12-12-2013

A Comparison Study of Second-Order Screening Designs and Their Extension

Shane A. Dougherty

A COMPARISON STUDY OF SECOND-ORDER SCREENING DESIGNS AND THEIR EXTENSION

DISSERTATION

Shane A. Dougherty, Lt Col

AFIT-ENS-DS-13-D-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-DS-13-D-01

A COMPARISON STUDY OF SECOND-ORDER SCREENING DESIGNS AND THEIR EXTENSION

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy (Simulation)

Shane A. Dougherty, BS, MS

Lt Col

December 2013

AFIT-ENS-DS-13-D-01

A COMPARISON STUDY OF SECOND-ORDER SCREENING DESIGNS AND THEIR EXTENSION

Shane A. Dougherty, BS, MS
Lt Col

Approved:

Raymond R. Hill, PhD (Chairman)	Date
James R. Simpson, PhD (Member)	Date
Joseph J. Pignatiello, Jr., PhD (Member)	Date
Edward D. White, III, PhD (Member)	Date

Accepted:

Dr. Heidi R. Ries　　Date
Interim Dean, Graduate School of Engineering and Management

AFIT-ENS-DS-13-D-01

Abstract

Recent literature has proposed employing a single experimental design capable of preforming both factor screening and response surface estimation when conducting sequential experiments is unrealistic due to time, budget, or other constraints. Military systems, particularly aerodynamic systems, are complex. It is not unusual for these systems to exhibit nonlinear response behavior. Developmental testing may be tasked to characterize the nonlinear behavior of such systems while being restricted in how much testing can be accomplished. Second-order screening designs provide a means in a single design experiment to effectively focus test resources onto those factors driving system performance. Sponsored by the Office of the Secretary of Defense (OSD) in support of the Science of Test initiative, this research characterizes and adds to the area of second-order screening designs, particularly as applied to defense testing. Existing design methods are empirically tested and examined for robustness. The leading design method, a method that is very run efficient, is extended to overcome limitations when screening for non-linear effects. A case study and screening design guidance for defense testers is also provided.

AFIT-ENS-DS-13-D-01

To my wife and children who supported me in every way imaginable as I pursued one more degree. I love you all and you fulfill my life in more ways than words can ever describe.

Acknowledgements

During my pursuit of a PhD, I have faced many trials and tribulations. At times, I truly questioned my ability to finish but my advisor, Dr. Raymond Hill, never lost faith in me. I sincerely thank him for his positive words of affirmation and guidance. My dissertation committee of Dr. Joseph Pignatiello, Dr. James Simpson, and Dr. Edward White have been invaluable. Through their ideas and comments on my draft work, I truly believe I have produced a better product than I ever could have produced on my own.

Shane A. Dougherty

Table of Contents

Page

List of Tables

Table Page

Table Page

Table Page

A COMPARISON STUDY OF SECOND-ORDER SCREENING DESIGNS AND THEIR EXTENSION

I. Introduction

1.1 Background

Shrinking budgets, in conjunction with the rising costs associated with replacing aging military hardware, have highlighted the necessity for Department of Defense (DOD) organizations to demonstrate fiscal responsibility while still maintaining core capabilities. As a result, the DOD continues to look for methods which promote efficiencies in all its operations. As such in April 2012, the Scientific Test and Analysis Techniques in Test & Evaluation Center of Excellence (STAT T&E COE) was established at the Air Force Institute of Technology Graduate School of Engineering and Management by the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation and Director, Air Force Test and Evaluation. Dr. Steven Hutchison, Principal Deputy, Office of the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation (DASD(DT&E)), stated "By applying scientific methods to the test design, we can not only achieve great efficiencies, but we can significantly improve confidence in our results. The STAT T&E COE will provide a critical venue for enhancing the test design for DOD acquisition programs."

Prior to the establishment of the STAT T&E COE, Dr. J. Michael Gilmore, Director of Operational Test and Evaluation (DOT&E), started an "initiative to increase the use of scientific and statistical methods in developing rigorous, defensible test plans and in evaluating their results" within OT&E (Gilmore, 2010). In a 2010 memorandum, Dr. Gilmore provided key policy guidance on the use of Design of Experiments (DOE) in OT&E. Furthermore the DOT&E Scientific Advisor (SA), Dr. Catherine Warner, highlighted the fact that while DOE is a structured, rigourous statistical tool for test planning and analysis, and it has been written about extensively within the academic setting, there are still many questions regarding how to apply DOE to T&E within DOD (Warner, 2011).

An ongoing effort, beginning in 2009, which focuses on transitioning basic science of test techniques and test methodology to DOD practice is the "Science of Test" initiative. Funded by OSD DOT&E in 2011, the member institutes which comprise this research consortium are Arizona State University, Virginia Tech, Naval Postgraduate School, and the AFIT Center of Operational Analysis.

This dissertation directly supports the "Science of Test" initiative. In particular, this research addresses the use of design of experiments and response surface designs to characterize the area of second-order screening designs, particularly as applied to defense testing. Extensions to existing designs are examined with respect to improvements in robustness and applicability to defense testing.

1.2 Problem Context

Response surface methodolgoy (RSM) is a collection of statistical design and numerical optimization techniques used to model a surface as an approximation for the relationship between a process or system response and its input factors (Myers and Anderson-Cook, 2009). The shape of the estimated surface is determined by the model selected to approximate the system and the response values recorded from various input factor settings. The assumption is that a response η is an unknown function of a set of design variables $x_1, x_2, ..., x_k$ and that the function can be approximated by a polynomial model. Prominent among the models considered are the first-order model

$$\eta = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k \tag{1.1}$$

and the second-order model

$$\eta = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j \tag{1.2}$$

Box and Wilson (1951) laid the foundation for RSM by outlining a philosophy of sequential experimentation which included experiments for screening, region seeking (such as steepest ascent), process/product characterization, and process/product optimization (Myers et al., 2004). Box and Liu (1999) illustrated a number of concepts which Box understood as the embodiment of RSM at the time to include the philosophy of sequential learning.

As such, the standard RSM approach is to use a three-stage process; however, there are times when the sequential nature can be a disadvantage, especially when the duration of an experiment is long or experimental preparation is time-consuming. In these instances, it would be better, if not necessary, to perform factor screening and response surface exploration on the same experiment vice conducting experiments sequentially.

1.3 Problem Statement

Military systems, particularly aerodynamic systems, are complex. It is not unusual for these systems to exhibit nonlinear behavior. Developmental testing may be tasked to characterize the nonlinear behavior of such systems but may also be restricted in how much testing can be accomplished. Second-order screening designs for nonlinear system responses provide a means to effectively focus test resources onto those factors driving system performance.

Second-order screening design methodology, sometimes referred to as One-Step Response Surface Methodology or Definitive Screening, is a relatively new focus in statistical research and effectively unknown to the defense test community. Important questions as to the method's usefulness and applicability remain unaddressed and so are examined in this research.

1.4 Research Objective and Scope

This research will characterize and add to the area of second-order screening designs, particularly as applied to defense testing. Existing design methods are tested and examined for robustness. Extensions to existing designs are examined with respect to improvements in robustness and applicability to defense testing.

We proceed with the following goals:

1. *Conduct an empirical study to characterize and better understand new proposed second-order screening designs.*

2. *Identify second-order screening design with favorable design parameter properties either through augmentation of existing designs or through creation of new designs.*

3. *Development of guidelines for use of second-order screening designs for DOD tests.*

By accomplishing these research goals, we can help make test managers within the DOD comfortable with implementing DOE techniques capable of examining the complex nature of military systems within fiscal, time, and resource constraints.

1.5 Overview

The remainder of this dissertation follows a scholarly article format. Chapter II contains a detailed literature review of screening and response surface designs, partitioned by sequential and single phase methods for fitting first order and second-order response surfaces. Chapters III, IV, and V are self-contained research articles on second-order screening designs. Each contains a literature review of the research relevant to that chapter. The original contribution of each chapter is as follows:

Chapter III formally examines the robustness of the two arguably best second-order screening designs with respect to the assumptions of both sparsity (factor or effect) and heredity (strong or weak). To date, evaluation of screening design performance has assumed both factor sparsity and strong effect heredity. The article is currently under review for publication in *Quality and Reliability Engineering International*.

Chapter IV describes a computer generated $D-$optimality design augmentation technique which uses a $k-$factor Definitive Screening Design (DSD) as a baseline fixed design and augments the design with $k-1$ additional runs. In a simulation study, the proposed augmented Definitive Screening Designs (DSD+) were able to increase the robustness of the original DSD to the principles of heredity and sparsity while also increasing the detection rate of second-order effects when both two-factor interactions and pure-quadratic effects are active. The article is currently under review for publication in the *Journal of Quality Technology*.

Chapter V presents the use of design of experiments and response surface designs in the area of second-order screening designs, particularly as applied to defense testing, through demonstrating the viable use of second-order screening designs in a wind tunnel case study. The article is currently targeted for publication in *Military Operations Research* or the *Journal of Defense Modeling and Simulation*.

And lastly, Chapter VI reiterates the importance of studying second-order screening designs, summarizes all original research contributions, and provides suggestions for future work.

II. Literature Review

This chapter covers the literature pertinent to this research effort. After a brief synopsis of Design of Experiments (DOE) and Response Surface Methodology (RSM), including common designs and terminology, the focus of the chapter shifts to relevant literature on second-order screening designs. Research on second-order screening designs falls into two broad categories; construction and design assessment/analysis methods. Both areas are extensively reviewed with gaps and limitations being discussed.

2.1 Design of Experiments (DOE)

DOE, or experimental design, is a statistical technique used to organize an experimental test or series of tests so that observed changes in an output response can be attributed to systematic changes made to the input variables of a process or system (Montgomery, 2013). While the designs are based upon statistical techniques, the actual design forms vary greatly dependent upon the experimental objective. For instance, the objectives of screening, modeling, or optimizing a process or system can result in vastly different designs.

Design selection also depends upon the form of the empirical model used to represent the process or system response. Typically, first-order polynomial models are used extensively in screening experiments while second-order polynomial models are commonly used in modeling and optimization experiments. Inherent within the designs and execution are data collection plans enabling the application of subsequent statistical analysis methods to reach valid and objective conclusions.

For the case of two independent factors, the **first-order polynomial** or **main effects** model is

$$y = \beta_0 + \beta_1 A + \beta_2 B + \varepsilon \tag{2.1}$$

where y is the response, A and B are the design factors, the βs are unknown estimable parameters, and ε is a random error term accounting for the experimental error in the system. An **interaction** term is usually added to the first-order model yielding

$$y = \beta_0 + \beta_1 A + \beta_2 B + \beta_{12} AB + \varepsilon \tag{2.2}$$

where the β_{12} represents the two-factor interaction effect between the design factors A and B. The **second-order polynomial** model with two factors is

$$y = \beta_0 + \beta_1 A + \beta_2 B + \beta_{12} AB + \beta_{11} A^2 + \beta_{22} B^2 + \varepsilon. \tag{2.3}$$

Second-order models are often used for response surface exploration (Montgomery, 2013). More general forms are given in Equations 1.1 and 1.2.

2.1.1 Screening Designs.

Many experiments may start by considering many factors, which in turn increases the overall size and cost of the experiment. Screening designs are a category of experimental designs, usually performed during the early stages of a process or system study, used to determine which of the many factors (if any) have a significant effect on the system or process. Screening designs usually assume a linear (main effects or main effects plus interaction) response function so factors can be studied at two levels and thereby conserving experimental resources.

Popular experimental designs used in screening experiments are full and fractional 2-level factorial designs, Plackett-Burman, and supersaturated designs. While all of these designs are capable of identifying the main effects, only the full factorial design is capable of identifying all interactions. To a varying degree, the remaining designs are capable of identifying some or all two-factor interaction effects.

2.1.2 Response Surface Designs (RSD).

Response surface designs are experimental designs used when the response surface is believed to possess significant curvature. In order to estimate curvature, each factor needs at least three levels. Response surface designs fulfill this requirement through augmentation of two-level regular designs or by specifying designs robust to the linear effect assumption. Response surface designs are called *second-order designs* because all $(k+1)(k+2)/2$ parameters in Equation 1.2 are estimable in the design.

A 3^k or 3^{k-p} fractional factorial design is often suggested to deal with response curvature. However, more efficient options are available including the Central Composite Design (CCD), Box-Behnken Design (BBD), and saturated/near-saturated Hoke, Hybrid, and Small Composite Designs (SCD).

2.1.2.1 Response Surface Methodology (RSM).

Since Box and Wilson (1951) laid the foundations for RSM, four comprehensive historical reviews have been written. Hill and Hunter (1966) provided a comprehensive bibliography while focusing on the practical applications of RSM in the fields of chemistry and chemical engineering. The Mead and Pike (1975) review focused on RSM as it applied to the modeling of biological data in the field of biometrics. Myers et al. (1989) reviewed important developments in RSM during the 1970s and 1980s while clearly defining RSM as "being confined to that of a collection of tools in design or data analysis that enhance the exploration of a region of design variables in one or more responses" (Myers et al., 1989).

Myers et al. (2004) provide the most current comprehensive review of RSM through discussions on advancements in robust parameter design and new developments in response surface design to include methods for evaluating response surface designs. Additionally, Myers et al. (2004) address both design and optimization issues for multiple responses and the application of generalized linear models.

Unfortunately, the nature of DOE and RSM sometimes makes it difficult to differentiate or draw a clear distinction between the two. Whereas DOE is comprised of RSDs used for response surface models which include quadratic terms for curvature, RSM employs DOE screening designs. As the RSM name implies, RSM is best viewed in context as a methodology which employs DOE elements with the goal of determining how changes in design variables can provide process improvement or optimization. As such, the standard RSM can be described as consisting of two stages: factor screening and response surface exploration.

Traditionally, the research for factor screening and for response surface exploration proceed not in concert but along separate avenues. The former involves concepts like design resolution, minimum aberration, power, and the number of clear (non-confounded) effects, and the latter involves the concepts like rotatability, alphabetical-optimality, and prediction variance.

2.1.3 Design Resolution.

Resolution is a measure of the degree of confounding for main effects and interactions in a fractional factorial design. Resolution is generally denoted in Roman numerals. The smallest useful resolution is III, and a design can technically have resolutions has high as $k + 1$. Designs of resolution III, IV, and V are most prevalently used because of the nature of confounding found within the designs. The confounding characteristics of these design resolutions are:

- Res III: Main effects clear of other main effects, at least one main effect is confounded with at least one two-way interaction.

- Res IV: Main effects are clear of two-way interactions, but at least one two-way interaction is confounded with at least one other two-way interaction.

- Res V: Main effects and two-way interaction are clear of any other main effect or two-way interaction, but at least one two-way interaction is confounded with at least one three-way interaction.

As an example, a design which confounds a variable A with a two-way interaction BC would at best be a resolution III design where A and BC are correlated or aliased, and therefore their effects cannnot be independently quantified. Usually the design that has the highest resolution possible, while meeting the required fractionation for design run size consideration, is employed.

There are times however that different designs can possess the same resolution and fractionation but have different confounding or aliasing structure. Fries and Hunter (1980) proposed the concept of design aberration for regular two-level designs as a means to differentiate between these designs. They defined a minimum aberration design as the design of maximum resolution R "which minimizes the number of words in the defining relation that are of minimum length". Since Fries and Hunter's initial work, the concept of minimum aberration criterion has been extended to two-level non-regular, multilevel, and mixed-level fractional-factorial designs (Guo et al., 2009).

2.1.4 Optimality Criteria.

Optimal designs are typically assessed based upon specific criteria like providing good estimation of model parameters or good prediction capacity within the design region. Alphabetic-optimality refers to the family of design optimality criteria that are characterized by a letter of the alphabet, currently $A-$, $D-$, $G-$, $V-$, or $I-$. These alphabetical-optimality criteria drive what constitutes an optimal design. These optimal designs are rather focused on a particular design characteristic. Two of the most popular methods of characterizing optimality are $I-$ and $D-$optimality.

$D-$optimality is based upon the notion of selecting design runs which maximize the determinant of $\mathbf{X}'\mathbf{X}$, denoted as $|\mathbf{X}'\mathbf{X}|$, where $\mathbf{X}$ is the model matrix consisting of the levels of the design matrix $\mathbf{D}$ expanded to model form. By selecting design runs which maximize $|\mathbf{X}'\mathbf{X}|$ or minimize $|(\mathbf{X}'\mathbf{X})^{-1}|$, $D-$optimal designs minimize the volume of the joint confidence region on the vector of the model regression coefficients $\hat{\boldsymbol{\beta}}$. Hence $D-$optimal designs focus on producing designs which provide good model parameter estimates.

$A-$optimality focuses on producing good model parameter estimates by minimizing the **trace** of $\mathbf{X}'\mathbf{X}^{-1}$, denoted $\text{tr}(\mathbf{X}'\mathbf{X})^{-1}$. In contrast to $D-$optimal designs which consider the covariances among coefficients through examining $|\mathbf{X}'\mathbf{X}|$, $A-$optimal designs deal with only the diagonals of $(\mathbf{X}'\mathbf{X})^{-1}$, which are related to the individual variances of the regression coefficients (Montgomery, 2013).

While $D-$ and $A-$optimal designs focus on good model parameter estimates, $G-$, $V-$, and $I-$optimal designs focus on good prediction capacity within the design region by focusing on the scaled prediction variance $N\text{Var}[\hat{y}(\mathbf{x})]/\sigma^2 = \nu(\mathbf{x})$. The $G-$optimality criterion is based upon the maximum $\nu(\mathbf{x})$ over the entire design region, the $I-$optimality criteria is based upon the $\nu(\mathbf{x})$ over a region of interest, and the $V-$optimality criteria is based upon the $\nu(\mathbf{x})$ for a specified set of points in the design region. A design is considered $G-$optimal if the maximum value of the $\nu(\mathbf{x})$ over the design region is a minimum, while a design is considered $V-$optimal if it minimizes the average $\nu(\mathbf{x})$ over a set of points of interest in the design region. Finally, a design is considered $I-$optimal if it minimizes the average $\nu(\mathbf{x})$ over the design region for the regression model.

Since $G-$, $V-$, and $I-$ criteria are prediction-oriented and $A-$ and $D-$ criteria are parameter-oriented criteria, the $G-$, $V-$, and $I-$ criteria are mostly used for second-order designs while the $A-$ and $D-$ criteria are mostly used for first-order designs. While the $G-$ and $D-$ criteria are widely seen throughout literature, the $G-$criterion can become computationally difficult as the design matrix grows. Fortunately, the $I-$criterion is computationally easier to implement than the $G-$criterion, and is available in several software programs (Montgomery, 2013). For more on alphabetic-optimality, please see Chapter 8 in (Myers and Anderson-Cook, 2009).

2.2 Full Model Estimable Designs

Designs which are full model estimable are designs which can estimate all factors within the form of the empirical model used to represent the process or system response. For a second-order polynomial model, the design must contain enough degrees of freedom to estimate p effects where

$$p = 1 + 2k + \frac{k(k-1)}{2} = \frac{(k+1)(k+2)}{2}$$

Recall a second-order polynomial model contains 1 intercept, k main effects, k pure quadratic, and $k(k-1)/2$ two-factor interaction terms for k factors (Myers and Anderson-Cook, 2009).

2.2.1 2^k and 3^k Factorial Designs.

In contrast to the *One Factor at Time (OFAT)* design strategy where factors are varied individually, *Factorial Designs* vary factors simultaneously thus allowing for estimates of interactions between factors. If measurements are made on the system or process response for all possible combinations of the values or levels of the different factors, the design plan is called a full factorial design experiment (Connor and Zelen, 1959). For example, if two factors A and B have a and b levels, respectively, each factorial design replication would contain ab treatment combinations while adding an additional factor C with c levels would require abc treatment combinations.

The 2^k Factorial Design consists of k factors each at only two levels and is a special case of the full factorial design with 2^k observations per replication. 2^k designs have many useful properties. In addition to being orthogonal, 2^k designs are $A-$, $G-$, $D-$ and $I-$optimal for fitting a first-order model or first-order model with interactions (Montgomery, 2013). The 2^k-type designs are widely used for factor screening as it provides the smallest number of runs for independently estimating all main effects and interactions for k factors. In total, the number of estimable effects for a 2^k design is $2^k - 1$ consisting of k main effects, $\binom{k}{2}$ two-way interactions, $\binom{k}{3}$ three-way interactions, ... , and 1 k-way interactions. A 2^3 design with three factors, denoted A, B, and C, can estimate $k = 3$ main effects (A, B, C), $\binom{3}{2} = 3$ two-way interactions (AB, AC, BC), and one 3-way interaction (ABC). See Table 1 for the 2^3 design matrix.

Table 1. A 2^3 Full Factorial Design

Run	A	B	C	AB	AC	BC	ABC
1	−	−	−	+	+	+	−
2	+	−	−	−	−	+	+
3	−	+	−	−	+	−	+
4	+	+	−	+	−	−	−
5	−	−	+	+	−	−	+
6	+	−	+	−	+	−	−
7	−	+	+	−	−	+	−
8	+	+	+	+	+	+	+

Note: Factor settings have been coded, replacing the low setting by − and the high setting by +.

The 3^k Factorial Design, which consists of k factors each at only three levels, is a special case of the full factorial design with $N = 3^k$ observations per replication. See Table 2 for a 3^3 design matrix.

Table 2. A 3^3 Full Factorial Design

Run	A	B	C	A^2	B^2	C^2	AB	AC	BC	ABC
1	0	0	0	2	2	2	2	2	2	0
2	0	0	1	2	2	1	2	1	1	1
3	0	0	2	2	2	2	2	0	0	2
4	0	1	0	2	1	2	1	2	1	1
5	0	1	1	2	1	1	1	1	1	1
6	0	1	2	2	1	2	1	0	1	1
7	0	2	0	2	2	2	0	2	0	2
8	0	2	1	2	2	1	0	1	1	1
9	0	2	2	2	2	2	0	0	2	0
10	1	0	0	1	2	2	1	1	2	1
11	1	0	1	1	2	1	1	1	1	1
12	1	0	2	1	2	2	1	1	0	1
13	1	1	0	1	1	2	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1
15	1	1	2	1	1	2	1	1	1	1
16	1	2	0	1	2	2	1	1	0	1
17	1	2	1	1	2	1	1	1	1	1
18	1	2	2	1	2	2	1	1	2	1
19	2	0	0	2	2	2	0	0	2	2
20	2	0	1	2	2	1	0	1	1	1
21	2	0	2	2	2	2	0	2	0	0
22	2	1	0	2	1	2	1	0	1	1
23	2	1	1	2	1	1	1	1	1	1
24	2	1	2	2	1	2	1	2	1	1
25	2	2	0	2	2	2	2	0	0	0
26	2	2	1	2	2	1	2	1	1	1
27	2	2	2	2	2	2	2	2	2	2

Note: Factor settings have been coded, replacing the low setting by 0, intermediate setting by 1, and the high setting by 2.

The addition of a third factor level over the 2^k design allows modeling the response surface as a quadratic function. Each main effect has 2 degrees of freedom used to estimate a first-order (linear) and second-order (quadratic) component. While each two-way interaction has 4 degrees of freedom, one for each linear×linear, linear×quadratic,

quadratic×linear, and quadratic×quadratic effect. In total, the number of estimable effects for a 3^k design is $3^k - 1$ consisting of k main effects, k pure quadratic effects, $\binom{k}{2}$ two-way interactions with four degrees of freedom, $\binom{k}{3}$ three-way interactions with eight degrees of freedom, ... , and 1 k-way interactions with 2^k degrees of freedom. For example, a design with three factors, denoted A, B, and C, can estimate $k = 3$ main effects (A, B, C), $k = 3$ pure quadratic effects (A^2, B^2, C^2), $\binom{3}{2} = 3$ two-way interactions (AB, AC, BC), and one 3-way interaction (ABC).

