Proceedings of the 2001 Winter Simulation Conference
B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds.

AN OVERVIEW OF NEWER, ADVANCED SCREENING METHODS FOR THE INITIAL PHASE IN AN EXPERIMENTAL DESIGN

Linda Trocine

Venutek, LLC
P.O. Box 622331
Oviedo, FL 32762-2331, U.S.A.

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

Screening is the first phase of an experimental study on systems and simulation models. Its purpose is to eliminate negligible factors so that efforts may be concentrated upon just the important ones. Successfully screening more than about 20 or 30 factors has been investigated only in the past 10 or 15 years with most improvements in the past 5 years. A handful of alternative methods including sequential bifurcation, iterated fractional factorial designs, and the Trocine Screening Procedure are described and evaluative and comparative results are presented.

1 INTRODUCTION

Simulation models, as with any system under study, involve many inputs, referred to as independent variables or factors, and one or more outputs, referred to as the dependent variables or responses. We focus on a single response in this paper. For a new approach to multiple responses, please see Grimes (2001).

In order to improve or optimize the response or output of the system, the best settings of the independent variables must be determined. This requires an understanding of the relationships between and among the variables. One approach to learning about a system is to passively observe it. A better approach is to systematically experiment on the system by setting the independent variables to levels and observing the response. Such an approach employs an experimental design followed by statistical analyses in order to make inferences about the underlying relationships between and among the inputs and outputs.

Myers and Montgomery (1995) refer to the initial phase of an experimental study as the screening phase. They call it Phase 0. After screening is done, a first-order linear model is constructed in Phase 1. Subsequent phases involve second-order terms and honing in on the optimum region. Screening is critically important because if it is not done well, all subsequent experimentation may yield erroneous results.

The most efficient screening designs use only two discrete levels for each factor. The full factorial designs in two levels require 2^k runs and can measure not only the effects of each of the k factors but can also measure all combinations of the interaction effects between and among the k factors. Because full factorial designs require a large number of runs (for as few as 5 factors 32 runs are required), fractional factorial designs were developed. Two level fractional factorial designs, denoted as 2^{k-p}, are very efficient and generally effective. These designs use many fewer runs than full factorial designs and can estimate main effects and some two-way interaction effects. However, they have not and cannot be used for more than about 20 factors. Software cannot generate them and their usefulness is limited because of the number of runs required. Ivanova, Malone, and Mollaghasemi (1999) showed that the results of screening 17 factors with a fractional factorial design were inconclusive when compared with a group screening experiment on the same 17 factors.

This paper provides an overview of screening methods available for screening more than 20 independent variables including two-stage group screening, sequential bifurcation, iterated fractional factorial designs (IFFD), supersaturated designs (SSDs) and a promising new method called the Trocine screening procedure (TSP). These methods will be compared and contrasted in terms of efficiency, effectiveness, and robustness defined in the next section on cases from the literature and on simulation models developed by the authors.

2 PURPOSE OF SCREENING

Like a sieve or a screen that is used to find chunks of gold in dust and dirt, the purpose of factor screening is to eliminate negligible factors in favor of concentrating experimental efforts on those factors that are important. This is

possible because of the parsimony principle that says that only a few factors are responsible for most of the effect in a response while most are not. This is equivalent to the Pareto rule used in quality studies. The vital few are selected with the largest expected returns. It is also unrealistic to believe that all factors affect a response equally.

After screening is completed, we expect to have reduced a large and complex problem into a simpler one with a few important factors.

2.1 Criteria for Screening Methods

In choosing a screening method there are three or four main criteria to consider. The criteria are efficiency, effectiveness, robustness, and ease of use.

The first criterion is how efficient the method is. Efficient screening methods are those that require a manageable number of runs. Efficiency is a qualitative measure that depends on the size of the problem (i.e. number of factors).

The second criterion is effectiveness. In the literature much less emphasis is placed on effectiveness because it is more difficult to measure. In practical problems, the underlying coefficients of the effects are unknown, so we cannot measure the effectiveness of a method directly. We can compare alternative methods and hope for confirmation of findings. Simulated cases, with known coefficients, can be used to measure effectiveness directly; however, this yields only empirically observed results. Analytic results are not generally tractable.

The third criterion is herein referred to as robustness. Some methods may only be applied if certain conditions of the problem are known to exist. In practice, however, the conditions of the problem are what is sought! Thus we desire screening methods that work well without prior knowledge of the problem. For example, sequential bifurcation requires that the direction of the signs of all the effects in the problem are oriented in the same way. In general this is not the case.

The last criterion is desirable but is not necessary. An easy to use method is certainly easier for the experimenter but can be forsaken in exchange for effective, efficient, and robust methods. Statistically designed experiments and analyses require a great deal of skill and interpretation and are not considered easy by non-statisticians!

