Stochastics and Statistics

Controlled sequential factorial design for simulation factor screening

Hua Shen\textsuperscript{a,1}, Hong Wan\textsuperscript{b,*}

\textsuperscript{a} Wachovia Corporation, Charlotte, NC 28288-0376, USA
\textsuperscript{b} School of Industrial Engineering, Purdue University, West Lafayette, IN 47907-2023, USA

\textbf{A R T I C L E   I N F O}

Article history:
Received 10 April 2007
Accepted 2 September 2008
Available online 12 September 2008

Keywords:
Simulation
Computer experiments
Design of experiments
Factor screening
Sequential factorial design

\textbf{A B S T R A C T}

Screening experiments are performed to eliminate unimportant factors efficiently so that the remaining important factors can be studied more thoroughly in later experiments. This paper proposes controlled sequential factorial design (CSFD) for discrete-event simulation experiments. It combines a sequential hypothesis testing procedure with a traditional (fractional) factorial design to control the Type I error and power for each factor under heterogeneous variance conditions. We compare CSFD with other sequential screening methods with similar error control properties. CSFD requires few assumptions and demonstrates robust performance with different system conditions. The method is appropriate for systems with a moderate number of factors and large variances.

© 2008 Elsevier B.V. All rights reserved.

\textbf{1. Introduction}

Stochastic simulation is one of the most widely used techniques for operations research and management science. Here the stochastic simulation refers to “the analysis of stochastic processes through the generation of sample paths (realizations) of the processes” (Henderson and Nelson, 2006). Simulation experiments are typically faster, cheaper, and more flexible than physical experiments. They are especially useful for pilot studies of complicated systems where the physical experiments are too expensive or time-consuming.

Despite the prevalence of simulation as a decision-support tool, there is a huge gap between the complexity of the systems under investigation and the experimental designs used to study these methods. Simulation models may take months or years to develop and have literally thousands of components (factors) for which the potential effects on system performance are unknown. Screening experiments in this case are designed to quickly eliminate these unimportant factors for a more compact model and faster, more transparent analysis.

Many screening methodologies have been developed to identify important factors with an economical number of observations. The most common ones are fractional factorial, central composite, and Plackett–Burman designs (Myers and Montgomery, 2002). However, these screening designs were developed for physical experiments. They typically involve fewer than 25 factors and do not take advantage of the sequential property of simulation experiments. Procedures developed for stochastic simulation experiments include one-factor-at-a-time designs (Campolongo et al., 2000); methods based on frequency domain analysis (Morrice and Bardhan, 1995); edge designs (Elster and Neumaier, 1995); iterated fractional factorial designs (Campolongo et al., 2000) and the Trocine screening procedure (Trocine and Malone, 2001). Kleijnen et al. (2005) gives a general review of the design and analysis of simulation experiments. Unfortunately, the analysis methods they have used typically assume equal variances across different factor settings – a rare occurrence for simulations of complex systems. More importantly, none of these designs for stochastic simulations give the error control we desire.

Controlled sequential bifurcation (CSB) (Wan et al., 2003, 2006) method is proposed to stress these problems. It combines a hypothesis testing procedure with the sequential bifurcation framework proposed by Bettonvil and Kleijnen (1997). CSB is a series of steps in which groups of factors are tested. If a group of factors is considered unimportant, then every factor in the group will be considered unimportant. If the group is considered important, then it is split (bifurcated) for further testing. Each factor will eventually be classified as important or unimportant. Whether the desired error control can be achieved depends on the statistical test employed at each step. Given an appropriate testing procedure, CSB can control the Type I error of each factor and power for each bifurcation step under heterogeneous variance conditions. The sequential nature of the method makes it a good fit for simulation experiments. CSB-X incorporates a fold-over design into the CSB framework to eliminate the effects of second-order terms and give unbiased screening results of main-effects (Wan et al., 2008). Examples show that CSB (and CSB-X) are highly efficient for large-scale problems...
when important factors are of small percentage and clustered, since it can eliminate unimportant factors in groups. On the other hand, the sequential bifurcation framework determines that CSB is a "one-at-a-time" design. Simulation replications generated in previous design points may not be useful in the later screening stages and new simulation replications are usually needed after each bifurcation step. This makes the method inefficient in many scenarios. In addition, the CSB method has to assume known direction of effects, which is not realistic in many scenarios.

Wan and Ankenman (2006) propose a different approach to controlled screening, namely two-stage controlled fractional factorial screening (TCFF). The user selects an appropriate (fractional) factorial design and collects a small number of observations at each design point in the first stage. More observations will be collected for those design points with large variances in the second stage and the weighted average of all observations generated at each design point is considered as a single pseudo-observation following a non-central t-distribution. The weight is the number of observations at each design point. The design is then treated as an unreplicated (fractional) factorial design. Because of the use of factorial designs, TCFF can test interaction effects and does not require the directions of the effects to be known. TCFF and CSB-X are tested under different scenarios. The results show that CSB-X requires fewer replications when only 1% of factors are important while the TCFF method has better efficiency when there are no less than 5% of the factors important. TCFF is also more robust in efficiency across different scenarios.

