Factor Screening Methods in Computer Simulation Experiments

Douglas C. Montgomery, Ph.D., Ginner Weatherby

School of Industrial and Systems Engineering,
Georgia Institute of Technology, Atlanta Georgia 30332

Abstract

The use of a computer simulation model may be viewed as an experiment in which a set of k controllable factors are varied according to an experimental design, and the effect of these factors on the output observed. Usually not all k factors will be active, and considerable efficiency in the use of the model will result if the subset of active factors can be identified. Factor screening methods are useful in identifying the set of active factors. This paper discusses experimental design methods useful for factor screening in computer simulation. The general strategy recommended is group screening methods combined with 2^{k-p} fractional factorial designs. Some variance reduction considerations are also discussed.

1. INTRODUCTION

1-1. Computer Simulation

A computer simulation may be viewed as an experiment in which a set of controllable input or independent variables are combined to produce at least one output variable, usually called the dependent variable or response. In performing a computer simulation experiment, the analyst will usually have one of two objectives in mind:

1. Investigate the relationships between the independent variables and the response, determining, if possible, which factors exert the greatest effect on the response, and the extent of interaction between the factors.

2. Determine the set of factor levels that, over some appropriate region of interest, optimize the response(s).

The use of experimental design methods in computer simulation results in significant advantages to the analyst, including simplicity of data interpretation and (usually) economic efficiency with respect to the total number of simulation runs required. For background reading in experimental design, consult Cochran and Cox [1957], Davies [1956], Hicks [1973], Montgomery [1976], or John [1971]. For discussion of applying experimental design methodology to computer simulation, see Burdick and Naylor [1966], Fishman [1973], Hunter and Naylor [1970], Ignall [1972], Kleijnen [1975a, part II], [1977], and Montgomery and Evans [1975].

1-2. The Need for Factor Screening

Assume that a computer simulation model may be described by a set of k controllable input variables or factors. These factors are generally of two types:

1. Factors that are controllable or subject to design in the "real world" system being modeled, such as inventory reorder quantities, service rates, or the rate of fire of a weapons system.

2. Factors that are not controllable in the real system, such as demand, weather effects, or the location of enemy troops or equipment. For purposes of conducting the experiment, however, all k factors will be assumed to be controllable in the simulation; that is, we may induce desired weather effects, or control the movements of an enemy submarine.

In general, not all of these k factors will be equally important with respect to their effect on the response variable(s). One frequently finds that only a subset, say $g < k$, of the original k factors are important in explaining the response variable. However, we usually do not know the value of g, nor do we know which g factors are important. This situation is discussed by Jacoby and Harrison [1962], who state that the problem is frequently encountered in computer simulation.

†Sponsored by the Office of Naval Research under Contract N0014-78-C-0312

CH1437-3/79/0347-0358$00.75

1979 Winter Simulation Conference

The problem of experimentation to discover the size and composition of the subset of active factors is called the factor screening problem. Failure to identify an active factor can lead to serious bias in the analysis and conclusions drawn from a model if that factor is ignored in subsequent experiments. On the other hand, experimentation with negligible factors is undesirable as it consumes the resources of experimentation needlessly, and may increase the noise level in the data to the point when real effects are more difficult to discover. Clearly, identification of the set of active factors plays a critical role in the successful use of this methodology.

This paper discusses the available statistical methodology useful in factor screening. Guidelines for development of a screening strategy are also provided. Other questions, including the implementation of variance reduction methods, choice of levels for factors thought to be negligible, and some details of parameter estimation in linear statistical models are discussed.

1-3. Factors, Levels, and Parameter Estimation

Suppose that $x_1, x_2, \ldots, x_k$ are the controllable factors in a computer simulation experiment and y is the (single) response. We assume that the general structure of the simulation is such that it can be expressed in the form

$$y = f(x_1, x_2, \ldots, x_k) + \varepsilon \qquad (1)$$

where f is a functional relationship that determines the mean value of the response y, and ε is an error term such that $E(\varepsilon) = 0$. In factor screening problems it is almost always sufficient to assume that f is linear in the unknown parameters that relate the response to the factors. For example, one possible model would be

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon \qquad (2)$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are unknown parameters. To perform an experiment with this system, we must choose a set of values or levels for each factor, and then run the computer simulation model at some subset (or possibly the full set) of the factor level combinations.

In most factor screening experiments, we are attempting to determine the effect of the factor, not necessarily trying to develop a useful predictive or interpolative equation. Consequently, a relatively small number of factor levels is generally employed. Often two levels, arbitrarily called high and low, are sufficient. The need for more than a small number of levels often indicates that the region of exploration for x is too large. The spacing of factor levels is also important. Levels should be far enough apart to measure anticipated effects, but not so far as to cause nonlinearities in the functional relationship to distort or mask significant effects.

The method of least squares can be used to estimate the main effects and interactions. Suppose that we can describe the system by a linear statistical model, say

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i, \quad i=1,2,\ldots,n \qquad (3)$$

where y_i is the i^{th} response, x_{ij} is the i^{th} level of factor j, and β_j, $j=1,\ldots,k$ are unknown parameters. Letting $\underline{y} = (y_1, y_2, \ldots, y_n)'$, $\underline{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$, $\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$, where the prime denotes transpose, and letting X denote an $n \times (k+1)$ matrix whose first column is all ones and whose $(i,j+1)^{st}$ element is x_{ij}, then it is well-known that (3) can be written as

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \qquad (4)$$

The least squares estimators of $\underline{\beta}$ are given by the solution to the normal equations

$$(X'X)\hat{\underline{\beta}} = X'\underline{y}, \qquad (5)$$

or

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y} \qquad (6)$$

assuming that $(X'X)^{-1}$ exists. If $E(\underline{\varepsilon}) = \underline{0}$ and are uncorrelated with constant variance σ^2, and the model is correct, then

$$E(\hat{\underline{\beta}}) = \underline{\beta} \qquad (7)$$

and

$$\mathrm{Cov}(\hat{\underline{\beta}}) = \sigma^2(X'X)^{-1} \qquad (8)$$

The assumption of independent observations with constant variance will likely not hold in a simulation experiment. There are cases where a variance reduction strategy induces a correlative structure between the observations. In cases where the assumption of uncorrelated errors with constant variance does not hold, the method of weighted least squares is useful. If V is a matrix of weights (chosen proportional to the variances and covariances of the errors) then the weighted least squares estimator of $\underline{\beta}$ is

$$\hat{\underline{\beta}}_{WLS} = (X'V^{-1}X)^{-1}X'V^{-1}\underline{y} \qquad (9)$$

$\hat{\underline{\beta}}_{WLS}$ is an unbiased estimator with covariance matrix

$$\mathrm{Cov}(\hat{\underline{\beta}}_{WLS}) = \sigma^2(X'V^{-1}X)^{-1}. \qquad (10)$$

Users of statistically designed experiments are accustomed to analyzing the resulting data by relatively formal methods, such as the analysis of variance. In factor screening problems this is usually not done, as the least squares estimates of the model parameters (or the effects) allow significant factors to be identified. Often it is not practical to conduct a formal analysis of variance because of the small number of degrees of freedom that remain for error.

