative values are shown in Table 6. A second order polynomial was derived from this data as it had been done for the performance measurements. The equation for the coded data which was obtained was:

$$\$ = 10.49 + 3.49\,V + 1.01\,H + 0.58\,N + 2.64\,VH + 1.26\,VN + 1.20\,HN + 0.412\,V^2 + 0.303\,H^2 + 0.622\,N^2 \qquad \text{(Equation 3)}$$

Table 6. Dependent and Independent Variables Related to Cost

Design Condition	Dependent(Y)	Independent(X)		
	Relative Costs	V	H	N
1)	12.63	1	-1	1
2)	17.50	1	1	-1
3)	10.86	0	0	0
4)	7.58	-1	1	1
5)	15.10	-1	-1	-1
6)	6.60	-1	1	-1
7)	7.27	-1	-1	1
8)	11.41	1	-1	-1
9)	19.49	1	1	1
10)	4.28	-1.63	0	0
11)	8.10	0	-1.63	0
12)	10.25	0	0	-1.63
13)	17.81	1.63	0	0
14)	13.41	0	1.63	0
15)	12.96	0	0	1.63

This equation, plotted for the N = 0 condition, is shown as the dashed contours overlaying the performance contours in Figure 7-B. Given this information, the engineer can make trade-offs between performance and costs for different display designs. The combined information in Figure 7-B could be interpreted, for example, as follows: reducing the number of lines per inch on the display from approximately 125 to 90 will not materially affect the detection range of 135 d, but would reduce costs from approximately \$14x to \$11x. Or, it will be necessary to spend at least \$11x to achieve maximum recognition range.

Optimization. Once the equation for the response surface has been derived, it can be used to seek the optimum combination of variables to produce the greatest yield. In the present study, the position of the maximum recognition range in the three dimensional coordinate system was found by differentiating the coded regression equation #1 with respect to V, H, and N in turn. The coordinates of the stationary point (maximum or minimum) are obtained by making these differentiated equations equal to zero, and arriving at the unique solution. In this example, the coordinates (coded) of the maximum point are:

V = -0.106	H = -1.17	N = -0.891
(72.3 lines/inch)	(0.16 micro-sec)	(6.22 volts RMS noise)

The numbers in parentheses represent the coordinates expressed in real world measurements. The approximate location of this optimum combination is shown by a star in Figure 7A (although that noise slice was -1.25 rather than the required -0.89).

In certain cases, the optimum point may not fall anywhere near the experimental region. The same caution expressed elsewhere, apply to this situation: beware of extrapolating too far beyond the region from which the original data were collected. One might use this estimated optimum (plus an observation of the rate and direction of change of the response surface) to suggest where a second experimental study might be located which hopefully would encompass the optimum point.

On the other hand, for some human factors studies, knowing the coordinates where performance is optimum may be of little interest. In certain cases, the experimental region is the only one of any concern because of other constraints outside of the experiment. For example, where range itself is an experimental variable in a target acquisition study, the knowledge that target recognition would be improved at closer ranges than were studied in the experiment may be irrelevant if that range were too small to allow an adequate time for missile launch. In other cases, the nature of the variables would permit the experimenter to guess the optimum combinations without need of experimentation. For example, an experiment is not needed to know that air-to-air detection ranges will increase as the size of the target increases, the contrast between target and sky increases, the cone of uncertainty as to target location becomes smaller, and so forth. Studies involving such variables are generally performed to obtain response surfaces from which to make quantified estimates of performance or from which the effects of trade-offs among certain variables can be determined.

Canonical Equations. When a polynomial involves more than three factors, simplified graphic representations are no longer possible and interpretation becomes difficult. Box suggested that second order polynomials be transformed to canonical form. Essentially, this transformation shifts the response surface around so the stationary points are shifted to the center of coordinate system (thereby eliminating the linear terms from the equation) and the axes are rotated so the cross-product terms are eliminated. This leaves a simplified equation composed of only the quadratic terms in a new coordinate system. While relating the new equation directly to the real world may be difficult, it does facilitate a visualization of the shape of the response surface of the complex, multivariate space. For each variable then, the sign of the quadratic term will indicate the direction of change in the response surface for each unit change of that variable to one side of center or the other. This information can be useful for estimating the approximate direction out of the experimental region in which further improvement in performance might be expected if sequential studies were to be performed.