2.2 Design and Analysis

Screening methods involve two major components. The first is the design. The design has coded levels for each of the factors. There are numerous construction methods for numerous types of designs. There are even numerous criteria for selecting designs! The second component is the analysis. After the responses are observed for each run in the design, the data must be analyzed in order to make inferences about which factors are important. Numerous analysis methods are available, some involving plots of the data and others computations on the data. Examples are normal probability plots, regression analyses, effects estimates, and so forth. All the many design construction methods and various analyses methods are beyond the scope of this paper and are discussed in a plethora of textbooks including Montgomery (2001), Wu and Hamada (2000), and Myers and Montgomery (1995). The five methods mentioned in section 1 are explained in some detail in the next section.

3 SCREENING METHODS

In the first section, we introduced the factorial designs which have a long history and proven record of being efficient, effective, and robust when screening problems with up to about 15 to 20 factors. With fractional factorial designs, a one-half or one-quarter fraction (of runs) of the larger full factorial design is selected for screening. Choosing the design and the arrangement of columns may be done in such a way as to improve the design. Regardless these designs are constructed before any observations are made. We call this prior design. After all the observations are made for the entire design, the entire data set is analyzed. Analyses include effects estimates, normal probability plots, and standard least squares regression. Regression is used to estimate the coefficients of effects up to one fewer than the number of runs in the design. These analyses allow us to make inferences about the importance of the factors.

Because the subject of this paper is screening large numbers of factors, which we define to be more than 15 or 20, we refer the reader to Myers and Montgomery (1995) for a more complete discussion of methods with fewer factors, including the fractional factorial and Plackett-Burman designs. Next we present an overview of five methods for larger numbers of factors. For each method we describe how the design is constructed and how the analyses are conducted.

3.1 Super Saturated Designs (SSDs)

Designs are called supersaturated when they aim to estimate more effects than they have runs. If n is the number of factors and m is the number of runs for screening then $m < n$. SSDs are prior constructed designs like both fractional factorial designs and Plackett-Burman designs. These are intentionally constructed to be supersaturated for the sake of efficiency. Thus we not only have a problem with confounding of effects but also of insufficient degrees of freedom to apply traditional analyses such as regression. We refer the reader to two papers by Holcomb, Montgomery, and Carlyle (2000a, 2000b) that describe the state of the art of supersaturated designs. The latter paper discusses several construction methods while the former compares sev-

eral alternative analyses methods. The options for analyses are stepwise regression, all models regression, normal probability plots, the contrast variance method and a bootstrapped contrasts method. Holcomb, Montgomery, and Carlyle (2000a) conclude that SSDs should be used with caution. Our interpretation of their findings is that you might get good results using them if you are lucky. You will have had to choose the right design, match the factors and columns in the design the right way, and choose and use the right analysis method in the right way. Hence we do not generally consider SSDs to be reliable.

3.2 Two-Stage Group Screening

Two-stage group screening has been used for screening roughly 15 to 50 factors. The experimenter uses experience and knowledge of the problem and the factors to arrange the factors into logical groups. Then a fractional factorial design is run on the groups. Upon identification of important groups, the factors within that group are separated into smaller groups or individual factors and a new fractional factorial design on the subgroups is run until the important factors are identified. The results of the first stage are analyzed and used to design the second stage. Hence, this screening method is iterative. In order for this to work, the factors and the interactions between and among factors in a group must not cancel each other through opposing signs-of-effects. Otherwise, the group may be considered unimportant and eliminated. Also, interactions between factors in different groups are not measured and if they exist may confound the results of the groups.

Two-stage group screening was successfully demonstrated on simulated cases by Mauro (1984) and Mauro and Smith (1984). Ivanova, Malone, and Mollaghasemi (1999) reported the results of comparing two screening methods on a simulation model of a 17 factor semiconductor manufacturing process. Two-stage group screening was performed, using a total of 64 runs, and led to a set of important factors. A separate 64 run fractional factorial design was run on the same simulation model resulting in a very different set of important factors. The authors also tried analyzing both data sets using different alpha values for inclusion of variables without improvement in matching the results of the two screening methods.

3.3 Sequential Bifurcation

Bettonvil and Kleijnen (1996) developed sequential bifurcation from an earlier idea by Jacoby and Harrison (1962). The design is constructed one point at a time using feedback from all prior points to direct the search for the important factors. Before the design begins, the factors are all arranged so that the direction of their effects is all positive (or negative). The method also requires that the response function is nondecreasing (or nonincreasing). Thus all main effects and all interaction effects are nonnegative (or nonpositive). If these conditions are not met, sequential bifurcation will give erroneous results. However, if they are met, as can often be done with simulation models, sequential bifurcation is extremely efficient.

First the point where all factors are set to the high level is run. Then the point where all factors are set to the low level is run. If the difference between the responses is considered to be significant, then some factors in the group (of all factors) are significant and the procedure continues. Cheng (1997) offers an approach for deciding whether the difference between two responses of these two points is significant or not by choosing an indifference value, δ, based on prior knowledge of the problem. We suggest that if there is not prior knowledge, that one or two replicates of one of these two points could be used for a rough estimate of the variability of the experimental region, similar to TSP in section 3.5 below.

If the difference is significant, then the point where the first half of the factors is set to the low level and the second half of the factors is set to the high level is run. If the response is about the same as one of the two earlier run points, then half of the factors can be eliminated. We characterize this as a divide and conquer approach.