1-4. Previous Work on Factor Screening in Simulation

There has been little analysis or interpretation of factor screening methodology in the computer simulation environment. Kleijnen [1975a,b], [1977] and Hunter and Naylor [1970] have suggested the use of fractional factorial designs and group screening (a procedure in which factors are arranged in sets) methods in simulation. However, they do not give any examples. Only Kleijnen [1975b] attempts to give any guidelines for the choice of a factor screening strategy. Nolan and Sovereign [1972] employ a group-screening strategy in a large-scale simulation model of airlift and sealift operations. However, they do not give any details of the screening methods used. Williams and Weeks [1974] have proposed using special types of p^n factorial designs for factor screening in simulation. In general, there does not presently seem to be any systematic collection or evaluation of factor screening methods available, nor is there much specific analysis of their use in computer simulation.

2. EXPERIMENTAL DESIGN METHODS IN FACTOR SCREENING

2-1. Full Factorial Designs

Full factorial experiments could be used for factor screening. The most efficient design to consider is the 2^k factorial; i.e., k factors each at two levels. The statistical model for a 2^k design would include k main effects, $\binom{k}{2}$ two-factor interactions, $\binom{k}{3}$ three-factor interactions, ..., one k-factor interaction.

For even a moderate number of factors the total number of runs in a 2^k factorial design is large. Since resources are usually limited, the number of replicates that the experimenter can employ may be restricted. Frequently, available resources will only allow a single replicate of the design to be run, unless the experimenter is willing to omit some of the original factors.

With only a single replicate of the 2^k it is impossible to compute an estimate of experimental error, that is, a mean square for error. However, the usual approach to the analysis of a single replicate of the 2^k is to assume that certain higher-order interactions are negligible. The statistical analysis of these designs is well-known (see John [1971] or Montgomery [1976]).

When a large number of effects are estimated, we may wish to find some formal basis for declaring which effects are significant. If there is either replication or insignificant factors pooled to estimate error, we could possibly use analysis of variance methods and conduct formal statistical tests. However, if variance reduction methods such as common random numbers have been used, the usual analysis of variance statistical tests may not be appropriate. For a discussion of this problem in simple designs, see Heikes, Montgomery, and Rardin [1976]. A useful approach is to plot the effects on normal probability paper. Negligible effects on such a display will fall approximately along a straight line, while real effects will lie far from the line. For examples of this methodology in a general experimental design setting, see Montgomery [1976].

The 2^k factorial series has a projection property useful in factor screening. For example, consider the 2^3 design in Figure 1. If factor A is negligible, we can collapse the 8 runs from the 2^3 design in factors A, B, and C into two replicates of a 2^2 in factors B and C. In general, if we have a single replicate of a 2^k and h(<k) factors can be dropped because they seem negligible, then the remaining data will always correspond to 2^h replicates of a full factorial in the remaining k-h factors. These replicated design points can be used to obtain an estimate of error.

2-2. The 2^{k-p} Fractional Factorial Design

2-2.1 General Results

If the experimenter can reasonably assume that certain high-order interactions are negligible, then information on main effects and low-order interactions may be obtained by running only a fraction of the complete 2^k design. For a general introduction to the construction and elementary properties of these designs refer to Montgomery [1976, ch. 10] or Box and Hunter [1961]. In a 2^{k-p} fractional factorial design, only a fraction of the 2^k treatment combinations are actually run. The designs discussed in this section are regular fractions, that is, estimates of the effects are orthogonal. The effects may be estimated by Yates' algorithm (John [1976], Daniel [1977], Montgomery [1976]) or by generating the contrast for any factor using the table of + and - signs for that design.

There are several methods of constructing these designs. One method of constructing a 2^{k-p} fractional factorial design is to select p independent generators (no chosen generator is a generalized interaction of the others), constructing the 2^p blocks associated with those generators, and then selecting one block as the fractional design. The defining relation for the design consists of the p generators initially chosen and their $2^p - p - 1$ generalized interactions. The alias structure may be found by multiplying each effect modulus 2 by the defining relation. Care should be exercised in choosing the generators so that effects of potential interest are not aliased with each other. Each effect has $2^p - 1$ aliases.

A second method of design construction is to consider the 2^{k-p} design as a full factorial in $h = k-p$ factors. Then the table of + and - signs for the full 2^k design is written down, and the

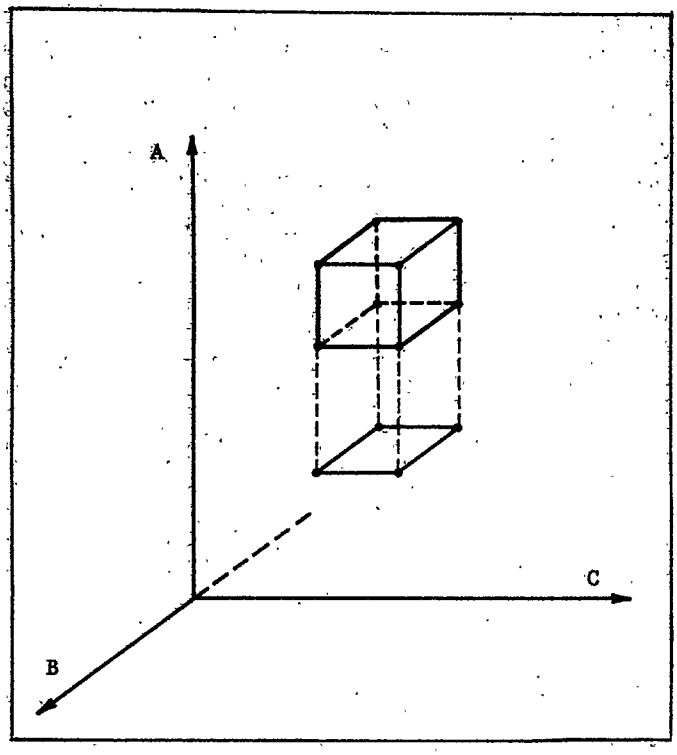

Figure 1. Projection of a 2^3 into a 2^2 Design in the Factors B and C

additional p factors added by equating their factor levels with the products of certain factor levels in the full 2^k.