2.2.2 Central Composite Designs (CCD).

Box and Wilson (1951) introduced a class of response surface designs as an alternative to the 3^k factorial designs. The Central Composite Designs (CCD) contain a 2^k or 2_V^{k-p} (see 2.3.1) design, axial/star runs, and center runs which are set at the middle of the factor range. One reason for the CCD being a popular class of second-order designs is because of the sequential nature in which they can be implemented. Typically, if the first-order response model associated with the 2^k or 2_V^{k-p} design proves to be a poor representation of the system response, center points are added to provide information on the overall curvature in the system while axial points are added to allow for the fitting of a second-order response model.

In addition to the number of runs associated with the 2^k or 2_V^{k-p} design, the CCD contain $2k$ runs per replication on the axis of each factor at a distance α from the center of the design. As such, the CCD typically involve k factors at 5 levels per factor. The value of α can be chosen so the design is rotatable, meaning the prediction variance for some point $\mathbf{x}$ is the same at all points that are equidistant from the design center. For CCD, the rotatable condition is satisfied by choosing $\alpha = \sqrt[4]{n_f}$, where n_f is the number of factorial runs. In other words, the variance of predicted response $\text{Var}[\hat{y}(\mathbf{x})]$ is constant on spheres (Montgomery, 2013). However, it is not necessary to have exact rotatability. By using $\alpha = \sqrt{k}$, the CCD is not necessarily rotatable, but the loss in rotatability is negligible while producing a more preferable design (e.g., more meaningful design-level settings (Myers and Anderson-Cook, 2009)).

Lastly, the CCD contains n_c center runs. The number of center runs affects the variance of the predicted response $\text{Var}[\hat{y}(\mathbf{x})]$. In the case of spherical or near spherical designs, ($\alpha = \sqrt{k}$ or $= \sqrt[4]{n_f}$), having $3-5$ center runs achieves a reasonable distribution of the scaled prediction variance, $SPV(\mathbf{x}) = N\text{Var}[\hat{y}(\mathbf{x})]/\sigma^2$ over the design region (Myers and Anderson-Cook, 2009).

2.2.3 Face-Centered Composite Designs (FCD).

A variant of the standard CCD is called the *face-centered* composite design (FCD). This design locates the axial points in the center of each face in the factorial space at a distance $\alpha = 1$ from the center. The face-centered design sacrifices rotatability but is useful in design situations that prevent larger axial distances, such as designs near the edges of performance envelopes. Additionally, as compared to the CCD, the FCD requires only $1-2$ center runs in order to achieve a reasonable distribution of the scaled prediction variance over the design region.

2.2.4 Computer Generated Designs.

Two important and useful concepts in statistical procedures used to assess experimental designs are optimality and robustness. Whereas the robustness of a design implies the design is insensitive to assumptions and/or models, optimal designs are generally developed for a specific set of assumptions and/or models.

Based upon the empirical model selected to represent the system response, available sample size, design factor values, a set of candidate points, and other constraints, "optimal" designs can be generated through the use of computer algorithms. While many criteria are available with which to generate designs, the criterion most often used due to its relatively simple computational nature is D-optimality (Myers and Anderson-Cook, 2009). However some computer packages use a criterion based upon good prediction capacity through examining scaled prediction variance. For instance, JMP can generate both $D-$optimal and $I-$optimal designs.

In contrast to algorithms where all possible sets of candidate points were evaluated, Meyer and Nachtsheim (1995) developed a coordinate exchange algorithm which systematically searched individual design coordinates to find the optimal settings thereby removing the candidate set of runs requirement (Montgomery, 2013).

The term "optimal" can be misleading as it implies the computer generated design is the single best design to use in a given situation. However, in truth, the "optimal" design is more likely to be one of a range of designs which can be used to meet a specific scientific objective. Both a benefit and disadvantage of computer generated "optimal" designs is while a custom design can be created for any specified model vice using a standard design, the design criterion is based upon the "correctness" of the model matrix. DuMouchel and Jones (1994) addressed the model-dependency problem by presenting "a Bayesian modification of $D-$optimality that allows the experimenter to 'hedge bets' about an assumed model."

While caution should be taken when dealing with computer generated designs there are times when they are helpful. For instance, when there are constraints on factor-level combinations and sample size, or unusual combinations of factor range, or there is the need to augment some current design with additional runs (Myers and Anderson-Cook, 2009).

2.3 Reduced Run Designs

As the number of factors k increases the run size requirement increases to a point which make full factorial designs sometimes impractical and inefficient. The *sparsity of effects principle* states the effects on a system or response of interest attributable to most high-order interactions are negligible when compared to some of the main effects and low-order interactions (Montgomery, 2013). For example, a full 2^7 design requires 128 runs for estimating 127 main effects and interactions but sparsity of effects means only a subset of the 7 main effects and 21 two-way interactions are likely significant. As such, only a fraction of the complete 2^7 runs are required to obtain estimates on significant effects. As a result, reduced run designs have been developed to be more efficient in terms of design size.

2.3.1 2^{k-p} and 3^{k-p} Fractional Factorial Designs (FFD).

The 2^{k-p} Fractional Factorial Design is comprised of a subset of the runs of the 2^k Factorial Design. Similar to the 2^k Factorial Design, the 2^{k-p} Fractional Factorial Designs consists of k factors each at only two levels. The value of p specifies the degree to which the design is fractionated, determined by $1/2^p$.

Table 3. A 2^{7-4} Fractional Factorial Design, Principle Fraction

Run	A	B	C	D=AB	E=AC	F=BC	G=ABC
1	+	+	+	+	+	+	+
2	+	+	−	+	−	−	−
3	+	−	+	−	+	−	−
4	+	−	−	−	−	+	+
5	−	+	+	−	−	+	−
6	−	+	−	−	+	−	+
7	−	−	+	+	−	−	+
8	−	−	−	+	+	+	−

For instance, a 2^{7-4} design (See Table 3) is a $1/2^4 = 1/16^{th}$ fraction of the 2^7 design. As such, the 2^{7-4} design contains 8 runs or $1/16^{th}$ of the 128 runs for the 2^7 design. A key issue is how should the fractional design be selected.

Generally, the first $k-p$ independent columns are generated by the runs in the 2^{k-p} design. In the 2^{7-4} design, the first 3 columns are generated by the runs associated with the 2^3 full factorial design. The remaining p columns can be generated as interactions of the first $k-p$ columns (Wu and Hamada, 2011). While these p columns are dependent upon the first $k-p$ columns, they are independent of each other. As such, the value of p determines the required number of independent design generators. Because the design generators were determined by column interactions, the p factor effect estimates are aliased, meaning the factor effects on the system response can not be estimated separately from factor interactions.

For the 2^{7-4} design in Table 3, the $p = 4$ design generators are $D = AB$, $E = AC$, $F = BC$, and $G = ABC$. Since $D = AB$, the estimate of the effect of factor D on the response is affected by the effects of A and B. The degree to which the effects are aliased is given by the design resolution. The 2^{7-4} design in Table 3 is of Resolution III (2^{7-4}_{III}) because main effects (D) are aliased with two-way interactions (AB). The technique used to generate the design in Table 3 will provide the "principle" fraction of a complete 2^p family of fractions. In practice any of the remaining $2^p - 1$ fractions may be used, each having the same design resolution.

While the design generators identify some of the alias structure, the complete design alias structure is determined by the complete defining relation for the design obtained by adding all combinations of the design generators. The defining relation is comprised of the $p = 4$ design generators and their $2^p - p - 1 = 11$ interactions (Montgomery, 2013). While 2^{k-p}_V designs would be more desirable because of their aliasing structure, 2^{k-p}_{IV} and 2^{k-p}_{III} designs are most commonly used for screening due to more economical run sizes.

When a large number of factors are being considered, the 3^k factorial can be excessively large, even more so than for the 2^k factorial. However, similar to the 2^k factorial, under sparsity of effect, fractional designs can be considered which still provide sufficient information for significant effect estimations.

The 3^{k-p} Fractional Factorial Designs consists of k factors each at three levels. The value of p again specifies the degree to which the design is fractionated, determined by $1/3^p$. For instance, a 3^{k-2} design is a $1/9^{th}$ fraction, while 3^{k-3} design is a $1/27^{th}$ fraction. A general procedure for constructing a 3^{k-p} fractional factorial design is given by Montgomery (2013). Connor and Zelen (1959) and Xu (2005) provide an extensive list of 3^{k-p} designs. Unfortunately, especially as compared to 2^{k-p} designs, the aliasing structure for 3^{k-p} designs is very complex especially as the level of fractioning increases. If effect interactions are not negligible, design results can be difficult, even nearly impossible to interpret because of the partial aliasing of two-degree-of-freedom components (Montgomery, 2013).

Regular designs are 2^{k-p} and 3^{k-p} designs constructed through defining relations among its factors. Nonregular designs lack such a defining relation. Two-level nonregular designs often used for factor screening are Plackett-Burman and Supersaturated designs (Cheng and Wu, 2001). Three-level nonregular response surface design are Box-Behnken designs.

2.3.2 Plackett-Burman Designs (PB).

Plackett and Burman (1946) developed nonregular two-level fractional factorial designs which can study $k = N - 1$ variables in N runs, where N is a multiple of 4. If

$N = 2^i$ for $i \geq 2$, PB designs are synonymous with 2^k factorial designs. An example design is presented in Table 4 where $N = 12$ runs for $k = 11$ factors.

Table 4. A 12-run Plackett-Burman Design for 11 factors

Run	A	B	C	D	E	F	G	H	I	J	K
1	+	−	+	−	−	−	+	+	+	−	+
2	+	+	−	+	−	−	−	+	+	+	−
3	−	+	+	−	+	−	−	−	+	+	+
4	+	−	+	+	−	+	−	−	−	+	+
5	+	+	−	+	+	−	+	−	−	−	+
6	+	+	+	−	+	+	−	+	−	−	−
7	−	+	+	+	−	+	+	−	+	−	−
8	−	−	+	+	+	−	+	+	−	+	−
9	−	−	−	+	+	+	−	+	+	−	+
10	+	−	−	−	+	+	+	−	+	+	−
11	−	+	−	−	−	+	+	+	−	+	+
12	−	−	−	−	−	−	−	−	−	−	−

The nonregular Plackett-Burman designs sacrifice a simple alias structure for better run economy and projectivity when compared to regular 2^{k-p} designs. A 2^{k-p}_{III} has *projectivity* 2, meaning it will collapse into a 2^2 factorial in a subset of any two of the original k factors, while $PB^{k=N-1}_{III}$ have projectivity 3 or 4 depending upon the design size. For instance, the Table 4 design will project into a full 2^3 factorial from 11 factors in 12 runs while a comparable 2^{11-7}_{III} design will only project into a full 2^2 factorial from 11 factors in 16 runs.

Unfortunately, PB designs have complex alias structures. In the Table 4 design, each main effect is partially aliased with 45 two-factor interactions while each main effect in a 2^{11-7}_{III} design is completely aliased with at most 4 two-factor interactions (Montgomery, 2013). Due to the complex aliasing structure, analysis of PB designs can become difficult. Hamada and Wu (1992) discuss methods for analyzing designs with complex aliasing based upon the *sparsity of effect* and *effect heredity* principles.

The *effects heredity* principle states if an interaction is significant the components of the interaction are significant. Under strong heredity all main effects within a significant interaction are themselves significant; however, under weak heredity not all the main effects are significant. In combination with effect sparsity, effect heredity would be concerned with only significant two-factor interactions. Thus if AB is significant, then under strong heredity, A and B would be significant while under weak heredity only A or B would be significant (Montgomery, 2013).

2.3.3 Box-Behnken Designs (BBD).

Box and Behnken (1960) developed a family of 17 efficient rotatable/near-rotatable spherical three-level designs suitable for fitting second-order (quadratic) response models. The BBD are formed by combining two-level factorials with balanced incomplete block designs (BIBD) or partially balanced incomplete block designs (PBIBD). In contrast to the CCD, the Box-Behnken design does not contain any points at the vertices or face-center of the design but rather at the center of the edges of the process space. As a result, the Box-Behnken designs avoid extreme values for factor-level combinations which may be impossible to test due to cost or physical process constraints (Montgomery, 2013).

Of the original 17 designs proposed by Box and Behnken, 10 were constructed from BIBDs while 7 were constructed from PBIBDs. BIBD are incomplete block designs where each factor appears an equal number of times with every other factor while with PBIBD each factor does not appear an equal number of times. The BBD are formed by varying p parameters in a full factorial manner while the remaining $k - p$ parameters are kept steady at the center factor level setting. For $k = 3 - 5$ and $6 - 7$, $p = 2$ and 3, respectively for the BBD designs. Additionally, the BBD uses three to five center runs to avoid singularity in the design matrix for $k = 4$ and 7 and to maintain favorable design qualities like a reasonable $\text{Var}[\hat{y}(\mathbf{x})]$ distribution (Myers and Anderson-Cook, 2009).

Overall, the design run requirements for both the BBD and CCD are comparable. For $k = 3$ and 5, the CCD with the full two-level factorial requires two more runs, not including center runs, than the BBD while for $k = 4$, the CCD and BBD require an equal number of runs. As a result, the benefit of employing a BBD design over a CCD is not necessarily due to run efficiency but rather the factor level combination location in the design space.

Through the years some of the original BBD have been improved upon in terms of rotatability, average prediction variance, $D-$ and $G-$efficiency (Nguyen and Borkowski, 2008). In addition, new Box-Behnken type designs with larger k (Mee, 2000) and differing orthogonally blocked solutions (Nguyen and Borkowski, 2008) than the original BBD have been proposed. More recently *small* Box-Behnken Designs (SBBD) have been proposed which reduce the run size requirement of the original BBD by replacing the full 2^k factorial designs within the balanced incomplete block designs (BIBD) or partially balanced incomplete block designs (PBIBD) partly by 2^{3-1}_{III} designs and partly by full factorial designs (Zhang et al., 2011). When compared to the original BBD, the SBBD possess smaller $D-$efficiency values but the values are still relatively high ($> 70\%$) for $k \leq 11$ while requiring fewer runs.

2.3.4 Other Reduced Run Designs.

Oehlert and Whitcomb (2002) proposed a class of equireplicated irregular fractions of 2^k factorials with resolution V where equireplicated means each factor occurs an equal number of times at their high and low levels. These designs, called Minimum-Run Res V Designs, are constructed using the Li and Wu (1997) columnwise-pairwise algorithm to optimize the $D-$optimal criteron. These may be used on their own if interested in a first-order response model or used as the factorial component of the CCD for a second-order response model estimation.

Morris (2000) proposed a method for constructing three-level designs, called augmented pairs designs, suitable for fitting second-order response models within a cuboidal region of interest. Starting with an initial two-level first-order design, the third level of each factor is determined by a linear combination of the levels of every pair of points. In comparison to BBD and CCD, Morris (2000) showed that the precision of model parameter and expected response estimates are favorable and requires fewer runs (Myers et al., 2004). In comparison to small composite designs, augmented pair designs show better model parameter and expected response estimate but do require more runs.

Gilmour (2006) introduced a class of three-level designs made up of subsets of 3^k factorial designs befittingly known as *subset designs*. Letting S_r represent the rth orbit, subset designs have the form $c_0S_0 + c_1S_1 + \cdots + c_kS_k$, where c_r, $r = 1, \ldots, k$, is the number of replicates of the points in S_r and an orbit is comprised of a subset of points of the 3^k factorial design on the hypersphere of radius $r^{\frac{1}{2}}$ about the center point, S^0. As such each subset design may contain points from any number of orbits with each subset S_r containing $\binom{k}{r}2^r$ points consisting of a 2^r factorial design at levels ± 1 for each combination of r factors and with the remaining $q - r$ factors at 0. In order for the subset design to be capable of fitting a second-order response, Gilmour (2006) stipulated two requirements:

- $c_r > 0$ for at least two r and $c_r > 0$ for at least one r with $1 \leq r \leq q-1$ so that all quadratic parameters can be estimated

- $c_r > 0$ for at least one $r \geq 2$ so that all interactions can be estimated.

Additionally, Gilmour (2006) specified *fractional subset* and *incomplete subset* reduced run designs where fractional subset designs replace all the 2^r factorials in at least one S_r by a fractional factorial and incomplete subset designs use a reduced number of the $\binom{k}{r}$ factorial sets of r factors.

2.4 Saturated/Near-Saturated Designs

Recall a second-order model contains $p = 1 + 2k + k(k-1)/2$ terms. A $k = 4$ factor BBD design has 15 terms to estimate but the design itself contains 24 points plus center runs. While reduced run designs like the CCD and the BBD provide more efficient designs than the full model estimable designs, these designs still can possess far more design points than needed to estimate the second-order response effects. As a result, the class of saturated or near-saturated designs have been developed. Saturated or near-saturated designs are designs such that the number of design points are equal to or near, but not less than, the number of terms in the design model.

2.4.1 Small Composite Designs (SCD).

In contrast to the CCD and FCD, which contain a 2^k or 2_V^{k-p} factorial design, Hartley (1959) suggested replacing the factorial design with a special resolution III factorial design, denoted III^* where two-factor interactions are not aliased with other two-factor interactions. As a result, the number of design runs is decreased resulting in *Small Composite Designs (SCD)*. The SCD sacrifices good prediction variance properties with the reduction in run size because main effects could be aliased with two-factor interactions. However, the SCD design still allows for the estimation of all main-effects because the star portion of the design provides additional information. While Hartley (1959) suggested replacing the 2^k or 2_V^{k-p} factorial with a III^* factorial, additional work included using irregular 2^k fractions (Westlake, 1965) and columns of Plackett-Burman designs (Draper, 1985). Draper and Lin (1990) improved upon the previous design work associated with modifying the composite structure of the SCD by adding a $k = 10$ design and reducing the run size on previous designs by deleting repeat runs (see Table 5). While deleting repeat runs reduced the design size, it also reduces the amount of information available to estimate pure error.

Table 5. Cube Points in Some Small Composite Designs (Draper and Lin, 1990)

	3	4	5	6	7	8	9	10
Coefficients								
$p = (k+1)(k+2)/2$	10	15	21	28	36	45	55	66
Star points $2k$	6	8	10	12	14	16	18	20
Minimal points in cube	4	7	11	16	22	29	37	46
Box and Hunter (1957)	8	16	16	32	64	64	128	128
	(2^3)	(2^4)	(2_V^{5-1})	(2_V^{6-1})	(2_V^{7-1})	(2_V^{8-2})	(2_V^{9-2})	(2_V^{10-3})
Hartley (1959)	4	8	–	16	32	–	64	–
	$(2_{III^*}^{3-1})$	$(2_{III^*}^{4-1})$	–	$(2_{III^*}^{6-2})$	$(2_{III^*}^{7-2})$	–	$(2_{III^*}^{9-3})$	–
Westlake (1965)	–	–	12	–	26	–	44	–
	–	–	$(3/8X2^5)$	–	$(13/64X2^7)$	–	$(11/128X2^9)$	–
Draper (1985)	–	–	12	–	28	–	44	
Draper and Lin (1990)	4	8	12	16	24	36	40	48
after elimination of repeat	4	8	11	16	22	30	38	46

2.4.2 Rechtschaffner Designs.

Rechtschaffner (1967) presented a class of saturated second-order designs for k factors in a cuboidal region of interest based upon 4 different design generators shown in Table 6. Each design generator is identified with the different terms of the second-order model.

Table 6. Design Generators for Saturated Fractions of 3^n Factorial Design

Number	Design Generator
I	(-1,$\cdots$, -1) for all n
II	(-1, 1, $\cdots$, 1) for all n
III	(-1, -1, 1) for $n = 3$
	(1, 1, -1, $\cdots$, -1) for $n > 3$
IV	(1, 0, $\cdots$, 0) for all n

For instance, Design Generator I identified with the intercept term while Design Generators II and III are identified with the main effect and two-way interaction effects, respectively. Treatment combinations are obtained by permuting the elements of each design generator to reach the desired saturated fraction, see Table 7. While the designs are not based upon $D-$optimality criterion, the signs of the design generators can be varied in order to get higher D values. Unfortunately, while Rechtschaffner designs are available for any k, they should be limited to small values of k because as k grows the designs can be shown to have an asymptotic $D-$efficiency of 0 with respect to the class of saturated designs (Notz, 1982).

Table 7. Saturated Fraction of a 3^5 Factorial Design

Run	Design Generator	A	B	C	D	E
1	(-1, -1, -1, -1, -1)	-1	-1	-1	-1	-1
2	(-1, 1, 1, 1, 1)	-1	1	1	1	1
3		1	-1	1	1	1
4		1	1	-1	1	1
5		1	1	1	-1	1
6		1	1	1	1	-1
7	(1, 1, -1, -1, -1)	1	1	-1	-1	-1
8		1	-1	1	-1	-1
9		1	-1	-1	1	-1
10		1	-1	-1	-1	1
11		-1	1	1	-1	-1
12		-1	1	-1	1	-1
13		-1	1	-1	-1	1
14		-1	-1	1	1	-1
15		-1	-1	1	-1	1
16		-1	-1	-1	1	1
17	(1, 0, 0, 0, 0)	1	0	0	0	0
18		0	1	0	0	0
19		0	0	1	0	0
20		0	0	0	1	0
21		0	0	0	0	1

2.4.3 Box-Draper Designs.

Box and Draper (1974) presented a class of saturated second-order designs for k factors in a cuboidal region of interest based on $D-$optimality. Although the designs are optimal for $k = 2$ and 3, they are not optimal for $k \geq 4$. Dubova and Federov (1972) found a better design for $k = 4$ and Notz (1982) found better designs for $k = 5$ or 6 by presenting alternative designs higher $D-$optimal criterion values. Additionally, while better designs for $k \geq 7$ have not been identified, Box and Draper (1974) designs were proved not optimal via an existence result. Therefore, the Box and

Draper designs are minimal $D-$optimal designs for $k = 2$ and 3, and are "minimal designs *of a simple form* for any k" for a cuboidal region of interest. Similar to the Rechtschaffner designs, the Box and Draper designs are available for any k, however they too should be limited to small values of k because as k grows the designs can be shown to have an asymptotic $D-$efficiency of 0 with respect to the class of saturated designs (Notz, 1982).

2.4.4 Hybrid Designs.

Roquemore (1976) presented a set of saturated or near-saturated second-order designs for $k = 3$ to 6 factors which are rotatable or near-rotatable while achieving the same degree of orthogonality as a CCD. The hybrid designs for k variables is constructed by first augmenting a $k - 1$ variable central composite design with an additional column for variable k. The design is then augmented with additional runs for variable k at different levels to create desirable design properties. Table 8 shows the design matrix for hybrid 310. In this instance, $k = 3$, so the hybrid design contains a $k = 2$ CCD augmented with a third column. These designs do suffer from having odd factor level settings. For instance, none of the 10 factor level settings for C in Table 8 are set to the typical values of 0 or ± 1.

