Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

FINDING IMPORTANT INDEPENDENT VARIABLES THROUGH SCREENING DESIGNS: A COMPARISON OF METHODS

Linda Trocine
Linda C. Malone

Department of Industrial Engineering
and Management Systems
P.O. Box 162450
University of Central Florida
Orlando, FL 32816-2450, U.S.A.

ABSTRACT

Once a simulation model is developed, designed experiments may be employed to efficiently optimize the system. Designed experiments are used on "real" production systems as well. The first step is to screen for important independent variables. Several screening methods are compared and contrasted in terms of efficiency, effectiveness, and robustness. These screening methods range from the classical factorial designs and two-stage group screening to new, more novel designs including sequential bifurcation (SB) and iterated fractional factorial designs (IFFD). Conditions for the use of the methods are provided along with references on how to use them.

1 INTRODUCTION

The first step in finding the important independent variables in a simulation model or production system is to conduct a screening experiment. These important variables can later be used to optimize the model. Simulation models typically represent complex and stochastic systems. Experimentation on these systems is assumed to be time consuming and can be expensive in terms of computation. Minimizing the number of experiments while maximizing information is the ultimate goal.

The parsimony principle says that some of the factors are important while others are not (Myers and Montgomery 1995, Kleijnen 1987). In other circles, this principle is known as the Pareto principle or the "80-20" rule. Said another way, a few variables are responsible for most of the effect on the response while most variables contribute little. Screening experiments aim to find these few variables that affect the response the most.

2 AVAILABLE METHODS

Response Surface Methodology: Process and Product Optimization Using Designed Experiments by Myers and Montgomery (1995) is a leading text on the design of experiments (DOE) and response surface methodology (RSM). The screening experiments emphasized are classical factorial designs and fractional factorial designs. These designs work well for screening a small number of variables (i.e., fewer than 20). Using them can detect main effects and some interaction effects depending on the resolution. They have many desirable features and can be augmented easily to optimize the response in subsequent experimentation.

One current method for screening large numbers of variables is group screening (Kleijnen, 1987). The analyst makes educated choices in grouping variables together and then performs a fractional factorial experiment on the groups. If a group is significant, subgroups or individual variables within the group are further screened in the same way. This can work well if the variables within a group have the same sign of effects and if there is no cancellation of effects due to interaction.

Myers and Montgomery (1995) also mention one-at-a time designs, fold-over designs, Plackett-Burman, and orthogonal designs. These designs are not as efficient nor as effective as the classical designs and will not be discussed here.

A more recent technique is the edge design discussed in Elster and Neumaier (1996). Edge designs look for significant differences in pairs of experimental runs where the settings differ for only one factor. An illustration of an edge design in three dimensions is shown in Figure 1 along with its design in Table 1. It appears from their discussion that exactly $2K$ runs are needed. Unfortunately they do not describe how to generate the designs easily.

Table 1: Edge Design in Three Variables

Run	A	B	C
1	+	+	-
2	+	-	+
3	-	-	+
4	-	+	-
5	+	+	+
6	-	-	-

Elster and Neumaier discuss that the difference in the responses for the pairs is due to noise unless there is a significant factor involved. They claim that these designs can detect nonlinearity (in the probability plots) better than classical designs. In other words, the significant factors are more obvious.

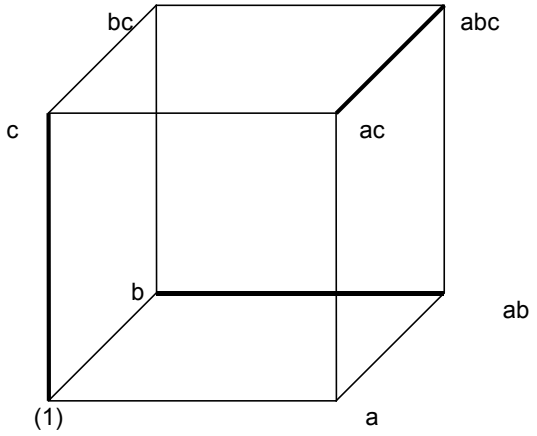

Figure 1: Edges of Three Dimension Design

Recently, Bettonvil and Kleijnen (1996) extended a novel approach introduced by Jacoby and Harrison (1962) known as sequential bifurcation that efficiently and effectively screens large numbers of variables in a restricted setting. In one deterministic example 128 variables were screened and the three most important variables were found in 16 observations. Another example given was a stochastic ecology model where in 144 observations, 15 important effects, main and 2-way, were found from 281 variables. However, the use of sequential bifurcation is limited to quantitative variables, same-sign effects and a monotone response function. These latter two limitations are very restrictive when little is known about the system under study.