In this paper, we propose a different approach of sequential factorial design for stochastic simulation experiments. We call it controlled sequential factorial design (CSFD). CSFD combines sequential hypothesis testing procedures with a full or fractional factorial design to provide simultaneous Type I error and power control for screening results under heterogeneous variance conditions. Unlike CSB methods, which usually need to generate new observations at each bifurcation step, CSFD can utilize all previously generated observations in later screening stages. In most cases, after the first few effects are classified, no more simulation runs are needed. Different from TCFF, the computational effort of CSFD is equally allocated to all design points. This seemingly inefficient approach turns out to be superior in many situations; numerical results show that CSFD is more effective than TCFF and CSB methods in many cases.

The paper is organized as follows: The underlying response model and the objective of screening are discussed in Section 2. Section 3 describes CSFD method and its performance in detail. Section 4 presents empirical evaluations of CSFD compared to CSB-X and TCFF. In Section 5, CSFD is implemented in a semiconductor manufacturing system. Conclusions and future research are presented in Section 6.

2. Model description

Suppose there are in total of \( L \) factors. A general metamodel including all main-effects and interactions is given as follows:

\[
Y = \beta_0 + \sum_{l=1}^{L} \beta_l z_l + \sum_{l_1,l_2 < L} \beta_{l_1l_2} z_{l_1} z_{l_2} + \cdots + \beta_{12\cdots L} z_1 z_2 \cdots z_L + \epsilon.
\]

Here \( \hat{\beta} = \{\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{12\cdots L}\} \) are the effect coefficients (\( l \) represents "transpose"). The level settings, \( z = (z_1, z_2, \ldots, z_L) \), are assumed to be deterministic. The error term, \( \epsilon \), on the other hand, is a random variable; in this paper we assume it is \( \text{Nor}(0, \sigma^2(z)) \), where \( \sigma^2(z) \) is unknown. In practice, the model may include any subset of the effects. Usually if an interaction exists, the main effects and lower-order interactions of all factors involved should also be included.

The objective of our screening procedure is to classify interested effects into two groups, important ones and unimportant ones. For effects with coefficients \( \leq \beta_0 \), CSFD should have \( \leq \alpha \) probability to declare them as important; for effects with coefficients \( \geq \beta_1 \), the power of identifying them as important should be \( \geq \gamma \). Those effects whose coefficients fall between \( \beta_0 \) and \( \beta_1 \) are considered important and we want CSFD to have reasonable, though not guaranteed, power to identify them. The parameters \( \beta_0 \) and \( \beta_1 \) are the thresholds of importance and criticality, respectively, with \( \beta_1 > \beta_0 \). Specifically, \( \beta_0 \) is the minimum change in the expected response that is practically important, and \( \beta_1 \) is a change in the expected response that we would not want to miss. The \( \alpha \) and \( \gamma \) are user-specified parameters. In practice, when we consider whether a change in the response is worth pursuing, the cost to produce the change is meaningful. Wan et al. (2003, 2006) proposed a cost model to determine the thresholds and factor settings so that all effects can be compared with thresholds without ambiguity. The selection of levels and thresholds, on the other hand, will not influence the performance of CSFD. After determining the levels of factors, CSFD will then code them from \(-1 \) to \(+1 \) (Montgomery, 2005).

3. Controlled sequential factorial design (CSFD)

The first step of CSFD is to select a full or fractional factorial design that will be sequentially implemented. Factorial designs are one of the most widely used experiment designs in practice (see, e.g., Montgomery, 2005). Full factorial designs contain all combinations of factor settings and can independently estimate all main effects and interactions. Fractional factorial designs contain a subset of all combinations of factor settings and some effects’ estimates will be confounded. If only main-effects and low-order interactions are of interests, fractional factorial designs are sufficient. In this research, we focus on \( 2^L \) full factorial designs and \( 2^{L/2} \) fractional factorial designs, i.e., each factor has two levels. A factorial design is characterized by its design matrix \( X \). Each row of the design matrix specifies an experiment setting of each factor, and each column specifies the settings of the specific effect across all experiments. The resolution of a fractional factorial design characterizes the degree of confounding. For example, a Resolution III design can independently estimate all main-effects, but some main effects are confounded with two-factor interactions. A Resolution V design has all main-effects and two-factor interaction estimators independent of each other. The higher the resolution, the more design points are required.

The factorial design is analyzed through regression technique. For a linear regression model \( Y = X\beta + \epsilon \), where \( X \) is the design matrix, \( Y = (y_1, y_2, \ldots, y_m) \) is the observation vector, \( \epsilon \) is the error vector, and the least square estimator of the effect coefficients is \( \hat{\beta} = (X'X)^{-1}X'Y \). For a full factorial design, all main and interaction effects can be estimated independently; for a fractional factorial design, some of the effects are confounded with others.

Many books (e.g., Wu and Hamada, 2000) and software packages provide recommended designs of different resolutions for 3–25 factors. For large-scale cases, Resolution III fractional factorial designs can be easily constructed. For a \( L \)-factor main-effects model, we will simply use a \( m \)-factor full factorial design with \( 2^m \) design points and confound the \( m + 1 \) to \( L \) factors’ main-effects with the interactions of the first \( m \) factors. Here, \( m \) is an integer satisfying \( 2^{m-1} \leq L < 2^m \) (Montgomery, 2005). Wan and Ankenman (2006) described the methods to construct Resolution IV designs. For constructing large-scale Resolution V designs, please see Sanchez and Sanchez (2005).

3.1. CSFD procedure

The generic structure of CSFD methods is presented in Table 1. Random observations are generated in batches and each batch is
Algorithm of CSFD

**Initialization:**

Create an empty group for effects of interest.
Add the effects \( \{b_1, b_2, \ldots, b_k\} \) to the group.
Select a factorial design with design matrix \( X \).
Generate \( n_0 \) replications of observations
\( \hat{b}(j) = (X'X)^{-1}X'Y(j), j = 1, 2, \ldots, n_0; n_1 - n_2 = \ldots = n_{K} - n_0; N = n_0 \)
While the group is not empty, do
Remove: remove an effect \( b_j \) from the group, \( n_1 - n_0 \).
**Qualified testing procedure:** calculate necessary statistics of \( \hat{b}_j \). Collect \( n \) new replications until specific error control requirements are achieved for \( \hat{b}_j \) (Table 2).
**Update:** Calculate \( \hat{b}(N + j) = (X'X)^{-1}X'Y(N + j), j = 1, 2, \ldots, n; N = n + n \).
Classify all effects in the group that can achieve the specific error control with \( N + n \).
End While

The qualified hypothesis procedure determines the number of the replications required to achieve desired error control (in other words, the stopping rules of the sequential factorial design). More detail is discussed in the following section.

### 3.2. Performance of CSFD

The process to classify the importance of desired effects is actually to sequentially test the following hypotheses:

\[
H_0 : |\hat{b}_k| \leq \Delta_0 \quad \text{vs} \quad H_1 : |\hat{b}_k| > \Delta_0.
\]

The selection of the testing procedure has a significant impact on the effectiveness and efficiency of CSFD. Wan et al. (2006) introduced the concept of a Qualified test. A testing procedure is called “qualified” if it guarantees that

- \( \Pr \{\text{ Declare an effect important } | \text{ effect } \leq \Delta_0 \} \leq \alpha, \)
- \( \Pr \{\text{ Declare an effect important } | \text{ effect } \geq \Delta_1 \} \geq \gamma. \)

Using this concept, Proposition 1 is straightforward:

**Proposition 1.** Given a qualified testing procedure, CSFD guarantees that

\[
\Pr \{\text{ Declare effect } k \text{ important } | |\hat{b}_k| \leq \Delta_0 \} \leq \alpha, \quad \text{and} \quad \Pr \{\text{Declare effect } k \text{ important } | |\hat{b}_k| \geq \Delta_1 \} \geq \gamma, \quad \text{for any } k = 1, 2, \ldots, K.
\]

In summary, given a qualified testing procedure, CSFD provides simultaneous Type I error and power control for each effect while CSB only provides step-wise power control. From this perspective, CSFD is superior.

We use a specific qualified fully sequential test proposed by Wan et al. (2006) and Wan et al. (2008). The procedure takes a small number of observations at the initial stage, and then adds one replication at a time and terminates as soon as a conclusion can be made. Numerical results show that the test is highly efficient in many cases. The structure of the fully sequential testing procedure is given in Table 2 along with the notations below:

- \( \hat{b}_1 = \frac{\sum_{i=1}^{n} \hat{b}_i(j)}{n}. \)
- \( S_b^2 = \frac{\sum_{j=1}^{n} (\hat{b}_j(j) - \hat{b}_1)^2}{n_0 - 1}; \) Sample variance \( S_b^2 \) is computed based on the first \( n_0 \) replications. \( S_b^2 \) will not be updated when more replications are generated.
- \( \lambda = (\Delta_1 - \Delta_0)/4. \)
- \( a(k) = a_0 S_b^2; \)
- \( M(k) = [a(k)/\lambda]. \)

There is no close-form solution for \( a_0 \) and \( r_0 \) in general cases and the values of these two constants have to be obtained numerically.

### Table 1
Algorithm of CSFD

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Set ( n = 0 ).</td>
</tr>
</tbody>
</table>
| 2.   | If \( N > M(k) \),
      | \( a(k) \) if \( |\hat{b}| < r_0 \), then classify the effect as unimportant.
      | \( a(k) \) else classify the effect as important. |
| 3.   | Else (i.e., \( N + n \) \( \leq M(k) \)),
      | \( a(k) \) if \( N + n |\hat{b}| - r_0 \) \( \leq a(k) \), then classify the effect as unimportant.
      | \( a(k) \) else if \( N + n |\hat{b}| - r_0 \) \( > a(k) \), then classify the effect as important. |
|      | \( a(k) \) else generate one more replication; update \( \hat{b} \); \( n = n + 1 \); go to (2). |

### Table 2
Fully sequential testing procedure

- **Step 1:** Set \( n = 0 \).
- **Step 2:** If \( N > M(k) \),
  - If \( |\hat{b}| < r_0 \), then classify the effect as unimportant.
  - Else classify the effect as important.
- **Step 3:** Else (i.e., \( N + n \) \( \leq M(k) \)),
  - If \( N + n |\hat{b}| - r_0 \) \( \leq a(k) \), then classify the effect as unimportant.
  - Else if \( N + n |\hat{b}| - r_0 \) \( > a(k) \), then classify the effect as important.
  - Else generate one more replication; update \( \hat{b} \); \( n = n + 1 \); go to (2).
We can easily see that $\eta = (\exp(\varphi) - 1)/2$, where $\varphi = -2\ln(2\pi)/(n_0 - 1)$.

Again, since we assume that the responses at each design point (for example, the batch means) follow an *i.i.d.* normal distribution (i.e., the response vectors of the $n$ replications are *i.i.d.* multivariate normal random variables), $\hat{\beta}_k(j)$ for any $j, k$, which is a linear combination of the responses of all design points, is also normally distributed; and $\hat{\beta}_1(1), \hat{\beta}_2(2), \ldots, \hat{\beta}_k(n_0)$ are *i.i.d.* normal random variables for any $k$. Therefore, all the proofs of the qualification of the fully sequential testing procedures in Wan et al. (2006) and Wan et al. (2008) are also valid for CSFD.

### 3.3. Illustrated example

To illustrate the procedure, let’s look at a toy problem with only two factors $x_1$ and $x_2$. The interested effect coefficients are $\beta_1$, $\beta_2$, and $\beta_{12}$, and their true values are 2, 3 and 0.5, respectively (for simplicity, we set $\beta_0 = 0$). We further set $A_0 = 2$ and $A_1 = 3$. A full factorial design with 4 runs are selected with the design matrix in Table 3.

We assume the standard deviation at each design point equals $1 + 0.5$ response. More specifically, the response at design point $m$ equals

$$y_m = \hat{\beta}_1 x_{1m} + \hat{\beta}_2 x_{2m} + \hat{\beta}_{12} x_{1m} x_{2m} + \epsilon_m,$$

where $\epsilon_m$ follows a normal distribution with mean 0 and standard deviation $1 + 0.5|\beta_1 x_{1m} + \beta_2 x_{2m} + \beta_{12} x_{1m} x_{2m}|$, with $m = 1, 2, 3, 4$.

We can easily see that $\hat{\beta}_1(j) = |y_1 - y_3 - y_2 + y_4|/4$, $\hat{\beta}_2(j) = |y_1 - y_2 - y_3 + y_4|/4$, and $\hat{\beta}_{12}(j) = |y_1 - y_2 - y_3 + y_4|/4, j = 1, 2, \ldots, N$. We pick $n_0 = 3$, and simulated the responses. The initial $Y = |y_4|$ are listed in Table 4 and the accordingly estimated effects are listed in Table 5.

### Table 3

$2^2$ Full factorial design

<table>
<thead>
<tr>
<th>Run</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>4</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
</tr>
</tbody>
</table>

### Table 4

Simulated responses

<table>
<thead>
<tr>
<th>Design points</th>
<th>Rep 1</th>
<th>Rep 2</th>
<th>Rep 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.76</td>
<td>-0.47</td>
<td>1.28</td>
</tr>
<tr>
<td>2</td>
<td>-2.66</td>
<td>-1.03</td>
<td>-2.90</td>
</tr>
<tr>
<td>3</td>
<td>-2.69</td>
<td>0.49</td>
<td>-5.52</td>
</tr>
<tr>
<td>4</td>
<td>1.04</td>
<td>1.55</td>
<td>3.24</td>
</tr>
</tbody>
</table>

### Table 5

Estimated effects of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_{12}$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>Rep 1</th>
<th>Rep 2</th>
<th>Rep 3</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.46</td>
<td>0.13</td>
<td>1.15</td>
<td>0.52</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1.41</td>
<td>0.41</td>
<td>3.23</td>
<td>1.43</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.45</td>
<td>0.88</td>
<td>-0.17</td>
<td>0.53</td>
</tr>
</tbody>
</table>

We can see that the estimations can be significantly different from the true values. Given $\alpha = 1 - \gamma = 0.05$, we can calculate the critical values $a(k)$ of the fully sequential test for $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_{12}$ (Table 6). For all three effects, $r_0 = 2.5$ and $\lambda = 0.33$. The fully

![Fig. 1. Screening of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_{12}$](image-url)
sequential test monitors whether the partial summation $\sum_{j=0}^{r-1}(\hat{\beta}_j - \tau)$ drifts out of a triangular region with upper bound $= a(k) - x\tau$ and lower bound $= -a(k) + x\tau$, where $r$ is the sample size. For this specific example, we can verify that no conclusion can be made for any of the three effects with the initial three replications. Therefore the experiments continued. Each replication contains three observations, one from each design point. It turns out that $\beta_1$ is classified as unimportant with six replications (three after the initial $n_0$ observations); $\beta_2$ is classified as important with 28 replications, and $\beta_{1,2}$ is classified as unimportant with four replications. The boundaries and partial summations are demonstrated in Fig. 1. This example also shows that the sequence of inspecting effects are irrelevant. For example, if $\beta_2$ is the first one to be classified, then no replications will be collected for classifying $\beta_1$ and $\beta_{1,2}$.