The 2^{k-p} fractional factorial design has the projection property noted previously for the full 2^k design. In general, a 2^{k-p} fractional factorial design can be projected into either a full factorial or a replicated fractional factorial in some subset of r = k-p of the original factors. This is particularly useful in screening experiments, when we suspect at the outset of the experiment that most of the original factors will have small effects. For example, the 2^{6-2} fractional factorial will collapse to a single replicate of a 2^4 design in any subset of four factors that is not a word in the defining relation. It will also collapse to a replicated one-half fraction of a 2^4 in any subset of four factors that is a word in the defining relation. Thus, a 2^{6-2} design with I = ABCE = ACDF = BDEF becomes two replicates of a 2^{4-1} in the factors ABCE, ACDF, and BDEF, since these are the words in the defining relation. There are 12 other combinations of the six factors, such as ABCD, ABCF, etc., for which the design projects to a single replicate of the 2^4. This design will also collapse to two replicates of a 2^3 in any subset of three of the six factors or four replicates of a 2^2 in any subset of two factors.

2-2.2 Resolution III Designs

It is useful to classify 2^{k-p} fractional factorial designs according to their resolution. The system is as follows:

(i) Resolution III Designs. These are designs in which no main effect is aliased with any other main effect, but main effects are aliased with two-factor interactions and two-factor interactions are aliased with each other.

(ii) Resolution IV Designs. These are designs in which no main effect is aliased with any other main effect or two-factor interaction, but two-factor interactions are aliased with other. The 2^{4-1} design with I = ABCD is of resolution IV.

(iii) Resolution V Designs. These are designs in which no main effect or two-factor interaction is aliased with any other main effect or two-factor interaction, but two-factor interactions are aliased with three-factor interactions. A 2^{5-1} design with I = ABCDE is a resolution V.

In general, the resolution of a design is equal to the smallest number of letters in any word in the defining relation. Resolution III and IV designs are particularly useful in factor screening studies.

We may construct resolution III designs for investigating up to k = N - 1 factors in N runs, where N is a multiple of 4. Designs in which N is a power of 2 can be constructed by the methods presented previously. Of particular importance are designs requiring 4 runs for up to 3 factors, 8 runs for up to 7 factors, 16 runs for up to 15 factors, and 32 runs for up to 31 factors. If k = N - 1 the fractional factorial design is said to be saturated. Designs for intermediate number of factors are produced by dropping columns from the saturated fraction.

By combining fractional factorial designs in which certain signs are switched, we can systematically isolate effects of potential interest. The alias structure for any fraction with the signs for one or more factors reversed is obtained by making changes of sign on the appropriate factors in the alias structure of the original fraction.

The designs due to Plackett and Burman [1946] are also two-level Resolution III fractional factorials. These designs can be used for studying k = N - 1 variables in N runs, where N is a multiple of 4. If N is a power of 2, these designs are identical to those presented earlier in this section. However, for N = 12, 20, 24, 28, and 36 the Plackett-Burman designs are frequently useful. The alias structure of these designs is complex. In general, all two-factor interactions not involving factor Q (say) are aliased with the estimate of Q. For example, in the 11 factor Plackett-Burman plan, each main effect is aliased with 45 two-factor interactions, and each two-factor interaction appears in 9 of the 11 estimates of main effects. This is somewhat less troublesome if fewer than 11 factors

are considered. Furthermore, the two-factor interactions could possibly be untangled by adding a second fraction with all signs reversed, provided that only a few of these effects were large.

EXAMPLE 1. We shall now illustrate some of these ideas with an example. The problem setting is inventory control, and we wish to determine the effect of various parameters on the average annual cost. We note that simulation methods are not required for this problem, as there are analytical models that can be used to describe the system. However, the problem has been kept simple deliberately to illustrate the experimental methods.

There are three items in the inventory. These items are military belts, such as used in jeans and other casual apparel. Item 1 is hardware, item 2 is dyed webbing, and item 3 is natural webbing. The following quantities are fixed:

	Item 1	Item 2	Item 3
Annual Demand D	500,000 doz.	300,000 doz.	200,000 doz.
Demand during a Lead time $X \sim N(\mu,\sigma^2)$	$\mu_1 = 20{,}000$ $\sigma_1 = 3{,}000$	$\mu_2 = 6{,}000$ $\sigma_2 = 900$	$\mu_3 = 4{,}000$ $\sigma_3 = 600$
Lead Time τ	2 weeks	1 week	1 week
Fixed Cost A	\$35/order	\$15/order	\$15/order
Unit Var. Cost C	\$6.25/doz.	\$3.10/doz.	\$2.80/doz.
Carrying Cost h	\$.20	\$.28	\$.28
Cost/unit short π	*	\$.40	\$.40

The following variables represent parameters that we would like to investigate to learn their effect on the system:

Variable	Level	Item 1	Item 2	Item 3
Order quantity Q	1	10,000	4,000	3,000
	2	20,000	8,000	6,500
Reorder point r	1	17,000	5,000	3,500
	2	35,000	11,000	7,000
*Cost/unit shbrt π	1	.3		
	2	.5		

Note that there are seven factors, each at two levels. A 2^{7-4}_{III} design is run, using the high and low levels of these factors shown above. Let factors A, B, and C denote the order quantities for items 1, 2, and 3; D, E, and F denote the reorder points for items 1, 2, and 3; and G denote the shortage cost for item 1. From the 2^{7-4}_{III} design, we obtain the following:

Treatment Combination	Response \$ X1000	Effect + Aliases	Estimate
def	4,626	A + BD + CE + FG	-65
afg	4,693	B + AD + CF + EG	50
beg	4,718	D + AB + CG + EF	-180
abd	4,655	C + AE + BF + DG	-66
cdg	4,662	E + AC + BG + DF	-72
ace	4,653	F + BC + AG + DE	-58
bcf	4,685	G + CD + BE + AF	80
abcdefg	4,626		