Table 8. Hybrid Design 310: $k = 3$ and $n = 10$

Run	A	B	C
1	0	0	1.2906
2	0	0	-0.1360
3	-1	-1	0.6386
4	1	-1	0.6386
5	-1	1	0.6386
6	1	1	0.6386
7	1.736	0	-0.9273
8	-1.736	0	-0.9273
9	0	1.736	-0.9273
10	0	-1.736	-0.9273

2.4.5 Hoke Designs.

Hoke (1974) presented a class of second-order designs for $k = 3$ to 6 factors at 3 levels based on saturated and near-saturated irregular fractions of the 3^k factorial. For each number of factors k, seven versions of the Hoke designs exist, denoted $\mathbf{D}_1$, $\mathbf{D}_2$, ..., $\mathbf{D}_7$, consisting of a mixture of factorial, axial, and edge points making the Hoke designs suitable for a cuboidal region of interest (Myers and Anderson-Cook, 2009). Tables 9 and 10 show the design matrices for two versions, one saturated $\mathbf{D}_2$ and one near-saturated $\mathbf{D}_6$, of Hoke designs for $k = 3$.

Table 9. Hoke Design $\mathbf{D}_2$: $k = 3$ and $n = 10$

Run	A	B	C
1	-1	-1	-1
2	1	1	-1
3	1	-1	1
4	-1	1	1
5	1	-1	-1
6	-1	1	-1
7	-1	-1	1
8	-1	0	0
9	0	-1	0
10	0	0	-1

Hoke compared his designs with Box-Behnken and SCD designs of comparable size based upon the $tr(\mathbf{X}'\mathbf{X})^{-1}$ ($A-$optimality) and $|\mathbf{X}'\mathbf{X}|$ ($D-$optimality) criteria and concluded that his designs compared favorably (Khuri and Cornell, 1996).

2.4.6 Other Minimal Run Designs.

Angelopoulos et al. (2009) presented a class of balanced, near rotatable second-order designs which minimized the number of factorial runs associated with a CCD suitable for a spherical region of interest. Their designs were determined by searching through designs with $0(mod2)$ factorial runs (i.e., keep an even number of runs)

Table 10. Hoke Design D_6: $k = 3$ and $n = 13$

Run	A	B	C
1	-1	-1	-1
2	1	1	-1
3	1	-1	1
4	-1	1	1
5	1	-1	-1
6	-1	1	-1
7	-1	-1	1
8	-1	0	0
9	0	-1	0
10	0	0	-1
11	1	1	0
12	1	0	1
13	0	1	1

associated with a CCD for k factors and selecting the design with the lowest possible correlation among main effects. In order to discriminate between near rotatable CCDs, the Draper-Pukelsheim measure $Q*$ was applied (Angelopoulos et al., 2009). Unfortunately, it was determined the α-value for the CCD axial runs which provided the maximum $Q*$ value, did so at the expense of efficiently estimating all the parameters for a second-order model.

2.5 Supersaturated Designs (SSD)

Supersaturated designs are a type of fractional factorial design where the number of factors k under investigation exceeds the number of available experimental runs N. Since $k > N - 1$, the degrees of freedom within the design are insufficient to estimate all the main effects and the design matrix cannot be orthogonal. Therefore in order for supersaturated designs to useful as screening designs only a few factors can be active. As such supersaturated designs are generally used when the number of potential factors is large but few are believed to have actual effects (effect sparsity) and either budget or time constraints limit the number of experimental runs.

Since Satterthwaite (1959) first introduced the supersaturated design as a random balanced design, research has focused in primarily three areas: design construction methods, development of criterion to assess supersaturated designs, and data analysis methods used to identify the important effects.

Booth and Cox (1962) provided seven supersaturated designs for two-level factors created via computer search using the $E(s^2)$ criterion, which measures the average correlation between design columns. Overall, a general design construction method did not exist until Lin (1993) developed a method based upon half fractions of Hadamard matrices. Subsequently, Lin (1995), Nguyen (1996) and Li and Wu (1997) have proposed methods based upon the $E(s^2)$ criterion while Jones et al. (2008) constructed designs using Bayesian $D-$optimality.

Yamada and Lin (1999) and Yamada et al. (1999) were the first to discuss and provide construction methods for three-level supersaturated designs. Fang et al. (2000) first addressed the construction of multi-level supersaturated designs. More recently, Yamada et al. (2006) detailed a general construction method for mixed-level supersaturated designs. Overall beyond construction methods, little else has been done with three-level SSD.

Closely related to the manner and method of which supersaturated designs are created is the evaluation criteria used to differentiate amongst these design methods. $E(s^2)$ optimality is still the most widely used criterion for selection of supersaturated designs, but a Bayesian D-optimality criteria has also been used. Beattie et al. (2002) detailed an alternative two-stage Bayesian model selection strategy by combining a stochastic search variable selection method and an intrinsic Bayes factor method.

Similar to the number of ways to assess design quality, there are many ways of analyzing the data recorded during the experimental runs to identify the important effects/factors. Three methods include stepwise selection procedures, the Gauss-

Dantzig selector, and model averaging. Candes and Tao (2007) developed the Gauss-Dantzig selector which was further expanded upon by Phoa et al. (2009) who proposed a graphical procedure and an automatic variable selection method to accompany the Dantzig selector. Marley and Woods (2010) evaluated the use of $E(s^2)-$optimal and Bayesian $D-$Optimal designs and the three analysis strategies through the use of simulation-based experiments.

Overall, SSDs offer a method of greatly reducing the amount of experimentation needed to screen important factors. However, they should not be used without a clear understanding of the risks involved. SSDs are nonregular factorial designs, in which, orthogonality is not obtainable. Most existing criteria for SSDs measure the non-orthogonality combinatorially between two factors. Some care must be exercised in the selection of a design, however, as a supersaturated design that departs considerably from an orthogonal design could produce misleading results. This can especially occur if the departure from orthogonality is more than slight. $E(s^2)$ gives an intuitive measure of nonorthogonality where smaller is better. In particular, stepwise regression is one method that has been used for identifying the effects that should be estimated, but stepwise regression can easily fail to make the appropriate determination when the correlation between the columns in the design are not quite small.

2.6 Second-Order Screening Designs

In contrast with the traditional sequential design approach of response surface methodology (RSM), recent literature has proposed employing a single experimental design capable of preforming both factor screening and response surface exploration when conducting multiple experiments is unrealistic due to time, budget, or other constraints. For instance, in agricultural settings the time duration of the design can be exceedingly long. Also within a manufacturing setting experimental preparation can be overly time-consuming. Directly applicable to the DOD, Lawson (2003) points out fixed deadlines for scale up and production of prototype engineering designs may not allow the possibility of follow-up experimentation.

Two important principles used in developing successful screening designs are sparsity and heredity. The sparsity principle stems from the Pareto principle which states that most of the variability in a system or process output is due to a small number of inputs. Traditionally, *factor sparsity* has led to the assumption in screening designs that only a small number of factors are present among the actual model terms, while *effect sparsity* indicates that the number of active effects compared to active factors is relatively small. Therefore, it is possible for the effect sparsity assumption to hold while factor sparsity does not.

Heredity, either strong or weak, is the second screening principle commonly used when considering model selection. Strong heredity implies that if a model includes a two-factor interaction, then its constituent main effects are included in the model. Conversely, weak heredity requires only one of the two constituent main effects be included in the model.

Initial attempts to use response surface designs capable of performing both factor screening and response surface exploration with a single design relied upon the design's projection capacity.

Cheng and Wu (2001), hereafter referred to as CW, introduced a two-stage analysis method where the first stage consisted of performing factor screening analysis to identify important factors and the second stage involved fitting a second-order model by assuming both *factor sparsity* and *strong effect heredity* held and the region chosen for factor screening contained the optimal response surface area.

For the first stage, CW recommended a main effect analysis method for simplicity. The key linkage between stage one and two was the ability to project the initial larger factor space onto a smaller factor space capable of fitting a second-order model. When the *factor sparsity* principle holds, any regular fractional factorial design of resolution R projects onto any subset of $R-1$ factors as a full factorial. For example, a 2^{3-1}_{III} design ($R = 3$) can project into a 2^2 design in every subset of two factors (Myers and Anderson-Cook, 2009). This projection property extends to nonregular designs like Plackett-Burman designs by Lin and Draper (1992) and Wang and Wu (1995).

Because a design can project onto many different combinations of factors, a projection-efficiency criterion was developed to compare orthogonal designs based upon (1) the number of eligible projected designs with lower-dimension projections being more important than higher-dimension projections and (2) the estimation efficiency for eligible projected designs determined by the ratio of each designs $D-$ and $G-$efficiences (Cheng and Wu, 2001). *Eligible* designs are designs which can fit a second-order model and the $D-$ and $G-$efficiences, denoted D_{eff} and G_{eff} respectively, criteria compare the performance of a design against a corresponding optimal design (Myers and Anderson-Cook, 2009).

CW studied three orthogonal array (OA) designs ($OA(18, 3^7)$, $OA(27, 3^8)$, and $OA(36, 3^{12})$) which demonstrated desirable projection properties. The $OA(N, 3^k)$ connotation shows the design's number of runs N and number of factors k. In contrast to 3^{n-k} designs which have defining contrast subgroups to describe the design structure, the $OA(N, 3^k)$ designs studied by CW required computer search to classify the possible projected designs. Fortunately, while more complex, the overall projection properties are better and generally required fewer runs. When compared to CCDs, the $OA(N, 3^k)$ designs studied exhibited good $D-$efficiences but poor $G-$efficiences as p, number of projected factors, increases. However, this should be expected because as p increases, the size of CCDs increases while the size for the $OA(N, 3^k)$ designs is fixed for any p.

Improving on the designs of CW, Xu et al. (2004), hereafter referred to by XCW, proposed a combinatorial method for constructing new and efficient OA designs and a design selection approach based upon a *projection aberration criterion* which combines the generalized word-length pattern of the generalized minimum aberration criterion (Xu and Wu, 2001) for factor screening and the projection-efficiency criteria (Cheng and Wu, 2001) for interaction detection. XCW assessed the projection performance of three combinatorially non-isomorphic $OA(18, 3^7)$s and three combinatorially non-isomorphic $OA(27, 3^{13})$s. Their three-step approach involves: (1) screening out poor orthogonal arrays (OA) for factor screening using the generalized word-length pattern, (2) applying the projection aberration criterion to select a best design from step 1, and (3) determining the best level permutations of the design from step 2 to improve design projection eligibility and estimation efficiency under the second-order model 1.2.

Ye et al. (2007), hereafter referred to as YTL, also examined 3-level 18-run and 27-run orthogonal designs. However, in addition to considering the projection properties of designs, their design choices were based on both model estimation and model discrimination criteria. The two model estimation criteria employed examine the proportion of estimable models, *Estimation Capacity (EC)*, and average $D-$efficiency of all models, *Information Capacity (IC)*. Defining the design D, the space of models F over D with F', $F' \subset F$, the subset of estimable models over D and $f_i \epsilon F'$ then Jones et al. (2007) proposed six non-Bayesian criteria for model discrimination of which YTL employed the *Average Expected Prediction Differences (AEPD)*

$$AEPD = \frac{1}{\binom{d'}{2}} \sum_{\mathbf{f}_i, \mathbf{f}_j \epsilon F'(D)} E(\|\hat{\mathbf{y}}_\mathbf{i} - \hat{\mathbf{y}}_\mathbf{j}\|^2 | \|\mathbf{y}\| = 1) \tag{2.4}$$

and *Minimum Maximum Prediction Difference (MMPD)*

$$MMPD = \mathbf{min}_{1 \leq i < j \leq n} \max_{\|\mathbf{y}=1\|} \|\mathbf{\hat{y}_i} - \mathbf{\hat{y}_j}\| \tag{2.5}$$

where d' is the number of estimable models, $\mathbf{y}$ is the response vector, and $\mathbf{\hat{y}_i}$ is the fitted value of the ith model Ye et al. (2007).

While previous work focused primarily on the design's projection capacity, Edwards and Truong (2011) applied the Jones and Nachtsheim (2011b) method for finding efficient designs with minimal aliasing between main effects and two-factor interactions. Deemed MA designs, Edwards and Truong (2011) constructed 18, 27, and 30-run designs for simultaneous screening and response surface optimization for $k = 4$ to 7, $k = 4$ to 13, and $k = 6$ to 14 factors, respectively, by minimizing the sum of squares of the elements of the alias matrix, $\mathbf{A}$, subject to a lower bound on the primary model $D-$efficiency. The optimization of interest is

$$\min_d \text{Tr}[\mathbf{A}(d)'\mathbf{A}(d)], \text{subject to} D_e(d) \geq l_D, \tag{2.6}$$

where $\mathbf{A}(d)$ is the alias matrix for design d, $D_e(d)$ is the D-efficiency of design d, and l_D denotes the lower bound for D-efficiency and $0 < l_D \leq 1$ (Jones and Nachtsheim, 2011b). Edwards and Truong (2011) compared the 27-run orthogonal arrays of XCW and YTL with MA designs generated with l_D values of 0.8 and 0.9 in terms of D-efficiency of projection and, via a simulation study, the proportion of active factors declared significant (Power 1) and the proportion of simulations in which only the true active factors are declared significant (Power 2). Although ranked last in terms of D-efficiency, the MA designs showed superior performance with their ability to detect active factors (Edwards and Truong, 2011).

A common thread connecting all CW, XCW, YTL, and MA designs is the use of a linear and quadratic main-effects only analysis for factor screening. Unfortunately, if the *strong effect heredity* principle fails to hold important interactions can be missed leading to a misspecified response surface model. Edwards and Truong (2011) confirmed this assertion for their designs and the XCW designs through a simulation model possessing only *weak effect heredity*. All four research efforts, CW, XCW, YTL, and MA, acknowledge that while the strong effect heredity assumption could be overly restrictive, they feel the inclusion of quadratic main effects diminishes the concern. However, if the concern exists where a factor's significance is only present in interactions with other factors, the authors proposed either the Bayesian approaches of Box and Meyer (1993) or Chipman et al. (1997) to account for significant factors outside of main effects when the *strong effect heredity* principle fails to hold (Cheng and Wu, 2001). Unfortunately, these methods are not readily available to practitioners in statistical software packages and are computationally intensive procedures, thus likely making their use impractical (Edwards and Truong, 2011). Therefore potential research efforts could focus on new or inventive analysis techniques.

Another area of concern for the CW, XCW, YTL, and MA designs is the projection of main and/or quadratic effects deemed significant during the first stage analysis does not always yield a second-order design. CW highlighted this concern using an illustrative example of a 27-run experiment with nine continuous factors (3^{9-6}). During the main and quadratic effects screening analysis, CW identified five important factors, unfortunately, there are no eligible projected designs of five factors in the 3^{9-6} design. As a result, a subset of the five factors must be considered when a single experiment is used and important effects could be missed.

Edwards and Mee (2011) introduced new spherical Fractional Box-Behnken designs (FBBD) aimed at overcoming the projection deficiencies and main/quadratic effect only analysis issues found in the CW/XCW/YTL/MA designs. The FBBD provide the ability to explore interactions during the screening stage and to fit second-order models via a backward elimination analysis strategy to each of the $(k-1)$-factor projections. Edwards and Mee (2011) questioned the applicability of the factor sparsity principle assumed by CW/XCW/YTL/MA designs preferring instead the idea of effect sparsity when many factors are under consideration. By effect sparsity, Edwards and Mee (2011) meant the number of active effects vice factors is relatively small. Since it is possible for effect sparsity to hold, even when factor sparsity does not, Edwards and Mee (2011) determined it was necessary to search for designs having larger factor eligible projections than the maximum $p = 5$ factor projections provided with the CW/XCW/YTL/MA designs.

The FBBDs are developed by taking subsets of the two-level fractional factorial designs which compose a BBD (Edwards and Mee, 2011). The number of runs associated with the FBBD vary depending upon the number of factors involved. For $k \leq 9$, the FBBD are saturated/near-saturated response surface designs, but for $k = 10 \ldots 13$, the FBBD are reduced run designs. While FBBDs require more runs than CW/XCW/YTL/MA designs, their ease of construction and aliasing structure facilitate an analysis strategy which cannot be applied to the CW/XCW/YTL/MA designs. Additionally, as k increases, the FBBD designs require fewer runs than CCD/BBD.

Jones and Nachtsheim (2011a) introduced a class of three-level designs referred to as "definitive screening designs" where main effects are not biased by second-order effects and all quadratic effects are estimable. Consisting of $2k+1$ runs for k factors, these designs were constructing using the same Jones and Nachtsheim (2011b) method used by Edwards and Truong (2011).

2.7 Chapter Summary

Experimental designs normally recommended for screening and optimization experiments differ. If an experimenter can afford to run only one experiment, a choice must be made between one objective or the other. If the experimenter chooses a classical response surface design, a subset of the factors must be selected to work with and the chance of missing other important factors and improvement possibilities increases. If the experimenter decides to conduct a screening experiment, important interactions and quadratic effects may be missed that could lead to further process improvements and cost reductions.

Examination of classical response surface designs show the CCD and FCD as efficient second-order designs, particularly when compared with the 3^k factorial, which can accommodate a spherical and cuboidal region, respectively, through appropriate design parameter selection while not requiring an unusually large number of design points. The efficiency of the second-order BBD is comparable to the CCD. However, the BBD only accommodates a spherical design region ignoring the “extreme” corner factorial points. This can be beneficial if the operational region does not permit corner points. Unfortunately, as the number of factors k increases, so does the run size of these designs. Whereas both the CCD and FCD can reduce their run size requirements through the use of fractional factorials, while sacrificing the number of estimable effects, the ability to reduce run size requirements for BBD has seen little work. One proposal makes use of replacing the standard 2^3 designs in a BBD with a combination of 2^3 and 2^{3-1}_{III} designs. However, the reduction in run size requirements has a corresponding reduction in parameter estimation efficiency (Zhang et al., 2011).

When cost constraints restrict the design size to levels at or equal to the number of parameters in a second-order design, hybrid designs for a spherical region and Hoke designs for a cuboidal region are typically a better option than either Box-Draper or SCD because they generally provide better efficiency (Myers and Anderson-Cook, 2009). Unfortunately, hybrid and Hoke designs have only been specified for up to $k = 7$ and 6 factors respectively. This could inhibit their usefulness, particularly if a factor screening design cannot be used to eliminate insignificant factors. Hybrid designs currently involve the use of a 2^k factorial. A potential expansion of hybrid designs could include using 2^{k-p} factorial designs for large values of k. Additionally, since these designs are currently using full factorials, the idea of design projectivity of the reduced run designs could be addressed. Unfortunately, the nature of hybrid designs results in odd factor levels which could be difficult to obtain in practice. It is in this instance that computer generated designs can prove most useful for even saturated designs.

Contrary to classical response surface designs, examination of screening designs finds focus primarily on examining main effects, potentially at the expense of important interactions and quadratic effects. While there is an abundant amount of research dealing with the construction methods, evaluation criteria, and data analysis methods used to identify the important effects for supersaturated designs, the majority of research is based upon the underlying response model being linear (first-order) in nature. Matsuura et al. (2011) constructed a supersaturated design using a Hadamard matrix and proposed its use in robust parameter design. The design was compared with a $D-$optimal design, a Central Composite design, and a Box-Behnken design. With considerably fewer experimental runs, as compared with CCDs and BBDs, the design demonstrated the capacity to identify main, two-factor interaction, and pure quadratic effects of active factors under the effect sparsity assumption. Matsuura et al. (2011) work currently is the only published findings which directly address the consideration of quadratic effects.

In the instances where neither a single screening nor a response surface design can fulfill experimental objectives, second-order screening designs have been proposed which screen factors beyond only main effects and provide the capacity to estimate a second-order response model within a subset of the original factors, simultaneously. In order to do so, assumptions such as factor or effect sparsity and effect heredity, to a varying degree, are made to facilitate data analysis. To what degree these assumptions hold has been debated and therefore could influence the use of these designs. As a result, a thorough comparison of these designs, as these assumptions are relaxed, could provide insight into the various designs' robustness.

III. Effect of Heredity and Sparsity on Second-Order Screening Design Performance

3.1 Introduction

Box and Wilson (1951) laid the foundation for response surface methodology (RSM) by outlining a philosophy of sequential experimentation which included experiments for screening, region seeking (such as steepest ascent), process/product characterization, and process/product optimization (Myers et al., 2004). Box and Liu (1999) illustrated a number of concepts which Box understood as the embodiment of RSM at the time to include the philosophy of sequential learning.

In contrast with the traditional sequential design approach of RSM, recent literature has proposed employing a single experimental design capable of preforming both factor screening and response surface exploration when conducting multiple experiments is unrealistic due to time, budget, or other constraints. For instance, in agriculture the time required to collect data specified by the design can be exceedingly long. Within a manufacturing setting experimental preparation can be overly time-consuming. Directly applicable to the DoD, Lawson (2003) points out fixed deadlines for scale up and production of prototype engineering designs may not allow the possibility of follow-up experimentation.

Military systems, particularly aerodynamic systems, are complex. It is not unusual for these systems to exhibit nonlinear behavior. Developmental testing may be tasked to characterize the nonlinear behavior of such systems but are also restricted in how much testing can be accomplished. In these instances, the single experimental design may be the preferred approach.

Second-order screening design methodology, sometimes referred to as One-Step RSM or Definitive Screening, is a relatively new focus in statistical research and effectively unknown to the defense test community. Important questions as to the methods' usefulness and applicability to Defense testing remain unaddressed but nonetheless second-order screening designs for nonlinear system responses provide a means to effectively focus test resources onto those factors driving system performance.

Two important principles used in developing successful screening designs are sparsity and heredity. The sparsity principle stems from the Pareto principle which states that most of the variability in a system or process output is due to a small number of inputs. Traditionally, *factor sparsity* has led to the assumption in screening designs that only a small number of factors are present among the actual model terms. However, the degree to which factor sparsity holds as the number of factors grows has resulted in a debate between *effect sparsity* and factor sparsity. Effect sparsity indicates that the number of active effects compared to active factors is relatively small. Therefore, it is possible for the effect sparsity assumption to hold while factor sparsity does not.

Heredity, either strong or weak, is another screening principle commonly used when considering model selection. Strong heredity implies that if a model includes a two-factor interaction, then both its constituent main effects are also included in the model. Conversely, weak heredity requires only one of the two constituent main effects be included in the model.

Edwards and Truong (2011) preformed a simulation study examining several second-order screening designs focusing on each design's ability to correctly identify active factors under a variety of conditions. While 5000 responses vectors were simulated for several combinations of coefficient magnitudes, the truth models used assumed both factor sparsity and strong effect heredity. This article formally examines the robustness of the two arguably best second-order screening designs with respect to the assumptions of both sparsity (factor or effect) and heredity (strong or weak).

The remainder of this paper is organized as follows. Section 3.2 discusses the literature relevant to second-order screening designs. In Section 3.3, we present an empirical study that quantifies the robustness of the two second-order screening designs to assumptions of heredity and sparsity. Section 3.4 provides a discussion on the tradeoffs in selecting among the two designs and Section 3.5 presents a summary of the conclusions.

3.2 Second Order Screening Designs

Initial attempts to use response surface designs capable of performing both factor screening and response surface exploration with a single design relied upon the design's projection capacity.

Cheng and Wu (2001), hereafter referred to as CW, introduced a two-stage analysis method where the first stage consisted of performing factor screening analysis to identify important factors and the second stage involved fitting a second-order model by assuming both *factor sparsity* and *strong effect heredity* held and that the region chosen for factor screening contained the optimal response surface area.

For the first stage, CW recommended a main effects analysis method for simplicity purposes. The key linkage between stage one and two was the ability to project the initial larger factor space onto a smaller factor space capable of fitting a second-order model. When the *factor sparsity* principle holds, any regular fractional factorial design of resolution R projects onto any subset of $R-1$ factors as a full factorial. For example, a 2^{3-1}_{III} design ($R = 3$) can project into a 2^2 design (Myers and Anderson-Cook, 2009). This projection property extends to nonregular designs like Plackett-

Burman designs as discussed by both Lin and Draper (1992) and Wang and Wu (1995).

