ABSTRACT

YAESOUBI, REZA. A Comparison of Factor Screening Methods for Simulation Models. (Under the direction of Dr. Stephen D. Roberts).

Computer simulation models that represent a real-world system consist of a large number of *input variables* which are generally referred to as *factors* in Design of Experiments (DOE). The large number of involved factors makes certain analyses which are usually conducted on the simulation models prohibitive or impractical. These analyses may include building predictive metamodels, finding the optimum factor settings for the simulated system, and etc. *Factor Screening experiments* are intended to examine all or some of the involved factors to identify those with significant effect on a selected response (output). The identified important factors can then be used in subsequent analyses.

This thesis is focused on factor screening methods with promising performance on simulation models from the medical decision making community with a relatively large number of factors. Two groups of factor screening methods are addressed: classical designs which are generally used for physical systems, and recent designs which have been exclusively developed for simulation models. Among the classic designs, 2^k Fractional Factorial (FF) Designs and Central Composite Designs are investigated in depth, because of their superior performance on the simulation models. Among the recent methods developed for simulation, Sequential Bifurcation (SB), folded-over SB (SB-X), Cheng's method, Controlled Sequential Bifurcation (CSB), folded-over CSB (CSB-X), Latin Hypercube Designs (LHD), and Nearly Orthogonal Latin Hypercube (NOLH) designs are addressed. In addition, two methods based on Cheng's method are developed in this thesis: the Modified Cheng's method, and the folded-over Modified Cheng's method (MCh-X). MCh-X is shown in this research that has superior performance compared with FF designs, Cheng's method, and CSB-X for situations where the response has *high homogeneous* variance.

Next, several criteria are considered for evaluating the factor screening methods, and the screening methods are compared based on the proposed criteria. Furthermore, the factor screening experiments are conducted on two available deterministic and stochastic simulation models. For the deterministic medical decision model, 2^k Fractional Factorial Designs, folded-over SB (SB-X), and Nearly Orthogonal Latin Hypercube (NOLH) designs are used; and for the stochastic medical decision model, 2^k Fractional Factorial Designs, folded-over Modified Cheng's (MChe-X), and folded-over CSB (CSB-X) were applied

Finally, based on *quantitative* measures, the performance of each method used for the available simulation models is evaluated in terms of its efficiency (requiring minimum number of runs), effectiveness (accuracy), and cost-effectiveness (achieving the highest accuracy with the least number of runs). Cost-effectiveness, which to the best of our knowledge has never been used as a criterion for evaluating factor screening methods, is introduced as a new measure encompassing both the concept of accuracy and efficiency. The research revealed that for the deterministic model, SB-X and for the stochastic model, MCh-X are the most cost-effective methods.

A Comparison of Factor Screening Methods for Simulation Models

By

Reza Yaesoubi

A Thesis submitted to the Graduate Faculty of
North Carolina State University
In partial fulfillment of the
requirements for the Degree of
Master of Science

Industrial and Systems Engineering

Raleigh

2006

APPROVED BY:

_____

Chair of Advisory Committee:
Dr. Stephen D. Roberts

_____ _____

Dr. James R. Wilson Dr. Yahya Fathi

_____

Dr. Charles E. Smith

Biography

Reza Yaesoubi is a graduate student with co-major in Industrial and Systems Engineering and Operations Research at North Carolina State University, where he will receive his Master of Science degree in December 2006. He is currently a PhD student and has served as both a teaching and research assistant. He received his bachelor's degree in Industrial Engineering from Sharif University of Technology in July of 2004. He is from Tehran, Iran.

Acknowledgments

I would like to express my gratitude to Dr. Stephen Roberts for his support and guidance through the long process of completing this thesis. I would also like to thank Dr. Wilson and Bob Klein for their valuable remarks on this thesis. Additional thanks are due to Ali Tafazzoli who provided extensive help along the way. Finally I would like to express my deep gratitude to my parents for their continuous support and inspiration.

This work was supported, in part, by a contract from Medical Decision Modeling, Inc., located in Indianapolis, IN.

Table of Contents

List of Tables

List of Figures

Terminology

- Experiment: a series of simulation runs in which purposeful changes are made to the input factors of a system in the form of treatments (scenarios) so that we may observe and identify the reasons for changes that may be observed in the output response.
- Design matrix: a matrix where the columns correspond to the input factors, the rows correspond to the design points, and the entries correspond to (possibly coded) levels of each of the input factor.
- Design point: same as treatment
- Observation: the outcome of a simulation run
- Replication: single observation (perhaps a summary observation) from a simulation corresponding to a particular treatment, but with different random number seeds. Multiple replications have different random number seeds and are needed to estimate "experimental error" in a stochastic simulation.
- Run: the outcome of executing the simulation model once with specific values for each factor and with specific random number seeds.
- Scenario: a simulation term that corresponds to an experimental treatment.
- Treatment: a combination of different levels of factors, corresponding to a row in a design matrix.
- Input factor: an independent variable subject to the control of the experimenters that is thought to affect the simulation output response of interest.
- Effect: a partitioning of the variables in the metamodel reflecting their degree of order. For example, there are main effects, interaction effects, second-order effects, and etc.

1. Introduction to Factor Screening

Computer simulation models that represent a real-world system generally consist of a large number of *inputs* or *parameters* which are referred to as *factors* in Design of Experiments (DOE). The output of a simulation model, which is called *response* in DOE, can be explored by using designed experiments. However, due to their size and running time, large-scale simulation models can become prohibitively costly and require time-consuming experimental designs to study their behavior. In many cases where the number of factors is relatively large, analysts assume that they know which factors are more likely to be important. They investigate only a few *intuitively* selected factors. Often they use an inefficient and ineffective one-factor-at-a-time design to check their assumptions regarding the importance of a factor or estimating the main effect of the factors.

A well-designed experiment allows the analyst to examine many more factors while providing insights and information that could not be gained from trial-and-error approaches or by using one-factor-at-a-time designs. *Screening experiments*, which assume that only a few factors are really important (*parsimony principle*), are experiments in which many factors are examined and the objective is to identify those factors (if any) that have significant (important) effects on the selected response. The parsimony principle states that some of the factors are important while others are not; this principle is equivalent to the *Pareto principle*: a few factors are responsible for most of the effect on the response while most factors contribute little.

Screening is generally necessary in the *pilot phase* of complicated simulation studies (Bettonvil et al. (1996)). The factors identified as being important can be further explored in later phases; e.g. the important factors might be cast as a metamodel and this metamodel can be used in optimization. Moreover, the results of factor screening can be used not only for confirming prior expectations (which is as an important step in validating the simulation model), but they are also informative when the simulation provides insights that do not match

expectations; for example, it is possible that a factor believed to be important by the subject-matter experts turns out to be statistically insignificant.

This thesis studies the factor screening methods in the context of some actual medical decision simulations. Both deterministic and stochastic models are addressed. This work does not present a comprehensive list of all available factor screening methods, but rather addresses those that seem most promising by virtue of their attention given in the literature and in simulation conferences. Furthermore, these are readily available or fairly easy to construct.

Many of the classical experimental designs can be used in simulation studies. However, the environments in which real-world (physical) experiments are performed are quite different from the simulation environments. Wan et al. (2003) point out that simulation experiments are significantly different from physical experiments in the following ways:

1. Screening problems in simulation can involve many more factors than real-world problems. In typical physical experiments it is difficult to control more than 20 factors, while in simulation experiments it is easy to control and simulate many combinations of factors because the experiments can be automated.
2. In traditional physical experiments a factor effect is compared to zero. If the effect is found to be statistically significantly different than zero, then the effect is considered to be important. But, when screening a simulation model, we usually require that the magnitude of an effect be greater than a specified threshold before it is considered to be important.
3. In physical experiments, switching from one factor setting to another can be costly (time and money). In simulation, however, switching is comparatively easy. This makes sequential methods especially attractive in simulation.
4. In simulation experiments, common random numbers (CRN) can be implemented to reduce the variance of estimated effects as compared to independent simulations. Controlling random number seeds is not applicable in physical experiments, although the concept is similar to "blocking."

Table 1.1 adapted from Sanchez et al. (2002), list some of the assumptions made in traditional Design of Experiments (DOE) settings, as well as features that characterize many simulation settings.

Table 1.1: The experimental environment

Traditional DOE Assumptions	Simulation Model Characteristics
Small or moderate number of factors	Large number of factors
Linear or low-order effects	Non-linear, non-polynomial behavior
Sparse effects	Many substantial effects
Negligible higher-order interactions	Substantial higher-order interactions
Homogeneous errors	Heterogeneous errors
Normally distributed errors	Various error distributions
Black box model	Substantial expertise exists
Univariate response	Many performance measures of interest

The objectives of this research are as follows:

1. Evaluating and comparing the performance of the factor screening methods on a number of medical simulation models is one of the main objectives of this research. This evaluation is based on criteria which are described in Section 2.5.1.
2. Among those criteria, efficiency (requiring minimum effort to determine the important factors) and effectiveness (accuracy) have the most attention. Effectiveness, however, is not easy to measure, mainly because no precise knowledge about the importance of a factor is available prior to the factor screening experiment. Therefore, generating a procedure for measuring the effectiveness of a factor screening method is an objective of this research.
3. Another goal of this research is to create a new measure that encompasses both concepts of effectiveness and efficiency.
4. Finally, we were interested in improving available factor screening methods and render them better suited to the characteristics of the medical simulation models which are of concern.

Section 1.1 will address the classic experimental designs including factorial, fractional factorial, and group screening designs. Those designs were initially introduced for physical systems, but they have also been used for screening the simulation models. Chapter 2 will present the methods exclusively designed for simulation models and address various issues associated with each. It further evaluates and compares the factor screening methods presented in this thesis. Chapters 3 and 4 will report the results of factor screening experiments on a deterministic model and a stochastic model, respectively. Finally, Chapter 5 presents the conclusions and recommendation for further study.

Throughout this thesis, JMP (2005), statistical software developed by SAS Institute Inc., has been used for generating the required experimental designs and conducting statistical analyses. For the factor screening methods whose designs cannot be generated by JMP, computer code, described in the Appendix, has been implemented in the Visual Studio .NET environment using VB. NET.

Sections 1.1 and 1.2 in this chapter review the following two common traditional statistical designs:

1. Factorial and Fractional Factorial (FF) designs: generally considered the classic factor screening method; as well as, 2^k - FF designs with different resolutions for different levels of complexity of the response. These designs can be augmented to incorporate quadratic terms into the metamodel by using Central Composite Designs (CCD).
2. Group Screening designs: which dates back to 1961. The key idea in this method is to divide factors into some groups and then perform a fractional factorial experiment on the groups. If a group turns out to be significant, subgroups or individual variables within the group are further screened in the same way.

1.1. Factorial and Fractional Factorial Designs

Factorial designs have always been considered an effective approach for screening a relatively small number of factors; i.e. they can detect important factors correctly, as long as

the assumptions are not violated. Factorial and Fractional Factorial (FF) Designs, also called gridded designs, are generally considered classic factor screening methods. Factorial designs are sometimes divided into two groups (Kleijnen et al. (2005)):

1. Fine grids: m^k factorial designs, where each factor can have $m > 2$ levels.
2. Course grids: 2^k factorial designs, where factors have only two levels.

Designs with fine grids are usually too expensive to apply to simulation models, where a large number of factors are involved. On the other hand, course grid designs are usually considered both effective and efficient for screening simulation models. The following sections investigate these methods and the issues related to them when they are applied to simulation models.

1.1.1. Factorial Designs

In a factorial design, in each complete trial of the experiment, all possible combinations of the levels of the factors are investigated. If we want to collect n samples of each treatment, then we need to run n replications for each treatment. In other words, if we want to investigate k factors each at m levels and make n replications for each treatment, then we will need a total of $n \cdot m^k$ runs. The larger the value of m for an m^k factorial design, the better the space-filling properties of the design.

Factorial designs are not generally used for factor screening the simulation models, because they are too inefficient in the scale of real-world experiments, where more than a handful of factors are involved. One of the major advantages of these designs is that they are capable of detecting important factors and interactions by using analysis of variance. In addition, since factors are studied at different levels, nonlinearity in the response function can be detected. Therefore, whenever the simulation run time is minimal, the benefit of the detailed information that this group of designs provides about the nature of the response surface may outweigh the cost of additional computation time.

If we assume that the response function can be approximated by function f, then for each observation y:

$$y = f(x_1, x_2, ..., x_K) + \varepsilon ,$$

where K is the total number of factors and ε is the error term. All factorial designs, including 2^k factorial and 2^k fractional factorial designs which will be discussed later in this chapter, assume that the error term (ε) in the response function is normally distributed with mean zero and constant standard deviation σ. More specifically it is assumed that:

1. Errors are approximately normally distributed.
2. The model is adequate: $E(\varepsilon) = 0$.
3. Error variance is homogenous: $Var(\varepsilon) = \sigma^2$ (applied only to stochastic models).
4. Errors are statistically independent: $Cov(\varepsilon_i, \varepsilon_j) = 0$, for observations $i \neq j$ (applied only to stochastic models).

Most screening methods assume all the preceding assumptions. Therefore, to avoid repetition from now on, these assumptions are referred to as the "*default error assumptions.*"

Analyzing the responses to factorial designs is well understood. There are two major analysis tools that are usually applied:

1. Analysis of Variance (ANOVA): ANOVA can be used as a statistical tool to effectively determine *significant* main effects and interactions.
2. Residual Analysis: Residual analysis is used to check the preceding four assumptions and the adequacy of the regression model (metamodel) generated by the factorial design.

Details about ANOVA and residual analysis techniques can be found in any Design of Experiments (DOE) textbook, such as Montgomery (2000). In Chapters 3 and 4 of this thesis, several residual analysis techniques have been applied to the results of the medical simulation experiments.

Investigating factors at different levels may result in a very expensive design. Therefore, factors are usually studied at only two levels. A factorial design where all factors are at two levels is called a 2^k factorial design, which is one of the most widely used screening methods in simulation.

1.1.2. 2^k Factorial Designs

The 2^k factorial design is the special case of factorial design where each factor has only two levels (lower and upper levels). A set of all the possible treatments of such a design requires 2^k observations and is called a 2^k factorial design. It provides the smallest number of runs with which k factors can be studied in a complete factorial design. For these designs, it is possible to fit a metamodel including all interactions, not only between pairs of factors, but also among triplets, etc. One of the advantages of 2^k factorial designs is that they are very easy to construct and already available in almost all statistical software.

2^k factorial designs are based on two assumptions:

1. Because there are only two levels for each factor, the response function is assumed to be approximately linear over the range of factor levels. This approximation can also be augmented by interactions terms as:

$$y = \beta_0 + \sum_{j=1}^{K} \beta_j x_j + \sum_{i<j} \sum \beta_{ij} x_i x_j + \sum_{i<j<k} \sum \sum \beta_{ijk} x_i x_j x_k + ... + \varepsilon$$

2. The error term (ε) satisfies the default error assumptions mentioned in Section 1.1.1.

For 2^k factorial designs as for all other factorial designs, the adequacy of model can be checked by residual analysis, and the factors and interactions with significant effect can be detected by using the following statistical methods:

3. In many experiments involving 2^k designs, we will examine the magnitude and direction of the factor effects to determine which variables are likely to be important. The *analysis of variance* can generally be used to create an interpretation.
4. *Constructing confidence intervals on the effects* is another approach to identify the significant factors and interactions. If the produced confidence interval for each factor or interaction includes zero, the factor or the interaction is considered to be insignificant; otherwise it is declared to be significant (Montgomery (2000)).

5. Using *normal probability plots* and *half-normal plots* are other approaches to assess the importance of the effects (Montgomery (2000)).

Since 2^k factorial design can estimate all main and interaction effects, it requires a large number of experimental runs to screen a small number of variables. A 2^k factorial design requires 2^k experimental points to estimate all k main effects of k variables, all C(k,2) 2-order interactions, C(k,3) 3-order interaction, …, and C(k,k) k-order interactions, where C(n,m) is the combination of m elements from a set of size n. However, experience suggests that high order interaction effects are not usually significant and a model including only main and low-order interaction effects can be a good fit for the response function. This interpretation leads us to a new set of designs called 2^k fractional factorial designs.

1.1.3. 2^k Fractional Factorial Designs

As the number of factors in a 2^k factorial design increases, the number of runs required for a complete factorial experiment grows rapidly. If the experimenter can reasonably assume that certain higher-order interactions are negligible, information on the main effects and low-order interactions can be obtained by running only a fraction of the complete factorial experiment. The number of required runs in a fractional factorial experiment is much smaller, but the ability to estimate interaction effects is also reduced. Consequently, these designs are widely used in factor screening experiments. Fractional factorial design can be augmented to detect curvature by using middle levels or center points; however they also increase the number of required runs.

The two assumptions made for 2^k factorial design also hold for fractional factorial designs. Furthermore, residual analysis and ANOVA can be used to check the adequacy of the model and to detect the important effects

What makes 2^k fractional factorial designs attractive in factor screening experiments is the efficient number of runs it requires, which is a direct result of factor confounding, i.e. when more effects are confounded, fewer parameters need be estimated and as a result fewer runs will be needed. One of the major concerns with fractional factorial designs is that this design

may confound a significant interaction effect with other effects; and therefore no information can be gained about the individual interaction effects within this confounded structure. The issue of confounding introduces the concept of *resolution* of a design. A design's resolution determines the complexity of metamodels that can be fit to the data if the design is used. Analysts may use a fractional factorial design of resolution III to focus on just finding important main effects, or use a resolution IV design to focus on finding main and selected 2-way interaction effects, or use a resolution V design for finding important main and 2-way interaction effects.

1.1.3.1 Resolution III, IV and V Designs

A design is of resolution *R* if no *p*-factor effect is aliased with another effect containing less than *R* - *p* factors. In other words, the resolution of a design is the minimum number of factors in each pair of *aliases*. As an example, assume that we have three factors: A, B and C. If for building a 2^{3-1} design, the generator C = AB is used, then C and AB are aliases and C = AB is a pair of design aliases. For this design, A = BC and B = AC are other pairs of aliases. Consequently, factor C is confounded with factor AB, factor B with AC, and factor A with BC. Confounding causes information about certain treatment effects (usually high-order interactions) to be indistinguishable from, or confounded with, other effects. If two effects are confounded they cannot be estimated separately. In the preceding example, an estimate for factor C (or AB) is actually an estimate for main effect C plus interaction effect AB. Therefore, neither the main effect C nor the interaction effect AB can be estimated separately.

As pointed out before, the following designs are of particular interest in fractional factorial experiments, especially in simulation:

1. Resolution III designs: No main effect is confounded with any other main effect, however main effects are confounded with two-factor interactions and two-factor interactions may be confounded with each other. Well-known designs of resolution III are *Plackett-Burman* designs which can estimate the main effects of *k* factors in only *k*+1 runs when specific requirements are met. These designs will be discussed later in this chapter.

2. Resolution IV designs: No main effect is confounded with any other main effect or two-factor interaction, but two-factor interactions are confounded with each other. This class of design can be constructed by using *fold-over* experiments with a resolution III design.
3. Resolution V designs: No main effect or two-factor interaction is confounded with any other main effect or two-factor interaction, but two-factor interactions are confounded with three-factor interactions.

Now, we address these designs in more detail.

1.1.3.2 Plackett-Burman Designs

Plackett-Burman designs are two-level fractional factorial designs for studying $k = N$-1 variables in N runs, where N is a multiple of 4 (i.e. $N = 4$, 8, 12, 16, 20, 28, 32 and 36). These designs are the same as 2^{k-p}_{III} fractional factorial designs when $k + 1$ is a power of 2. For $N =$ 12, 20, 24, 28 and 36, the Placket-Burman designs are sometimes of special interest. They are sometimes called *nongeometric* designs (Montgomery (2000)) because these designs cannot be represented as cubes.

The nongeometric Plackett-Burman designs have very messy confounding structures. For instance, in the 12-run design, every main effect is partially confounded with every two-factor interaction not including itself. For example, the AB interaction is confounded with the nine main effects C, D, …, K. Moreover, each main effect in partially confounded with 45 two-factor interactions. Thus, experimenters are usually advised to use these designs *very carefully* (Montgomery (2000)).

If we assume the absence of interactions, and $N = k + 1$ is a multiple of 4, then this design can estimate the main effects in only N runs. If N is not a multiple of 4, N should be rounded up (e.g., for $k = 2$, $N = 4$). Plackett-Burman designs can be found in any textbook on experimental designs such as: Montgomery (2000). Some authors, additionally, have tabulated Plackett-Burman designs for N equals any multiple of 4 from 40 to 100 (Kleijnen (1975b))

Although they are efficient in terms of number of runs required, Placket-Burman designs are not usually *independently* used for factor screening, mainly due to the weaknesses mentioned above. Nevertheless, as we will see later, this design has been used with other screening designs (e.g. group screening designs).

1.1.3.3 Fold-over Designs (Resolution IV Designs)

By combining fractional factorial designs in which certain signs are switched, we can systematically isolate effects of potential interest. This type of experiment is called a *fold-over* of the original design. In general, if we add to a design of resolution III a further design with the signs of a single factor reversed, the combined design will provide estimates of the main effect of that factor and its two-factor interactions.

More generally, if we add to a design of resolution III a second design in which the signs for all the factors are reversed, we will have a type of fold-over which breaks the alias links between main effects and two-factor interactions. So by using this combined design we can estimate all the main effects clear of any two-factor interactions. Resolution IV designs exist for k factor and $N = 2k$ when N is a multiple of 8 (Kleijnen (1975a)).

JMP, can generate resolution IV designs for more than 1000 factors. The number of required runs for different number of factors is summarized in Table 1.2.

Table 1.2: Data requirement for R4 designs

k	No. of design points
4	$2^3 = 8$
5-8	$2^4 = 16$
9-16	$2^5 = 32$
17-31	$2^6 = 64$
32-63	$2^7 = 128$
64-127	$2^8 = 256$
128-200	$2^9 = 512$

1.1.3.4 Resolution V Designs

Until recently it was difficult to construct a very efficient R5 fractional factorial for more than a dozen factors. For example, the largest R5 fractional factorial in Montgomery (2000) is a 2^{8-3}. Sanchez et al. (2005a) have recently developed a method, based on discrete-valued Walsh functions, for rapidly constructing very large R5 fractional factorial designs – a simple Java program generates designs up to $2^{120-105}$ in under a minute. These allow all main effects and two-way interactions to be fit, and may be more useful for simulation analysts than saturated or nearly-saturated designs. The sizes of the resulting designs are given in Table 1.3.

Table 1.3: Data requirement for R5 designs

k	No. of design points	k	No. of design points
1	$2^1 = 2$	18-21	$2^9 = 512$
2	$2^2 = 4$	22-29	$2^{10} = 1{,}024$
3	$2^3 = 8$	30-38	$2^{11} = 2{,}048$
4-5	$2^4 = 16$	39-52	$2^{12} = 4{,}096$
6	$2^5 = 32$	53-69	$2^{13} = 8{,}192$
7-8	$2^6 = 64$	70-92	$2^{14} = 16{,}384$
9-11	$2^7 = 128$	93-120	$2^{15} = 32{,}768$
12-17	$2^8 = 256$		

1.1.3.5 Further discussion on 2^k Fractional Factorial Designs

Factorial designs have several attractive properties. Since more than one factor can be examined at a time, these designs can identify important interaction effects. They are also *orthogonal* designs, i.e. the pairwise correlation between any two columns (factors) in the design matrix is equal to zero. This simplifies the analysis of the output (response) we get from running our experiment, because estimates of the factors' effects (β_i's) and their contribution to the explanatory power of the regression metamodel will not depend on what other explanatory terms are placed in the regression metamodel (Sanchez (2005b)). In addition, ordinary least squares regression assumes that the ε's are also identically distributed, but the regression coefficients are still unbiased estimators even if the underlying variance is not constant. However, examining each factor at only two levels (the low and high values) does not reveal how the simulation output behaves for factor combinations in

the interior of the experimental region. Moreover, it is possible that the choice of low and high level for factors cancels the interaction (Trocine et al. (2000)). This is depicted in Figure 1.1. The chart in left side shows how the interaction would appear given a 2^k design while the chart in right side shows a strong interaction when another level of factor A is included. Factorial Designs (finer grids) can reveal complexities in the region of interest.

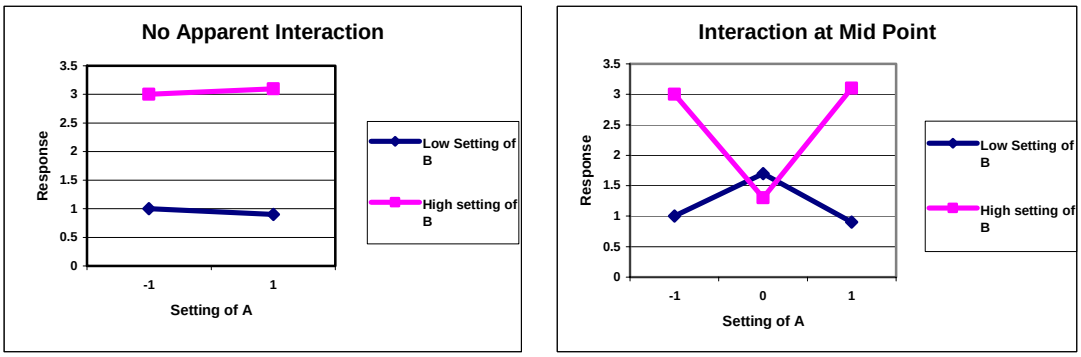


Figure 1.1: Interaction effect in 2^k factorial designs

Clearly, fractional factorial designs are more efficient than factorial designs, but it is more complicated to appropriately design a fractional factorial. One of the major concerns about the fractional factorial design is that this design may confound the interaction of significant importance with other effects; thus the result is combined with other main effects or interactions and nothing can be determined about the individual interactions within this confounded structure. Consequently, designing a fractional factorial should be done with care, so that none of the effects in which we are interested are confounded with other effects.

Another potential concern with the use of two-level factorial designs is the assumption of linearity in the factor effects. Of course, perfect linearity is unnecessary, and the 2^k system will work quite well even when the linearity assumption holds only very approximately. However, it is noticed that if the interaction terms are added to the main effects or first-order model, then we have a model capable of representing some curvature in the response function (Montgomery (2000)). This curvature results from the twisting of the plane induced by the interaction term $\beta_{ij}x_ix_j$. The first-order model is:

$$y = \beta_0 + \sum_{j=1}^{K} \beta_j x_j + \sum_{i<j} \sum \beta_{ij} x_i x_j + \varepsilon$$

In the situations where the curvature in the response function is not adequately modeled by the above equation, a second-order response surface model can be used:

$$y = \beta_0 + \sum_{j=1}^{K} \beta_j x_j + \sum_{i<j} \sum \beta_{ij} x_i x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \varepsilon \quad (1.1)$$

where the β_{jj} represent pure second-order or quadratic effects.

Adding *center points* to 2^k design provides protection against curvature from second order effects. This method consists of n replications at the point $(x_1,...,x_k) = (0,...,0)$.

It is not possible to estimate the unknown parameters in this model because the number of independent runs is less than the number of unknown parameters. A simple and highly effective solution to this problem is to augment the 2^k design with *axial runs*. The resulting design, called a *central composite design*, can now be used to fit the second-order metamodel.

1.1.4. Central Composite Designs (CCD)

Since 2^k factorials or fractional factorials sample each factor at only two levels, only main effects and second-order interactions can be estimated by these designs, and it does not reveal what happens to the simulation's response in the middle of the factor ranges. Using a 3^k factorial design lets us estimate quadratic effects, but it requires more design points, especially when k is large. Central Composite Designs (CCD) are the most popular class of designs used for fitting a second-order model. In practice, a 2^k design can be used to fit a first-order model, and if the model exhibits lack of fit, axial runs are then added to allow the quadratic terms to be incorporated into the model (Montgomery (2000)).

Generally, the CCD consists of a 2^k factorial (or fractional factorial of resolution V) with n_f factorial runs, $2k$ axial or star runs, and n_c center runs. In the coded designs, if -1 and $+1$ are the low and high levels, respectively, then the center point occurs at (0, 0, ..., 0), the first pair of axial (star) points are ($-c$, 0, ..., 0) and (c, 0, ..., 0); the second pair of axial (star) points

are $(0, -c, 0, \ldots, 0)$ and $(0, +c, 0, \ldots, 0)$, and so on. There are two parameters in the design that must be specified: the distance c of the axial runs from the design center and the number of center points n_c. For n_c, Montgomery (2000) recommends using three to five center runs. It is possible to determine an optimal value of c if the white-noise assumption (i.e. the residuals of the model are normally distributed and IID) holds. Since this assumption does not hold for most simulation experiments, Kleijnen et al. (2005) believe there is no need to worry too much about the choice of c. They, however, suggest that analysts choose an intermediate value for better space filling. More information about the CCD can be found in any DOE textbook, such as Montgomery (2000).

1.2. Group Screening

Group screening is usually used for screening a large number of factors. The key idea in this method is to divide factors into groups and then perform a fractional factorial experiment on the groups. If a group turns out to be significant, subgroups or individual variables within the group are further screened in the same way.

In group screening designs k factors are grouped into g groups and each group is considered as a single factor. All factors in that group are at their high (low) level. Then these g group-factors are tested and if a group-factor is found to be insignificant, all factors within that group-factor are regarded insignificant and can be dropped from further investigation. If a group-factor is found to be significant, then one or more factors in that group are significant, and further analysis should be done on that group-factor.

Group screening can be done in two stages or multi stages:

1. In two-stage group screening which was introduced by Watson (1961), the g group-factors are tested in the first stage in order to identify important group-factors and then in the second stage, each factor in the significant groups is tested individually.

2. In multi-stage group screening which was introduced independently by Patel (1962) and Li (1962), groups that are found to be significant in the first stage (or more generally in stage *i*) are repartitioned into smaller groups that will be tested in the next stage.

Now, we discuss two-stage and multi-stage group screening in more details.

1.2.1. Two-Stage Group Screening

Watson (1961) used the following assumptions:

I. all factors have, independently, the same prior probability of being effective, p,
II. effective factors have the same effect, $\Delta > 0$,
III. there are no interactions present,
IV. the required designs exist,
V. the directions of possible effects are known,
VI. the errors of all observations are independently normal with a constant known variance σ^2,
VII. $k = g*f$, where g is number of groups and f is number of factors per group.

By assumption (V), we can define the upper (lower) level of each factor the level which may lead to the highest (lowest) response. The upper (lower) level of a group-factor is defined when all the factors within the group are set at their upper (lower) levels. In this way, and with assumption (III), there is no chance of the cancellation of effects, thus a group-factor with at least one significant factor will give a nonzero effect.

If $g + 1$ is divisible by 4, then assumptions (III) and (IV) allow us to unbiasedly estimate the main effect of each group-factor, by using Plackett-Burman designs in only $g + 1$ runs. Otherwise, 2^{k-p}_{III} would be used in order to identify the significant group-factors. Watson (1961) discussed an example of applying group screening on an experiment with 9 factors.

An important question here would be: what level should be assigned to factors within the insignificant groups in Stage 2? Kleijnen (1975b) proved that it is best to set an insignificant factor either at its low or high level in *all* runs of Stage 2. He shows that in this way a

possible main effect of this factor is confounded with the grand mean. However, setting the factor at its low level for some runs and at its high level in some other runs of Stage 2 confounds its main effect with the main effect of the significant factors.

Later, Kleijnen (1975b) showed that the assumptions (I) through (VII) are not very restrictive:

I. Assumption (I) should be interpreted as follows: we need some *prior rough estimate* for the number of factors thought to be significant among the total of k factors. If we have some prior rough estimate of the number of significant factors, the optimal group-size f, derived by Watson (1961), will be:

$$f_0 = \frac{1}{\sqrt{(1-\alpha_1)p}} \qquad (1.2)$$

where: f_0 is the size of each group in the first stage, α_1 is the significance level in the first stage (i.e. the probability of declaring an unimportant group-factor to be important), and p is the probability of a factor being significant.

II. Assumption (II) is needed to derive the optimal group size and therefore is not critical.

III. For assumption (III), Kleijnen (1975b) showed that a two-factor interaction β_{xy} biases the estimator of the main effect of a factor z, only if the factors x, y, and z belong to three different group-factors (in addition, they believe that pure quadratic effects β_{ii} never bias the estimator of the main effect.) Therefore, if we assume that a two-factor interaction exists only between particular factors, then we should place those *related* factors in the same group. It has been further shown that if we examine the group-factors of Stage 1 in a resolution IV design, then main effects are not biased by any two-factor interaction (but they are still confounded with each other within the same group-factor). For proof, refer to Kleijnen (1975b) pages 419-422.

IV. Assumption (IV) was introduced in order to derive the group-size that by using Plackett-Burman designs minimizes the number of runs in stages one and two. However, in practice the number of group-factors in Stage 1 and number of factors in Stage 2 are not always a multiple of four. Thus, the number of required runs is not always optimal; however, this does not invalidate the procedure.

V. Assumption (V) can be weakened, because first, the probability of having two important factors in the same group is insignificant and second, factors with unknown direction can be studied separately.

VI. Assumption (VI) is needed for derivation of the optimal design and also makes ANOVA possible. It has been shown that ANOVA is robust with regard to nonnormality and heteroscedasticity.

VII. Equation 1.2 implies that factors with high probability of being significant should be placed in small groups. So different estimates of p (which are more realistic) can be incorporated into the group screening method and as a result, the group-size is no longer a constant; so assumption (VII) can be replaced by the less restrictive assumption:

$$k = \sum_{i=1}^{J} g_i f_i$$

where g_i is the number of groups of size f_i, and J is the total number of group-factors in stage (1).

Furthermore, unequal group sizes make it possible to test a factor individually when we do not know the direction of its effect. In addition, the number of group-factors tested in stage one can be chosen such that $\sum_{i=1}^{J} g_i$ is a multiple of four, in which case a saturated Plackett-Burman design is possible.

1.2.2. Multi-Stage Group Screening

In multi-stage group screening, groups that are found to be significant in the first stage (or more generally in stage i) are repartitioned into smaller groups that will be tested in the next stage. This procedure continues until all the remaining groups from the previous stage are of

size one. The assumptions of multi-stage group screening are the same as those of two-stage group screening.