Obviously, the effect of D (and its aliases) is large. Since this is the only large effect, we might stop and conclude that over the range of variation, that only item 1's reorder point seriously affects the system. However, to be more certain of these results, we run a second 2^{7-4}_{III} fraction from the same family, obtained by switching signs on all 8 treatment combinations. This gives

Treatment Combination	Response \$ X1000	Effect + Aliases	Estimate
abcg	4,683		
bcde	4,632	-A + BD + CE + FG	66
acdf	4,656	-B + AD + CF + EG	114
cefg	4,704	-D + AB + CG + EF	182
abef	4,647	-C + AE + BF + DG	-32
bdfg	4,640	-E + AC + BG + DF	72
adeg	4,640	-F + BC + AG + DE	24
(1)	4,716	-G + CD + BE + AF	-16

Combining the results from the two fractions, we obtain

i	From $\frac{1}{2}(\ell_i + \ell_i^*)$	From $\frac{1}{2}(\ell_i - \ell_i^*)$
A	BD + CE + FG = 1	A = -65
B	AD + CF + EG = 82	B = -32
C	AE + BF + DG = -49	C = -17
D	AB + CG + EF = 1	D = -181
E	AC + BG + DF = 0	E = -72
F	BC + AG + DE = -17	F = -41
G	CD + BE + AF = 32	G = 48

Clearly the main effect of D is large. Since the effect of D is over twice as large as the next largest effect, we are tempted to conclude that it is the only significant factor. This is confirmed by viewing the normal probability plot of the estimates of the effects, Figure 2. Point 1 on this plot is D. It is significantly off the straight line formed by the other effects.

2-2.3 Resolution IV Designs

A 2^{k-p} fractional factorial design is of resolution IV if main effects are clear of two-factor interactions and some two-factor interactions are aliased with each other. Thus, if three-factor and higher interactions are suppressed, main effects may be estimated directly in a 2^{k-p}_{IV} design. Note that the two combined fractions of the 2^{7-4}_{III} design in Example 1 is a 2^{7-3}_{IV} design.

Any 2^{k-p}_{IV} design must contain at least 2k runs. Resolution IV designs that contain exactly 2k runs are called minimal designs. Resolution IV designs may be obtained from resolution III designs by the process of fold over. To fold over a 2^{k-p}_{III} design simply add to the original fraction a second fraction with all signs reversed. Then the plus signs in the identity column I in the first fraction are switched in the second fraction, and a $(k+1)^{st}$ factor associated with this column. The result is a 2^{k+1-p}_{IV} fractional factorial design.

Any resolution IV design will contain a 2^3 complete factorial design. That is, it will provide r replicates of a 2^7 design in any 3 of the original factors, provided the design contains $r2^3$ points. For example a 2^{8-4}_{IV} plan provides two replicates of a 2^3 in any subset of 3 of the original 8 factors. This concept often has important applications in screening.

2-3. Irregular Fractions of the 2^k Design

There are some multifactor screening situations in which higher saturation of the design than can be accomplished with regular fractions is justified. This would be the case when computer runs are very time-consuming or expensive. In these situations, certain irregular fractional factorial designs may be useful. Often in these designs, the experimenter will only be able to estimate certain parameters in the model and will have few remaining degrees of freedom. Furthermore, the estimates of the effects will generally be nonorthogonal.

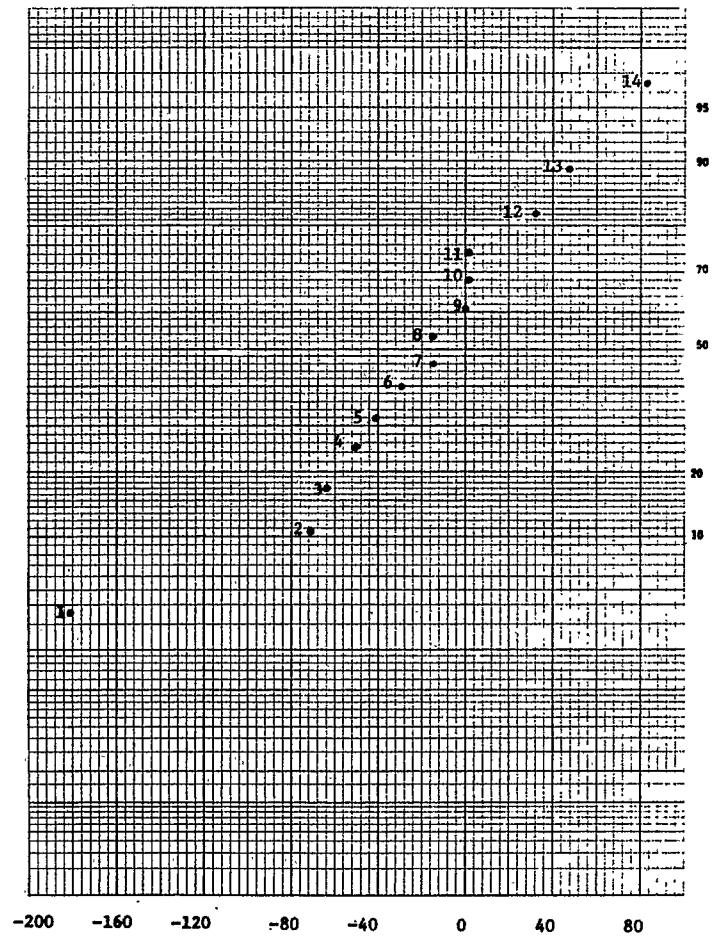

Figure 2. Normal Probability Plot, Example 1.

The simplest irregular fractions result from augmentation of a balanced 2^{k-p} fraction. One may view the process of combining fractions from the same family in the 2^{k-p}_{III} series as augmentation designs, where the augmented set is as large as the initial set. The methods presented here are based on smaller augmented sets, usually 1, 2, 4, or 8 runs, added with the objective of estimating two-factor interactions.