Andres and Hajas (1993) introduced another novel approach for screening large numbers of variables called iterated fractional factorial designs (IFFD). Campolongo, Kleijnen, and Andres (2000) describe it as a type of group screening where the variables and their signs are assigned randomly to groups; this is repeated over several iterations. The number of groups is usually either 8 or 16 (some power of 2) and should be slightly larger than the number of important factors. The sign (+1 or –1) is assigned with equal probability (0.5). The total number of experimental runs is dependent on the budget available but is typically 100 to 500 simulation runs. IFFD is insensitive to the total number of initial factors considered, *K*, so long as *K* is a large number (1000 or more). It was unclear what the incremental improvement of this algorithm is for each additional iteration completed.

IFFD can estimate certain interaction and quadratic effects as well (Campolongo, Kleijnen, and Andres, 2000). IFFD requires that very few factors dominate. In a nuclear waste disposal case, 18 significant effects were found from over 3000 variables; this is less than 1% of the total, very few indeed (Andres, 1997). Software is available called SAMPLE2. The software generates the design.

2.1 Issues with Screening

Factorial designs require a large number of experimental runs to screen a small number of variables. A 2^k factorial design requires 2^k experimental points to estimate linear effects of k variables and their interactions. For example, if k=5, then 32 experiments are conducted. All five main effects, all ten 2-way interactions, all ten 3-way interactions, all five 4-way interactions and the 5-way interaction can all be estimated. Practitioners may opt for a fractional factorial design and focus just on finding important main effects and 2-way interactions (resolution V), or just main effects (resolution III) or main and selected 2-way interactions (resolution IV). With k=5, a resolution V design requires 16 experimental points (Myers and Montgomery 1995). The number of runs required for fractional factorial designs is much more realistic but the ability to estimate interaction effects is reduced.

If k=15, a full factorial design would require over 32,000 experiments, impossible to carry out. Instead a 2^{15-8} fraction requires 128 experiments, still a large number. This fraction can estimate main effects and some 2 way interactions (resolution IV).

Another important issue with screening experiments is the possibility of failing to retain significant effects as important (Trocine, 2000). For example, a two-way interaction is shown in Figure 2. This interaction should be included in further analysis because the combination of the settings of the two variables dramatically affects the response. Two possible situations exist where this interaction may be overlooked. First, the design may confound this particular interaction with other effects (such as in a fractional factorial design). The results are combined with other interactions and nothing can be determined about individual interactions within the alias structure. The second situation is where the choice of low and high values effectively cancels the interaction. This is depicted in Figure 3. The upper chart shows how the interaction would appear given a 2^k design while the lower

chart shows a strong interaction when another level of factor A is included. This is a dangerous situation because these interactions may significantly affect the response in the optimum region and if they are excluded during screening, the model will be misleading.

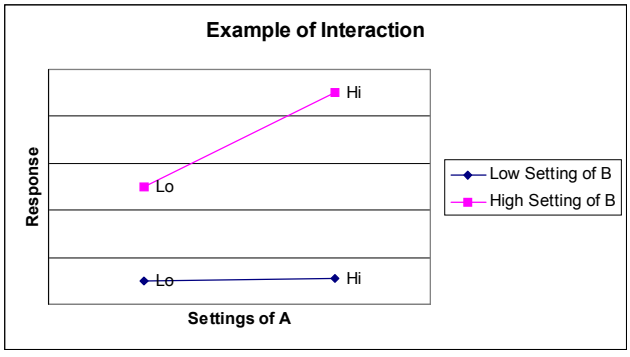

Figure 2: Interaction Example

In the case of group screening, the interaction might be missed because another factor within the group cancels the interaction by having equal and opposing effect on the response. Although this is a violation of the main assumption in group screening that the sign of all effects in a group is the same, many times we may not know enough about the system at the screening stage to be able to assign variables to groups correctly (i.e. satisfying the underlying assumptions for group screening). In general, if the sum of the effects within a group is close to zero, important effects within that group will be missed.