Earlier research on multi-stage group screening concerned mostly the optimum number of groups and group-size in each stage, and estimating the prior probability (p) of a factor being important. Patel (1962) derived that the expected number of runs over all (n + 1) stages is minimized by choosing the number of groups as calculated in Equations 1.3 and 1.4.

$$g_1 \approx kp^{n/(n+1)} \tag{1.3}$$

$$g_2 = g_3 = \ldots = g_n = g_{n+1} \approx p^{-1/(n+1)} \tag{1.4}$$

where, g_i does not denote the total number of groups in Stage I, but the number of groups into which g_{i-1} is split, and optimal group-sizes can be found by:

$$f_i \approx p^{-\{n-(i-1)\}/(n+1)}$$

Patel (1962) and Li (1962), who originally developed multi-stage group screening, assumed that groups in each stage have the same size (f_i), but group-size may differ from one stage to the next one. Later, Kleijnen (1987) argued that having groups with different sizes would be beneficial in some senses: first, it would reduce the required number of runs if we can use Plackett-Burman designs, and more importantly, we can study a factor with unknown sign in a group of size one.

1.2.3. Further discussion on Group Screening designs

As discussed previously, Kleijnen (1975a) showed that the most crucial assumptions for group screening designs are:

1. There is no interaction; however, he showed that two-factor interactions do not bias the main effects if we use a design of resolution IV or higher for group-factors.
2. The direction of main effects are known; yet, we can study a factor with unknown sign in a group of size one to overcome this restriction.

An important issue with group screening is how to group factors at each stage; however, in practice we have prior knowledge about the simulated system and we can use that knowledge

to form groups of related factors. In grouping factors, Ivanova et al. (1998) recommend: (1) a factor with unknown direction should be placed in a single group, (2) factors with assumed (positive) important effects should be placed in one group, and (3) factors with assumed (positive) small effects should be placed in one group.

Another issue is that interactions between variables that are in different groups are not measured. Finally, Trocine et al. (2000) have shown that the results of a group screening experiment and a fractional factorial experiment on a set of fairly small number of variables (17 factors in their example) can result in two different sets of important factors. Thus, a decision about how to group factors is of great importance in yielding correct results.

Mauro (1986) mentioned that there are two major advantages to group screening designs. First, since factors within a group are completely confounded and factors in different groups are not confounded we can, to a certain extent, control the confounding pattern. Secondly, the grouping process reduces the dimensionality of the model and enables the use of orthogonal main effect designs, such as Plackett-Burman designs, to test the significance of group-factors. Moreover, such designs can be analyzed by the usual analysis of variance procedures for factorial experiments.

1.3. Conclusion

This chapter described the factorial, fractional factorial and group screening designs, which have been used as the original factor screening methods for physical systems. This chapter also discussed the subsequent issues that usually occur when these methods are applied.

More recently, several factor screening methods have been introduced which are exclusively designed for simulation models. In the next chapter, those designs will be addressed.

2. Factor Screening Methods for Simulation

This chapter reviews several factor screening methods which have been recently applied to simulation models. The procedure used by each method as well as the issues that may arise when applying the methods are also discussed in this chapter. The methods addressed are as follows:

1. Deterministic Sequential Bifurcation (SB): has gained a great deal of attention since its introduction in 1997. It is fundamentally based on the group screening design but uses a sequential procedure to determine the important factors.
2. Cheng's method, the Modified Cheng's method and the folded-over Modified Cheng's method (MCh-X): all are versions of SB designed for stochastic simulations.
3. Controlled Sequential Bifurcation (CSB): is another version of SB for stochastic simulations which has less strict assumptions than Cheng's or the Modified Cheng's method.
4. Latin Hypercube Design (LHD): the first design specifically proposed for deterministic computer simulations in 1979, known for its good space-filling property.

In addition, different factor screening methods have different characteristics in terms of structure, capability and performance. In this chapter, several criteria are considered for evaluating the factor screening methods. Then, based on the proposed criteria, the screening methods are evaluated and compared.

2.1. Deterministic Sequential Bifurcation

2.1.1. Introduction

Originally, Sequential Bifurcation (SB) was originally developed in the doctoral dissertation by Bettonvil in 1990, and summarized by Bettonvil et al. (1996). SB is designed to find the important factors in simulation models that have many (for example, 300) factors. The SB procedure introduced by Bettonvil et al. (1996) was mainly designed for deterministic simulation models.

SB uses group screening, but instead of applying a fractional factorial design, it uses a binary search to determine important factors. SB, like all other group screening techniques assumes a low-order polynomial metamodel for the response function of the simulation model, and known signs or directions for the first order or main effects. Known signs of the main effects are assumed by all group screening techniques, in order to assure that individual effects do not compensate for the effects of each other within a group.

The criterion that SB uses for declaring a factor as important is the absolute value of the factor's main effect. In other words, in SB, a factor is called important if and only if its main effect is "important." SB is both effective and efficient; that is, it does find important factors, and it requires relatively few simulation runs.

SB is a combination of two types of screening designs: group screening designs and sequential step-down designs:

- Group screening methods have been widely used for the situation with a large number of factors. As described in Section 1.2, the basic idea behind group screening is simple: if several factors can be aggregated into a group, and a screening test on that group shows that this group of factors does not have any significant effect on the output, then all the factors in the group can be considered as unimportant and eliminated from the list of potentially important factors. On the other hand, if the screening test shows that the group of factors has a significant effect, it is inferred that at least one of the factors in that group is important; therefore, the groups should be split into subgroups or individual factors for further screening tests.
- In a sequential design, the design point (factor combination to be studied) at each stage is determined as the experimental results become available. Therefore, as the experiment progresses, based on the result of screening at each stage, the next design point or group of design points are selected.

SB starts with all factors of interest in a single group and tests that group's effect. If the group's effect is important, indicating that at least one factor in the group may have an important effect, then the group is split into two subgroups. The main effects of these two subgroups are then tested and again each subgroup is either classified as unimportant or split into two subgroups for further testing. This procedure continues until eventually all factors that have not been classified as unimportant are tested individually, where there will be only one factor in each remaining subgroups.

The procedure discussed above is the same for all variants of SB. Here we address two variants of SB: SB in the absence of interactions and SB in the presence of interactions (SB-X)

2.1.2. Sequential Bifurcation in the Absence of Interactions

2.1.2.1 Introduction

The simplest variant of SB assumes that the simulation output can be modeled by a first-order polynomial function. In this situation, SB is considered a supersaturated design, where the total number of required runs (n) is smaller that the total number of involved factors (K). This variant of SB requires the following assumptions:

Assumption (SB-1): First-order polynomial approximation gives "negligible" approximation error over the experimental domain of the simulation model.

It is convenient to transform an original, quantitative factor into a standardized variable (say x) that has the value -1 and +1, where -1 (and +1, respectively) correspond to the level that generates a low (and high) output. The simplest approximation of the response function is a first-order polynomial in terms of the standardized variables, which has main effects β_j and overall mean β_0. Hence the response function can be estimated by:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_K x_K \qquad (2.1)$$

where:

y: response of the metamodel (the sequential bifurcation approximation)

K: total number of factors in the experiment

β_j : first-order or main effect of factor j with $j = 1, \ldots, K$

x_j: value of factor j, standardized to lie in [-1,+1]

Based on the Taylor approximation, every differentiable function at point x can be approximated by a Taylor expansion around point x. Thus the polynomial approximation seems reasonable for representing the metamodel in a *vicinity* around x. A polynomial approximation implies that the underlying simulation model is treated as a black box. The advantage of a black box is that it can be applied simply to all types of random and deterministic simulations. The disadvantage is that it cannot exploit the special structure of the simulation model at hand (Bettonvil et al. (1996)). However the assumption (SB-1)

usually does not hold over the entire factor space, but it may be a reasonable assumption for, e.g., small variations in a region of interest.

If for factor z_j, we define L_j as the level of factor j that generated a low value for the output y, and H_j as the level that generates a higher value, provided that factor z_j has any effect at all, then the transformation of a factor to its corresponding standardized factor (x_j) will be as follows:

$$x_j = \frac{z_j - (H_j + L_j)/2}{(H_j - L_j)/2} \qquad (2.2)$$

The scaling in Equation 2.2 implies that in the first-order polynomial metamodel, the most important factor is the one with the largest absolute value of its first-order effect or main effect (β_j); the least important factor is the one with the effect closest to zero. Apparently, the importance of factors depends on the experimental domain (experimental area to be explored). Kleijnen et al. (2003a) noticed that the larger the range of an untransformed factor is, the larger the response difference and hence the main effect of the transformed factor is.

As mentioned before, in the metamodel given by Equation 2.1, it is assumed that approximation errors are negligible. Bettonvil et al. (1996) defines *negligible approximation errors* as errors that are "small" relative to the factor effects. In other words, it is assumed that Equation 2.1 represents a perfect fit to the real output.

Assumption (SB-2): The direction of the effect that a factor may have on the response function is known.

All group screening methods, including SB, assume known signs for the main effects. In order to satisfy this assumption, L_j can be defined as the level of factor j that generated a low value for the output y, and H_j as the level that generates a higher value, provided that factor j has any effect at all. Consequently, if an increase in factor j reduces the output, then the upper and lower levels of that factor should be switched such that $H_j < L_j$. Assumption (SB-2) together with the definition of H_j and L_j, implies that all main effects in Equation 2.1 are non-negative: $\beta_j \geq 0$.

The values of the β_i typically will be unknown, but in most situations it is quite realistic to assume that their signs are known. For example consider a queuing model where Y represents the average waiting time of customers in queue, and x_i is the service rate, it is fairly clear that the main effect of x_i is negative even if the magnitude is unknown. Please refer to Section 2.1.4.8 for cases where this assumption is not met.

2.1.2.2 Design of SB

The symbol $y_{(j)}$ denotes the response value when the factors 1, ..., j are set at their upper levels (H_j) and the remaining factors (j+1, ..., K) are set at their lower levels (L_j). Therefore, in the standardized Metamodel 2.1, the values of the first j factors are +1 and the values of the remaining factors are -1. So the polynomial in Equation 2.1 yields:

$$y_{(j)} = \beta_0 + \beta_1 + ... + \beta_{j-1} + \beta_j - \beta_{j+1} - ... - \beta_K, \qquad j = 0, 1, 2, ..., K$$

The symbol $\beta_{j\text{-}j'}$ denotes the sum of individual effects β_j through $\beta_{j'}$ with $j' > j$. Therefore:

$$\beta_{j-j'} = \frac{y_{(j')} - y_{(j-1)}}{2}$$

2.1.2.3 SB Procedure

SB is a *sequential* procedure where the selection of the next factor combination to be simulated depends on the outputs of the previous combinations already simulated. At the start (Stage 0) of the procedure, SB always observes the two responses corresponding to the two extreme factor combinations, namely $y_{(0)}$ (all the factors at their low levels) and $y_{(K)}$ (all the factors at their high levels). The presence of important factors implies that at the end of Stage 0, SB gives $y_{(0)} < y_{(K)}$. In this situation, SB split the factor into two different groups of factors. If a group size is a power of 2, then SB splits that group into two equal groups; otherwise, SB splits that group at a point such that the first subgroup gets a size equal to the largest possible power of two.

As an example, assume a situation where among 16 factors, factors 5 and 8 are important. The SB procedure applied on this group of factors is presented in Figure 2.1. Each row depicts a stage of the SB procedure, starting with Stage 0.

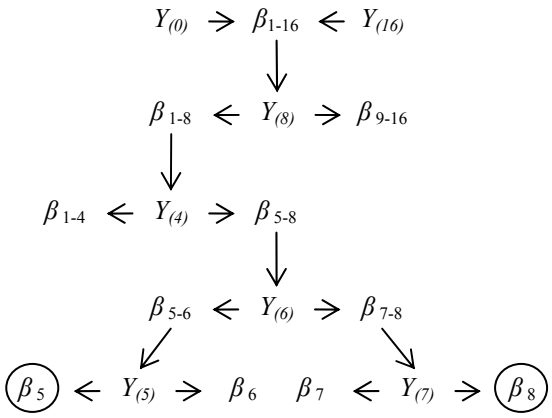


Figure 2.1: SB for $K = 16$

In this example, the presence of important factors implies that at the end of Stage 0, SB gives $y_{(0)} < y_{(16)}$. Hence, SB infers that the sum of all individual main effects is important: $\beta_{1\text{-}16} > 0$. Then, SB splits the group of factors into two groups with equal sizes. At Stage #1, the comparison between $y_{(0)}$ and $y_{(8)}$ implies that there should be at least one important factor among the first 8 factors: $\beta_{1\text{-}8} > 0$. Group-factor $\beta_{1\text{-}8}$ is then split into $\beta_{1\text{-}4}$ and $\beta_{5\text{-}8}$ .This procedure continues until all important factors (5 and 8 in this example) are detected.

2.1.2.4 Metamodel

SB uses a first-order polynomial, as presented in Equation 2.1, to approximate the output. Therefore, in order to build a metamodel for SB we only need to have estimates of main effects. SB, however, only provides the main effect estimates of the important factors. So an approximate metamodel in this case will be:

$$y = \beta_0 + \sum_{j \in I} \beta_j x_j \qquad I = \text{set of important factors}$$

SB does not estimate the overall mean (β_0).To estimate the β_0, one more observation should be taken at central points (i.e. all the factors set at their central level).

2.1.3. Sequential Bifurcation in the Presence of Interactions (SB-X)

2.1.3.1 Introduction

By definition, an interaction means that the effect of a specific factor depends on the levels of other factors. In order to estimate the main effects in the presence of interaction, Bettonvil et al. (1996) used a fold-over design which is considered a resolution IV design. As previously discussed, the classic group screening designs assume a first-order polynomial. Kleijnen (1975b), however, proved that two-factor interactions do not bias the main effect estimators if a resolution IV design is used for the group screening. It means that groups of main effects are estimated without bias from two-factor interactions if sequential bifurcation uses a fold-over design. It should be noted that using a fold-over design when applying SB doubles the number of required runs. SB-X requires the following assumptions:

Assumption (SBX-1): A first-order polynomial metamodel augmented with two-factor interactions gives zero approximation error. Therefore, the simulation output can be approximated by a second-order polynomial function as follows:

$$y = \beta_0 + \sum_{j=1}^{K} \beta_j x_j + \sum_{j=1}^{K-1} \sum_{j'=j+1}^{K} \beta_{j,j'} x_j x_{j'} \quad (2.3)$$

where,

y: response of the metamodel (the sequential bifurcation approximation)

K: total number of factors in the experiment

β_j : first-order or main effect of factor j with j =1, …, K

$\beta_{j;j'}$: interaction effect of the factors j' and j with $1 \leq j < j' \leq k$

x_j: value of factor j, standardized to lie in [-1,+1]

Assumption (SBX-2): the signs of all main effects are known so $\beta_j \geq 0$ (j = 1, …, k)

In the first-order approximation (Equation 2.1) all main effects were assumed non-negative: $\beta_j \geq 0$. But, what does this assumption mean in the presence of interactions? In a 2^k factorial design, the main effect of factor j is defined as the difference between:

1. the average output over all 2^{k-1} combinations of other factors, when the factor j is at its upper level, and

2. the average output over all 2^{k-1} combinations of other factors, when the factor j is at its lower level.

It can be verified that in a metamodel with interactions as presented in Equation 2.3, the main effect of factor j is given by β_j. Therefore, the known direction assumption leads to $\beta_j \geq 0$ (j = 1, …, k)

Assumption (SBX-3): If factor j is unimportant, then $\beta_{ij} = 0$, for any $i \neq j$. SB-X may give misleading results if two factors have unimportant main effects but an important interaction. Therefore, this assumption is added.

2.1.3.2 Design of SB-X

The symbol $y_{(j)}$ denotes the response value when the factors 1, …, j are set at their upper levels (H_j) and the remaining factors (j + 1, …, K) are set at their lower levels (L_j). Therefore, in the standardized metamodel of Equation 2.3, the values of first j factors are +1 and the values of the remaining factors are -1. So the polynomial in Equation 2.3 yields:

$$y_{(j)} = \beta_0 + \beta_1 + \ldots + \beta_{j-1} + \beta_j - \beta_{j+1} - \ldots - \beta_K + \sum_{i<i'\leq j}\sum \beta_{i,i'} + \sum_{j<i<i'}\sum \beta_{i,i'} - \sum_{i\leq j<i'}\sum \beta_{i,i'}$$

for j= 0, 1, 2, …, K

The symbol $y_{-(j)}$, called the mirror observation of $y_{(j)}$, denotes the response value when the factors 1, …, j are set at their lower levels (L_j), and the remaining factors (j + 1, …, K) are set at their upper levels (H_j). Therefore, in the standardized metamodel of Equation 2.3, the values of the first j factors are -1 and the values of the remaining factors are +1. So the polynomial in Equation 2.3 yields:

$$y_{(-j)} = \beta_0 - \beta_1 - \ldots - \beta_{j-1} - \beta_j + \beta_{j+1} + \ldots + \beta_K + \sum_{i<i'\leq j}\sum \beta_{i,i'} + \sum_{j<i<i'}\sum \beta_{i,i'} - \sum_{i\leq j<i'}\sum \beta_{i,i'}$$

for j= 0, 1, 2, …, K

Obviously, the definitions of $y_{(j)}$ and $y_{-(j)}$ imply that $y_{-(0)} = y_{(K)}$ and $y_{-(K)} = y_{(0)}$.

Based on the definitions above, the following equation yields an unbiased estimate for the main effect of factor j as:

$$\beta_j = \frac{(y_{(j)} - y_{-(j)}) - (y_{(j-1)} - y_{-(j-1)})}{4} \qquad (2.4)$$

2.1.3.3 Procedure

The SB-X procedure fundamentally follows the same logical procedure used by SB. In Stage 0, SB-X takes four observations, two of which are equal: $y_{(0)} = y_{-(K)}$ and $y_{(K)} = y_{-(0)}$. Equation 2.4 yields:

$$\beta_{1-K} = \frac{(y_{(K)} - y_{-(K)}) - (y_{(0)} - y_{-(0)})}{4} = 2(y(K) - y(0))$$

The $\beta_{1-K} > 0$ implies that there is at least one important factor among K factors and SB-X proceeds to Stage 1, where the group factor will be split into two groups of factors: group (1) includes factors 1 to $K/2$ and group (2) includes factors $K/2+1$ to K. If we assume the K is an even number then:

$$\beta_{1-K/2} = \frac{(y_{(K/2)} - y_{-(K/2)}) - (y_{(0)} - y_{-(0)})}{4}$$

$$\beta_{(K/2+1)-K} = \frac{(y_{(K)} - y_{-(K)}) - (y_{(K/2)} - y_{-(K/2)})}{4}$$

If any subgroup yields a zero main effect, it will be dropped from further screening. On the other hand, each subgroup with main effect greater than zero will be split into two subgroups, and this procedure continues until the sizes of all remaining subgroups are one.

The fold-over design does not enable us to estimate individual interactions, but it does enable us to estimate whether interactions are important or not. If the fold-over design and the "original" design (the one without considering the mirror scenarios) give the same conclusions, then interactions are unimportant (Kleijnen et al. (2003a)). If both the original and fold-over designs find different point estimates, but select the same factors as important, it implies that interactions are present in the model, but the main effects of the important factors are larger than the two-factor interaction effects (Kleijnen et al. (2003b)).

2.1.3.4 Metamodel

SB-X provides the unbiased estimates of *main effects* of important factors. Kleijnen et al. (2003a) suggested that in order to estimate the individual *interaction* effects between the factors declared as important, a resolution-V design should be used. Eventually the metamodel can be approximated by the following equation:

$$y = \beta_0 + \sum_{j \in I} \beta_j x_j + \sum_{i,j \in I} \sum \beta_{ij} x_i x_j \qquad I = \text{set of important factors}$$

2.1.4. Further Discussions on SB and SB-X

Issues that are addressed in this section are common among all the screening methods that are based on SB, including Cheng's method (Cheng (1997)) and Controlled Sequential Bifurcation (Wan et al. (2003)) which will be discussed later in this chapter. In Chapters 3 and 4 where we have reported the results of our factor screening experiments on the available simulation models, we will augment the discussion of some of the following issues according to the method being used.

2.1.4.1 Quantifying Importance

In the previous sections, SB might have declared a factor j important if its effect β_j was positive (not zero). In practice, however, adopting this policy would result in declaring all factors as important, because all factors have some non-zero effects. Therefore, before calling a factor important, its effect should be greater than a certain value, like δ. For example, at the first stage, if $y_{(K)} - y_{(0)} > \delta$, we conclude that at least one of the factors 1 through K is important. For quantifying importance, two policies may be adopted by SB users:

1. In the first policy, the analyst determines a value for δ that if a main effect of a factor is greater than that limit, the factor should be considered as important. For this policy, at each stage, SB uses a *constant* δ for finding if a group of factors includes any important factor.

2. In the second policy, analysts do not need to quantify a priori how big a factor effect should be in order to be called important. As simulation outputs become available, SB updates the upper limits for the factor effects; the analysts can stop the factor screening experiment as soon as they find these limits sharp enough. In this policy, in Stage 0, beta limit (β_{Lim}) is equal to $\beta_{0\text{-}K}$, implying that all individual main effects are smaller than $\beta_{Lim}= \beta_{0\text{-}K}$. In Stage 1, β_{Lim} will be updated to $\min\{\beta_{0-K/2}, \beta_{K/2+1-K}\}$. The beta limit of Stage 1 is certainly equal to or less than $\beta_{0\text{-}K}$. For the following stages, β_{Lim} at each stage is the minimum of beta values (main effects) corresponding to all remaining group-factors at that stage. If at a certain stage, the user recognizes that the β_{Lim} is too small, SB will eliminate the group-factor with main effect equal to β_{Lim} and update the β_{Lim} to the minimum of the remaining group-factor effects. If the user finds the β_{Lim} large enough, SB proceeds to the next Stage.

Adopting the first policy leads SB to find all the factors with main effects equal or greater than δ. However, if the user chooses a low value for δ, then SB needs to make a large number of observations in order to detect all the factors with main effects equal or greater than δ. On the other hand, adopting the second policy enables SB to have interaction with user. So, the users can eliminate the group-factors with main effect not large enough to be considered important by the users.

2.1.4.2 Rescaling the Effect Coefficient

In practice, when we consider whether a change in the response is worth pursuing, the cost to achieve the change is often critical. In other words, when we compare the effects of two different factors, the comparison may have little meaning if the cost to change the factors is very different. Therefore, rescaling the effect coefficients with respect to the cost of changing the factors' settings can insure that the results have a useful interpretation. As a result, in the new formulation, the cost of deviating from each factor's nominal level is also incorporated into the metamodel.

Wan et al. (2003) proposed the following approach to rescale the effect coefficients: Let c_i be the cost per unit change of factor i, for $i = 1, 2, \ldots, K$. Further, let $c^* = \max_{i \in D} c_i$, where D is the set of indices of all of the factors whose settings can only be changed in discrete units (e.g., number of machines at a workstation, or pills per prescription). Let

$$\delta_i = \begin{cases} c^* / c_i, & i \notin D \\ \lfloor c^* / c_i \rfloor, & i \in D \end{cases}$$

which is the maximum change in factor i that can be achieved without exceeding a cost c^*; and let $\omega_i = \delta_i c_i / c^*$, which is the fraction of a full-cost move, c^*/c_i, that can actually be made for factor i. If factor i can be changed continuously ($i \notin D$), or $i \in D$ but c^*/c_i is an integer, then $\omega_i = 1$. If $i \in D$ and c^*/c_i is not an integer, then $\omega_i < 1$.

Recall that the main-effects model is

$$Y = \tilde{\beta}_0 + \sum_{i=1}^{K} \tilde{\beta}_i z_i + \varepsilon$$

For screening with a main-effects model, a two-level experimental design is adequate. Let the nominal low setting of z_i be z_i^0 and let the high setting be $z_i^0 + \delta_i$, for $i = 1, 2, \ldots, K$. Define the transformed variables $x_i = \omega_i (z_i - z_i^0) / \delta_i = (c_i / c^*)(z_i - z_i^0)$. Then Y can be expressed as a linear regression on x_i, $i = 1, 2, \ldots, K$, as:

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

where the low setting of x_i is 0, the high setting is ω_i, and $\beta_i = \delta_i \tilde{\beta}_i / \omega_i$, for $i = 1, 2, \ldots, K$. Now each β_i, $i > 0$, has a practical interpretation: it represents the change in the expected response when spending at most c^* to change the setting of factor i.

2.1.4.3 Verifying the Factor Screening Result

In order to verify the important factor shortlist reported by SB or SB-X, Kleijnen et al. (2003a) tested the effects of the remaining *unimportant* factors for the following two scenarios: first, they set all unimportant factors at their low values, while keeping the important factors fixed; and secondly, they switched all unimportant factors to their high values, while keeping the important factors fixed. For both scenarios, they fixed the important factors at their central (base) values. A considerable difference between the outputs

of these two scenarios reveals that there is at least one important factor among the reported set of unimportant factors.

An alternative approach is to build a metamodel in terms of factors identified as important, and then check how well the metamodel fits the real output of the simulation model. This approach has been employed in Chapters 3 and 4.

2.1.4.4 Checking Nonlinearity in the Output

A potential concern in the use of SB or SB-X is the assumption of linearity in the factor effects. Of course, perfect linearity is unnecessary, and SB works quite well even when the linearity assumption holds only approximately. In fact, it is noted that Equation 2.3 is capable of representing some curvature in the response function. This curvature of course, results from the twisting of the plane induced by the interaction term $\beta_{ij}x_ix_j$.

In running SB-X, we usually anticipate fitting the second-order polynomial, given by Equation 2.3. Nevertheless, we should be aware that a second-order polynomial with pure quadratic effects may be more appropriate.

In SB-X, for each design corresponding to $y_{(j)}$, there exists exactly one and only one symmetric design corresponding to $y_{-(j)}$. So the average of $y_{(j)}$ and $y_{-(j)}$ can be an estimate of the output corresponding to the center points (all factors at their central levels). Therefore, the central point analysis can use the outputs without requiring any further observations at the upper and lower levels of the factors. The only required observation is the response value when all the factors are at their central levels (center points). Let $\bar{y}_F$ be the average of responses observed by SB-X in the process of factor screening, and y_C be the response value corresponding to the center points. If SB and SB-X are being applied to a deterministic simulation model, only one observation is needed at the center point. If the difference $\bar{y}_F - y_C$ is small, then the center points lie on or near the plane passing through the factorial points, and the existence of the quadratic curvature is very unlikely. On the other hand, if $\bar{y}_F - y_C$ is large, then quadratic curvature is present.

2.1.4.5 Factor's Effects on Response

Three cases may happen if a factor changes from its low level *L* to its high level *H*.

1. Monotonic response surface: the simulation output increases as the factor increases. This case can be handled well by SB.
2. Non-monotonic response surface: the simulation output does not necessarily increases as the factor increases. This situation may happen in the presence of quadratic effects. In this case, SB may falsely infer that an important factor is unimportant.
3. Interaction Only: When a factor does not have any main effect but its interaction with some other factors are significant, SB is not able to declare that factor as important. A simple example would be a metamodel like: $y = \beta_{1,2} x_1 x_2$.

Conclusions from (2) and (3) are that SB is not always applicable to all simulation models. In case (2), where the output exhibits considerable quadratic effects, in order to get valid results from SB, the experimental area should be restricted such that the interaction terms in a metamodel can properly represent the curvature in the output. SB is never appropriate for case (3).

2.1.4.6 Approximation Errors

In both SB and SB-X, assumptions (SB-1) and (SBX-1) also mean that the approximation errors are zero, which implies a perfect fit. In order to model the approximation errors, Bettonvil et al. (1996) suggested two other alternative approaches:

1. The approximation errors (say ε) can be assumed to be normally and independently distributed (n.i.d) with zero expectation and constant variance σ^2 (*white noise* error).
2. Use *covariance stationary process* estimates, instead of white noise, to model the approximation errors.

If we use either approach (1) or approach (2) to model the error term, after doing the factor screening experiment, we have to ensure that the actual residuals are consistent with the

assumption we have made. This issue will be discussed thoroughly in the applications in Chapters 3 and 4.

2.1.4.7 Efficiency of SB

Generally, the efficiency of a screening method is measured by the number of required runs. Bettonvil et al. (1996) suggested the following strategies for improving the efficiency of SB:

1. When the size of a group is not a power of two, then its first subgroup should have a size equal to the largest possible power of two. For example, if a group has 24 factors, then it should be split into two groups, one with 16 factors, and the other with 8 factors.
2. Clustering important and unimportant factors improves the efficiency of SB. For example, factors can be labeled from 1 to K in *increasing* order of their importance, so that after Stage #1, the important factors are clustered in the second groups.

Bettonvil et al. (1996) gave a formula for finding the maximum number of runs (*maxn*) needed to find the k important factors among $K = 2^m$ factors:

$$maxn = 1 + k[\log_2(2K/k)] \qquad (2.5)$$

For SB, the expected number of runs, *E(n)*, with probability of p for a factor to be important, is derived by Bettonvil et al. (1996) as follows:

$$E(n) = 1 + K - \sum_{j=1}^{m} 2^{m-j}(1-p)^{2^j}$$

2.1.4.8 Issues with Information Requested from Users

In order to apply SB to a simulation model, the client must supply information on the experimental domain (experimental area to be explored), including realistic ranges of the individual factors and limits on the admissible scenarios or combinations of factor levels; for example, some factor values must add up to 100% (Kleijnen et al. (2003a)).

For SB and SB-X, assumptions (SB-1) and (SBX-1) usually does not hold over the entire factor space, and it may be a reasonable assumption for only small variations in a region of interest. Therefore, the situation where one of the observed responses seems to be unrealistically low or high implies that the user has failed to reasonably specify the factor's upper and lower levels. Thus, restricting the experimental region, by changing the upper or lower levels of some factors, might be considered a solution.

Assumptions (SB-2) and (SBX-2) imply that knowing the signs of main effects is necessary for applying SB or SB-X on a simulation system. Bettonvil et al. (1996) used a single-factor-at-a-time design to discover the direction of the factors that the expert was unable to determine. If we assume that β_0 is the response when all the factors are set at their central level, then $y(x_j) - \beta_0$ yields an estimate of β_j, where $y(x_j)$ is the response value when all factors are set at their central levels except factor j which is set at its upper level. The sign of $y(x_j) - \beta_0$ is the same as the main effect sign of factor j. Thus SB requires one more run for each factor with unknown direction and therefore, as the number of factors with unknown direction increases, the efficiency of SB decreases significantly.

Sanchez et al. (2005b) relaxed the assumption of known directions by suggesting the use of efficient fractional factorial experiments as Phase (1) of the factor screening experiment, to estimate the signs and magnitudes of the effects. Their results show that this strategy greatly reduces the possibility of erroneously concluding that important effects are unimportant because of incorrect groupings. In addition, they showed that re-labeling factors from 1 to K in *increasing* order of their importance (main effects estimated in phase (1)), improves the efficiency of SB.

2.2. Sequential Bifurcation under Uncertainty

2.2.1. Introduction

The sequential bifurcation (SB) method developed by Bettonvil et al. (1996) requires the assumptions that the simulation response is deterministic and contains negligible random

error. In this section we extend SB to handle simulations where the response is stochastic and subject to significant error. The main idea of the proposed procedure in this section is derived from Cheng (1997). Their method, however, assumes that the response function can be represented by a first-order linear function. We make the required changes in their method for handling the situations where a first-order polynomial function augmented with second-order interaction terms is a better approximation for the response function.

One of the differences between deterministic and stochastic output is that in the latter case a disproportionate amount of effort can be expended on investigating factors which are borderline in importance. The method proposed by Cheng (1997) allows for an "indifference-zone." If the importance of a factor is estimated to fall inside this indifference zone, then no further effort is made to estimate this factor effect more accurately. However, it will be shown that the proposed "indifference-zone" does not have a significant effect in improving the efficiency of the method. As an alternative, we have based our statistical comparisons on hypothesis testing, which is shown to be more efficient.

2.2.2. Cheng's SB under Uncertainty

Cheng (1997) extended the SB method, which was originally developed for deterministic simulation models, to handle stochastic simulation models where the response is subject to significant error. They assumed a first-order polynomial function for the response function with known signs of main effects.

2.2.2.1 Model Structure and Assumptions

If y denotes the output of interest from a run, then there are two quantities which affect the value of y:

1. a vector of decision variables: $\mathbf{x} = (x_1, x_2, ..., x_K)$ which are under the control of the experimenter
2. a set of random numbers: $\mathbf{u} = (u_1, u_2, ..., u_n)$ which form the stochastic input driving the simulation run

Thus, y can be regarded as a function of only **x** and **u**:

$$y = y(\mathbf{x},\mathbf{u}) \qquad (2.11)$$

Cheng (1997) proposes adopting the following regression metamodel:

$$y(\mathbf{x},\mathbf{y}) = \eta(\mathbf{x}) + e(\mathbf{x},\mathbf{u}) \qquad (2.12)$$

where e(**x**,**u**) is regarded as an "error" term containing all the variation of y caused by stochastic inputs. Moreover we assume that:

$$\mathrm{E}[e(\mathbf{x},\mathbf{u})] = 0\,,$$

$$\mathrm{Var}[e(\mathbf{x},\mathbf{u})] = \sigma^2(\mathbf{x})\,.$$

The precise distribution of e(**x**,**u**) is not known but can be expected to be approximately normally distributed. The objective is to investigate the local behavior of η about some notional operating point $\mathbf{x}^0 = (x_1^0, x_2^0, ..., x_K^0)$. If attention is confined to a small change of **x** about $\mathbf{x}^0$, one of the following functions can approximately represent η around $\mathbf{x}^0$:

1. First-order polynomial function:

$$\eta(\mathbf{x}) = \eta_0 + \sum_{i=1}^{K} \beta_i (x_i - x_i^0)$$

2. Second-order polynomial function without quadratic effects

$$\eta(\mathbf{x}) = \eta_0 + \sum_{i=1}^{K} \beta_i (x_i - x_j^0) + \sum_{i>j} \beta_{ij} (x_i - x_j^0)(x_j - x_j^0)$$

Bettonvil et al. (1996) assume that the simulation error e(**x**,**u**) is small and can be neglected. Cheng (1997) considered the situation where e(**x**,**u**) cannot be neglected but locally the variability of e(**x**,**u**) does not depend on x, that is, he assumes:

$$\sigma^2(\mathbf{x}) = \sigma^2\,, \qquad \text{is independent of x.}$$

Moreover, although the values of the β_i are unknown, in most situations it is quite realistic to assume that their signs are known. In addition, by simply reversing the sign of x_i where necessary, it may be assumed that:

$$\beta_i \geq 0\,, \qquad \text{for all i.}$$

In order to satisfy this assumption, L_j can be defined as the level of factor j that generated a low value for the output y, and H_j as the level that generate a higher value, provided that

factor j has any effect at all. Consequently, if an increase in factor j reduces the output, then the upper and lower levels of the factor should be switched such that $H_j < L_j$, and thus the definition of H_j and L_j implies that all main effects in equation are non-negative: $\beta_j \geq 0$. For more explanation about how to relax this assumption please refer to Section 2.1.4.8.