As an elementary example, consider the 2^{3-1}_{III} design. If only the A effect is large, then an estimate of the A effect clear of the BC interaction can be obtained with only one additional run. Thus if I = -ABC, the runs made are (1), ab, ac, and bc. Now consider observation a. Since E(a) = μ + A - B - C - AB - AC + BC, we have, if B = C = AB = AC = 0,

$$E(a) = \mu + A + BC \quad .$$

If we have an estimate $\hat{\mu}$ from the original fraction, then A + BC is estimated by $\ell^* = a - \mu$. We can estimate A - BC directly from the first fraction as $\ell = -(1) + ab + ac - bc$. Then $\ell^* + \ell$ estimates A and $\ell^* - \ell$ estimates BC.

Similar augmentation schemes can be derived for most other designs in the 2^{k-p} series, either to separate a single two-factor interaction, a pair of two-factor interactions, or four such interactions. Daniel [1962] is the basic reference in this area. Addelman [1969] discusses the same problem, in more detail than Daniel [1962], but with less adaptation of results to special cases.

Three-quarter replicates of the 2^{k-p} series are often highly useful. These designs may be viewed as constructed by either omitting a quarter-fraction from the full 2^k or by adding a quarter-fraction to a one-half fraction. A good survey of these designs is in John [1971].

Addelman and Kempthorne [1961] have developed a series of orthogonal main effect plans. These designs are useful in cases where only main effects are of interest. In many cases factors with either 2 or 3 levels can be considered. Much other work has been done on irregular fractions of the $2^k 3^m$ series. Margolin [1968] [1972] has done much of the work in this area. Webb [1965] [1971] has also developed very compact mixed fractional factorials from this series, involving 20 or fewer runs. These plans all have very heavy 2 factor interaction aliasing. Of related interest is Webb [1968].

2-4. Supersaturated Plans

These are two-level designs devised by Booth and Cox [1962]. In these designs, each of k factors appears at the high and low levels N/2 times, where $N \leq k$. We assume that N is even. Clearly not all estimates of the effects can be orthogonal, since $N \leq k$. Booth and Cox [1962] generated these designs to obtain "near-orthogonality" by using the design criterion

$$\min(\max_{i \neq j} |\underline{d}_i' \underline{d}_j|)$$

where $\underline{d}_i$ is a row vector denoting the levels of factor i. The vector $\underline{d}_i$ will consist of N/2 +1's and N/2 -1's. Booth and Cox [1962] tabulate designs for N = 12 and $k \leq 16$, 20, 24; N = 18 and $k \leq 24$, 30, 36; and N = 24 and $k \leq 30$. They describe an algorithm for generating other designs, although the procedure may be very inefficient.

A major disadvantage of a supersaturated design is that if following the initial experiment, several effects seem to be potentially active, there is no simple additional set of experiments that can be run to isolate the factors of interest. Moreover, the aliasing that is present in the contrasts from a supersaturated design is very heavy and irregular, and this will frequently present a confusing picture to the analyst.

2-5. Group Screening Designs

2-5.1 General Approach

These designs are intended for use in situations where the following conditions apply:

1. The number of factors k is relatively large,
2. All factors have the same prior probability of being active,
3. There are no interactions between active factors,
4. The direction of all effects is known,
5. The errors associated with the observations are NID(0,σ^2).

A group screening design is conducted by forming the original k factors into g groups. Then each group is considered as a single factor and investigated through a design such as the 2^{g-p}. If a group-factor is negligible, than all factors within that group are considered insignificant. Group factors that exhibit significant effects are then divided into smaller groups for subsequent experimentation.

2-5.2 Two-stage Group Screening

The k factors will be divided into g groups. Watson [1961] originally suggested that all groups be of the same size, although this assumption is unnecessary. Because the direction of effects is known, we can label the high level of each factor as the level producing the largest response. The upper level of a group factor consists of running each factor in the group at the high level. If this arrangement is not followed, some factor effects may cancel.

Watson [1961] derives the optimum group size to be

$$f^* = [(1-\alpha_1)p]^{-\frac{1}{2}} \qquad (11)$$

where p is an estimate of the fraction of active factors and α_1 is the significance level used for the first-stage statistical analysis. This formula attempts to minimize the total number of runs required in both stages. It also implies that groups will be of equal size. If we have no prior estimate of p, or if the direction of some effects are not known, then (11) is invalid.

Generally, we would expect p to vary from factor to factor. That is, we would have considerable knowledge about some factors, and little knowledge about others. Note that as p increases, the optimum group size decreases. Therefore, it would seem reasonable to use groups of different sizes, depending on our knowledge of p for each factor. Factors that we strongly suspect are significant would be run in very small groups (perhaps of size 1). Furthermore, factors for which we do not know the direction of the effect could be tested in groups of size 1 to prevent the cancellation effect.

As an example of group screening, suppose we have 17 factors. The direction of factor 1 is unknown, and we are almost positive that factor 2 is active. The possible directions of the other is factors are known. Therefore, a logical arrangement of the groups would be:

Group Factors	Original Factors
A	1
B	2
C	3,4,5,6,7
D	8,9,10,11,12
E	13,14,15,16,17 .

These five factors could be investigated in the first stage using a 2^{5-2}_{III} design (8 runs). This would permit investigation of all main group effects, but these effects would be aliased with the two-factor interactions of the group effects. If we wanted to use 16 runs, the 2^{5-1}_{IV} design would allow estimation of all main effects and two-factor interactions of the group factors.

If the assumption of no active two-factor interactions between the original factors holds, then the factors may be formed into groups on an arbitrary basis. However, some choices of grouping arrangements will lead to more easily interpreted results, or to smaller sets of active factors to be investigated at the second stage. Sometimes we can use our knowledge of the problem to form the groups. For example, we might place all similar factors in the same groups. Thus, if we are simulating an inventory system, all reorder quantities could form one group, all reorder levels a second group, etc. Generally, a significant two-factor interaction (say AB) biases the estimates of a third factor (say C) if and only if all three factors belong to separate group factors. Therefore, if we suspect that some two-factor interactions are active, then all the factors involved in those interactions should be placed in the same group. For a proof of this result, see Kleijnen [1975a, b].