Supplementing the Basic Study

There are relatively few experiments which really provide all of the required answers. If we were interested in mapping a response surface and had successfully picked the correct area of greatest practical interest, we might still wish to make additional measurements to supplement the original data.

One may wish to supplement a basic study in a number of ways. One might collect additional data at points adjoining the original design to see how the surface changes in that expanded area. One might wish to replicate within the design, possibly in the region of optimum performance, in order to obtain more precise information about that part of the space. One might wish to study the effect on the response surface when new factors were added.

With human observers, running additional conditions later than the original runs creates the same types of problems that can occur when a study is blocked. Relatively little experience has been accumulated as to the best way to proceed for running additional points. Overlapping data points with the original design can provide a basis for fitting the parts of the experiment together. When it can be anticipated that some additional data will be wanted (such as certain corners of a rectangular space which were omitted with the spherical shape of the central composite designs), these might best be run along with the points of the original data. The basic analysis of the central composite design can be made first, and the effects of the additional points can be examined later.

Box and others have warned of the dangers of attempting to examine too large a space (not in terms of the number of variables, but in the range covered by each variable). This warning is based on the assumption that the further apart the data collection points are, the less likely the second order polynomial will make an adequate fit.

What would happen if the second order polynomial had not adequately represented the observed data? The data might be transformed in order to simplify the relationship (much as a log transformation may linearize what was originally a curved relationship between subjective judgements of brightness and light intensity in foot lamberts). Or one might add additional data points to the original design in a number and location sufficient to isolate the third order effects. There are some experimental designs which permit this to be done sequentially much as the original central-composite design is built from first order to second order models.

SUMMARY AND CONCLUSIONS

A three-factor target recognition study was carried out using the central-composite design for selecting the coordinates of the experimental data collection points. This study was used to illustrate some of the advantages and some of the limitations of response surface methodology for human factors engineering research.

Some advantages are:

1. It provides information in a form which an engineer can use best. Results are expressed quantitatively as multivariate functions approximated by second order polynomials. Linear, quadratic, and interaction effects are determined.

2. It collects the information economically, permitting more comprehensive studies to be performed. The minimum number of data points are used to express the functional relationship, to provide some estimate of error, and to provide some additional data from which the fit of the equation can be evaluated. By collecting data in a spherical region, the center of the space is emphasized and certain irrelevant conditions at the corners of the experimental space are eliminated.

3. It lends itself to collecting the data in incomplete blocks. This permits a large multi-variate experiment to be broken into manageable size, it reduces unwanted sources of variability, and it permits the more efficient utilization of subjects and materials when these are limited in number. Blocking enables a study to be carried out in a series of sequential steps which enable the experimenter to change the characteristics of the experimental design after the study has begun and even terminate the study with meaningful data before the originally planned design has been completed.

4. It facilitates both the analysis and the interpretation of results. With the results presented in equation form rather

Preceding page blank

than as an acceptance or rejection of a hypothesis, many questions can be asked of the same data. Coding the independent variables simplifies both the analysis and the interpretation of the results. Interpretation is further simplified when the results are presented graphically or in canonical form.

Some disadvantages are:

1. The central-composite design requires a rather rigid pattern of data collection points which do not always fit the needs of human factors engineering studies. Five levels of each factors are required. They must be spaced symmetrically about the center at particular locations on a scale, which changes as the number of factors in the study change.

2. Existing designs are limited primarily to studying first and second order response surfaces. They were never intended for use with qualitative variables, and they do not lend themselves to the investigation of the effects of single terms.

This paper attempted to show, however, that the advantages override the limitations. Furthermore, since the original central-composite designs were introduced, other designs suitable for response surface exploration have been developed. What Box did was to provide a total methodology, a philosophy of applied research, of which the pattern of the data collection design is only one part. He has demonstrated an approach which will permit more factors to be included economically and reasonably into a single experiment, enabling the human factors investigator to obtain an overview rather than a piecemeal examination of a problem. It represents a systems approach to engineering design. Furthermore, it forces an experimenter to become involved in his experiment and to make decisions for improving his data, rather than allowing the all too common situation to exist in which studies are carried out in cookbook fashion.

Central-composite designs were planned originally for chemical research. It is natural that certain modifications of the method should be expected in research involving human observers. Problems of presentation order, the need for counterbalancing among observers, the economical use of replication, the special problems of data transformation, and the separation of observer effects from equipment effects must all be considered for human factors engineering experimentation. The problems arise less from the technique and methodology and more from the lack of experience in using them. The paucity of attempts to make full use of these designs makes it difficult to anticipate what must be done to maintain their positive qualities and at the same time fit them to studies involving human subjects. Kempthorne, at the Tenth Conference on the Design of Experiments in Army Research Development and Testing, 1965, stated it best: "What we really lack are accounts of actual experiences with the various methods. Perhaps a good practical strategy is to use the 'deterministic' schemes at first, and then turn to the stochastic schemes when the former ceases to give advances."

BIBLIOGRAPHY

The following articles and books each include well-written chapters on the development of response surface methodology and the central-composite designs. All contain examples with the analyses worked out in detail.

Box, G. E. P., and Hunter, J. S., "Multi factor experimental designs for exploring response surfaces." Ann. Math. Stat., 1957, 28, 195 - 241.

Box, G. E. P., and Wilson, K. B., "On the experimental attainment of optimum conditions", Journal of the Royal Statistical Society, Series B, 1951, 18, pp. 1 - 45.

Chew (ed.), V., Experimental Designs in Industry, Wiley, New York, 1956. Part I, Chapter 5, Box and Hunter, "Experimental designs for the exploration and exploitation of response surfaces".

Cochran, W. G., and Cox, G. M. Experimental Designs, Wiley, New York, 1956. 1957 (2nd Edition). Chapt. 8-A, "Some Methods for the Study of Response Surfaces."

Davies, O. L., (ed.), The Design and Analysis of Industrial Experiments, Hafner Pub. Co., New York, 1956, Chapter 11, "The determination of optimum conditions". pp. 495 - 578.

Hill, W. J., and Hunter, W. G., "A review of response surface methodology: A literature survey." Technometrics, 1966, 8, pp. 571 - 590.

The following papers clearly and succinctly describe the central-composite design and its analysis in step-by-step detail.

Bradley, R. A., "Determination of optimum operating conditions by experimental methods. I. Mathematics and statistics fundamental to the fitting of response surfaces", Industrial Quality Control, 1958, 15, pp. 1 - 5.

Hunter, J. S., "Determination of optimum operating conditions by experimental methods, models and methods", Industrial Quality Control, Part II-1, 1958, 15, December, pp. 16 - 24; Part II-2, 1959, 15 January; Part II-3, 1959, 15, February, pp. 6 - 15.

Preceding page blank

In addition to the above, the following books and papers represent more general references for understanding some of the statistical concepts discussed in the present paper.

Draper, N. R., and Smith, H., Applied Regression Analysis. John Wiley and Sons, N. Y., 1968.

Hays, W. L., Statistics. Holt, Rinehart, and Winston, N. Y., 1963.

Mendenhall, W., Introduction to Linear Models and the Design and Analysis of Experiments. Wadsworth Publishing Co., Belmont, Calif., 1968.

Simon, C. W., Reducing Irrelevant Variance Through the Use of Blocked Experimental Designs. Hughes Aircraft Co., Culver City, Calif., Tech. Report No. AFOSR-70-5, November 1970.