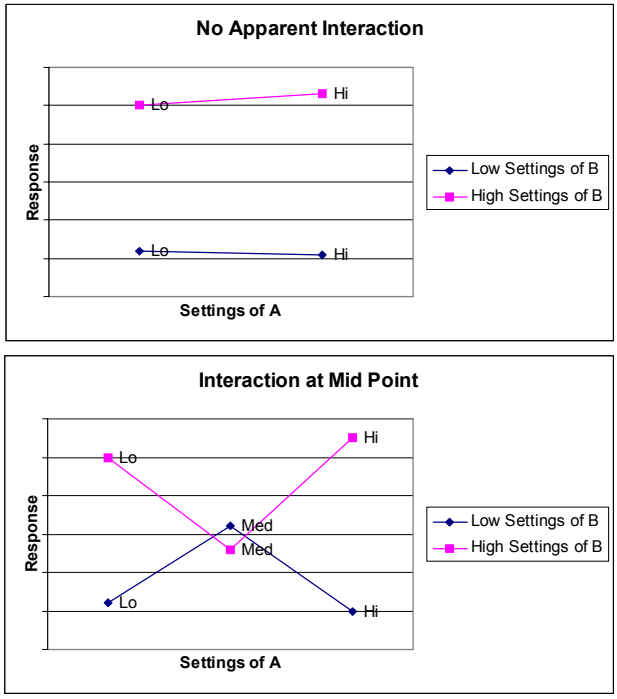

Figure 3: Missed Interaction Effect

An issue with two-stage group screening is that interactions between variables that are in different groups are not measured. Ivanova, Malone and Mollaghasemi (1999) showed that the results of a group screening experiment and a fractional factorial experiment on 17 variables can result in two different sets of important variables. No conclusions could be made in this semiconductor manufacturing process example about what variables were truly important.

2.2 Desirable Screening Design Features

Screening experiments can be evaluated on a number of criteria to more adequately assess these issues. Efficiency is necessary to screen large numbers of variables in a reasonable number of experiments. As shown earlier, classical factorial designs are exponential in the number of variables and are thus inefficient for more than a few variables. Effectiveness is defined as the ability to find the important independent variables regardless of the interactions among the variables. Screening designs should cover both deterministic and stochastic systems, both monotonic and non-monotonic response surfaces, and both small or large numbers of variables. All of this together is referred to as robustness and provides versatility in the application of a design.

Table 2 summarizes the effectiveness of the screening methods discussed here. The first column, "main effects estimated" shows that all methods can be used to estimate main effects. The next column shows whether the method can be used for 2-way interactions. The third column shows whether the method can easily be adapted to measure curvature (by introducing a third level of a variable). The next column, "interactions without main effects", shows that some methods are only able to measure interactions if there is a significant main effect. Finally, the last of the columns shows whether the method is robust to canceling of effects by two effects having opposite signs and making the combined effect appear to be negligible. The note '1' indicates that factorial designs may easily be augmented with middle levels or center points to detect curvature. Table 3 summarizes the efficiency, robustness and issues with the screening methods. The column for number of experiments shows known values in terms of "*K*" the total number of variables in the original problem. The columns under "robustness" show whether the method can be used for large numbers of variables and whether monotonicity is required. A monotone function is either non-decreasing or non-increasing. That is, an increase in the level of a significant effect will always result in an increase in the response in the case of a non-decreasing function. A rough guideline is provided to show the intended number of variables that the method operates upon. Small is usually less than 10 and certainly less than 20. Large is certainly 20 or more but can

include 15 or more. The last column under "issue" shows the requirement for like signs-of-effects. This is not always known in advance. All of the methods discussed here result in unbiased estimators and operate on stochastic systems.

Table 2: Effectiveness of Screening Methods

Existing Screening Experiment Methods					
Design	**Effectiveness in finding significant effects**				
	Main Effect Estimateds	Interaction Effects Estimated	Quadratic Effects Estimated	No Interactions without main effects	Robust to Cancelled Effects
Desired	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
One-at-a-time	Yes	No	No	No	Yes
Full Factorial	Yes	Yes	1	Yes	Yes
Fractional Factorial	Yes	Yes	1	Yes	Yes
Edge Designs	Yes	No	No	No	Yes
Two-Stage Group Screening	Yes	Some	No	Some	No
Sequential Bifurcation	Yes	Some	No	No	No
IFFD	Yes	Some	Yes	No	Yes

Table 3: Efficiency and Robustness of Screening Methods

Existing Screening Experiment Methods				
Design	**Efficiency**	**Robustness**		**Issue**
	Number of experiments needed	Monotonicity required	Number of variables	Sign of Effect Required
Desired	**Small**	**No**	**Large**	**No**
One-at-a-time	K	No	Small	No
Full Factorial	2^k Large	No	Small	No
Fractional Factorial	2^{k-p} Large	No	Small	No
Edge Designs	2K	No	Small	No
Two-Stage Group Screening	Varies	Yes	Medium	Yes
Sequential Bifurcation	O(k log K)	Yes	Large	Yes
IFFD	100-500	No	Large	No

3 CONDITIONS FOR USE

The choice of which screening plan to use depends on the number of variables in the model. Other considerations include how much the practitioner knows about the underlying model (e.g. polynomial model of linear function, exponential model, other nonlinear model) whether the sign of the effects are known or not, how many experiments may be run (the budget) and how much is known about the system or model under study. For very general circumstances, the following decision tree may be used to guide the practitioner in choosing the best design.