Since a larger design can project onto many different combinations of factors, a projection-efficiency criterion was developed to compare orthogonal designs based upon (1) the number of eligible projected designs with lower-dimension projections being more important than higher-dimension projections and (2) the estimation efficiency for eligible projected designs determined by the ratio of each design's $D-$ and $G-$efficiences (Cheng and Wu, 2001). *Eligible* designs are designs which can fit a second-order model and have calculated $D-$ and $G-$efficiences, denoted D_{eff} and G_{eff}, respectively, to compare the performance of that design against a corresponding optimal design (Myers and Anderson-Cook, 2009).

CW studied three orthogonal array (OA) designs ($OA(18, 3^7)$, $OA(27, 3^8)$, and $OA(36, 3^{12})$) each of which demonstrated desirable projection properties. The $OA(N, 3^k)$ connotation shows the design's number of runs N and number of factors k. In contrast to 3^{n-k} designs which have defining contrast subgroups to describe the design structure, the $OA(N, 3^k)$ designs studied by CW required computer search to classify the possible projected designs. Fortunately, while design generation is more complex, the overall projection properties are better and generally required less experimental runs. When compared to Central Composite Designs (CCD), the $OA(N, 3^k)$ designs studied exhibited good $D-$efficiences but poor $G-$efficiences as p, number of projected factors, increases. However, this is to be expected because as p increases, the size of CCDs increases while the size for the $OA(N, 3^k)$ designs is fixed for any p.

Improving on the designs of CW, Xu et al. (2004), hereafter referred to by XCW, proposed a combinatorial method for constructing new and efficient OA designs and a design selection approach based upon a *projection aberration criterion* which combines the generalized word-length pattern of the generalized minimum aberration criterion

(Xu and Wu, 2001) for factor screening and the projection-efficiency criteria (Cheng and Wu, 2001) for interaction detection. XCW assessed the projection performance of three combinatorially non-isomorphic $OA(18, 3^7)$s and three combinatorially non-isomorphic $OA(27, 3^{13})$s. Their three-step approach involves: (1) screening out poor orthogonal arrays for factor screening using the generalized word-length pattern, (2) applying the projection aberration criterion to select a best design from step (1), and (3) determining the best level permutations of the design from step (2) to improve design projection eligibility and estimation efficiency under the second-order model:

$$\eta = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j. \tag{3.1}$$

Ye et al. (2007), hereafter referred to as YTL, also examined 3-level 18-run and 27-run orthogonal designs. However, in addition to considering the projection properties of designs, their design choices were based on both model estimation and model discrimination criteria. The two model estimation criteria employed examine the proportion of estimable models, *Estimation Capacity (EC)*, and average $D-$efficiency of all models, *Information Capacity (IC)*. Defining the design D, the space of models F over D with F', $F' \subset F$, the subset of estimable models over D and $f_i \epsilon F'$ then Jones et al. (2007) proposed six non-Bayesian criteria for model discrimination of which YTL employed the *Average Expected Prediction Differences (AEPD)*

$$AEPD = \frac{1}{\binom{d'}{2}} \sum_{\mathbf{f}_i, \mathbf{f}_j \epsilon F'(D)} E(\|\hat{\mathbf{y}}_\mathbf{i} - \hat{\mathbf{y}}_\mathbf{j}\|^2 | \|\mathbf{y}\| = 1) \tag{3.2}$$

and *Minimum Maximum Prediction Difference (MMPD)*

$$MMPD = \mathbf{min}_{1 \leq i < j \leq n} \max_{\|\mathbf{y}=1\|} \|\hat{\mathbf{y}}_\mathbf{i} - \hat{\mathbf{y}}_\mathbf{j}\| \tag{3.3}$$

where d' is the number of candidate models, $\mathbf{y}$ is the response vector, and $\hat{\mathbf{y}}_{\mathbf{i}}$ is the fitted value of the ith model (Ye et al., 2007).

While previous work focused primarily on the design's projection capacity, Edwards and Truong (2011) applied the Jones and Nachtsheim (2011b) method for finding efficient designs with minimal aliasing between main effects and two-factor interactions. Deemed MA designs, Edwards and Truong (2011) constructed 18, 27, and 30-run designs for simultaneous screening and response surface optimization for $k = 4$ to 7, $k = 4$ to 13, and $k = 6$ to 14 factors, respectively, by minimizing the sum of squares of the elements of the alias matrix, $\mathbf{A}$, subject to a lower bound on the primary model $D-$efficiency. Their optimization of interest is

$$\min_d \text{Tr}[\mathbf{A}(d)'\mathbf{A}(d)], \text{subject to} D_e(d) \geq l_D, \tag{3.4}$$

where $\mathbf{A}(d)$ is the alias matrix for design d, $D_e(d)$ is the D-efficiency of design d, and l_D denotes the lower bound for D-efficiency with $0 < l_D \leq 1$ (Jones and Nachtsheim, 2011b). Edwards and Truong (2011) compared the projection D-efficiency and, via a simulation study, the proportion of active factors declared significant (Power 1) and the proportion of simulations in which only the true active factors are declared significant (Power 2) of the 27-run orthogonal arrays of XCW, YTL, and MA designs generated with l_D values of 0.8 and 0.9. Although ranked last in terms of D-efficiency, the MA designs showed superior performance in their ability to detect active factors (Edwards and Truong, 2011).

A common thread connecting all CW, XCW, YTL, and MA designs is the use of a linear and quadratic main-effects only analysis for factor screening. Unfortunately, if the *strong effect heredity* principle fails to hold important interactions can be missed leading to a misspecified response surface model. Edwards and Truong (2011) confirmed this assertion for their designs and the XCW designs using a simulation model possessing only *weak effect heredity*. All four research efforts, CW, XCW, YTL, and MA, acknowledge that while the strong effect heredity assumption could be overly restrictive, they feel the inclusion of quadratic main effects diminishes the concern. However, Edwards and Truong (2011) proposed either the Bayesian approaches of Box and Meyer (1993) or Chipman et al. (1997) to account for significant factors outside of main effects when the *strong effect heredity* principle fails to hold (Cheng and Wu, 2001). Unfortunately, these methods are not readily available in statistical software packages and are computationally intensive procedures, thus likely limiting widespread use (Edwards and Truong, 2011). Therefore potential research efforts could focus on new analysis techniques.

Another area of concern for the CW, XCW, YTL, and MA designs is that the projection of main and/or quadratic effects deemed significant during the first stage analysis do not always yield a second-order design. CW highlighted this concern using an illustrative example of a 27-run experiment with nine factors (3^{9-6}). During the main and quadratic effects screening analysis, CW identified five important factors; unfortunately, there are no eligible projected designs of five factors in the 3^{9-6} design. As a result, a subset of the five important factors must be considered in order to have enough degrees of freedom to fit a full second-order model. Since not all the important factors can be used important effects could be missed.

Edwards and Mee (2011) introduced new spherical Fractional Box-Behnken designs (FBBD) aimed at overcoming the projection deficiencies and main/quadratic effect only analysis issues found in the CW/XCW/YTL/MA designs. The FBBD provides the ability to explore interactions during the screening stage and to fit second-order models via a backward elimination analysis strategy to each of the $(k-1)$-factor projections. Edwards and Mee (2011) questioned the applicability of the factor sparsity principle assumed by CW/XCW/YTL/MA designs preferring instead the idea of effect sparsity when many factors are under consideration. Since it is possible for effect sparsity to hold, even when factor sparsity does not, Edwards and Mee (2011) determined it was necessary to search for designs having larger factor eligible projections than the maximum $p = 5$ factor projections provided with the CW/XCW/YTL/MA designs.

The FBBDs are developed by taking subsets of the two-level fractional factorial designs which compose a Box-Behnken Design (BBD) (Edwards and Mee, 2011). The number of runs associated with the FBBD vary depending upon the number of factors involved. For $k \leq 9$, the FBBD are saturated/near-saturated response surface designs, but for $k = 10, \ldots, 13$, the FBBD are reduced run designs. While FBBDs require more runs than CW/XCW/YTL/MA designs, their ease of construction and aliasing structure facilitate an analysis strategy which cannot be applied to the CW/XCW/YTL/MA designs. Additionally, as k increases, the FBBD designs require fewer runs than CCD/BBD designs.

Jones and Nachtsheim (2011a) introduced a class of three-level designs referred to as "definitive screening designs" where main effects are not biased by second-order effects and all quadratic effects are estimable. Consisting of $2k+1$ runs for k factors, these designs were constructed using the same Jones and Nachtsheim (2011b) method used by Edwards and Truong (2011).

3.3 Empirical Study

Our empirical study examines the nine-factor designs, identified as $(1/2)BB9.1$ in Table 4 of Edwards and Mee (2011) and the definitive screening design generated using conference matrices based on Xiao et al. (2012). The study focus is on each design's robustness to detect important effects in models exhibiting different combinations of heredity and sparsity. A single replication is investigated in depth for each scenario where the truth model terms and coefficients were chosen to be a variation on the original nine-factor model considered by Edwards and Mee (2011). Additionally, each design is analyzed using the author's recommended analysis methodologies. Summary statistics involving replications for each model are then provided and discussed.

The Edwards and Mee (2011) analysis methodology involves performing a factor-based backward elimination to identify a possible second-order model. Since there are not enough degrees of freedom to fit a full second-order model for all the factors under consideration, the first step assumes at least one factor can be omitted from the second-order model containing all $k = 9$ factors. Therefore the root mean-square error (RMSE) for each of the 9 $(k - 1 = 8)$-factor second-order models are compared with the model yielding the smallest RMSE being selected and thus identifying which factor is omitted. Subsequent steps involve determining if any additional factors may be removed based upon whether all effects, to include main, quadratic, and two-way interactions, in the second order model containing the factor are negligible.

Jones and Nachtsheim (2011a) suggest using a forward stepwise regression, which considers all terms in a second-order model of $k = 9$ factors. With a p-value of 0.1 to enter, effects are added into the second-order model while forcing a strong heredity model. As such, when either two-factor interactions or pure-quadratic effects are included in the model, the lower order terms must also be included.

Four cases are considered to represent different combinations of model heredity (strong or weak) and sparsity (factor or effect). In addition, each model is examined with four different noise level scenarios. The noise level vector used in each scenario is identical across each model for each design. The 49 treatment combinations for the Edwards and Mee (2011) design and the 21 treatment combinations for the Jones and Nachtsheim (2011a) design are given in Tables 11 and 12, respectively. Tables 13 and 14 show the simulated response values for the 16 combinations of case and noise level scenario for Edwards and Mee (2011) and Jones and Nachtsheim (2011a), respectively.

Case 1 data was simulated based on the model

$$y_i = 2A_i - 1.5E_i + 2G_i - 3A_i^2 + 2.5E_i^2 - 4G_i^2 + 4A_iE_i + 3.5A_iG_i - 5E_iG_i + \varepsilon_i, \quad (3.5)$$

thereby representing a response which exhibits factor sparsity and strong heredity between active two-factor interactions or pure quadratic effects and main effects. The model exhibits factor sparsity as only 3 of the 9 factors are active within the 9 effects contained in the model.

The Edwards and Mee (2011) analysis method first performs a factor-based backward elimination to identify a possible second-order model. Table 15 shows the backward elimination steps for the Table 13, Case 1 data for all four noise level scenarios using $\alpha = 0.05$.

For example, when considering Case 1, Scenario 3, where $\varepsilon_i \sim N(0, 3)$, the eight-factor second-order model that omits F has the smallest RMSE among all the eight-factor models and factor J contributions as a main, quadratic, or as a part of a two-way interaction are negligible.

After identifying which factors can be removed, Edwards and Mee (2011) fit a full second-order model in the remaining factors. Again when considering scenario three, a full second-order model is fit using the remaining factors: A, B, C, D, E, G, and H.

In contrast to Edwards and Mee (2011) factor-based backward elimination analysis method, Jones and Nachtsheim (2011a) perform forward stepwise regression with a p-value of 0.1 to enter while forcing a strong heredity model. Table 16 shows the forward stepwise regression steps for the Case 1 data for all four noise level scenarios.

Since the "combined" option rule is used for the forward stepwise regression, the inclusion of two-way interaction or pure quadratic effects result in the inclusion of

Table 11. Nine-Factor Fractional Box-Behnken Design (FBBD)

A	B	C	D	E	F	G	H	J
1	1	1	0	0	0	0	0	0
1	-1	-1	0	0	0	0	0	0
-1	1	-1	0	0	0	0	0	0
-1	-1	1	0	0	0	0	0	0
1	0	0	1	0	0	1	0	0
1	0	0	-1	0	0	-1	0	0
-1	0	0	1	0	0	-1	0	0
-1	0	0	-1	0	0	1	0	0
1	0	0	0	1	0	0	0	1
1	0	0	0	-1	0	0	0	-1
-1	0	0	0	1	0	0	0	-1
-1	0	0	0	-1	0	0	0	1
1	0	0	0	0	1	0	1	0
1	0	0	0	0	-1	0	-1	0
-1	0	0	0	0	1	0	-1	0
-1	0	0	0	0	-1	0	1	0
0	1	0	1	0	0	0	0	1
0	1	0	-1	0	0	0	0	-1
0	-1	0	1	0	0	0	0	-1
0	-1	0	-1	0	0	0	0	1
0	1	0	0	1	0	0	1	0
0	1	0	0	-1	0	0	-1	0
0	-1	0	0	1	0	0	-1	0
0	-1	0	0	-1	0	0	1	0
0	1	0	0	0	1	1	0	0
0	1	0	0	0	-1	-1	0	0
0	-1	0	0	0	1	-1	0	0
0	-1	0	0	0	-1	1	0	0
0	0	1	1	0	0	0	1	0
0	0	1	-1	0	0	0	-1	0
0	0	-1	1	0	0	0	-1	0
0	0	-1	-1	0	0	0	1	0
0	0	1	0	1	0	1	0	0
0	0	1	0	-1	0	-1	0	0
0	0	-1	0	1	0	-1	0	0
0	0	-1	0	-1	0	1	0	0
0	0	1	0	0	1	0	0	1
0	0	1	0	0	-1	0	0	-1
0	0	-1	0	0	1	0	0	-1
0	0	-1	0	0	-1	0	0	1
0	0	0	1	1	1	0	0	0
0	0	0	1	-1	-1	0	0	0
0	0	0	-1	1	-1	0	0	0
0	0	0	-1	-1	1	0	0	0
0	0	0	0	0	0	1	1	1
0	0	0	0	0	0	1	-1	-1
0	0	0	0	0	0	-1	1	-1
0	0	0	0	0	0	-1	-1	1
0	0	0	0	0	0	0	0	0

Table 12. Nine-Factor Definitive Screening Design (DSD)

A	B	C	D	E	F	G	H	J
0	1	1	1	1	1	1	1	1
0	-1	-1	-1	-1	-1	-1	-1	-1
1	0	-1	-1	-1	-1	1	1	1
-1	0	1	1	1	1	-1	-1	-1
1	-1	0	-1	1	1	-1	-1	1
-1	1	0	1	-1	-1	1	1	-1
1	-1	-1	0	1	1	1	1	-1
-1	1	1	0	-1	-1	-1	-1	1
1	-1	1	1	0	-1	-1	1	-1
-1	1	-1	-1	0	1	1	-1	1
1	-1	1	1	-1	0	1	-1	1
-1	1	-1	-1	1	0	-1	1	-1
1	1	-1	1	-1	1	0	-1	-1
-1	-1	1	-1	1	-1	0	1	1
1	1	-1	1	1	-1	-1	0	1
-1	-1	1	-1	-1	1	1	0	-1
1	1	1	-1	-1	1	-1	1	0
-1	-1	-1	1	1	-1	1	-1	0
1	1	1	-1	1	-1	1	-1	-1
-1	-1	-1	1	-1	1	-1	1	1
0	0	0	0	0	0	0	0	0

Table 13. Nine-Factor FBBD Simulated Response

Scenario	$\varepsilon \sim N(0,1)$				$\varepsilon \sim N(0,2)$				$\varepsilon \sim N(0,3)$				$\varepsilon \sim N(0,5)$			
Case	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
	-2.2647	4.7353	-0.7647	4.7353	-0.7154	6.2846	0.7846	6.2846	-6.2028	0.7972	-4.7028	0.7972	-0.1891	6.8109	1.3109	6.8109
	1.2486	0.2486	2.7486	0.2486	3.4210	2.4210	4.9210	2.4210	1.8927	0.8927	3.3927	0.8927	1.9998	0.9998	3.4998	0.9998
	-5.4854	-6.4854	-3.9854	1.5146	-5.5177	-6.5177	-4.0177	1.4823	-4.9026	-5.9026	-3.4026	2.0974	-8.6821	-9.6821	-7.1821	-1.6821
	-5.3763	1.6237	-3.8763	-6.3763	-4.3831	2.6169	-2.8831	-5.3831	-5.4348	1.5652	-3.9348	-6.4348	2.0404	9.0404	3.5404	1.0404
	-0.1641	3.3359	3.8359	3.3359	-0.1264	3.3736	3.8736	3.3736	4.5276	8.0276	8.5276	8.0276	-0.5919	2.9081	3.4081	2.9081
	-11.0650	-0.5650	-4.0650	-0.5650	-13.1572	-2.6572	-6.1572	-2.6572	-2.9519	7.5481	4.0481	7.5481	-13.2355	-2.7355	-6.2355	-2.7355
	-6.7809	-3.2809	1.2191	-3.2809	-6.9533	-3.4533	1.0467	-3.4533	-6.4563	-2.9563	1.5437	-2.9563	5.1335	8.6335	13.1335	8.6335
	-10.7867	-0.2867	-7.7867	-0.2867	-8.8595	1.6405	-5.8595	1.6405	-9.9887	0.5113	-6.9887	0.5113	-15.4562	-4.9562	-12.4562	-4.9562
	4.4535	3.4535	2.4535	3.4535	7.3352	6.3352	5.3352	6.3352	2.5334	1.5334	0.5334	1.5334	-6.3020	-7.3020	-8.3020	-7.3020
	-1.6281	5.3719	3.3719	5.3719	0.4617	7.4617	5.4617	7.4617	-2.8404	4.1596	2.1596	4.1596	0.5246	7.5246	5.5246	7.5246
	-7.3386	-0.3386	4.6614	-0.3386	-3.5807	3.4193	8.4193	3.4193	-8.5033	-1.5033	3.4967	-1.5033	-13.9843	-6.9843	-1.9843	-6.9843
	2.3133	1.3133	-6.6867	1.3133	3.1108	2.1108	-5.8892	2.1108	-0.5989	-1.5989	-9.5989	-1.5989	5.9438	4.9438	-3.0562	4.9438
	-0.8864	-0.8864	0.6136	-0.8864	1.9894	1.9894	3.4894	1.9894	-1.7556	-1.7556	-0.2556	-1.7556	7.1774	7.1774	8.6774	7.1774
	-0.2521	5.7479	1.2479	-0.2521	-0.0167	5.9833	1.4833	-0.0167	-1.2407	-7.2407	-5.7407	-7.2407	11.3126	5.3126	6.8126	5.3126
	-5.5337	0.4663	-4.0337	-5.5337	-2.7187	3.2813	-1.2187	-2.7187	2.4725	8.4725	3.9725	2.4725	-3.0930	2.907	-1.5930	-3.0930
	-5.8397	-5.8397	-4.3397	-5.8397	-3.8245	-3.8245	-2.3245	-3.8245	-8.0828	-8.0828	-6.5828	-8.0828	-4.2479	-4.2479	-2.7479	-4.2479
	1.7650	1.7650	1.7650	1.7650	-1.5778	-1.5778	-1.5778	-1.5778	-0.3906	-0.3906	-0.3906	-0.3906	4.6624	4.6624	4.6624	4.6624
	0.7927	0.7927	0.7927	0.7927	1.0561	1.0561	1.0561	1.0561	1.2474	1.2474	1.2474	1.2474	-0.0077	-0.0077	-0.0077	-0.0077
	-0.4940	-0.4940	-0.4940	-0.4940	-3.4910	-3.4910	-3.4910	-3.4910	-2.5399	-2.5399	-2.5399	-2.5399	-3.6832	-3.6832	-3.6832	-3.6832
	0.4495	0.4495	0.4495	0.4495	-3.1934	-3.1934	-3.1934	-3.1934	0.7770	0.7770	0.7770	0.7770	0.8996	0.8996	0.8996	0.8996
	1.8955	2.3955	5.3955	2.3955	2.0240	2.5240	5.5240	2.5240	2.0791	2.5791	5.5791	2.5791	-2.6628	-2.1628	0.8372	-2.1628
	2.2551	8.7551	-1.2449	2.7551	5.6349	12.1349	2.1349	6.1349	7.4457	13.9457	3.9457	7.9457	-2.1867	4.3133	-5.6867	-1.6867
	0.5852	0.0852	4.0852	-5.9148	0.9653	0.4653	4.4653	-5.5347	6.4065	5.9065	9.9065	-0.0935	0.7924	0.2924	4.2924	-5.7076
	5.3797	-1.1203	1.8797	-1.1203	5.7621	-0.7379	2.2621	-0.7379	0.9254	-5.5746	-2.5746	-5.5746	0.6438	-5.8562	-2.8562	-5.8562
	-3.0146	0.9854	-1.0146	0.9854	-1.9416	2.0584	0.0584	2.0584	-4.0317	-0.0317	-2.0317	-0.0317	-2.9282	1.0718	-0.9282	1.0718
	-7.3938	-3.3938	-1.3938	-3.3938	-5.8685	-1.8685	0.1315	-1.8685	-5.1637	-1.1637	0.8363	-1.1637	-5.8083	-1.8083	0.1917	-1.8083
	-3.6797	0.3203	2.3203	0.3203	-5.0174	-1.0174	0.9826	-1.0174	-10.5673	-6.5673	-4.5673	-6.5673	-12.6945	-8.6945	-6.6945	-8.6945
	-2.1660	1.8340	-0.1660	1.8340	-5.5268	-1.5268	-3.5268	-1.5268	1.1700	5.1700	3.1700	5.1700	-2.5402	1.4598	-0.5402	1.4598
	1.0891	2.0891	1.0891	-1.9109	-3.6814	-2.6814	-3.6814	-6.6814	-1.0757	-0.0757	-1.0757	-4.0757	-1.8311	-0.8311	-1.8311	-4.8311
	-0.0751	6.9249	-0.0751	-3.0751	2.1934	9.1934	2.1934	-0.8066	-4.9229	2.0771	-4.9229	-7.9229	-0.4678	6.5322	-0.4678	-3.4678
	-1.5848	-2.5848	-1.5848	-4.5848	1.5025	0.5025	1.5025	-1.4975	0.7182	-0.2818	0.7182	-2.2818	-0.5162	-1.5162	-0.5162	-3.5162
	-1.3903	-8.3903	-1.3903	-4.3903	-1.3487	-8.3487	-1.3487	-4.3487	1.8304	-5.1696	1.8304	-1.1696	-1.5580	-8.5580	-1.5580	-4.5580
	-6.7667	1.2333	-1.2667	-2.7667	-4.7074	3.2926	0.7926	-0.7074	-5.6048	2.3952	-0.1048	-1.6048	-6.8367	1.1633	-1.3367	-2.8367
	-7.7144	10.2856	-5.2144	6.2856	-9.3188	8.6812	-6.8188	4.6812	-11.0459	6.9541	-8.5459	2.9541	-7.0710	10.9290	-4.5710	6.9290
	-0.0974	-10.0974	9.4026	-6.0974	1.7609	-8.2391	11.2609	-4.2391	0.4167	-9.5833	9.9167	-5.5833	-4.2179	-14.2179	5.2821	-10.2179
	5.0307	5.0307	3.5307	9.0307	8.4558	8.4558	6.9558	12.4558	9.2045	9.2045	7.7045	13.2045	11.7458	11.7458	10.2458	15.7458
	-0.9775	3.0225	-0.9775	-0.9775	-2.0740	1.9260	-2.0740	-2.0740	-5.0188	-1.0188	-5.0188	-5.0188	-6.1467	-2.1467	-6.1467	-6.1467
	0.0149	4.0149	0.0149	0.0149	-2.0162	1.9838	-2.0162	-2.0162	-1.5257	2.4743	-1.5257	-1.5257	-4.3396	-0.3396	-4.3396	-4.3396
	0.6214	-3.3786	0.6214	0.6214	-0.1437	-4.1437	-0.1437	-0.1437	-2.4711	-6.4711	-2.4711	-2.4711	7.1107	3.1107	7.1107	7.1107
	-0.3686	-4.3686	-0.3686	-0.3686	-0.2011	-4.2011	-0.2011	-0.2011	-3.1096	-7.1096	-3.1096	-3.1096	6.5486	2.5486	6.5486	6.5486
	0.3221	0.3221	3.8221	0.3221	-1.0949	-1.0949	2.4051	-1.0949	-3.5126	-3.5126	-0.0126	-3.5126	10.8751	10.8751	14.3751	10.8751
	4.5270	4.5270	1.0270	4.5270	4.3275	4.3275	0.8275	4.3275	7.0925	7.0925	3.5925	7.0925	-4.1615	-4.1615	-7.6615	-4.1615
	0.8172	0.8172	4.3172	0.8172	4.1982	4.1982	7.6982	4.1982	1.2551	1.2551	4.7551	1.2551	7.1728	7.1728	10.6728	7.1728
	3.0821	3.0821	-0.4179	3.0821	4.7183	4.7183	1.2183	4.7183	4.4092	4.4092	0.9092	4.4092	4.7061	4.7061	1.2061	4.7061
	-1.2233	-4.2233	0.7767	-4.2233	0.2228	-2.7772	2.2228	-2.7772	-4.6713	-7.6713	-2.6713	-7.6713	-7.9199	-10.9199	-5.9199	-10.9199
	-1.9041	9.0959	0.0959	3.0959	-0.2148	10.7852	1.7852	4.7852	0.4453	11.4453	2.4453	5.4453	-1.4767	9.5233	0.5233	3.5233
	-7.0454	-2.0454	-1.0454	-2.0454	-6.2335	-1.2335	-0.2335	-1.2335	-7.4782	-2.4782	-1.4782	-2.4782	-2.2255	2.7745	3.7745	2.7745
	-6.0199	-9.0199	-0.0199	-9.0199	-4.0375	-7.0375	1.9625	-7.0375	-8.3122	-11.3122	-2.3122	-11.3122	0.0425	-2.9575	6.0425	-2.9575
	-0.2319	-0.2319	-0.2319	-0.2319	-0.3494	-0.3494	-0.3494	-0.3494	-2.5439	-2.5439	-2.5439	-2.5439	3.3081	3.3081	3.3081	3.3081