In the second stage of a group screening design, in addition to investigating the set of potentially active factors, we must also choose levels for the negligible factors identified in the first stage. The linear model can be written as

$$\underline{y} = X_1\underline{\beta}_1 + X_2\underline{\beta}_2 + \underline{\varepsilon}$$

where $\underline{\beta}_1$ contains the set of potentially active factors and $\underline{\beta}_2$ contains the set of factors tentatively identified as negligible at the first stage. The matrix X_1 consists of the levels assigned to the active factors in the second stage and X_2 consists of the factors assigned to the negligible factors. Now, the expected value of the least squares estimate of $\underline{\beta}_1$, $\hat{\underline{\beta}}_1 = (X_1'X_1)^{-1}X_1'\underline{y}$, is

$$E(\hat{\underline{\beta}}_1) = \underline{\beta}_1 + (X_1'X_1)^{-1}X_1'X_2\underline{\beta}_2 \ .$$

Clearly, if all the factors thought to be insignificant from stage 1 really are insignificant, then $\underline{\beta}_2 = \underline{0}$ and $\hat{\underline{\beta}}_1$ is an unbiased estimator of $\underline{\beta}_1$. However, if one or more of these effects is active, then $\underline{\beta}_2 \neq 0$ and $\hat{\underline{\beta}}_1$ is a biased estimator of $\underline{\beta}_1$.

The extent of the bias in $\hat{\underline{\beta}}_1$ is given by the alias matrix $A = (X_1'X_1)^{-1}X_1'X_2$. This may be controlled by the choice of factor levels for the variables in X_2. Assuming that two-level factors are employed, then if all levels in X_2 are identical (say +1, the high level) then the coefficients in $\underline{\beta}_2$ will bias only the intercept or overall mean term in $\underline{\beta}_1$. No other effects in $\underline{\beta}_1$ will be biased by factors in $\underline{\beta}_2$.

To prove this, suppose that X_1 is n×p $\underline{\beta}_1$ is p×1, X_2 is n×r, and $\underline{\beta}_2$ is r×1. If the second-stage design is a 2^k or an orthogonal fraction of the 2^k, then $(X_1'X_1)^{-1} = (1/n)\ I_p$. Furthermore, if all of the negligible factors in X_2 are set at their high levels, then X_2 is a n×r matrix of 1's. Now X_1 is an n×p matrix, the first column of which consists of 1's (to account for the overall mean μ) and the remaining p-1 columns consist of the +1 and -1 levels from the orthogonal 2^k design. Therefore, $X_1'X_2$ is a p×r matrix, the first row of which consists of n's, all the remaining elements are all zero. Therefore,

$$(X_1'X_2)^{-1}X_1'X_2 = (1/n)\ I_p\begin{bmatrix}\underline{n'}\\ \underline{0}\end{bmatrix} = \begin{bmatrix}\underline{1'}\\ \underline{0}\end{bmatrix}$$

and the alias structure is

$$E(\hat{\beta}_0) = \beta_0 + \sum_{i=p}^{r-1}\beta_i$$

$$E(\hat{\beta}_i) = \beta_i,\ i=1,2,\ldots,p-1 \ .$$

Thus the r elements in $\underline{\beta}_2$ bias only the estimate of the intercept $\hat{\beta}_0$. Strictly speaking, all of the r negligible factors do not all have to be here at the high level. However, each factor must be held at the same level throughout the experiment.

EXAMPLE 2. Consider the inventory problem in Example 1. We add a fourth item to the inventory, with parameters D = 350,000, σ_4 = 2,000, A = \$25, C = \$4.30, n = \$0.45, and π = \$0.50. There are 13 factors of interest, Q_1, r_1, μ_1, π_1, Q_2, r_2, μ_2, Q_3, μ_3, Q_4, r_4, and μ_4. We will arrange these factors in 4 groups, according to item, as follows:

Group Factor	Original Factor
A	Q_1, r_1, μ_1, π_1
B	Q_2, r_2, μ_2
C	Q_3, r_3, μ_3
C	Q_4, r_4, μ_4

A 2^{4-1} design is used to analyze these four group factors. The results are summarized below:

Treatment Combination	Response	Effect	Estimate
(1)	6207		
ad	6164	A + BCD	-180
bd	6183	B + ACD	-116
ab	6134	AB + CD	-10
cd	6210	C + ABD	6
ac	6168	AC + BD	4
bc	6181	BC + AD	-8
abcd	6135	D + ABC	2

Note that the two largest effects are A and B (and other aliases). Group factors C and D, and consequently the factors for item 3 and 4 are negligible. Therefore, following the initial 8 rows, we have reduced the set of potentially active factors from 13 to 7. The 7 remaining factors, Q, r_1, μ_1, π_1, Q_2, r_2, and μ_2 could be investigated using a 2^{7-4}_{III} or 2^{7-3}_{IV} plan, such as illustrated earlier.

2-5.3 Group Screening With More Than Two Stages

Patel [1962] and Li [1962] have generalized Watson's results to more than two stages. Patel showed that the total number of runs is minimized if we choose the number of groups according to

$$g_1 \simeq kp^{n/(n+1)}$$

$$g_2 = g_3 = \ldots = g_n = g_{n+1} \simeq p^{-1/(n+1)},$$

where g_i is the number of groups into which each of the groups at stage i-1 is split. He also notes that an n-stage procedure is preferable to an n-1 stage procedure if

$$p < [1-(1/n)]^{n(n-1)}.$$

Group sizes decrease geometrically with parameter $p^{1/(n+1)}$. Note that if we suspect that if more than one-fourth of the factors are active (p > .25), then the optimum number of stages is one. If between one-twelfth and one-fourth of the factors are active, then two stages should be used. Similarly, a three-stage procedure would be used if between one-thirtieth and one-twelfth of the factors are active. Clearly, these designs will be useful only in situations where p (the ratio of active to total factors) is thought to be very small.

3. VARIANCE REDUCTION CONSIDERATIONS IN FACTOR SCREENING

An important consideration in the design of a computer simulation experiment is the incorporation of variance reduction methods into the design. Two common variance reduction methods are the use of common pseudorandom numbers and antithetic pseudorandom numbers for different points in the design. These methods have application to factor screening. Early work on this problem was by Fishman [1974]. Recently, a comprehensive treatment of the subject was published by Schruben and Margolin [1978].

We assume that when common random number streams are used at two design points, the two output statistics exhibit positive correlation, and when antithetic random number streams are used at any two points, negative correlation between outputs is induced. These assumptions are, of course, not always met in practice, but they are satisfied relatively often, as has been confirmed by numerous investigations (see Kleijnen [1975a], pp. 197-198).