Using Bettonvil's notation, we denote $y_{(j)}$ as the point where all factors 1 through j are set to low and factors j+1 to n are set to high. A group is eliminated if the parameter β_{ij} is insignificant. A group β_{ij} is divided in half again and responses are compared until the significant factors are isolated. We show an example of the process on a simulated problem from Trocine (2001). See Equation 1. There are 20 factors in this problem, A, B, C,…,T. Factors C, G, and M have nonzero coefficients. The tree of decisions made to identify the three important factors is shown in Figure 1. Each $y_{(j)}$ is a run. There were a total of 12 runs to find 3 significant out of 20 initial factors. We chose δ to be 1, which worked correctly and consistently for this problem. Results may be less accurate with a different δ value.

$$y = 4 \times C + 4 \times G + 4 \times M + \mathcal{N}(-2,\ 0.4) \quad (1)$$

Cheng (1997) reused a 24 factor case of unknown origin first reported in Bettonvil and Kleijnen (1996). Bettonvil and Kleijnen (1996) also reported a 281 factor ecology model screened in 144 runs finding 15 important effects. A 128 factor problem screened in 16 runs was reported in Bettonvil and Kleijnen (1996).

3.4 Iterated Fractional Factorial Designs

The Canadian agency AECL had constructed a simulation model of a nuclear waste disposal problem with over 3000 independent variables (Goodwin et al, 1994). AECL

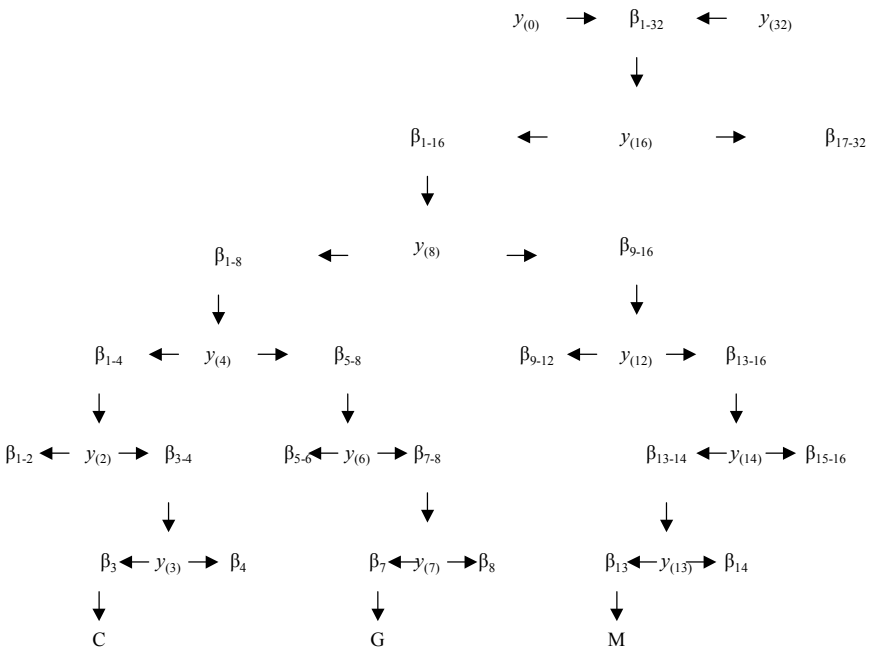

Figure 1: Branches taken by Sequential Bifurcation

needed a method to screen the 3000 variables. Andres and Hajas (1993) described a new method called IFFD, for iterated fractional factorial designs, that was used to screen the variables in this problem. They used 512 runs altogether and found 8 significant factors.

The method starts with the construction of a fractional factorial design and with factors randomly assigned to groups. A third level is introduced by setting 25% of the runs to 0 while the remaining runs are equally split between +1 and –1. The process is repeated with factors assigned again randomly to (different) groups in the same way. After several iterations the data set is analyzed using forward stepwise regression. See Andres (1997) for a thorough explanation of how to use IFFD. Upon completion of the experiments and analysis, the significant factors are identified.

Andres (1997) used a novel approach to check the accuracy of IFFD in finding the significant factors in the simulation model. Three sets of runs were observed. The first set randomly assigned the levels of all the factors. The second set had randomly assigned levels for the factors that were not selected by IFFD while the selected ones were held at constant values. The third set of runs had randomly assigned levels for the selected factors while the other factors were set to constant values. Next pairwise comparisons were made between results of the first and second set and the first and third set. The results were then plotted. The plot showed that changes only to the selected factors resulted in a plot accounting for the variation in the responses. The plot comparing the first and second set showed that little variation in the responses was due to this set. See Andres (1997) Table 2, Figure 2 and Figure 3.

According to Andres (2000), other cases demonstrating the use of IFFD on large problems have not been published. Also, the method is designed for very large problems such as the nuclear waste disposal simulation model and may not be efficient on problems with as few as 100 factors. Andres and Hajas (1993) also stated that the method works best when a small number of factors dominate. We note that 8 of 3000 is less than 1% of the total.