Table 14. Nine-Factor DS Simulated Response

Scenario	$\varepsilon \sim N(0,1)$				$\varepsilon \sim N(0,2)$				$\varepsilon \sim N(0,3)$				$\varepsilon \sim N(0,5)$			
Case	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
	-6.5718	-2.0718	-1.0718	-6.0718	-5.0282	-0.5282	0.4718	-4.5282	-8.8568	-4.3568	-3.3568	-8.3568	-5.1739	-0.6739	0.3261	-4.6739
	-8.9271	-6.4271	-6.4271	-8.4271	-6.3521	-3.8521	-3.8521	-5.8521	-8.4283	-5.9283	-5.9283	-7.9283	-5.8325	-3.3325	-3.3325	-5.3325
	6.7540	-0.2460	14.2540	-0.2460	8.5632	1.5632	16.0632	1.5632	5.3980	-1.6020	12.8980	-1.6020	0.6315	-6.3685	8.1315	-6.3685
	-5.3611	1.6389	13.1389	-12.3611	-7.5396	-0.5396	10.9604	-14.5396	-4.3540	2.6460	14.1460	-11.3540	-2.8272	4.1728	15.6728	-9.8272
	-1.0722	-4.0722	2.4278	-10.0722	-0.9806	-3.9806	2.5194	-9.9806	-1.5015	-4.5015	1.9985	-10.5015	0.8681	-2.1319	4.3681	-8.1319
	3.0372	-5.9628	-4.4628	-5.9628	-2.1549	-11.1549	-9.6549	-11.1549	-0.2920	-9.2920	-7.7920	-9.2920	4.4056	-4.5944	-3.0944	-4.5944
	0.9647	2.9647	1.4647	2.9647	-2.2766	-0.2766	-1.7766	-0.2766	-0.9184	1.0816	-0.4184	1.0816	5.0621	7.0621	5.5621	7.0621
	-5.1812	10.8188	-7.6812	-3.1812	-7.7627	8.2373	-10.2627	-5.7627	-10.6842	5.3158	-13.1842	-8.6842	-6.1450	9.8550	-8.6450	-4.1450
	-11.5682	8.9318	-4.5682	8.9318	-9.7422	10.7578	-2.7422	10.7578	-10.5137	9.9863	-3.5137	9.9863	-19.1293	1.3707	-12.1293	1.3707
	-11.6201	6.8799	-8.6201	8.8799	-11.1763	7.3237	-8.1763	9.3237	-7.5815	10.9185	-4.5815	12.9185	-15.6390	2.8610	-12.6390	4.8610
	5.9775	17.9775	13.4775	11.9775	5.9904	17.9904	13.4904	11.9904	1.9805	13.9805	9.4805	7.9805	14.9630	26.9630	22.4630	20.9630
	-6.1405	-8.1405	12.3595	-0.1405	-3.1020	-5.1020	15.3980	2.8980	-3.1339	-5.1339	15.3661	2.8661	-1.7023	-3.7023	16.7977	4.2977
	-0.4898	9.0102	4.5102	3.0102	0.0662	9.5662	5.0662	3.5662	-3.6061	5.8939	1.3939	-0.1061	1.4916	10.9916	6.4916	4.9916
	-8.4915	3.0085	3.5085	-4.9915	-5.7396	5.7604	6.2604	-2.2396	-7.1792	4.3208	4.8208	-3.6792	-11.9173	-0.4173	0.0827	-8.4173
	-0.3253	-7.8253	3.1747	-7.8253	-3.9942	-11.4942	-0.4942	-11.4942	0.6261	-6.8739	4.1261	-6.8739	1.4588	-6.0412	4.9588	-6.0412
	1.6734	2.1734	-5.8266	-5.8266	-0.9510	-0.4510	-8.4510	-8.4510	4.5618	5.0618	-2.9382	-2.9382	0.3469	0.8469	-7.1531	-7.1531
	-13.0033	12.9967	-2.5033	12.9967	-17.4189	8.5811	-6.9189	8.5811	-14.8327	11.1673	-4.3327	11.1673	-18.0066	7.9934	-7.5066	7.9934
	-17.0536	6.9464	-3.5536	8.9464	-15.4480	8.5520	-1.9480	10.5520	-24.6111	-0.6111	-11.1111	1.3889	-23.4302	0.5698	-9.9302	2.5698
	-0.3768	6.6232	0.1232	0.6232	1.1869	8.1869	1.6869	2.1869	-1.4119	5.5881	-0.9119	-0.4119	-3.3727	3.6273	-2.8727	-2.3727
	-3.3919	-10.3919	-5.8919	-2.3919	-6.2247	-13.2247	-8.7247	-5.2247	-4.5994	-11.5994	-7.0994	-3.5994	-3.8509	-10.8509	-6.3509	-2.8509
	1.8669	1.8669	1.8669	1.8669	-1.1295	-1.1295	-1.1295	-1.1295	0.0071	0.0071	0.0071	0.0071	4.3526	4.3526	4.3526	4.3526

Table 15. **FBBD Stepwise Backward Elimination Results: Case 1**

	Scenario			
	$\varepsilon \sim N(0,1)$	$\varepsilon \sim N(0,2)$	$\varepsilon \sim N(0,3)$	$\varepsilon \sim N(0,5)$
Step	**Factors Removed**			
1	H	J	F	E
2	–	F	J	F
3	–	–	–	H
4	–	–	–	D
5	–	–	–	C
6	–	–	–	B

Table 16. **DSD Forward Stepwise Results: Case 1**

	Scenario			
	$\varepsilon \sim N(0,1)$	$\varepsilon \sim N(0,2)$	$\varepsilon \sim N(0,3)$	$\varepsilon \sim N(0,5)$
Step	**Effects Added**			
1	EG	AG	EG	CH
2	AG	EG	AG	EJ
3	AE	EJ	AE	F^2
4	G^2	AE	DF	AG
5	AJ	–	H^2	D
6	DH	–	–	–
7	AD	–	–	–
8	F	–	–	–
9	C	–	–	–

all the factors which comprise the two-way interaction or pure quadratic effects. For example, when considering Scenario 3, where $\varepsilon_i \sim N(0,3)$, the EG and H^2 effects, which entered the regression model in steps 1 and 5, respectively, would require the E, G, and H factors also be in the model.

Case 2 data was simulated using the model

$$y_i = 2A_i - 1.5E_i + 2G_i + 4C_i - 3H_i + 2.5E_i^2 - 4G_iH_i + 3.5E_iH_i - 5C_iG_i + \varepsilon_i, \quad (3.6)$$

to represent a response exhibiting effect sparsity and strong heredity between active two-factor interactions or pure quadratic effects and their associated main effects. The model is considered to exhibit effect sparsity because although over 50% of the factors (5 of 9) are active only 9 of 54 total effects are active. Cases 1 and 2 both have the same number of active effects but differ in the number of active factors contained within the second-order portion of the model.

Table 17 provides Edwards and Mee (2011) factor-based backward elimination results and Table 18 Jones and Nachtsheim (2011a) forward stepwise regression, respectively, using the Case 2 response data associated with each design for all four noise level scenarios in Tables 13 and 14.

Table 17. FBBD Stepwise Backward Elimination Results: Case 2

	Scenario			
	$\varepsilon \sim N(0,1)$	$\varepsilon \sim N(0,2)$	$\varepsilon \sim N(0,3)$	$\varepsilon \sim N(0,5)$
Step	**Factors Removed**			
1	H	J	F	E
2	–	F	J	B
3	–	B	–	D

Case 3 data was simulated using the model

$$y_i = 2A_i + 2E_i - 1.5A_i^2 + 2.5E_i^2 - 3.5A_iE_i + 4A_iG_i - 5E_iG_i + \varepsilon_i, \quad (3.7)$$

Table 18. DSD Forward Stepwise Results: Case 2

	Scenario			
	$\varepsilon \sim N(0,1)$	$\varepsilon \sim N(0,2)$	$\varepsilon \sim N(0,3)$	$\varepsilon \sim N(0,5)$
Step	**Effects Added**			
1	CG	C^2	C^2	EH
2	GH	GH	E^2	CE
3	EH	CG	DH	AE
4	A	CJ	A^2	CF
5	E^2	A	DG	AG
6	J^2	GJ	AF	D
7	DH	–	–	–
8	CF	–	–	–

thereby representing a response exhibiting factor sparsity and weak heredity between active two-factor interactions or pure quadratic effects and main effects. The model exhibits factor sparsity because only 3 of the 9 factors are active within the 7 effects contained in the model. Since not all factors, which comprise the two-factor interactions, are present as a main effect, the model exhibits weak heredity. For instance, although factor G is significant within two two-factor interactions, factor G by itself is not significant.

Table 19 provides Edwards and Mee (2011) factor-based backward elimination results and Table 20 Jones and Nachtsheim (2011a) forward stepwise regression, respectively, using the Case 3 response data associated with each design for all four noise level scenarios in Tables 13 and 14.

Case 4 data was simulated using the model

$$y_i = 2A_i - 1.5E_i + 2G_i - 3H_i^2 + 2.5E_i^2 + 4A_iC_i + 3.5E_iH_i - 5C_iG_i - 4G_iH_i + \varepsilon_i \quad (3.8)$$

represents a response exhibiting effect sparsity and weak heredity between active two-factor interactions or pure quadratic effects and main effects. The case 4 model is identical to the model used by Edwards and Mee (2011). However, the response data differs even for the $\varepsilon_i \sim N(0,1)$ scenario.

Table 21 provides Edwards and Mee (2011) factor-based backward elimination results and Table 22 Jones and Nachtsheim (2011) forward stepwise regression, respectively, using the Case 4 response data associated with each design for all four noise level scenarios in Tables 13 and 14.

Table 19. FBBD Stepwise Backward Elimination Results: Case 3

	Scenario			
	$\varepsilon \sim N(0,1)$	$\varepsilon \sim N(0,2)$	$\varepsilon \sim N(0,3)$	$\varepsilon \sim N(0,5)$
Step	**Factors Removed**			
1	H	J	A	E
2	–	F	H	C
3	–	B	F	H
4	–	–	J	B
5	–	–	C	J
6	–	–	B	–

Table 20. DSD Forward Stepwise Results: Case 3

	Scenario			
	$\varepsilon \sim N(0,1)$	$\varepsilon \sim N(0,2)$	$\varepsilon \sim N(0,3)$	$\varepsilon \sim N(0,5)$
Step	**Effects Added**			
1	AE	AE	EG	F^2
2	BF	CH	AG	E^2
3	J^2	EJ	AE	FH
4	A^2	J^2	DF	AG
5	FH	E^2	H	DE
6	DJ	–	AF	–
7	E^2	–	–	–

3.4 Case Comparison

Tables 23, 24, 25, and 26 show which effects from Cases 1 through 4's four different noise level scenarios were properly identified, incorrectly identified (Type I error), and not identified (Type II error), for both nine-factor designs.

Table 21. FBBD Stepwise Backward Elimination Results: Case 4

	Scenario			
	$\varepsilon \sim N(0,1)$	$\varepsilon \sim N(0,2)$	$\varepsilon \sim N(0,3)$	$\varepsilon \sim N(0,5)$
Step	**Factors Removed**			
1	F	J	F	A
2	–	F	J	G
3	–	B	–	C

Table 22. DSD Forward Stepwise Results: Case 4

	Scenario			
	$\varepsilon \sim N(0,1)$	$\varepsilon \sim N(0,2)$	$\varepsilon \sim N(0,3)$	$\varepsilon \sim N(0,5)$
Step	**Effects Added**			
1	GH	GH	GH	F^2
2	AH	AE	AH	AC
3	AF	EG	DE	BG
4	EF	HJ	AD	CF
5	G^2	E^2	DG	BF
6	AC	J^2	FH	E
7	DF	–	A^2	D
8	J	–	–	–

Table 23. Second Order Screening Design Results: Case 1

Strong Heredity, Factor Sparsity Model: $2A - 1.5E + 2G - 3A^2 + 2.5E^2 - 4G^2 + 4AE + 3.5AG - 5EG + \varepsilon$			
Scenario		**DSD**	**FBBD**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, G, G^2, AE, AG, EG	$A, E, G, A^2, E^2, G^2, AE, AG, EG$
	Type I errors	$C, D, F, H, J, AD, AJ, DH$	$B, C, D, J, B^2, C^2, J^2, AB, AC, AD, AF, AJ, BE, BF, BG, CD, CE, CF, DE, DF, DG, DJ, EJ, FG, FJ$
	Type II errors	A^2, E^2	NONE
$\varepsilon \sim N(0,2)$	**Identified**	A, E, G, AE, AG, EG	$A, E, G, A^2, E^2, G^2, AE, AG, EG$
	Type I errors	J, EJ	H^2, CD, DG
	Type II errors	A^2, E^2, G^2	NONE
$\varepsilon \sim N(0,3)$	**Identified**	A, E, G, AE, AG, EG	$E, G, E^2, G^2, AE, AG, EG$
	Type I errors	D, F, H, H^2, DF	B^2, D^2, AH, BG, DG
	Type II errors	A^2, E^2, G^2	A, A^2
$\varepsilon \sim N(0,5)$	**Identified**	A, E, G, AG	G^2, AG
	Type I errors	$C, D, F, H, J, F^2, CH, EJ$	AJ
	Type II errors	A^2, E^2, G^2, AE, EG	$A, E, G, A^2, E^2, AE, EG$

Table 24. Second Order Screening Design Results: Case 2

Strong Heredity, Effect Sparsity Model: $2A - 1.5E + 2G + 4C - 3H + 2.5E^2 - 5CG + 3.5EH - 4GH + \varepsilon$			
Scenario		**DSD**	**FBBD**
$\varepsilon \sim N(0,1)$	**Identified**	A, C, E, G, H, E^2, CG EH, GH	A, C, E, G, E^2, CG
	Type I errors	D, F, J, J^2, CF, DH	$B, H, J, A^2, B^2, C^2, G^2, J^2, AB,$ $AC, AD, AE, AF, AJ, BD, BE,$ $BF, BG, BJ, CD, CE, CF, DE, DF,$ $DG, DJ, EG, EJ, FG, FJ, GJ$
	Type II errors	NONE	H, EH, GH
$\varepsilon \sim N(0,2)$	**Identified**	A, C, G, H, CG, GH	$A, C, E, G, H, E^2, CG, EH, GH$
	Type I errors	J, C^2, CJ, GJ	D, A^2, H^2, AC, CD, DG
	Type II errors	E, E^2, EH	NONE
$\varepsilon \sim N(0,3)$	**Identified**	A, C, E, G, H, E^2	$C, E, G, H, E^2, CG, EH, GH$
	Type I errors	$D, F, A^2, C^2, AF, DG, DH$	$A^2, B^2, D^2, G^2, AH, BG, DG$
	Type II errors	CG, EH, GH	A
$\varepsilon \sim N(0,5)$	**Identified**	A, E, G, C, H, EH	C, H, CG
	Type I errors	D, F, AE, AG, CE, CF	AJ, FJ
	Type II errors	E^2, CG, GH	A, G, E, E^2, EH, GH

Table 25. Second Order Screening Design Results: Case 3

Weak Heredity, Factor Sparsity Model: $2A + 2E - 1.5A^2 + 2.5E^2 - 3.5AE + 4AG - 5EG + \varepsilon$			
Scenario		**DSD**	**FBBD**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, A^2, E^2, AE	$A, E, A^2, E^2, AE, AG, EG$
	Type I errors	$B, D, F, H, J,$ J^2, BF, DJ, FH	$B, C, D, G, J, B^2, C^2,$ $G^2, J^2, AB, AC, AD, AF, AJ,$ $BE, BF, BG, CD, CE, CF, DE,$ DF, DG, DJ, EJ, FG, FJ
	Type II errors	AG, EG	none
$\varepsilon \sim N(0,2)$	**Identified**	A, E, E^2, AE	A, E, E^2, AE, AG, EG
	Type I errors	C, H, J, J^2, CH, EJ	H^2, CD, DG
	Type II errors	A^2, AG, EG	A^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E, AE, AG, EG	E, E^2, EG
	Type I errors	D, F, G, H, AF, DF	DG
	Type II errors	A^2, E^2	A, A^2, AE, AG
$\varepsilon \sim N(0,5)$	**Identified**	A, E, E^2, AG	A, AG
	Type I errors	D, F, G, H, F^2, DE, FH	F^2, AD, DF, DG
	Type II errors	A^2, AE, EG	E, A^2, E^2, AE, EG

Table 26. **Second Order Screening Design Results: Case 4**

Weak Heredity, Effect Sparsity Model: $2A - 1.5E + 2G + 2.5E^2 - 3H^2 + 4AC - 5CG + 3.5EH - 4GH + \varepsilon$			
Scenario		**DSD**	**FBBD**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, G, AC, GH	$A, E, G, E^2, H^2, AC, CG, EH, GH$
	Type I errors	$C, D, F, H, J, G^2, AF, AH, DF, EF$	$B, C, D, H, J, B^2, C^2, G^2, J^2, AB, AD, AE, AJ, BD, BE, BG, BH, CD, DE, DG, DH, DJ, EG, EJ, HJ$
	Type II errors	E^2, H^2, CG, EH	none
$\varepsilon \sim N(0,2)$	**Identified**	A, E, G, E^2, GH	$A, E, G, E^2, H^2, AC, CG, EH, GH$
	Type I errors	H, J, J^2, AE, EG, HJ	A^2, CD, DG
	Type II errors	H^2, AC, CG, EH	none
$\varepsilon \sim N(0,3)$	**Identified**	A, E, G, GH	$E, G, E^2, H^2, CG, EH, GH$
	Type I errors	$D, F, H, A^2, AD, AH, DE, DG, FH$	$A^2, B^2, D^2, G^2, AH, BG, DG$
	Type II errors	E^2, H^2, AC, CG, EH	A, AC
$\varepsilon \sim N(0,5)$	**Identified**	A, E, G, AC	E, H^2
	Type I errors	$B, C, D, F, F^2, BF, BG, CF$	$B, D, E, H, J, B^2, F^2, H^2, AB, AD, AE, BD, BF, BH, BJ, DE, DF, EF, FH, FJ, HJ$
	Type II errors	E^2, H^2, CG, EH, GH	$A, G, E^2, AC, CG, EH, GH$

As expected for both the FBBD and DSD, Type II errors increased in all four cases as the noise level increased. However, whereas the increase in the number of Type II errors across the cases for the DSD were on the order of 1 to 3, the increase in Type II errors for the FBBD were far greater, from 5 to 7, suggesting the DSD may be more robust to noise effects. Unfortunately, the DSD did not exhibit robust results when it came to whether or not weak or strong heredity and factor or effect sparsity assumption held. When comparing strong heredity to weak heredity for DSD, the DSD performed better when strong heredity was exhibited, particularly when effect sparsity was present. Similarly, when comparing factor sparsity to effect sparsity for DSD, the DSD performed better when factor sparsity was exhibited, particularly when strong heredity was present. Overall, this is not surprising as the analysis method for DSD forces a strong heredity model and the DSD has more power in determining active effects when fewer factors are active Jones and Nachtsheim (2011a). The DSD performance is inferior to the FBBD, in terms of Type II errors, at the lower noise levels, Scenarios 1 and 2, in all but Case 2 which represented strong heredity and effect sparsity. However, in all but one scenario in one case (Case 2, Scenario 2), the Type II errors were limited to mostly pure quadratic effects and a few two-factor interactions. This result carries across all case and scenario combinations for the DSD and is likely a by-product of the design which focuses on main effects which are unbiased by any second-order effect (pure quadratic or two-factor interaction) and where second-order effects have some correlation but are not completely confounded with other second-order effects.

In contrast, the FBBD has no discernible pattern in Type II errors. This implies the FBBD is just as likely to miss important main effects as second-order effects, especially at the higher noise levels. However at the lower noise levels, Scenarios 1 and 2, the FBBD made only a few Type II errors. In so doing, the FBBD consistently over specified a model, particularly at the lowest noise level, Scenario 1. With regards to robustness to heredity and sparsity, the FBBD performed equally well for weak and strong heredity and factor and effect sparsity, excluding Case 2. However, even with nearly double the number of runs, the FBBD is susceptible to excluding an active main effect during the initial stages of design analysis based upon RMSE, as noise level increases as evident by Case 2, Scenario 1.