2.2.2.2 Design of SB under Uncertainty

Cheng (1997) assumes that the response function (y) can be modeled by a first-order polynomial around point $\mathbf{x}^{(0)}$. Hence the response function can be estimated by:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_K x_K + e(\mathbf{x}^{(0)}, \mathbf{u}) \qquad (2.13)$$

where:

y: response of the metamodel (the sequential bifurcation approximation)

K: total number of factors in the experiment

β_j : first-order or main effect of factor j with $j = 1, \ldots, K$

x_j: value of factor j, standardized to lie in [-1,+1]

$e(\mathbf{x}^{(0)},\mathbf{u})$: the approximate error, assumed to be approximately normally distributed with mean zero and constant variance σ^2.

The symbol $y^{(j)}$ denotes the response value when the factors 1, …, j are set at their upper levels (H_j) and the remaining factors ($j + 1, \ldots, K$) are set at their lower levels (L_j). Therefore, in the standardized Metamodel 2.13, the values of the first j factors are +1 and the values of the remaining factors are -1. So the polynomial in Equation 2.13 yields:

$$y^{(j)} = \beta_0 + \beta_1 + ... + \beta_{j-1} + \beta_j - \beta_{j+1} - ... - \beta_K + e(\mathbf{x}, \mathbf{u}), \qquad j = 0, 1, 2, \ldots, K$$

If $j < k$ then the scaled difference

$$D(j,k) = [y^{(k)} - y^{(j-1)}]/2 \qquad (2.14)$$

has the expectation

$$E[D(j,k)] = \sum_{i=j}^{k} \beta_i \qquad (2.15)$$

so that $D(j, k)$ can be regarded as an estimator of the sum of the main effects for factors j to k. If the $y^{(j)}$ and $y^{(k)}$ are independent, then

$$\mathrm{Var}[D(j,k)] = \frac{1}{2}\sigma^2 . \qquad (2.16)$$

The symbol β_{j-k} denotes the *expected* sum of individual effects β_j through β_k with $k > j$. Therefore:

$$\beta_{j-k} = \mathrm{E}[D(j,k)] .$$

Improving the accuracy of $y^{(j)}$ can simply be done by replicating this design point. In general, if there are $r^{(j-1)}$ observations at level j-1: $y_i^{(j-1)}$, $i = 1, 2, \ldots, r^{(j-1)}$, and $r^{(k)}$ observations at level k: $y_i^{(k)}$, $i = 1, 2, \ldots, r^{(k)}$, then:

$$\overline{D}(j,k) = [\frac{1}{r^{(k)}}\sum_{i=1}^{r^{(k)}} y_i^{(k)} - \frac{1}{r^{(j-1)}}\sum_{i=1}^{r^{(j-1)}} y_i^{(j-1)}]/2 \qquad (2.17)$$

and

$$\mathrm{Var}[\overline{D}(j,k)] = \frac{1}{4}\sigma^2 (\,1\!\!\diagup_{r^{(j-1)}} + 1\!\!\diagup_{r^{(k)}}) . \qquad (2.18)$$

2.2.2.3 Cheng's SB Procedure under Uncertainty

The primary objective of factor screening is to divide the factors into just two groups: important factors and unimportant factors. For calling a factor important, its effect should be greater than a certain value, like δ. That is, if $\overline{D}(j,k) > \delta$, we conclude that at least one of the factors j through k is important. Thus we can divide the main effects of factors into two groups as follows:

$$I = \{\beta_j : \beta_j > \delta\} \quad \text{and} \quad U = \{\beta_j : \beta_j \le \delta\}$$

As mentioned before, much simulation effort can be spent on borderline cases, where β_j is close to δ. Cheng (1997) introduced an indifference-zone $(0, \delta+a)$ for coping with this problem. If β_j is estimated as being located within this zone, it is automatically classified as being in U. If a is set small relative to δ, then a misclassification where β_j is wrongly put in U, when actually $\beta_j > \delta$, can be assumed to be of little practical concern, as we know that $\beta_j < \delta + a$.

Similar to deterministic SB, the key idea here is that if we find that for $k > j$:

$$\sum_{i=j}^{k} \beta_i < \delta$$

then, since we assumed that $\beta_i > 0$ for all i, we must have $\beta_i < \delta$ for $i = j$ to k. Thus we can classify all the β_i in the expression above as unimportant.

The following procedure is taken from Cheng (1997) with some minor changes in notation and equations so as to make it consistent with the procedures presented in the previous sections for deterministic SB and SB-X.

Step (0): To initiate the process we make $r^{(0)} > 1$ and $r^{(K)} > 1$ runs at levels 0 and K respectively (at this step, typically $r^{(0)} = r^{(K)}$ = a small number, between 2 and 5). Thus at the start of Step (1) we have two sets of observations:

Observations at level 0: $\{y_j^{(0)} : j = 1,..., r^{(0)}\}$

Observations at level K: $\{y_j^{(K)} : j = 1,..., r^{(K)}\}$

We also place all the coefficients in the single, unclassified set $G_1 = \{\beta_1, \beta_2, \ldots, \beta_K\}$. Thus initially the number of sets is $p_1 = 1$. We also assign 1 to s and proceed to the next step.

Step (s) ($s > 0$): At the beginning of Step (s), factors are partitioned into p_s sets:

$$G_{i,s} = \{\beta_j : k_{(i-1),s} < j \le k_{i,s}\} \qquad ,i = 1,2,..., p_s \qquad (2.19)$$

For example, at Step (1) ($s = 1$ and $p_1 = 1$), there is only one set of factors and Equation 2.19 results in $G_{i,1} = \{\beta_j : k_{(i-1),1} < j \le k_{i,1}\}, i = 1$, where $k_{0,1} = 1$ and $k_{1,1} = K$. Thus $G_1 = G_{1,1} = \{\beta_1, \beta_2, \ldots, \beta_K\}$.

At the beginning of Step (s), some of the sets are already classified, some are unclassified. If all sets have been classified, then the algorithm ends. Otherwise we select any unclassified set: call this $G_{i,s}$.

1. If $G_{i,s}$ is not a singleton set (set of size one) we see if all the coefficients can be classified as unimportant. This can be done by considering the expected main effect of set $G_{i,s}$, which is $\overline{D}(k_{(i-1),s}, k_{i,s})$. Based on Equation 2.19, set $G_{i,s}$ includes

factors $k_{(i\text{-}1),s}$ to $k_{i,s}$, therefore, for estimating $\overline{D}(k_{(i-1),s}, k_{i,s})$, two sets of observations are needed:

Observations at level $k_{i,s}$: $\{y_j^{(k_{i,s})} : j = 1,..., r^{(k_{i,s})}\}$ (2.20-a)

Observations at level $k_{(i\text{-}1),s}$-1: $\{y_j^{(k_{(i-1),s}-1)} : j = 1,..., r^{(k_{(i-1),s}-1)}\}$. (2.20-b)

From the assumption that $e(\mathbf{x},\mathbf{u})$ is normally distributed, $\overline{D}(k_{(i-1),s}, k_{i,s})$ is also normally distributed with mean and variance:

$$\mu(G_{i,s}) = \sum_{i=k_{(i-1),s}}^{k_{i,s}} \beta_i , \qquad \upsilon(G_{i,s}) = \frac{1}{4}\sigma^2 \left(1 \Big/ r^{(k_{(i-1),s}-1)} + 1 \Big/ r^{(k_{i,s})} \right).$$

For each new observation sets (2.20-a) and (2.20-b), σ^2 can be estimated by:

S^2 for observation set (2.20-a): $S^2_{k_{i,s}} = \dfrac{\sum_{j=1}^{r^{(k_{i,s})}} (y_j^{(k_{i,s})} - \bar{y}^{(k_{i,s})})^2}{r^{(k_{i,s})} - 1}$

S^2 for observation set (2.20-b): $S^2_{k_{(i-1),s}-1} = \dfrac{\sum_{j=1}^{r^{(k_{(i-1),s}-1)}} (y_j^{(k_{(i-1),s}-1)} - \bar{y}^{(k_{(i-1),s}-1)})^2}{r^{(k_{(i-1),s}-1)} - 1}$

and pooling these estimates gives an overall estimator of σ^2 for set $G_{i,s}$ as:

$$S^2_{G_{i,s}} = \frac{(r^{(k_{(i-1),s}-1)} - 1)S^2_{k_{(i-1),s}-1} + (r^{(k_{i,s})} - 1)S^2_{k_{i,s}}}{r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2} \quad (2.21)$$

and pooling the estimates of σ^2 over all stages gives an overall estimator of σ^2 at the start of Step (s) as:

$$S_s^2 = \frac{\sum_{i=1}^{s} (r^{(k_{(i-1),s}-1)} + r^{(k_{is})} - 2) S^2_{G_{is}}}{\sum_{i=1}^{s} (r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2)}, \qquad (2.22)$$

then with probability approximately (1-α):

$$\mu(G_{i,s}) < \overline{D}(k_{(i-1),s}, k_{i,s}) + \frac{1}{2} z_\alpha S_s \sqrt{\left(1 \Big/ r^{(k_{(i-1),s}-1)} + 1 \Big/ r^{(k_{i,s})} \right)}$$

If therefore

$$\overline{D}(k_{(i-1),s}, k_{i,s}) < \delta - \frac{1}{2} z_\alpha S_s \sqrt{(1/r^{(k_{(i-1),s}-1)} + 1/r^{(k_{i,s})})} \quad (2.23)$$

then with confidence (1- α):

$$\mu(G_{i,s}) < \delta \quad \text{i.e.} \quad \sum_{i=k_{(i-1),s}}^{k_{i,s}} \beta_i < \delta .$$

And since β_j are all positive this means:

$$\beta_j < \delta, \qquad j = \{k_{(i-1),s}, k_{(i-1),s} + 1, ..., k_{i,s}\}$$

and all the factors in set of $G_{i,s}$ can be classified as unimportant.

If Expression 2.23 is not satisfied, then $G_{i,s}$ will be split into two groups at $k = \left\lceil (k_{(i-1),s} + k_{i,s})/2 \right\rceil$. The two new sets that will replace $G_{i,s}$ are:

$$\{\beta_{k_{(i-1),s}}, ..., \beta_k\} \quad and \quad \{\beta_{k+1}, ..., \beta_{k_{i,s}}\}$$

2. If $G_{i,s} = \{\beta_k\}$, i.e. it is a singleton set, then we proceed to fully classify β_k. This can be done by considering the expected main effect of set β_k, which is $\overline{D}(k,k)$. for estimating $\overline{D}(k,k)$, two sets of observations are needed:

Observations at level k: $\{y_j^{(k_s)} : j = 1, ..., r^{(k_s)}\}$ (2.24-a)

Observations at level k-1: $\{y_j^{(k_s-1)} : j = 1, ..., r^{(k_s-1)}\}$. (2.24-b)

From the assumption that $e(\mathbf{x},\mathbf{u})$ is normally distributed, $\overline{D}(k,k)$ is also normally distributed with mean and variance:

$$\mu(\beta_k) = \beta_k, \qquad \upsilon(\beta_k) = \frac{1}{4}\sigma^2 (1/r^{(k_s-1)} + 1/r^{(k_s)}).$$

For each new observation set (2.24-a) and (2.24-b), σ^2 can be estimated by:

S^2 for observation set (2.24-a): $S_{k_s}^2 = \dfrac{\sum_{j=1}^{r^{(k_s)}} (y_j^{(k_s)} - \bar{y}^{(k_s)})^2}{r^{(k_s)} - 1}$

S^2 for observation set (2.24-b): $S^2_{k_s-1} = \dfrac{\sum_{j=1}^{r^{(k_s-1)}} (y_j^{(k_s-1)} - \bar{y}^{(k_s-1)})^2}{r^{(k_s-1)} - 1}$

and pooling these estimates gives an overall estimator of σ^2 for factor k at stage s as:

$$S^2_{k_s} = \frac{(r^{(k_s-1)} - 1)S^2_{k_s-1} + (r^{(k_s)} - 1)S^2_{k_s}}{r^{(k_s-1)} + r^{(k_s)} - 2} \qquad (2.25)$$

and pooling the estimates of σ^2 over all stages gives an overall estimator of σ^2 at the start of Step (s) as:

$$S^2_s = \frac{\sum_{i=1}^{s} (r^{(k_s-1)} + r^{(k_s)} - 2)S^2_{k_s}}{\sum_{i=1}^{s} (r^{(k_s-1)} + r^{(k_s)} - 2)} \qquad (2.26)$$

then a two-sided (1- α) confidence interval for β_k with upper and lower limits given by:

$$\beta_k^{\pm} = \overline{D}(k,k) \pm \frac{1}{2} z_{\alpha/2} S_s \sqrt{(1/r^{(k_s-1)} + 1/r^{(k_s)})} \ . \qquad (2.27)$$

If δ is contained in this interval we make additional runs at the levels k-1 and k. If $r^{(k_s-1)}$ and $r^{(k_s)}$ are not initially equal, add the runs at the level with the smaller number of runs, until $r^{(k_s-1)} = r^{(k_s)}$. Then add runs at both levels k-1 and k, keeping $r^{(k_s-1)} = r^{(k_s)}$. As runs are added, the length of the confidence interval Equation 2.27 decreases. We stop when either:

i. $\delta < \beta_k^-$, then β_k is classified as important,

ii. $\beta_k^+ < \delta$, then β_k is classified as unimportant,

iii. $\beta_k^+ < \delta + a$ (where a is small relative to δ), then β_k is regarded to be sufficiently close to δ to be classified as unimportant.

Once β_k is classified, we increment s and go to Step (s + 1).

It will be seen that at the end of the process every β_j will have been classified as either important or unimportant. Moreover all β_j classified as important will have a corresponding confidence interval calculated.

2.2.3. Modified SB under Uncertainty

This section first addresses the shortcomings and the problems inherent in the Cheng's method and then presents an improved algorithm for SB under Uncertainty.

As mentioned in the previous section, Cheng (1997) introduced an indifference zone $(0, \delta+a)$ where if β_j is estimated as being located within this zone, it is automatically classified as being unimportant. However, the method uses the indifference zone only when it is classifying a *single* factor, and does not use the indifference zone when comparing the *group-factor*'s main effect with δ. But in the process of applying SB, most comparisons are made by deciding about the significance of a *group-factor* not a *single* factor. For instance, we consider the example presented in Bettonvil et al. (1996), where 3 important factors were detected among 128 factors. In this example, in the process of factor screening, 29 main effects (either group-factor's or single factor's main effect) were compared with δ. Only 6 of these comparisons were used for deciding about the importance of a *single* factor, whereas the other 23 were made by comparing a *group-factor*'s main effect with δ, for which the "indifference zone" is not applied. And more importantly, when there is only one factor in a group, no matter if it is important or not, two runs (with certain numbers of replications) must be made in order to classify the factor. Thus, comparing the main effect of this factor with $\delta + \alpha$, only reduces the accuracy of the method and does not improve its efficiency, because we have already made 2 runs for this comparison!

Even if we apply the indifference-zone concept for all the required comparisons, this method classifies a main effect, say k, as unimportant when either $\beta_k^+ < \delta$ or $\beta_k^+ < \delta + a$. Thus, if Cheng's method is performed once with parameter (δ, a) and once with parameter $(\delta+a, 0)$, the results do not differ. In addition, even determining an appropriate value for threshold δ is challenging for many users, and they usually want to see the results of factor screening for different value of δ. Introduction of another parameter into the method will be even more

confusing for the users. They may want to see the factor screening result for different combinations of δ and a values, which is often impractical, or completely destroys the efficiency of SB.

Another important issue with Cheng's method is the way it eliminates unimportant factors. As mentioned before, the method uses Equation 2.23 for classifying a factor as unimportant. As we will show in the following discussion, Equation 2.23 is lenient in classifying unimportant factors, i.e. even if a main effect of a factor (or a group-factor) is less than δ, it is possible that the method does not eliminate it. Although this strategy does not affect the correctness of the method, it is counter to the fundamental assumption of most factor screening techniques: only a *few* factors (k) are important among *many* potentially important factors (K) ($k << K$). As a result, assuming that each factor is unimportant unless otherwise is proven, improves the efficiency of the factor screening procedure, without sacrificing the accuracy of the method.

Finally, although the variance of the error term $e(\mathbf{x}, \mathbf{u})$ is unknown, Cheng's method uses the normal distribution percentage point for making statistical comparisons. While the t-distribution approaches a normal distribution as its degrees of freedom increases, the SB method does not make a sufficient number of runs to make this approximation legitimate. For example, Bettonvil et al. (1996) report the number of required runs when performing SB on models with 1024 involved factors and 0 to 8 important factors. The results summarized in Table 2.1, show that even in the worst case SB requires only 65 runs, as a result the variance of the error term can be estimated with 64 degrees of freedom. But the t-distribution does not follow normal distribution perfectly with this value for degrees of freedom. Therefore, the t-distribution percentage points should be used in making comparisons.

Table 2.1: Total number of required runs when applying SB on $K = 1024$ factors

	Number of important factor k								
	0	1	2	3	4	5	6	7	8
Number of required runs	2	12	21	29	37	44	51	58	65

2.2.3.1 Modified Algorithm for SB under Uncertainty

The modified algorithm is based on all the assumptions made in Section 2.2.2.1.

Step (0): to initiate the process we make $r^{(0)} > 1$ and $r^{(K)} > 1$ runs at levels 0 and K respectively (at this step, typically $r^{(0)} = r^{(K)}$ = a small number, between 2 and 5). Thus at the start of Step (1) we have two sets of observations:

Observations at level 0: $\{y_j^{(0)} : j = 1,..., r^{(0)}\}$

Observations at level K: $\{y_j^{(K)} : j = 1,..., r^{(K)}\}$.

We also place all the coefficients in the single, unclassified set $G_1 = \{\beta_1, \beta_2, \ldots, \beta_K\}$. Thus initially the number of sets is $p_1 = 1$. We also assign 1 to s and proceed to the next step

Step (s) ($s > 0$): At the beginning of Step (s), factors are partitioned into p_s sets:

$$G_{i,s} = \{\beta_j : k_{(i-1),s} < j \le k_{i,s}\}, \qquad i = 1,2,..., p_s \qquad (2.28)$$

For example, at Step (1) ($s = 1$ and $p_1 = 1$), there is only one set of factors and Equation 2.19 results in $G_{i1} = \{\beta_j : k_{(i-1),1} < j \le k_{i,1}\}, i = 1$, where $k_{0,1} = 1$ and $k_{1,1} = K$. Thus $G_1 = G_{1,1} = \{\beta_1, \beta_2, \ldots, \beta_K\}$.

At the beginning of Step (s), some of the sets are already classified, some are unclassified. If all sets have been classified, then the algorithm ends. Otherwise we select any unclassified set: call this $G_{i,s}$.

1. If $G_{i,s}$ is not a singleton set (set of size one) we see if all the coefficients can be classified as unimportant. This can be done by considering the expected main effect of set $G_{i,s}$, which is $\overline{D}(k_{(i-1),s}, k_{i,s})$. Based on Equation 2.19, set $G_{i,s}$ includes factors $k_{(i\text{-}1),s}$ to $k_{i,s}$, therefore, for estimating $\overline{D}(k_{(i-1),s}, k_{i,s})$, two sets of observations are needed:

   Observations at level k_{is}: $\{y_j^{(k_{i,s})} : j = 1,..., r^{(k_{i,s})}\}$ (2.29-a)

   Observations at level $k_{(i\text{-}1),s}$-1: $\{y_j^{(k_{(i-1),s}-1)} : j = 1,..., r^{(k_{(i-1),s}-1)}\}$ (2.29-b)

From the assumption that $e(\mathbf{x},\mathbf{u})$ is normally distributed, $\overline{D}(k_{(i-1),s},k_{i,s})$ is also normally distributed with mean and variance:

$$\mu(G_{i,s}) = \sum_{i=k_{(i-1),s}}^{k_{i,s}} \beta_i \,, \qquad \upsilon(G_{i,s}) = \frac{1}{4}\sigma^2 \left(1 \Big/ r^{(k_{(i-1),s}-1)} + 1 \Big/ r^{(k_{i,s})} \right).$$

For each new observation sets (2.29-a) and (2.29-b), σ^2 can be estimated by:

S^2 for observation set (2.29-a): $S^2_{k_{i,s}} = \dfrac{\sum_{j=1}^{r^{(k_{i,s})}} (y_j^{(k_{i,s})} - \bar{y}^{(k_{i,s})})^2}{r^{(k_{i,s})} - 1}$

S^2 for observation set (2.29-b): $S^2_{k_{(i-1),s}-1} = \dfrac{\sum_{j=1}^{r^{(k_{(i-1),s}-1)}} (y_j^{(k_{(i-1),s}-1)} - \bar{y}^{(k_{(i-1),s}-1)})^2}{r^{(k_{(i-1),s}-1)} - 1}$

and pooling these estimates gives an overall estimator of σ^2 for set $G_{i,s}$ as:

$$S^2_{G_{i,s}} = \frac{(r^{(k_{(i-1),s}-1)} - 1)S^2_{k_{(i-1),s}-1} + (r^{(k_{i,s})} - 1)S^2_{k_{i,s}}}{r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2} \qquad (2.30)$$

and pooling the estimates of σ^2 over all stages gives an overall estimator of σ^2 at the start of Step (s) as:

$$S^2_s = \frac{\sum_{i=1}^{s} (r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2)S^2_{G_{i,s}}}{\sum_{i=1}^{s} (r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2)} \,. \qquad (2.31)$$

Now, in order to see if the main effect of set $G_{i,s}$ is significant, we test the following one-tail hypothesis:

$$\begin{aligned} &\mathrm{H}_0 : \ \mu(G_{i,s}) \le \delta \\ &\mathrm{H}_1 : \ \mu(G_{i,s}) > \delta \end{aligned} \qquad (2.32)$$

If H_0 is rejected, the factor will be classified as important, whereas if H_0 is not rejected, the implication is that the factor does not have significant effect. Because rejecting H_0 is a strong conclusion, this formulation forces the factor to demonstrate that its main effect exceeds δ. In the other word, this formulation assumes that the effect is unimportant unless there is strong evidence to the contrary.

If the null hypothesis is correct, $\mu(G_{i,s}) = \delta$, then the quantity

$$t_0 = \frac{\overline{D}(j,k) - \delta}{\sqrt{\mathrm{Var}[\overline{D}(j,k)]}} \Big/ \sqrt{\frac{SS_s}{\sigma^2}\Big/\omega} = \frac{\overline{D}(j,k) - \delta}{\sqrt{\frac{1}{4}\sigma^2\left(1/r^{(j-1)} + 1/r^{(k)}\right)}} \Big/ \sqrt{\frac{S_s^2}{\sigma^2}} = \frac{\overline{D}(j,k) - \delta}{\frac{1}{2}S_s\sqrt{1/r^{(j-1)} + 1/r^{(k)}}} \qquad (2.33)$$

has a t-distribution with ω degrees of freedom, where $\omega = \sum_{i=1}^{s}(r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2)$

and $SS_s = \omega S_s^2$.

If $t_{\alpha,\omega}$ denotes the upper α percentage point of the t-distribution with ω degrees of freedom, then if $t_0 > t_{\alpha,\omega}$, then we reject the null hypothesis H_0 and conclude that the corresponding factor or group-factor is important. The condition $t_0 > t_{\alpha,\omega}$ results in

$$\overline{D}(k_{(i-1),s}, k_{i,s}) > \delta + \frac{1}{2}t_{\alpha,\omega}S_s\sqrt{\left(1/r^{(k_{(i-1),s}-1)} + 1/r^{(k_{i,s})}\right)} \qquad (2.34)$$

Therefore, if the inequality in Equation 2.34 holds, then with confidence (1- α):

$$\mu(G_{is}) > \delta \quad \text{i.e.} \quad \sum_{i=k_{(i-1),s}}^{k_{i,s}}\beta_i > \delta$$

and thus, with confidence (1- α) we can consider set $G_{i,s}$ as important. Comparing inequalities in both Equations 2.23 and 2.34 reveals how lenient the Cheng's formulation is on eliminating factors. In the Cheng's formulation, a group-factor may remain in the experiment even if its true main effect is slightly less than δ, whereas in the new formulation a group-factor can remain in the experiment only if it can demonstrate its significant effect.

If Equation 2.34 is satisfied, then $G_{i,s}$ will be split into two groups at $k = \left\lceil (k_{(i-1),s} + k_{i,s})/2 \right\rceil$. The two new sets will replace $G_{i,s}$:

$$\{\beta_{k_{(i-1),s}}, \ldots, \beta_k\} \quad \textit{and} \quad \{\beta_{k+1}, \ldots, \beta_{k_{i,s}}\}$$

and we increment s and go to Step (s + 1).

2. If $G_{i,s} = \{\beta_k\}$, i.e. it is a singleton set, then we proceed to fully classify β_k. This can be done by considering the expected main effect of set β_k, which is $\overline{D}(k,k)$. For estimating $\overline{D}(k,k)$, two sets of observations are needed:

Observations at level k: $\{y_j^{(k_s)} : j = 1,...,r^{(k_s)}\}$ (2.35-a)

Observations at level k-1: $\{y_j^{(k_s-1)} : j = 1,...,r^{(k_s-1)}\}$. (2.35-b)

From the assumption that $e(\mathbf{x},\mathbf{u})$ is normally distributed, $\overline{D}(k,k)$ is also normally distributed with mean and variance:

$$\mu(\beta_k) = \beta_k, \qquad \upsilon(\beta_k) = \frac{1}{4}\sigma^2(1/r^{(k_s-1)} + 1/r^{(k_s)}).$$

For each new observation set (2.35-a) and (2.35-b), σ^2 can be estimated by:

S^2 for observation set (2.35-a): $S_{k_s}^2 = \dfrac{\sum_{j=1}^{r^{(k_s)}}(y_j^{(k_s)} - \bar{y}^{(k_s)})^2}{r^{(k_s)} - 1}$

S^2 for observation set (2.35-b): $S_{k_s-1}^2 = \dfrac{\sum_{j=1}^{r^{(k_s-1)}}(y_j^{(k_s-1)} - \bar{y}^{(k_s-1)})^2}{r^{(k_s-1)} - 1}$

and pooling these estimates gives an overall estimator of σ^2 for factor k at stage s as:

$$S_{k_s}^2 = \frac{(r^{(k_s-1)} - 1)S_{k_s-1}^2 + (r^{(k_s)} - 1)S_{k_s}^2}{r^{(k_s-1)} + r^{(k_s)} - 2} \qquad (2.36)$$

and pooling the estimates of σ^2 over all stages gives an overall estimator of σ^2 at the start of Step (s) as:

$$S_s^2 = \frac{\sum_{i=1}^{s}(r^{(k_s-1)} + r^{(k_s)} - 2)S_{k_s}^2}{\sum_{i=1}^{s}(r^{(k_s-1)} + r^{(k_s)} - 2)}. \qquad (2.37)$$

If we set $\omega = \sum_{i=1}^{s}(r^{(k_s-1)} + r^{(k_s)} - 2)$, then a two-sided (1- α) confidence interval for β_k with upper and lower limits given by:

$$\beta_k^{\pm} = \overline{D}(k,k) \pm \frac{1}{2} t_{\alpha/2,\omega} S_s \sqrt{(1/r^{(k_s-1)} + 1/r^{(k_s)})} \qquad (2.38)$$

In order to make decision about the importance of factor *k*, two different strategies are used, depending on the expense of making an additional run.

If simulation runs are expensive, we classify the factor without making any additional runs, so

i. if $\delta < \beta_k^-$, then β_k is classified as important,

ii. otherwise, β_k is classified as unimportant.

On the other hand, if simulation runs are not too expensive, we can obtain a better estimate of factor *k*'s main effect by making additional runs. If δ is contained in this interval additional runs at the levels *k*-1 and *k* are made. If $r^{(k_s-1)}$ and $r^{(k_s)}$ are not initially equal add the runs at the level with the smaller number of runs, until $r^{(k_s-1)} = r^{(k_s)}$, and then add runs at both levels *k*-1 and *k*, keeping $r^{(k_s-1)} = r^{(k_s)}$. As runs are added, the length of the confidence interval calculated as Equation 2.38 decreases. We stop when either:

i. $\delta < \beta_k^-$, then β_k is classified as important,

ii. $\beta_k^+ < \delta$, then β_k is classified as unimportant.

Once β_k is classified, we increment *s* and go to Step (*s* + 1).

2.2.4. Folded-over Modified Cheng's Method (MCh-X)

In this section, we modify the SB algorithm under uncertainty for cases where the second-interaction effects are not negligible. The basic idea of SB-X is exactly the same as SB under uncertainty: If we find that for $k > j$,

$$\sum_{i=j}^{k} \beta_i < \delta$$

then, since we assumed that $\beta_i > 0$ for all *i*, we must have $\beta_i < \delta$ for $i = j$ to *k*. Thus we can classify all the β_i in the expression above as unimportant.

SB under uncertainty (SB-X) assumes that the response function (y) can be modeled by a second-order polynomial function around $\mathbf{x}^{(0)}$ as follows:

$$y = \beta_0 + \sum_{j=1}^{K} \beta_j x_j + \sum_{j=1}^{K-1} \sum_{j'=j+1}^{K} \beta_{j,j'} x_j x_{j'} + e(\mathbf{x}^{(0)}, \mathbf{u}) \quad (2.41)$$

where:

y: response of the metamodel (the sequential bifurcation approximation)

K: total number of factors in the experiment

β_j : first-order or main effect of factor j with $j = 1, \ldots, K$

$\beta_{j';j}$: interaction effect of the factors j' and j with $1 \leq j' < j \leq K$

x_j: value of factor j, standardized to lie in [-1,+1]

$e(\mathbf{x}^{(0)},\mathbf{u})$: the approximate error, assumed to be approximately normally distributed with mean zero and constant variance σ^2.

2.2.4.1 Design of Folded-over Modified Cheng's Method

The symbol $y^{(j)}$ denotes the response value when the factors $1,\ldots,j$ are set at their upper levels (H_j) and the remaining factors ($j + 1, \ldots, K$) are set at their lower levels (L_j). Therefore, in the standardized metamodel from Equation 2.41, the values of first j factors are +1 and the values of the remaining factors are -1. So the polynomial in Equation 2.41 yields:

$$y^{(j)} = \beta_0 + \beta_1 + \ldots + \beta_{j-1} + \beta_j - \beta_{j+1} - \ldots - \beta_K + \sum_{i<i'\leq j} \sum \beta_{i,i'} + \sum_{j<i<i'} \sum \beta_{i,i'} - \sum_{i\leq j<i'} \sum \beta_{i,i'} + e(\mathbf{x},\mathbf{u}) \qquad ,j = 0, 1, 2, \ldots, K$$

The symbol $y^{-(j)}$, called the mirror observation of $y^{(j)}$, denotes the response value when the factors $1, \ldots, j$ are set at their lower levels (L_j) and the remaining factors ($j + 1, \ldots, K$) are set at their upper levels (H_j). Therefore, in the standardized metamodel (Equation 2.41), the values of first j factors are -1 and the values of the remaining factors are +1. So the polynomial in Equation 2.41 yields:

$$y^{-(j)} = \beta_0 - \beta_1 - \ldots - \beta_{j-1} - \beta_j + \beta_{j+1} + \ldots + \beta_K + \sum_{i<i'\leq j} \sum \beta_{i,i'} + \sum_{j<i<i'} \sum \beta_{i,i'} - \sum_{i\leq j<i'} \sum \beta_{i,i'} + e(\mathbf{x},\mathbf{u}) \qquad ,j = 0, 1, 2, \ldots, K$$

If $j < k$ then the scaled difference

$$D(j,k) = \frac{(y^{(k)} - y^{-(k)}) - (y^{(j-1)} - y^{-(j-1)})}{4} \qquad (2.42)$$

has the expectation

$$\mathrm{E}[D(j,k)] = \sum_{i=j}^{k} \beta_i \qquad (2.43)$$

so that $D(j, k)$ can be regarded as an estimator of the sum of the main effects for factors j to k. If the $y^{(j)}$ and $y^{(k)}$ are independent, then

$$\mathrm{Var}[D(j,k)] = \frac{1}{4}\sigma^2 . \qquad (2.44)$$

The symbol $\beta_{j\text{-}k}$ denotes the *expected* sum of individual effects β_j through β_k with $k > j$. Therefore:

$$\beta_{j-k} = E[D(j,k)] .$$

Improving the accuracy of $y^{(j)}$ can be done by replicating. In general, if we have $r^{(j\text{-}1)}$ observations at level j-1: $y_i^{(j-1)}$ and its mirror: $y_i^{-(j-1)}$, $i = 1, 2, \ldots, r^{(j\text{-}1)}$, and $r^{(k)}$ observations at level k: $y_i^{(k)}$, and its mirror: $y_i^{-(k)}$, $i = 1, 2, \ldots, r^{(k)}$, then:

$$\overline{D}(j,k) = [\frac{1}{r^{(k)}} \sum_{i=1}^{r^{(k)}} (y_i^{(k)} - y_i^{-(k)}) - \frac{1}{r^{(j-1)}} \sum_{i=1}^{r^{(j-1)}} (y_i^{(j-1)} - y_i^{-(j-1)})] / 4 \qquad (2.45)$$

and

$$\mathrm{Var}[\overline{D}(j,k)] = \frac{1}{8}\sigma^2 (1/r^{(j-1)} + 1/r^{(k)}) . \qquad (2.46)$$

2.2.4.2 Algorithm for the Folded-over Modified Cheng's Method (MCh-X)

Step (0): to initiate the process we make $r^{(0)} > 1$ runs at levels 0 and its mirror level -0, and $r^{(K)} > 1$ runs at level K and its mirror level -K. Thus at the start of Step (1) we have four sets of observations:

Observations at level 0: $\{y_j^{(0)} : j = 1, \ldots, r^{(0)}\}$

Mirror observations at level -0: $\{y_j^{-(0)} : j = 1, \ldots, r^{(0)}\}$

Observations at level K: $\{y_j^{(K)} : j = 1,..., r^{(K)}\}$

Mirror observations at level -K: $\{y_j^{-(K)} : j = 1,..., r^{(K)}\}$

We also place all the coefficients in the single, unclassified set $G_1 = \{\beta_1, \beta_2, \ldots, \beta_K\}$. Thus initially the number of sets is $p_1 = 1$. We assign 1 to s and proceed to the next step.