3.5 Trocine Screening Procedure

Simulation is an accepted and viable method of understanding and experimenting with large, complex systems, safely and with good application of results back to the physical problem. Simulation has allowed experimenters to study larger and more complex problems. But experimental designs have not kept the pace with these larger problems except for the work by Bettonvil and Kleijnen and Andres and Hajas, and those studying supersaturated designs. As we have shown, each of these three approaches has its own limitations as to the size or type of problem, and the accuracy of the results. We therefore have been researching the problem of screening many factors in an efficient and effective way while being robust to various types of problems.

The ideas behind TSP (Trocine screening procedure) are to use a genetic algorithm to generate points to observe and to experiment iteratively using feedback from prior observations. During the process of researching the approach, a handful of other heuristics were employed, and a new iterative analysis method was developed. All the many pieces are put together in a framework for screening many factors.

Figure 2 shows a high-level flow diagram of TSP's framework. In the first step, 3 replicates of the single point where all factors are set to the high level are run. The mean and range of the three responses are computed and used both in computing the fitness of points for the genetic algorithm and in analyzing the results. In the second step, an initial set of points is run. The number of points is dependent on the total number of factors in the problem. For 15 factors, 4 initial points are run. For 16 to 31 factors, 5 initial points are run. These points are constructed in such a way that no two factors are *totally positively aliased* with each other. This means that no two columns of levels of +1 and –1 are identical. Whereas there are factors that are *totally negatively aliased* because the column of one is –1 times the column of the other. See Table 1. Column N is totally negatively aliased with C but by inspection, no two columns are identical.

In the third step of TSP, new points are generated. First the fitness of the observed points is computed. The fitness is defined to be the absolute difference of the response from the endpoint of the range about the mean of the replicates'. Thus the larger the difference of the response from the replicates' the more information is being derived by the point, thus it has a higher fitness. Points that have a response value within the range about the mean do not provide any useful information about which factors might be significant and hence their fitness is set to zero. The genetic algorithm applies the standard operators of selection, mutation, and crossover. Parents are selected with probability proportional to their fitness over the total fitness of all observed points. Mutation occurs by flipping a level of a single factor between +1 and –1 of a single factor. Crossover occurs between two selected parents by selecting factor levels from either parent with 50% probability for each.

Other heuristics are applied to the offspring points after the genetic algorithm is run. These heuristics balance the number of instances of +1 and –1 in each factor's column so that no level appears 6 more times than the other level. Aliases among the four most highly aliased factors are artificially broken in the new runs and replicates of previously observed points are prevented. For each iteration four new points are added to the design and observed. The process repeats after the experimenter enters the responses.

Also at each iteration, the data set is analyzed to see which factors appear to be significant, which appear not to be significant, and which are still questionable. Those that are not significant are discarded so that subsequent experimentation will concentrate on the questionable ones. For each factor a data structure called the scorecard is kept that contains four tallies and one accumulator. Comparisons are made for every pair of points to see how different their responses are and whether the factor levels are different as well. If a factor's level changes but the response did not then that factor may have negligible effect and so an increment to a corresponding tally is made. On the other hand, if the responses are widely different and a factor

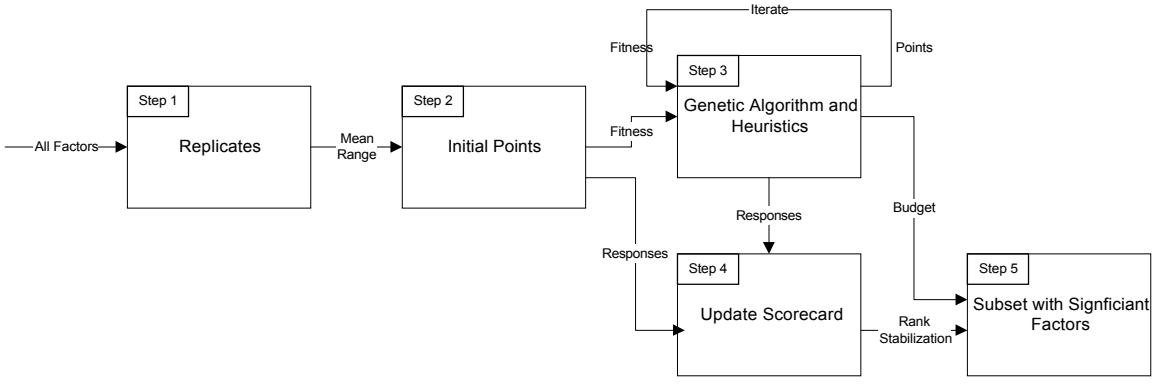

Figure 2: Flow of TSP

Table 1: Initial Points for 15 Factor Problem in TSP

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
-1	-1	-1	-1	1	-1	-1	-1	1	1	1	-1	1	1	1
-1	-1	-1	1	-1	-1	1	1	-1	-1	1	1	-1	1	1
-1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	1	-1	1
-1	1	-1	-1	-1	1	1	-1	1	-1	-1	1	1	1	-1

level changed, then a different tally is incremented to record the fact. Two other tallies track the direction of the effect of a factor and the accumulator adds up the portion of the change in the response attributed to a factor. All these values taken together form a total score for each factor.