Table 27 displays average results of all active effects, second-order effects and pure quadratic effects correctly identified based on five independent replications for all four cases considered and for each level of random noise. Clearly, the larger FBBD does quite well compared to DSD, but its performance degrades as noise levels increase. The smaller DSD does reasonably well under strong heredity but does seem to struggle finding the interactions and quadratic effects.

3.5 Conclusions

Regardless of the heredity (weak or strong), sparsity (effect or factor), and noise level combination, the DSD is robust in its ability to correctly identify active main effects. At lower noise levels, the DSD performs favorably in identifying active two-factor interactions but as the noise level increases the DSD performance suffers. Additionally, regardless of case or scenario, the DSD struggled finding active quadratic effects. However, if the experimenter has prior knowledge regarding the importance of second-order effects, especially pure quadratic effects, and wishes to maintain the requirement for a single design without follow up design runs, augmenting the DSD could reduce the correlation between a factor's second-order effect without sacrificing too much in the way of design run efficiency. For instance, within many physical models of complex aerodynamic systems, a quadratic "velocity" factor is often present.

Table 27. Second Order Screening Design Results: Average

Strong Heredity, Factor Sparsity Model:			5 Rep Avg
Scenario		**DSD**	**FBBD**
$\varepsilon \sim N(0,1)$	**Identified**	$67\%, 50\%, 20\%$	$100\%, 100\%, 100\%$
$\varepsilon \sim N(0,2)$	**Identified**	$58\%, 37\%, 20\%$	$96\%, 97\%, 100\%$
$\varepsilon \sim N(0,3)$	**Identified**	$51\%, 33\%, 27\%$	$80\%, 90\%, 100\%$
$\varepsilon \sim N(0,5)$	**Identified**	$49\%, 23\%, 0\%$	$53\%, 57\%, 67\%$
Strong Heredity, Effect Sparsity Model:			**5 Rep Avg**
Scenario		**DSD**	**FBBD**
$\varepsilon \sim N(0,1)$	**Identified**	$98\%, 95\%, 80\%$	$91\%, 90\%, 100\%$
$\varepsilon \sim N(0,2)$	**Identified**	$78\%, 55\%, 20\%$	$96\%, 100\%, 100\%$
$\varepsilon \sim N(0,3)$	**Identified**	$84\%, 70\%, 40\%$	$62\%, 65\%, 100\%$
$\varepsilon \sim N(0,5)$	**Identified**	$69\%, 30\%, 0\%$	$53\%, 50\%, 40\%$
Weak Heredity, Factor Sparsity Model:			**5 Rep Avg**
Scenario		**DSD**	**FBBD**
$\varepsilon \sim N(0,1)$	**Identified**	$80\%, 72\%, 50\%$	$100\%, 100\%, 100\%$
$\varepsilon \sim N(0,2)$	**Identified**	$60\%, 44\%, 20\%$	$86\%, 80\%, 60\%$
$\varepsilon \sim N(0,3)$	**Identified**	$51\%, 32\%, 0\%$	$69\%, 64\%, 80\%$
$\varepsilon \sim N(0,5)$	**Identified**	$60\%, 44\%, 20\%$	$46\%, 48\%, 40\%$
Weak Heredity, Effect Sparsity Model:			**5 Rep Avg**
Scenario		**DSD**	**FBBD**
$\varepsilon \sim N(0,1)$	**Identified**	$49\%, 23\%, 0\%$	$96\%, 97\%, 100\%$
$\varepsilon \sim N(0,2)$	**Identified**	$47\%, 23\%, 10\%$	$91\%, 93\%, 90\%$
$\varepsilon \sim N(0,3)$	**Identified**	$44\%, 20\%, 10\%$	$76\%, 77\%, 100\%$
$\varepsilon \sim N(0,5)$	**Identified**	$51\%, 27\%, 30\%$	$53\%, 50\%, 60\%$

Note: Identified percentages correspond to percentage of active effects, second-order effects, and pure quadratic effects.

The FBBD has no discernible pattern in Type II errors. This implies the FBBD is just as likely to miss important main effects as second-order effects. Thus, modifications to the FBBD design to ensure at least main effect Type II errors are eliminated are not readily apparent. In addition, since the FBBD over specifies models, as indicated by the large number of Type I errors particularly at lower noise levels, it seems further fractionation of the FBBD is possible, which can reduce design run size requirements without sacrificing Type II error performance.

Each design was examined using the authors' recommended analysis method. Employing different analysis methods may yield improved performance of the design. For instance, it is possible analyze the FBBD with the forward stepwise regression method used on the DSD.

Whenever a screening design is employed, analytical tradeoffs must be accepted. Overall, both designs performed in the environment to which they were designed. The DSD is run size efficient when strong heredity and factor sparsity are present and when few second-order effects are active. In contrast, the FBBD diminishes the importance of the heredity and sparsity assumption but at the cost of additional design runs. Depending upon subject matter expertise regarding a system under study, selection or modification of one or both of the designs could certainly be useful within many high-technology industries.

IV. Augmentation of Definitive Screening Designs (DSD+)

4.1 Introduction

Response surface methodology (RSM) focuses on approximating a real world system response typically with either a first-order or second-order polynomial model. While the choice of experimental designs for first-order models is fairly straight forward depending upon the shape of the experimental design region and number of available experimental runs, choosing an experimental design to fit a second-order model,

$$\eta = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j, \tag{4.1}$$

is more complex due to the variety of design criteria and characteristics to consider.

Usually, the experimenter does not have *a priori* knowledge regarding the appropriate polynomial model to use to approximate the system response. As such it is common practice in RSM to employ experiments sequentially. Box and Liu (1999) illustrated the RSM philosophy of sequential learning where first-order designs are typically used to perform factor screening and second-order designs are used to fit a response surface exhibiting some degree of curvature. Since the *a posteriori* knowledge about a system response possessing curvature comes from analysis of the first-order design, the typically sequential nature of RSM allows developing second-order designs by augmenting first-order designs with additional experimental runs.

Whether due to time, budget, or other constraints, there are times when conducting multiple experiments is unrealistic. For instance, Lawson (2003) points out fixed deadlines for scale up and production of prototype engineering designs may not allow the possibility of follow-up experimentation. Couple this with the fact that military systems, particularly aerodynamic systems, are complex and often exhibit nonlinear behavior, there are times when a single experimental design capable of performing both factor screening and higher order response surface exploration may be required.

Recent literature has proposed second-order screening design methodologies, sometimes referred to as One-Step RSM or Definitive Screening, employing a single experimental design capable of both factor screening and fitting a second-order polynomial model.

Edwards and Truong (2011) preformed a simulation study examining several second-order screening designs focusing on the design's ability to correctly identify active factors under a variety of conditions. The truth models used assumed both factor sparsity and strong effect heredity.

Sparsity and heredity are two important principles considered during the development of successful screening designs. The sparsity principle stems from the Pareto principle which has led to an assumption in screening designs that only a small number of factors, *factor sparsity*, are significant in their contribution to an appropriate polynomial model approximation of a system response. However, the degree to which factor sparsity holds as the number of factors being investigated grows has been debated. The term *effect sparsity* has been used to identify with the assumption that instead of the number of active factors being relatively small in the polynomial model approximation, the number of active effects is relatively small. As a result, it is possible for the assumption of effect sparsity to hold while factor sparsity does not.

Heredity, either strong or weak, is another screening principle considered during model selection. Strong heredity means that if a model includes a two-factor interaction, then both its constituent main effects are also included in the model. Conversely, weak heredity requires only one of the two constituent main effects be included in the model.

Dougherty et al. (2013b) examined the robustness of Definitive Screening Designs (DSD) and Fractional Box-Behken Designs (FBBD), two second-order screening designs, with respect to the assumptions of sparsity (factor or effect) and heredity (strong or weak). Dougherty et al. (2013b) showed that regardless of the heredity (weak or strong), sparsity (effect or factor), or noise level combination, the DSD is robust in its ability to correctly identify active main effects. At lower noise levels, the DSD performs favorably in identifying active two-factor interactions but as the noise level increases the DSD performance suffers. Additionally the DSD had trouble identifying active pure quadratic effects when two-factor interactions are present. As a result, if the experimenter has *a priori* knowledge regarding the importance of a particular factor, or that factor's second-order effects, augmentation of the DSD could reduce the correlation between a factors' second-order effects without sacrificing too much in the way of design run efficiency while maintaining the requirement for a single design. Conversely, if the experimenter has *a posteriori* knowledge about a particular factor or factors' second-order effects, augmenting the DSD demonstrates the feasibility of follow-up design runs for DSD.

The remainder of this paper is organized as follows, Section 4.2 briefing discusses the literature relevant to second-order screening designs while Section 4.3 focuses on the Definitive Screening Designs generation and augmentation. In Section 4.4, we present a side-by-side comparison of the Definitive Screening Design examined in Dougherty et al. (2013b) with an augmented design focusing on improved robustness to the assumptions of heredity and sparsity and significant second-order factor identification. Section 4.5 examines the effect of replicating the analysis on the designs' ability to identify important factors of interest and Section 4.6 concludes the article.

4.2 Second Order Screening Designs

Initial attempts at identifying second-order screening designs relied upon the design's projection capacity. When the *factor sparsity* principle holds any regular fractional factorial design of resolution R, projects onto any subset of $R-1$ factors as a full factorial. For example, a 2^{3-1}_{III} design ($R = 3$) can project into a 2^2 design (Myers and Anderson-Cook, 2009). This projection property extends to nonregular designs like the Plackett-Burman designs discussed in Lin and Draper (1992) and Wang and Wu (1995).

Cheng and Wu (2001), hereafter referred to as CW, studied three orthogonal array (OA) designs ($OA(18, 3^7)$, $OA(27, 3^8)$, and $OA(36, 3^{12})$). The $OA(N, 3^k)$ connotation shows the design's number of runs N and number of factors k. In contrast to 3^{n-k} designs which have defining contrast subgroups to describe the design structure, the $OA(N, 3^k)$ designs studied by CW required computer search to classify the possible projected designs.

Because a design can project onto many different combinations of factors, CW developed a projection-efficiency criterion to compare designs based upon (1) the number of eligible projected designs and (2) the estimation efficiency for eligible projected designs determined by the ratio of each designs $D-$ and $G-$efficiences (Cheng and Wu, 2001). *Eligible* designs are designs to fit a second-order model and the $D-$ and $G-$efficiences, denoted D_{eff} and G_{eff}, respectively, criteria compare the performance of a design against a corresponding optimal design (Myers and Anderson-Cook, 2009).

Under the assumptions of factor sparsity and strong heredity, CW introduced a two-stage analysis method. The first stage consisted of performing a main effect factor screening analysis and the second stage involved fitting a second-order model with the identified main effects from the first stage. The key linkage between stage one and two was the ability to project the initial larger factor space onto a smaller factor space capable of fitting a second-order model. Unfortunately, the designs CW studied have no guarantee as to their ability to project down to a specific subset of the original factors and no flexibility in modifying the number of design runs.

Improving on the designs of CW, Xu et al. (2004), hereafter referred to by XCW, proposed a combinatorial method for constructing new and efficient OA designs and a design selection approach based upon a *projection aberration criterion* which combines the generalized word-length pattern of the generalized minimum aberration criterion (Xu and Wu, 2001) for factor screening and the projection-efficiency criteria (Cheng and Wu, 2001) for interaction detection. XCW assessed the projection performance of three combinatorially non-isomorphic $OA(18, 3^7)$s and three combinatorially non-isomorphic $OA(27, 3^{13})$s. Their three-step approach involves: (1) screening out poor orthogonal arrays for factor screening using the generalized word-length pattern, (2) applying the projection aberration criterion to select a best design from step 1, and (3) determining the best level permutations of the design from step 2 to improve design projection eligibility and estimation efficiency under the second-order polynomial model.

Ye et al. (2007), hereafter referred to as YTL, also examined 3-level 18-run and 27-run orthogonal designs; however, in addition to considering the projection properties of designs, their design choices were based on both model estimation and model discrimination criteria. The two model estimation criteria employed examine the proportion of estimable models, *Estimation Capacity (EC)*, and average $D-$efficiency of all models, *Information Capacity (IC)*. YTL employed two of the six non-Bayesian criteria, *Average Expected Prediction Differences (AEPD)* and *Minimum Maximum Prediction Difference (MMPD)*, proposed by Jones et al. (2007) for model discrimination.

While previous work focused primarily on the designs projection capacity, Edwards and Truong (2011) applied Jones and Nachtsheim (2011b) method for finding efficient designs with minimal aliasing between main effects and two-factor interactions. Deemed MA designs, Edwards and Truong (2011) constructed 18, 27, and 30-run designs for simultaneous screening and response surface optimization for $k = 4$ to 7, $k = 4$ to 13, and $k = 6$ to 14 factors, respectively, by minimizing the sum of squares of the elements of the alias matrix, $\mathbf{A}$, subject to a lower bound on the primary model $D-$efficiency. Edwards and Truong (2011) compared the 27-run OAs of XCW and YTL with MA designs in terms of D-efficiency of projection and, via a simulation study, the proportion of active factors declared significant (Power 1) as well as the proportion of simulations in which only the true active factors are declared significant (Power 2). Although ranked last in terms of D-efficiency, the MA designs showed superior performance in their ability to detect active factors (Edwards and Truong, 2011).

For simplicity, the CW, XCW, YTL, and MA designs use linear and quadratic main-effects only analysis for factor screening but the Bayesian approaches of Box and Meyer (1993) or Chipman et al. (1997) can also be used to screen for significant factors outside of main effects. However, these methods are not readily available in statistical software packages and are computationally intensive procedures, thus likely making their use impractical (Edwards and Truong, 2011). Unfortunately, as shown by Truong (2010), if the strong heredity principle fails to hold important effects can be missed leading to a misspecified second-order polynomial model.

Edwards and Mee (2011) introduced the spherical FBBD aimed at overcoming the projection deficiencies and main/quadratic effect only analysis issues found in the CW/XCW/YTL/MA designs. The FBBD provide the ability to explore interactions during the screening stage and to fit second-order models via a backward elimination analysis strategy to each of the $(k-1)$-factor projections. In contrast to the CW/XCW/YTL/MA designs, Edwards and Mee (2011) assumed an effect sparsity vice factor sparsity model and searched for designs having larger factor eligible projections than the CW/XCW/YTL/MA designs by taking subsets of the two-level fractional factorial designs which compose a BBD. While FBBDs require more runs than CW/XCW/YTL/MA designs, their ease of construction and aliasing structure facilitate an analysis strategy which cannot be applied to the CW/XCW/YTL/MA designs.

Jones and Nachtsheim (2011a) introduced a class of three-level designs referred to as *Definitive Screening Designs* (DSD) where main effects are not biased by second-order effects and all quadratic effects are estimable. For $k \geq 6$, the DSD can project down to a full quadratic model in any three factors.

4.3 Definitive Screening Design Augmentation

Jones and Nachtsheim (2011a) used a computerized search algorithm to create the DSD, with $2k+1$ runs to investigate k factors. The DSD consist of k fold-over pairs for k factors and a single center point. The search algorithm forces each run to maintain a single factor at its center point while forcing the remaining factors to their extremes (± 1). The DSD is constructed using a variant of the coordinate exchange algorithm of Meyer and Nachtsheim (1995) to maximize the determinant of the information matrix of the main effects model while maintaining the desired design structure.

To guard against local maxima, Jones and Nachtsheim (2011a) use multiple random starting designs for each k-factor design; however Xiao et al. (2012) demonstrate a method for generating global optimum DSD for an even value of k through the use of conference matrices. Table 28 shows the nine-factor DSD generated by JMP

10. JMP 10 uses the conference matrices method of Xiao et al. (2012) even when k is odd by producing a DSD for $k+1$ factors and removing the $k+1$ column factor settings. As a result when k is odd, the DSD has $2k+3$ runs. When k is even, the DSD maintains the $2k+1$ number of runs original proposed by Jones and Nachtsheim (2011a).

Table 28. Nine-Factor Definitive Screening Design (DSD)

A	B	C	D	E	F	G	H	J
0	1	1	1	1	1	1	1	1
0	-1	-1	-1	-1	-1	-1	-1	-1
1	0	-1	-1	-1	-1	1	1	1
-1	0	1	1	1	1	-1	-1	-1
1	-1	0	-1	1	1	-1	-1	1
-1	1	0	1	-1	-1	1	1	-1
1	-1	-1	0	1	1	1	1	-1
-1	1	1	0	-1	-1	-1	-1	1
1	-1	1	1	0	-1	-1	1	-1
-1	1	-1	-1	0	1	1	-1	1
1	-1	1	1	-1	0	1	-1	1
-1	1	-1	-1	1	0	-1	1	-1
1	1	-1	1	-1	1	0	-1	-1
-1	-1	1	-1	1	-1	0	1	1
1	1	-1	1	1	-1	-1	0	1
-1	-1	1	-1	-1	1	1	0	-1
1	1	1	-1	-1	1	-1	1	0
-1	-1	-1	1	1	-1	1	-1	0
1	1	1	-1	1	-1	1	-1	-1
-1	-1	-1	1	-1	1	-1	1	1
0	0	0	0	0	0	0	0	0

The $2k+1$ or $2k+3$ runs for when k is even or odd, respectively, provide a sufficient number of degrees of freedom for estimates of the intercept, all k main effects, and all k pure quadratic effects. However, Dougherty et al. (2013b) showed when both two-factor interactions and pure-quadratic effects are active, regardless of heredity (strong or weak) or sparsity (factor or effect), the standard DSD may not have enough degrees of freedom to decouple the correlation between two-factor interactions and pure-quadratic effects. As a result, the DSD when used as a single experimental design is susceptible to making Type-II errors particularly with regards to active pure-quadratic effects. Because the DSD is very run efficient when compared to other second-order screening designs, augmenting the original DSD to improve detection of active quadratic effects (both two-factor interactions and pure-quadratic) is desirable.

If the experimenter has *a priori* knowledge regarding the importance of a particular factor or factors' second-order effects, augmentation of the DSD, hereafter referred to as DSD+, could reduce the correlation between a factor's second-order effects without sacrificing too much in the way of design run efficiency while maintaining the requirement for a single design. Conversely, if the experimenter has *a posteriori* knowledge about a particular factor or factors' second-order effects, augmenting the DSD demonstrates the feasibility of follow-up design runs for DSD.

Common approaches to design augmentation to clarify model ambiguity involves the augmentation of the design with runs specifically designed to de-alias a specific alias chain or using complete or fractional foldovers of the design. Since the DSD are basically already full foldover designs, using the foldover approach on DSD does not reduce aliasing between second-order effects. Additionally, the alias chains for DSD are very complex due to the nature of the design construction. Therefore an alternative approach using a D-optimal strategy for selecting augmentation points is employed.

Similar to Jones and Nachtsheim (2011a), a computerized search algorithm is used to add $k-1$ runs to the DSD. However, instead of the information matrix being only a main effects model, the information matrix contains the main effects and the $k-1$ two-factor interactions involving a particular factor. The DSD+ were constructed using a variant of the coordinate exchange algorithm of Meyer and Nachtsheim (1995) to maximize the determinant of the updated information matrix. Multiple random starting designs for each $k-$factor design were explored to guard against local maxima; however, the generated designs were still not unique. Multiple designs were generated which were equivalent based upon both *D-optimal* and *I-efficient* criteria; although, as k increased the number of different designs decreased.

Table 29 shows the $k = 9$ factor DSD generated by JMP 10 plus $k - 1 = 8$ augmentation runs after updating the information matrix to include the 8 two-way interactions involving factor A.

4.4 Case Comparison

Dougherty et al. (2013b) conducted an empirical study of the nine-factor definitive screening design generated using conference matrices based on Xiao et al. (2012) focusing on the design's robustness to detect important effects in models exhibiting different combinations of heredity and sparsity. Using Jones and Nachtsheim (2011a) recommended analysis methodology, the cases and scenarios studied are reexamined using the DSD+.

Jones and Nachtsheim (2011a) suggest performing a forward stepwise regression, which considers all terms in a second-order model of $k = 9$ factors. With a p-value of 0.1 to enter, effects are added into the second-order model while forcing a strong heredity model. As such, when either two-factor interactions or pure-quadratic effects are included in the model, the lower order terms must also be included.

Four cases were considered to represent different combinations of model heredity (strong or weak) and sparsity (factor or effect). In addition, each model was examined with four different noise levels scenarios; however, the noise level vector used for each scenario was identical across each model for each design. The 21 and 29 treatment combinations for the DSD and DSD+ designs are given in Tables 28 and 29, respectively. Table 30 shows the simulated response values for the 16 combinations of case

Table 29. Nine-Factor Augmented Definitive Screening Design (DSD+)

A	B	C	D	E	F	G	H	J
0	1	1	1	1	1	1	1	1
0	-1	-1	-1	-1	-1	-1	-1	-1
1	0	-1	-1	-1	-1	1	1	1
-1	0	1	1	1	1	-1	-1	-1
1	-1	0	-1	1	1	-1	-1	1
-1	1	0	1	-1	-1	1	1	-1
1	-1	-1	0	1	1	1	1	-1
-1	1	1	0	-1	-1	-1	-1	1
1	-1	1	1	0	-1	-1	1	-1
-1	1	-1	-1	0	1	1	-1	1
1	-1	1	1	-1	0	1	-1	1
-1	1	-1	-1	1	0	-1	1	-1
1	1	-1	1	-1	1	0	-1	-1
-1	-1	1	-1	1	-1	0	1	1
1	1	-1	1	1	-1	-1	0	1
-1	-1	1	-1	-1	1	1	0	-1
1	1	1	-1	-1	1	-1	1	0
-1	-1	-1	1	1	-1	1	-1	0
1	1	1	-1	1	-1	1	-1	-1
-1	-1	-1	1	-1	1	-1	1	1
0	0	0	0	0	0	0	0	0
1	-1	1	-1	-1	-1	-1	-1	-1
-1	1	1	1	1	-1	1	1	1
-1	-1	1	-1	-1	-1	-1	-1	1
-1	1	-1	-1	-1	-1	-1	1	-1
1	-1	1	1	1	1	-1	1	1
1	1	1	1	-1	1	1	1	-1
-1	-1	-1	1	1	-1	-1	-1	-1
1	1	1	-1	1	-1	1	1	1

and noise level scenario for the original DSD runs and the eight additional runs for the DSD+.

Case 1 data was simulated based on the model

$$y_i = 2A_i - 1.5E_i + 2G_i - 3A_i^2 + 2.5E_i^2 - 4G_i^2 + 4A_iE_i + 3.5A_iG_i - 5E_iG_i + \varepsilon_i, \quad (4.2)$$

thereby representing a response which exhibits factor sparsity and strong heredity between active two-factor interactions or pure quadratic effects and main effects. The model exhibits factor sparsity because only 3 of the 9 factors are active within the 9 effects contained in the model.

Jones and Nachtsheim (2011a) perform forward stepwise regression with a p-value of 0.1 to enter while forcing a strong heredity model. Table 31 shows the forward stepwise regression steps for the Case 1 data for all four noise level scenarios of Table 30.

Since the "combined" option rule is used for the forward stepwise regression, the inclusion of two-way interaction or pure quadratic effects result in the inclusion of all the factors which comprise the two-way interaction or pure quadratic effects. For example, when considering the original DSD Scenario 3, where $\varepsilon_i \sim N(0, 3)$, the EG and H^2 effects, which entered the regression model in steps 1 and 5, respectively, would require the E, G, and H factors to also be in the model.