Step (s) ($s > 0$): At the beginning of Step (s), factors are partitioned into p_s sets:

$$G_{is} = \{\beta_j : k_{(i-1),s} < j \le k_{i,s}\}, \qquad i = 1,2,..., p_s \qquad (2.47)$$

For example, at Step (1) ($s = 1$ and $p_1 = 1$), there is only one set of factors and Equation (2.47) results in $G_{i,1} = \{\beta_j : k_{(i-1),1} < j \le k_{i,1}\}, i = 1$, where $k_{0,1} = 1$ and $k_{1,1} = K$. Thus $G_1 = G_{1,1} = \{\beta_1, \beta_2, \ldots, \beta_K\}$.

At the beginning of Step (s), some of the sets are already classified, some are unclassified. If all sets have been classified, then the algorithm ends. Otherwise we select any unclassified set: call this $G_{i,s}$.

1. If $G_{i,s}$ is not a singleton set (set of size one) we see if all the coefficients can be classified as unimportant. This can be done by considering the expected main effect of set $G_{i,s}$, which is $\overline{D}(k_{(i-1),s}, k_{i,s})$. Based on Equation 2.47, set $G_{i,s}$ includes factors $k_{(i\text{-}1),s}$ to $k_{i,s}$, therefore, for estimating $\overline{D}(k_{(i-1),s}, k_{i,s})$, the following four sets of observations are needed:

   Observations at level $k_{i,s}$: $\{y_j^{(k_{i,s})} : j = 1,..., r^{(k_{i,s})}\}$ (2.48-a)

   Mirror observations at level $k_{i,s}$: $\{y_j^{-(k_{i,s})} : j = 1,..., r^{(k_{i,s})}\}$ (2.48-b)

   Observations at level $k_{(i\text{-}1),s}$-1: $\{y_j^{(k_{(i-1),s}-1)} : j = 1,..., r^{(k_{(i-1),s}-1)}\}$ (2.48-c)

   Mirror observations at level $k_{(i\text{-}1),s}$-1: $\{y_j^{-(k_{(i-1),s}-1)} : j = 1,..., r^{(k_{(i-1),s}-1)}\}$. (2.48-d)

   From the assumption that $e(\mathbf{x},\mathbf{u})$ is normally distributed, $\overline{D}(k_{(i-1),s}, k_{i,s})$ is also normally distributed with mean and variance:

$$\mu(G_{i,s}) = \sum_{i=k_{(i-1),s}}^{k_{i,s}} \beta_i \,, \qquad \upsilon(G_{i,s}) = \frac{1}{8}\sigma^2 \left(1 \Big/ r^{(k_{(i-1),s}-1)} + 1 \Big/ r^{(k_{i,s})} \right).$$

For each new observation set (2.48-a), (2.48-b), (2.48-c) and (2.48-d), σ^2 can be estimated by:

S^2 for observation set (2.48-a): $S^2_{k_{i,s}} = \dfrac{\sum_{j=1}^{r^{(k_{i,s})}} (y_j^{(k_{i,s})} - \bar{y}^{(k_{i,s})})^2}{r^{(k_{i,s})} - 1}$

S^2 for observation set (2.48-b): $S^2_{-k_{i,s}} = \dfrac{\sum_{j=1}^{r^{(k_{i,s})}} (y_j^{-(k_{i,s})} - \bar{y}^{-(k_{i,s})})^2}{r^{(k_{i,s})} - 1}$

S^2 for observation set (2.48-c): $S^2_{k_{(i-1),s}-1} = \dfrac{\sum_{j=1}^{r^{(k_{(i-1),s}-1)}} (y_j^{(k_{(i-1),s}-1)} - \bar{y}^{(k_{(i-1),s}-1)})^2}{r^{(k_{(i-1),s}-1)} - 1}$

S^2 for observation set (2.48-d): $S^2_{-(k_{(i-1),s}-1)} = \dfrac{\sum_{j=1}^{r^{(k_{(i-1),s}-1)}} (y_j^{-(k_{(i-1),s}-1)} - \bar{y}^{-(k_{(i-1),s}-1)})^2}{r^{(k_{(i-1),s}-1)} - 1}$

and pooling these estimates gives an overall estimator of σ^2 for set $G_{i,s}$ as:

$$S^2_{G_{i,s}} = \frac{(r^{(k_{(i-1),s}-1)} - 1)(S^2_{k_{(i-1),s}-1} + S^2_{-(k_{(i-1),s}-1)}) + (r^{(k_{i,s})} - 1)(S^2_{k_{i,s}} + S^2_{-k_{i,s}})}{2(r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2)} \qquad (2.49)$$

and pooling the estimates of σ^2 over all stages gives an overall estimator of σ^2 at the start of Step (s) as:

$$S^2_s = \frac{\sum_{i=1}^{s} (r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2) S^2_{G_{i,s}}}{\sum_{i=1}^{s} (r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2)} \,. \qquad (2.50)$$

Now, in order to see if the main effect of set $G_{i,s}$ is significant, we test the following one-tail hypothesis:

$$\begin{aligned} &\mathrm{H}_0 : \ \mu(G_{i,s}) \le \delta \\ &\mathrm{H}_1 : \ \mu(G_{i,s}) > \delta \end{aligned} \qquad (2.51)$$

If H_0 is rejected, then the factor will be classified as important. If H_0 is not rejected, then the implication is that the factor does not have significant effect. Because rejecting H_0 is a strong conclusion, this formulation forces the factor to demonstrate that its main effect exceeds δ. In the other words, this formulation assumes that the effect is unimportant unless there is strong evidence to the contrary.

If the null hypothesis is correct, $\mu(G_{i,s}) = \delta$, then the quantity

$$t_0 = \frac{\overline{D}(j,k) - \delta}{\sqrt{\mathrm{Var}[\overline{D}(j,k)]}} = \frac{\overline{D}(j,k) - \delta}{\sqrt{\frac{1}{8}\sigma^2(1/r^{(j-1)} + 1/r^{(k)})}} = \frac{\overline{D}(j,k) - \delta}{\frac{1}{2\sqrt{2}} S_s \sqrt{1/r^{(j-1)} + 1/r^{(k)}}}$$

$$\sqrt{\frac{SS_s}{\sigma^2}/\omega} \qquad \sqrt{\frac{S_s^2}{\sigma^2}}$$

(2.52)

has a t-distribution with ω degrees of freedom, where $\omega = 2\sum_{i=1}^{s}(r^{(k_{(i-1),s}-1)} + r^{(k_{i,s})} - 2)$ and $SS_s = \omega S_s^2$.

If $t_{\alpha,\omega}$ denotes the upper α percentage point of the t-distribution with ω degrees of freedom, then if $t_0 > t_{\alpha,\omega}$, we reject the null hypothesis H_0 and conclude that the corresponding factor or group-factor is important. $t_0 > t_{\alpha,\omega}$ results in

$$\overline{D}(k_{(i-1),s}, k_{i,s}) > \delta + \frac{1}{2\sqrt{2}} t_{\alpha,\omega} S_s \sqrt{(1/r^{(k_{(i-1),s}-1)} + 1/r^{(k_{i,s})})} \qquad (2.53)$$

Therefore, if Inequality 2.53 holds, then with confidence (1- α):

$$\mu(G_{i,s}) > \delta \quad \text{i.e.} \sum_{i=k_{(i-1),s}}^{k_{i,s}} \beta_i > \delta$$

and thus, with confidence (1- α) we can consider set $G_{i,s}$ as important.

If the Inequality 2.53 is satisfied, then $G_{i,s}$ will be split into two groups at $k = \left\lceil (k_{(i-1),s} + k_{i,s})/2 \right\rceil$. The two new sets will replace $G_{i,s}$:

$$\{\beta_{k_{(i-1),s}}, ..., \beta_k\} \quad and \quad \{\beta_{k+1}, ..., \beta_{k_{i,s}}\}$$

and we increment *s* and go to Step (*s* + 1)

2. If $G_{is} = \{\beta_k\}$, i.e. it is a singleton set, then we proceed to fully classify β_k. This can be done by considering the expected main effect of set $\{\beta_k\}$, which is $\overline{D}(k,k)$. For estimating $\overline{D}(k,k)$, four sets of observations are needed:

$$\text{Observations at level } k\text{: } \{y_j^{(k_s)} : j = 1,\ldots, r^{(k_s)}\} \qquad (2.54\text{-a})$$

$$\text{Mirror observations at level } k\text{: } \{y_j^{-(k_s)} : j = 1,\ldots, r^{(k_s)}\} \qquad (2.54\text{-b})$$

$$\text{Observations at level } k\text{-1: } \{y_j^{(k_s-1)} : j = 1,\ldots, r^{(k_s-1)}\} \qquad (2.54\text{-c})$$

$$\text{Mirror observations at level } k\text{-1: } \{y_j^{-(k_s-1)} : j = 1,\ldots, r^{(k_s-1)}\} \qquad (2.54\text{-d})$$

Suppose that $S_{k_s}^2$, $S_{-k_s}^2$, $S_{k_s-1}^2$, and $S_{-(k_s-1)}^2$ are the sample variances of observations for Equations 2.54-a, 2.54-b, 2.54-c, and 2.54-d, respectively, then pooling these estimates gives an overall estimator of σ^2 for factor *k* in stage *s* as:

$$S_{k_s}^2 = \frac{(r^{(k_s-1)} - 1)(S_{k_s-1}^2 + S_{-(k_s-1)}^2) + (r^{(k_s)} - 1)(S_{k_s}^2 + S_{-k_s}^2)}{2(r^{(k_s-1)} + r^{(k_s)} - 2)} \qquad (2.55)$$

and pooling the estimates of σ^2 over all stages gives an overall estimator of σ^2 at the start of Step (*s*) as:

$$S_s^2 = \frac{\sum_{i=1}^{s} (r^{(k_s-1)} + r^{(k_s)} - 2) S_{k_s}^2}{\sum_{i=1}^{s} (r^{(k_s-1)} + r^{(k_s)} - 2)} \qquad (2.56)$$

If we set $\omega = 2\sum_{i=1}^{s} (r^{(k_s-1)} + r^{(k_s)} - 2)$, then a two-sided (1- α) confidence interval for β_k with upper and lower limits given by:

$$\beta_k^{\pm} = \overline{D}(k,k) \pm \frac{1}{2\sqrt{2}} t_{\alpha/2,\omega} S_s \sqrt{(1/r^{(k_s-1)} + 1/r^{(k_s)})} \qquad (2.57)$$

Similar to SB, now in order to decide on the importance of factor *k*, we follow two different strategies depending on the expense of making an additional run.

If simulation runs are expensive it is preferred to classify the factor without making any additional runs, so

i. if $\delta < \beta_k^-$, then β_k is classified as important,

ii. otherwise, β_k is classified as unimportant,

On the other hand, if simulation runs are not too expensive, we can have a better estimate of factor *k*'s main effect by making additional runs. If δ is contained in this interval we make additional runs at the levels *k*-1 and *k*. If $r^{(k_s-1)}$ and $r^{(k_s)}$ are not initially equal we add the runs at the level with the smaller number of runs, until $r^{(k_s-1)} = r^{(k_s)}$. Then we add runs at both levels *k*-1 and *k*, keeping $r^{(k_s-1)} = r^{(k_s)}$. As runs are added, the length of the confidence interval presented in Equation 2.57 decreases. We stop when either:

i. $\delta < \beta_k^-$, then β_k is classified as important,

ii. $\beta_k^+ < \delta$, then β_k is classified as unimportant,

Once, β_k is classified we increment *s* and go to Step (*s* + 1).

2.2.5. Further Observations

The key controlling factor in SB under uncertainty is setting the probability level in the hypothesis formulations for Equations 2.32 and 2.51. A high level (i.e. small α value) increases the probability of correctly determining the important and unimportant sets *I* and *U* but at the expense of additional observations being needed. A low level (i.e. large value of α) makes the overall determination process faster but with a higher risk of variables being assigned to the incorrect set.

Moreover, Cheng's method does not guarantee to control Type I Error (declaring an unimportant factor to be important) for each factor or power (declaring an important factor to be important) at any step. Wan et al. (2003) developed a procedure that controls Type I and Type II error during a factor screening experiments. This method is discussed in Section 2.3.

In addition, the performance of Cheng's method depends on the case considered. When the variances are large or unequal, Cheng's method loses control of the Type I Error and power (Wan et al. (2003)). In addition, Cheng's procedure is not valid under Common Random Numbers (CRN) because it assumes independence and equal variance for each observation.

Having applied Cheng's method to a simulation model, the next step would be to check the assumptions that Cheng's method holds. As mentioned before, Cheng's method assumes:

1. Errors are approximately normally distributed
2. The model is adequate: $E(\varepsilon) = 0$
3. Error variance is homogenous: $Var(\varepsilon) = \sigma^2$
4. Errors are statistically independent: $Cov(\varepsilon_i, \varepsilon_j) = 0$, for observations $i \neq j$

Assumptions 1 and 2 can be checked by using normal probability plots, assumption 3 by plotting residuals versus actual fits and assumption 4 by plotting residuals in run order or plotting e_i versus e_{i-1} (Tunali et al. (2000)). Details of these tests will be discussed during its application in Chapter 4.

2.3. Controlled Sequential Bifurcation (CSB)

The SB proposed by Bettonvil et al. (1996) has no performance guarantee for the stochastic case. Controlled Sequential Bifurcation (CSB), proposed by Wan et al. (2003) is a procedure that incorporates a two-stage hypothesis-testing approach into SB to control error and power, when factor screening a stochastic simulation model. Similar to basic SB, Wan et al. (2003) do not assume that the main-effects model holds across the entire range of the factors x; rather, they assume that it is a good local approximation for modest deviations from a nominal level, typically the center of the design space. CSB controls the power at each bifurcation step and Type I Error for each factor under *heterogeneous* variance conditions.

CSB assumes that the response function (y) can be modeled by a first-order polynomial function around point $\mathbf{x} = (x_1, x_2, \ldots, x_K)$ as follows:

$$y = \beta_0 + \sum_{j=1}^{K} \beta_j x_j + \varepsilon(\mathbf{x}) \qquad (2.61)$$

where:

y: response of the metamodel (the sequential bifurcation approximation)

K: total number of factors in the experiment

β_j : first-order or main effect of factor j with j =1, …, K

x_j: value of factor j, standardized to lie in [-1,+1]

$\varepsilon(\mathbf{x})$: the approximate error, assumed to be approximately normally distributed with mean zero and variance $\sigma^2(\mathbf{x})$.

In CSB, the analyst must specify two thresholds. The lower threshold (Δ_0) indicates the level that the main effect of a factor must reach to be considered *important*, while factors with main effects larger than the higher threshold (Δ_1) are considered *critical*. In other words, for factor j if $\beta_j \leq \Delta_0$, then factor j is classified as unimportant and if $\beta_j \geq \Delta_1$, factor j is classified critical, and otherwise if $\Delta_0 < \beta_j < \Delta_1$, it is classified as being important.

CSB assumes that the sign of each factor effect is known so that we can set the levels of the factors to have $\beta_i \geq 0$ for all $i > 0$. Further, we assume that for a fixed factor setting, $\mathbf{x} = (x_1, x_2, \ldots, x_K)$, replications of Expression 2.61 are independent and identically distributed (IID); dependence of outputs across different factor settings due to CRN is permitted.

CSB uses a hypothesis-testing approach to control the probability of Type I error (i.e., the probability an effect is classified as important when it is not) and power (i.e., the probability that an important effect is correctly classified). More specifically, for those factors with effects $\leq \Delta_0$, CSB controls the Type I Error by declaring them important to be less than α; and for those factors with effects $\geq \Delta_1$, CSB provides power for identifying them as important to be greater than γ. Those factors whose effects fall between Δ_0 and Δ_1 are considered important and the CSB procedure has reasonable, though not guaranteed, power to identify them (Wan et al. (2003)). In summary, the CSB procedure controls the Type I Error for each factor individually and guarantees the power at each step.

Like other group screening designs, CSB begins with all factors placed in a single group and the group's accumulated effect is tested. If the group's effect is classified as unimportant,

then all factors within the group are classified as unimportant. Otherwise, the group is split into two smaller ones for further testing. If the group contains only one factor, this factor is classified as important. This procedure continues until all factors have been classified. One important advantage of CSB is that it rarely misclassifies a critical factor as unimportant, or an unimportant factor as important (Sanchez et al. (2005b)). A detailed procedure of CSB is presented in Wan et al. (2003).

For the special case, where $\alpha = 1 - \gamma$ (Type I error is equal to one minus power), Wan et al. (2005) implements a fully sequential test in CSB that has the same error control as the two-stage testing procedure. The test adds one replication at a time to both the upper and lower levels of the group being tested until a decision is made. Wan et al. (2005) mentioned that in most cases the sequential test is more efficient than the two-stage testing procedure.

Wan et al. (2003) observed that CSB has superior performance to Cheng's method in large and unequal variance cases. CSB has guaranteed performance with different parameter and factor configurations, which makes it attractive for problems with limited prior knowledge, especially about the variance of the error term. Cheng's method, on the other hand, assumes variance homogeneity to gain advantages in degrees of freedom and it can be effective when this assumption is satisfied.

Wan et al. (2004) improved the CSB procedure by incorporating a fold-over design in the hypothesis test to identify important main effects even when two-factor interactions and quadratic terms are present. The new procedure, called CSB-X, still has the same error control for screening main effects. At each step, the design gives an unbiased estimate of the main effect of the group, which is defined as the summation of the main effects of all factors in the group, even when two-factor interactions and quadratic terms are present. However, CSB-X is not able to estimate interaction and quadratic effects.

2.4. Latin Hypercube Sampling (LHS)

Latin Hypercube Sampling (LHS) was proposed by McKay (McKay et al. (1979)) for situations involving a relatively large number of factors. Latin hypercube (LH) sampling

provides a flexible way of constructing efficient designs for quantitative factors. Let k be the number of factors and let n denote the number of design points desired ($n \geq k$), and define n levels per factor. The low and high levels for factor x_i are coded as 1 and n, respectively, and the set of coded factor levels are $\{1, 2, \ldots, n\}$. In *Random LH* each column of the design matrix is a random permutation of the factor levels. So in one replication, each of the k factors will be sampled exactly once at each of its n levels. LH designs have good space-filling properties; that is, the design points are scattered throughout the experimental region with minimal unsampled regions. Unlike the 2^k factorial design, the LH design provides some information about the interior of the experimental region. The main benefit of LH sampling is its efficiency in terms of the number of required runs. The smallest LH designs are square, with $n = k$, so the number of design points grows linearly with k rather than exponentially (Sanchez (2005b)).

LH designs can be used in factor screening. One approach to determine the important factors by using LH sampling is to build a metamodel for the LH designs. Then based on the estimated parameters, the factors with important effects can be classified as important. In order to have uncorrelated estimates for the metamodel parameters, the design matrix should have orthogonal columns, or equivalently zero correlations. This property is critical for factor screening experiments, because if the parameter estimates are biased, the factors may not be classified correctly as important or unimportant. Cioppa (2005) measures the degree of orthogonality of a design matrix with the maximum pairwise correlation of the columns of the design matrix, denoted by ρ. A design matrix with $\rho = 0$ is truly orthogonal, and a design matrix with $\rho = 1$ has at least one column that is a linear combination of the remaining columns.

It is not always easy to generate an orthogonal LH. Random LH designs, where the elements of each column are permutated randomly, have good orthogonality properties if n is much larger than k, but for smaller designs some factors might have high pairwise correlations. One approach often taken is to randomly generate many LH designs and then choose a good one (Sanchez (2005b)). Ye (1998) described a procedure to construct an Orthogonal Latin Hypercube (OLH) when its number of rows n is a power of 2 or a power of 2 plus 1; i.e. for $n = 2^m$ or $2^m + 1$, an OLH with $2m - 2$ columns can be constructed. Later, Cioppa (2005)

extended the Ye's procedure (Ye (1998)) to incorporate more factors into the design matrix. Table 2.2 from Cioppa (2005) shows the number of factors that can be screened by Ye's method and extended Ye's method.

Table 2.2: Number of factors for Ye's method and Ye's extended method

		Maximum number of factors by	
Number of runs	m	Ye's method	Extended Ye's method
17	4	7	6
33	5	11	8
65	6	16	10
129	7	22	12

They, however, noticed that the space-filling of these new designs is poor; therefore, they suggested sacrificing some of the orthogonality intentionally in order to achieve better space-filling while incorporating a greater number of factors. These designs are called Nearly Orthogonal Latin Hypercubes (NOLH). They define a NOLH design as a design which has a maximum pairwise correlation no greater than 0.03 and a *condition number* no greater than 1.13 (For more information about the condition number of a matrix, please refer to Cioppa (2005).) Based on their procedure, Sanchez has implemented an Excel file which produces Nearly Orthogonal Latin Hypercube (NOLH) for less than 29 factors (Sanchez (2005a)). The numbers of design points required for investigating $k \leq 29$ factors are provided in Table 2.3.

Table 2.3: Number of runs for NOLH designs

Number of factors	Number of design points
2-7	17
8-11	33
12-16	65
17-22	129
23-29	257

The major disadvantage of the OLH and NOLH designs is that these designs are not yet available for more than 29 factors. In addition, despite having a good space-filling property, these designs require considerably more runs than Resolution IV designs.

2.5. Comparing Factor Screening Methods

Different factor screening methods have different characteristics in terms of structure, capability and performance. In this section, several criteria are considered for evaluating the factor screening methods. Then, based on the proposed criteria, the screening methods are evaluated and compared.

2.5.1. Criteria for Evaluating Designs

This section describes criteria for evaluating experimental designs which are generally applied to simulation models. We will use these criteria to compare the factor screening methods to understand the strength and weakness of each.

2.5.1.1 Number of Runs

One of the important attributes on which a design can be evaluated is the number of runs it requires for estimating the metamodel parameters or determining the important factors. However, for stochastic models, where a number of observations should be obtained for each scenario, it is more realistic to evaluate a method based on the number of required observations instead of number of required scenarios.

A design is called *saturated* if the number of required runs, n, equals the number of metamodel parameters, q. For example, if the metamodel is a first-order polynomial in k factors, then $q = k + 1$ (where 1 refers to the grand or overall mean, often denoted by β_0); and a saturated design for this model requires $n = k + 1$ runs to estimate the metamodel parameters (here only main effects).

A screening method is called *saturated* if it can determine the important factors among the K involved factors with K runs. And it is called *supersaturated* if it can determine the important factors with less that K runs.

2.5.1.2 Accuracy

In general, in screening experiments we want (1) to detect as many important factors as possible, and (2) to declare important as few unimportant factors as possible. Accuracy, which also has been called 'effectiveness' (Trocine et al. (2001)), is often difficult to measure, because in practical problems the underlying coefficients of the effect are unknown. In this thesis, we measure the accuracy of the factor screening results by generating a metamodel for the factors identified as being important. Then the goodness of fit of the metamodel is evaluated.

2.5.1.3 Orthogonality

A design is said to be *orthogonal* if the columns of the design matrix are orthogonal (i.e., the inner product of any two columns is zero). Orthogonality has long been a desirable criterion for evaluating designs. For an orthogonal design, since the input factors are uncorrelated, it simplifies interpretation of the results. Lack of orthogonality, also called *multicolinearity*, implies that the effect estimates are not independent or cannot be computed at all.

Unfortunately, building an orthogonal design also has limitations. In reality, some factor level combinations may not be feasible. For example, in an M/M/1 queue the expected steady-state waiting time is infinite if the arrival rate exceeds the service rate. In general, forcing orthogonal designs may limit many factors to narrower ranges. Unfortunately, in complex models it may not be possible to know a priori which factor-level combinations are problematic; therefore, in these situations, in order to avoid getting unrealistic response values, all the factor ranges should be narrowed.

2.5.1.4 Space Filling

Space-filling designs sample not only at the edges of the hypercube that defines the experimental area, but also in the interior. A design with good space-filling properties frees analysts from making many assumptions about the behavior of the response surface. Kleijnen et al. (2005) believes that space-filling designs, such as the Latin Hypercube Design (LHD),

currently provide the best way of exploring surfaces where we do not expect to have smooth metamodels.

2.5.1.5 Strictness of Assumptions

As the number of factors increases, to be more efficient, factor screening methods tend to require more restrictive assumptions. A first-order polynomial function is generally assumed for the response by methods that screen a large number of factors, such as SB and PB designs. If the runs are not prohibitively expensive, the first-order polynomial function can be augmented by second-order interaction terms. SB-X and Resolution IV and V designs require this assumption. Moreover, it is desirable that a factor screening method can be easily adapted to account for curvature, for example by introducing a central point for factors.

Additionally, the group screening methods, including SB and all its variants, assume that the interaction between two factors is important only if both factors have important main effects. This assumption may not always be true in practice; therefore, a method which does not require this assumption is more desirable.

In addition, most of the classical factor screening methods such as Fractional Factorials, SB, and Cheng's method assume variance homogeneity over the entire experimental region. The newly developed methods such as CSB, however, allow the assumption of variance heterogeneity. Although this more general assumption makes CSB less efficient, its result is more reliable for situations where the typical error assumptions are violated.

In general, although hard to achieve, it is desirable for factor screening methods to be able to screen a large number of factors with the least restrictive assumptions, while maintaining reasonable efficiency.

2.5.1.6 Ability to Handle Constraints on Factor-Level Combinations

In many situations, the value that a factor can attain is completely dependent on the values of other factors. For example, the values of a number of factors must add up to 100%. Or, for example in many queuing situations, certain combinations of factor settings give unstable

outputs. The classic DOE literature presents *mixture* designs for these situations (Montgomery (2000)). Many designs exist for exploring experimental regions (i.e., permissible combinations of design points) that are either hypercubes or spheres. In simulation experiments, restricting factor values to realistic combinations may complicate the design process dramatically. Until designs that can handle such situations are available, Kleijnen et al. (2005) believes that the visual presentation of the results may be the most appropriate ways of determining whether these situations exist.

2.5.1.7 Ease of Design Construction and Implementation

Simulation models usually involve a large number of factors. Consequently, the number of runs required for the factor screening experiment is too large to be made manually. Therefore, it is necessary to implement computer code to automate the process of factor screening a simulation model. As a result, the factor screening methods need to be simple enough to be implemented in a computer code.

In addition, the results of factor screening experiments are generally used by the simulation modelers who are not particularly familiar with the statistical aspects of the factor screening experiments. Therefore, it is desirable for a factor screening method to generate results that can be interpreted by the modelers without requiring too much statistical knowledge.

2.5.2. Evaluating Screening Methods

Based on the criteria described for evaluating screening methods in Section 2.5.1 and the structures of factor screening methods explained in Chapter 2, we can now classify the methods according to the number of factors they can screen and the assumptions they require. Figure 2.2, which is partially driven from Kleijnen et al. (2005) shows suitable methods for particular situations.

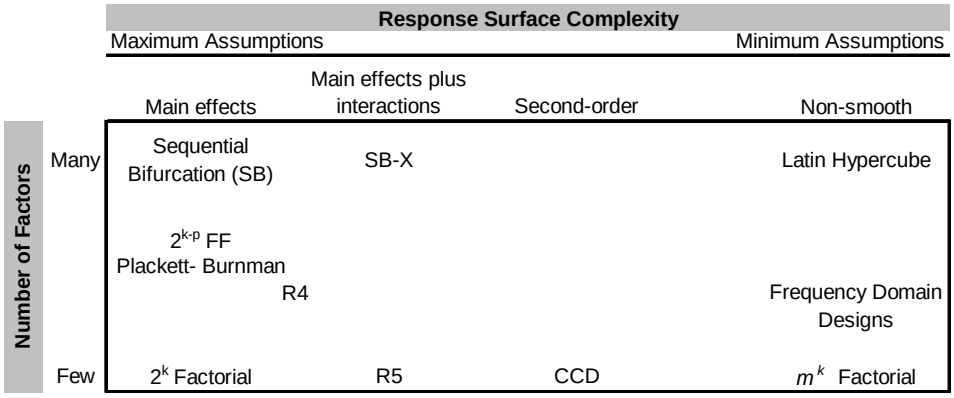


Figure 2.2: Recommended designs for simulation models

In Figure 2.2, the horizontal axis represents a continuum from simple to complex response surfaces. A simple response surface can be modeled by a first-order polynomial function and identically and independently distributed error terms, whereas a complex response surface cannot simply be modeled by low-order polynomial functions, and the error terms can present a more complex nature. The vertical axis represents the number of factors that the method can screen.

In Figure 2.2, the lower left represents simple response surfaces with only a handful of factors, while the upper right represents very complex response surfaces with many factors. An analyst willing to make simplifying assumptions can start from the left of the figure. If little is known about the response surface, the analyst can start from the upper right of the figure for an initial experiment. In the present thesis, we have employed all the methods shown in Figure 2.2 except the Frequency Domain Methodology (FDM). More information about FDM can be found in papers by Jacobson et al. (1991), Sanchez et al. (1987), Sanchez et al. (2003), and Schruben et al. (1987). Now, we evaluate and compare the factor screening methods discussed in this chapter based on some of the criteria introduced in Section 2.5.1.

For deterministic simulation models, the factor screening methods require only one observation for each run (or scenario). For these models, therefore, the number of required runs is equal to the number of obtained observations. On the other hand, for stochastic simulation models, depending on the factors screening method being used, different number of observations may be needed for each run (or scenario).

Figure 2.3 compares the number of required runs for each factor screening method for the deterministic models. The number of runs for R4 is based on Table 1.2, for R5 is based on Table 1.3, and for the Orthogonal Latin Hypercube Design (OLHD) is based on Table 2.3. The number of needed runs by SB is calculated based on Equation 2.5, with the probability of a factor to be important equal to 0.15. Moreover, as observed in Section 2.1.3.3, SB-X requires twice as many runs as SB.

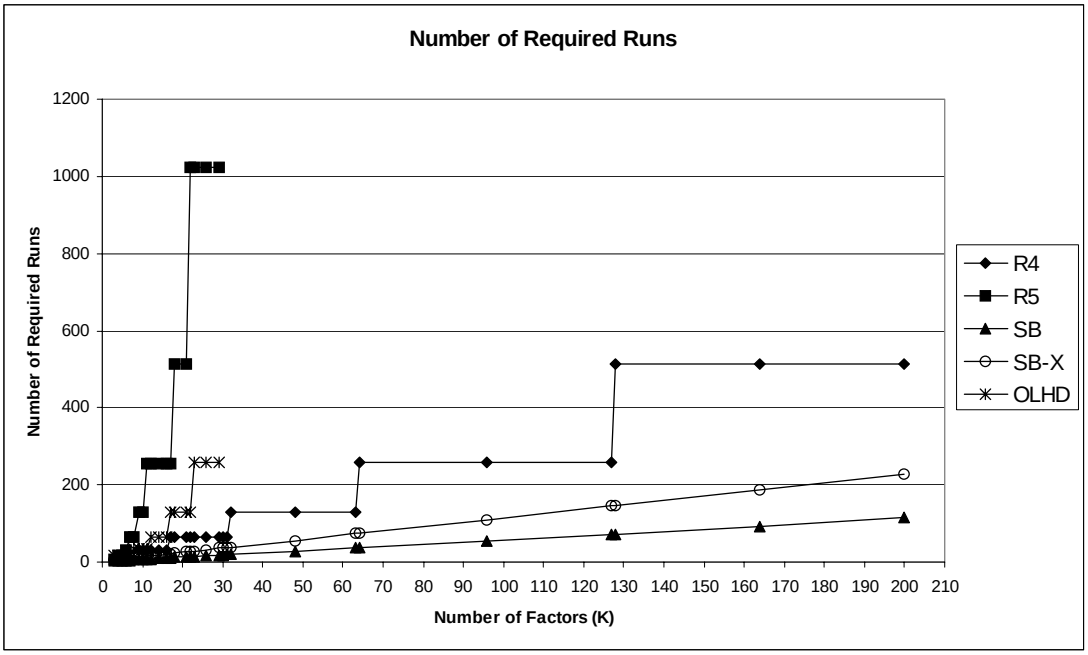


Figure 2.3: Number of required runs for the factor screening methods

Figure 2.3 indicates that SB and SB-X are the most efficient methods in terms of required number of runs. However, it should be noted that one assumption of SB and SB-X is the known signs of main effects; when this assumption cannot be easily satisfied, prior to applying SB or SB-X, a Resolution III designs (such as BP) should be used to determine the directions of the main effects. The number of runs required by this initial experiment is almost equal to the number of involved factors.

Orthogonal Latin Hypercube (OLH) designs are only available for up to 29 factors (based on the procedure proposed by Sanchez (2005a)), and even in that range it is considered an

expensive design as compared with other methods. R5 designs are generally available for up to 120 factors in the statistical software; but they are so expensive that they are infrequently used for screening more than a handful of factors.

Figure 2.3 is also applicable for the stochastic simulation models, with the difference that instead of SB and SB-X, Cheng's method, Modified Cheng's method, CSB, and CSB-X are used. In addition, for the stochastic models, the number of observations obtained for each treatment (run) should also be taken into consideration. For R4, R5 and OLH designs, depending on the desired level of accuracy for the parameter estimates, a certain number of observations are obtained for each treatment. For Cheng's and Modified Cheng's methods, however, the number of observations should be determined such that the required statistical comparison at each step can be conducted. CSB and CSB-X, depending on the response variance, may obtain different numbers of observations for different treatments.

From the perspective of the necessary number of observations, the performance of Cheng's method, the Modified Cheng's method, CSB and CSB-X, deteriorates as the response variance increases. This deterioration occurs because they need more observations at each step to be able to conduct the necessary statistical comparisons. As a result, one major advantage of R4 designs over the sequential designs is that R4 designs do not require an enormous number of observations to detect the important factors in the case of a high response variance. Moreover, although high variance reduces the accuracy of the parameter estimates, but for factor screening purposes, the parameters are not required to be estimated with high level of accuracy. After conducting a factor screening experiment, a metamodel can be generated using the important factors in which the parameters are estimated with the desired degree of accuracy.

In terms of the metamodel complexity assumed for the response, LH designs are the most flexible, because they do not assume any specific type of response function. A stepwise approach is usually used to fit a metamodel to the data obtained according to a LH design. Using this approach, a simple metamodel (e.g. first-order polynomial) is formed and then, more complex terms (e.g. second or third-order interactions or pure quadratic) are added to the metamodel until the fitted metamodel shows an appropriate goodness-of-fit or all the

degrees of freedom are consumed. Likewise, R4 and R5 designs can easily be augmented by Central Composite Designs (CCD) to incorporate the quadratic terms. It is therefore possible for the LH, R4 and R5 designs to detect the important factors that do not present significant main effect but present important quadratic effects. On the other hand, SB and its variants (generally sequential designs) are not able to detect an important factor with insignificant main effects and important quadratic effects because they classify factors only based on the main effect estimates.

Among the described methods, the LH designs have the best space-filling property. For the R4 and R5 design, this property can be improved significantly by incorporating CC designs into the R4 or R5 designs. In contrast, SB and CSB and their variants have the poorest space-filling property, because they only consider upper and lower levels for factor screening.

Variance homogeneity is always required by R4, R5 and LH designs. However, when this assumption is violated, a corrective action, such as transformation of the response function, may be considered. Although in the transformed metamodel the parameter estimates do not convey perceptible meaning, the important factors can still be detected as if no transformation had been applied. CSB and CSB-X are claimed to function correctly even if the variance homogeneity assumption is violated. But Cheng's method and the Modified Cheng's method may lead to false results if this assumption is not met. Moreover, transformation is not very useful for these two methods, because determining an appropriate delta limit (δ) when the transformation has been applied can be impractical.

2.6. Conclusion

This chapter described several factor screening methods which are generally used for screening with simulation models, as well as the subsequent issues usually occurring when these methods are applied. Each of these methods has advantages and disadvantages, and are suitable only for certain situations. Several criteria were addressed in this chapter for evaluating the factor screening methods among which, the number of required observations, space-filling property, and the strictness of assumptions appeared to be of greatest

importance. None of the factor screening methods are claimed to be superior in all the criteria. A method with good space-filling property or a complex metamodel requires an excessive number of runs; or a method with a moderate number of required runs has to make several simplifying and strict assumptions.

This chapter indicated that each factor screening method is suitable for certain situations specified by the complexity of the response function and the number of factors to be screened. In the next two chapters, we apply the methods discussed previously to two medical simulation models, and then evaluate the performance of each method.

3. Factor Screening on Deterministic Models

This chapter provides the results from several factor screening experiments on an available deterministic model, called the Drug Model. The methods will be evaluated in terms of the number of required runs and the goodness of fit of the metamodel built for the factors detected as important.

A major issue with using *statistical* tools such as ANOVA in analyzing a *deterministic* simulation model is the fact that these statistical tests do not convey reasonable meaning in the context of deterministic models. In other words, F-ratio, *p*-value, and standard errors calculated in ANOVA are meaningful only when the data are obtained from a stochastic response. Therefore, through this chapter, whenever we apply a statistical test on the data, we assume that the data are obtained according to a single replicate design.

3.1. Factor Screening on Drug Model

The Drug Model is a proprietary deterministic simulation model which consists of 34 factors; but according to an expert's opinion, due to dependency, some of the factors should be confounded. The values of the confounded factors cannot be changed independently; thus, we treat the confounded factors together as a single factor. After appropriately combining confounded factors, we performed the factor screening experiment on 25 factors. Since the number of factors is not excessive in the Drug model and the model executes quickly, a 2^k-Fractional Factorial with resolution IV is a reasonable design to be used as a factor screening method. It is efficient in terms of number of required runs and it provides unbiased estimates for all main effects. Sequential Bifurcation augmented with a fold-over design (SB-X) and Latin Hypercube Design (LHD) are other methods that are used for factor screening the Drug model.

As mentioned previously, a computer code has been implemented for factor screening the available simulation models. The complete structure of the code is discussed in Appendix A. For SB-X, which is a sequential procedure, the code generates the appropriate design at each Stage, whereas for 2^k Fractional Factorial and Latin Hypercube designs, the user should provide the code with the proper designs. The code can read those designs via an Excel file and returns the obtained observations for each treatment. We used JMP and Sanchez (2005a) to generate the 2^k Fractional Factorial and LH designs, respectively.

3.1.1. 2^k Fractional Factorial Design

JMP was used to generate a 2^k fractional factorial design of resolution IV. This design required 64 runs for estimating the factors' main effects. Since the Drug model is a deterministic model, only one observation is needed for each scenario.

It is possible that for some combinations of upper and lower levels of factors, the model generates an unrealistic response. Therefore, after making 64 observations, we showed the distribution plot of observed responses to our client to make sure that all the response values are at a reasonable range. The plot of the response distribution is shown in Figure 3.1.

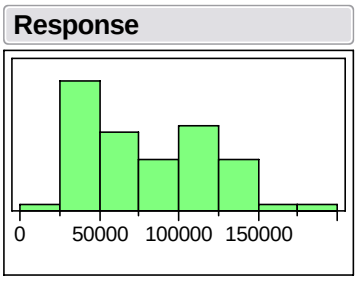


Figure 3.1: Response distribution of the Drug Model

Analysis of variance and fit for the 2^k fractional factorial design are shown in Table 3.1 and Table 3.2.

Table 3.1: Summary of fit for the 2^k - FF design

Summary of Fit	
RSquare	0.999916
RSquare Adj	0.999243
Root Mean Square Error	1096.687
Mean of Response	79994.67
Observations (or Sum Wgts)	64

Table 3.2: ANOVA for the 2^k - FF design

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	56	1.0013e+11	1.788e+9	1486.609
Error	7	8419054.33	1202722	Prob > F
C. Total	63	1.0014e+11		<.0001*

Both the R^2_{Adj} and the model *p*-value shows that the generated metamodel satisfactorily fits the data. $R^2_{Adj} = 0.999$ implies that 99.9% of the variation in the data can be accounted for by the model. Therefore, a first-order polynomial function augmented with second-order interactions effects can be a good approximation for the response function. It should, however, be noted that for the deterministic models the F-test and *p*-value do not carry any meaningful interpretation.

The plot of actual versus predicted response is shown in Figure 3.2 and the normal probability plot for residuals is depicted in Figure 3.3. Both of these plots indicate that the normal distribution assumption for residuals seems reasonable and there is no significant evidence for the violation of this assumption.

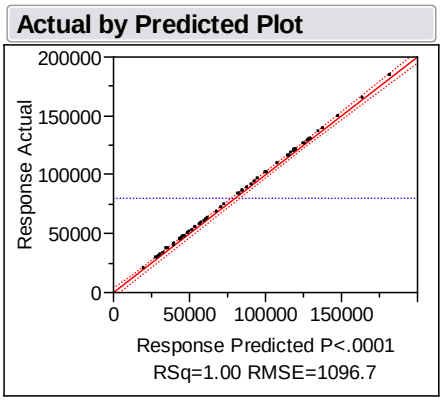


Figure 3.2: Actual by predicted data for 2^k - FF design

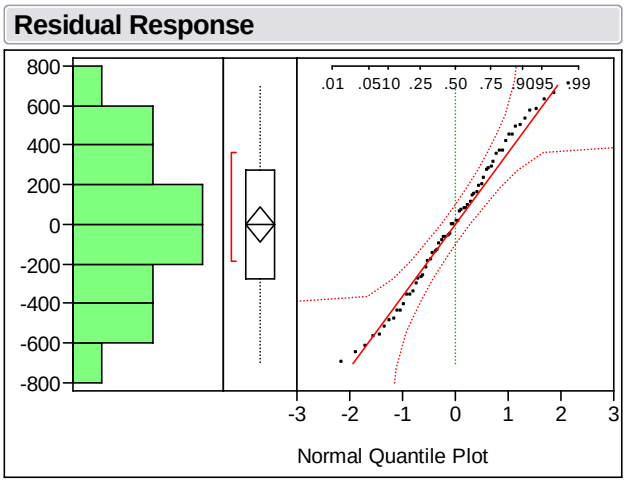


Figure 3.3: Residual distribution for 2k - FF design

3.1.1.1 Adding center point

Before judging the significance of each factor, the existence of nonlinearity effects in the response function should be checked by adding a center point (a point with all factors at their central levels). The analysis of variance and summary of fit for the new model with the center point added are shown in Table 3.3 and Table 3.4

Table 3.3: Summary of Fit for 2^k - FF with center point

Summary of Fit	
RSquare	0.999856
RSquare Adj	0.998847
Root Mean Square Error	1343.113
Mean of Response	79956.65
Observations (or Sum Wgts)	65

Table 3.4: ANOVA for 2^k - FF with center point

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	56	1.0013e+11	1.788e+9	991.1448
Error	8	14431616.9	1803952.1	Prob > F
C. Total	64	1.0014e+11		<.0001*

Adding the center point reduces R^2_{Adj} from 0.9992 to 0.9988 and F-Ratio from 1486.609 to 991.1448. Although the reduction in R^2_{Adj} is not significant, the decrease in F-Ratio is not negligible and requires further investigation.

As shown in Figure 3.4 and Figure 3.5, the center point acts as an outlier. Both figures show that the residual for the center point is remarkably greater than the residuals for other design points. All of this evidence suggests that a first-order polynomial augmented by second-order interactions may not be a good approximation of the response function. Therefore, a Central Composite Design (CCD) is used to estimate the quadratic effects in the data.

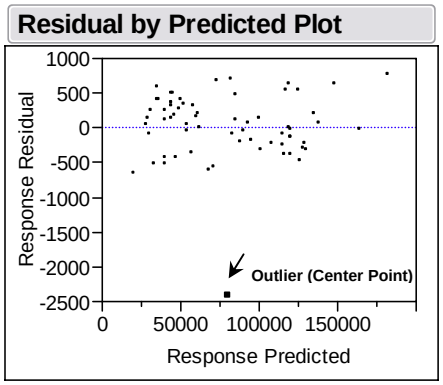


Figure 3.4: Residual by predicted response for 2^k - FF with center point

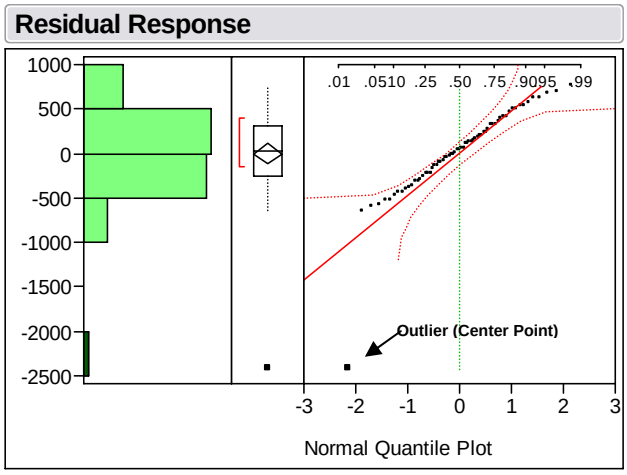


Figure 3.5: Residual Distribution for 2^k - FF with center point

3.1.1.2 Central Composite Design

The Central Composite Design (CCD) is the most popular class of designs used for fitting a second-order model. The created "on-face CCD" requires 2^{25-19} runs for a fractional factorial of resolution IV, and 2×25 runs for axial designs, and 1 run for central point; thus a total of 115 runs are needed. An on-face CCD provides more information about each factor. In this design, two more observations are added for each factor *j*. For the first observation, factor *j* is at its upper level while other factors are at their center level, and for the second observation, factor *j* is at its lower level while other factors are at their center level. The analysis of variance and fit for the CCD are shown in Table 3.5 and Table 3.6.

Table 3.5: Summary of fit for CCD

Summary of Fit	
RSquare	0.99989
RSquare Adj	0.999621
Root Mean Square Error	585.7111
Mean of Response	78940.56
Observations (or Sum Wgts)	115

Table 3.6: ANOVA for CDD

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	81	1.0317e+11	1.2738e+9	3712.961
Error	33	11320897.4	343057.5	Prob > F
C. Total	114	1.0319e+11		<.0001*

Both the R^2_{Adj} and the model *p*-value show that the generated metamodel satisfactorily fits the data. $R^2_{Adj} = 0.999$ implies that 99.9% of the variations in the data can be accounted for by the model.

The plot of actual versus predicted response is shown in Figure 3.6 and the normal probability plot for residuals is depicted in Figure 3.7. Both of these plots show that the normal distribution assumption for residual seems reasonable and there is no significant evidence that implies a violation of this assumption.

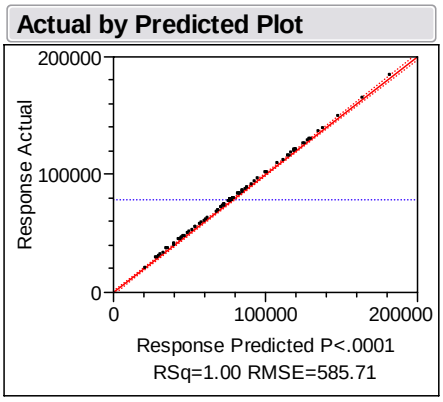


Figure 3.6: Actual by predicted data for CCD

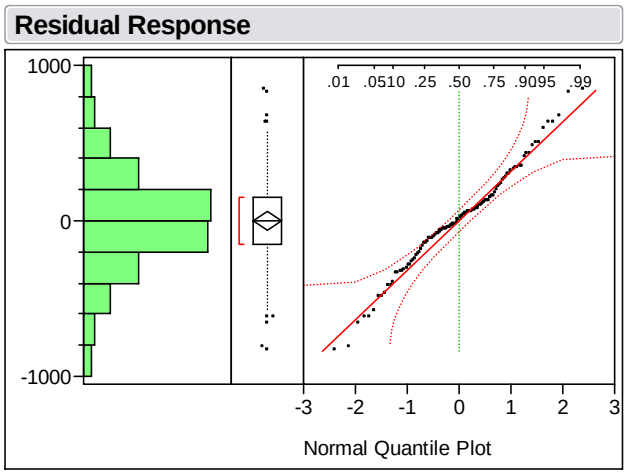


Figure 3.7: Residual distribution for CCD

3.1.1.3 Finding Important Factors

The client wanted to find the factors that would produce a change in the response function of $5000 or more. Therefore, for the Drug model an *important* factor was defined as a factor able to cause a $5000 change in the output.

As a first step to identify the important factors, the "Rule of 2" is employed. This criterion drops any effect with F-Ratio less than 2 from the model. For more information about "Rule of 2", refer to Wallace (1977). The analysis of variance and fit for the restricted model are shown in Table 3.7 and Table 3.8.

Table 3.7: Summary of fit for the restricted model

Summary of Fit	
RSquare	0.999625
RSquare Adj	0.999397
Root Mean Square Error	738.4825
Mean of Response	78940.56
Observations (or Sum Wgts)	115

Table 3.8: ANOVA for the restricted model

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	43	1.0315e+11	2.3988e+9	4398.535
Error	71	38720308.7	545356.46	Prob > F
C. Total	114	1.0319e+11		<.0001*

Both the R^2_{Adj} and the model F-Ratio have increased. The increase in R^2_{Adj} is not significant, but the model F-ratio has increased noticeably.

Table 3.9 displays the estimated effects of the generated metamodel. Recall that an *important* factor is one that will cause a $5000 change in the output. Therefore, if a factor has a main effect greater than $5000/2 = $2500, or a quadratic effect greater than $5000, or is involved in a second-order interaction effect greater that $5000/2 = $2500, it is declared to be important. Table 3.10 lists the important factors.

Table 3.9: Parameter estimates for CCD

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	77582.572	105.4975	735.40	<.0001*
X1	34729.802	90.90092	382.06	<.0001*
X2	323.20964	90.90092	3.56	0.0007*
X3	-1015.555	90.90092	-11.17	<.0001*
X5	-5370.167	90.90092	-59.08	<.0001*
X6	-579.185	90.90092	-6.37	<.0001*
X7	9008.4596	90.90092	99.10	<.0001*
X8	1352.0727	90.90092	14.87	<.0001*
X9	-5217.852	90.90092	-57.40	<.0001*
X11	-6844.832	90.90092	-75.30	<.0001*
X13	302.22846	90.90092	3.32	0.0014*
X16	-9907.881	90.90092	-109.0	<.0001*
X18	-224.8697	90.90092	-2.47	0.0158*
X19	1185.3053	90.90092	13.04	<.0001*
X20	1544.9978	90.90092	17.00	<.0001*
X21	1041.3688	90.90092	11.46	<.0001*
X23	486.47313	90.90092	5.35	<.0001*
X24	-1298.365	90.90092	-14.28	<.0001*
X27	-268.6111	90.90092	-2.95	0.0042*
X29	-4363.485	90.90092	-48.00	<.0001*
X30	268.38015	90.90092	2.95	0.0043*
X34	347.05847	90.90092	3.82	0.0003*
X1*X2	132.02197	92.31032	1.43	0.1570
X1*X5	-150.7269	92.31032	-1.63	0.1069
X1*X7	3344.7731	92.31032	36.23	<.0001*
X1*X8	1037.7573	92.31032	11.24	<.0001*
X1*X9	-1879.518	92.31032	-20.36	<.0001*
X1*X11	-2368.665	92.31032	-25.66	<.0001*
X1*X13	474.01233	92.31032	5.13	<.0001*
X1*X16	-3311.14	92.31032	-35.87	<.0001*
X1*X19	670.2912	92.31032	7.26	<.0001*
X1*X20	1144.0938	92.31032	12.39	<.0001*
X1*X21	601.8327	92.31032	6.52	<.0001*
X1*X24	-971.5914	92.31032	-10.53	<.0001*
X1*X27	-1187.343	92.31032	-12.86	<.0001*
X1*X29	-2091.306	92.31032	-22.66	<.0001*
X1*X30	791.26544	92.31032	8.57	<.0001*
X1*X34	649.04553	92.31032	7.03	<.0001*
X2*X6	-941.7017	92.31032	-10.20	<.0001*
X2*X9	-186.0884	92.31032	-2.02	0.0476*
X2*X11	-456.553	92.31032	-4.95	<.0001*
X3*X8	344.54431	92.31032	3.73	0.0004*
X7*X34	242.55652	92.31032	2.63	0.0105*
X16*X16	2366.1858	139.2577	16.99	<.0001*

Table 3.10: List of important factors for the Drug Model

Screening ID	Effect	Main effect	Expert Guess
1	X1	34729.802	Important
5	X5	-5370.167	Important
7	X7	9008.4596	Important
9	X9	-5217.852	Unknown
11	X11	-6844.832	Important
16	X16	-9907.881	Important
29	X29	-4363.484	Important

3.1.1.4 Generating the Metamodel

A metamodel can now be constructed using the on-face central composite design composed of the 7 factors determined to be important. This design requires 2^{7-1} factorial runs for a resolution V design, 14 axial and 1 center point run. The analysis of variance and fit for the model are shown in Table 3.11 and Table 3.12.

Table 3.11: Summary of fit for the generated metamodel

Summary of Fit	
RSquare	0.999798
RSquare Adj	0.999633
Root Mean Square Error	687.7485
Mean of Response	79290.21
Observations (or Sum Wgts)	79

Table 3.12: ANOVA for the generated metamodel

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	35	1.0055e+11	2.8729e+9	6073.863
Error	43	20338912.3	472997.96	Prob > F
C. Total	78	1.0057e+11		<.0001*

Both the R^2_{Adj} and the model *p*-value show that the generated metamodel satisfactorily fits the data.

The plot of actual versus predicted response is shown in Figure 3.8 and the normal probability plot for residuals is given in Figure 3.9. Both of these plots show that the normal distribution assumption for residuals appears reasonable and there is no significant evidence that implies the violation of this assumption.

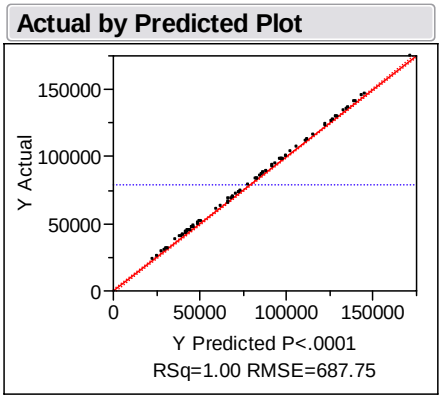


Figure 3.8: Actual by predicted responses

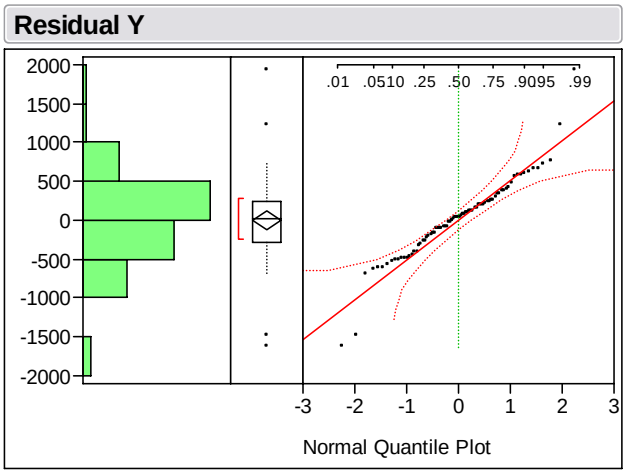


Figure 3.9: Residual distribution for the generated metamodel

Now, in order to have more accurate estimates for effects, all the effects with F-Ratio less than 2 are eliminated. The analysis of variance and fit for the restricted model are shown in Table 3.13 and Table 3.14.

Table 3.13: Summary of fit for the restricted metamodel

Summary of Fit	
RSquare	0.999762
RSquare Adj	0.999656
Root Mean Square Error	665.5421
Mean of Response	79290.21
Observations (or Sum Wgts)	79

Table 3.14: ANOVA for the restricted metamodel

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	24	1.0055e+11	4.1895e+9	9458.331
Error	54	23919103	442946.35	Prob > F
C. Total	78	1.0057e+11		<.0001*

Both the R^2_{Adj} and the model F-Ratio has increased. The increase in R^2_{Adj} is not significant, but the model F-ratio has been almost doubled. The estimates of effects are shown in Table 3.15 and the effect tests for each factor are tabulated in Table 3.16.

Table 3.15: Parameter Estimates for the restricted model

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	77694.757	184.5882	420.91	<.0001*
X1	34541.257	81.92258	421.63	<.0001*
X5	-5266.857	81.92258	-64.29	<.0001*
X7	8995.4255	81.92258	109.80	<.0001*
X9	-5103.579	81.92258	-62.30	<.0001*
X11	-6819.673	81.92258	-83.25	<.0001*
X16	-9775.387	81.92258	-119.3	<.0001*
X29	-4312.412	81.92258	-52.64	<.0001*
X1*X7	3246.6237	83.19277	39.03	<.0001*
X5*X7	-170.4662	83.19277	-2.05	0.0453*
X1*X9	-1787.298	83.19277	-21.48	<.0001*
X7*X9	-473.1196	83.19277	-5.69	<.0001*
X1*X11	-2357.947	83.19277	-28.34	<.0001*
X9*X11	132.21765	83.19277	1.59	0.1178
X1*X16	-3188.866	83.19277	-38.33	<.0001*
X7*X16	-734.4593	83.19277	-8.83	<.0001*
X9*X16	531.34936	83.19277	6.39	<.0001*
X11*X16	771.27978	83.19277	9.27	<.0001*
X1*X29	-1872.782	83.19277	-22.51	<.0001*
X5*X29	283.82892	83.19277	3.41	0.0012*
X7*X29	-489.505	83.19277	-5.88	<.0001*
X9*X29	277.68439	83.19277	3.34	0.0015*
X11*X29	372.13995	83.19277	4.47	<.0001*
X16*X29	531.14008	83.19277	6.38	<.0001*
X16*X16	1909.7029	201.9507	9.46	<.0001*

Table 3.16: Effect Tests for the restricted model

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
X1	1	1	7.8744e+10	177774.3	<.0001*
X5	1	1	1830825372	4133.289	<.0001*
X7	1	1	5340566933	12056.92	<.0001*
X9	1	1	1719070397	3880.990	<.0001*
X11	1	1	3069524133	6929.788	<.0001*
X16	1	1	6306840626	14238.38	<.0001*
X29	1	1	1227395113	2770.979	<.0001*
X1*X7	1	1	674596172	1522.975	<.0001*
X5*X7	1	1	1859759.45	4.1986	0.0453*
X1*X9	1	1	204443877	461.5545	<.0001*
X7*X9	1	1	14325896.1	32.3423	<.0001*
X1*X11	1	1	355834450	803.3353	<.0001*
X9*X11	1	1	1118816.48	2.5259	0.1178
X1*X16	1	1	650807608	1469.270	<.0001*
X7*X16	1	1	34523549	77.9407	<.0001*
X9*X16	1	1	18069257.4	40.7933	<.0001*
X11*X16	1	1	38071840.3	85.9514	<.0001*
X1*X29	1	1	224467991	506.7611	<.0001*
X5*X29	1	1	5155766.64	11.6397	0.0012*
X7*X29	1	1	15335369.8	34.6213	<.0001*
X9*X29	1	1	4934951.74	11.1412	0.0015*
X11*X29	1	1	8863241.24	20.0097	<.0001*
X16*X29	1	1	18055026	40.7612	<.0001*
X16*X16	1	1	39608811.4	89.4212	<.0001*

As mentioned before, CCD assumes a second-order polynomial function for the response function:

$$y = \beta_0 + \sum_{j=1}^{K} \beta_j x_j + \sum_{j<j'} \sum \beta_{j,j'} x_j x_{j'} + \sum_{j=1}^{K} \beta_{jj} x_j^2$$

where, each variable x_j is standardized and lies in [-1,+1]. K is the total number of important factors.

The estimates for parameters for the preceding metamodel are shown in Table 3.17. The intercept (β_0) can be found in Table 3.15.

Table 3.17: Coefficient of the CCD second-order metamodel

Response Surface

Coef

	X1	X5	X7	X9	X11	X16	X29	Y
X1	0	0	3246.6237	-1787.298	-2357.947	-3188.866	-1872.782	34541.257
X5	.	0	-170.4662	0	0	0	283.82892	-5266.857
X7	.	.	0	-473.1196	0	-734.4593	-489.505	8995.4255
X9	.	.	.	0	132.21765	531.34936	277.68439	-5103.579
X11	.	.	.	.	0	771.27978	372.13995	-6819.673
X16	.	.	.	.	.	1909.7029	531.14008	-9775.387
X29	.	.	.	.	.	.	0	-4312.412

3.1.1.5 Verifying the Result of Factor Screening

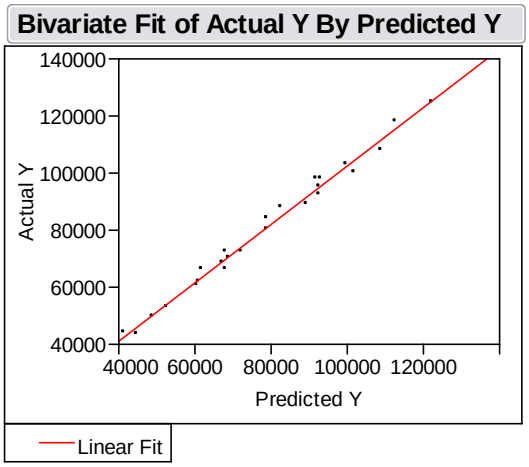
One way to verify that the factors detected as important are truly important is to investigate how well the metamodel generated in Section 3.1.1.4 represents the response function over the experimental region. To examine the metamodel over the entire experimental region, Latin Hypercube Sampling (LHS) is used to sample the region, because this design has a very good space filling property. The actual response, predicted response, and the residual corresponding to the 25 samples are summarized in Table 3.18.

Table 3.18: Actual Y and Predicted Y for the LH design

Run	Actual Y	Predicted Y	Residual	Run	Actual Y	Predicted Y	Residual
1	83704.84	78877.15	4827.68	14	80054.54	78601.54	1453.00
2	43371.03	44620.36	-1249.33	15	117914.22	112402.38	5511.84
3	87827.41	82294.57	5532.84	16	69784.64	68691.96	1092.68
4	68500.53	67193.99	1306.54	17	97714.04	91606.70	6107.34
5	72249.55	72008.55	241.01	18	44076.85	41243.52	2833.33
6	66189.77	61537.24	4652.52	19	99746.05	101527.67	-1781.62
7	72278.91	68081.51	4197.40	20	97846.81	93081.43	4765.38
8	108016.13	108776.57	-760.44	21	102907.90	99778.30	3129.60
9	66004.85	67779.36	-1774.51	22	60654.60	60427.11	227.49
10	61662.48	60965.39	697.08	23	95239.70	92315.00	2924.69
11	92160.59	92433.84	-273.26	24	88636.54	89297.45	-660.91
12	124421.96	122110.74	2311.22	25	49374.06	48907.52	466.54
13	52563.30	52421.58	141.73				

Now, to see how well the Actual Y is estimated by the Predicted Y, we fit the Actual Y to the Predicted Y for the data shown in Table 3.18. The fit of Actual Y by Predicted Y is shown in Figure 3.10.

Figure 3.10: Fit of Actual Y by Predicted Y



Actual Y = -56.1248 + 1.0242 Predicted Y

Ideally, the intercept and the slope of the fitted line should be 0 and 1, respectively. For this metamodel, however, the intercept of -56.1248 is not very critical when compared with the average of the Actual Y's, which is 80116.05. The slope of 1.0242 seems acceptable, too. Therefore, we can conclude that the generated metamodel is a good fit for the Drug model over the entire range of factors.

Another measure for evaluating the goodness of fit is Mean Absolute Relative Error (MARE):

$$MARE = \frac{1}{n}\sum_{i=1}^{n}\frac{|e_i|}{|Y_i|}$$

where e_i is the residual and Y_i is the observed response value for ith observation, and n is the total number of observations. However, since *MARE* always decreases as more regression variables are included in the metamodel, the adjusted *MARE*, defined as follows, is a better criterion for evaluating the goodness of fit:

$$MARE_{Adj} = (MARE)\frac{N-1}{N-q}$$

where, N is the total number of runs and q is the number of regression variables.

For the generated metamodel, $N = 25$ and $q = 7$. The *MARE* and $MARE_{Adj}$ for this metamodel are: 0.02901 and 0.03869, respectively. The small *MARE* represents a good-fit.

3.1.2. Sequential Bifurcation with Interactions (SB-X)

The only parameter that SB-X requires from a user is the delta limit (δ), which is the value that the main effect of a factor should reach to be considered important. As mentioned in Section 3.1.1.3 the client was interested in the factors that were able to change the response function as much as $5000. Thus, $2500 was selected as the delta limit. As discussed earlier, SB-X assumes that the direction for main effects is known. A Plackett-Burman design was used to discover the signs of main effects. This design required 28 runs. Although SB-X assumes a second-order polynomial for the response function, it cannot estimate the second-order interaction and quadratic effects. It, however, can provide unbiased estimates for main effects (β_j). In order to estimate the interaction and quadratic effects, a fractional factorial design of resolution V and central composite design (CCD) is used for the factors identified to be important.

3.1.2.1 Results of SB-X

It is possible that for some combinations of upper and lower levels of factors, the model generates an unrealistic response. Therefore, we showed the distribution plot of observed responses obtained by SB-X, to our client to make sure that all the response values are at a reasonable range. The plot of the response distribution is shown in Figure 3.11

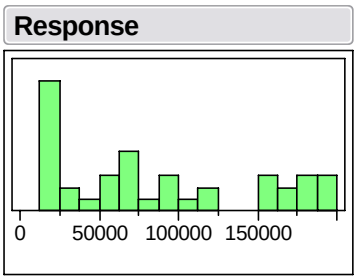


Figure 3.11: Response distribution of Drug Model

SB-X identified 7 factors to be important by making 40 runs. The factor screening results are shown in Table 3.19.

Table 3.19: List of Important Factors for the Drug Model

Screening ID	Effect	Main effect	Expert Guess
1	X1	38,412.39	Important
5	X5	-5,635.76	Important
7	X7	8,411.16	Important
9	X9	-4,952.41	Unknown
11	X11	-6,538.56	Important
16	X16	-10,198.10	Important
29	X29	-5,658.13	Important

3.1.2.2 Generating Metamodel

Since the factors identified as important are exactly the same for SB-X and Fractional Factorial methods, the metamodel generated in Section 3.1.1.4 can be used here, too.

3.1.2.3 Verifying the Result of Factor Screening

A similar procedure to that used in Section 3.1.1.5 should be followed in this section too, but since the generated metamodel for both methods are the same the results will be identical. Therefore, the *MARE* and $MARE_{Adj}$ for this metamodel are: 0.03869 and 0.02901, respectively.

3.1.3. Latin Hypercube Design

As mentioned in Section 2.4, to use LH designs to find the important factors, a metamodel should be fitted to the data observed according to the selected LH design. Sanchez (2005a) has implemented an Excel file containing the Nearly Orthogonal Latin Designs (NOLD) for up to 29 factors. These designs have both good orthogonality and space-filling properties. We used this design for generating the metamodel from which we can determine the important factors.

3.1.3.1 Fitting the Metamodel

For the 25 factors involved in the Drug model, an NOLD requires 256 runs. Before fitting a metamodel to the collected data, we should ensure that the responses for all treatments are at a reasonable range. Therefore, we showed the distribution plot of observed responses obtained by the NOLD to our client to make sure that this requirement is met. The plot of the response distribution is shown in Figure 3.12.

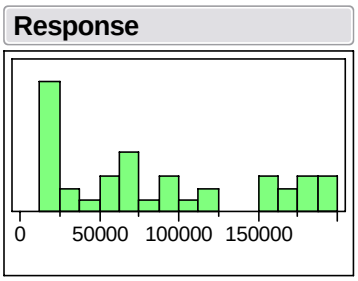


Figure 3.12: Response distribution of the Drug Model

With 256 observations, we can fit a metamodel with at most 256 unknown parameters. Fitting a full second-order metamodel for 25 factors needs 351 runs: 1 run for the intercept, 25 runs for the main effects, 25 runs for the quadratic effects, and 300 runs for second-order interactions. Therefore, we begin by fitting a second-order polynomial - without the interaction terms - to the collected data. The summary of fit and ANOVA for this metamodel are shown in Table 3.20 and Table 3.21.

Table 3.20: Summary of fit for the LHD

Summary of Fit	
RSquare	0.990751
RSquare Adj	0.988506
Root Mean Square Error	2441.057
Mean of Response	80513.4
Observations (or Sum Wgts)	257

Table 3.21: ANOVA for the LHD

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	50	1.3149e+11	2.6298e+9	441.3369
Error	206	1227504490	5958759.7	Prob > F
C. Total	256	1.3272e+11		<.0001*

Then, we declare any factor with both main and quadratic effect's F-ratio less than 2 to be unimportant; because it is very unlikely that a factor is truly important but its main and

quadratic effects are insignificant. Consequently, factors 18, 22, 26, 28, and 34 will be eliminated and the further analyses will be done on the remaining 20 factors.

Next, in order to fit a metamodel, we used the "stepwise fit" feature of JMP. In this approach, at each step we have an option to add as many terms (main, second-order interaction, or quadratic) as necessary to the metamodel. We started by including all the main effects except factors 18, 22, 26, 28, and 34 in the metamodel; and then among all the interaction and quadratic terms, we entered a term only if it increased the R^2_{Adj} of the fit. Next, we need to ensure that all the included terms have F-ratio greater than 2, and if not, the term should be dropped from the model. In addition, all the terms left out should be investigated to ensure that they will not improve the R^2_{Adj} if included in the metamodel.

Analysis of variance and fit for the metamodel generated according to the preceding procedure are shown in Table 3.22 and Table 3.23.

Table 3.22: Summary of fit for the restricted model

Summary of Fit	
RSquare	0.997034
RSquare Adj	0.996469
Root Mean Square Error	1353.05
Mean of Response	80513.4
Observations (or Sum Wgts)	257

Table 3.23: ANOVA for the restricted model

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	41	1.3232e+11	3.2274e+9	1762.910
Error	215	393609887	1830743.7	Prob > F
C. Total	256	1.3272e+11		<.0001*

Both the R^2_{Adj} and the model *p*-value show that the generated metamodel satisfactorily fits the data. $R^2_{Adj} = 0.997$ implies that 99.7% of the variation in the data can be accounted for by the model.

The estimates for the following metamodel parameters are shown in Table 3.24 and the effect tests are tabulated in Table 3.25.

$$y = \beta_0 + \sum_{j=1}^{K} \beta_j x_j + \sum_{j<j'} \sum \beta_{j,j'} x_j x_{j'} + \sum_{j=1}^{K} \beta_{jj} x_j^2$$

In this metamodel, each variable x_j is standardized and lies in [-1,+1]. K is the total number of factors.

Table 3.24: Parameter Estimates for the restricted model

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	80172.131	198.9432	402.99	<.0001*
X1	34864.449	145.6134	239.43	<.0001*
X2	465.72545	145.6134	3.20	0.0016*
X3	-1035.484	145.6134	-7.11	<.0001*
X5	-5376.57	145.614	-36.92	<.0001*
X6	-708.5214	145.6128	-4.87	<.0001*
X7	8499.0135	145.6134	58.37	<.0001*
X8	1421.4935	145.6132	9.76	<.0001*
X9	-5089.421	145.6134	-34.95	<.0001*
X11	-6314.29	145.6157	-43.36	<.0001*
X13	773.66232	145.6151	5.31	<.0001*
X16	-10214.39	145.6142	-70.15	<.0001*
X19	1092.1924	145.6139	7.50	<.0001*
X20	1184.8533	145.6143	8.14	<.0001*
X21	1031.0479	145.6137	7.08	<.0001*
X23	479.1223	145.6152	3.29	0.0012*
X24	-1405.969	145.6151	-9.66	<.0001*
X29	-4734.318	145.6132	-32.51	<.0001*
X30	427.39727	145.6135	2.94	0.0037*
X1*X2	-521.4689	258.1597	-2.02	0.0446*
X1*X7	2722.9529	360.7702	7.55	<.0001*
X1*X8	734.26507	275.9473	2.66	0.0084*
X1*X9	-1331.023	264.3433	-5.04	<.0001*
X1*X11	-2215.342	266.3509	-8.32	<.0001*
X1*X13	727.65015	271.999	2.68	0.0080*
X1*X16	-3188.903	271.2033	-11.76	<.0001*
X1*X19	547.74223	259.8577	2.11	0.0362*
X1*X20	518.94735	291.2034	1.78	0.0761
X1*X23	401.15382	241.0155	1.66	0.0975
X1*X24	-676.8848	292.5005	-2.31	0.0216*
X1*X29	-1789.718	274.7106	-6.51	<.0001*
X3*X6	524.66434	250.1992	2.10	0.0372*
X3*X23	678.27722	375.0764	1.81	0.0719
X7*X16	-711.1261	241.5344	-2.94	0.0036*
X7*X29	-357.5484	242.9194	-1.47	0.1425
X8*X16	-513.5008	273.3815	-1.88	0.0617
X9*X16	532.51364	247.4955	2.15	0.0325*
X16*X24	684.03782	264.6109	2.59	0.0104*
X16*X29	550.70957	253.2914	2.17	0.0308*
X8*X8	-518.3683	311.3772	-1.66	0.0974
X16*X16	1040.6807	303.9734	3.42	0.0007*
X29*X29	492.94926	313.5033	1.57	0.1173

Table 3.25: Effect Tests for the restricted model

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
X1	1	1	1.0495e+11	57327.51	<.0001*
X2	1	1	18727724.6	10.2296	0.0016*
X3	1	1	92578764.6	50.5689	<.0001*
X5	1	1	2495927979	1363.341	<.0001*
X6	1	1	43344506.9	23.6759	<.0001*
X7	1	1	6236796947	3406.701	<.0001*
X8	1	1	174468116	95.2990	<.0001*
X9	1	1	2236463294	1221.615	<.0001*
X11	1	1	3442391309	1880.324	<.0001*
X13	1	1	51679424.1	28.2287	<.0001*
X16	1	1	9008341747	4920.592	<.0001*
X19	1	1	102995965	56.2591	<.0001*
X20	1	1	121212799	66.2096	<.0001*
X21	1	1	91786941.6	50.1364	<.0001*
X23	1	1	19820153	10.8263	0.0012*
X24	1	1	170673548	93.2263	<.0001*
X29	1	1	1935266125	1057.093	<.0001*
X30	1	1	15772039	8.6151	0.0037*
X1*X2	1	1	7469774.81	4.0802	0.0446*
X1*X7	1	1	104290902	56.9664	<.0001*
X1*X8	1	1	12962269.6	7.0803	0.0084*
X1*X9	1	1	46415365.9	25.3533	<.0001*
X1*X11	1	1	126648686	69.1788	<.0001*
X1*X13	1	1	13102019.7	7.1567	0.0080*
X1*X16	1	1	253116379	138.2588	<.0001*
X1*X19	1	1	8134085.91	4.4431	0.0362*
X1*X20	1	1	5814084.61	3.1758	0.0761
X1*X23	1	1	5071769.09	2.7703	0.0975
X1*X24	1	1	9804013.59	5.3552	0.0216*
X1*X29	1	1	77704443.2	42.4442	<.0001*
X3*X6	1	1	8050425.65	4.3974	0.0372*
X3*X23	1	1	5986907.57	3.2702	0.0719
X7*X16	1	1	15869489.9	8.6683	0.0036*
X7*X29	1	1	3966182.59	2.1664	0.1425
X8*X16	1	1	6459095.29	3.5281	0.0617
X9*X16	1	1	8475286.68	4.6294	0.0325*
X16*X24	1	1	12234106.1	6.6826	0.0104*
X16*X29	1	1	8654296.83	4.7272	0.0308*
X8*X8	1	1	5073768.58	2.7714	0.0974
X16*X16	1	1	21458098.6	11.7210	0.0007*
X29*X29	1	1	4526345.17	2.4724	0.1173

3.1.3.2 Finding Important Factors

Table 3.24 displays the estimated effects of the generated metamodel. Recall that an *important* factor is one that will cause a $5000 change in the output. Therefore, if a factor has a main effect greater than $5000/2 = $2500, or a quadratic effect greater than $5000, or is involved in a second-order interaction effect greater that $5000/2 = $2500, it is declared to be important. Table 3.26 lists the important factors.

Table 3.26: List of important factors for the Drug Model

Screening ID	Effect	Main effect	Expert Guess
1	X1	34864.449	Important
5	X5	-5376.57	Important
7	X7	8499.0135	Important
9	X9	-5089.421	Unknown
11	X11	-6314.29	Important
16	X16	-10214.39	Important
29	X29	-4734.318	Important

3.1.3.3 Verifying the Result of Factor Screening

The similar procedure used in Sections 3.1.1.5 and 3.1.2.3 should be followed in this section too, but since the factors identified as to be important by the LH method are exactly the same as those detected by the other two methods, the generated metamodels will be also identical; therefore, the *MARE* and $MARE_{Adj}$ for this metamodel will be: 0.03869 and 0.02901, respectively.

3.2. Comparing Results

To identify important factors, the Fractional Factorial (FF) design generates a metamodel first. For the Drug model, FF needed 64 factorial runs for a resolution IV design, 50 runs for axial points, and 1 run for center points. Therefore, it detected the important factors by making 115 runs. SB-X needs the signs of all main effects before starting its factor screening procedure. Therefore, the Plackett-Burman design with 28 runs was used to determine the main effect signs. Then, SB-X required 40 runs to detect the important factors. LHD requires 256 runs to generate a metamodel by which the important factors can be detected. Unlike SB-X, it does not need any prior information.

The necessary information for comparing these three methods is summarized in Table 3.27.

Table 3.27: Summary of factor screening results

	Fractional Factorial	Sequential Bifurcation	Latin Hypercube
# of Initiation runs	0	28	0
# of runs used for factor screening	115	40	256
# of factors identified to be important	7	7	7
Total # of runs used for factor screening	115	68	256
MARE	0.02901	0.02901	0.02901
$MARE_{Adj}$	0.03869	0.03869	0.03869
Goodness of fit	Actual Y = -56.1248 + 1.0242 Predicted Y	Actual Y = -56.1248 + 1.0242 Predicted Y	Actual Y = -56.1248 + 1.0242 Predicted Y
Mean of Actual Y	80116.05	80116.05	80116.05

Based on *MARE* values, and the intercepts and slopes of "Goodness of fit", we can conclude that all these three methods are accurate and effective for the Drug model; that is, all the (relatively) important factors have been accurately detected. Thus, in terms of effectiveness, all these three methods have the same performance on the Drug model.

In terms of efficiency, SB-X requires the least number of runs, even though we assumed that no prior information exists about the direction of the factors and therefore we had to make 28 additional runs to determine the main effect directions.

3.3. Conclusion

We applied three factor screening methods to a deterministic simulation model. These methods were: Fractional Factorial of Resolution IV, SB-X, and Latin Hypercube Sampling. All the methods identified the same 7 factors as being important. Next, a metamodel was generated for the 7 important factors; 25 observations were obtained according to a Latin Hypercube Designs and the goodness of fit of the generated metamodel was measured by the Mean Average Relative Error (*MARE*) and the adjusted *MARE* ($MARE_{Adj}$). All three methods performed equally in terms of accuracy; however, SB-X required the least number of runs to determine the important factors.

4. Factor Screening on Stochastic Models

This chapter provides the results from several factor screening experiments on an available stochastic model, called Prophy Model. The methods include Fractional Factorial design, Cheng's method, the Modified Cheng's method, and Controlled Sequential Bifurcation (CSB). The methods are evaluated in terms of number of required runs and the goodness of fit of the metamodel built from the factors detected as important.

4.1. Necessary Steps before Factor Screening the Prophy Model

The Prophy Model is a proprietary stochastic simulation model built in TreeAge® (2006) and consists of 76 factors. Among those factors, 54 factors correspond to 27 *distribution-based variables* and the rest are *constant factors*. Distribution-based *variables* are those whose values are determined according to a distribution function whose parameters are referred to as *factors*. For example, in the Prophy model variable-1 follows a standard Beta (0,1) distribution with shape factor-1 and shape factor-2 as distribution parameters. In almost all cases, the factors associated with a common distribution variable should be confounded, i.e. their values cannot change independently. After confounding the necessary factors, the factor screening experiment is performed on 40 factors.

Before performing factor screening on stochastic models, it is usually suggested that the experimental region be explored by using an inexpensive design such as small Latin Hypercube Design (LHD) or a Plackett-Burnman (PB) design. JMP is not able to generate an LHD for more than 25 factors; therefore since the Prophy model consists of 40 factors, LHD could not be used to explore the response over the region. Alternatively, a PB design augmented with a center point could be used. For 40 factors, PB requires 44 runs in addition to one run that is added for the center point. Since the response was believed to have high variance, 20 replications of each run were made. This small experiment can be used for the following analyses:

1. To insure that the treatments are statistically different at this number of replications an ANOVA is conducted on the observed data. If treatments are not statistically different with this number of replications, the number of replications needs to be increased.
2. To insure that the response function has equal variance over the experimental region, when the variances across treatments are not equal, the usual analysis of variance assumptions are not satisfied. Thus the ANOVA F-test is not valid. As discussed previously, most of the factor screening methods designed for stochastic models assume homogeneous variance over the entire experimental region, i.e. equal variances for all treatments. Therefore, prior to performing a factor screening experiment, constant variance over the region must exist. If not, some variance-stabilizing transformation is applied and the experiment is run on the transformed data.

As mentioned previously, a computer code has been implemented for factor screening the simulation models. The complete structure of the code is discussed in Appendix A. For Cheng's method, the Modified Cheng's methods, and CSB-X, which are sequential procedures, the code generates the appropriate design at each Stage, whereas for the 2^k Fractional Factorial and the BP, the user needs to provide the code with the generated designs. We used JMP to produce the 2^k Fractional Factorial and PB designs and then the code can read the designs via an Excel file.

4.1.1. ANOVA on the Observed Data

Initially we chose 20 observations for each treatment. The observed responses for each treatment are plotted in Figure 4.1. The diamonds encompass two-sided 95% confidence intervals for observations corresponding to each treatment. This figure reveals that with a 95% confidence level based on 20 replications for each treatment, many treatments are statistically different.

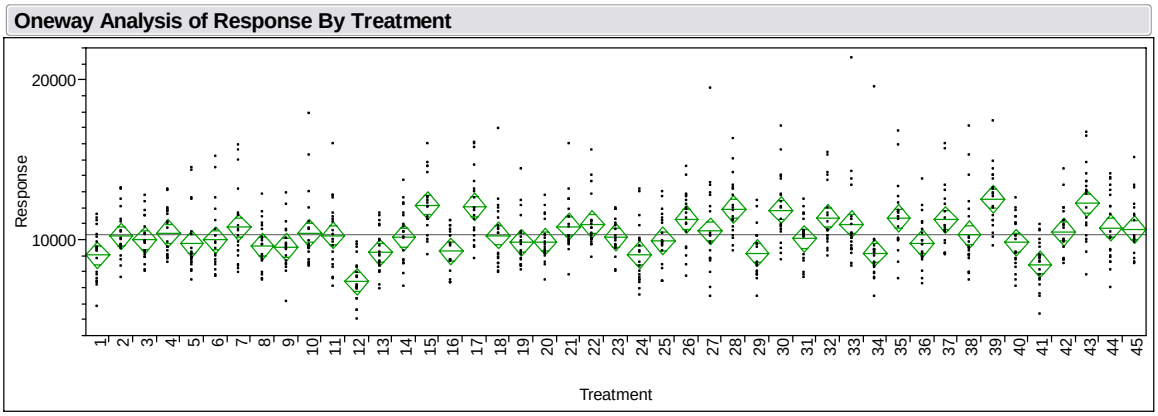


Figure 4.1: Response versus Treatment

The analysis of variance is shown in Table 4.1. The *p*-value is less than 0.0001, confirming that the treatments are statistically different.

Table 4.1: ANOVA for the BP designs

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Treatment	44	978400116	22236366	6.0389	<.0001*
Error	855	3148279050	3682197.7		
C. Total	899	4126679167			

The preceding analyses suggest that 20 replications for each treatment is sufficient for making statistical comparisons between treatments. However, as Figure 4.1 indicates, the response of the Prophy model seems to have a relatively high variance. The mean standard deviation for all treatments is 1872.46, which is quite high compared to the overall mean, 10321.94.

High standard deviation would result in inaccurate estimation of effects and misleading detection of important factors. Two approaches to cope with this problem are to either increase the number of observations or to use batch means. These methods and their influence on our analyses will be addressed in the next sections.

4.1.2. Testing the Equality of Variance

A plot of residual versus fitted values is usually a good visual aid for detecting irregular patterns among data. Nonconstant variance can often be observed on this plot. Sometimes the variance of the observations increases as the magnitude of the observations increase. If this is the case, the residuals get larger as the observed data gets larger, and the plot of residual versus fitted values will look like an outward-opening funnel. Nonconstant variance also arises in cases where the data follow a nonnormal, skewed distribution because in skewed distributions the variance tends to be a function of the mean (Montgomery (2000)).

Although residual plots are frequently used to diagnose inequality of variance, several statistical tests have also been proposed. A formal test for the following hypotheses is:

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_a^2$$
$$H_1 : \text{ above not true for at least one } \sigma_i^2$$

where a is the number of treatments. Montgomery (2000) recommends using "Bartlett's test" when the normality assumption holds and "Modified Levene test" when the normality assumption is violated.

Bartlett's test computes a statistic whose sampling distribution is closely approximated by the chi-square distribution with a - 1 degrees of freedom, when the a random samples are from independent normal populations. The test statistics is:

$$\chi_0^2 = 2.3026 \frac{q}{c}$$

where

$$q = (N-a)\log_{10} S_p^2 - \sum_{i=1}^{a}(n_i - 1)\log_{10} S_i^2$$

$$c = 1 + \frac{1}{3(a-1)}\left(\sum_{i=1}^{a}\frac{1}{n_i - 1} - \frac{1}{N-a}\right)$$

$$S_p^2 = \frac{\sum_{i=1}^{a}(n_i - 1)S_i^2}{N-a}$$

Here a is the number of treatments, n_i is the number of observations for treatment i, N is the total number of observations, and S_i^2 is the sample variance of the ith treatment. The test should reject H_0 when values of χ_0^2 that are too large; that is H_0 is rejected only when

$$\chi_0^2 > \chi_{\alpha,a-1}^2$$

where $\chi_{\alpha,a-1}^2$ is the upper α percentage point of the chi-square distribution with a - 1 degrees of freedom.

Because Bartlett's test is sensitive to the normality assumption, there may be situations where the Modified Levene test is more appropriate. To test the hypothesis of equal variances in all treatments, the Modified Levene test uses the absolute deviation of the observations y_{ij} in each treatment from the treatment median, say $\tilde{y}_i$. These deviations can be denoted by

$$d_{ij} = \left| y_{ij} - \tilde{y}_i \right| \text{ for } \begin{cases} i = 1,2,...,a \\ j = 1,2,...n_i \end{cases}$$

The Modified Levene test then evaluates whether or not the means of these deviations are equal for all treatments. It turns out that if the mean deviations are equal, the variances of the observations in all treatments will be the same. The test statistic for Levene's test is simply the usual ANOVA F statistic for testing equality of means applied to the absolute deviations.

To check for the equality of variance in the Prophy model, a Plackett-Burnman design is augmented with a center point to observe the behavior of the response function over the experimental region.

Bartlett's Test

The Bartlett's test is usually conducted at the significant level 0.01 or 0.001. Anderson et al. (1974) suggests rejecting the variance homogeneity assumption if the test is rejected at the significant level 0.001 and if the homogeneity test is accepted at $\alpha = 0.01$ level, no transformation is required.

The Bartlett's test statistic for the observed data in the BP designs is $\chi_0^2 = 81.3556$. Since $\chi^2_{0.001,44} = 78.7495$, at the confidence level of 0.1%, we reject the null hypothesis and conclude that not all the variances are statistically equal. However, the test statistic is slightly greater than the critical level; therefore, we tend to also rely on the result of Modified Levene test.

Modified Levene Test

The ANOVA for the Modified Levene test is shown in Table 4.2. The large *p*-value implies that there is no significant difference between the treatment variances. Therefore, the Levene test fails to reject the null hypothesis and we conclude that there is not enough evidence to accept variance heterogeneity.

Table 4.2: ANOVA for the Levene test

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Treatment	44	61955257.9	1408074	0.8038	0.8159
Error	855	1497857400	1751880		
C. Total	899	1559812657			

Plot of Residual versus Predicted Data

As previously noted, one way to test variance equality is to plot the residual versus predicted data. To draw this plot, a first-order metamodel for the data collected is fit according to the Plackett-Burnman designs discussed previously in Section 1.1.3.2 . The plot of residual versus predicted data is shown in Figure 4.2. No irregular pattern is observed in this plot and the data variation is almost the same among all treatment.

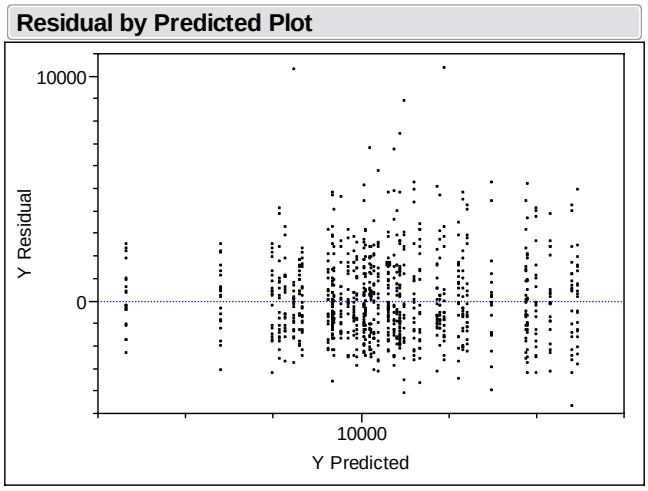


Figure 4.2 : Residuals by Predicted Plot

Homogeneous Variance Decision

The Bartlett's test rejected the homogeneous variance for treatments, but the Levene test and the plot of residual versus predicted did not reject the variance homogeneity. However, for the Prophy model we can not accept the variance homogeneity assumption with certainty, and therefore in future experiments, we'll retest this assumption and take corrective action if necessary.

4.2. Factor Screening the Prophy Model

There are 40 factors in the Prophy model that need to be screened. We applied three different factor screening methods to determine the important factors. Since the number of factors is not excessive and the model executes quickly, a 2^k-Fractional Factorial with resolution IV is a reasonable design to be used as the factor screening method. Cheng's method and Controlled Sequential Bifurcations, which both are based on the Sequential Bifurcation augmented with a fold-over design (SB-X), were used for the Prophy model.

4.2.1. 2^k Fractional Factorial Design

JMP was used to generate a 2^k fractional factorial design of resolution IV. This design required 128 runs for estimating the factors' main effects. A total of 20 observations were obtained for each scenario.

It is possible that for some combinations of upper and lower levels of the factors, the model generates an unrealistic response. Therefore, after making 128×20 = 2560 observations, we showed the distribution plot of observed responses to our client to make sure that all the response values are within a reasonable range. The plot of the response distribution is shown in Figure 4.3.

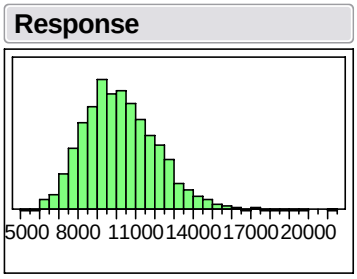


Figure 4.3: Response distribution of Prophy Model

Analysis of variance and fit for the 2^k fractional factorial design are shown in Table 4.3 and Table 4.4. Table 4.5 summarizes the lack of fit analysis.

Table 4.3: Summary of fit for the 2^k - FF design

Summary of Fit	
RSquare	0.200048
RSquare Adj	0.187345
Root Mean Square Error	1764.16
Mean of Response	10259.69
Observations (or Sum Wgts)	2560

Table 4.4: ANOVA for the 2^k - FF design

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	40	1960528627	49013216	15.7484
Error	2519	7839785538	3112261	Prob > F
C. Total	2559	9800314165		<.0001*

Table 4.5: Lack of fit for the 2^k - FF design

Lack Of Fit				
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	87	276650407	3179890	1.0225
Pure Error	2432	7563135131	3109842	Prob > F
Total Error	2519	7839785538		0.4236
				Max RSq
				0.2283

The small R^2_{Adj} implies that only a small amount of variation can be accounted for by the model. However, the large *p*-value for lack of fit (0.4236) indicates that low R^2_{Adj} is not because of bad fit, but it is due to high degree of variation existing in the observed data corresponding to each treatment.

With 20 observations for each treatment, the 95% half-width can be constructed as follows:

$$t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} = t_{0.025,19}\frac{1764.16}{\sqrt{20}} = 825.65$$

It should be noted that the main goal of this experiment is not to estimate the main effects with high accuracy, but to detect the factors with important effects. In spite of that, in order to get more accurate results for the factor screening experiment, we increase the number of observations to 90, or equivalently decrease the half-width to 400.

Analysis of variance and fit for the 2^k fractional factorial design with 90 observations for each treatment are shown in Table 4.6 and Table 4.7

Table 4.6: Summary of fit for the 2^k - FF design

Summary of Fit	
RSquare	0.206529
RSquare Adj	0.203764
Root Mean Square Error	1826.2
Mean of Response	10288.1
Observations (or Sum Wgts)	11520

Table 4.7: ANOVA for the 2^k - FF design

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	40	9964379840	249109496	74.6953
Error	11479	3.8283e+10	3335007.3	Prob > F
C. Total	11519	4.8247e+10		0.0000*

The F-Ratio has increased but the R^2_{Adj} has changed only slightly. One approach to reduce the variation among the observations is to put a number of observations in a batch and treat the batch mean as one observation. For each treatment, a total of 90 observations were obtained, and every 15 observations can be placed in a batch. Now, for each treatment, the 6 batch means are used for fitting the metamodel.

4.2.1.1 Test the equality of variance for the batched data

When batched data are used, there are 6 observations corresponding to each treatment. Before fitting a metamodel to the batched data, the variance of observations should be checked to ensure all the treatments have equal variances.

As previously mentioned in Section 4.1.2, the Bartlett's test is usually conducted at the significant level of 1% or 0.1%. The Bartlett's test statistic for the batched data is $\chi_0^2 = 170.3905$, and $\chi^2_{0.01,44} = 168.1332$ and $\chi^2_{0.001,44} = 183.1864$; at the confidence level of 1%, , χ_0^2 is only slightly greater than $\chi^2_{0.01,44}$ and at the level 0.1%, the homogeneity assumption is not rejected. Therefore, to make an accurate decision, the Modified Leven test is also conducted on the data.

The ANOVA for the Modified Levene test is shown in Table 4.8. The large *p*-value implies that there is not a significant difference between the treatment variances. Therefore, the Modified Levene test fails to reject the null hypothesis and we can conclude that the variance is homogeneous for batched data.

Table 4.8: ANOVA for Levene test

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Treatment	128	10893411	85104.8	0.9433	0.6529
Error	645	58193799	90222.9		
C. Total	773	69087210			

For the 2^k Fractional Factorial design, the plot of residuals versus predicted value is shown in Figure 4.4. This plot reveals no irregularity in the data, leading us to the conclusion that the treatments do not have statistically different variances.

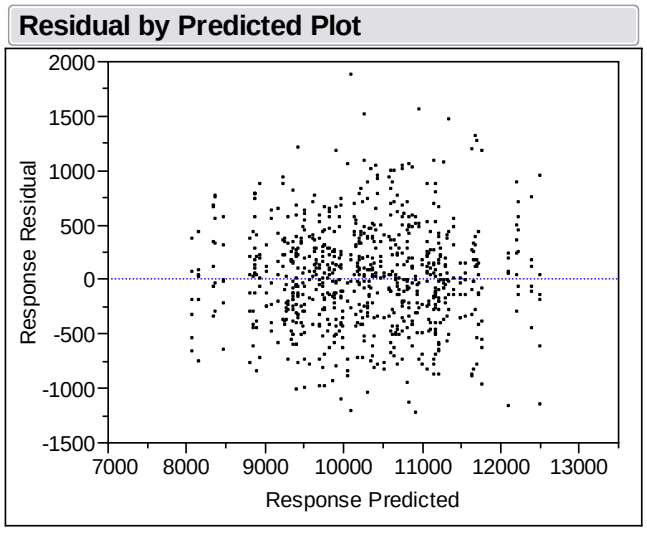


Figure 4.4 : Residuals by Predicted

Figure 4.4 demonstrates that the variation among observations is proportional to the power of the mean of *y*; i.e.

$$\sigma_y \propto \mu^{\alpha}$$

or more specifically for observation *i*:

$$\sigma_{y_i} \propto \mu_i^{\alpha} = \theta\mu_i^{\alpha},$$

where θ is a constant of proportionality. We may take logs to obtain

$$\log\sigma_{y_i} = \log\theta + \alpha\log\mu_i .$$

Therefore, a plot of $\log\sigma_{y_i}$ versus $\log\mu_i$ would be a straight line with slope α. Typically, the standard deviation S_i and the average $\bar{y}_i$ of the ith treatment can be used to estimate σ_{y_i} and μ_i . The plot of $\log\sigma_{y_i}$ versus $\log\mu_i$ is shown in Figure 4.5. This plot reveals that even though the α is estimated to be 0.5109, but the σ_{y_i} fluctuation is significant and it does not increase or decrease regularly as μ_i increases. Therefore, even if a transformation is applied to the data, its effect will not be noticeable, because the main goal of transforming y is to yield a constant variance. When σ_{y_i} changes almost independent of μ_i (which is the case here, since the correlation coefficient between σ_{y_i} and μ_i is estimated to be 0.1156) constant variance cannot be achieved.

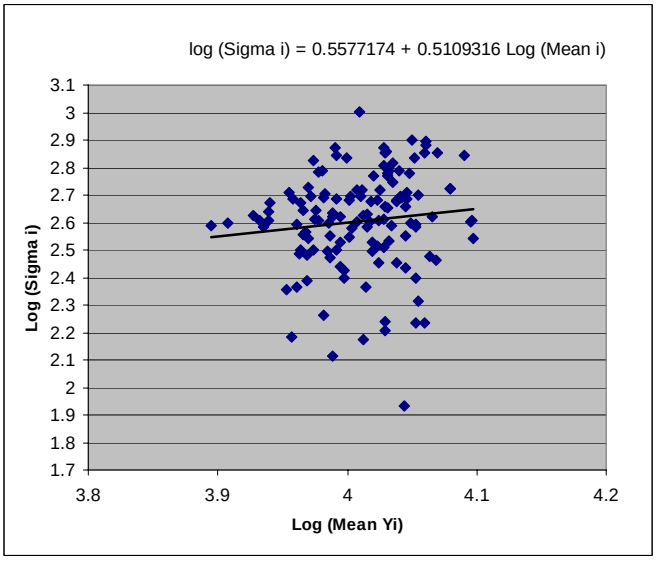


Figure 4.5: plot of Log(Sigma) versus Log(Mean)

4.2.1.2 Analyses for the 2^k Fractional Factorial Design with Batched Means

Analysis of variance and fit for the 2^k fractional factorial design with batched data are shown in Table 4.9 and Table 4.10. Table 4.11 summarizes the lack of fit analysis.

Table 4.9: Summary of fit for the 2^k - FF design for batched data

Summary of Fit	
RSquare	0.804456
RSquare Adj	0.793697
Root Mean Square Error	474.44
Mean of Response	10290.72
Observations (or Sum Wgts)	768

Table 4.10: ANOVA for the 2^k - FF design for batched data

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	40	673215065	16830377	74.7707
Error	727	163642864	225093.35	Prob > F
C. Total	767	836857930		<.0001*

Table 4.11: Lack of fit for the 2^k - FF design for batched data

Lack Of Fit				
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	87	24681895	283700	1.3066
Pure Error	640	138960970	217127	Prob > F
Total Error	727	163642864		0.0400*
				Max RSq
				0.8339

The R^2_{Adj}, and F-Ratio have increased significantly compared with Table 4.3 and Table 4.4, which implies that when the batching strategy is employed, the generated metamodel represents a better fit to the data. Therefore, from now on, we use only the batched data for our analyses.

The normal probability plot for residuals is depicted in Figure 4.6. This plot indicates that, when batched means are used, the normal distribution assumption for residuals seems reasonable and there is no significant evidence for the violation of this assumption.

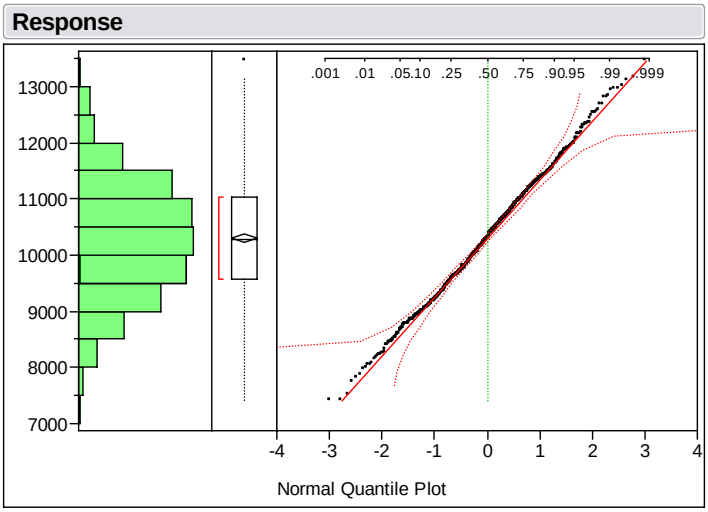


Figure 4.6 : Residual distribution for 2^k - FF design with batched data

4.2.1.3 Adding center point

Before judging the significance of each factor, the existence of nonlinearity effects in the response function should be checked by adding a center point (a point with all factors at their central levels). The analysis of variance, fit, and lack of fit for the new model with three center points added are shown in Table 4.12, Table 4.13 and Table 4.14, respectively.