After each iteration the total score of each factor is ranked. If the ranks of the factors change significantly between iterations then TSP continues to derive useful information about the problem. If, on the other hand, the ranks are not changing much, no new information is being derived so the procedure stops. It selects those factors that have greater than average proportion of the total of all scores. This is about 25% of the total number of factors in the problem. Alternatively, TSP will stop if a preset budget is reached and select the factors in the same way. See Trocine (2001) for a complete discussion of how TSP works.

4 SIMULATED CASES

In order to demonstrate the performance of TSP some simulated cases were constructed. These cases were implemented as simulation models in Microsoft Excel. Each factor has a coefficient. The response is computed by summing the product of the coefficient with the factor's level and adding the error term to it. This is based on the assumption made in screening that a first order linear model is a good approximation of the underlying model for the purpose of screening. The error terms were normally distributed in all cases but one where exponential error terms were used. Again in screening we assume that error terms are normally distributed. The variability of the responses were constant throughout the experimental region in these simulated cases. This is consistent with our practice of ensuring that a process is in control before we tinker with it.

The cases are listed in Table 2 and ordered roughly by the number of significant factors in the problem. The cases range in total number of factors from 16 to 100. They range in number of significant factors from 1 to 15. And the range of the proportion of significant factors to total factors was from 2% to 37%. Recall the parsimony principle that states that only a few are important and most are not. We would not employ screening at all if we did not believe that this principle holds true. There are columns that show whether the signs of the effects are in the same direction or opposing directions, the coefficient of variation, and whether comparisons were made with other methods. These cases were constructed to represent a variety of problem sizes and conditions so that TSP could be evaluated across these conditions.

Additionally, cases 15 and 16 were taken from the literature. Case 15-Cheng was published in Cheng (1997) and Bettonvil and Kleijnen (1996). This problem had 24 factors; the coefficients were all positive and are plotted in Figure 3. Bettonvil and Kleijnen applied sequential bifurcation deterministically on this case (without error terms) to find the most significant factors. Cheng (1997) proposed an improvement to sequential bifurcation where a δ value is chosen by the experimenter based on prior knowledge of the problem. The δ is referred to as the indifference zone and serves the purpose of distinguishing responses as significantly different or not. In TSP we use the range of the replicates responses' for this same purpose. Thus we propose that δ may be chosen without prior knowledge when using sequential bifurcation by running 3 replicates of a single point and computing the range or some other measure of the variability.

Case 16-Holcomb was taken from Holcomb, Montgomery, and Carlyle (2000a). It was based on the Williams (1968) rubber making process that was again published by Lin (1993). In this case a Plackett-Burman design in 23 factors was constructed and then analyzed with alternative methods for dealing with supersaturated design sets. Williams concluded there were 7 factors significant and published the coefficients. As Holcomb, Montgomery, and Carlyle (2000a) did, we simulated the case using the published coefficients and their own error terms. The coefficients are shown in Table 2. Both of these cases have a high proportion of significant factors (37.5 % and 30.4%, respectively).

For each of the 16 cases, five replications of the use of TSP on the case were run. The results were consistent across the replications. Thus averages of the five replications are presented next in the evaluation of performance. For more details, please see Trocine (2001).

5 EVALUATION OF PERFORMANCE

In this section, we present several charts. These charts illustrate how TSP performed on the 16 cases described in the last section on both efficiency and effectiveness. Both Type I and Type II errors were recorded for the cases in much the same way as described in Holcomb, Montgomery, and Carlyle (2000). Ideally, a screening method will always select the significant factors and never select the insignificant factors. However, if a screening method selects an insignificant factor, this is called a Type I error. Type I errors are computed by the taking the number of selected insignificant factors and dividing them by the total number of insignificant factors. In all cases the average Type I errors were less than 19% with TSP. A Type II error occurs when we fail to select a significant factor. Type II errors are computed by dividing the number of significant factors that the method failed to select by the total number of significant factors in the problem. Note that the scales of Type I and Type II measures are different.

There are always fewer significant factors than insig nificant ones so when a significant factor fails to be selected the error rate is much higher than when one insignificant factor is selected.