Case 2 data was simulated according to the model

$$y_i = 2A_i - 1.5E_i + 2G_i + 4C_i - 3H_i + 2.5E_i^2 - 4G_iH_i + 3.5E_iH_i - 5C_iG_i + \varepsilon_i, \quad (4.3)$$

to represent a response exhibiting effect sparsity and strong heredity between active two-factor interactions or pure quadratic effects and their associated main effects. The model exhibits effect sparsity vice factor sparsity because although over 50% of

Table 30. Nine-Factor Simulated Response

Scenario	$\varepsilon \sim N(0,1)$				$\varepsilon \sim N(0,2)$				$\varepsilon \sim N(0,3)$			
Case	1	2	3	4	1	2	3	4	1	2	3	4
DSD	-6.5718	-2.0718	-1.0718	-6.0718	-5.0282	-0.5282	0.4718	-4.5282	-8.8568	-4.3568	-3.3568	-8.3568
	-8.9271	-6.4271	-6.4271	-8.4271	-6.3521	-3.8521	-3.8521	-5.8521	-8.4283	-5.9283	-5.9283	-7.9283
	6.7540	-0.2460	14.2540	-0.2460	8.5632	1.5632	16.0632	1.5632	5.3980	-1.6020	12.8980	-1.6020
	-5.3611	1.6389	13.1389	-12.3611	-7.5396	-0.5396	10.9604	-14.5396	-4.3540	2.6460	14.1460	-11.3540
	-1.0722	-4.0722	2.4278	-10.0722	-0.9806	-3.9806	2.5194	-9.9806	-1.5015	-4.5015	1.9985	-10.5015
	3.0372	-5.9628	-4.4628	-5.9628	-2.1549	-11.1549	-9.6549	-11.1549	-0.2920	-9.2920	-7.7920	-9.2920
	0.9647	2.9647	1.4647	2.9647	-2.2766	-0.2766	-1.7766	-0.2766	-0.9184	1.0816	-0.4184	1.0816
	-5.1812	10.8188	-7.6812	-3.1812	-7.7627	8.2373	-10.2627	-5.7627	-10.6842	5.3158	-13.1842	-8.6842
	-11.5682	8.9318	-4.5682	8.9318	-9.7422	10.7578	-2.7422	10.7578	-10.5137	9.9863	-3.5137	9.9863
	-11.6201	6.8799	-8.6201	8.8799	-11.1763	7.3237	-8.1763	9.3237	-7.5815	10.9185	-4.5815	12.9185
	5.9775	17.9775	13.4775	11.9775	5.9904	17.9904	13.4904	11.9904	1.9805	13.9805	9.4805	7.9805
	-6.1405	-8.1405	12.3595	-0.1405	-3.1020	-5.1020	15.3980	2.8980	-3.1339	-5.1339	15.3661	2.8661
	-0.4898	9.0102	4.5102	3.0102	0.0662	9.5662	5.0662	3.5662	-3.6061	5.8939	1.3939	-0.1061
	-8.4915	3.0085	3.5085	-4.9915	-5.7396	5.7604	6.2604	-2.2396	-7.1792	4.3208	4.8208	-3.6792
	-0.3253	-7.8253	3.1747	-7.8253	-3.9942	-11.4942	-0.4942	-11.4942	0.6261	-6.8739	4.1261	-6.8739
	1.6734	2.1734	-5.8266	-5.8266	-0.9510	-0.4510	-8.4510	-8.4510	4.5618	5.0618	-2.9382	-2.9382
	-13.0033	12.9967	-2.5033	12.9967	-17.4189	8.5811	-6.9189	8.5811	-14.8327	11.1673	-4.3327	11.1673
	-17.0536	6.9464	-3.5536	8.9464	-15.4480	8.5520	-1.9480	10.5520	-24.6111	-0.6111	-11.1111	1.3889
	-0.3768	6.6232	0.1232	0.6232	1.1869	8.1869	1.6869	2.1869	-1.4119	5.5881	-0.9119	-0.4119
	-3.3919	-10.3919	-5.8919	-2.3919	-6.2247	-13.2247	-8.7247	-5.2247	-4.5994	-11.5994	-7.0994	-3.5994
	1.8669	1.8669	1.8669	1.8669	-1.1295	-1.1295	-1.1295	-1.1295	0.0071	0.0071	0.0071	0.0071
DSD+	-16.4003	14.5997	-5.9003	8.5997	-14.3990	16.6010	-3.8990	10.6010	-12.5880	18.4120	-2.0880	12.4120
	-18.5988	-3.5988	-5.0988	-11.5988	-15.7068	-0.7068	-2.2068	-8.7068	-10.7264	4.2736	2.7736	-3.7264
	-4.2611	11.7389	-6.7611	-2.2611	-2.9713	13.0287	-5.4713	-0.9713	-3.9842	12.0158	-6.4842	-1.9842
	-4.0588	-11.0588	-6.5588	-3.0588	-3.5441	-10.5441	-6.0441	-2.5441	-0.1862	-7.1862	-2.6862	0.8138
	-1.2761	13.7239	2.2239	13.7239	0.7377	15.7377	4.2377	15.7377	-1.8346	13.1654	1.6654	13.1654
	4.3426	-4.6574	11.8426	-4.6574	5.5804	-3.4196	13.0804	-3.4196	4.6927	-4.3073	12.1927	-4.3073
	-5.3180	-16.3180	13.1820	-14.3180	-3.7730	-14.7730	14.7270	-12.7730	-2.7244	-13.7244	15.7756	-11.7244
	0.9107	0.9107	1.4107	0.9107	1.0700	1.0700	1.5700	1.0700	0.8791	0.8791	1.3791	0.8791

Table 31. Forward Stepwise Results: Case 1

	Scenario					
	$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,2)$		$\varepsilon \sim N(0,3)$	
Design	**DSD**	**DSD+**	**DSD**	**DSD+**	**DSD**	**DSD+**
Step	**Effects Added**					
1	EG	EG	AG	AG	EG	EG
2	AG	AG	EG	EG	AG	AG
3	AE	AE	EJ	AE	AE	AE
4	G^2	G^2	AE	G^2	DF	H^2
5	AJ	AJ	–	DF	H^2	AH
6	DH	GJ	–	CJ	–	BG
7	AD	A^2	–	–	–	DE
8	F	AD	–	–	–	FH
9	C	CH	–	–	–	AF
10	–	–	–	–	–	B^2

the factors (5 of 9) are active only 9 of 54 total effects are active, not coincidentally the same number as Case 1.

Table 32 provides Jones and Nachtsheim (2011a) forward stepwise regression using the Case 2 response data associated with each design for all four noise level scenarios in Tables 30.

Case 3 data was simulated according to the model

$$y_i = 2A_i + 2E_i - 1.5A_i^2 + 2.5E_i^2 - 3.5A_iE_i + 4A_iG_i - 5E_iG_i + \varepsilon_i, \qquad (4.4)$$

thereby representing a response which exhibits factor sparsity and weak heredity between active two-factor interactions or pure quadratic effects and main effects. The model exhibits factor sparsity because only 3 of the 9 factors are active within the 7 effects contained in the model. Since not all factors, which comprise the two-factor interactions, are present as a main effect, the model exhibits weak heredity. For instance, although factor G is significant within two two-factor interactions, factor G by itself is not significant.

Table 32. **Forward Stepwise Results: Case 2**

	Scenario					
	$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,2)$		$\varepsilon \sim N(0,3)$	
Design	**DSD**	**DSD+**	**DSD**	**DSD+**	**DSD**	**DSD+**
Step	**Effects Added**					
1	CG	CG	C^2	CG	C^2	CG
2	GH	GH	GH	GH	E^2	GH
3	EH	EH	CG	EH	DH	EH
4	A	A	CJ	AE	A^2	A^2
5	E^2	E^2	A	H^2	DG	DE
6	J^2	J^2	GJ	DF	AF	E^2
7	DH	DH	–	DE	–	AH
8	CF	CH	–	BJ	–	AE
9	–	CD	–	–	–	CF
10	–	CJ	–	–	–	–
11	–	D^2	–	–	–	–

Table 33 provides Jones and Nachtsheim (2011a) forward stepwise regression using the Case 3 response data associated with each design for all four noise level scenarios in Tables 30.

Table 33. **Forward Stepwise Results: Case 3**

	Scenario					
	$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,2)$		$\varepsilon \sim N(0,3)$	
Design	**DSD**	**DSD+**	**DSD**	**DSD+**	**DSD**	**DSD+**
Step	**Effects Added**					
1	AE	EG	AE	AE	EG	EG
2	BF	AG	CH	AG	AG	AG
3	J^2	AE	EJ	EG	AE	AE
4	A^2	E^2	J^2	D^2	DF	–
5	FH	J^2	E^2	AF	H	–
6	DJ	CE	–	DF	AF	–
7	E^2	–	–	BE	–	–

Case 4 data was simulated according to the model

$$y_i = 2A_i - 1.5E_i + 2G_i - 3H_i^2 + 2.5E_i^2 + 4A_iC_i + 3.5E_iH_i - 5C_iG_i - 4G_iH_i + \varepsilon_i \quad (4.5)$$

to represent a response which exhibits effect sparsity and weak heredity between active two-factor interactions or pure quadratic effects and main effects.

Table 34 provides Jones and Nachtsheim (2011a) forward stepwise regression using the Case 4 response data associated with each design for all four noise level scenarios in Tables 30.

Table 34. Forward Stepwise Results: Case 4

	Scenario					
	$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,2)$		$\varepsilon \sim N(0,3)$	
Design	**DSD**	**DSD+**	**DSD**	**DSD+**	**DSD**	**DSD+**
Step	**Effects Added**					
1	GH	GH	GH	AC	GH	GH
2	AH	CG	AE	CG	AH	CG
3	AF	AC	EG	EH	DE	AC
4	EF	EH	HJ	GH	AD	EH
5	G^2	J^2	E^2	AE	DG	DE
6	AC	E^2	J^2	DF	FH	A^2
7	DF	H^2	–	BH	A^2	E^2
8	J	EJ	–	–	–	AH
9	–	FH	–	–	–	FJ
10	–	CD	–	–	–	–
11	–	EF	–	–	–	–
12	–	DE	–	–	–	–

Tables 35, 36, 37, and 38 show which effects from Cases 1 through 4's four different noise level scenarios were properly identified, incorrectly identified (Type I error), and not identified (Type II error), for both the DSD and DSD+ based upon Jones and Nachtsheim (2011a) suggested analysis methodology.

In all four Cases, regardless of noise level, the DSD+ performance in identifying active effects met or exceeded the DSD performance. However, similar to the DSD, the DSD+ was still susceptible to increased Type II errors as the noise level increased. Fortunately, the DSD+ was more robust to the heredity (strong or weak) or sparsity (factor or effect) assumption than the DSD. When comparing strong heredity to weak heredity for DSD, the DSD performed better when strong heredity was exhibited, particularly when effect sparsity was present. In contrast, the DSD+ performed equally well under the heredity assumption. With regards to the sparsity assumption, the DSD+ showed better performance under effect sparsity than factor sparsity which was counter to the DSD. However, the DSD+ performance under factor sparsity assumption was still better than the DSD. Interestingly, all the Type II errors across all Scenarios and Cases made by the DSD+ involved not identifying active pure-quadratic effects.

Table 35. Second Order Screening Design Results: Case 1

Strong Heredity, Factor Sparsity Model: $2A - 1.5E + 2G - 3A^2 + 2.5E^2 - 4G^2 + 4AE + 3.5AG - 5EG + \varepsilon$			**Rep 1**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, G, G^2, AE, AG, EG	$A, E, G, A^2, G^2, AE, AG, EG$
	Type I errors	$C, D, F, H, J, AD, AJ, DH$	$C, D, H, J, AD, AJ, CH, GJ$
	Type II errors	A^2, E^2	E^2
$\varepsilon \sim N(0,2)$	**Identified**	A, E, G, AE, AG, EG	A, E, G, G^2, AE, AG, EG
	Type I errors	J, EJ	C, D, F, J, CJ, DF
	Type II errors	A^2, E^2, G^2	A^2, E^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E, G, AE, AG, EG	A, E, G, AE, AG, EG
	Type I errors	D, F, H, H^2, DF	B^2, D^2, AH, BG, DG
	Type II errors	A^2, E^2, G^2	A^2, E^2, G^2

4.5 Analysis Replication Results

In order to insure the increased performance exhibited by the DSD+ over the DSD was not limited to a single instance, the response data was replicated four additional times. Table 39 displays the average percentage of all active effects, second-order effects, and pure-quadratic effects correctly identified from five replications of all four Cases and three Scenarios. For instance, Case 3 (Weak Heredity, Factor Sparsity Model), Scenario 1 ($\varepsilon \sim N(0,1)$) shows on average the DSD correctly identified 80% of the active effects in model, 72% of the active second-order effects (two-way interactions and pure-quadratic effects), and 50% of the active pure-quadratic effects.

Table 36. Second Order Screening Design Results: Case 2

Strong Heredity, Effect Sparsity Model: $2A - 1.5E + 2G + 4C - 3H + 2.5E^2 - 5CG + 3.5EH - 4GH + \varepsilon$			Rep 1
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	$A, E, C, G, H, E^2, CG, EH, GH$	$A, E, C, G, H, E^2, CG, EH, GH$
	Type I errors	D, F, J, J^2, CF, DH	$D, J, D^2, J^2, CD, CH, CJ, DH$
	Type II errors	NONE	NONE
$\varepsilon \sim N(0,2)$	**Identified**	A, C, G, H, CG, GH	$A, E, C, G, H, CG, EH, GH$
	Type I errors	J, C^2, CJ, GJ	$B, D, F, J, H^2, AE, BJ, DE, DF$
	Type II errors	E, E^2, EH	E^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E, C, G, H, E^2	$A, E, C, G, H, E^2, CG, EH, GH$
	Type I errors	$D, F, A^2, C^2, AF, DG, DH$	$D, F, A^2, AE, AH, CF, DE$
	Type II errors	CG, EH, GH	NONE

Table 37. Second Order Screening Design Results: Case 3

Weak Heredity, Factor Sparsity Model: $2A + 2E - 1.5A^2 + 2.5E^2 - 3.5AE + 4AG - 5EG + \varepsilon$			Rep 1
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, A^2, E^2, AE	A, E, E^2, AE, AG, EG
	Type I errors	$B, D, F, H, J, J^2, BF, DJ, FH$	C, G, J, J^2, CE
	Type II errors	AG, EG	A^2
$\varepsilon \sim N(0,2)$	**Identified**	A, E, E^2, AE	A, E, AE, AG, EG
	Type I errors	C, H, J, J^2, CH, EJ	$B, D, F, G, D^2, AF, BE, DF$
	Type II errors	A^2, AG, EG	A^2, E^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E, AE, AG, EG	A, E, AE, AG, EG
	Type I errors	D, F, G, H, AF, DF	G
	Type II errors	A^2, E^2	A^2, E^2

Table 38. Second Order Screening Design Results: Case 4

Weak Heredity, Effect Sparsity Model: $2A - 1.5E + 2G + 2.5E^2 - 3H^2 + 4AC - 5CG + 3.5EH - 4GH + \varepsilon$			**Rep 1**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, G, AC, GH	$A, E, G, E^2, H^2, AC, CG, EH, GH$
	Type I errors	$C, D, F, H, J, G^2,$ AF, AH, DF, EF	$C, D, F, H, J, J^2,$ CD, DE, EF, EJ, FH
	Type II errors	E^2, H^2, CG, EH	NONE
$\varepsilon \sim N(0,2)$	**Identified**	A, E, G, E^2, GH	A, E, G, AC, CG, EH, GH
	Type I errors	H, J, J^2, AE, EG, HJ	$B, C, D, F, H, AE, BH, DF$
	Type II errors	H^2, AC, CG, EH	E^2, H^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E, G, GH	$A, E, G, E^2, AC, CG, EH, GH$
	Type I errors	$D, F, H, A^2, AD, AH,$ DE, DG, FH	$C, D, F, H, J, A^2,$ AH, DE, FJ
	Type II errors	E^2, H^2, AC, CG, EH	H^2

Additionally, Table 39 displays the average number of Type I errors made. Overall, the percentages show the DSD+ outperforms the DSD with regards to identifying active effects and their various subsets across the board with little to no increase in Type I errors. However, when the noise level increases neither the DSD nor the DSD+ are consistently finding the active pure-quadratic effects. The individual replication results are found in Tables 40 through 55 in the Appendix.

4.6 Conclusions

For a second-order polynomial model, if a factor screening design is not used, a design must contain enough degrees of freedom to estimate all effects. For k factors this equates to $\frac{(k+1)(k+2)}{2}$ design runs. As k increases, the number of required runs will quickly exceed the number of available runs provided to an experimenter, particularly within the DOD testing realm. As such as k increases, a screening design must be employed while maintaining the ability to estimate a second-order polynomial model when constraints dictate a single experiment. Jones and Nachtsheim (2011a)

Table 39. **Second Order Screening Design Results: Average**

Strong Heredity, Factor Sparsity Model:			**5 Rep Avg**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	67%, 50%, 20%	91%, 87%, 73%
	Type I errors	9.6	10.6
$\varepsilon \sim N(0,2)$	**Identified**	58%, 37%, 20%	84%, 77%, 53%
	Type I errors	7.4	5.6
$\varepsilon \sim N(0,3)$	**Identified**	51%, 33%, 27%	62%, 43%, 13%
	Type I errors	6.4	4.8
Strong Heredity, Effect Sparsity Model:			**5 Rep Avg**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	98%, 95%, 80%	98%, 95%, 80%
	Type I errors	6.6	7.6
$\varepsilon \sim N(0,2)$	**Identified**	78%, 55%, 20%	91%, 80%, 20%
	Type I errors	4.0	5.4
$\varepsilon \sim N(0,3)$	**Identified**	84%, 70%, 40%	93%, 85%, 40%
	Type I errors	4.2	4.2
Weak Heredity, Factor Sparsity Model:			**5 Rep Avg**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	80%, 72%, 50%	94%, 92%, 80%
	Type I errors	9.8	10
$\varepsilon \sim N(0,2)$	**Identified**	60%, 44%, 20%	77%, 68%, 20%
	Type I errors	8.0	7.0
$\varepsilon \sim N(0,3)$	**Identified**	51%, 32%, 0%	77%, 68%, 20%
	Type I errors	7.4	4.6
Weak Heredity, Effect Sparsity Model:			**5 Rep Avg**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	49%, 23%, 0%	93%, 90%, 70%
	Type I errors	10.6	11.2
$\varepsilon \sim N(0,2)$	**Identified**	47%, 23%, 10%	84%, 77%, 30%
	Type I errors	8.4	7.0
$\varepsilon \sim N(0,3)$	**Identified**	44%, 20%, 10%	82%, 73%, 20%
	Type I errors	7.6	6.6

Note: Identified percentages correspond to percentage of active effects, second-order effects, and pure quadratic effects.

proposed the economical three-level DSD for screening quantitative factors in the presence of active second-order effects. Dougherty et al. (2013b) showed the DSD were effective in identifying active main effects regardless of the heredity and sparsity assumption but lacked the power to differentiate between active second-order effects when both two-factor interactions and pure-quadratic effects are active. We introduce a way to augment the DSD, deemed DSD+, with $k-1$ runs which increased the detection performance of active second-order effects involving a particular factor of interested. The $k-1$ additional runs can be run as part of a single experiment with the original DSD if the experimenter has *a priori* knowledge or as part of a follow-on experiment based upon *a posteriori* knowledge. Furthermore, while the additional runs are optimized for two-factor interactions, the impact of adding additional center point runs on identifying active pure-quadratic effects requires further investigation.

While the $k-1$ runs are associated with the $k-1$ two-factor interactions of a single factor of interest in a k factor experiment, the manner in which the DSD is augmented can easily be extended to additional factors. For instance, the DSD can be augmented with $k-1+k-2=2k-3$ runs for all the two-factors interactions of two factors and so on until a total of $\frac{(k)(k-1)}{2}$ runs are added for all the two-factor interactions in a k factor experiment. As such, DSD can be tailored with augmentation runs which take the DSD from the standard $2k+1$ runs all the way to $\frac{(k+1)(k+2)}{2}$ runs for a saturated second-order design.