4. Empirical evaluation

In this section, we present the numerical results of artificial examples to compare CSFD, TCFF and CSB (CSB-X) methods. Because the performance of CSB-X is usually as good as, if not better than, CSB (Wan et al., 2008), we will only run CSB-X in the comparison. The fully sequential test is used for both CSB-X and CSFD. In all cases, normal errors are assumed with mean 0 and standard deviation $\sigma = a \cdot (1 + |E[Y]|)$, i.e., the standard deviation is proportional to the expected response. The variance parameter $a$ takes different values at different cases. Common random numbers were not employed (the effect of CRN is discussed in Section 4.3). For each case considered, all three methods are repeatedly applied 1000 times and the percentage of times each effect is classified as important is recorded, which is denoted by $P(\text{DI})$; for the kth effect, this percentage is an unbiased estimate of $Pr(\text{Declare effect } k \text{ important})$. We use “$\#$ of runs” to represent the average number of simulation runs required for each method.

### Table 7

Parameters for small-scale case with interactions

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_1$</td>
<td>10</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>2</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>4</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.05</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.95</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.1, 0.3, 1</td>
</tr>
</tbody>
</table>

### Table 8

Small-scale case with 2nd-order interactions

<table>
<thead>
<tr>
<th>Effect</th>
<th>$a = 0.1$</th>
<th>$a = 0.3$</th>
<th>$a = 1.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CSFD</td>
<td>TCFF</td>
<td>CSB-X</td>
</tr>
<tr>
<td>$\beta_1 = 0.2$</td>
<td>0.000</td>
<td>0.058</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_2 = 0.2$</td>
<td>0.000</td>
<td>0.042</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_1 = 2.4$</td>
<td>0.000</td>
<td>0.167</td>
<td>0.006</td>
</tr>
<tr>
<td>$\beta_2 = 2.7$</td>
<td>0.000</td>
<td>0.303</td>
<td>0.174</td>
</tr>
<tr>
<td>$\beta_1 = 3.0$</td>
<td>0.495</td>
<td>0.486</td>
<td>0.494</td>
</tr>
<tr>
<td>$\beta_2 = 3.3$</td>
<td>1.000</td>
<td>0.696</td>
<td>0.751</td>
</tr>
<tr>
<td>$\beta_1 = 3.6$</td>
<td>1.000</td>
<td>0.858</td>
<td>0.897</td>
</tr>
<tr>
<td>$\beta_2 = 3.8$</td>
<td>1.000</td>
<td>0.901</td>
<td>0.953</td>
</tr>
<tr>
<td>$\beta_3 = 4.0$</td>
<td>1.000</td>
<td>0.951</td>
<td>0.968</td>
</tr>
<tr>
<td>$\beta_1 = 4.2$</td>
<td>1.000</td>
<td>0.977</td>
<td>0.987</td>
</tr>
<tr>
<td>$\beta_2 = 4.0$</td>
<td>0.000</td>
<td>0.055</td>
<td>n/a</td>
</tr>
<tr>
<td>$\beta_1 = 4.0$</td>
<td>1.000</td>
<td>0.953</td>
<td>n/a</td>
</tr>
<tr>
<td>$\beta_2 = 4.0$</td>
<td>1.000</td>
<td>0.941</td>
<td>n/a</td>
</tr>
<tr>
<td>$# \text{of runs}$</td>
<td>512</td>
<td>640</td>
<td>290</td>
</tr>
</tbody>
</table>

4.1 Small-scale case with two-factor interactions

We first compare the effectiveness and efficiency of CSFD, CSB-X and TCFF methods on a small-scale case (factor number $l = 10$) with second-order interactions. The simulation models in this scale (or smaller) can be found in various disciplines. For example, a hospital scheduling simulation model with four factors are discussed in Kopach et al. (2007). The simulation parameters are listed in Table 9 and the effect coefficients of the model are set up as follows:

- **Main-effects:** $\langle \beta_1, \beta_2, \ldots, \beta_{10} \rangle = (2.0, 2.0, 2.4, 2.7, 3.0, 3.3, 3.6, 3.8, 4.0, 4.2)$.
- **Interaction effects:** $\langle \beta_{1,2}, \beta_{3,5}, \beta_{4,8} \rangle = (2.0, 4.0, -4.0)$.
- **All other second or higher order interactions are zero.**

A $2^{10-3}$ design is used in CSFD and TCFF. For CSB-X, factor effects are tested in groups and the design points are determined sequentially (Wan et al., 2008). All three methods are repeatedly run with different $n_0$ (Recall that for CSFD and TCFF, $n_0$ is the initial number of replications at each design point; for a $2^{10-3}$ factorial design each replications contains one observation from $2^{10-3} = 128$ design points; for CSB-X, $n_0$ is the initial number of replications generated at each design point). The selection of $n_0$ has little influence on the screening results but does change the total simulation runs required. Table 9 presents the result of each case with smallest average number of simulation runs required.