Table 2: Simulated Cases Used to Evaluate TSP

Case Name	Verification/ Contrived/ Literature	Number of Signficant Factors	Total Number of Factors	Proportion of Signficant Factors	Interaction Terms	Signs	Smallest Coef/StDev	Replications Run
1-EN50-1	C	1	50	2.0%	--	Equal	8.33	5
2-ON16-3	C	3	16	18.8%	--	Oppose	8.33	5
3-ON20-2	V	2	20	10.0%	--	Oppose	8.57	5
4-EN20-3	V	3	20	15.0%	--	Equal	10.00	5
5-EE20-3	C	3	20	15.0%	--	Equal	--	5
6-VN50-3	C	3	50	6.0%	--	Same	18.00	5
7-VN16-5	C	5	16	31.3%	--	Same	10.00	5
8-ON20-5	C	5	20	25.0%	--	Oppose	13.33	5
9-ON50-5	C	5	50	10.0%	--	Oppose	8.57	5
10-VN50-10	C	10	50	20.0%	--	Same	7.14	5
11-ON100-10	C	10	100	10.0%	--	Oppose	6.00	5
12-VN50-15	C	15	50	30.0%	--	Same	6.25	5
13-VN100-15	C	15	100	15.0%	--	Same	7.14	5
14-VN20-3I	C	--	25	--	one 2-way	Same	10.00	5
15-Cheng	L	9	24	37.5%	--	Same	4.33	5
16-Holcomb	L	7	23	30.4%	--	Oppose	11.38	5

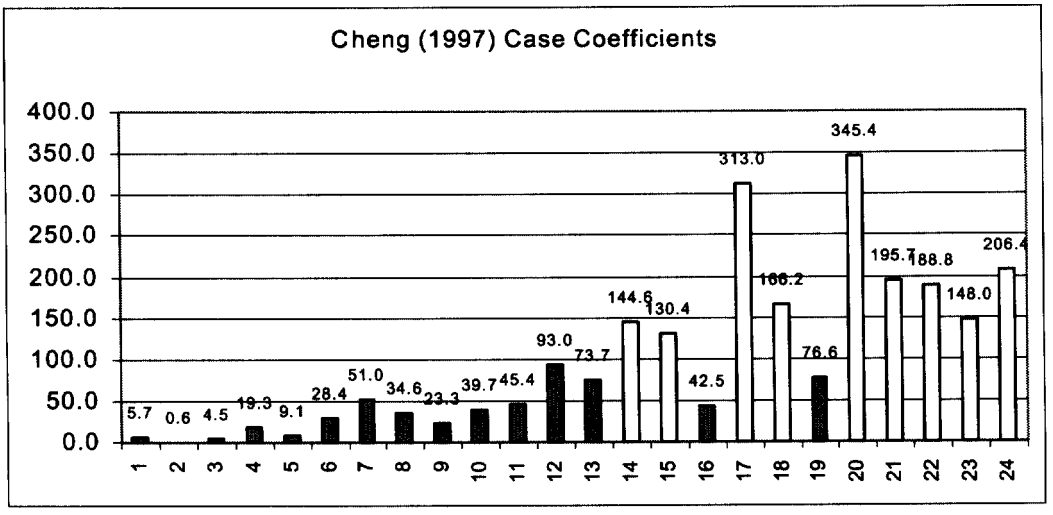

Figure 3: Magnitude of Coefficients in 24 Factor Case

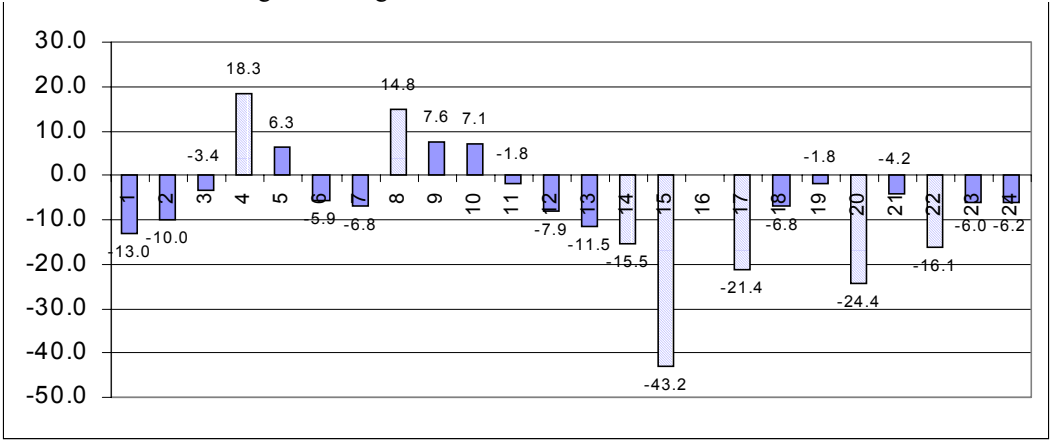

Figure 4: Magnitude of Coefficients of 23 Factor Case

Figure 5 shows the average Type I and Type II errors by the total number of factors in the problem. Figure 6 shows the average Type I and Type II errors by the number of significant factors in the problem. The interpretation of these two charts is that the Type I errors in TSP depend on the total number of factors in the problem while Type II errors do not. As the total number of factors increases, the Type I errors also increase. The Type II errors on the other hand increase when the number of significant factors increases as shown in Figure 6. We point out, however, that when there were 15 significant factors of 100, TSP identified about 6 on average. Also, in all cases but 16-Holcomb, TSP always identified 3 of the significant factors. And in problems with 3 or fewer significant factors, TSP always selected them. Thus TSP works best on these problems and opportunities exist to improve it when the number of significant factors is greater.