Appendix

Table 40. Second Order Screening Design Results: Case 1

Strong Heredity, Factor Sparsity Model: $2A - 1.5E + 2G - 3A^2 + 2.5E^2 - 4G^2 + 4AE + 3.5AG - 5EG + \varepsilon$			**Rep 2**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, G, A^2, EG	$A, E, G, A^2, E^2, G^2, AE, AG, EG$
	Type I errors	$B, C, D, F, B^2, D^2,$ BE, BF, CE, DF	$B, C, F, H, F^2, BE,$ BF, CG, EH, FH
	Type II errors	E^2, G^2, AE, AG	NONE
$\varepsilon \sim N(0,2)$	**Identified**	A, E, G, EG	$A, E, G, A^2, G^2, AE, AG, EG$
	Type I errors	$B, D, F, J, F^2, J^2,$ AF, BD, BF, DJ, FG	D, J, J^2, DG, DJ
	Type II errors	A^2, E^2, G^2, AE, AG	E^2
$\varepsilon \sim N(0,3)$	**Identified**	E, E^2	A, E, G, AE, AG, EG
	Type I errors	C, H, J, CH, EJ	B, F, BF
	Type II errors	$A, G, A^2, G^2, AE,$ AG, EG	A^2, E^2, G^2

Table 41. Second Order Screening Design Results: Case 2

Strong Heredity, Effect Sparsity Model: $2A - 1.5E + 2G + 4C - 3H + 2.5E^2 - 5CG + 3.5EH - 4GH + \varepsilon$			**Rep 2**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	$A, E, C, G, H, E^2,$ CG, EH, GH, EH, GH	$A, E, C, G, H, E^2,$ CG, EH, GH, EH, GH
	Type I errors	B, F, AF, BC, BG	$B, D, F, J, AH,$ BE, DJ, EG, FG
	Type II errors	NONE	NONE
$\varepsilon \sim N(0,2)$	**Identified**	A, E, C, G, H	$A, E, C, G, H, CG, EH, GH$
	Type I errors	$B, D, J, AD, BE, CE,$ CH, DG, EJ	B, D, AB, AD, BD, DG
	Type II errors	E^2, CG, EH, GH	E^2
$\varepsilon \sim N(0,3)$	**Identified**	E, C, G, H, CG, EH, GH	$A, E, C, G, H, CG, EH, GH$
	Type I errors	CH	A^2, CH
	Type II errors	A, E^2	E^2

Table 42. Second Order Screening Design Results: Case 3

Weak Heredity, Factor Sparsity Model: $2A + 2E - 1.5A^2 + 2.5E^2 - 3.5AE + 4AG - 5EG + \varepsilon$			**Rep 2**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	$A, E, A^2, E^2, AE, AG, EG$	$A, E, A^2, E^2, AE, AG, EG$
	Type I errors	B, C, F, G, BE, CE, CF BE, BF, CG, FH	$B, C, F, G, H, J, J^2,$
	Type II errors	NONE	NONE
$\varepsilon \sim N(0,2)$	**Identified**	A, E, A^2, AE, AG, EG	A, E, A^2, AE, AG, EG
	Type I errors	B, D, G, J, B^2, BD, EJ	D, G, J, J^2, DG
	Type II errors	E^2	E^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E, AE, AG, EG	A, E, AE, AG, EG
	Type I errors	C, G, H, CH	B, F, G, BF
	Type II errors	A^2, E^2	A^2, E^2

Table 43. **Second Order Screening Design Results: Case 4**

Weak Heredity, Effect Sparsity Model: $2A - 1.5E + 2G + 2.5E^2 - 3H^2 + 4AC - 5CG + 3.5EH - 4GH + \varepsilon$			**Rep 2**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, G, EH, GH	$A, E, G, E^2, H^2, AC, CG, EH, GH$
	Type I errors	$B, C, F, H, G^2, AF, AH, BE, CF, EF$	$B, C, F, H, B^2, AB, BC, BE, BG, CF$
	Type II errors	E^2, H^2, AC, CG	NONE
$\varepsilon \sim N(0,2)$	**Identified**	A, E, G, GH	$A, E, G, E^2, AC, CG, EH, GH$
	Type I errors	$B, C, D, F, H, J, D^2, AB, AJ, BF, CE$	$B, C, D, H, J, G^2, AB, AH, BD, BG, BH$
	Type II errors	E^2, H^2, AC, CG, EH	H^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E, G	A, E, G, AC, CG, EH, GH
	Type I errors	B, D, AD, BD, BE	C, H, CH
	Type II errors	E^2, H^2, AC, CG, EH, GH	E^2, H^2

Table 44. **Second Order Screening Design Results: Case 1**

Strong Heredity, Factor Sparsity Model: $2A - 1.5E + 2G - 3A^2 + 2.5E^2 - 4G^2 + 4AE + 3.5AG - 5EG + \varepsilon$			**Rep 3**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, G, AE, AG, EG	A, E, G, G^2, AE, AG, EG
	Type I errors	$B, C, D, F, H, J, J^2, AB, DH, EJ$	C, D, F, H, J, AH, CF
	Type II errors	A^2, E^2, G^2	A^2, E^2
$\varepsilon \sim N(0,2)$	**Identified**	A, E, G, G^2, EG	A, E, G, G^2, AE, AG, EG
	Type I errors	$B, D, F, H, B^2, AH, BF, EF, EH$	$B, C, D, F, AF, BC, CD, CF$
	Type II errors	A^2, E^2, AE, AG	A^2, E^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E, G, A^2, G^2, EG	A, E, G
	Type I errors	$B, C, D, F, H, J, BF, CJ, DF, EH$	$B, F, H, J, J^2, BJ, EH, EJ, FG$
	Type II errors	E^2, AE, AG	$A^2, E^2, G^2, AE, AG, EG$

Table 45. **Second Order Screening Design Results: Case 2**

Strong Heredity, Effect Sparsity Model: $2A - 1.5E + 2G + 4C - 3H + 2.5E^2 - 5CG + 3.5EH - 4GH + \varepsilon$			**Rep 3**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	$A, E, C, G, H, E^2, CG, EH, GH$	$A, E, C, G, H, E^2, CG, EH, GH$
	Type I errors	B, D, F, BH, CD, FG	$B, D, F, J, G^2, AB, AJ, FG$
	Type II errors	NONE	NONE
$\varepsilon \sim N(0,2)$	**Identified**	$A, E, C, G, H, CG, EH, GH$	$A, E, C, G, H, CG, EH, GH$
	Type I errors	F, AF	$B, D, F, G^2, AE, AF, AH, BF, DH$
	Type II errors	E^2	E^2
$\varepsilon \sim N(0,3)$	**Identified**	$A, E, C, G, H, E^2, CG, EH, GH$	$A, E, C, G, H, E^2, CG, EH, GH$
	Type I errors	B, F, J, AB, CE, CF	A^2
	Type II errors	NONE	NONE

Table 46. **Second Order Screening Design Results: Case 3**

Weak Heredity, Factor Sparsity Model: $2A + 2E - 1.5A^2 + 2.5E^2 - 3.5AE + 4AG - 5EG + \varepsilon$			**Rep 3**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, AE, AG, EG	A, E, E^2, AE, AG, EG
	Type I errors	$D, F, G, H, J, D^2, AJ, DH$	$B, C, D, F, G, H, J, B^2, D^2, AJ, CD, DG, FH, GJ$
	Type II errors	A^2, E^2	A^2
$\varepsilon \sim N(0,2)$	**Identified**	A, E, AE	A, E, AE, AG, EG
	Type I errors	$B, D, F, G, J, G^2, AD, BE, BF, DJ, EF$	$B, C, D, F, G, B^2, AC, AD, AF, BC, BF, DF$
	Type II errors	A^2, E^2, AG, EG	A^2, E^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E	A, E, E^2, AE, AG, EG
	Type I errors	$B, C, D, F, G, H, J, B^2, F^2, J^2, AD, AJ, CE, DH$	G, H, EH
	Type II errors	A^2, E^2, AE, AG, EG	A^2

Table 47. Second Order Screening Design Results: Case 4

Weak Heredity, Effect Sparsity Model: $2A - 1.5E + 2G + 2.5E^2 - 3H^2 + 4AC - 5CG + 3.5EH - 4GH + \varepsilon$			**Rep 3**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, G, GH	A, E, G, AC, CG, EH, GH
	Type I errors	$C, D, F, H, AD, AH, CE, DE, DG$	$B, C, D, F, H, J, G^2, AB, AG, AJ, DJ, FJ, GJ$
	Type II errors	E^2, H^2, AC, CG, EH	E^2, H^2
$\varepsilon \sim N(0,2)$	**Identified**	A, E, G	A, E, G, AC, CG, EH, GH
	Type I errors	$B, C, D, F, D^2, AG, BD, BF, CE, CF, FG$	C, H, J, J^2
	Type II errors	E^2, H^2, AC, CG, EH, GH	E^2, H^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E, G, AC, CG, EH, GH	A, E, G, AC, CG, EH, GH
	Type I errors	$B, C, D, F, H, J, AF, CE, DJ$	$B, C, F, H, AB, AH, BF, CF, EF, FH$
	Type II errors	E^2, H^2	E^2, H^2

Table 48. Second Order Screening Design Results: Case 1

Strong Heredity, Factor Sparsity Model: $2A - 1.5E + 2G - 3A^2 + 2.5E^2 - 4G^2 + 4AE + 3.5AG - 5EG + \varepsilon$			**Rep 4**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, G, E^2, AG, EG	$A, E, G, E^2, G^2, AE, AG, EG$
	Type I errors	$B, C, D, F, H, J, AD, BF, BG, CH, EJ, FJ$	$B, C, D, F, H, J, F^2, H^2, AC, BC, CD, CG, DE, EH$
	Type II errors	A^2, G^2, AE	A^2
$\varepsilon \sim N(0,2)$	**Identified**	A, E, G, G^2, AE, EG	A, E, G, G^2, AE, AG, EG
	Type I errors	B, D, F, AB, BF, DF	B, F, H, BF, FH, GH
	Type II errors	A^2, E^2, AG	A^2, E^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E, G, G^2, AG	A, E, G, E^2, AG, EG
	Type I errors	B, F, BE, BF	J, AJ, GJ
	Type II errors	A^2, E^2, AE, AG	A^2, G^2, AE

Table 49. Second Order Screening Design Results: Case 2

Strong Heredity, Effect Sparsity Model: $2A - 1.5E + 2G + 4C - 3H + 2.5E^2 - 5CG + 3.5EH - 4GH + \varepsilon$			**Rep 4**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	$A, E, C, G, H, CG, EH, GH$	$A, E, C, G, H, CG, EH, GH$
	Type I errors	$B, D, F, J, A^2, J^2, BG, EG, FG$	J, A^2, G^2, J^2, AC
	Type II errors	E^2	E^2
$\varepsilon \sim N(0,2)$	**Identified**	$A, E, C, G, H, CG, EH, GH$	$A, E, C, G, H, CG, EH, GH$
	Type I errors	B	NONE
	Type II errors	E^2	E^2
$\varepsilon \sim N(0,3)$	**Identified**	$A, E, C, G, H, CG, EH, GH$	$A, E, C, G, H, CG, EH, GH$
	Type I errors	F, AF, EF, FH	D, F, CD, FH
	Type II errors	E^2	E^2

Table 50. Second Order Screening Design Results: Case 3

Weak Heredity, Factor Sparsity Model: $2A + 2E - 1.5A^2 + 2.5E^2 - 3.5AE + 4AG - 5EG + \varepsilon$			**Rep 4**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, E^2, AE, AG, EG	$A, E, A^2, E^2, AE, AG, EG$
	Type I errors	$B, C, D, F, H, J, J^2, AC, BD, BJ$	$B, C, D, F, G, H, J, B^2, C^2, G^2, J^2, BG, CE, DF, EH, EJ, HJ$
	Type II errors	A^2	NONE
$\varepsilon \sim N(0,2)$	**Identified**	A, E, AE	A, E, AE, AG, EG
	Type I errors	$B, F, H, J, B^2, J^2, AH, BF$	G, H, GH
	Type II errors	A^2, E^2, AG, EG	A^2, E^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E, AE	A, E, AE, AG, EG
	Type I errors	B, F, BE, BF	F, G, J, J^2, FJ, GJ
	Type II errors	A^2, E^2, AG, EG	A^2, E^2

Table 51. Second Order Screening Design Results: Case 4

Weak Heredity, Effect Sparsity Model: $2A - 1.5E + 2G + 2.5E^2 - 3H^2 + 4AC - 5CG + 3.5EH - 4GH + \varepsilon$			**Rep 4**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, G, GH	$A, E, G, H^2, AC, CG, EH, GH$
	Type I errors	$B, C, D, F, H, J, B^2, AD, AH, CD, DE, DG, GJ$	$C, D, H, J, A^2, AD, AH, CE, CJ$
	Type II errors	E^2, H^2, AC, CG, EH	E^2
$\varepsilon \sim N(0,2)$	**Identified**	A, G, CG, GH	A, E, G, AC, CG, EH, GH
	Type I errors	B, C, H, J, A^2, AB, CJ	C, D, H, AD, CD
	Type II errors	E, E^2, H^2, AC, EH	E^2, H^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E, E^2	$A, E, G, H^2, AC, CG, EH, GH$
	Type I errors	F, H, AF, AH, FH	C, D, F, H, CD, FH
	Type II errors	G, H^2, AC, CG, EH, GH	E^2

Table 52. Second Order Screening Design Results: Case 1

Strong Heredity, Factor Sparsity Model: $2A - 1.5E + 2G - 3A^2 + 2.5E^2 - 4G^2 + 4AE + 3.5AG - 5EG + \varepsilon$			**Rep 5**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, G, AE, AG, EG	$A, E, G, A^2, E^2, G^2, AE, AG, EG$
	Type I errors	$B, C, F, J, B^2, J^2, CE, EJ$	$B, C, D, F, H, J, C^2, D^2, AB, AC, BJ, DE, FJ, HJ$
	Type II errors	A^2, E^2, G^2	$NONE$
$\varepsilon \sim N(0,2)$	**Identified**	A, E, G, A^2, EG	$A, E, G, A^2, E^2, G^2, AE, AG, EG$
	Type I errors	$B, C, D, F, J, F^2, BF, CJ, DF$	C, F, AF
	Type II errors	E^2, G^2, AE, AG	$NONE$
$\varepsilon \sim N(0,3)$	**Identified**	A, E, G, EG	A, E, G, A^2, AE, AG, EG
	Type I errors	$C, D, J, D^2, AJ, CE, CG, EJ$	D, J, J^2, EJ
	Type II errors	A^2, E^2, G^2, AE, AG	E^2, G^2

Table 53. Second Order Screening Design Results: Case 2

Strong Heredity, Effect Sparsity Model: $2A - 1.5E + 2G + 4C - 3H + 2.5E^2 - 5CG + 3.5EH - 4GH + \varepsilon$			**Rep 5**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	$A, E, C, G, H, E^2, CG, EH, GH$	$A, E, C, G, H, E^2, CG, EH, GH$
	Type I errors	B, D, F, J, BD, CH, FG	$B, F, J, A^2, AF, BC, BG, EJ$
	Type II errors	NONE	NONE
$\varepsilon \sim N(0,2)$	**Identified**	$A, E, C, G, H, E^2, CG, EH,$	$A, E, C, G, H, E^2, CG, EH, GH$
	Type I errors	F, AC, CE, EF	F, G^2, AF
	Type II errors	GH	NONE
$\varepsilon \sim N(0,3)$	**Identified**	$A, E, C, G, H, CG, EH, GH$	$A, E, C, G, H, CG, EH, GH$
	Type I errors	J, J^2, AJ	B, D, J, J^2, AE, BG, CD
	Type II errors	E^2	E^2

Table 54. Second Order Screening Design Results: Case 3

Weak Heredity, Factor Sparsity Model: $2A + 2E - 1.5A^2 + 2.5E^2 - 3.5AE + 4AG - 5EG + \varepsilon$			**Rep 5**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, A^2, AE, EG	$A, E, A^2, E^2, AE, AG, EG$
	Type I errors	$B, C, D, F, G, J,$ B^2, J^2, BD, BF, CE	B, C, F, G, J, BJ, CF
	Type II errors	E^2, AG	NONE
$\varepsilon \sim N(0,2)$	**Identified**	A, E, AE, AG, EG	A, E, E^2, AE, AG, EG
	Type I errors	$C, F, G, J, F^2, AC, AJ, CF$	C, F, G, H, AF, CF, GH
	Type II errors	A^2, E^2	A^2
$\varepsilon \sim N(0,3)$	**Identified**	A, E, AE	A, E, A^2, AE, AG, EG
	Type I errors	$B, C, F, G, J, G^2, BF, EJ, GJ$	$B, C, G, J, J^2, AB, BC, BG, EJ$
	Type II errors	A^2, E^2, AG, EG	E^2

Table 55. **Second Order Screening Design Results: Case 4**

Weak Heredity, Effect Sparsity Model: $2A - 1.5E + 2G + 2.5E^2 - 3H^2 + 4AC - 5CG + 3.5EH - 4GH + \varepsilon$			**Rep 5**
Scenario		**DSD**	**DSD+**
$\varepsilon \sim N(0,1)$	**Identified**	A, E, G, GH	$A, E, G, E^2, H^2, AC, CG, EH, GH$
	Type I errors	$B, C, F, H, J, B^2, G^2, AE, AF, AH, EF$	$B, C, D, F, H, J, D^2, AF, BE, CE, DF, DH, FJ$
	Type II errors	E^2, H^2, AC, CG, EH	NONE
$\varepsilon \sim N(0,2)$	**Identified**	A, E, G, AC, CG	$A, E, G, E^2, H^2, AC, CG, EH, GH$
	Type I errors	$C, F, C^2, G^2, AF, CE, FG$	$C, F, H, A^2, G^2, AF, CE$
	Type II errors	E^2, H^2, EH, GH	NONE
$\varepsilon \sim N(0,3)$	**Identified**	A, E, G	A, E, G, AC, CG, EH, GH
	Type I errors	$B, C, D, F, J, G^2, AB, BD, BF, CE$	C, H, J, J^2, EG
	Type II errors	E^2, H^2, AC, CG, EH, GH	E^2, H^2

V. Nonlinear Screening Designs for Defense Testing: An Overview and Case Study

5.1 Introduction

"Necessity is the Mother of Invention." Plato is often credited with authoring this quote, but whether he is the true author or not remains unknown. However, not knowing the author does not diminish the meaning and impact this simple quote has with regards to the situation the Department of Defense (DOD) and in particular the Defense Acquisition Test and Evaluation community find themselves in today. Available resources, whether they be personnel, budgets, or facilities, are continuing to shrink. Meanwhile, acquisition efforts are reducing timelines, even though systems are becoming increasingly more complex. As a result, testing methodologies which optimize the employment of resources are gaining emphasis and acceptance.

In a 2010 memorandum, Dr. Gilmore provided key policy guidance on the use of Design of Experiments (DOE) in OT&E. Furthermore the DOT&E Scientific Advisor (SA), Dr. Catherine Warner, highlighted the fact that while DOE is a structured, rigorous statistical tool for test planning and analysis, and it has been written about extensively within the academic setting, there are still many questions regarding how to apply DOE to T&E within DOD (Warner, 2011).

In April 2012, Dr. Steven Hutchison, Principal Deputy, Office of the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation (DASD(DT&E)), stated, "By applying scientific methods to the test design, we can not only achieve great efficiencies, but we can significantly improve confidence in our results. The Scientific Test and Analysis Techniques in Test & Evaluation Center of Excellence

(STAT T&E COE) will provide a critical venue for enhancing the test design for DOD acquisition programs."

Finally, in July 2013, Dr. Gilmore published a best practices memorandum for the statistical adequacy of operational test and evaluation (Gilmore, 2013). Included as an attachment to the memo was a white paper of best practices. One of the four identified test objectives was screen for important factors that affect system performance. While an important characteristic of test, the included potentially useful experimental designs focused just on linear effects.

This paper presents the use of design of experiments and response surface designs in the area of second-order screening designs, particularly as applied to defense testing. Extensions to existing designs are examined with respect to improvements in robustness and applicability to defense testing. A wind tunnel case study to demonstrate a viable use of a second-order screening design.

5.2 Background

Military systems, particularly aerodynamic systems, are complex. It is not unusual for these systems to exhibit nonlinear behavior. Developmental testing may be tasked to characterize the nonlinear behavior of such systems while being asked to reduce the amount of testing accomplished.

The one-factor-at-a-time (OFAT) experimentation strategy consists of successively varying each factor independently over its range while holding all the remaining factors at baseline settings. The OFAT method saturates the experimental design space and provides the capacity to determine how each factor affects the response variable while all other factors are held constant. An overwhelming disadvantage of the OFAT strategy is that it does not consider any possible interaction between factors. Additionally, as Hill et al. (2011) point out, the OFAT strategy is not cost effective nor does it control experimental uncertainty or produce minimum variance predictions.

Despite the disadvantages and deficiencies of implementing an OFAT experimentation strategy, conventional wind tunnel tests, which are a critical factor in the Developmental Test and Evaluation (DT&E) of aeronautical systems, are usually conducted in a manner consistent with the OFAT methodology. Hill et al. (2011) point out current high-performance military aircraft development programs require up to 3700 wind tunnel test hours in the conceptual design phase and up to 18,500 hours in the development/validation phase. Such requirements become tenuous during a time when the United States Air Force (USAF) and DOD are seeking reductions in developmental schedules and budgetary requirements.

DOE, or experimental design, is a statistical technique used to organize an experimental test or series of tests in a manner such that observed changes in an output response can be attributed to systematic changes made to the input variables of a process or system (Montgomery, 2013). While the designs are based upon statistical techniques, the actual design forms vary greatly depending upon the form of the empirical model used to represent the process or system response. Typically, first-order polynomial models are used extensively in screening experiments while second-order polynomial models are commonly used in modeling and optimization experiments.

When a system or process is new, screening designs are usually performed to determine which of the many factors (if any) have a significant effect on the system or process response. Screening designs usually assume a linear (main effects or main effects plus interaction) response so as to not waste valuable resources when experimenters do not know much about the system or process being studied. The assumption that the response is approximately linear for many factor screening experiments is reasonable when a system or process is just starting to be studied. However, there are times when subject matter expertise or historical data indicates a second-order polynomial response is more reasonable.

Traditionally, when a response is suspected of exhibiting second-order behavior, experiments are conducted sequentially. First, screening for important factors is conducted assuming a linear response in main effects and then follow-on experiments focus on second-order models using those important factors identified in the screening experiment. However, there are times when conducting multiple experiments sequentially is unrealistic due to time, budget, or other constraints. For instance, within the agricultural field the time duration of the design can be exceedingly long and/or within a manufacturing setting experimental preparation can be overly time-consuming. Directly applicable to the DOD, Lawson (2003) points out fixed deadlines for scale-up and production of prototype engineering designs may not allow the possibility of follow-up experimentation.

In these instances, it would be better, if not necessary, to perform factor screening and response surface exploration on the same experiment vice conducting experiments sequentially. This has significant implications for experimental screening designs. Second-order screening designs are extremely important when working with new or existing systems/technologies believed to exhibit nonlinear system responses so valuable resources will not be wasted using best guess and one-factor-at-a-time (OFAT) approaches. In the following sections, we provide an overview of screening designs and illustrate the use of a single augmented Definitive Screening Designs (DSD+) for determining significant factors associated with transonic and supersonic subspace wind tunnel testing data.

5.3 Screening Designs Overview

Many experiments start by considering many factors, which in turn increases the overall size and cost of the experiment. Since in reality no two experiments are exactly the same, a multitude of screening designs are available for use depending upon any number of variables. For the novice practitioner, the most obvious variables are the number of factors being considered and the number of available experimental runs. However, variables such as the range over which factors will be varied and the number of distinct levels at which runs will be made are equally important. More importantly are issues associated with whether or not follow-up experiments will be available, the shape of the design region, the ease of which statistical analysis of data can be done and subject matter expertise affect screening design selection. Lastly, design selection depends greatly upon the form of the empirical model used to represent the process or system response.

Traditionally, research has involved concepts like design resolution, minimum aberration, power, the number of clear (non-confounded) effects, concepts like rotatability, alphabetical-optimality, and prediction variance.

Resolution, generally denoted in roman numerals, is the measure of the degree of complete confounding for main effects and interactions in a fractional factorial design. The confounding characteristics of these design resolutions are:

- Res III: Main effects clear of other main effects, at least one main effect is confounded with at least one two-way interactions.

- Res IV: Main effects are clear of two-way interactions, but at least one two-way interaction is confounded with at least one other two-way interaction.

- Res V: Main effects and two-way interaction are clear of any other main effect or two-way interaction, but at least one two-way interaction is confounded with at least one three-way interaction.

There are times however that different designs can possess the same resolution and fractionation but have different confounding or aliasing structure. Fries and Hunter (1980) proposed the concept of design aberration for regular two-level designs as a means to differentiate between these designs. Since Fries and Hunter initial work, the concept of minimum aberration criterion has been extended to two-level non-regular, multilevel, and mixed-level fractional-factorial designs (Guo et al., 2009).

Optimal designs are typically assessed based upon specific criteria like providing good estimation of model parameters or good prediction capacity within the design region. *Alphabetic*-optimality refers to the family of design optimality criteria that are characterized by a letter of the alphabet, currently $A-$, $D-$, $G-$, $V-$, or $I-$. These *alphabetical*-optimality criteria drive what constitutes an optimal design. These "optimal" designs are rather focused on a particular design characteristic. Two of the most popular methods of assessing optimality are $I-$ and $D-$optimality. Where $D-$optimal designs focus on good model parameter estimates, and $I-$optimal designs focus on good prediction capacity within the design region by focusing on the scaled prediction variance. Since $I-$ criteria are prediction-oriented and $D-$ criteria are parameter-oriented, they are mostly used for second-order and first-order designs, respectively. For more on *alphabetic*-optimality, please see Chapter 8 in (Myers and Anderson-Cook, 2009).

Screening designs usually assume a linear (main effects or main effects plus interaction) response so factors can be studied at two levels thereby conserving experimental resources. Popular experimental regular and nonregular designs used in screening experiments are full and fractional 2-level factorial designs, Plackett-Burman, and supersaturated designs. Regular designs are designs constructed through defining relations among its factors, whereas, nonregular designs lack such a defining relation.

The 2^k Factorial Design consists of k factors each at only two levels and is a special case of the full factorial design with 2^k observations per replication. 2^k designs have many useful properties. In addition to being orthogonal, 2^k designs are $I-$optimal for fitting a first-order model or first-order model with interactions (Montgomery, 2013). The 2^k-type designs are widely used for factor screening as it provides the smallest number of runs for independently estimating all main effects and interactions for k factors.

The 2^{k-p} Fractional Factorial Design uses a subset of the runs of the 2^k Factorial Design. Similar to the 2^k Factorial Design, the 2^{k-p} