Calhoun: The NPS Institutional Archive
DSpace Repository

Faculty and Researchers

Faculty and Researchers' Publications

2009

A Hybrid Method for Simulation Factor Screening

Shen, Hua; Wan, Hong; Sanchez, Susan M.

Monterey, California: Naval Postgraduate School.

Wiley Periodicals, Inc., Naval Research Logistics, Volume 57, pp. 4557, 2010
<https://hdl.handle.net/10945/43593>

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

Downloaded from NPS Archive: Calhoun

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

<http://www.nps.edu/library>

A Hybrid Method for Simulation Factor Screening

Hua Shen,¹ Hong Wan,² Susan M. Sanchez^{3,4}

¹ Bank of America, Charlotte, NC 28255

² School of Industrial Engineering, Purdue University, West Lafayette, Indiana 47907-2023

³ Department of Operations Research, Graduate School of Operational and Information Sciences,
Naval Postgraduate School, Monterey, California 93943

⁴ Graduate School of Business and Public Policy, Naval Postgraduate School,
Monterey, California 93943

Received 24 July 2008; revised 23 July 2009; accepted 6 August 2009

DOI 10.1002/nav.20382

Published online 28 October 2009 in Wiley InterScience (www.interscience.wiley.com).

Abstract: Factor screening is performed to eliminate unimportant factors so that the remaining important factors can be more thoroughly studied in later experiments. Controlled sequential bifurcation (CSB) and controlled sequential factorial design (CSFD) are two new screening methods for discrete-event simulations. Both methods use hypothesis testing procedures to control the Type I Error and power of the screening results. The scenarios for which each method is most efficient are complementary. This study proposes a two-stage hybrid approach that combines CSFD and an improved CSB called CSB-X. In Phase 1, a prescreening procedure will estimate each effect and determine whether CSB-X or CSFD will be used for further screening. In Phase 2, CSB-X and CSFD are performed separately based on the assignment of Phase 1. The new method usually has the same error control as CSB-X and CSFD. The efficiency, on the other hand, is usually much better than either component method. © 2009 Wiley Periodicals, Inc. *Naval Research Logistics* 57: 45–57, 2010

Keywords: computer experiments; discrete-event simulation; factor screening

1. INTRODUCTION

Screening experiments allow the researcher to eliminate unimportant factors so that more detailed investigations can focus on the most influential ones. Many strategies have been proposed for factor screening purposes [2, 18, 19]. However, much of the current research has concentrated on designs for physical experiments, which typically involve fewer than 25 factors and do not take advantage of the sequential nature of simulation experiments. In addition, because of the high cost of conducting physical experiments, the traditional screening methods usually emphasize using the lowest number of runs to estimate as many effects as possible. In contrast, simulation models in a wide variety of application areas involve dozens, hundreds, or even thousands of factors that can be manipulated during an experiment. Screening designs appropriate for these high-dimensional situations are extremely valuable for gaining insights. For example, those factors identified as important using simulation screening experiments may be

useful for specifying conditions for subsequent live tests, i.e., physical experiments on the system modeled by the simulation. Other application areas include logistics systems, population dynamics modeling, homeland security, and more. For a case study of a model of organizational performance, see Ref. 9.

Controlled sequential bifurcation (CSB) is a new group screening method specifically designed for stochastic simulation experiments [21]. Factors are tested in groups. If the group effect is unimportant, all factors in the group are considered to be unimportant; if the group effect is important, the factors in the group will be split into two smaller subgroups for further testing. CSB is the first screening method to control the probabilities of misclassifications. With a qualified hypothesis testing procedure [21, 22] at each bifurcation step, CSB can control the Type I Error for each factor (i.e., the probability of an unimportant factor being classified as important) and the power at each bifurcation step (i.e., the probability of a critical group being identified) under heterogeneous variance conditions. Since CSB can eliminate unimportant factors in groups, it is particularly well suited

Correspondence to: H. Wan (hwan@purdue.edu)

Table 1. Structure of CSB and CSB-X.

Initialization:
Create an empty LIFO queue for groups. Add all factors as one group to the 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

for cases involving a large number of factors when only a small percentage of them are important. Wan et al. [22] later proposed an improved version of CSB, called CSB-X (“X” represents the interactions), using a fold-over design. CSB-X relaxes the assumption of a main-effects model and gives the same error control for main effects as does CSB, even when two-factor interactions and quadratic terms exist.

Although CSB methods are attractive for many simulation applications, they also have serious limitations: (1) the methods can only screen main effects and the results can be misleading when higher-order interactions exist; (2) the signs of main effects are assumed to be known to avoid effect cancellation, but this knowledge is not always available in practice; (3) in CSB methods, simulation observations generated in previous bifurcation steps may not be useful in the later screening stages, so new observations are usually needed at each bifurcation step; and (4) the efficiency of CSB methods is sensitive to the index order of factors (important factors are preferably clustered) and the variances of the response surface. In reality, the optimal setting is rarely achieved since prior information is often faulty.

Controlled sequential factorial design (CSFD) [14, 16] was proposed to overcome the limitations of CSB methods. CSFD combines a sequential hypothesis testing procedure with a traditional factorial design to provide simultaneous Type I Error and power control for each effect of interest under heterogeneous variance conditions (the power control is stronger than CSB). CSFD can screen any main effects and interactions without assuming the directions of effects to be known a priori. In addition, unlike CSB methods, it can use all previously generated observations in the later screening process. In most cases, after the first few effects are classified, there is enough data to classify all of the remaining effects. On the

other hand, when the number of factors is large and interactions exist, CSFD would have to repeat a huge factorial design and the simulation effort might be prohibitive.

The structures of CSB (CSB-X) and CSFD are demonstrated in Tables 1 and 2, respectively. Numerical evaluations show that a complementary relationship exists between the CSB and CSFD methods [16, 21]; that is, the strength of CSB methods is usually the weakness of CSFD, and vice versa.

This complementary relationship inspires us to propose a hybrid method that combines CSB-X and CSFD to achieve better efficiency. The underlying idea is to apply CSFD to screen those effects likely to be important (typically a small percentage of all effects), and CSB-X to screen those effects likely to be unimportant. Since each method will be conducted in its favorable settings, the efficiency of the screening process can be significantly improved. The challenge is to coordinate the different procedures and provide overall error control for the screening results. Note that CSB can be used in place of CSB-X within the hybrid framework; however, in this study we focus on using the improved procedure.

This article is organized as follows. We discuss the underlying response model and the objective of screening in Section 2. In Section 3, we describe the proposed hybrid procedure and its error control ability. In Section 4, we present empirical evaluations of the hybrid method and compare it with the existing methods. We discuss our conclusions and directions for future research in Section 5.

2. MODEL DESCRIPTION

Suppose there are L factors. A general linear model that includes all main effects and interactions, but no polynomial terms, is shown below:

$$Y = \beta_0 + \sum_{i=1}^L \beta_i z_i + \sum_{i < j} \beta_{ij} z_i z_j + \sum_{i < j < k} \beta_{ijk} z_i z_j z_k + \cdots + \beta_{12 \dots L} z_1 z_2 \dots z_L + \varepsilon. \quad (1)$$

Table 2. Structure of CSFD.

Initialization:
Form a queue of effects of interest. Select a factorial design.
Generate $N = n_0$ replications of observations.
While queue is not empty, do
Remove: Remove an effect from the queue.
Compute: Compute sample mean and sample variance of the effect coefficient. Sample size = N .
While the effect is not classified, do
If the effect cannot be classified with the specified error control, then generate new replication(s). Update sample size N and sample mean.
End If
End While
End While

Table 3. 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 $\hat{\beta}_1, \dots, \hat{\beta}_k$. Order the estimates so that $\hat{\beta}_{[1]} \leq \dots \leq \hat{\beta}_{[z]} < 0 \leq \hat{\beta}_{[z+1]} \dots \leq \hat{\beta}_{[L]}$. Add factors $\{[1], \dots, [z]\}$ to the NEG LIFO queue, and factors $\{[z+1], \dots, [L]\}$ to the POS LIFO queue.
Phase 2:
Apply CSB on two LIFO queues, NEG and POS, separately.

Here $\beta = \{\beta_1, \beta_2, \dots, \beta_{12 \dots L}\}$ is the effect coefficient vector and $\mathbf{z} = (z_1, z_2, \dots, z_L)$ stands for the deterministic level settings. In practice, the effects of interest can be any subset of β . Heterogeneous variances are allowed and the error term, ε , is assumed to be a $\text{Nor}(0, \sigma^2(\mathbf{z}))$ random variable whose variance is unknown and may depend on $\mathbf{z}$. We also assume that the error terms are independent. A screening experiment will classify each factor as “important” or “unimportant.” We want to simultaneously control the Type I Error $\leq \alpha$ for those effects $\leq \Delta_0$ (unimportant effects), and the power $\geq \gamma$ for those effects $\geq \Delta_1$ (critical effects). We consider the effects between Δ_0 and Δ_1 to be important, but no power control will be offered for the screening results. Here, the parameters Δ_0 and Δ_1 are the thresholds of importance and criticality, respectively; α and γ are user-specified error control parameters.

Consider a typical complex simulation model with a large number of factors, possible important interactions, and little prior knowledge of the system. Neither CSB-X nor CSFD alone will be both effective and efficient. CSB-X cannot screen interactions and the scattered important factors may impede the elimination of unimportant factors in large groups. CSFD will require a huge factorial design; even if this design is replicated only a few times, too many runs may be required. The proposed hybrid method is intended to target this situation. The “sparsity of effects” principle [8] is still assumed to be valid (namely, only a small percentage of factors are responsible for most of the response variation); otherwise, screening experiments would have little benefit. Note that in this article, we focus on polynomial response surfaces. Those interested in other model estimation techniques, such as kriging, may still require screening methods prior to their primary experiment to reduce the number of factors to a manageable size.

3. HYBRID METHODOLOGY

3.1. Hybrid Procedures

FF-CSB [13, 14], where “FF” stands for fractional factorial design, is the first effort on the hybrid screening approach,

and Table 3 demonstrates its structure. FF-CSB still assumes a main-effects response model. To drop the assumption of known effect directions in CSB, FF-CSB adds a prescreening stage to estimate the directions and magnitudes of the effects. All factors are then divided into positive and negative groups, and factors are sorted within each group. In the second stage, the original CSB procedure is applied separately on the “positive” and “negative” groups. Although the prescreening is not guaranteed to accurately estimate the directions of all effects, the numerical evaluation shows that FF-CSB usually screens factors effectively without assuming the known directions of effects. Furthermore, sorting the factors significantly improves the efficiency of the following CSB procedure. Similar results hold for a variant called FF-CSBX that uses CSB-X in the second phase, allowing the screening for main effects to occur even in the presence of two-way interactions.

We extend the aforementioned hybrid strategy by integrating the prescreening, CSB-X, and CSFD procedures into one screening method. Table 4 gives the generic structure of our proposed hybrid method. It consists of two phases. Prescreening is performed in Phase 1 to obtain the initial estimates of all desired effects. Based on these estimates, all factors are explicitly assigned into one of the three groups: factors associated with one or more potentially important main or interaction effects are assigned to group IMP; the rest of the factors, the potentially unimportant ones, are assigned to either group POS or group NEG, based on the estimated effect directions. The cut off value that divides potentially important and unimportant effects is called the factor assignment threshold. Typically, this value will be less than Δ_0 since we know that our estimates will have noise, and therefore many factors with effects just above Δ_0 might have effect estimates below Δ_0 . The prescreening procedure used in Phase 1 should be able to estimate all desired effects with a small number of simulation runs. For example, in FF-CSB, saturated or nearly

Table 4. Structure of hybrid method.

Initialization:
Create three empty groups: IMP, POS, and NEG.
Phase 1:
Prescreen: Conduct a prescreening procedure to estimate the coefficients of all effects of interest.
Divide: Assign factors related to potential important effects to group IMP; assign other factors to either group POS or NEG based on the directions of their estimated effect coefficients.
Phase 2:
Sort: Sort factors in groups POS and NEG, respectively, based on estimated effect coefficients.
CSB-X: Apply CSB-X to classify factors in groups POS and NEG separately.
CSFD: Apply CSFD to classify factors in group IMP (main effects and interactions).

saturated fractional factorial designs are used. In Phase 2, the CSFD procedure is applied on group IMP, and the CSB-X procedure is applied separately on groups POS and NEG. The factors in groups POS and NEG will first be sorted based on the estimated main-effects coefficients within each group. Note that because of the stochastic nature of the response, it is possible for factors to be assigned to the wrong groups during Phase 1. The influence of such misassignments will be discussed in Section 3.3.

3.2. Prescreening Procedures

Different prescreening procedures could be used in Phase 1. When response models can be predetermined (for example, by expert opinions or past experience), a saturated or nearly saturated design can be used as the prescreening procedure to estimate the coefficients of the desired effects. A saturated design has the number of runs equal to the number of effects of interest plus one. For example, a Resolution III fractional factorial design allows $N - 1$ main effects to be estimated in only N runs, where N is a power of 2. The (non geometric) Plackett-Burman designs extend N to be a multiple of 4 [7]. They have complicated alias structures, and in special cases, they may allow both main effects and a limited number of interactions to be estimated. However, they require special analysis methods [20], and these designs are typically only available for $N \leq 28$ which limits their application for simulation screening. For general response models with interaction effects, fractional factorial designs with higher resolutions can be used. For example, Resolution V factorial designs can estimate all main and two-factor interaction effects independently; Resolution VII factorial designs can estimate up to three-factor interaction effects independently, etc. Recommended factorial designs for different resolutions and number of factors can be found in most experimental design textbooks (e.g., Refs. 7 and 23) and statistical software packages. Typically, these are available only for a relatively small number of factors (e.g., less than 16). Sanchez and Sanchez [13] described a fast, easy way to generate Resolution VI fractional factorials up to $2^{443-423}$. Sanchez [16] extends this for Resolution VII designs involving up to 81 factors. For other orthogonal-array-based saturated designs involving interaction effects, see Refs. 1, 6, and 17. If there is no prior response model information available, then the prescreening needs to explore the response surface and an appropriate model form as well as the initial grouping of factor effects. In this situation, space-filling designs such as Latin hypercube sampling (LHS) would be appropriate. For details of LHS, see Refs. 3, 4, and 5.

3.3. Factor Assignments

After obtaining the estimated effect coefficients, the overall efficiency and error control of the hybrid method depend

critically on how the factors are divided into the three groups. If the factor assignment threshold is too small, then too many factors will be assigned to group IMP. The resulting CSFD will have to replicate a large factorial design, and the simulation effort in Phase 2 could be too much. On the other hand, if the threshold is too large, many potentially important factors will be assigned to unimportant groups, which may not only seriously affect the efficiency of CSB-X but also impede the effectiveness of the hybrid method, especially the classification of important effects. In other words, the selection of the factor assignment threshold needs to balance the trade-off between the overall efficiency and effectiveness of the hybrid method.

With a qualified hypothesis test [21, 22], which refers to a test that simultaneously controls the Type I and Type II errors at an individual step, both CSFD and CSB-X can provide specified Type I Error and power control of the classification results in Phase 2. However, the overall error control of the hybrid method is not automatically guaranteed. Factor misassignments may occur during the prescreening phase, since it is not intended to provide accurate estimates. There are three kinds of misassignments: (1) important factors assigned to unimportant groups POS or NEG; (2) unimportant factors assigned to group IMP; and (3) positive factors assigned to group NEG or vice versa. Notice that none of these misassignments would, in fact, damage the overall Type I Error control ability of the hybrid method; that is, the probability of any unimportant effect being classified as important would not increase, even when misassignments of factors occur. This is because for any (main or interaction) effect that is eventually identified as important by the hybrid method, it must have been tested individually by either the CSB-X or CSFD procedure. The qualified test, implemented in these procedures, guarantees the Type I Error of classifying any unimportant effect as important to be less than the specified target error.

There are two kinds of factor misassignments, however, that can affect the power of the hybrid method: (1) a factor associated with a critical interaction effect assigned to one of the unimportant groups; as well as (2) a factor with critical positive main effect assigned to group NEG, or a factor with critical negative main effect assigned to group POS. In the first scenario, when a factor associated with a critical interaction effect is assigned to an unimportant group, there is no chance that this interaction effect could be identified as important in Phase 2, since the CSB-X procedure cannot screen interactions. In the second scenario, when a factor with critical main effect is assigned to an unimportant group associated with the opposite effect direction, it is possible that effect cancellation occurs and this factor, and other factors grouped with it, can be classified simultaneously as unimportant. In the rest of this chapter, we will refer to the factors associated with at least one critical (main or interaction) effect

as critical factors and the misassignments of critical factors as critical misassignments.

The possibility of critical misassignments prompts us to be conservative in selecting the factor assignment threshold. Being “conservative” means that the factor assignment threshold should be well below Δ_0 to place almost all effects that are actually greater than Δ_0 into the IMP group, even though this might cause extra simulation effort in the CSFD procedure of Phase II. On the other hand, the factor assignment threshold cannot be too small, otherwise most of the effects (including unimportant ones) will be assigned to CSFD and the benefit of CSB-X will be too trivial (see Section 4 for more discussions). In our empirical evaluation, we choose the factor assignment threshold to be $\Delta_0/2$ rather than using the threshold of importance Δ_0 . The reasons for this choice are as follows: (1) the smaller factor assignment threshold results in fewer critical misassignments and better overall power of the hybrid method; (2) the smaller factor assignment threshold also means fewer factors with non zero effects are placed in unimportant groups, which may significantly increase the efficiency of the CSB-X procedure; (3) a moderate increase of the number of factors in group IMP is unlikely to result in a significantly larger factorial design required in each replication of CSFD; and finally, (4) the factor assignment threshold should not be too small, otherwise most of the factors will be assigned to the IMP group and the hybrid method will basically be reduced to the CSFD method. Numerical results in Section 4 will show that with a threshold of $\Delta_0/2$, it is very unlikely for critical misassignments to occur unless the variances are extremely large. More specifically, the misassignment rates for critical factors in our empirical evaluation are all less than 0.05% (Table 9 in Section 4). In most cases, the hybrid method has almost the same effectiveness as that of the CSFD method and is more efficient than both the CSFD and FF-CSB methods.

3.4. Error Control

The error control abilities of the hybrid method are given in the following theorems.

THEOREM 1: Given a prescreening method and a qualified testing procedure, the hybrid method guarantees that

$$\Pr \{ \text{Declare effect } k \text{ important} \mid |\beta_k| \leq \Delta_0 \} \leq \alpha$$

and

$$\Pr \{ \text{Declare effect } k \text{ important} \mid |\beta_k| \geq \Delta_1 \} \geq (1 - \zeta)\gamma$$

for each effect where α and γ are the user-specified target Type I Error and power rates and $\zeta = \Pr(\text{a critical effect is not assigned to IMP group})$, which depends on the prescreening procedure.

The details of the proof and the determination of ζ , when an orthogonal factorial design is used as the prescreening procedure of the hybrid method, are discussed in the Appendix. From Theorem 1, we can see that the hybrid method has the same Type I Error control ability as the CSB-X and CSFD methods. It also has stronger power control ability than the CSB methods when ζ is small. Recall that the CSB-X method only controls the power at each bifurcation step and the overall power of the classification of a critical effect is the product of the powers of the hypothesis tests this effect goes through. When there are many bifurcation steps, the overall power could be much less than the specified power rate. In the hybrid method, this lower bound of the overall power, $(1 - \zeta)\gamma$, depends only on the prescreening results and the power control of the CSFD procedures. It is independent of the number of bifurcation steps taken by the CSB-X procedure in Phase 2.

REMARK: The error control calculations for the prescreening method require independent errors. However, the individual component methods CSB-X and CSFD both allow the application of common random numbers. This may reduce the total sampling effort [16].

We can further prove the following theorem:

THEOREM 2: The overall power of the hybrid method is non increasing in terms of the factor assignment threshold T , as is the lower bound of the power given in Theorem 1.

This is straightforward because the smaller the threshold is, the less likely critical misassignments occur, and thus larger is the overall power. Theorem 2 implies that the lower bound of the overall power of the hybrid method is controllable in Phase 1 of the screening process. After the value of ζ is obtained at the end of Phase 1, the lower bound $(1 - \zeta)\gamma$ can be calculated. If the lower bound does not meet the specified requirement, the screening procedure can automatically decrease the value of the factor assignment threshold, reassign the factors, and obtain a new value of ζ accordingly, without generating new observations. In fact, a one-to-one correspondence can be set up between the lower bound of the power and the value of T , and it could be used as a guide to select the factor assignment threshold, based on the target lower bound of power control. However, this approach may not lead to the threshold that makes the hybrid method most efficient.

4. EMPIRICAL EVALUATION

To study the performance of the proposed hybrid method, we conduct a series of numerical experiments to compare

Table 5. Simulation experiment parameters.

Parameter	Value
L	50, 200, 500
Δ_0	2
Δ_1	4
α	0.05
γ	0.95
σ	$m \times (1 + \text{size of the group effect})$
m	0.01, 0.1, 0.3, 1

it with CSFD and FF-CSB. Table 5 lists the simulation experimental parameters. In all cases, the presented results are the averages of 1000 independent trials. Fractional factorial designs are used as prescreening procedures (Resolution III designs for main-effects models and Resolution V designs for second-order models). Unless stated otherwise, there is only one replication per design point in the prescreening procedures. For the CSB-X and CSFD procedures, different initial sample sizes (n_0) are used, and for each case, the results with the minimal number of simulation runs required for screening are presented for comparison. For the relationship between the initial sample sizes and the efficiency of CSB-X and CSFD, see Ref. 16.

4.1. Main-Effects Model

We first consider the examples with main-effects response models. Effect coefficients are randomly generated and the distribution of their absolute values is as follows: 2.5% are equal to $\Delta_1 = 4$, 2.5% are uniformly distributed on $(\Delta_0, \Delta_1) = (2, 4)$, 2.5% are equal to $\Delta_0 = 2$, 2.5% are uniformly distributed on $(0, \Delta_0) = (0, 2)$, and all others are equal to zero. The signs for factors with non zero effects are also randomly distributed. Each non zero coefficient has equal probability of being positive or negative. Two cases with 200 and 500 factors, respectively, are randomly generated and their non zero effect coefficients are listed as follows:

- 200-factor: $(\beta_9, \beta_{19}, \beta_{63}, \beta_{77}, \beta_{123}) = (4.0, 4.0, 4.0, -4.0, -4.0)$, $(\beta_4, \beta_{17}, \beta_{71}, \beta_{118}, \beta_{127}) = (2.38, 2.88,$

$3.35, 2.74, -3.39)$, $(\beta_{25}, \beta_{55}, \beta_{96}, \beta_{140}, \beta_{144}) = (2.0, 2.0, -2.0, -2.0, -2.0)$, $(\beta_{10}, \beta_{35}, \beta_{39}, \beta_{174}, \beta_{182}) = (1.28, -0.88, -0.35, 1.61, 0.48)$.

- 500-factor: $(\beta_{22}, \beta_{26}, \beta_{37}, \beta_{80}, \beta_{147}, \beta_{338}, \beta_{340}, \beta_{355}, \beta_{359}, \beta_{366}, \beta_{402}, \beta_{463}) = (4.0, -4.0, 4.0, 4.0, -4.0, 4.0, -4.0, 4.0, -4.0, -4.0, -4.0, -4.0)$, $(\beta_{58}, \beta_{141}, \beta_{169}, \beta_{174}, \beta_{178}, \beta_{191}, \beta_{293}, \beta_{301}, \beta_{302}, \beta_{308}, \beta_{350}, \beta_{395}, \beta_{480}) = (-3.57, -3.96, -3.51, -2.41, 3.15, -2.17, 3.95, -2.76, 2.27, 2.65, 2.65, 2.69, -2.32)$, $(\beta_{10}, \beta_{57}, \beta_{111}, \beta_{131}, \beta_{171}, \beta_{209}, \beta_{215}, \beta_{327}, \beta_{342}, \beta_{396}, \beta_{416}, \beta_{451}) = (-2.0, -2.0, 2.0, 2.0, -2.0, -2.0, -2.0, -2.0, 2.0, -2.0, 2.0, 2.0)$, $(\beta_{18}, \beta_{50}, \beta_{52}, \beta_{87}, \beta_{110}, \beta_{144}, \beta_{156}, \beta_{206}, \beta_{250}, \beta_{274}, \beta_{309}, \beta_{331}, \beta_{475}) = (1.26, 0.57, 0.17, -0.26, 1.57, -1.26, -0.62, 1.25, 0.08, 1.19, 0.91, -0.71, 1.50)$.

Orthogonal resolution III factorial designs are used to estimate the effect coefficients. For the 200-factor and 500-factor experiments, these designs need 256 and 512 simulation runs, respectively, in a single replication to provide an independent estimate for each main effect in the prescreening. The same approach is also used to construct the factorial design used in the Phase 2 CSFD procedure, and the size of this factorial design depends on how many factors are assigned to group IMP.

Table 6 presents the average numbers of simulation runs (with standard errors in parentheses) required by the CSFD, the FF-CSB, and the hybrid method in selected cases with main-effects response models. The hybrid method is clearly the most efficient one in all scenarios. Its relative savings of simulation efforts when compared with the CSFD method are listed in the last column. When the variance factor is large, the efficiency of the hybrid method approaches that of the CSFD. This is mainly because, with larger variances, the effectiveness of the prescreening procedure drops (since the estimates are less precise), which then affects the efficiency of CSB-X and CSFD in Phase 2 of the hybrid method. When compared with FF-CSB, the benefit of incorporating the CSFD procedure into the hybrid method becomes more obvious when the variance increases.

Table 6. Simulation runs required in examples with main-effects model.

Case	Variance factor (m)	CSFD	FF-CSB	Hybrid	Relative savings (hybrid vs. CSFD) %
200-factor	0.01	512 (0)	392 (0)	336 (0)	34.4
	0.1	559 (127)	798 (97)	371 (13)	33.6
	0.3	792 (98)	2817 (392)	475 (54)	40.0
	1.0	1807 (434)	22523 (3415)	1475 (337)	18.4
500-factor	0.01	1024 (0)	1379 (247)	656 (0)	35.9
	0.1	1304 (420)	4500 (406)	734 (32)	43.7
	0.3	1678 (339)	26986 (2239)	998 (98)	40.5
	1.0	4169 (856)	275101 (37729)	4064 (819)	2.5

Table 7. Selected P(DI)s of the 200-factor example with main-effects models.

Effect	$m = 0.1$			$m = 0.3$			$m = 1.0$		
	CSFD	FF-CSB	Hybrid	CSFD	FF-CSB	Hybrid	CSFD	FF-CSB	Hybrid
$\beta_9 = 4.0$	1.000	0.974	1.000	1.000	0.957	1.000	1.000	0.955	0.990
$\beta_{19} = 4.0$	1.000	0.965	1.000	1.000	0.956	1.000	1.000	0.943	0.996
$\beta_{63} = 4.0$	1.000	0.958	1.000	1.000	0.943	1.000	1.000	0.947	0.995
$\beta_{77} = -4.0$	1.000	1.000	1.000	1.000	0.996	1.000	0.998	0.966	0.994
$\beta_{123} = -4.0$	1.000	1.000	1.000	1.000	0.991	1.000	0.999	0.955	0.998
$\beta_{127} = -3.39$	1.000	0.843	0.999	0.998	0.745	0.933	0.896	0.721	0.841
$\beta_{71} = 3.35$	1.000	0.773	0.994	0.994	0.745	0.880	0.877	0.720	0.790
$\beta_{17} = 2.88$	0.018	0.353	0.207	0.197	0.398	0.383	0.349	0.359	0.397
$\beta_{118} = 2.74$	0.000	0.233	0.011	0.025	0.325	0.176	0.203	0.285	0.232
$\beta_4 = 2.38$	0.000	0.046	0.000	0.000	0.095	0.022	0.028	0.099	0.081
$\beta_{25} = 2.0$	0.000	0.001	0.000	0.000	0.008	0.000	0.001	0.035	0.007
$\beta_{55} = 2.0$	0.000	0.000	0.000	0.000	0.013	0.000	0.003	0.017	0.007
$\beta_{140} = 2.0$	0.000	0.025	0.000	0.000	0.051	0.000	0.000	0.024	0.000
$\beta_{96} = -2.0$	0.000	0.025	0.000	0.000	0.048	0.001	0.001	0.021	0.007
$\beta_{144} = -2.0$	0.000	0.017	0.000	0.000	0.043	0.000	0.002	0.022	0.003
$\beta_{174} = 1.61$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.001
$\beta_{10} = 1.28$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{35} = -0.88$	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
$\beta_{182} = -0.48$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{39} = -0.35$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of runs	559	798	371	792	2817	475	1807	22523	1475
n_0	2	6	1/2/3 ^a	3	20	1/3/5 ^a	6	60	1/8/10 ^a

^aThe order of the initial numbers is “Prescreen/CSB-X/CSFD.”

Selected P(DI)s, i.e., the percentage of times that each effect is declared important, are presented in Table 7 for the 200-factor case (for an effective method, P(DI)s should be greater than $\gamma = 0.95$ for critical effects and less than $\alpha = 0.05$ for unimportant effects). We can see that the CSFD method and the hybrid method have similar effectiveness, and the classification results are much more conservative than the error control requirements for all scenarios. The FF-CSB method is less conservative than the other two methods. Its efficiency deteriorates in the $m = 1.0$ case due to the sensitivity of the CSB-X to the large, heterogeneous variance [16, 21]. When considering both the efficiency and effectiveness, the hybrid method is clearly the best; it achieves the desired error control with significantly fewer runs.

We now investigate the effectiveness of the prescreening procedure. Given $\Delta_0/2$ as the factor assignment threshold, Tables 8 and 9 present the average numbers and percentages of misassignments of factors in Phase 1 of the hybrid method in different scenarios. For example, in the 500-factor case, when the variance factor $m = 1.0$ and the prescreening sample size is 1, “3.96 (9.2%)” means that there are, on average, 3.96 potentially important factors ($|\beta| \geq \frac{\Delta_0}{2}$) mistakenly assigned to unimportant groups, which is 9.2% of all potentially important factors. We can see that when variances are not too large, few misassignments occur. Furthermore, as shown in Table 9, critical misassignments (i.e., assigning critical effects $|\beta| \geq \Delta_1$ to unimportant groups) are even

more unlikely. In the same case above, of the 12 critical factors, the average number of critical misassignments is 0.005 (0.04%); and of the 13 important, but not critical, factors ($\Delta_0 < |\beta| < \Delta_1$), the number of misassignments is 0.395 (3.29%). On the other hand, when the variances are large, there are many unimportant factors assigned to the important group IMP. As we discussed earlier, this kind of misassignment has no influence on the error control, but may decrease the efficiency of the hybrid method.

From Tables 8 and 9, we can also see that the larger the prescreening sample size, the fewer misassignments is in

Table 8. Misassignments of IMP and UNIMP factors in main-effects examples.

Case	Variance factor (m)	Prescreening sample size	Misassignment	
			IMP to UNIMP	UNIMP to IMP
200-factor	0.1	1	0.00 (0.0%)	0.08 (0.0%)
		1	0.14 (0.8%)	0.32 (0.2%)
		1	1.23 (6.8%)	40.7 (22.4%)
	1.0	2	0.70 (3.9%)	15.7 (6.6%)
		3	0.47 (2.6%)	6.76 (3.7%)
500-factor	0.1	1	0.02 (0.0%)	0.20 (0.0%)
		1	0.77 (1.8%)	0.79 (0.2%)
		1	3.96 (9.2%)	119.4 (26.1%)
	1.0	2	2.53 (5.9%)	51.8 (11.3%)
		3	1.92 (4.5%)	24.5 (5.4%)

Table 9. Misassignments of critical and important factors in main-effects examples ($m = 1.0$).

Case	Prescreening sample size	Misassignment	
		Critical factor	Important but not critical factor
200-factor	1	0.002 (0.04%)	0.088 (1.76%)
	2	0.000 (0.00%)	0.013 (0.26%)
	3	0.000 (0.00%)	0.002 (0.04%)
500-factor	1	0.005 (0.04%)	0.395 (3.29%)
	2	0.000 (0.00%)	0.099 (0.83%)
	3	0.000 (0.00%)	0.026 (0.22%)

Phase 1. The prescreening sample size refers to the number of replications of the prescreening factorial design. With fewer misassignments, both the effectiveness and overall efficiency improve CSB-X and CSFD procedures in Phase 2. The optimal prescreening sample size is unknown. However, since the prescreening only accounts for a small percentage of the total simulation effort in main-effects models, especially for cases with large variances, using two or three replications in prescreening could be a safer approach if little knowledge is known about the variance.

We next study the influence of the selection of the factor assignment threshold on the performance of the hybrid method. Table 10 shows the change of the efficiency of the hybrid method with different thresholds in the 500-factor case, with the variance factor $m = 1.0$. In Table 10, the numbers in parentheses are the standard errors of corresponding statistics. Sample sizes of the CSB-X and CSFD procedures in Phase 2 are 15 and 10, respectively. Eleven different thresholds are used. In each case, columns 2–4 list the average numbers of factors assigned to each group in Phase 1, and the average numbers of critical misassignments of factors are given in column 5. Columns 6 and 7 list the simulation runs required by the Phase 2 CSFD and CSB-X procedures, respectively. The total numbers of simulation runs, including the 512 runs in prescreening, are given in the

last column. From Table 10, we can see that if the threshold is too large, many factors with non zero effect coefficients, including some potentially important ones, are assigned to unimportant groups; this may not only result in misclassifications of critical effects, but also significantly increases the simulation effort required by the CSB-X procedure. On the other hand, when the threshold is too small, the majority of the factors, including most of the unimportant ones, are assigned to the important group IMP; thus, the Phase 2 CSFD procedure will have a large design to repeat and CSB-X's ability to identify unimportant factors in a group has little benefit. As shown in Table 10, $\Delta_0/2$ seems to be a good choice of the factor assignment threshold from the perspectives of both the overall efficiency and overall effectiveness, i.e., coming close to minimize the total number of runs required for screening while achieving the desired error control. In addition, these numerical results show that unless the threshold is too large, different thresholds seems to have little impact on the effectiveness of the hybrid method.

4.2. Second-Order Model

We now compare the CSFD method and the hybrid method on a 50-factor example where two-factor interactions exist. Main-effect coefficients are randomly generated in the same way as those in main-effects models, except that the percentage of each non zero effect category is increased from 2.5% to 5.0%. Non zero interaction effects are generated independently of each other. The probability that a two-factor interaction exists is assumed to be 0.2, if both parent factors are important; 0.05, if only one parent factor is important; and 0, if neither parent factor is important. The probability of a non zero interaction effect coefficient being positive (or negative) is 0.5, and the distribution of the absolute values of non zero interaction effect coefficients is as follows: 25% are equal to $\Delta_1 = 4$, 25% are uniformly distributed on $(\Delta_0, \Delta_1) = (2, 4)$, 25% equal to $\Delta_0 = 2$, and 25% of them are uniformly distributed on $(0, \Delta_0) = (0, 2)$. The results

Table 10. Efficiency of the hybrid method with different factor assignment thresholds, 500-factor example ($m = 1.0$).

Threshold	Number of factors			Critical misassignment	Number of runs		Total number of runs
	IMP	POS	NEG		CSFD	CSB-X	
$0.25 \times \Delta_0$	304 (13)	98 (10)	98 (9)	0 (0)	5131.8 (117)	170.9 (174)	5814.7 (209)
$0.30 \times \Delta_0$	269 (13)	116 (10)	115 (10)	0 (0)	4846.1 (761)	198.6 (203)	5556.7 (781)
$0.35 \times \Delta_0$	238 (13)	132 (10)	130 (10)	0 (0)	3226.1 (809)	246.7 (272)	3984.8 (837)
$0.40 \times \Delta_0$	209 (13)	146 (11)	145 (11)	0.001 (0.032)	3005.7 (554)	323.8 (418)	3841.5 (671)
$0.45 \times \Delta_0$	182 (13)	160 (11)	159 (11)	0.005 (0.071)	3019.5 (548)	416.0 (494)	3947.6 (707)
$0.50 \times \Delta_0$	158 (13)	172 (11)	170 (11)	0.005 (0.071)	3021.3 (572)	530.2 (599)	4063.5 (819)
$0.55 \times \Delta_0$	137 (12)	182 (11)	180 (11)	0.006 (0.077)	2872.2 (584)	733.1 (804)	4117.3 (952)
$0.60 \times \Delta_0$	119 (12)	191 (10)	190 (11)	0.007 (0.083)	2521.6 (595)	981.4 (971)	4015.0 (1090)
$0.65 \times \Delta_0$	103 (11)	199 (10)	198 (10)	0.009 (0.094)	2383.1 (484)	1328.2 (1244)	4223.3 (1285)
$0.70 \times \Delta_0$	89 (10)	206 (10)	205 (10)	0.017 (0.129)	2368.0 (508)	1817.4 (1704)	4697.4 (1736)
$0.75 \times \Delta_0$	77 (7)	213 (10)	210 (10)	0.028 (0.177)	2339.3 (519)	2459.9 (2120)	5311.2 (2108)

Table 11. Simulation runs required in the second-order example.

Variance factor (m)	CSFD	Hybrid method				Relative saving %
		Prescreen	CSB-X	CSFD	Total	
0.1	8192 (0)	4096 (0)	16.3 (1.3)	592 (198)	4704 (198)	42.6
0.3	8385 (959)	4096 (0)	24.1 (0.6)	807 (141)	4928 (141)	41.2
1.0	12390 (714)	4096 (0)	40.8 (16)	1925 (560)	6062 (561)	51.1

presented in this section are based on a randomly generated 50-factor case, where there are 10 non zero main effects ($\beta_4, \beta_6, \beta_{14}, \beta_{16}, \beta_{18}, \beta_{20}, \beta_{25}, \beta_{28}, \beta_{33}, \beta_{48}$) = (3.28, 2.22, 1.76, 1.35, 2.0, -0.90, -2.0, -4.0, -4.0, -3.86) and 11 non zero interactions ($\beta_{3,33}, \beta_{4,28}, \beta_{4,33}, \beta_{4,48}, \beta_{6,29}, \beta_{6,33}, \beta_{27,28}, \beta_{28,33}, \beta_{31,33}, \beta_{33,37}, \beta_{33,48}$) = (-4.0, 1.73, -2.18, -0.84, 2.0, 3.82, -3.88, -3.01, 4.0, 0.56, -3.27). All other effects are zero. The CSFD method and the prescreening procedure of the hybrid method run a 2^{50-38} factorial design, which requires 4096 runs in each replication, to provide a set of independent estimates of all main effects and two-factor interactions. For details of the construction of this design, see Ref. 13. The threshold of factor assignment threshold is still $\Delta_0/2$ in this example.

For the example with second-order model, Table 11 presents the average simulation runs required by the CSFD method and the hybrid method. In each case, the averages based on 1000 independent trials and the standard errors are listed in parentheses. The advantage of the hybrid method over the CSFD method is even more obvious when interactions are allowed in the response model. In this 50-factor case, the hybrid method saves more than 40% of simulation effort when compared with the CSFD method, and the saving is not affected by the change of variances. With the same number of factors, experiments with second-order models require a much larger prescreening factorial design than that in experiments with main-effects models. For a 50-factor case, when compared with 64 runs for the main-effects model, second-order models needs 4096 runs in one replication, which is more than half of the total simulation runs required by the hybrid method. Therefore, when the number of factors is not small and interaction effects exist, having more than one replication in prescreening is not recommended.

On the other hand, the increase of the sample size within one replication also provides more accurate effect estimation because the variances of the estimated effect coefficients are substantially smaller. Table 12 presents the prescreening results of the hybrid method in this second-order example, which used a Resolution V design with 4096 design points in prescreening. Compared with Table 8 of main-effects models with Resolution III factorial design for prescreening (with 256 and 512 design points for 200-factor and 500-factor cases, respectively), the percentages of misassignments in Table 12 are significantly smaller with a larger sample size.

In fact, in this 50-factor case, even when the variance factor $m = 1.0$, there are no important factors (whose true effect coefficients are no less than $\Delta_0 = 2.0$) assigned to unimportant groups in all 1000 trials. This significant improvement of effectiveness in Phase 1 makes it possible for the hybrid method to take full advantage of the strength of the CSB-X and the CSFD procedures in Phase 2 to obtain the best efficiency. In fact, in all cases, the CSFD procedure only needs to repeat a Resolution V factorial design of size 256, and the CSB-X procedure needs no more than 1 bifurcation in all 1000 trials.

Selected P(DI)s of this second-order example are presented in Table 13. Both the CSFD method and the hybrid method show no difference in classifying main effects and interactions. The two methods have similar effectiveness results; that is, error control requirements are met at all effects and the classification results are, in fact, much more conservative in effectiveness than the error control requirements. The only difference between the CSFD method and the hybrid method is that the former has slightly higher power for identifying the important, but not critical, effects.

Table 14 presents the values of the lower bounds that the hybrid method provides with the factor assignment threshold $T = \Delta_0/2$ in all scenarios of the numerical experiments discussed in this section. In all examples presented, we assume unequal variances. For each scenario, 1000 independent trials are run and we present the minimal, mean, and maximum values of ζ and $(1 - \zeta)\gamma$ values. We can see that in almost all cases, the values of ζ are very small. Therefore, the lower bound on the overall power of the hybrid method, $(1 - \zeta)\gamma$, is very close to $\gamma = 0.95$, the original power control rate of CSFD procedures. When the variance factor is $m = 1.0$, the values of the upper bound of $\Pr(E_1)$ are consistent with the statistics of critical misassignments presented in Table 9.

Table 12. Prescreening results of hybrid method in the second-order example.

Variance factor (m)	Number of factors			Misassignment	
	IMP	POS	NEG	IMP to UNIMP	UNIMP to IMP
0.1	13.00	17.81	19.19	0.00 (0.00%)	0.00 (0.00%)
0.3	13.07	17.81	19.12	0.00 (0.00%)	0.08 (0.44%)
1.0	13.31	17.84	18.85	0.05 (0.28%)	0.60 (3.33%)

Table 13. Effectiveness results (P(DI)s) in the second-order example.

Effect	$m = 0.1$		$m = 0.3$		$m = 1.0$	
	CSFD	Hybrid	CSFD	Hybrid	CSFD	Hybrid
$\beta_{28} = -4.0$	1.000	1.000	1.000	1.000	1.000	0.999
$\beta_{33} = -4.0$	1.000	1.000	1.000	1.000	1.000	0.999
$\beta_{3,33} = -4.0$	1.000	1.000	1.000	1.000	1.000	1.000
$\beta_{31,33} = 4.0$	1.000	1.000	1.000	1.000	1.000	0.999
$\beta_{27,28} = -3.88$	1.000	1.000	1.000	1.000	1.000	0.995
$\beta_{48} = -3.86$	1.000	1.000	1.000	1.000	1.000	0.998
$\beta_{6,33} = 3.82$	1.000	1.000	1.000	1.000	1.000	0.994
$\beta_4 = 3.28$	1.000	1.000	1.000	0.966	0.986	0.805
$\beta_{33,48} = -3.27$	1.000	1.000	1.000	0.959	0.988	0.818
$\beta_{28,33} = -3.01$	0.754	0.590	0.590	0.542	0.541	0.522
$\beta_6 = 2.22$	0.000	0.000	0.000	0.000	0.000	0.010
$\beta_{4,33} = -2.18$	0.000	0.000	0.000	0.000	0.000	0.001
$\beta_{18} = 2.00$	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{25} = -2.00$	0.000	0.000	0.000	0.000	0.000	0.001
$\beta_{6,29} = 2.00$	0.000	0.000	0.000	0.000	0.000	0.001
$\beta_{14} = 2.76$	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{4,28} = 1.73$	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{16} = 1.35$	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{20} = -0.90$	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{4,48} = -0.84$	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{33,37} = 0.56$	0.000	0.000	0.000	0.000	0.000	0.000
Number of runs	792	2817	478	1807	22294	1474
n_0	2	1/2/2 ^a	2	1/3/3 ^a	3	1/4/6 ^a

^aThe order of the initial numbers is "Prescreen/CSB-X/CSFD."

Moreover, it is easy to see that $(1 - \zeta)\gamma$ is a conservative lower bound of the overall power of the hybrid method. First, even though a critical interaction effect is not considered as potentially important in Phase 1, there is still some chance that the factors associated with this effect are both assigned to the important group IMP because of the existence of other potentially important main or interaction effects. Second, even if a factor with a critical main effect is assigned to an unimportant group, the CSB-X procedure still has a high power for identifying it as important. Effect cancelation may happen if this factor is assigned to the unimportant group with the incorrect sign, i.e., events

$\{-T < \hat{\beta}_k < 0 \mid \beta_k \geq \Delta_1\}$ or $\{0 < \hat{\beta}_k < T \mid \beta_k \leq -\Delta_1\}$ occur, but both are extremely unlikely. Third, both the CSB-X and CSFD procedures (especially with the favorable configurations in Phase 2 of the hybrid method) can provide classification results with power higher than the specified power rate. Therefore, as shown in the numerical results, the hybrid method typically provides much higher power than the user-specified power control rate, γ .

5. CONCLUSIONS

By integrating the prescreening, the CSB-X, and the CSFD procedures into one screening method, the hybrid approach provides a general framework for efficiently and effectively performing factor screening. The existence of the prescreening stage not only eliminates the requirement for prior knowledge of the simulation system but also allows the subsequent screening procedures to perform in their favorable conditions; thus resulting in the hybrid method having a better overall efficiency than its component screening methods in general circumstances. Moreover, the hybrid structure still maintains most of the mathematical error control properties of the component methods. More specifically, the empirical results show that the hybrid method provides the same or even better error control than its component screening methods. Although we implement only fractional factorial designs as the prescreening procedure and in CSFD to screen important effects in the empirical evaluation, other experimental designs such as central composite designs, 3^k factorial designs, and Latin hypercube designs can all be easily incorporated into the hybrid framework on different scenarios. Similarly, the hybrid framework also allows the incorporation of other screening/analysis methods in place of the CSB-X and the CSFD methods.

In summary, the hybrid method we propose is an effective and efficient factor screening strategy for discrete-event simulation. It has obvious advantages over other existing screening methods, especially for experiments that involve

Table 14. Low bounds of the power control rates ($T = \Delta_0/2$).

m	n	Statistics	Main-effects models						2nd-order model		
			500-factor 512			200-factor 256			50-factor 4096		
			Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
0.1	ζ	ζ	0	$4.0e-247$	$1.2e-80$	0	$3.7e-297$	$1.9e-145$	0	0	0
		$(1 - \zeta)\gamma$	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
0.3	ζ	ζ	$1.9e-164$	$2.5e-29$	$1.3e-10$	$1.9e-82$	$6.1e-35$	$6.1e-18$	0	0	0
		$(1 - \zeta)\gamma$	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
1.0	ζ	ζ	$1.3e-16$	$4.0e-4$	$2.8e-2$	$4.2e-9$	$1.5e-4$	$5.1e-3$	$1.7e-49$	8.2-44	$3.8e-38$
		$(1 - \zeta)\gamma$	0.95	0.9496	0.9233	0.95	0.9499	0.9451	0.95	0.95	0.95

a large number of factors and little prior information. The hybrid method has wider applicability because it requires few assumptions and demonstrates a robust performance under different system conditions.

APPENDIX

PROOF OF THEOREM 1: As discussed earlier, misassignments of factors in Phase 1 have no influence on the classification of the unimportant effects. Therefore, the overall Type I Error control property of the hybrid method follows automatically from the Type I Error control properties of CSB-X and CSFD. That is, for any individual unimportant effect, the probability of the hybrid method classifying it as important is no greater than α . For the power control:

$$\begin{aligned} & \Pr \{ \text{Declare effect } k \text{ important} \mid |\beta_k| \geq \Delta_1 \} \\ & \geq \Pr \{ \text{Declare effect } k \text{ important} \mid |\beta_k| \\ & \geq \Delta_1 \text{ and the effect is assigned to IMP group} \} \\ & \times \Pr \{ \text{the effect is assigned to IMP group} \mid |\beta_k| \geq \Delta_1 \} = (1 - \zeta) \gamma \end{aligned}$$

□

Now, we discuss the determination of ζ when an orthogonal factorial design is used as the prescreening method. We need the following proposition:

PROPOSITION 1: If the rank of X is p , then $X^T X$ is a $p \times p$ invertible matrix.

PROOF: Consider linear spaces $L_1 = \{x \in \mathbb{R}^n : x \text{ is a linear combination of the column vectors of } X\}$ and $L_2 = \{x \in \mathbb{R}^n : \langle x, y \rangle = 0, \text{ for any } y \in L_1\}$. That is, any vector from L_2 is orthogonal to L_1 . Then, the dimensions of L_1 and L_2 are p and $n - p$, respectively. Let $\tilde{X}$ be a $n \times (n - p)$ matrix such that its column vectors are a base of L_2 . Then, $X^T \tilde{X} = \tilde{X}^T X = 0$. Let $B = (X, \tilde{X})$. Thus, B is an $n \times n$ matrix and $\text{Rank}(B) = n$. Therefore, $B^T B = \begin{pmatrix} X^T & \tilde{X}^T \\ \tilde{X} & \tilde{X}^T \tilde{X} \end{pmatrix} = \begin{pmatrix} X^T X & 0 \\ 0 & \tilde{X}^T \tilde{X} \end{pmatrix}$ has a rank of n . Thus, $X^T X$ is a matrix of rank p , i.e., it is invertible.

Given a fixed factor assignment threshold, T , if the absolute value of an estimated effect coefficient is no less than the factor assignment threshold, all factors associated with this effect would be assigned to the important group IMP. In other words, the happening of the event $E_1 = \{-T < \hat{\beta}_k < T \mid |\beta_k| \geq \Delta_1\}$ is a necessary condition of a critical misassignment. Therefore, if we have an upper bound of the probability of event E_1 , we can obtain an upper bound of the probability of a critical misassignment, which would further help to obtain an lower bound of the overall power of the classification of critical effects in the hybrid method.