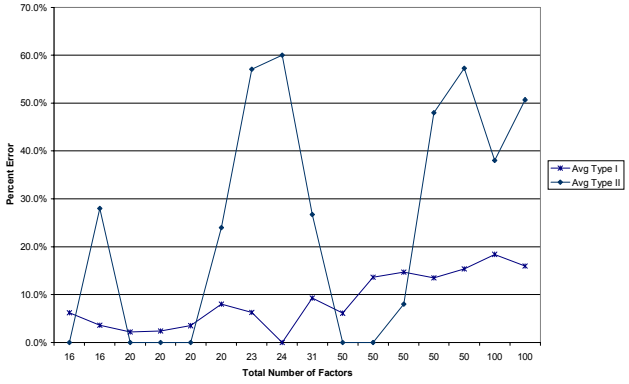

Figure 5: Average Observed Errors

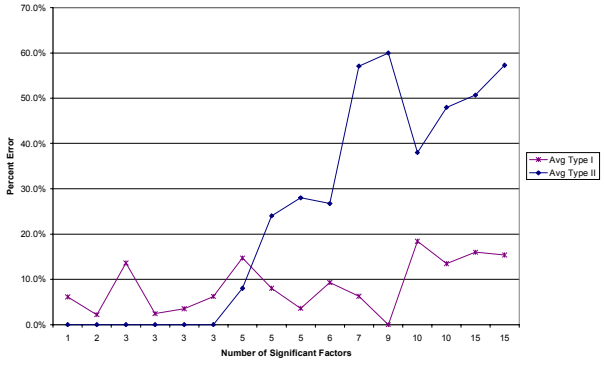

Figure 6: Average Observed Errors

Figure 7 and Figure 8 show that TSP performed equally well regardless of the signs of the effects. Figure 7 plots Type I errors for all runs of cases in two groups, one with all same signs of effects (n=45), the second with mixed signs of effects (n=33). Though the number of instances is unequal, the shapes of both histograms is the same. Figure 8 shows similar curves for Type II errors. This is important to note because sequential bifurcation, which is more efficient, cannot work on problems with mixed signs or when the signs are unknown.

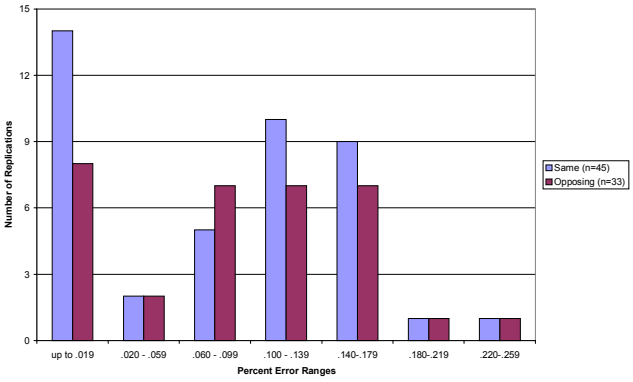

Figure 7: Type I Errors by Signs of Effects

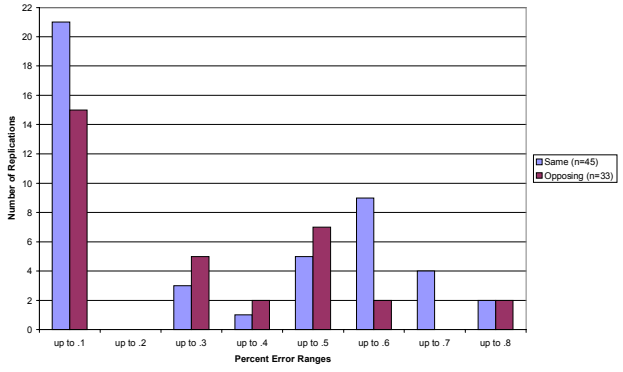

Figure 8: Type II Errors by Signs of Effects

Figure 9 shows that TSP is acceptably efficient for all cases. The vertical axis is the average number of observations per case which ranged from 32 runs to 70 runs. The horizontal axis corresponds to the individual cases ordered by the total number of factors in the case. As expected, the average number of runs increases as the total number of factors increases.

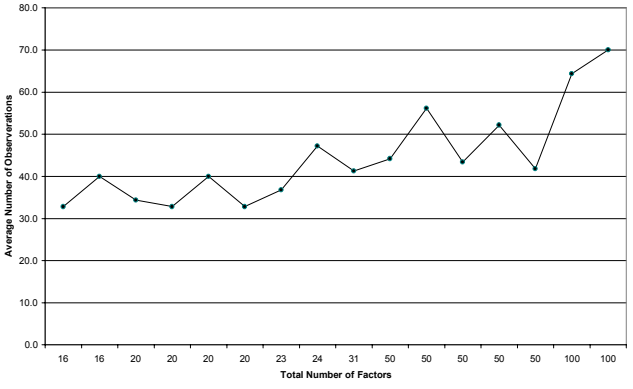

Figure 9: Efficiency of TSP

A result that is not shown in these charts is that TSP is by far much easier to use than all the other methods discussed herein. The experimenter only enters the number of factors, the budget, and the responses. The entire design is constructed by the software and all the analyses and interpretations are conducted by the TSP software. The results presented in the section did not involve construction of plots, interpretation of plots, nor running regression analyses, choosing alpha values, and so forth. TSP produces a list of factors. The charts show how the list TSP produced matched the factors with significant coefficients.