Table 4.12: Summary of fit for the 2^k - FF design for batched data

Summary of Fit

RSquare	0.802361
RSquare Adj	0.791749
Root Mean Square Error	471.7918
Mean of Response	10289.99
Observations (or Sum Wgts)	786

Table 4.13: ANOVA for the 2^k - FF design for batched data

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	40	673215065	16830377	75.6124
Error	745	165827685	222587.5	Prob > F
C. Total	785	839042750		<.0001*

Table 4.14: Lack of fit for the 2^k - FF design for batched data

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	88	24699714	280679	1.3067
Pure Error	657	141127971	214807	Prob > F
Total Error	745	165827685		0.0390*
				Max RSq
				0.8318

Adding the center point does not show any significant effect on the R^2_{Adj} and the model F-Ratio. In addition, there is a very negligible change in the *p*-value of the lack of fit analysis, which confirms the observation that the model does not present nonlinearities.

4.2.1.4 Finding Important Factors

The client wanted to find the factors that would produce a change in the response function by $300 or more. Therefore, for the Prophy model an *important* factor was defined as a factor whose main effect is $150 or more. Based on this criterion, the list of important factors is shown in Table 4.15.

Table 4.15 : List of important factors for the Prophy model

Screening ID	Effect	Main effect	Expert Guess
6	X6	-150.350	Unknown
9	X9	256.125	Unknown
21	X21	-576.997	Unknown
25	X25	-152.659	Important
32	X32	-270.680	Unknown
37	X37	355.171	Unknown
39	X39	212.579	Unknown
43	X43	-272.534	Unknown

4.2.1.5 Generating the Metamodel

A metamodel can now be constructed using a CCD design for the 8 important factors listed in Table 4.15. This design requires 2^{8-2} factorial runs for a Resolution V design, 2×8 runs for axial designs, and 3 runs for central point; thus a total of 83 runs were needed. For each run, 90 observations were obtained, which were divided into 6 batches, each containing 15 observations. The analysis of variance, fit, and lack of fit for the model are shown in Table 4.16, Table 4.17 and Table 4.18.

Table 4.16: Summary of fit for the CCD

Summary of Fit

RSquare	0.784159
RSquare Adj	0.763194
Root Mean Square Error	453.6651
Mean of Response	10286.58
Observations (or Sum Wgts)	498

Table 4.17: ANOVA for the CCD

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	44	338718810	7698155	37.4038
Error	453	93232855	205812	Prob > F
C. Total	497	431951665		<.0001*

Table 4.18: Lack of fit for the CCD

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	36	10366823	287967	1.4491
Pure Error	417	82866033	198720	Prob > F
Total Error	453	93232855		0.0488*
				Max RSq
				0.8082

The normal probability plot for residuals is shown in Figure 4.7. This plot indicates that the normal distribution assumption for residuals appears reasonable and there is no significant evidence that implies the violation of this assumption. Moreover, Figure 4.8 implies that the assumption of homogeneous variance is also valid for the generated metamodel.

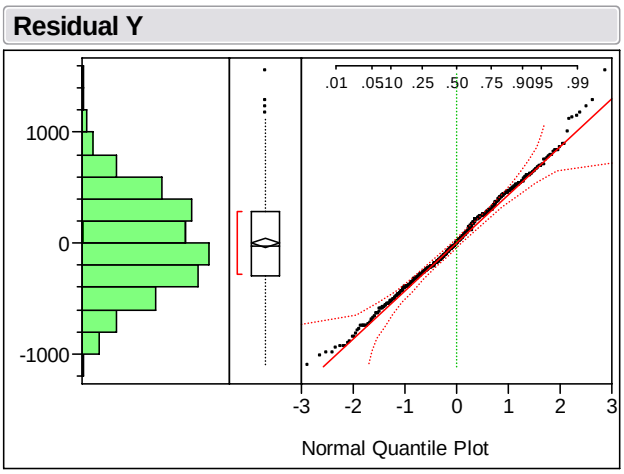


Figure 4.7: Residual distribution for CCD

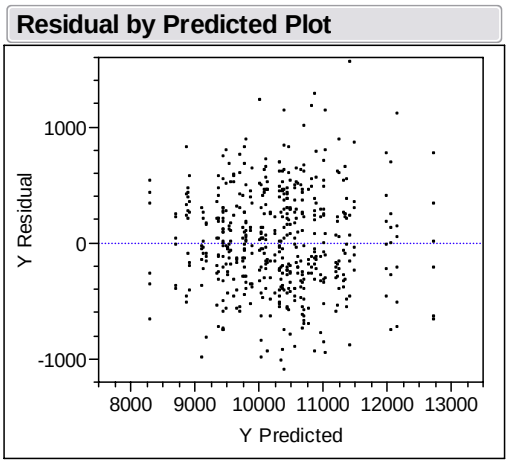


Figure 4.8: Residuals by Predicted Plot

Now, in order to have more accurate estimates for effects, all the effects with F-Ratio less than 2 are eliminated. The analysis of variance and fit for the restricted model are shown in Table 4.19 and Table 4.20.

Table 4.19: Summary of fit for the restricted metamodel

Summary of Fit

RSquare	0.77607
RSquare Adj	0.768621
Root Mean Square Error	448.4372
Mean of Response	10286.58
Observations (or Sum Wgts)	498

Table 4.20: ANOVA for the restricted metamodel

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	16	335224517	20951532	104.1867
Error	481	96727149	201095.94	Prob > F
C. Total	497	431951665		<.0001*

Both the R^2_{Adj} and the model F-Ratio have increased. The increase in R^2_{Adj} is not significant, but the model F-ratio has been almost doubled.

The estimates of effects are shown in Table 4.21 and the effect tests for each factor are tabulated in Table 4.22. The CCD design assumes the following second-order polynomial function for the response function:

$$y = \beta_0 + \sum_{j=1}^{K} \beta_j x_j + \sum_{j<j'} \sum \beta_{j,j'} x_j x_{j'} + \sum_{j=1}^{K} \beta_{jj} x_j^2$$

where, each variable x_j is standardized and lies in [-1,+1]. K is the total number of important factors.

Table 4.21: Parameter Estimates for the restricted model

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	10286.58	20.09495	511.90	0.0000*
X6	-121.4572	22.53482	-5.39	<.0001*
X9	258.50315	22.53482	11.47	<.0001*
X21	-591.3126	22.53482	-26.24	<.0001*
X25	-137.1693	22.53482	-6.09	<.0001*
X32	-325.7138	22.53482	-14.45	<.0001*
X37	341.8811	22.53482	15.17	<.0001*
X39	265.24942	22.53482	11.77	<.0001*
X43	-295.1558	22.53482	-13.10	<.0001*
X9*X25	-48.25613	22.88422	-2.11	0.0355*
X21*X32	35.353818	22.88422	1.54	0.1230
X9*X37	39.312477	22.88422	1.72	0.0865
X25*X37	-45.31833	22.88422	-1.98	0.0482*
X32*X37	-65.05988	22.88422	-2.84	0.0047*
X9*X39	-39.75623	22.88422	-1.74	0.0830
X32*X39	-37.32949	22.88422	-1.63	0.1035
X21*X43	46.422043	22.88422	2.03	0.0431*

Table 4.22: Effect Tests for the restricted model

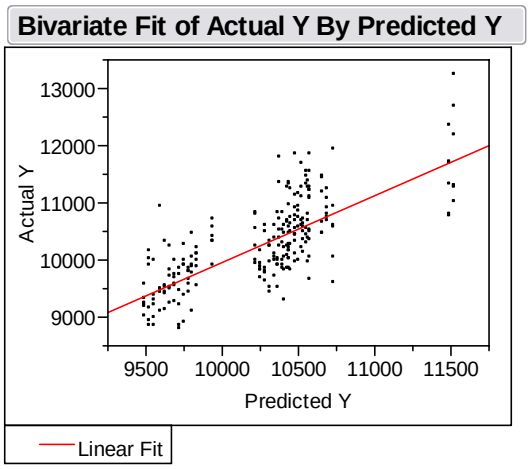
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
X6	1	1	5841734	29.0495	<.0001*
X9	1	1	26462256	131.5902	<.0001*
X21	1	1	138461648	688.5353	<.0001*
X25	1	1	7450908	37.0515	<.0001*
X32	1	1	42011433	208.9124	<.0001*
X37	1	1	46285543	230.1665	<.0001*
X39	1	1	27861473	138.5482	<.0001*
X43	1	1	34498312	171.5515	<.0001*
X9*X25	1	1	894203	4.4466	0.0355*
X21*X32	1	1	479959	2.3867	0.1230
X9*X37	1	1	593461	2.9511	0.0865
X25*X37	1	1	788640	3.9217	0.0482*
X32*X37	1	1	1625390	8.0827	0.0047*
X9*X39	1	1	606934	3.0181	0.0830
X32*X39	1	1	535100	2.6609	0.1035
X21*X43	1	1	827522	4.1151	0.0431*

4.2.1.6 Verifying the Result of Factor Screening

One way to verify that the factors detected as important are truly important is to investigate how well the metamodel generated in Section 4.2.1.5 represents the response function over the experimental region. To examine the metamodel over the entire experimental region, Latin Hypercube Sampling (LHS) is used to sample the region. For the 40 treatments of the LH designs, 6 batched means, each calculated from 15 observations, were obtained. Now, to see how well the Actual Y is estimated by the Predicted Y, the Actual Y is fit to the Predicted Y. The fit of Actual Y by Predicted Y is shown in Figure 4.9.

Figure 4.9: Fit of Actual Y by Predicted Y



Actual Y = -1738.626 + 1.1702 Predicted Y

Ideally, the intercept and the slope of the fitted line should be 0 and 1, respectively, and all the observations should be as close as possible to the fitted line. For this metamodel the intercept of -1738.626 would be of concern because it is noticeably high compared with the average of Actual Y's, which is 10271.016. Moreover, the slope of 1.1702 also seems to be problematic.

For stochastic models, however, a quantitative criterion for evaluating the metamodel's goodness of fit is more preferable. Kleijnen et al. (2000) suggest using the *coefficient of determination* which is defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\sum_{r=1}^{m_i}(\hat{y}_i - y_{i,r})^2}{\sum_{i=1}^{n}\sum_{r=1}^{m_i}(y_{i,r} - \bar{y})^2}$$

where,

$\hat{y}_i$: predicted value for treatment *i*.

$y_{i,r}$: actual response for *r*th observation of treatment *i*.

$\bar{y}$: mean of all observations

m_i: number of observations for treatment *i*.

n: total number of observations for all treatment

R^2 equals one (perfect fit) if all *n* metamodel output equals their corresponding simulation outputs ($\hat{y}_i = y_{i,r}$, for all *i* and *r*). Because R^2 always increase as more regression variable are added (higher *q*), the adjusted R^2 is introduced

$$R^2_{Adj} = 1 - (1 - R^2)\frac{N-1}{N-q}$$

where, *N* is the total number of obtained runs and *q* is the number of regression variables.

For the fitted metamodel, where *q* is 8 and *N* is 240, we obtained $R^2 = 0.5460$ and $R^2_{Adj} = 0.5323$.

4.2.2. Cheng's Method Augmented with Fold-over Design

The only parameter that Cheng's method requires from a user is the delta limit (δ), which is the value that the main effect of a factor should reach to be considered important. As mentioned in Section 4.2.1.4 the client was interested in the factors that were able to change the response function as much as \$300, or equivalently have the main effect of \$150 or more. Thus, \$150 was selected for delta limit. As discussed earlier, SB-X assumes that the direction for main effects is known. A Plackett-Burman design was used to discover the signs of the main effects. This design required 44 runs and as before, for each treatment 6 batched means were obtained, with 15 observations in each batch.

In Section 2.2.4.2 it was shown that a (1-α)% half-width for the estimate of main effect β_i can be calculated by:

$$\frac{1}{2\sqrt{2}} t_{\alpha/2,\omega} S_s \sqrt{(1/r^{(k_s-1)} + 1/r^{(k_s)})} \qquad (4.1)$$

If we assume the same number of observations for each Stage, Equation 4.1 can be approximated by:

$$(1\text{-}\alpha)\% \text{ half-width for } \beta_i\text{: } \frac{1}{2} z_\alpha \frac{S}{\sqrt{n}} \qquad (4.2)$$

Since we did not want to miss an important factor with main effect \$150 greater that the delta limit, we required that the half-width be less than 150; i.e.

$$\frac{1}{2} z_\alpha \frac{S}{\sqrt{n}} < 150 \Rightarrow n > \frac{(z_\alpha S)^2}{300^2}$$

The average of variances for the 44 treatments of the BP design is 1874.01 and therefore with $\alpha = 0.05$, the minimum number of observations for each treatment will be 106. Therefore, to use Cheng's method, for each treatment we obtained 120 observations and then split them into 10 batches, each of which includes 12 observations.

4.2.2.1 Results of the Modified Cheng's Method

The original Cheng's method classifies a group (or a factor) as unimportant if the upper bound of the confidence interval for its main effect is greater than delta limit (δ). For the Prophy model even when 120 observations are obtained for each run, since the standard deviation is dramatically high, the above criterion is always satisfied and no factor is detected as unimportant. Alternatively, the Modified Cheng's method was used for the Prophy model.

This method identified 5 factors as important by making 34 runs. As mentioned before, for each run 120 observations were obtained which were split into 12 batches. The factor screening results are shown in Table 4.23. For both the original and the Modified Cheng's method, the delta limit (δ) and the alpha (α) were set at \$150 and 0.05, respectively.

Table 4.23: List of important factors for the Prophy model

Screening ID	Effect	Main effect	Expert Guess
21	X21	-517.498	Unknown
32	X32	-329.886	Unknown
37	X37	339.302	Unknown
39	X39	237.679	Unknown
43	X43	-385.458	Unknown

It is possible that for some combinations of upper and lower levels of factors, the model generates an unrealistic response. Therefore, we showed the distribution plot of observed responses at all stages to our client to make sure that all the response values are reasonable. The plot of the response distribution is shown in Figure 4.10.

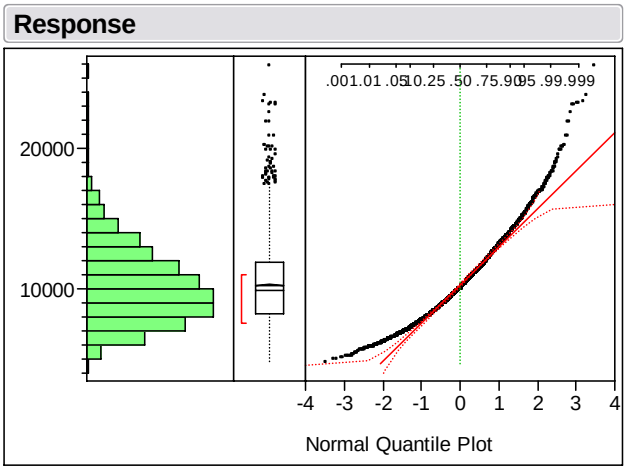


Figure 4.10: Response distribution of Prophy Model

4.2.2.2 Generating a Metamodel

To build a metamodel using the factors detected as important by the Modified Cheng's method, a CCD was used. The created on-face CCD required 2^{5-1} runs for a 2^{5-1}-fractional factorial of resolution V, and 2×5 runs for axial designs, and 3 runs for central point; thus a total of 29 runs were needed. For each run, 90 observations were obtained, which were divided into 6 batches, each contains 15 observations. The analysis of variance, fit, and lack of fit for the model are shown in Table 4.24, Table 4.25 and Table 4.26.

Table 4.24: Summary of fit for the CCD

Summary of Fit

RSquare	0.680072
RSquare Adj	0.638251
Root Mean Square Error	435.9494
Mean of Response	10307.12
Observations (or Sum Wgts)	174

Table 4.25: ANOVA for the CCD

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	20	61811022	3090551	16.2616
Error	153	29077941	190052	Prob > F
C. Total	173	90888963		<.0001*

Table 4.26: Lack of fit for the CCD

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	6	2433888	405648	2.2380
Pure Error	147	26644053	181252	Prob > F
Total Error	153	29077941		0.0427*
				Max RSq
				0.7069

CCD assumes a second-order polynomial function for the response function:

$$y = \beta_0 + \sum_{j=1}^{K} \beta_j x_j + \sum_{j<j'} \sum \beta_{j,j'} x_j x_{j'} + \sum_{j=1}^{K} \beta_{jj} x_j^2$$

where, each variable x_j are standardized and lies in [-1,+1] and K is the total number of important factors. The estimates of effects are shown in Table 4.27.

Table 4.27: Parameter Estimates for the restricted model

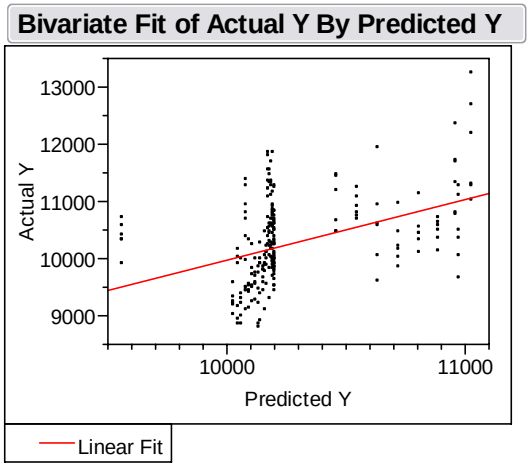
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	10320.614	58.53732	176.31	<.0001*
X21	-451.1752	41.94925	-10.76	<.0001*
X32	-279.8647	41.94925	-6.67	<.0001*
X37	362.66966	41.94925	8.65	<.0001*
X39	206.47766	41.94925	4.92	<.0001*
X43	-294.7945	41.94925	-7.03	<.0001*
X21*X32	-85.92191	44.4939	-1.93	0.0553
X21*X37	-12.59868	44.4939	-0.28	0.7774
X32*X37	-37.97773	44.4939	-0.85	0.3947
X21*X39	27.117666	44.4939	0.61	0.5431
X32*X39	-30.10281	44.4939	-0.68	0.4997
X37*X39	-29.61529	44.4939	-0.67	0.5067
X21*X43	36.55545	44.4939	0.82	0.4126
X32*X43	-36.63309	44.4939	-0.82	0.4116
X37*X43	30.279246	44.4939	0.68	0.4972
X39*X43	-9.328217	44.4939	-0.21	0.8342
X21*X21	323.21785	113.6332	2.84	0.0051*
X32*X32	-15.82731	113.6332	-0.14	0.8894
X37*X37	-110.7738	113.6332	-0.97	0.3312
X39*X39	-87.68628	113.6332	-0.77	0.4415
X43*X43	-130.6781	113.6332	-1.15	0.2519

4.2.2.3 Verifying the Result of Factor Screening

Again we used Latin Hypercube Sampling (LHS) to investigate the goodness of fit of the generated metamodel. The fit of Actual Y by Predicted Y is shown in Figure 4.11.

Figure 4.11: Fit of Actual Y by Predicted Y



Actual Y = -618.826 + 1.0594 Predicted Y

Figure 4.11 reveals that the generated metamodel is barely able to account for the variation existing in the obtained observations. The values of the *coefficient of determination* (R^2) and the *adjusted* R^2 (R^2_{Adj}), which were introduced in Section 4.2.1.6, substantiates this claim; for the fitted metamodel, where q is 5 and N is 240, $R^2 = 0.1838$ and $R^2_{Adj} = 0.16996$, which both are too low. Therefore, in order to have a more accurate metamodel we will decrease the delta limit to \$100 allowing the procedure to declare more factors important.

4.2.3. Modified Cheng's Method with Smaller Delta Limit

The Modified Cheng's method was applied again on the Prophy model with delta limit (δ) and alpha (α) set at \$100 and 0.05, respectively. The method identified 8 factors as important by making 40 runs. As mentioned before, for each run 120 observations were obtained, which were split into 12 batches. The factor screening results are shown in Table 4.28.

Table 4.28: List of important factors for the Prophy model

Screening ID	Effect	Main effect	Expert Guess
8	X8	-191.8425	Unknown
21	X21	-732.6497	Unknown
24	X24	-159.1982	Unknown
25	X25	-240.658	Important
32	X32	-178.6335	Unknown
37	X37	395.9072	Unknown
39	X39	178.9597	Unknown
43	X43	-236.7192	Unknown

4.2.3.1 Generating Metamodel

As previously done, to build a metamodel using the factors detected as important by the Modified Cheng's method, a CCD was used. The created on-face CCD required 2^{8-2} runs for a 2^{8-2}-fractional factorial of resolution V, and 2×8 runs for axial designs, and 3 runs for central point; thus a total of 83 runs were needed. For each run, 90 observations were obtained, which were divided into 6 batches, each contains 15 observations. The analysis of

variance, fit, and lack of fit for the model are shown in Table 4.29, Table 4.30 and Table 4.31.

Table 4.29: Summary of fit for the CCD

Summary of Fit

RSquare	0.751334
RSquare Adj	0.727181
Root Mean Square Error	451.53
Mean of Response	10333.74
Observations (or Sum Wgts)	498

Table 4.30: ANOVA for the CCD

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	44	279054005	6342136	31.1073
Error	453	92357341	203879	Prob > F
C. Total	497	371411346		<.0001*

Table 4.31: Lack of fit for the CCD

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	36	6149662	170824	0.8263
Pure Error	417	86207679	206733	Prob > F
Total Error	453	92357341		0.7535
				Max RSq
				0.7679

The normal probability plot for residuals is shown in Figure 4.12. This plot indicates that the normal distribution assumption for residuals appears reasonable and there is no significant evidence that implies the violation of this assumption. Moreover, Figure 4.13 implies that the assumption of homogeneous variance is also valid for the generated metamodel.

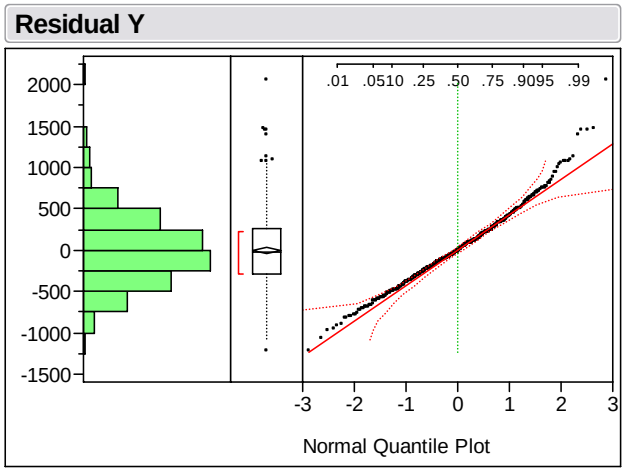


Figure 4.12: Residual distribution

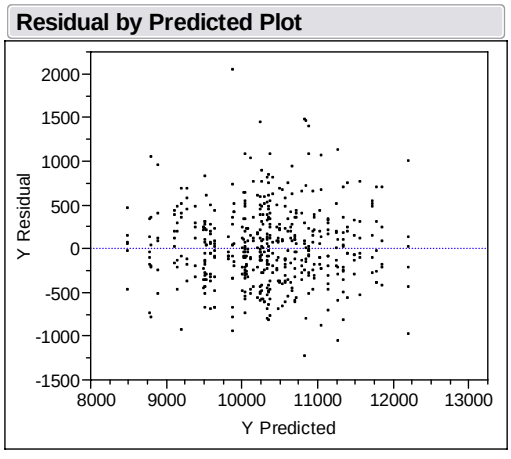


Figure 4.13: Residuals by Predicted Plot

For the generated metamodel, a large number of effects have F-Ratios less than 2; therefore, applying Rule of 2 would result in too much *bias* in the parameter estimations. On the other hand, dropping the factors with small F-Ratio would result in more accurate estimates for the metamodel parameters. We, therefore, dropped the effects with F-Ratio close to zero (i.e. less than 0.4). The parameter estimates for the restricted metamodel are shown in Table 4.32.

Table 4.32: Parameter Estimates for the restricted model

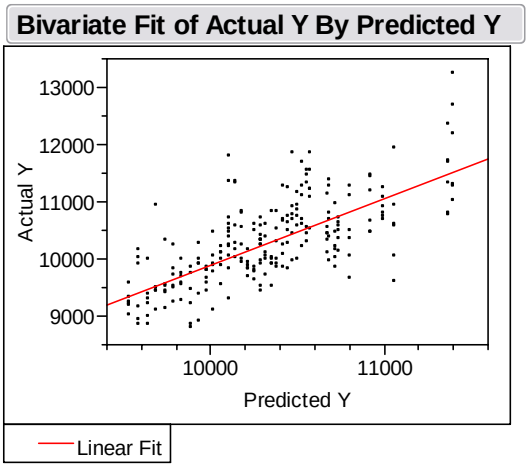
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	10352.76	46.50979	222.59	0.0000*
X8	-27.10869	22.43735	-1.21	0.2276
X21	-578.239	22.43735	-25.77	<.0001*
X24	-47.41162	22.43735	-2.11	0.0351*
X25	-184.5671	22.43735	-8.23	<.0001*
X32	-293.5304	22.43735	-13.08	<.0001*
X37	348.26082	22.43735	15.52	<.0001*
X39	208.36859	22.43735	9.29	<.0001*
X43	-245.0957	22.43735	-10.92	<.0001*
X8*X21	32.437016	22.78524	1.42	0.1552
X8*X25	-16.51391	22.78524	-0.72	0.4690
X21*X25	18.457188	22.78524	0.81	0.4183
X24*X25	29.156129	22.78524	1.28	0.2013
X8*X32	-16.20413	22.78524	-0.71	0.4773
X21*X32	-48.27628	22.78524	-2.12	0.0346*
X8*X37	18.296469	22.78524	0.80	0.4224
X32*X37	-31.22004	22.78524	-1.37	0.1713
X21*X39	-39.44733	22.78524	-1.73	0.0841
X24*X39	-18.88629	22.78524	-0.83	0.4076
X25*X39	27.09384	22.78524	1.19	0.2350
X32*X39	16.321285	22.78524	0.72	0.4742
X37*X39	-17.44857	22.78524	-0.77	0.4442
X21*X43	-53.29804	22.78524	-2.34	0.0198*
X24*X43	-20.05805	22.78524	-0.88	0.3791
X25*X43	24.131357	22.78524	1.06	0.2901
X32*X43	23.05709	22.78524	1.01	0.3121
X37*X43	-45.63904	22.78524	-2.00	0.0458*
X39*X43	-47.73658	22.78524	-2.10	0.0367*
X21*X21	92.547186	115.7678	0.80	0.4245
X24*X24	-200.982	115.7678	-1.74	0.0832
X25*X25	167.53083	115.7678	1.45	0.1485
X32*X32	-178.0756	115.7678	-1.54	0.1247
X39*X39	95.062405	115.7678	0.82	0.4120

4.2.3.2 Verifying the Result of Factor Screening

Again we used Latin Hypercube Sampling (LHS) to investigate the goodness of fit of the generated metamodel. The fit of Actual Y by Predicted Y is shown in Figure 4.14.

Figure 4.14: Fit of Actual Y by Predicted Y



Actual Y = -1681.946 + 1.1572 Predicted Y

For the fitted metamodel, where q is 8 and N is 40, the *coefficient of determination* (R^2) and the *adjusted* R^2 (R^2_{Adj}) are $R^2 = 0.49199$ and $R^2_{Adj} = 0.47667$. Comparing Figure 4.11 and Figure 4.14, as well as the values of R^2 and R^2_{Adj} in this Section and Section 4.2.2.3 confirms that reducing the Delta limit has led the Modified Cheng's method to produce a more accurate and reliable result.

4.2.4. Controlled Sequential Bifurcation with Interactions (CSB-X)

The Controlled Sequential Bifurcation with fold-over design (CSB-X) was applied on the Prophy model with the following parameters:

$N_0 = 5$

Alpha (α): 0.2

Gamma (γ): 0.8

Delta zero (Δ_0): 100

Delta zero (Δ_1): 200

For each treatment, 10 observations were made and placed in a single batch. It should be noted that we first ran the method with alpha 0.1, but due to high response variance, for most

of the treatments, the CSB-X required more than 1000 observations; therefore, we set alpha at 0.2 to make the procedure less demanding in terms of number of required runs.

The method identified 9 factors as important by making 125 runs, and equivalently 1802 batched means. The factor screening results are shown in Table 4.33.

Table 4.33: List of important factors for the Prophy model

Screening ID	Effect	Main effect	Expert Guess
6	X6	-274.1726	Unknown
9	X9	229.9934	Unknown
21	X21	-754.4293	Unknown
25	X25	-176.5533	Unknown
32	X32	-223.5743	Unknown
37	X37	415.7047	Unknown
43	X43	-354.549	Unknown
55	X55	-168.9961	Unknown
65	X65	218.452	Unknown

4.2.4.1 Generating a Metamodel

As was done before, to build a metamodel using the factors detected as important by the CSB-X method, a CCD was used. The created on-face CCD required 2^{9-2} runs for a 2^{9-2}-fractional factorial of resolution V, and 2×9 runs for axial designs, and 3 runs for central point; thus a total of 149 runs were needed. For each run, 90 observations were obtained, which were divided into 6 batches, each contains 15 observations. The analysis of variance, fit, and lack of fit for the model are shown in Table 4.34, Table 4.35 and

Table 4.36.

Table 4.34: Summary of fit for the generated metamodel

Summary of Fit	
RSquare	0.754109
RSquare Adj	0.738283
Root Mean Square Error	482.8478
Mean of Response	10291.03
Observations (or Sum Wgts)	894

Table 4.35: ANOVA for the generated metamodel

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	54	599893759	11109144	47.6497
Error	839	195606111	233141.97	Prob > F
C. Total	893	795499870		<.0001*

Table 4.36: Lack of fit for the CCD

Lack Of Fit				
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	92	24108994	262054	1.1414
Pure Error	747	171497117	229581	Prob > F
Total Error	839	195606111		0.1838
				Max RSq
				0.7844

The normal probability plot for residuals is shown in Figure 4.15. This plot indicates that the normal distribution assumption for residuals appears reasonable and there is no significant evidence that implies the violation of this assumption. Moreover, Figure 4.16 implies that the assumption of homogeneous variance is also valid for the generated metamodel.

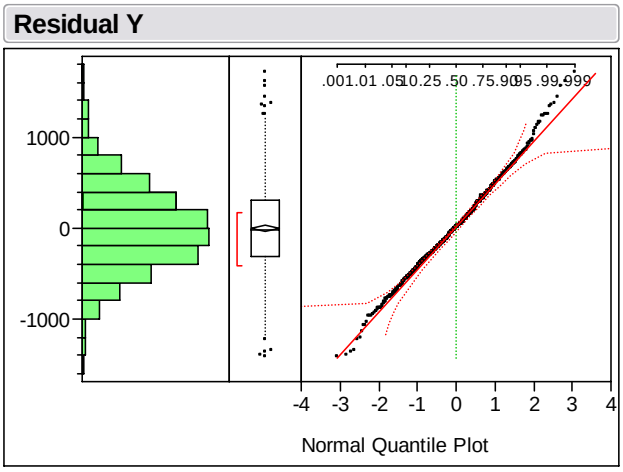


Figure 4.15: Residual distribution

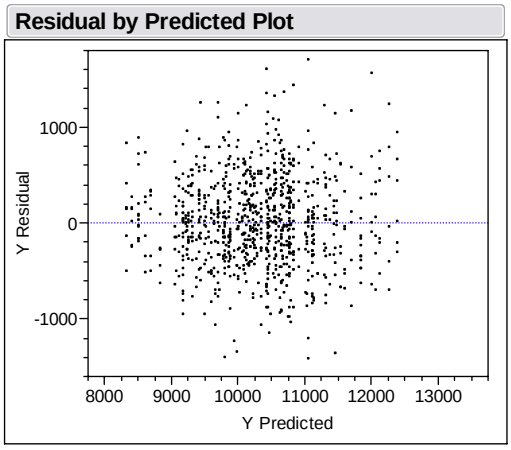


Figure 4.16: Residuals by Predicted Plot

For the generated metamodel, a large number of effects have F-Ratios less than 2; therefore, applying Rule of 2 would result in too much *bias* in the parameter estimations. On the other

hand, dropping the factors with small F-Ratio would result in more accurate estimates for the metamodel parameters. We, therefore, dropped the effects with F-Ratio close to zero (i.e. less than 0.4). The parameter estimates for the restricted metamodel are shown in Table 4.37.

Table 4.37: Parameter Estimates for the restricted model

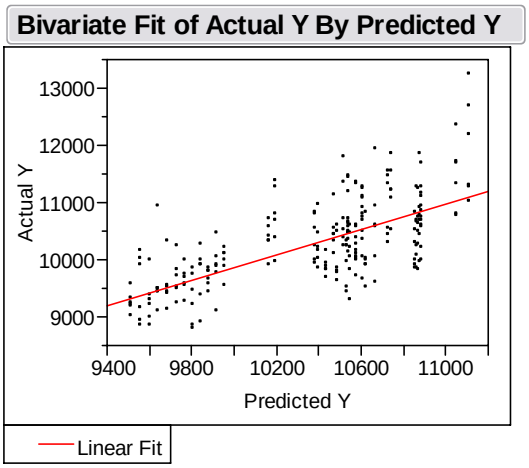
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	10458.941	46.75247	223.71	0.0000*
X6	-118.1683	17.09458	-6.91	<.0001*
X9	282.54309	17.09458	16.53	<.0001*
X21	-572.4173	17.09458	-33.49	<.0001*
X25	-149.7358	17.09458	-8.76	<.0001*
X32	-283.5212	17.09458	-16.59	<.0001*
X37	324.68864	17.09458	18.99	<.0001*
X43	-242.4545	17.09458	-14.18	<.0001*
X55	-164.5489	17.09458	-9.63	<.0001*
X65	150.01449	17.09458	8.78	<.0001*
X6*X9	20.50843	17.22762	1.19	0.2342
X6*X21	-20.62032	17.22762	-1.20	0.2317
X21*X25	-14.9209	17.22762	-0.87	0.3867
X9*X32	-37.22404	17.22762	-2.16	0.0310*
X6*X37	-16.15516	17.22762	-0.94	0.3486
X9*X37	56.492328	17.22762	3.28	0.0011*
X25*X37	-42.73096	17.22762	-2.48	0.0133*
X32*X37	-32.30425	17.22762	-1.88	0.0611
X6*X43	45.185836	17.22762	2.62	0.0089*
X21*X43	38.622501	17.22762	2.24	0.0252*
X25*X43	28.930459	17.22762	1.68	0.0935
X6*X55	-41.90435	17.22762	-2.43	0.0152*
X25*X55	-24.48767	17.22762	-1.42	0.1556
X43*X55	-69.4505	17.22762	-4.03	<.0001*
X6*X65	19.084622	17.22762	1.11	0.2683
X37*X65	15.045846	17.22762	0.87	0.3827
X43*X65	59.851868	17.22762	3.47	0.0005*
X55*X65	28.017094	17.22762	1.63	0.1043
X6*X6	-114.3567	123.6816	-0.92	0.3554
X21*X21	98.854557	123.6816	0.80	0.4244
X37*X37	-142.7883	123.6816	-1.15	0.2486
X43*X43	-144.0156	123.6816	-1.16	0.2446
X65*X65	109.84837	123.6816	0.89	0.3747

4.2.4.2 Verifying the Result of Factor Screening

Again we used Latin Hypercube Sampling (LHS) to investigate the goodness of fit of the generated metamodel. The fit of Actual Y by Predicted Y is shown in Figure 4.17.