Consider the matrix formation of the response model (1) of an L -factor factorial design used in prescreening, $Y = X\beta + \epsilon$, where

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, X = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1,p-1} \\ 1 & X_{21} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{n,p-1} \end{pmatrix}_{n \times p},$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}, \text{ and } \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Here, n is the number of design points, $p - 1$ is the number of effects of interest, including both main and interaction effects, and $p < n$. $\beta_1, \beta_2, \dots, \beta_{p-1}$ are the unknown effect coefficients. X is the design matrix, and the i th row of X gives the deterministic settings of the factors at i th design point, $z_i = (X_{i1}, X_{i2}, \dots, X_{i,p-1})$, $i = 1, 2, \dots, n$. Random errors are independent, normally distributed random variables, i.e., $\epsilon_i \sim \text{Nor}(0, \sigma^2(z_i))$, $i = 1, 2, \dots, n$, whose variances are unknown and may depend on z_i . Y_i is the observation at i th design point, which is also normally and independently distributed. Notice that it is assumed there is only one replication of observations in prescreening, and therefore there is only one observation at each design point. The least square estimate of the effect coefficients β minimizes $(Y - X\beta)^T(Y - X\beta)$. Then, $\frac{d}{d\beta}(Y - X\beta)^T(Y - X\beta) = -2(Y - X\beta)^T X = 0$ leads to $X^T Y = X^T X \beta$. With Proposition 1, the least square estimates of the effect coefficients are $\hat{\beta} = (X^T X)^{-1} X^T Y$ and $\hat{\beta}$ is an unbiased estimate of β .

For an orthogonal factorial design, (e.g., the Resolution III designs discussed in Section 3.2 and the Resolution V designs proposed in Ref. 13), the rank of the design matrix X is p . Moreover, all entries of an orthogonal design matrix X are either 1 or -1 and the inner product of any pair of columns of X is zero, i.e., $\sum_{i=1}^n X_{li} X_{ji} = 0$, $\sum_{i=1}^n X_{li}^2 = n$, $i, j = 1, 2, \dots, p - 1$, $i \neq j$. Then, $X^T X = nI_{p \times p}$ and $(X^T X)^{-1} = \frac{1}{n}I_{p \times p}$. Thus, $\hat{\beta} = (X^T X)^{-1} X^T Y = \frac{1}{n} X^T Y$, i.e., $\hat{\beta}_k = \frac{1}{n} \sum_{i=1}^n X_{ik} Y_i$, $k = 1, 2, \dots, p - 1$. Since $\hat{\beta}_k$ is a linear combination of Y_i s, which are all normally distributed, $\hat{\beta}_k$ is also normally distributed, i.e., $\hat{\beta}_k \sim \text{Nor}(\beta_k, \text{Var}(\hat{\beta}_k))$. Further, since Y_i s are independent with each other and all X_{ij} 's equal to either 1 or -1 , the variance of the k th estimated effect coefficient is,

$$\begin{aligned} \text{Var}(\hat{\beta}_k) &= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_{ik} Y_i\right) = \frac{1}{n^2} \sum_{i=1}^n X_{ik}^2 \text{Var}(Y_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(Y_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2(z_i). \end{aligned}$$

That is, all estimated effect coefficients in prescreening are normally distributed with the same variance, $\hat{\beta}_k \sim \text{Nor}(\beta_k, \tau^2 = \sum_{i=1}^n \sigma^2(z_i)/n^2)$, $k = 1, 2, \dots, p - 1$.

First, let us assume that the variances of the responses are the same at all design points, i.e., $\sigma^2(z_1) = \sigma^2(z_2) = \dots = \sigma^2(z_n) = \sigma^2$. Based on the regression model, the residual, the residual sum of squares, and the error sum of squares are defined as follows: $\hat{\epsilon} = Y - \hat{Y} = Y - X\hat{\beta} = (I - X(X^T X)^{-1} X^T)Y$, $\text{SS}_E = \hat{\epsilon}^T \hat{\epsilon} = Y^T (I - X(X^T X)^{-1} X^T)Y$, $\text{MS}_E = \text{SS}_E/(n - p) = Y^T (I - X(X^T X)^{-1} X^T)Y/(n - p)$. With equal variances assumption, the following results are standard, $\text{Var}(\hat{\beta}_k) = \sigma^2/n$, $\text{SS}_E \sim \sigma^2 \chi_{n-p}^2$, $\text{MS}_E/n \sim \frac{\sigma^2}{n(n-p)} \chi_{n-p}^2$, and $(\hat{\beta}_k - \beta_k)/\sqrt{\text{MS}_E/n} \sim t_{n-p}$, where t_{n-p} is t -distributed random variable with $n - p$ degrees of freedom. We then have

$$\begin{aligned} \Pr(E_1) &= \Pr(-T < \hat{\beta}_k < T \mid |\beta_k| \geq \Delta_1) \leq \Pr(-T < \hat{\beta}_k < T \mid \beta_k = \Delta_1) \\ &= \Pr\left(\frac{-T - \Delta_1}{\sqrt{\text{MS}_E/n}} < \frac{\hat{\beta}_k - \beta_k}{\sqrt{\text{MS}_E/n}} < \frac{T - \Delta_1}{\sqrt{\text{MS}_E/n}} \mid \beta_k = \Delta_1\right) \\ &= \Pr\left(\frac{-T - \Delta_1}{\sqrt{\text{MS}_E/n}} < t_{n-p} < \frac{T - \Delta_1}{\sqrt{\text{MS}_E/n}}\right) = \text{UB}_e. \end{aligned}$$

Therefore, as long as the factorial design used in the prescreening procedure is not a saturated factorial design, we have enough degrees of freedom to compute the error sum of squares, MS_E , and further obtain UB_e , an upper bound of the probability of event E_1 . For a saturated design, if we are willing to assume that only a fraction of the factors are important, we can use the effect(s) closest to zero to estimate the mean square error.

If the variances are not the same at all design points, then the definition of MS_E is different with that above and the approach above cannot be used

directly. If the variance matrix or the ratios of the variances are known, that is $\epsilon \sim \text{Nor}(0, \sigma^2 W^{-1})$, where W is a known positive definite $n \times n$ matrix, then the weighted least square method could be used to estimate the effect coefficients and the variance of the effect coefficients. Because random errors are assumed to be independent of each other, W is a diagonal matrix. Define $Z = W^{\frac{1}{2}} Y$ and $\tilde{\epsilon} = W^{\frac{1}{2}} \epsilon$, then $Z = W^{\frac{1}{2}} X \beta + \tilde{\epsilon}$ and $\tilde{\epsilon} \sim \text{Nor}(0, \sigma^2 I_{n \times n})$. Thus, the least square estimators of the effect coefficients are $\hat{\beta} = (X^T W X)^{-1} X^T W Y$ and $\text{SE} = Y^T (W - W X (X^T W X)^{-1} X^T W) Y$. It is easy to see that $\hat{\beta}$ is still normally distributed and all analysis for the equal variances case can be paralleled to this known unequal variances case to obtain an upper bound of $\Pr(E_1)$, $\text{UB}_{u(1)}$.

