Proceedings of the 2005 Winter Simulation Conference
M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds.

A TWO-PHASE SCREENING PROCEDURE FOR SIMULATION EXPERIMENTS

Susan M. Sanchez

Operations Research Department and
Graduate School of Business & Public Policy
Naval Postgraduate School
Monterey, CA, 93943-5219, U.S.A.

Hong Wan

Industrial Engineering Department
Purdue University
West Lafayette, IN, 47907-2023, U.S.A.

Thomas W. Lucas

Operations Research Department
Naval Postgraduate School
Monterey, CA, 93943-5219, U.S.A.

ABSTRACT

Analysts examining complex simulation models often conduct screening experiments to identify the most important factors. Controlled sequential bifurcation (CSB) is a screening procedure, developed specifically for simulation experiments, that uses a sequence of hypothesis tests to classify the factors as either important or unimportant. CSB controls the probability of Type I error for each factor, and the power at each bifurcation step, under heterogeneous variance conditions. CSB does, however, require the user to correctly state the directions of the effects prior to running the experiments. Experience indicates that this can be problematic with complex simulations.

We propose a hybrid two-phase approach, FF-CSB, to relax this requirement. Phase 1 uses an efficient fractional factorial experiment to estimate the signs and magnitudes of the effects. Phase 2 uses these results in controlled sequential bifurcation. We describe this procedure and provide an empirical evaluation of its performance.

1 INTRODUCTION

Screening experiments are intended to eliminate unimportant factors quickly, leaving a short list of important factors that can be studied in more detail via higher-resolution experimental designs. They are useful tools for examining simulation models that involve a large number of factors. The most well-known screening designs are saturated fractional factorials (Box et al. 1978, Montgomery 2000, NIST/SEMATECH 2005), but other screening methods have also been developed (e.g., Trocine and Malone 2001). Some procedures are specifically intended to facilitate large-scale experiments on simulation systems by taking advantage of the sequential nature of simulation experiments. Kleijnen et al. (2005) provide a general discussion—and numerous examples—of the design and analysis of simulation experiments. The challenge for those proposing sequential methods is establishing (either theoretically or empirically) the “correctness" of the screening results.

Group screening approaches can be efficient and practical when there are many factors but only a few important ones. The basic idea behind group screening is straightforward: if several factors can be aggregated into a group for testing, and the results indicate that this group of factors has no significant effect on the outcome, then all factors in the group can be eliminated from the list of potential important factors without further testing. Group screening has been used for years in physical experiments when tests are expensive, such as in screening a large number of new soldiers for syphilis during World War II in only a few tests (Dorfman 1943).

More recently, group screening has been proposed for simulation experiments. One such procedure is sequential bifurcation (SB), developed by Bettonvil and Kleijnen (1997) for deterministic simulation models. They assume important factors are sparse, that the direction of all effects is known, and that a main-effects metamodel is a reasonable approximation of the simulation response over the region of exploration. SB was extended to stochastic simulations by Cheng (1997), who assumes that the errors are normally distributed with constant variance and uses an indifference-zone approach to avoid excessive sampling for factors deemed unimportant. Kleijnen, Bettonvil, and

Persson (2005) also discuss the use of SB for experiments involving stochastic simulations. Examples and empirical investigations have shown that SB can be very efficient (i.e., require a relatively small number of runs) when the factor effects are sparse. Deterministic SB performs best if the factors are initially ordered according to increasing (or decreasing) values of the unknown factor effects. However, there are no theoretical guarantees of the performance, either in terms of the number of runs required or the probabilities of correct classification, in the stochastic case.