CSFD and TCFF have similar efficiency, although CSFD is a little more efficient when the variance is small. We can also see when the variance is small, CSB-X can be more efficient than CSFD and TCFF; however, when the variance gets larger, the simulation effort required by CSB-X increases dramatically and CSFD and TCFF are much more efficient than CSB-X.

For all three cases, $P(\text{DI})$ should be less than $z = 0.05$ for effects $\beta_1, \beta_2, \beta_3$, and $\beta_{1,2}$ since their coefficients are $< A_0$, and greater than $\gamma = 0.95$ for effects $\beta_3, \beta_1, \beta_2, \beta_3, \beta_{1,2}$, and $\beta_{1,3}$ since their coefficients are $\geq A_1$. We can see that CSFD strictly meets both requirements and is conservative. CSB-X has conservative Type I errors but the powers are lower for critical factors when variances are large. TCFFs results are close to the specified error control. CSFD and TCFF show no difference in classifying main-effects and interaction effects.

4.2 Large-scale cases

As discussed in Saeger and Hinch (2001), the Army runs a series of simulation experiments to generate tactics and simultaneously...
assess the value of various systems. The simulations that are used are very large in size, sometimes taking many hours to run a single experiment and having a large number of user selectable input values. We will consider two large-scale cases with 200 factors and 500 factors, respectively. Main-effects models are assumed in both cases. The simulation parameters of the large-scale cases are given in Table 9. We use a Resolution III fractional factorial designs discussed in Section 3. The size of the design is 256 for 200-factor case and 512 for 500-factor case. In both cases, the effect coefficients are set up as follows:

- 5% of the factors are assumed to be critically important and their effect coefficients equal to \( \beta_1 = 4 \).
- There are another 5% of factors whose effect coefficients are set to be \( \beta_0 = 2 \).
- All other factors have effect coefficients set to be 0.

For each case, three scenarios of factor distribution will be studied:

- The first scenario has all factors with non-zero effect coefficients clustered together with the smallest indices so that the number of important groups is as small as possible at each step of CSB-X.
- The second scenario has factors with non-zero effect coefficients evenly spread so there are maximum number of important groups remaining at each step.
- The third scenario has those factors randomly spread.

Again, various initial numbers \( n_0 \) are used. Table 10 presents the results of the most efficient case for each scenario. We can draw similar conclusions on these three methods as in small-scale case results of the most efficient case for each scenario. We can draw the following conclusions:

<table>
<thead>
<tr>
<th>Extra effects increase, CSB-X becomes very inefficient.</th>
</tr>
</thead>
<tbody>
<tr>
<td>When variance increases, CSB-X becomes very inefficient.</td>
</tr>
<tr>
<td>and the efficiency advantage of CSFD is more obvious. When variance increases, CSB-X becomes very inefficient.</td>
</tr>
<tr>
<td>Similarly conclusions on these three methods as in small-scale case results of the most efficient case for each scenario. We can draw the following conclusions:</td>
</tr>
</tbody>
</table>

\[
\begin{array}{|c|c|c|c|c|c|c|}
\hline
\text{Case} & \text{Scenario} & \text{CSFD} & \text{TCFF} & \text{CSB-X} & \text{CSFD} & \text{TCFF} & \text{CSB-X} \\
\hline
\text{200-factor} & 1st & 1024 & 1280 & 161 & 1024 & 1280 & 2060 \\
 & 2nd & 1024 & 1280 & 355 & 1024 & 1280 & 3424 \\
 & 3rd & 1024 & 1280 & 333 & 1024 & 1280 & 3260 \\
\hline
\text{500-factor} & 1st & 2048 & 2560 & 772 & 2048 & 2561 & 25462 \\
 & 2nd & 2048 & 2560 & 1550 & 2048 & 2561 & 38580 \\
 & 3rd & 2048 & 2560 & 1418 & 2048 & 2561 & 36510 \\
\hline
\end{array}
\]

4.3. Summary and further discussions

**Efficiency:** The numerical evaluation shows that overall CSFD has superior performance compared to TCFF or CSB-X. Compared with CSB-X, CSFD is more efficient in most circumstances (except for cases with very small variances). CSFD has similar efficiency as TCFF in small-scale case, but is more effective than TCFF in large-scale cases. This seems counter-intuitive since in large-scale cases, each replication of CSFD requires a large number of simulation runs. However, CSFD usually only needs a few replications to screen all effects. This is because the variances of the effect estimates are small. Consider CSFD with a \( 2^4 \) full factorial design and assume equal variances \( \left( \sigma^2 \right) \) across the response surface. It is easy to verify that \( \text{Var} (\beta_k) = \sigma^2 / 2^4 \), \( k = 1, 2, \ldots, K \). For CSB-X, the variance of a group effect estimate will be \( \sigma^2 / 2 \) in the same situation. Thus, compared to CSB-X, CSFD significantly reduces the variances of the effect estimates in most cases. This is why CSFD is not as sensitive to the change of variance as CSB-X and is overall more effective. Compared with TCFF, which may require a large number of simulation runs at design points with large variances, CSFD takes equal numbers of runs at each design point and the resulting effect estimates share the same variance.