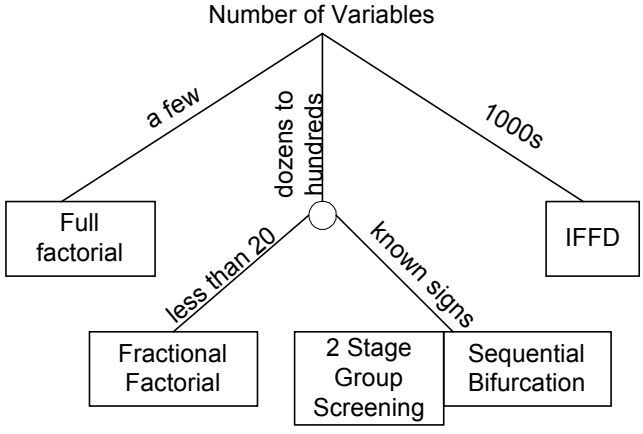

Figure 4: Decision Tree for Selecting a Screening Design

For a few variables, the classical factorial designs are best. For thousands of variables, the only choice is IFFD. For fewer than 20 variables, the fractional factorial is probably the best design. To screen more than 15 or 20 variables, either two-stage group screening or sequential bifurcation can be used cautiously. These are the best screening alternatives for each category.

For more information about sequential bifurcation see Campolongo, Kleijnen, and Andres (2000), Bettonvil and Kleijnen (1996), or Kleijnen (1995). For more information on IFFD see Campolongo, Kleijnen, and Andres (2000), Andres and Hajas (1993), Andres (1997), and Saltelli et al (1995). Elster and Neumaier (1996) is the source for information on edge designs. Both IFFD and edge designs are based on Hadamard matrices. For practical information about group screening see Kleijnen (1987). Other sources for group screening are Mauro (1984), Mauro and Smith (1984) and Watson (1961). For a case study comparing fractional factorial designs to two-stage group screening, see Ivanova, Malone and Mollaghasemi (1999). For all others see Myers and Montgomery (1995).

4 ANALYSIS

Once the screening experiments are completed an analysis to find the functional form and ultimately to optimize the response must be completed. These analyses include techniques discussed in Myers and Montgomery (1995) for response surface methodology: normality and interaction plots, construction of confidence intervals, hypothesizing and fitting regression models. Other techniques include nonlinear regression, training neural networks in place of a regression equation, using stepwise regression, and multinomial analysis of variance. See Myers and Montgomery (1995), Kleijnen (1987), or Ratkowsky (1990) for material on these analyses. Smith (1996) and Christodoulou and Georgiopoulos (2000) discuss neural networks and how they may be used to fit a nonlinear model in place of regression.

5 CONCLUSIONS

There are several methods available for screening. However, each method is limited in one aspect or another and the analyst should make informed decisions about which method to apply. Tradeoffs must be made by the analyst when choosing a screening method according to the criteria outlined above and the circumstances of the simulation model.

Table 4 lists the same screening methods yet again along with their relative ease of use in building the design, conducting the analysis, and whether software is available to assist in the use of the method.

Table 4: Relative Ease of Screening Designs

Existing Screening Experiment Methods				
Design	**Ease of Use**			
	Overall	Software Avaialble	Design Phase	Analysis Phase
Desired	**Easy**	**Yes**	**Easy**	**Easy**
One-at-a-time	Easy	No	Easy	Easy
Full Factorial	Easy	Yes	Easy	Moderate
Fractional Factorial	Easy	Yes	Moderate	Moderate
Edge Designs	Moderate	No	Difficult	Moderate
Two-Stage Group Screening	Moderate	Yes	Difficult	Moderate
Sequential Bifurcation	Moderate	No	Easy	Moderate
IFFD	Moderate	Yes	Easy	Difficult

ACKNOWLEDGMENTS

Many thanks and sincere appreciation to Terry Andres for copies of several publications and for many answers to our questions on IFFD; to Jack Kleijnen for the pre-publication copy of the chapter in *Mathematical and Statistical Methods for Sensitivity Analysis of Model Output* edited by A. Saltelli and K. Chan; and to Bert Bettonvil for copies of technical reports regarding Sequential Bifurcation.

REFERENCES

Andres, T. H. 1997. Sampling methods and sensitivity analysis for large parameter sets. *Journal of Statistical Computation and Simulation*. 57(1-4): 77-110.

Andres, Terry H. and Wayne C. Hajas. 1993. Using iterated fractional factorial design to screen parameters in sensitivity analysis of a probabilistic risk assessment model. *Joint International Conference on Mathematical Methods and Supercomputing in Nuclear Applications*. ed. Kusters, Stein, and Werner. 2:328-337.

Bettonvil, Bert and Jack P.C. Kleijnen. 1996. Searching for important factors in simulation models with many factors: sequential bifurcation. *European Journal of Operational Research*. 96:180-194.

Campolongo, F., J. Kleijnen, and T. Andres. 2000. Screening methods in sensitivity analysis. In *Mathematical and statistical methods for sensitivity analysis of model output.* ed. A. Saltelli and K. Chan. Chichester, England: Wiley.

Christodoulou, C. and M. Georgiopoulos. 2000. *Applications of neural networks in electromagnetics.* Artech House, in print.

Elster, Clemens and Arnold Neumaier. 1995. Screening by conference designs. *Biometrika*. 82(3):589-602.

Ivanova, Theodora, Linda Malone, and Mansooreh Mollaghasemi. 1999. Comparison of a two-stage group-screening design to a standard 2^{k-p} design for a whole-line semiconductor manufacturing simulation model. In *Proceedings of the Winter Simulation Conference v.1* ed. P.A. Farringtom, H.B. Nembhard, D.T. Sturrock and G.W. Evans, 640-646.

Jacoby, J. E. and S. Harrison. 1962. Multi-variable experimentation and simulation models. *Naval Research Logistics Quarterly*. 9:121-136.

Kleijnen, Jack P. C. 1987. *Statistical tools for simulation practitioners*. New York: Marcel Dekker.

Kleijnen, Jack P. C. 1995. Sensitivity analysis and optimization in simulation: design of experiments and case studies. In *Proceedings of the Winter Simulation Conference*, ed. C. Alexopoulos, K. Kang, W.R. Lilegdon, and D. Goldsman, 133-140.

Mauro, Carl A. 1984. On the performance of two-stage group screening experiments. *Technometrics*. 26(3):255-264.

Mauro, Carl C., and Dennis E. Smith. 1984. Factor screening in simulation: evaluation of two strategies based on random balance sampling. *Management science*. 30(2):209-221.

Myers, Raymond H. and Douglas C. Montgomery. 1995. *Response surface methodology: process and product optimization using designed experiments*. New York: John Wiley & Sons.

Ratkowsky, David A. 1990. *Handbook of nonlinear regression models*. New York: Marcel-Dekker.

Saltelli, A., T. H. Andres, and T. Homma. 1993. Some new techniques in sensitivity analysis of model output. *Computational statistics & data analysis*. 15:211-27.

Saltelli, A., T. H. Andres, , and T. Homma. 1995. Sensitivity analysis of model output: performance of iterated fractional factorial design (IFFD) method. *Computational statistics & data analysis*. 20:387-407.

Smith, Murray. 1996. *Neural networks for statistical modeling*. London: International Thomson Computer Press.

Trocine, Linda. June 20, 2000. *Designing screening experiments using modern heuristics: dissertation proposal*. Unpublished report, University of Central Florida, Orlando.

Watson, G.S. 1961. A study of the group screening method. *Technometrics*. 3(3):371-388.

AUTHOR BIOGRAPHIES

LINDA TROCINE is a doctoral candidate in UCF's industrial engineering program. Her current research is in effective and efficient screening of large number of variables through modern heuristic search methods such as genetic algorithms, simulated annealing, and tabu search. Her other interests are in industrial statistics, optimization, and simulation modeling and analysis. She earned a B.S. in mathematics from the State University of New York, College at New Paltz and an M.S. in engineering management from UCF. She has 12 years professional experience in information systems, project management, and software development. She is a student member of the American Statistical Association and the Society for Computer Simulation. Her email and web addresses are <ljt49602@pegasus.cc.ucf.edu> and <pegasus.cc.ucf.edu/~ljt49602>.

LINDA C. MALONE is a professor in the Department of Industrial Engineering and Management Systems at UCF. She received a B.S. in mathematics from Emory and Henry College, an M.S. in mathematics from the University of Tennessee in Knoxville, and a Ph.D. in statistics from Virginia Polytechnic Institute and State University in 1975. Her major interests are in design of experiments, response surface analysis, and regression. She has a number of publications and has co-authored an introductory statistics textbook. Her email and web addresses are <lmalone@mail.ucf.edu> and <ie.engr.ucf.edu/faculty/malone/>.