To address this shortcoming, Wan, Ankenman, and Nelson (2003, 2005a) propose a variant of SB called the controlled sequential bifurcation (CSB) procedure. In CSB, the analyst must specify two thresholds. The lower threshold (Δ_0) indicates the level the effect must reach to be considered *important*, while effects larger than the higher threshold (Δ_1) are considered *critical*. They also discuss a cost model which associates the thresholds and factor settings with a benchmark cost so the effectiveness of the screening procedure is not influenced by the sometimes arbitrary choices of thresholds and factor settings (Wan, Ankenman, and Nelson 2003, 2005a). CSB uses a hypothesis-testing approach to control the probability of Type I error (i.e., the probability an effect is classified as important when it is not) and power (i.e., the probability an important effect is correctly classified). Factors begin in a single group and the group's accumulated effect is tested. If the group's effect is classified as unimportant, then all factors within the group are classified as unimportant. Otherwise, the group is split into two smaller ones for further testing; if the group contains only one factor, this factor is classified as important. This procedure continues until all factors have been classified. Wan, Ankenman, and Nelson (2005a) provide proof of the CSB's performance even when the underlying variance is heterogeneous. Variance heterogeneity is a pervasive characteristic of large-scale simulation experiments.

One assumption of CSB (as for SB) is that the *direction* of the effects is known *a priori* so that factors with opposite effects are not included in the same group. This avoids the problem of full or partial cancellation of factor effects, which might cause the analyst to overlook one or more key factors. Unfortunately, for models of complex systems with several hundred factors, it may be unreasonable to expect that an analyst (or even a subject-matter expert) can correctly identify the signs of all potential factor effects. Experience has also shown that even experts may not be able to correctly identify the three to five most influential factors before the study commences: some factors may be more interesting that originally anticipated, while others thought to be important might not have significant effects on the response (Lucas et al. 2002).

In this paper, we propose a hybrid procedure for sequential screening. An efficient fractional factorial conducted in phase 1 is used to classify the factors into groups according to the signs and magnitudes of their estimated effects. This classification is the basis for applying sequential CSB in the second phase of the experiment. We describe the procedure in detail in Section 2, and provide an empirical evaluation of its performance in Section 3. Our results show that even if phase 1 simply classifies the factors as having negative or non-negative effects, rather than also making use of the magnitudes of the estimated effects, the hybrid procedure greatly reduces the possibility of erroneously concluding that important effects are unimportant because of incorrect groupings. The additional computational effort is minimal, so the hybrid procedure is a viable, efficient screening approach for simulation experiments even when little or nothing is known about the factor effects. Preliminary results also indicate that sorting the factors after phase 1 does not affect classification rates, but greatly improves the efficiency of the procedure. We finish with a discussion of issues for further research.

2 SCREENING PROCEDURE DESCRIPTIONS

2.1 CSB Procedure

We begin with a description of the CSB procedure. Suppose there are K factors of interest with effect coefficients $\beta_1, \ldots, \beta_K$. The output from a simulation replication is denoted by Y, and the underlying metamodel assumption for employing CSB is the main effects model:

$$Y = \beta_0 + \sum_{i=1}^{K} \beta_i x_i + \varepsilon, \tag{1}$$

where the ε's are distributed as $N\left(0, \sigma^2_{\mathbf{x}}\right)$ and may depend on the values of $\mathbf{x} = (x_1, \ldots, x_k)$. The settings of $\mathbf{x}$ are deterministic and are controlled by the analyst during the experiment. Note that the assumption of a main-effects model usually does not hold over the entire factor space, but it may be a reasonable assumption for, e.g., small variations in a region of interest. Wan, Ankenman, and Nelson (2005b) also proposed a version of CSB that gives unbiased screening results for main effects even if two-factor interactions exist, although the interactions effects are not themselves estimated.

The CSB procedure, like the SB procedure of Bettonvil and Kleijnen (1997) and the SB-under-uncertainty procedure of Cheng (1997), goes through a series of steps in which groups of factors are tested. If a group is determined to be important, then it is split into smaller groups for additional testing. If a group is determined to be unimportant, then all factors within that group are considered unimportant and need not be examined further. The procedure continues until each factor is classified as either important (i.e., a factor is in a group by itself and that group is determined to be important), or its group is classified as unimportant.

Table 1: Structure of CSB

Initialization:
Create an empty LIFO queue for groups. Add the group $\{1,\ldots,K\}$ to the LIFO queue.

While queue is not empty, do
Remove: Remove a group from the queue.
Test:
Unimportant:
If the group is unimportant, then classify all factors in the group as unimportant.
Important (size=1):
If the group is important and of size 1, then classify the factor as important.
Important (size>1):
If the group is important and the size is greater than 1, then split the group into two subgroups such that all factors in the first subgroup have smaller indices than those in the second subgroup. Add each subgroup to the LIFO queue.
End Test
End While

A general description of the algorithm appears in Table 1, adapted from Wan, Ankenman and Nelson (2003).

CSB is a screening procedure which guarantees that the probability of Type I error is less than α for any effect with $|\beta_i| < \Delta_0$, and that the power of detection is greater than γ for any effect with $|\beta_i| > \Delta_1$. Here α and γ are user-specified classification error bounds. The error control of CSB is determined by the error control of the hypothesis testing that occurs at each bifurcation step. Wan, Ankenman and Nelson (2005a, 2005b) show that as long as the basic hypothesis testing procedure is capable of guaranteeing a maximum probability of Type I error and a minimum power, CSB can guarantee the Type I error for each factor and power for each bifurcation step; they propose both a two-stage and a fully sequential version of CSB which satisfy the criteria. Both procedures will initially take a small number of observations (n_0, usually $n_0 \leq 5$). If no conclusions can be made, more observations are collected. The fully sequential testing procedure is typically more efficient since it takes one observation each time and will terminate as soon as the effect can be classified. Details of these two tests, and comparisons of their performances, appear in Wan, Ankenman, and Nelson (2005a, 2005b). We use the fully sequential version of CSB in this paper.

In CSB, the assumption that the signs of potential factor effects are accurately known before the experiment begins means that the factors (x_i's) associated with negative effects can be redefined to have positive effects. Thus, without loss of generality, it can be assumed that $\beta_1,\ldots,\beta_K$ are all nonnegative. In fact, the efficiency of the SB or CSB procedures is highest if the β_i's are ordered so that $\beta_1 \leq \beta_2 \leq \ldots \leq \beta_K$ or, equivalently, $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_K$. In reality, the directions of the effects can be unknown even for experts, especially for novel, complex systems where little prior knowledge exists. The hybrid procedure FF-CSB, discussed below, was developed to overcome this limitation.

2.2 FF-CSB Procedure

As Table 2 indicates, the FF-CSB procedure begins with a saturated or nearly-saturated fractional factorial experiment. We then explicitly divide the factors into two groups after phase 1 of experimentation according to their estimated effects $\hat{\beta}_i$ ($i = 1,\ldots,K$). Two separate groups are constructed: one contains all factors that yielded negative $\hat{\beta}_i$ during phase 1; the other contains all factors that yielded zero or positive $\hat{\beta}_i$. CSB is then performed separately on each of these two groups. At the end of phase 2, every one of the K factors will either be classified as *important* or as *unimportant*.

Note that the goal of phase 1 is not to obtain accurate estimates of the $\hat{\beta}$. If it did so, there would be no need for phase 2. However, even without ranking the estimated factor effects, the fractional factorial design conducted during phase 1 reduces the chance that two critical effects with opposite signs are included in the same group. The initial groups need not be of equal size, but instead reflect the preponderance of negative (or positive) $\hat{\beta}$'s. Because of the stochastic nature of the response, factors may sometimes be placed in the wrong initial group after phase 1.

3 EMPIRICAL PERFORMANCE EVALUATION

To assess the screening capabilities of the procedure, we conduct empirical experiments to compare the performance of FF-CSB with the original CSB for various values of K ($K = 2^m - 1$ for $m = 3,\ldots,9$). We fix the threshold for important factors $\Delta_0 = 2$, and the threshold for critical factors $\Delta_1 = 4$. The required maximum Type I error is specified

Table 2: Structure of FF-CSB

Initialization:
Create two empty LIFO queues for groups, NEG and POS.

Phase 1:
Conduct a saturated or nearly-saturated fractional factorial experiment and estimate $\hat{\beta}_1,\ldots,\hat{\beta}_k$. Order the estimates so that $\hat{\beta}_{[1]} \le \ldots \le \hat{\beta}_{[z]} < 0 \le \hat{\beta}_{[z+1]} \ldots \le \hat{\beta}_{[K]}$. Add factors $\{[1],\ldots,[z]\}$ to the NEG LIFO queue, and factors $\{[z+1],\ldots,[K]\}$ to the POS LIFO queue.