**Initial sample size selection:** Fig. 2 demonstrates the relationship between the initial replication number \( n_0 \) and the average simulation runs required for screening (CSFD, 200-factor case, 3rd scenario, and \( a = 1.0 \)). This figure is consistent with Fig. 18.1 in Kim and Nelson (2006), which presents the typical form of expected simulation runs required as a function of initial number \( n_0 \) in ranking-and-selection problems. There often exists an unknown “optimal” initial sample size; an \( n_0 \) too small may lead to huge penalty (as shown in Fig. 2) and an \( n_0 \) too large is usually unnecessary. The reason is that when \( n_0 \) is too small, it may produce a large sample variance, which further leads to a large total number of replications. When \( n_0 \) is too large, all of the sequential testing procedures reach conclusions in the first stage with more than necessary observations. This is why the right side of the curve is actually a straight line. As the numerical results show, with other system parameters fixed, the larger the variance, usually the larger the optimal \( n_0 \) is. A small deviation from the true optimal initial number usually has little influence on the simulation effort.

**Common random numbers:** The impact of common random numbers (CRN) in CSFD is discussed as follows. For a \( 2^4 \) full factorial design, \( \beta_k = \frac{1}{2} X_k^t Y \) and the entries of \( X_k \), are equal number of 1s and -1s. Suppose \( X_{ij} = (a_1, a_2, \ldots, a_5) \), where \( a_i = +1 \) or \( -1 \), then...
\[
\text{Var}(\hat{\beta}_i) = \frac{1}{2\pi} \left( \sum_{i=1}^{2^n} \Var(Y_i) \right)
\]
\[
= \frac{1}{2\pi} \left( \sum_{i=1}^{2^n} a_i^2 \Var(Y_i) + \sum_{i<j} a_i a_j \text{Cov}(Y_i, Y_j) \right)
\]
\[
= \frac{1}{2\pi} \left( \sum_{i=1}^{2^n} \Var(Y_i) + \sum_{i<j} a_i a_j \text{Cov}(Y_i, Y_j) \right).
\]

The effect of the CRN is determined by the magnitude of each Cov(Y_i, Y_j). In a special case where all covariances are equal to \(\rho > 0\), since there are \(2^n\) positive \(a/\)'s and \(2^n\) negatives \(a/\)'s in \(X_i^2\), \(2^n(2^{n-1})\) \(a/\)'s will be negative and the other \(2^n(2^{n-1} - 1)\) \(a/\)'s will be positive. Therefore

\[
\text{Var}(\hat{\beta}_i) = \frac{1}{2\pi} \left( \sum_{i=1}^{2^n} \Var(Y_i) - 2^n \text{Cov}(Y_1, Y_2) \right)
\]
\[
= \frac{1}{2\pi} \sum_{i=1}^{2^n} \Var(Y_i) - \frac{1}{2^n} \rho,
\]

and \(\frac{1}{2^n} \rho\) is the benefit of implementing CRN. Because of the structure of contrasts, when the covariances are not too much different from each other, the effect of CRN will be favorable.

5. Case study

Wan et al. (2006) and Wan et al. (2008) have implemented CSB and CSB-X, respectively, in a simplified semiconductor manufacturing system to identify important machines and transporters that are worthy of investment. In this section, we will implement CSFD with fully sequential testing procedure on the system. Here we will only give a brief description of the system. For details, please refer to Wan et al. (2006) and Wan et al. (2008).

As shown in Fig. 3, the semiconductor manufacturing system consists of two basic steps, diffusion and lithography, each of which contains multiple stations. Raw material is released into the system at the rate of 1 cassette/hour and processed in single-cassette loads. It begins with the diffusion process and then proceeds to the lithography process. For each product, the two processes alternate until all passes required are completed. The percentages of the product mix and the number of passes of each product type requires are given in Table 11. Transporters (AGV and CONVEYOR) move products between diffusion and lithography and are also considered as stations. The time of moving products within each process step is negligible. The mean processing times of each cassette at different stations are listed in Table 12. The performance measure of the system is the mean cycle time of product processing weighted by the percentages of different products. The factors of interests are the numbers of fast and slow machines.

\[
\text{Var}(\hat{\beta}_i) = \frac{1}{2\pi} \left( \sum_{i=1}^{2^n} \Var(Y_i) - 2^n \text{Cov}(Y_1, Y_2) \right)
\]

\[
= \frac{1}{2\pi} \sum_{i=1}^{2^n} \Var(Y_i) - \frac{1}{2^n} \rho,
\]

and \(\frac{1}{2^n} \rho\) is the benefit of implementing CRN. Because of the structure of contrasts, when the covariances are not too much different from each other, the effect of CRN will be favorable.