If variances are unequal and unknown, we denote $S^2 = Y^T (I - X(X^T X)^{-1} X^T) Y$. We want to show that when the design matrix X is orthogonal, $\mathbf{E}(S^2) = n(n-p)\tau^2$, and $\frac{S^2}{n(n-p)} = \frac{Y^T (I - X(X^T X)^{-1} X^T) Y}{n(n-p)}$ is an unbiased estimate of τ^2 , the variance that all estimated effect coefficients share. Plugging in $Y = X\beta + \epsilon$, it is easy to verify that $S^2 = \epsilon^T (I - X(X^T X)^{-1} X^T) \epsilon$. Denote $A = (a_{ij})_{n \times n} = I - X(X^T X)^{-1} X^T$, then $S^2 = \epsilon^T A \epsilon = \sum_{i,j=1}^n a_{ij} \epsilon_i \epsilon_j$, and $\mathbf{E}(S^2) = \sum_{i=1}^n a_{ii} \mathbf{E}(\epsilon_i^2) + \sum_{i \neq j}^n a_{ij} \mathbf{E}(\epsilon_i \epsilon_j)$. Since $\mathbf{E}(\epsilon_i) = 0, i = 1, 2, \dots, n$, $\mathbf{E}(\epsilon_i \epsilon_j) = \mathbf{E}((\epsilon_i - \mathbf{E}(\epsilon_i))(\epsilon_j - \mathbf{E}(\epsilon_j))) = \text{Cov}(\epsilon_i, \epsilon_j)$, and $\mathbf{E}(\epsilon_i^2) = \mathbf{E}((\epsilon_i - \mathbf{E}(\epsilon_i))(\epsilon_i - \mathbf{E}(\epsilon_i))) = \text{Var}(\epsilon_i) = \sigma^2(z_i)$. Because ϵ_i s are independent of each other, $\mathbf{E}(S^2) = \sum_{i=1}^n a_{ii} \sigma^2(z_i)$. Since $(X^T X)^{-1} = \frac{1}{n} I_{p \times p}$, we have $A = (a_{ij})_{n \times n} = I - X(X^T X)^{-1} X^T = I - X(\frac{1}{n} I_{p \times p})^{-1} X^T = I - \frac{1}{n} X X^T$. Because all entries in X are either 1 or -1, for $i = 1, 2, \dots, n$, $a_{ii} = 1 - \frac{1}{n} X_i X_i^T = 1 - \frac{1}{n} (1 \times 1 + \sum_{k=1}^{p-1} X_{ik}^2) = 1 - \frac{p}{n} = \frac{n-p}{n}$, where X_i is the i th row vector of the design matrix X . Because $\tau^2 = \sum_{i=1}^n \sigma^2(z_i)/n^2$, we have $\mathbf{E}(S^2) = \sum_{i=1}^n \frac{n-p}{n} \sigma^2(z_i) = n(n-p) \sum_{i=1}^n \frac{\sigma^2(z_i)}{n^2} = n(n-p)\tau^2$. Thus, $\hat{\beta}_k \sim \text{Nor}(\beta_k, \frac{S^2}{n(n-p)})$. Therefore, in an experiment with unknown and unequal variances, at the end of Phase 1 of the hybrid method, we can compute $\tau_1^2 = \frac{S^2}{n(n-p)}$ and use $\text{Nor}(\beta_k, \tau_1^2)$ to approximate the distribution of the estimated effect coefficients $\hat{\beta}_k$, and further to obtain an approximate upper bound of $\Pr(E_1)$ as follows:

$$\begin{aligned} \Pr(E_1) &= \Pr(-T < \hat{\beta}_k < T \mid |\beta_k| \geq \Delta_1) \leq \Pr(-T < \hat{\beta}_k < T \mid \beta_k = \Delta_1) \\ &= \Pr\left(\frac{-T - \Delta_1}{\sqrt{S^2/(n(n-p))}} < \frac{\hat{\beta}_k - \beta_k}{\sqrt{S^2/(n(n-p))}}\right. \\ &\quad \left.< \frac{T - \Delta_1}{\sqrt{S^2/(n(n-p))}} \mid \beta_k = \Delta_1\right) \\ &\approx \Pr\left(\frac{-T - \Delta_1}{\sqrt{S^2/(n(n-p))}} < Z < \frac{T - \Delta_1}{\sqrt{S^2/(n(n-p))}}\right) = \text{UB}_{u(2)}. \end{aligned}$$

Let ζ equal to UB_e , if variances are equal; $\text{UB}_{u(1)}$, if unequal variances are known; and $\text{UB}_{u(2)}$, if unequal variances are unknown. The definition of ζ implies that for each individual critical effect (main or interaction), the probability of the effect being considered as potentially important in the Phase 1 of the hybrid method (and thus having all factors associated with it assigned to the important group IMP) is no less than $(1 - \zeta)$. Further, CSFD procedure has a power of no less than γ to classify this effect as important in Phase 2. Therefore, for any critical effect, the overall power of the hybrid method identifying such an effect as important is no less than $(1 - \zeta)\gamma$. $\square$

ACKNOWLEDGMENTS

This research was partially supported by Purdue Research Foundation (PRF), Naval Postgraduate School Award (N00244-06-C-0002), the Office of the Secretary of Defense's Test Resource Management Center, and the Joint IED Defeat Organization. Preliminary results from this article appeared in Ref. 15.

REFERENCES

- [1] R.M. Block and R.W. Mee, Second-order saturated resolution IV designs, *J Stat Theory Appl* 2 (2003), 96–112.
- [2] F. Campolongo, J.P.C. Kleijnen, and T. Andres, "Screening methods," Sensitivity analysis, A. Saltelli, K. Chan, and E.M. Scott (Editors), Wiley, New York, 2000, pp. 65–80.
- [3] T.M. Cioppa and T.W. Lucas, Efficient nearly orthogonal and space-filling Latin hypercubes, *Technometrics* 49 (2007), 45–55.
- [4] A.S. Hernandez, M. Carlyle, and T.W. Lucas, Breaking barriers to design dimensions in nearly orthogonal Latin hypercubes, Department of Operations Research, Naval Postgraduate School, Monterey, CA, 2009.
- [5] M.D. McKay, R.J. Beckman, and W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* 21 (1979), 239–245.
- [6] R.W. Mee, Efficient two-level designs for estimating all main effects and two-factor interactions, *J Qual Technol* 36 (2004), 400–412.
- [7] D.C. Montgomery, Design and analysis of experiments 6th ed., Wiley, New York, 2005.
- [8] R.H. Myers and D.C. Montgomery, Response surface methodology: Process and product optimization using designed experiments, 2nd ed., Wiley, New York, 2002.
- [9] R. Oh, S.M. Sanchez, T.W. Lucas, H. Wan, and M.E. Nissen, Efficient experimental design tools for exploring large simulation results spaces, Department of Operations Research, Naval Postgraduate School, Monterey, CA, 2009.
- [10] P.J. Sanchez, Efficient generation of resolution VII fractional factorial designs, Department of Operations Research, Naval Postgraduate School, Monterey, CA, 2009.
- [11] S.M. Sanchez, H. Wan, and T.W. Lucas, "A two-phase screening procedure for simulation experiments," Proceedings of the 2005 Winter Simulation Conference, M.E. Kuhl, N.M. Steiger, F.B. Armstrong, and J.A. Joines (Editors), Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2005, pp. 223–230.
- [12] S.M. Sanchez, H. Wan, and T.W. Lucas, A two-phase screening procedure for simulation experiments, *ACM Trans Model Comput Simul* 19 (2009), 1–24.
- [13] S.M. Sanchez and P.J. Sanchez, Very large fractional factorial and central composite designs, *ACM Trans Model Comput Simul* 15 (2005), 362–377.
- [14] H. Shen and H. Wan, "Controlled sequential factorial design for simulation factor screening," Proceedings of the 2005 Winter Simulation Conference, M.E. Kuhl, N.M. Steiger, F.B. Armstrong, and J.A. Joines (Editors), Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2005, pp. 467–474.

- [15] H. Shen and H. Wan, "A hybrid method for simulation factor screening," Proceedings of the 2006 Winter Simulation Conference, L.F. Perrone, F.P. Wieland, J. Liu, B.G. Lawson, D.M. Nicol, and R.M. Fujimoto (Editors), Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2006, pp. 382–389.
- [16] H. Shen and H. Wan, Controlled sequential factorial design for simulation factor screening, *European J Oper Res* 198 (2009), 511–519.
- [17] R.D. Tobias, Saturated second-order two-level designs: An empirical approach. Available at <<http://support.sas.com/rnd/app/papers/minres5.pdf>>. Accessed on 2005.
- [18] L. Trocine and L. Malone, "Finding important independent variables through screening designs: A comparison of methods," Proceedings of the 2000 Winter Simulation Conference, J.A. Joines, R.R. Barton, K. Kang, and P.A. Fishwick (Editors), Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2000, pp. 749–754.
- [19] L. Trocine and L. Malone, "An overview of newer, advanced screening methods for the initial phase in an experimental design," Proceedings of the 2001 Winter Simulation Conference, B.A. Peters, J.S. Smith, D.J. Medeiros, and M.W. Rohrer (Editors), Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2001, pp. 169–178.
- [20] J. Tyssedal, H. Griade, and C.C. Rostad, The use of a 12-run Plackett-Burman design in the injection moulding of a technical plastic component, *Qual Reliab Eng Int* 22 (2006), 651–657.
- [21] H. Wan, B.E. Ankenman, and B.L. Nelson, Controlled sequential bifurcation: A new factor-screening method for discrete-event simulation, *Oper Res* 54 (2006), 743–755.
- [22] H. Wan, B.E. Ankenman, and B.L. Nelson, Simulation factor screening with controlled sequential bifurcation in the presence of interactions, *INFORMS J Comput* (in press).
- [23] J.C.F. Wu and M. Hamada, *Experiments: Planning, analysis, and parameter design optimization*, Wiley, New York, 2000.