Figure 4.17: Fit of Actual Y by Predicted Y



Actual Y = -1273.59 + 1.1137 Predicted Y

For the fitted metamodel, where q is 9 and N is 40, the *coefficient of determination* (R^2) and the *adjusted* R^2 (R^2_{Adj}) are $R^2 = 0.44747$ and $R^2_{Adj} = 0.42833$.

4.3. Comparing Results

To identify important factors, the Fractional Factorial (FF) design must generate a metamodel first. For the Prophy model, for a Resolution IV design, the FF design needed 128 factorial runs and for each run, 90 observations were obtained and grouped into 6 batches.

As mentioned previously, due to the high response variance, Cheng's method failed to detect important factors for the Prophy model; therefore, the Modified Cheng's method was applied. Since the Modified Cheng's method needs the signs of all main effects to be known, a Plackett-Burman (PB) design with 44 runs was used to determine the main effect signs. For each treatment of the PB design, 90 observations were obtained and divided into 6 batches. Then, the Modified Cheng's method required 40 runs, each with 120 observations grouped into 12 batches, to detect the important factors.

Similar to the Cheng's method, the CSB-X requires that the main effect's signs be known. Therefore, the BP design generated for Cheng's method was also used for this method. Then,

having defined 125 scenarios, the CSB-X determined the important factors with 1802 batched means or equivalently 18020 observations.

The necessary information for comparing these three methods is summarized in Table 4.38. Cost-effectiveness shown in the last row is the ratio of the total number of required observations to the Adjusted R^2.

$$\text{Cost} - \text{effectiveness} = \frac{\text{Adjusted R}^2}{\text{Total \# of required observations}}$$

Table 4.38: Summary of the factor screening on the Prophy model

	Fractional Factorial	Modified Cheng's Method	CSB-X
Parameters	-	$\alpha = 0.05$, $\delta = 100$	$\alpha = 0.2$, $\gamma = 0.8$, $\Delta_0 = 100$, $\Delta_1 = 200$
# of initiation observations	0	3960	3960
# of observations used for factor screening	11520	4080	18020
# of factors identified to be important	8	8	9
Total # of observations used for factor screening	11520	8040	21980
Coefficient of determination (R^2)	0.5460	0.49199	0.44747
Adjusted R^2	0.5323	0.47667	0.42833
Goodness of fit	Actual Y = -1738.626 + 1.1702 Predicted Y	Actual Y = -1681.946 + 1.1572 Predicted Y	Actual Y = -1273.59 + 1.1137 Predicted Y
Mean of Actual Y	10271.02	10271.02	10271.02
Cost-effectiveness (×10000)	0.462	0.593	0.1949

Based on the values of coefficient of determination (R^2) and adjusted R^2, the Fractional Factorial (FF) design produced the most accurate results; in other words, for the Prophy model, the FF design is the most effective method among the methods studied in this chapter. In terms of efficiency, the Modified Cheng's method requires the least number of observations, even though we assumed that no prior information exists about the direction of the factors and therefore we had to make additional 3960 observations to determine the main effect directions. However, it should be noted that FF design could have been used with a fewer number of observations for each treatment. Because, although reducing the number of observations decrease the accuracy of the estimated parameter, the important factor can still be detected with reasonable degree of accuracy, if a *sufficient* number of observations are

obtained. Yet, determining the *sufficient* number of observations to maintain the desired accuracy is itself a challenging issue.

From a cost-effectiveness perspective, it turned out that for the Prophy model, the Modified Cheng's method can detect the truly important factors (effectiveness) with the least number of required observations (cost).

4.4. Conclusion

We applied three factor screening methods to a stochastic simulation model. These methods were: Fractional Factorial of Resolution IV, Modified Cheng's method, and CSB-X. These methods identified different factors as being important. For each method, a metamodel was generated for the identified important factors. For verifying the results of each method, a Latin Hypercube Design was used. The goodness of fit for the generated metamodels was measure by the coefficient of determination (R^2) and Adjusted R^2. For the Prophy model, the Modified Cheng's method required the least and the CSB-X required the most number of observations. Moreover, the FF design produced the most and CSB-X produced the least accurate results. And finally, in terms of cost-effectiveness, the Modified Cheng's method outperformed the other two approaches.

5. Conclusion and Recommendation

Computer simulation models that represent a real-world system generally consist of a large number of factors. Finding the factors with significant effects on a selected response has always been of great interest. Often, the analysts use an inefficient and ineffective one-factor-at-a-time design to check the importance of a factor or estimating the main effect of the factors. On the other hand, *screening experiments*, which are more efficient and effective experiments, examine many factors to identify those factors (if any) that have significant effects on the selected response.

It should be noted that the factor screening experiment is a back and forth procedure; i.e. some assumptions are made about the response surface, an appropriate screening methodology is applied and then the validity of the assumptions needs to be checked. If the assumptions are proved to be valid, the factor screening result is reliable; otherwise, the assumptions should be revised and the methodology should be augmented to incorporate the new assumptions. As a result, the expert prior knowledge about the response surface and the importance of the factors has remarkable effect on the efficiency and effectiveness of the screening experiment.

For the available deterministic model (the Drug model), three methods were applied: 2^k Fractional Factorial (FF) Design, Nearly Orthogonal Latin Hypercube (NOLH) Design, and folded-over Sequential Bifurcation (SB-X). For this model, the research showed that all the three methods have equal accuracy (effectiveness), but the SB-X was the most efficient in terms of required number of runs.

In general, for the deterministic models, the following recommendations are usually made:

1. For models with moderate number of factors (less than 29) and very little knowledge about the response surface, NOLH design is usually recommended. These designs are more expensive than the other two methods but have better space-filling property. Moreover, since these designs allow a stepwise fit to the obtained data, no specific assumption about the response surface is required. Therefore, the analysts can begin with any kind of response function (usually first-order polynomial) and then, if necessary, revise their assumption.
2. For models with a large number of factors and little prior knowledge about the response surface, 2^k FF designs of resolution IV are recommended. Because, while maintaining good efficiency, they can easily be augmented to incorporate quadratic effects into the metamodel by using Central Composite Designs (CCD).
3. For models whose response can be approximated by a first-order polynomial augmented with second-order interaction terms, SB-X is superior. However, it requires known directions for the main effects. This assumption can be satisfied by the expert judgment or using a Plackett-Burnman design.

For the available stochastic model (the Prophy model), the following methods were applied: 2^k Fractional Factorial (FF) Design, folded-over Cheng's method and the Modified Cheng's method, and folded-over Controlled Sequential Bifurcation (CSB-X). For this model, due to high variance, Cheng's method failed to operate; the FF design turned out to be the most accurate method; and the Modified Cheng's method was the most efficient method in terms of the required number of runs. From a cost-effectiveness perspective, it turned out that the folded-over Modified Cheng's method (MCh-X) can detect the truly important factors (effectiveness) with the least number of required observations (cost). In general, from a cost-effectiveness perspective, MCh-X is believed to have superior performance to FF designs and CSB-X for situations where the response has homogeneous high variance.

For screening the stochastic models, the following recommendations should be considered:

1. As the response variance gets higher, the accuracy of FF designs, Cheng's method, and MCh-X deteriorates, but CSB-X can still maintain a good level of accuracy. On the other hand, the high variance results in remarkable impairment in the efficiency of Cheng's method, MCh-X and particularly the CSB-X method; however, for these situations, NOLH and 2^k FF designs can still classify the factors correctly while requiring a moderate number of observations.
2. If Common Random Numbers (CRN) are employed, Cheng's and the MCh-X method may lead to misleading results; while CSB-X, LHD and 2^k FF can still be used.
3. With the same reasoning as provided for the deterministic models, for stochastic models with moderate number of factors (less than 29) and very little knowledge about the response surface, NOLH design is usually preferred.
4. For stochastic models with a large number of factors and little prior knowledge about the response surface, 2^k FF designs of resolution IV are recommended.
5. For models whose response can be approximated by a first-order polynomial augmented with second-order interaction terms, MCh-X is superior in terms of then required number of runs. CSB-X is also considered being more efficient than 2^k FF and NOLH designs for cases where the response variance is not too high.

The contributions of this research are as follows:

1. Cheng's method was modified to make it work for the simulation models where the response variance is high. The new method, called the Modified Cheng's method (MCh-X), was shown to be the most cost-effective method among the other stochastic methods studied in this research.
2. A procedure was proposed for measuring the effectiveness (accuracy) of a factor screening method.
3. A new criterion, called cost-effectiveness, was introduced that encompasses both concepts of efficiency and effeteness.
4. The performance of some well-known factor screening methods was evaluated in terms of efficiency, effectiveness, and cost-effectiveness.

5. A factor screening code has been implemented, called NCSU-FSC, which is able to do factor screening on a simulation model based on the methods addressed in this research. In addition, so far we have able to incorporate this class into the simulation models built in Excel, TreeAge, and Microsoft Visual Studio .NET.

Recommendation for Future Study

One possible area for future research is to conduct the factor screening experiment on more than one response. In selecting the *best strategy* (sometimes called *system* or *scenario*) among a group of strategies that share a common set of input factors, say $I = \{x_1,...,x_K\}$, it is of great interest to detect those factors with significant influence on the decision about selecting the best system. That is, to find factors that a change in their values results in a different selection as the *best strategy*. The problem of "Factor Screening between Strategies" is aimed at detecting the factors that have the most important effect in the superiority of one strategy over the other or vice versa.

As mentioned in Section 2.5.1.6 some treatments may result in unrealistic response values and therefore they may cause the result of factor screening to be unreasonable or not reliable. In those situations, since the users do not exactly know which factor(s) have caused the disturbance, they have to narrow the range of some *intuitively* selected factors to eliminate the outliers. Therefore, it is of great interest to generate a methodology which is able to detect the disturbing factor(s) only based on the obtained responses and their corresponding design points.

Conducting sensitivity analysis on the parameters of the sequential factor screening methods is another interesting area for future research. SB-X has only one parameter which is delta limit (δ). Cheng's method and the Modified Cheng's method have three parameters: Type I error (α), delta limit (δ), and number of observations for each design point (n). CSB-X needs 5 parameters: initial number of observations (N_0), Type I error (α), Type II error (γ), Delta zero (Δ_0), and Delta zero (Δ_1). It appeared that Cheng's method, Modified Cheng's method,

and CSB-X are not very sensitive to the Type I error (α); this conjecture, however, should be investigated more precisely.

Finally, at the outset of this research, medical decision making models were the target applications. It was thought that these models may have special characteristics that could be exploited to improve factor screening. However, this conjecture could not be founded and further work may again attempt to extract those characteristics.

References

1. JMP User Guide, Release 6, 2005. SAS Institute Inc.

2. TreeAge Pro 2006 User's Manual, 2006. TreeAge Software, Inc.

3. Anderson, Virgil L. and Robert A. McLean. 1974. *Design of Experiments - A Realistic Approach*.: Marcel Dekker, Inc.

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

5. Cheng, Russell. C. H. 1997. Searching For Important Factors: Sequential Bifurcation under Uncertainty. In *Proceedings of the 1997 Winter Simulation Conference*, 275-280.

6. Cioppa, Thomas M. 2005. Experimental Designs for High-Dimensional Complex Models, United State Army Training and Doctrine Command Analysis Center.

7. Ivanova, Theodora, Linda C. Malone, and Mansooreh Mollaghasemi. 1998. Comparison of a Two-Stage Group-Screening Design to a Standard 2^{k-p} Design for a Whole-Line Semiconductor Manufacturing Simulation Model. In *Proceedings of the 1999 Winter Simulation Conference*, 640-646.

8. Jacobson, Sheldon H., Arnold H. Buss, and Lee W. Schruben. 1991. Driving Frequency Selection for Frequency Domain Simulation Experiments. *Operations Research* 39, (6): 917-924.

9. Kleijnen, Jack. P. C. 1975a. Screening Designs for Poly-factor Experimentation. *Technometrics* 17, (4): 487-493.

10. Kleijnen, Jack. P. C. 1975b. *Statistical Techniques in Simulation, Part II*.: Marcel Dekker.

11. Kleijnen, Jack. P. C. 1987. *Statistical Tools for Simulation Practitioners*.: Marcel Dekker.

12. Kleijnen, Jack. P. C., Bert W. M. Bettonvil, and F. Persson. 2003a. Finding the Important Factors in Large Discrete-Event Simulation: Sequential Bifurcation and Its Applications, Tilburg University Press.

13. Kleijnen, Jack. P. C., Bert W. M. Bettonvil, and F. Persson. 2003b. Robust Solutions for Supply Chain Management: Simulation, Optimization, and Risk Analysis.

14. Kleijnen, Jack. P. C., Susan M. Sanchez, Thomas W. Lucas, and Thomas M. Cioppa. 2005. A User's Guide to the Brave New World of Designing Simulation Experiments. *Informs Journal on Computing* 17, (3): 263-289.

15. Kleijnen, Jack. P. C. and Robert G. Sargent. 2000. A Methodology for Fitting and Validating Metamodels in Simulation. *European Journal of Operational Research* 120: 14-29.

16. Li, C. H. 1962. A Sequential Method for Screening Experimental Variables. *Journal of the American Statistical Association* 57: 455-477.

17. Mauro, C. A. 1986. Efficient Identification of Important Factors in Large Scale Simulations. In *Proceedings of the 1984 Winter Simulation Conference*, 296-305.

18. McKay, M. D., Beckman R.J., and W. J. Conover. 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* 21, (2): 239-245.

19. Montgomery, Douglas. C. 2000. *Design and Analysis of Experiments*. 5 ed.: John Wiley & Sons.

20. Patel, M. S. 1962. Group-screening with more than two stages. *Technometrics* 4: 209-217.

21. Sanchez, Paul J. and Arnold H. Buss. 1987. A Model for Frequency Domain Experiments. In *Proceedings of the 1987 Winter Simulation Conference*, 424-427.

22. Sanchez, Susan M. 2005a. "NOLHdesigns spreadsheet." Available on-line via http://diana.cs.nps.navy.mil/SeedLab/ [accessed July 1,2005a].

23. Sanchez, Susan M. 2005b. Work Smarter, Not Harder: Guidelines for Designing Simulation Experiments. In *Proceedings of the 2005 Winter Simulation Conference*, 69-82.

24. Sanchez, Susan M. and T. W. Lucas. 2002. Exploring the World of Agent-Based Simulations: Simple Models, Complex Analyses. In *Proceedings of the 2002 Winter Simulation Conference*, 116-126.

25. Sanchez, Susan M. and Paul J. Sanchez. 2005a. Very Large Fractional Factorial and Central Composite Designs. *ACM Transactions on Modeling and Computer Simulation* 15, (4): 362-377.

26. Sanchez, Susan M., Hong Wan, and T. W. Lucas. 2005b. A Two-Phase Screening Procedure for Simulation Experiments. In *Proceedings of the 2005 Winter Simulation Conference*, 223-230. *Proceedings of the 2005 Winter Simulation Conference*.

27. Sanchez, Susan M. and Hsin-Fu Wu. 2003. Frequency-Based Designs for Terminating Simulation: A Peace-Enforcement Example. In *Proceedings of the 2003 Winter Simulation Conference*, 952-959.

28. Schruben, Lee W. and V. James COGLIANO. 1987. An Experimental Procedure for Simulation Response Surface Model Identification. *Communications of the ACM* 30, (8).

29. Trocine, Linda and Linda C. Malone. 2000. Finding Important Independent Variables through Screening Designs: A Comparison of Methods. In *Proceedings of the 2000 Winter Simulation Conference*.

30. Trocine, Linda and Linda C. Malone. 2001. An Overview of Newer, Advanced Screening Methods for the Initial Phase in an Experimental Design. In *Proceedings of the 2001 Winter Simulation Conference*.

31. Tunali, S. and I. Batmaz. 2000. Dealing with the Least Squares Regression Assumptions in Simulation Modeling. *Computers and Industrial Engineering* 38: 307-320.

32. Wallace, T. Dudley. 1977. Pretest Estimation in Regression: A Survey. *American Journal of Agricultural Economics* 59, (3): 431-443.

33. Wan, Hong, Bruce E. Ankenman, and Barry L Nelson. 2003. Controlled Sequential Bifurcation: A New Factor-Screening Method For Discrete-Event Simulation. In *Proceedings of the 2003 Winter Simulation Conference*, 565-573.

34. Wan, Hong, Bruce E. Ankenman, and Barry L Nelson. 2005. Controlled Sequential Bifurcation: A New Factor-Screening Method for Discrete-Event Simulation.

35. Wan, Hong, Bruce E. Ankenman, and Barry L. Nelson. 2004. Simulation Factor Screening with Controlled Sequential Bifurcation in the Presence of Interactions, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208-3119, U.S.A.

36. Watson, G. S. 1961. A Study of the Group Screening Method. *Technometrics* 3, (3): 371-388.

37. Ye, Kenny Q. 1998. Orthogonal Column Latin Hypercubes and Their Application in Computer Experiments. *Journal of American Statistical Association* 93, (444): 1430-1439.

Appendix

Appendix A: Computer Code for Factor Screening

Factor screening experiment can be conducted either manually or automatically. In a manual factor screening experiment, the factor screening method is usually implemented step by step by user. The interaction between the method and the simulation model is possible through active review of each result by the user. In the other words, based on what the factor screening method dictates, the user manually changes the factor levels in the simulation model, and then depending on the response value, the factor screening method specifies a new setting for the simulation model to run. Therefore, in the manual factor screening experiment, all the interaction between the method and the simulation models is done by the user. On the other hand, in an automatic factor screening experiment, all the required interactions between the factor screening method and the simulation model, including the assignments of factor levels and receiving the corresponding response value, are performed by a separate module, called a Factor Screening Class (FSC).

This Appendix addresses the general structure of the factor screening class used in the NCSU Factor Screening Class (NCSU-FSC) library developed for this thesis. The NCSU-FSC library can be imported into any simulation model implemented in a language which is executable by Microsoft Visual Studio .NET. The library will also interact with Microsoft Excel by using a simple bridge written in Microsoft Visual Studio .NET and the VS .NET collection. Moreover, NCSU-FSC requires that all the involved factors in the simulation model be listed in a database with a special structure. Detailed descriptions of the factor database are discussed in the Section A.4.

A.1. Simulation Model Structure in the Context of Factor Screening Experiment

In the context of factor screening experiment, a simulation model can be assumed to consist of three major components as described below and depicted in Figure A.1:

1. Input factors: values of involved factors used by the simulation model in order to generate the result. The involved factors can be divided into two main groups:
    a. Constants: global constant factors common among all the possible scenarios and usually don't change from one scenario to another.
    b. Scenario variables: variables whose values are assigned by the code or user according to the defined scenario.
2. Simulation procedures: includes the procedures required for generating the outputs by using the input factors.
3. Output: depending on the type of the simulation model, the output can be either deterministic or probabilistic:
    a. Deterministic output: the output of a model that has no random input; as a result, all the required analyses for factor screening are done based on only one replication of the model.
    b. Stochastic output: the output of a model that operates with at least some random inputs. To do analyses on the response of stochastic simulation models, based on the desired degree of accuracy, an appropriate number of replications should be taken.

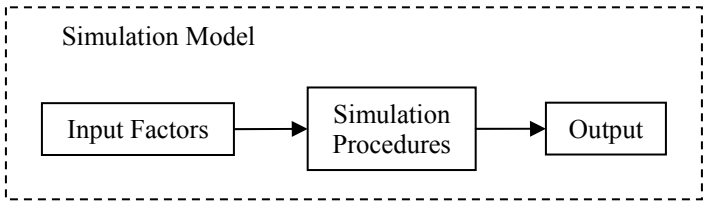


Figure A.1: Simulation Model in the Context of Factor Screening Experiment

A.2. Incorporating the Factor Screening Class into a Simulation Model

Each factor screening method, regardless of the design it employs, sets input factors at some levels, run the simulation model, get the model response, designs another treatment and continue the procedure until it stops according to a certain criteria specified by the method. Therefore, the factor screening module must be able to control the values of input factors, run the simulation model whenever necessary, and from the generated outputs, and get the information required for factor screening, such as the mean and the standard deviation of the response selected for factor screening. The general concept of this process is depicted in Figure A.2.

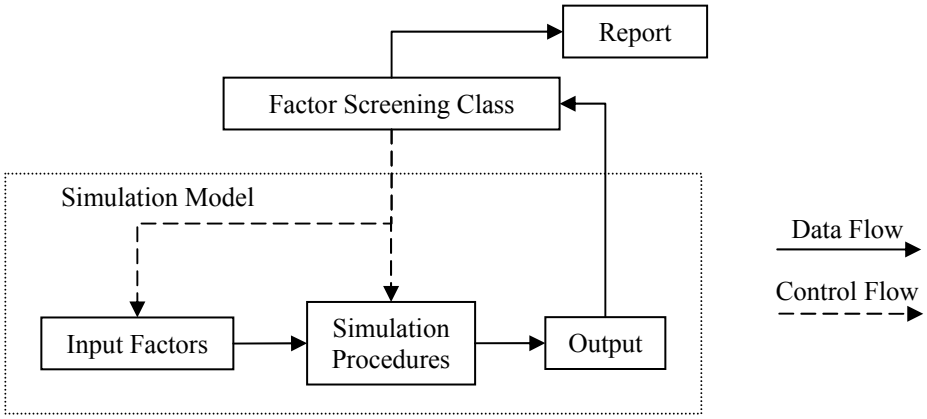


Figure A.2: Factor Screening Module installed on a simulation model

In practice, to incorporate the factor screening procedure into a simulation model, some components of the structure presented in Figure A.2 needs to be modified and some additional elements should be added. These adjustments are:

1. Use "Factor Database" instead of "Input Factors"
2. Add "Interaction Manager" component
3. Use "Response" instead of "Output"
4. Add "Signal-to-Run" component

The new conceptual model is shown in Figure A.3 and its components are described below:

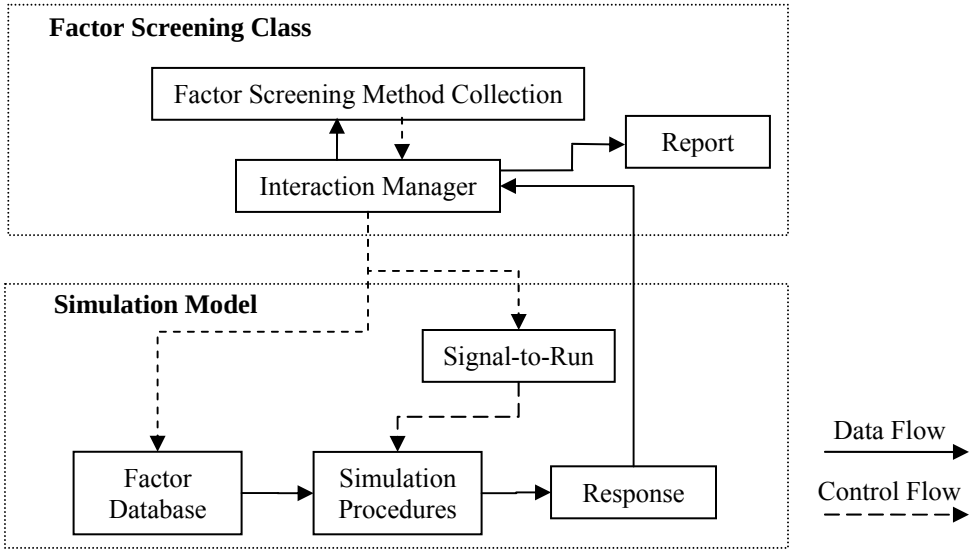


Figure A.3: Factor Screening Module installed on a simulation model

A.2.1. Factor Database

Generally, in a simulation model, all the involved factors are embedded in the simulation code. Therefore, it is completely impossible for the factor screening code to gain control over all factors for assigning them the appropriate values. In order to solve this problem, all the involved factors which the user is interested in screening, must be listed in a special-structured database and the simulation model must be coded such that it can read the required input factor from the Factor Database. The structure of the Factor Database is discussed in Section A.4.

A.2.2. Interaction Manager

The main function of the "Interaction Manager" class is to make the communication between FSC and the simulation model possible. Most of the factor screening methods use a "coded" language, where scaled values are used instead of the factor's real values. Therefore, one of the main rules of "Interaction Manager" class is to function as a translator between the model

and the factor screening class. In addition, the interaction manager has control over the factor database to change the factor levels as dictated by the screening method.

Employing this strategy has the following additional advantages:

1. Any necessary modification or improvement can be made in the screening methods regardless of the simulation model on which the factor screening is being applied.
2. To have a factor screening code that can work with different types of Factor Databases created by different applications, such as Excel, Access, and etc., we only need to add the required features in the Interaction Manager class and no change in the factor screening methods is required.
3. Many factor screening methods need to exchange information with the user before the start of the factor screening experiment and while the screening is in progress. The Interaction Manager can facilitate this type of interactions without requiring the users to change the code in their simulation models.

A.2.3. Response

Factor screening experiments can be performed on *only* one response function, which should be specified by the user. Therefore, the simulation model must be able to provide the factor screening code with the information about the response function. This information usually includes the response mean and variance, and the number of replication used to produce the response value.

The "Interaction Manager" class in NCSU-FSC has a MustOverride function called "`GetModelResponse`" which can be overridden by user with the function that returns the value of the response function on which the factor screening is done.

A.2.4. Signal-to-Run Component

One of the most desirable features of the factor screening class is its ability to operate with least interference in the simulation model. In other words, it should not require the model

developers to modify their code so as to incorporate the factor screening class into the simulation model. In order to achieve this superiority, an overridable function should be included in the Interaction Manager class which can be overridden by user with the subroutine that triggers the simulation model to run. This overridable function acts as a "Signal-to-Run" component. Therefore, whenever the factor screening class needs a new observation of the response function, it calls this function to signal the simulation model to run and then the Interaction Manager can get the response value via the "Response" component.

A.2.5. Factor Screening Methods Collection (FSMC)

The fundamental component of any factor screening library is the collection of the methods that can be employed for factor screening experiments. These methods can generally be divided into two groups:

1. Sequential methods, in which the design of each stage is determined based on the response of the previous stage; such as: SB, Cheng's method and CSB.
2. Predetermined methods such as Fractional Factorial and BP designs, where the data are collected according to a predetermined design matrix.

Almost all the statistical software, like SAS, JMP, and etc., are able to generate 2^k Fractional Factorial, Plackett-Burman, and Latin Hypercube Designs. For these methods, the FSC needs to only obtain the desired number of observations for each row of the designs matrix, and the consequent analyses are conducted by the statistical software. NCSU-FSC can provide the user with an Excel worksheet where the user can input the design matrix and then the code obtains the required number of observations for each design row and returns the values to the Excel worksheet.

It should be noted that in the implementation of these methods, factors are generally assumed that vary in a scaled range, like between -1 to 1 or 0 to 1. Therefore, by using a suitable interaction management these methods can be applied on any simulation models.

NCSU-FSC includes the following sequential factor screening methods:

1. Deterministic Sequential Bifurcation (SB)
2. Deterministic Sequential Bifurcation augmented with fold-over design (X-SB)
3. Stochastic Sequential Bifurcation (Cheng's method and the Modified Cheng's method)
4. Stochastic Sequential Bifurcation augmented with fold-over design (Cheng's method and the Modified Cheng's method augmented with fold-over design)
5. Controlled Sequential Bifurcation augmented with fold-over designs (CSB-X)

A.3. Structure of the Factor Database

In the factor screening experiments, the simulation model should be run for the different combinations of certain factor levels as specified by the screening design. For example in a 2^k factorial design, the simulation model should be run for 2^k combinations of factor levels, which requires the analyst to run the simulation model for 2^k times and each time assign different values to all the involved factors. On the other hand, simulation models are generally used for studying large-scale or complicated systems which usually include a large number of variables. As a result, when performing factor screening on a simulation model, even if an efficient screening method is employed, the required number of factor combinations is so high that changing factor levels for each simulation run by hand is prohibitive.

As a solution, it is desirable that a computer program be in charge of assigning values to the factors according to a certain screening design, at each run. In order to achieve this, it is necessary that all the involved factors in the model be explicitly listed in a database which can be accessed by the FSC module.

This section addresses the structure and configuration of the Factor Database that each simulation model with moderately large number of factors is encouraged to have in order to make the factor screening experiments possible or much easier. Table A.1 shows the fields of the factor database and the following sections provide explanation for each field.

Table A.1: Factor database

Factor Information												
General Information						Values			Expert Judgment			
Variable ID	Factor ID	Screening ID	Variable Name	Input Distribution	Parameter Name	Center Level	Low Value	High Value	Involved	Effect On	Factor Direction	Expert Guess
1	1	1	X1	Beta	Alpha	3.955	3.2765	4.4330	TRUE	"Cost/QALY"	Positive	Important
1	2	1	X1	Beta	Beta	3.955	4.4330	3.2765	TRUE	"Cost/QALY"	Positive	Important
2	3	2	X2	Beta	Alpha	4	3.5595	4.3505	TRUE	"Cost/QALY"	Negative	Unknown
2	4	2	X2	Beta	Beta	4	4.3505	3.5595	TRUE	"Cost/QALY"	Negative	Unknown
3	5	3	X3	Constant	value	0.02	0	0.03	TRUE	"Cost/QALY"	Unknown	Unknown

A.3.1 Variable ID, Factor ID, and Screening ID

We can group the factors involved in a simulation model into *distribution-based factor* and *constant factors*. Distribution-based *variables* are those whose values are determined according to a distribution function whose parameters are referred as *factors*.

For example, in Table A.1, variable-1 follows a standardized (0,1) Beta distribution with shape factor-1 as shape (alpha) and shape factor-2 as parameters. And variable-3 is a constant variable and can be referred as factor-5.

Obviously, in each simulation model there are a numbers of factors that are highly dependant on each other. It means that when changing the value of one of them, we have to change the value of the others too. A well-known example would be the parameters of random variables. For instance, assume a uniform random variable with parameter *a* as min and *b* as max. If we want to set the mean of this random variable at its upper level, we have to change the value of *a* and *b* simultaneously, equally and in the same direction. Therefore, for all the factors believed to be dependent, the same value must be assigned to the corresponding '`Screening ID`' column, so that the FSC treats them as a single factor. Otherwise, the FSC will change the value of those dependent factors as the screening method dictates, which can be completely unreasonable. For example, assign greater value to min parameter and the smaller value to max parameter!

A.3.2. Variable Name

This field can store the name of the variables.

A.3.3. Input Distribution and Parameter Name

For distribution-based variables, the name of the distribution and its parameters, and for constant variables, ‘constant’ and ‘value’ are assigned to these two fields.

A.3.4. Central, Upper and Lower Values

For each factor, central, upper, and lower levels are desired to be assigned by the expert. For some special types of factors, like random variables, FSC is not able to assign the upper and lower levels, thus the expert is requested to provide the appropriate values for the corresponding lower and upper levels.

A.3.5. Is the variable involved?

The first decision about each factor is to whether include it in the screening experiment or not. This decision is of great importance due to the following reasons:

1. Eliminating factors that have no role in the model will reduce considerably the number of required runs for factor screening experiment.
2. Eliminating factors that play an important role in the model will definitely lead to misleading results.

In addition, in the simulation models that allow users to choose different scenarios to run (like CRC model), a great number of factors are not even involved when a specific scenario is chosen. Ideally, the simulation model should be able to assign the appropriate value (i.e. involved or not involved) to this field as soon as the user defines the scenario of interest. Otherwise, the context expert would be the best source to fill out this field. By default, NCSU-FSC will consider all variables as involved in the model.

A.3.6. Effect on Which Response Function(s)?

Simulation models usually generate statistics for a various number of response functions, and it is possible that the analyst would like to perform factor screening experiment on more than

one output function. However, not all the factors have an effect on all the response functions; by contrast, a large number of factors affect only certain response functions. Therefore, having chosen the response, we need to identify the factors that have influence on it. For that reason, the analyst will be requested to specify for the FSC the different response functions on which to do factor screening. Then, for each factor, the expert should determine the response function that the factor might have effect on, and assign the function's name to the `EffectOn` field. This allows the FSC to conduct factor screening on only the factors that are believed to have effect on the selected response.

A.3.7. Factor Direction

As previously discussed, some of the screening methods assume that the signs of the main effects are known. Violation of this assumption would result in ignoring the main effects of some important factors and eventually leads to misleading results. If the expert cannot specify the main effect directions, the FSC should be able to determine the directions by running a Resolution III or a one-factor-at-a-time design.

A.3.8. Expert Guess

This field provides the experts with opportunity to import their insight about the significance of a factor into the factor screening experiment. The user can assign one of the values "Important", "Unknown", and "Unimportant" to the field `ExpertGuess`. This information can be used either for checking the results of the factor screening experiment, or for improving the efficiency of the method. The FSC can place all the factors named "Unimportant" in a single group and then apply an inexpensive design, such as a Resolution III, to judge about the significance of this group-factor.

A.4. Conclusion

Automating the factor screening experiment demands a specific structure for the simulation code and the database in which the factors and their corresponding information are stored. In this chapter we proposed an exclusive structure for the Factor Screening Class (FCS) and the factor database. These structures are believed to noticeably facilitate the process of factor screening.