5. Case study

Wan et al. (2006) and Wan et al. (2008) have implemented CSB and CSB-X, respectively, in a simplified semiconductor manufacturing system to identify important machines and transporters that are worthy of investment. In this section, we will implement CSFD with fully sequential testing procedure on the system. Here we will only give a brief description of the system. For details, please refer to Wan et al. (2006) and Wan et al. (2008).

As shown in Fig. 3, the semiconductor manufacturing system consists of two basic steps, diffusion and lithography, each of which contains multiple stations. Raw material is released into the system at the rate of 1 cassette/hour and processed in single-cassette loads. It begins with the diffusion process and then proceeds to the lithography process. For each product, the two processes alternate until all passes required are completed. The percentages of the product mix and the number of passes of each product type requires are given in Table 11. Transporters (AGV and CONVEYOR) move products between diffusion and lithography and are also considered as stations. The time of moving products within each process step is negligible. The mean processing times of each cassette at different stations are listed in Table 12. The performance measure of the system is the mean cycle time of product processing weighted by the percentages of different products. The factors of interests are the numbers of fast and slow machines.

![Fig. 3. Production process of the semiconductor manufacturing system (Wan et al., 2006, 2008).](image-url)
at each station and the numbers of each kind of transporter. The simulation of the semiconductor manufacturing system was done in simlib (Law and Kelton, 2000). For each observation of mean cycle time, 365 days of operations were simulated with a 300-hour warm-up period. Table 13 gives the factor description and the level settings, which we pick the extreme levels as in Wan et al. (2008). A $2^{11}$ factorial design is used, which requires 512 experimental points in each replication, such that no main and second-order interaction effects are confounded with each other. The initial sample size is set to be 3.

The result for different combinations of $(a_0, a_1)$ are given in Table 14. The two-factor interactions are represented by two factors in parenthesis. For example (13, 15) represents the interactions between factor 13 and 15. The screening results are overall consistent with the CSB and CSB-X screening results reported in Wan et al. (2006) and Wan et al. (2008). Main effect of factor 16 is identified as the most important effect. When the thresholds are reduced, additional effects are identified as important. When the thresholds are reduced to (0.5, 1.0), CSFD gives insightful information on second-order interactions, which is not available in other methods. All four cases require 1536 replications, which implies that the initial replications have been sufficient to draw conclusions in all cases.

6. Conclusions and future research

CSFD is a sequential factor screening approach that provides simultaneous Type 1 error and power control. It requires few assumptions on the model and little prior information about the system. With the option of using fractional factorial designs, CSFD can handle large-scale cases efficiently. The performance is robust across different factor configurations. Numerical evaluation shows that there is a complementary relationship between CSFD and CSB-X. That is, the weakness of CSB-X is usually the strength of CSFD, and vice versa. Our future research will concentrate on developing a hybrid method combining CSB-X and CSFD methods together to take the advantages of both methods. In addition, the replications collected at each design point will allow us to estimate not only the means, but also the variances. In the future we are interested in exploring screening designs for robustness studies.

Acknowledgements

This research was partially supported by Purdue Research Foundation (PRF) and Naval Postgraduate School Grant No. 00244-06-C-0002. Preliminary results from this paper appeared in Shen and Wan (2005) and Shen (2006). Additional thanks go to Professor Barry L. Nelson and Professor Bruce E. Ankenman from Northwestern University for their insightful comments.

Appendix

For the fully sequential testing procedure introduced in 3.2, when $\alpha = 1 - \gamma$, constants $a_0$ and $r_0$ are the solutions of the following equations:

$$
\int_{a_0}^{\infty} \int_{a_0}^{\infty} \frac{\gamma(y)}{1 - \gamma(y)} \left( \frac{Y(\mathbf{x}) - (r_0 - a_0)}{\theta(\mathbf{x})} \right) f(x, n_0 - 1) \, dx \, dy = \alpha, \tag{2}
$$

and

$$
\int_{0}^{a_0} \int_{0}^{a_0} \frac{\gamma(y)}{1 - \gamma(y)} \phi(\theta(\mathbf{x}) - (a_1 - r_0) \theta(\mathbf{x})) f(x, n_0 - 1) \, dx \, dy = 1 - \gamma, \tag{3}
$$

where $\gamma(y) = e^{-y^2/2}$, $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$, $\theta(\mathbf{x}) = \sqrt{n_0 - 1}/\sqrt{x_0/\lambda}$, and

$$
f(x, n) = \frac{1}{\Gamma(n/2)^{1/2}} \cdot \left( \frac{x}{\Gamma(n/2)^{1/2}} \right)^{n/2 - 1} e^{-x}, \quad x > 0
$$

which is the probability density function of $\chi^2$ distribution with $n$ degrees of freedom, and $f(1/2) = \int_{0}^{\infty} x^{n/2 - 1} e^{-x} \, dx$

Numerical methods are needed to evaluate Eqs. (2) and (3) and search for solutions. The following lemma can be proved, which makes the search relatively easy (Wan et al., 2008).

Lemma 1. For $a_0$ fixed, both power and Type I error are decreasing in $r_0$.

Lemma 2. If the required $\alpha < 1/2$ and $a_1 > 1$ then $a_0 < r_0 < a_1$.

References