6 OPPORTUNITIES AND CONCLUSIONS

Bert Gunter (2001) made a comment at the Spring Research Conference on Statistics in Industry and Technology this year that 90% of what we do is screening. His point is that it is therefore very important to get screening right. We believe there are many more opportunities to improve screening methods for large numbers of variables and that this is a valuable research area. The methods we presented: fractional factorial designs, supersaturated designs, sequential bifurcation, iterated fractional factorial designs, and the Trocine screening procedure, all make an impact and all can be improved. TSP, however, has been shown to be robust to the various types of cases used herein particularly in identifying the top 3 most significant factors. Because of its ease of use, we recommend those with limited statistical training consider using TSP and those with statistical training to apply it to cases, evaluate its performance, and compare its performance with other methods or with known coefficients. Software will be made available on the website www.venutekllc.com.

Simulation models will continue to be constructed for larger and more complex problems. Experimentation on these models should always be tackled in a systematic way to maximize information derived from a minimum number of runs. We therefore strongly encourage practitioners to use an available screening method.

ACKNOWLEDGMENTS

Many thanks to Bert Bettonvil, Jack Kleijnen, Terry Andres, Russell Cheng, and Don Holcomb for sharing papers and materials discussed herein.

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. May 18, 2000. Personal communications.

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. *Proceedings of the Joint International Conference on Mathematical Models and Supercomputing in Nuclear Applications, 2.* Karlsruhe, Germany, 19-23 April 1993. Eds. H. Küsters, E. Stein, and W. Werner. 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.

Cheng, Russell. 1997. Searching for important factors: sequential bifurcation under uncertainty. In *Proceedings of the Winter Simulation Conference.* ed. S. Andradóttir, K.J. Healy, D.H. Withers, and B.L. Nelson. 275-280.

Goodwin, B.W., D.B. McConnell, T.H. Andres, W.C. Hajas, D.M. LeNeveu, T.W. Melnyk, G.R. Sherman, M.E. Stephens, J.G. Szekely, P.C. Bera, C.M. Cosgrove, K.D. Dougan, S.B. Keeling, C.I. Kitson, B.C. Kummen, S.E. Oliver, K. Witzke, L. Wojciechowshki, and A.G. Wikjord. 1994. *The Disposal of Canada's Nuclear Fuel Waste: Postclosure Assessment of a Reference System.* Atomic Energy of Canada Report AECL-10717 (COG-93-7). Pinawa, Manitoba: Whiteshell Laboratories.

Grimes, Steven R. 2001. Multiple Response Surface Methodology. Dissertation. Unpublished report. University of Central Florida.

Gunter, Bert 2001. Comment in his presentation. June 20, 2001, Spring Research Conference on Statistics in Industry and Technology, Roanoke, Virginia, June 18-20, 2001.

Holcomb, Don R., Douglas C. Montgomery, and W. Matthew Carlyle. 2000a. Analysis of Supersaturated Designs. Technical Report. Arizona State University.

Holcomb, Don R., Douglas C. Montgomery, and W. Matthew Carlyle. 2000b. Some Combinatorial Aspects, Construction Methods, and Evaluation Criteria for Supersaturated Designs. Technical Report. Arizona State University.

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 wholeline semiconductor manufacturing simulation model. In *Proceedings of the Winter Simulation Conference v.1.*ed. P.A. Farrington, H.B. Nembard, 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.

Lin, D. K. J. 1993. A new class of supersaturated designs. *Technometrics,* 35. 28-31.

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

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

Montgomery, Douglas C. 2001. *Design and analysis of experiments 5th edition*. New York: John Wiley.

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

Trocine, Linda. 2001. An efficient, effective, and robust procedure for screening more than 20 independent variables employing a genetic algorithm. Bell and Howell Dissertations Online. University of Central Florida.

Williams, K. R. 1968. Designed experiments. *Rubber Age,* 100. 65-71.

Wu, C.F. Jeff and Michael Hamada 2000. *Experiments: planning, analysis, and parameter design optimization.* New York: John Wiley.

AUTHOR BIOGRAPHIES

LINDA TROCINE is a process improvement consultant for Venutek, LLC. She also researches, develops, and markets solutions, systems, and software. She received a Ph.D. in industrial engineering and an M.S. in engineering management from the University of Central Florida. She also holds a B.S. in mathematics from the State University of New York, College at New Paltz. Her research interests are in applied statistics, artificial intelligence, and simulation, for improving processes and systems. She is a member of the American Statistical Association and enjoys teaching on the side, as well. Her email and web addresses are <ltrocine@cfl.rr.com> and <www.venutekllc.com>.

LINDA C. MALONE is a professor in the Industrial Engineering and Management Systems department at the University of Central Florida. 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/>.