Phase 2:
For queue = POS and queue = NEG, do
While queue is not empty, do
Remove: Remove a group from the queue.
Test:
Unimportant:
If the group is unimportant, then classify all factors in the group as unimportant.
Important (size=1):
If the group is important and of size 1, then classify the factor as important.
Important (size>1):
If the group is important and the size is greater than 1, then split the group into two subgroups such that all factors in the first subgroup have smaller $[i]$'s (ordered indices) than those in the second subgroup. Add each subgroup to the LIFO queue.
End Test
End While
End For

to be $\alpha = 0.05$, the power requirement for critical effects is fixed at $\gamma = 0.95$, and the initial sample size for CSB is $n_0 = 5$. We assume that a main effects model suffices, and that the random errors are normally distributed with mean 0 and (common) variance 1.

Factor effect values β_i, $i = 1,\ldots,K$ are set as follows:

$$\beta_i = \begin{cases} (-1)\left(-5+10\left(\frac{i-1}{K-1}\right)\right) & \text{if } i < p \\ -5+10\left(\frac{i-1}{K-1}\right) & \text{otherwise,} \end{cases} \qquad (2)$$

for several values of $p \le (K+1)/2$.

If $p = 0$ then roughly half of the factor effects are negative. This is an extremely bad situation for CSB since the positive and negative effects will essentially cancel each other and CSB will conclude that most of the factors are not important. On the other hand, if $p = (K+1)/2$ then all factor effects are positive, and CSB will work well without adding the initial fractional factorial experiment. We also consider other values of p that correspond to intermediate situations for CSB. To facilitate comparisons, we let p be a function of K, rather than a constant. The five cases we consider will be referred to as follows:

- *none negative*: $p = (K+1)/2$,
- *small negative*: $p = 3(K+1)/8$,
- *medium negative*: $p = (K+1)/4$,
- *large negative*: $p = (K+1)/8$, and
- *half negative*: $p = 0$.

The negative effects are assigned to *smaller* values of β first, to reflect the possibility that subject-matter experts might be more likely to know the magnitude of critical factors (and so the factor levels could be set so that the corresponding β's were positive). Regardless of p, approximately 20% of all factors are critical, 40% are important (but not critical), and 40% are unimportant. This approximation is more accurate for larger K.

Ideally, the FF-CSB procedure will meet or exceed the probabilistic guarantees for CSB regardless of the signs of the β_i's. Since one of our motivations for this work was our belief that good indications about the signs and magnitudes of the effects might not be available before the experiment, we randomly reorder the initial values of $\beta_1,\ldots,\beta_k$ for each replication. 1000 replications are conducted for experiments with $K \le 127$, and 400 replications are conducted for experiments with $K = 255$ and $K = 511$.

We begin by investigating a simplification of the FF-CSB procedure, called *unsorted* FF-CSB, where the sampling during phase 1 is used only to classify the $\hat{\beta}_i$'s as negative or non-negative, rather than to rank them within these categories. This allows us to determine whether or not estimating the signs of the factor effects can, by itself,

Table 3: Performance of CSB and Unsorted FF-CSB Procedures for Randomly Ordered Factor Effects

Pattern of β values	K	CSB Correct Classification Proportions			CSB Avg. Runs	Unsorted FF-CSB Correct Classification Proportions			Unsorted FF-CSB Avg. Runs
		Critical	Important	Unimp.		Critical	Important	Unimp.	
None Negative	7	1.000	0.788	0.999	100	1.000	0.809	0.999	110
	15	1.000	0.425	1.000	248	1.000	0.432	1.000	268
	31	0.999	0.503	1.000	610	0.998	0.493	0.999	656
	63	1.000	0.492	1.000	1,488	0.999	0.490	1.000	1,563
	127	1.000	0.506	1.000	3,559	1.000	0.507	1.000	3,692
	255	1.000	0.495	0.000	8,192	1.000	0.497	1.000	8,704
	511	1.000	0.500	1.000	19,099	1.000	0.497	1.000	19,528
Small Negative	7*	1.000	0.788	0.999	100	1.000	0.809	0.999	110
	15	0.998	0.401	1.000	241	0.999	0.431	1.000	251
	31	0.991	0.458	1.000	581	0.999	0.498	0.999	605
	63	0.994	0.453	1.000	1,424	1.000	0.487	1.000	1,461
	127	0.993	0.469	1.000	3,428	1.000	0.506	1.000	3421
	255	0.992	0.460	1.000	7,824	1.000	0.496	1.000	8,024
	511	0.992	0.461	1.000	17,958	1.000	0.499	1.000	18,132
Medium Negative	7	0.972	0.671	1.000	92	1.000	0.804	0.999	100
	15	0.909	0.332	1.000	202	0.999	0.432	1.000	250
	31	0.879	0.384	1.000	491	0.999	0.502	0.999	586
	63	0.880	0.381	1.000	1,169	0.999	0.494	1.000	1,407
	127	0.879	0.393	1.000	2,811	1.000	0.508	1.000	3,307
	255	0.878	0.382	1.000	6,568	1.000	0.496	1.000	7,550
	511	0.877	0.386	1.000	15,361	1.000	0.499	1.000	17,703
Large Negative	7	0.744	0.297	1.000	66	1.000	0.785	0.999	100
	15	0.684	0.142	1.000	140	0.999	0.427	1.000	234
	31	0.664	0.164	0.999	339	0.997	0.495	0.999	558
	63	0.663	0.187	1.000	781	0.999	0.489	1.000	1,329
	127	0.662	0.201	1.000	1,827	1.000	0.505	1.000	3,171
	255	0.636	0.185	1.000	4,219	1.000	0.494	1.000	7,440
	511	0.661	0.200	1.000	10,041	1.000	0.499	1.000	17,595
Half Negative	7	0.000	0.000	1.000	10.7	1.000	0.770	1.000	99
	15	0.000	0.000	1.000	10.7	0.999	0.431	1.000	231
	31	0.000	0.000	1.000	10.6	0.998	0.494	0.999	556
	63	0.000	0.000	1.000	10.6	0.999	0.489	1.000	1,313
	127	0.000	0.000	1.000	10.7	1.000	0.507	1.000	3,143
	255	0.000	0.000	1.000	10.7	1.000	0.495	1.000	7,441
	511	0.000	0.000	1.000	10.6	1.000	0.499	1.000	17,370

*Same as the "none negative" case

yield a procedure that performs well without requiring the analyst to specify the directions of the factor effects before conducting the experiment.

The results are summarized in Table 3. The proportions of correct classifications for the critical, important, and unimportant factors are provided for both CSB and the unsorted FF-CSB procedures. Ideal values for all these proportions are 1.00. Table 3 also reports the average numbers of runs required, under various patterns of the underlying β_i values, for both the CSB and the FF-CSB procedures. If the procedure is able to meet or exceed the guaranteed classification probabilities, then a smaller average number of runs required indicates a more efficient procedure.

We first discuss the classification results for CSB. When the β's are all non-negative, CSB exceeds the probability and power specifications, as expected (since the β's are not in the least-favorable configuration for error and power calculations). CSB rarely misclassifies a critical factor as unimportant, or an unimportant factor as important, and it correctly classifies about half of the important, non-critical factors. CSB also performs well when only a small proportion of effects are negative, but its performance deteriorates rapidly as the number of negative effects increases. When roughly 25% of the effects are negative, the classification probabilities for critical factors drop to around 88%—significantly below the nominal value of 0.95 (p-value< 0.001); the classification probabilities for important factors also drop significantly from their values when all $\hat{\beta}_i \geq 0$ (p-value< 0.001). Perhaps the most striking result from Table 3 is that CSB is completely unsuccessful at classifying important factors when half the β's are negative. Not once in the 5800 trials is *any* factor classified as important.

Next, consider how FF-CSB performs in terms of classifying factors. When all the β_i are non-negative, its classification probabilities are indistinguishable from that of CSB (p-values> 0.50). The classification rates are insensitive to the proportion of negative factors in the study; regardless of the initial pattern of the β_i, FF-CSB correctly identifies essentially all the critical and unimportant factors, and about half of the important, non-critical factors. FF-CSB may place some factors (particularly unimportant ones) in the wrong initial group. For example, 13% of the 5800 experiments involving only non-negative $\hat{\beta}_i$'s classify two or more factors as having negative effects after phase 1. Nonetheless, FF-CSB appears to be a procedure for which only the magnitudes (not the signs) of the factor effects influence its classification rates.

Table 3 also provides the average number of runs required to complete the experiment. Note that for the hybrid procedure, the runs include both the phase 1 sampling ($K+1$) and the phase 2 sampling (using CSB). Clearly, experiments involving more factors require a greater number of runs. The effects are not sparse, so both procedures require substantial sampling to completely classify the factors in the cases where they correctly identify at least 95% of the critical factors (the none negative and small negative situations for CSB, and all situations for FF-CSB). FF-CSB takes slightly more samples than CSB when all $\beta_i > 0$, but the results for FF-CSB also indicate a small, but statistically significant decrease in the average number of runs as the proportion of negative effects increases. For example, if half of the effects are negative and $K = 511$, using FF-CSB requires 9% less sampling, on average, than is needed if all factor directions could be accurately determined. This might at first appear counter-intuitive, since one would expect not to improve on the performance when all factors are known to have non-negative effects. However, the fractional factorial does impose a partial ordering on the factor effects which may account for the improved performance. The expected range of β_i within the NEG group is less than the range for CSB ($\max(\beta_i) - \min(\beta_i)$).

The variability in the number of runs required is also a useful measure of FF-CSB's performance, since an analyst running a single experiment might be interested in how likely it will be to take a extremely long time to finish. In all cases, the coefficient of variation (CV, equal to the ratio standard deviation / mean) associated with the total runs ranges from 0.15 to 0.33, with an average of 0.26 (values associated with the $K = 7$ range from 0.38 to 0.48).

We now examine the effects of sorting in more detail. The results in Table 3 are intended to demonstrate the potential effectiveness of FF-CSB, relative to CSB. Sequential bifurcation is known to be most efficient when the proportion of important and critical effects is lower than the cases in Table 3. We conduct another set of experiments to assess the impact of sorting the $\hat{\beta}_i$'s (over and above the impact of determining their direction), in a situation more favorable to CSB. In this set of experiments, we take the β_i's from equation (2) (with $p = 0$) and modify them as follows:

- critical effects are set to -5 or +5, according to the sign of the original β_i;
- all other effects (both important and unimportant) are set to zero.

All other conditions, such as the standard deviations and the random ordering of the β's prior to each replication, remain the same. Once again, 1000 replications are made for $K = 15, 31, 63$ and 127, and 400 replications are made for $K = 255$ and 511. (We do not consider $K = 7$ since it would have only a single critical negative and a single critical positive effect, so sorting will not influence the results.) The classification results for these experiments indicate that both the sorted and unsorted version of the FF-CSB procedure correctly identify all effects in over 99.9% of the cases, easily exceeding the Type I error and power

requirements. The efficiency results, both with and without sorting the $\hat{\beta}$'s after phase 1, appear in Table 4. The relative efficiency of the sorted FF-CSB to the unsorted FF-CSB is also provided; values less than 1.00 indicate that the sorted FF-CSB procedure is more efficient.

Table 4: Efficiency of Unsorted and Sorted FF-CSB

	Average No. of Runs		Relative
No. of Factors	Unsorted	Sorted	Efficiency
K	N_U	N_S	N_S/N_U
15	105	82	0.79
31	212	137	0.64
63	420	230	0.55
127	865	425	0.49
255	1,908	849	0.44
511	4,159	1,761	0.42

The benefits of sorting are apparent from Table 4. The sorted FF-CSB requires no more than 79% of the data of the unsorted FF-CSB when $K = 15$, and improves to 42% as the number of factors increases. This improvement in efficiency occurs because the critical effects tend to be grouped closer together at the beginning of phase 2, so large groups of unimportant factors can be eliminated in early bifurcation steps. The coefficients of variation range from approximately 0.20 (for $K = 15$) to 0.10 (for $K = 511$). This means that the sorted FF-CSB procedure is not only more efficient, but also has less variation in the number of runs required, and that the standard deviation of the number of runs required grows more slowly than the mean as the number of factors increases.

4 DISCUSSION

These results are the initial part of a larger empirical study investigating the how CSB performs under various conditions listed below:

- Δ_0: the threshold below which effects are considered unimportant;
- Δ_1: the threshold above which effects are considered critical;
- other patterns of standard deviations, since variance heterogeneity is pervasive in complex simulations; and
- other patterns of β's, including different proportions of critical and important factors.

The underlying β's used in our study are small enough (relative to the error variances) that the fractional factorial experiment conducted in phase 1 is unlikely to definitively identify any effects as important. For situations where a few critical factors dominate the results, some factors might be classified at the end of phase 1, or at least be separated from other factors (i.e., placed in their own initial groups) for phase 2 testing. The procedure's performance when a less-saturated fractional factorial experiment is used during phase 1 (i.e., $K+1$ is not a power of 2) is also of interest.

Currently, we are expanding the empirical investigation to better understand the performance of the sorted FF-CSB over a broader range of conditions. We are also exploring better ways to utilize the information from phase 1, such as other ways of handling effects that are obviously important (or unimportant) after phase 1.

5 CONCLUSIONS

Group screening approaches have the potential to provide valuable information to analysts exploring complex simulation models. Yet, to be truly useful, the methods should be applicable to a broad range of simulation studies while requiring few assumptions about the simulation model's performance. The CSB procedure has been shown to control the probability of Type I error for each factor, as well as the power of detecting critical factor effects, for stochastic simulations with heterogeneous variance. However, it does require that the direction of all factor effects be known before experimentation begins. The new FF-CSB procedure overcomes the limitation of CSB that the signs of factor effects have to be known beforehand. FF-CSB combines a fractional factorial design with the CSB procedure, and the resulting hybrid method can effectively screen mixed positive and negative main effects. The procedure is most efficient when the factors are sorted by their estimated effects after phase 1.

A major benefit is that the improvement in efficient classification will not depend on accurate subject-matter expertise regarding the directions and magnitudes of effects for a large number of factors, so the gains in efficiency are likely to be realized for practical applications. This makes it a more flexible and useful tool for analysts who seek to explore simulation models when they have little information about the nature of its response surface. Modifications to FF-CSB that make better use of the results from phase 1 of the study are currently under investigation. The resulting procedure will be even more efficient and adaptive.

ACKNOWLEDGMENTS

This work was supported in part by the U.S. Marine Corps' Project Albert. The authors also wish to express their appreciation to Jack Kleijnen, on the occasion of his retirement, for his foundational work in sequential bifurcation.

REFERENCES

Bettonvil, B. and J. P. C. Kleijnen. 1997. Searching for important factors in simulation models with many factors: sequential bifurcation. *European Journal of Operational Research* 96: 180-194.

Box, G. E. P., W. G. Hunter, and J. S. Hunter. 1978. *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*. New York: Wiley.

Campolongo, F., J. P. C. Kleijnen, and T. Andres. 2000. Screening methods. In *Sensitivity Analysis*, eds. A. Saltelli, K. Chan, and E. M. Scott, 65–89. New York: Wiley.

Cheng, R. C. H. 1997. Searching for important factors: sequential bifurcation under uncertainty. In *Proceedings of the 1997 Winter Simulation Conference*, eds. S. Andradottir, K. J. Healy, D. H. Withers, and B. L. Nelson, 275–280. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available via <http://www.informs-sim.org/wsc97papers/0275.PDF> [accessed July 1, 2005].

Cioppa, T. M., T. W. Lucas, and S. M. Sanchez. 2004. Military applications of agent-based models. In *Proceedings of the 2004 Winter Simulation Conference*, eds. R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 171–180. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available via <http://www.informs-sim.org/wsc04papers/020.pdf> [accessed July 1, 2005].

Dorfman, R. 1943. The detection of defective numbers of large populations. *The Annals of Mathematical Statistics* 14: 436-440.

Hedayat, A., J. Sloane, and N. J. A. Stufken. 1999. *Orthogonal Arrays: Theory and Applications*. New York: Springer.

Kleijnen, J. P. C., B. Bettonvil, and F. Persson. 2005. Finding the important factors in large discrete-event simulation: sequential bifurcation and its applications. In *Screening*, eds. A. M. Dean and S. M. Lewis. Screening. New York: Springer-Verlag, forthcoming.

Kleijnen, J. P. C., S. M. Sanchez, T. W. Lucas, and T. M. Cioppa. 2005. A user's guide to the brave new world of simulation experiments. *INFORMS Journal on Computing* 17(3), 263–289.

Law, A. M. and W. D. Kelton. 2000. *Simulation Modeling and Analysis*. 3rd ed. New York: McGraw-Hill.

Lucas, T. W., S. M. Sanchez, L. Brown, W. Vinyard. Better designs for high-dimensional explorations of simulations. In *Maneuver Warfare Science 2002*, eds. G. Horne and S. Johnson, 17–46. Quantico, Virginia: USMC Project Albert.

Montgomery, D. C. 2000. *Design and Analysis of Experiments*. 5th ed. Wiley, New York.

NIST/SEMATECH. 2005. *e-Handbook of Statistical Methods*. Available via <http://www.itl.nist.gov/div898/handbook/> [accessed July 1, 2005].

Trocine L. and L. C. Malone. 2001. An overview of newer, advanced screening methods for the initial phase in an experimental design. In *Proceedings of the 2001 Winter Simulation Conference*, eds. B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 169-178. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available via <http://www.informs-sim.org/wsc01papers/020.PDF> [accessed July 1, 2005].

Wan, H., B. Ankenman, and B. L. Nelson. 2003. Controlled sequential bifurcation: a new factor-screening method for discrete-event simulation. In *Proceedings of the 2003 Winter Simulation Conference*, eds. S. Chick, P. J. Sanchez, D. Ferrin, and D. J. Morrice, 565-573. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available via <http://www.informs-sim.org/wsc03papers/070.pdf> [accessed July 1, 2005].

Wan, H., B. Ankenman, and B. L. Nelson. 2005a. Controlled sequential bifurcation: a new factor-screening method for discrete-event simulation. *Operations Research*, forthcoming.

Wan, H., B. Ankenman, and B. L. Nelson. 2005b. Simulation factor screening with controlled sequential bifurcation in the presence of interactions. Working paper, School of Industrial Engineering, Purdue University, West Lafayette, Indiana.

AUTHOR BIOGRAPHIES

SUSAN M. SANCHEZ is a Professor of Operations Research at the Naval Postgraduate School, where she holds a joint appointment in the Graduate School of Business and Public Policy. Her research interests include experimental design, data-intensive statistics, and robust selection. She has a Ph.D. in Operations Research from Cornell University. Sanchez has served the simulation community in several roles, including the transition of the INFORMS College on Simulation to the INFORMS Simulation Society during her term as President of the College. She is currently the Simulation Area Editor for the *INFORMS Journal on Computing* and the ASA representative to the WSC Board of Directors. Her e-mail and web addresses are <ssanchez@nps.edu> and <http://www.nps.navy.mil/orfacpag/resumePages/sanchs.htm>.

HONG WAN is an Assistant Professor in the School of Industrial Engineering at Purdue University. Her current research is in factor-screening methods for discrete-event simulation. The other projects in which she has partic-

ipated include healthcare management, industrial product design and quality control; and software for strategic management of transportation systems. Her research interests include experimental design and quality control, simulation of complicated systems, and industrial statistics. She has taught a variety of courses and is a member of INFORMS and ASA. Her e-mail and web addresses are <hwan@purdue.edu> and <http://web.ics.purdue.edu/^hwan/index.htm>.

THOMAS W. LUCAS is an Associate Professor in the Department of Operations Research at the Naval Postgraduate School. His research interests include robust Bayesian statistics, design of experiments, and combat modeling. He has a Ph.D. in Statistics from the University of California at Riverside. His e-mail address is <twlucas@nps.edu> and his web page is <http://www.nps.navy.mil/orfacpag/resumePages/lucas.htm>.