Two possible estimation methods can be used, ordinary least squares (OLS), or weighted least squares (WLS). If V is the correlation matrix induced on the responses, then the covariance matrices for these estimators are $\text{Cov}(\hat{\underline{\beta}}_{OLS}) = (X'X)^{-1}X'V^{-1}X(X'X)^{-1}$ and $\text{Cov}(\hat{\underline{\beta}}_{WLS}) = (X'V^{-1}X)^{-1}$. A widely used criterion for comparing designs for estimating $\underline{\beta}$ is the determinant of the covariance matrix of the estimator. Designs that minimize this criterion are called D-optimal designs. The determinants of the covariance matrices associated with the OLS and WLS estimators are

$$D_{OLS} = |X'X|^{-2}|X'V^{-1}X|$$

and

$$D_{WLS} = |(X'V^{-1}X)^{-1}| .$$

The WLS estimator has smallest generalized variance among the class of linear unbiased estimators. However, it is often impossible to calculate the WLS estimate because the matrix V is unknown.

There are some situations in which the OLS and WLS estimators are equivalent, and, hence, these two estimators would produce the same covariance matrix. The two estimators are equivalent for the cases of the random number assignment schemes that minimize D_{WLS}. That is, an induced correlation structure that would minimize D_{WLS} is also one for which the estimators $\hat{\beta}_{OLS}$ and $\hat{\beta}_{WLS}$ are identical. Therefore, the OLS estimator can be used.

Schruben and Margolin [1978] propose the following rule. If an experimental design admits orthogonal blocking into two blocks, then if for all points in block 1 we use the same common set of pseudorandom numbers, and for all points in block 2 we use the antithetic set of random numbers, then the OLS estimator of β will have minimum generalized variance. Specifically, this assignment rule will produce an estimator of β_0 that is superior to that obtained by common random numbers, and equivalent in terms of dispersion to common random numbers for estimating the remaining parameters in β. In general, the best results are obtained if the block sizes are the same. The positive and negative correlations induced do not have to be equal.

There are some special results that can be stated for the 2^{k-p} series of designs. If the induced positive and negative correlations are equal in magnitude, then the assignment rule above produces a minimum generalized variance for the class of 2^{k-p} designs assuming that the linear model contains a mean (β_0) plus a subset of $r < 2^{k-p}$ effects. This assignment rule also minimizes the trace of the covariance matrix of $\hat{\beta}$ (that is, the sum of the variances of $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_r$ is minimized).

Occasionally, factor screening experiments will make use of saturated designs. For a saturated design, <u>any</u> induced correlative structure between the observations results in an improvement with respect to the generalized variance criterion over that obtained from independently seeking each design point. Furthermore, the OLS and WLS estimators are equivalent in this case also.

These results have direct application to factor screening. Any 2^k or 2^{k-p} design that is not saturated can be run in two orthogonal blocks by identifying the blocks with the + and - levels of one of the k columns in the design. Thus, only k-1 factors could be investigated.

We now give some illustrations. First consider the 2^{6-2} design. We can run this design in two blocks, say

block 1: (1), aef, bcf, abce, bdef, abd, cde, acdf

block 2: be, abf, cef, ac, af, ade, bcd, abcdef

The treatment combinations in block 1 would be run with one set of common random numbers and those in block 2 would be run with the antithetic set of random numbers.

As a second example, consider the 2^{7-4}_{III} design run in Example 1. Since 7 factors are considered in only 8 runs, this is a saturated fractional factorial. If only this fraction is to be run, any induction of correlation is superior to independent observations, so running all 8 observations with common random number streams would be an appropriate strategy. Now, if any fraction from the same family is added to the original fraction, the new fraction should be run using the antithetic random number stream. Clearly, this is an optimal strategy, since the two fractions together can be viewed as a fold-over design with the random number stream effect taking the levels of the eighth factor (which is + in fraction 1 and - in fraction 2).

As a third example, consider the 2^{6-3}_{III} design obtained by writing down a 2^3 in factors A, B, and C, and generating three additional columns as D = AB, E = AC, and F = BC. This design investigates 6 factors in 8 runs, and since it is not a saturated fraction, we could obtain a minimum generalized variance by decomposing the design into two orthogonal blocks of 4 runs each. Now any nonsaturated Resolution III plan can be run in two blocks by identifying the + and - levels of a single additional variable with the blocks. Thus, in our example, add a seventh factor to the design by setting the signs in that column equal to G = ABC. Consequently, we run treatment combinations def, abd, ace, and bcf in block 1 (-) with a common set of random numbers, and treatment combinations af, be, cd, and abcdef in block 2 with the antithetic set of random numbers.

Now suppose upon examining the estimates of the effects from this fraction, it is decided to add a second fraction from the same family to separate main effects and two-factor interactions. The appropriate second fraction found by setting D = -AB, E = -AC, and F = -BC. In this new fraction, block 1 would consist of bcde, acdf, abef, and (1). These runs would be made with the same set of random numbers used in block 1 from the first fraction. Block 2 in the new fraction would consist of abc, cef, bdf, and ade. These runs would be made with the antithetic stream of random numbers used in block 2 in the original fraction. It is easy to verify that the final design is a 2^{6-2}_{IV} plan, with generation I=BCDE=ACDF=ABEF. The estimators from the combined design have minimum generalized variance.

4. EVALUATION AND CHOICE OF SCREENING DESIGNS

The 2^{k-p} fractional factorial design has many advantages in factor screening. If we can afford N runs, where N is a power of 2, Resolution III plans can be derived that incorporate up to N-1 factors. These plans require the experimenter to assume that two-factor and higher interactions are negligible. However, the assumptions regarding interactions can, to some extent, be checked by combining the original 2^{k-p}_{III} design with a second fraction from the same family. If the experimenter can afford up to N=2k runs, the 2^{k-p} Resolution III and IV plans are highly recommended. The Plackett-Burman plans, also of

Resolution III, are not generally recommended for factor screening unless the analyst knows in advance that all but a few two-factor interactions are negligible. The heavy aliasing of main effects and two-factor interactions is an undesirable property of these designs.

The supersaturated plans of Booth and Cox, like the Plackett-Burman designs, assume that only main effects are active. If this assumption is false, then the alias structure generated by a supersaturated design would be extremely difficult to untangle. The group screening methods of Watson and Patel are recommended instead. This approach would seem to have the economic efficiency required in simulation, without the overly-restrictive assumptions regarding interactions. For the vast majority of screening problems, either two or three stages will be sufficient. Once groups of factors are formed, it is recommended that 2^{k-p} fractional factorials be used to investigate the group factors.

REFERENCES

1. Addelman, S. [1969], "Sequences of Two-Level Fractional Factorial Plans," Technometrics, Vol. 11, No. 3, pp. 477-509.

2. Addelman, S. and O. Kempthorne [1961], Orthogonal Main-Effect Plans, ASTIA Arlington Hall Station.

3. Booth, K. H. Y. and Cox, D. R. [1962], "Some Systematic Supersaturated Designs," Technometrics, Vol. 4, No. 4, pp. 489-495.

4. Box, G. E. P. and Hunter, J. S. [1961], "The 2^{k-p} Fractional Factorial Designs," Parts I and II. Technometrics, Vol. 3, pp. 311-352, pp. 449-458.

5. Budne, T. A. [1959], "The Application of Random Balance Designs," Technometrics, Vol. 1, No. 2, pp. 139-155.

6. Burdick, D. S. and T. H. Naylor [1966], "Design of Computer Simulation Experiments for Industrial Systems," Communications of the ACM, Vol. 9, No. 8.

7. Cochran, W. G. and G. M. Cox [1957], Experimental Designs, Second Edition, John Wiley & Sons, New York.

8. Daniel, C. [1962], "Sequences of Fractional Replicates in the 2^{p-q} Series," Journal of the American Statistical Association, Vol. 58, No. 246, pp. 403-429.

9. Daniel, C. [1973], "One at a Time Plans," Journal of the American Statistical Association, Vol. 68, No. 285, pp. 353-360.

10. Daniel, C. [1976], Applications of Statistics to Industrial Experimentation, John Wiley & Sons, New York.

11. Davies, O. L. [1956], The Design and Analysis of Industrial Experiments, Hafner, New York.

12. Fishman, G. S. [1973], Concepts and Methods in Discrete Event Digital Simulation, John Wiley & Sons, New York.

13. Fishman, G. S. [1974], "Correlated Simulation Experiments," Simulation, Vol. 23, pp. 177-180.

14. Heikes, R. G., D. C. Montgomery, and R. L. Rardin [1976], "Using Common Random Numbers in Simulation Experiments - An Approach to Statistical Analysis," Simulation, Vol. 26, No. 3, pp. 81-85.

15. Hicks, C. R. [1973], Fundamental Concepts in the Design of Experiments, Second Edition, Holt, Rinehart, and Winston, New York.

16. Hunter, J. S. [1964], "Sequential Factorial Estimation," Technometrics, Vol. 6, No. 1, pp. 41-55.

17. Hunter, J. S. and T. H. Naylor [1970], "Experimental Designs for Computer Simulation Experiments," Management Science, Vol. 16, No. 7.

18. Ingall, E. J. [1972], "On Experimental Designs for Computer Simulation Experiments," Management Science, Vol. 18, pp. 384-388.

19. Jacoby, J. E. and S. Harrison [1962], "Multivariable Experimentation and Simulation Models," Naval Research Logistics Quarterly, Vol. 9, pp. 121-136.

20. John, P. M. W. [1971], Statistical Design and Analysis of Experiments, McMillan, New York.

21. Kleijnen, J. P. C. [1975a], Statistical Techniques in Simulation, Parts I and II, Marcel Dekker, New York.

22. Kleijnen, J. P. C. [1975b], "Screening Designs for Poly-Factor Experimentations," Technometrics, Vol. 17, No. 4, pp. 487-493.

23. Kleijnen, J. P. C. [1977], "Design and Analysis of Simulations: Practical Statistical Considerations," Simulation, Vol. 27, No. 9, pp. 81-90.

24. Li, C. H. [1962], "A Sequential Method for Screening Experimental Variables," Journal of the American Statistical Association, Vol. 57, pp. 455-457.

25. Margolin, B. H. [1968], "Orthogonal Main Effect $2^n 3^m$ Designs and Two-Factor Interaction Aliasing," Technometrics, Vol. 10, No. 3, pp. 559-573.

26. Margolin, B. H. [1972], "Non-Orthogonal Main-Effect Designs," Journal of the Royal Statistical Society, B, Vol. 34, No. 3, pp. 431-440.

27. Montgomery, D. C. [1979], Methods for Factor Screening in Computer Simulation Experiments, Technical Report, Office of Naval Research, Alexandria, Virginia.

28. Montgomery, D. C. [1976], Design and Analysis of Experiments, John Wiley & Sons, New York.

29. Montgomery, D. C. and D. M. Evans [1975], "Second Order Response Surface Designs in Computer Simulation," Simulation, Vol. 25, No. 6, pp. 169-178.

30. Naylor, T. H. (ed.) [1968], The Design of Computer Simulation Experiments, Duke University Press, Durham.

31. Nolan, R. L. and M. G. Sovereign [1972], "A Recursive Optimization and Simulation Approach to Analysis with an Application to Transportation Systems," Management Science, Vol. 18, No. 12, pp. 676-690.

32. Patel, M. S. [1962], "Group Screening with More than Two Stages," Technometrics, Vol. 4, No. 2, pp. 209-217.

33. Plackett, R. L. and Burman, J. P. [1946], "The Design of Optimum Multifactorial Experiments," Biometrica, Vol. 33, pp. 305-325.

34. Satterthwaite, F. D. [1959], "Random Balance Experimentation," Technometrics, Vol. 1, No. 2, p. 111.

35. Schruben, L. W. and B. H. Margolin [1978], "Pseudorandom Number Assignment in Statistically Designed Simulation and Distribution Sampling Experiments," Journal of the American Statistical Association, Vol. 73, No. 363, pp. 504-520.

36. Watson, G. S. [1961], "A Study of the Group Screening Method," Technometrics, Vol. 3, No. 3, pp. 371-388.

37. Webb, S. R. [1965], Design, Testing and Estimation in Complex Experimentation, Office of Aerospace Research, ARL 65-116.

38. Webb, S. R. [1968], "Saturated Sequential Factorial Designs," Technometrics, Vol. 10, No. 3, pp. 535-550.

39. Webb, S. R. [1971], "Small Incomplete Factorial Experimental Designs for Two-Level and Three-Level Factors," Technometrics, Vol. 13, No. 2, pp. 243-256.

40. Williams, D. R. and D. L. Weeks [1974], "A Technique for Designating and Augmenting Computer Simulation Experiments," Management Science, Vol. 20, No. 10.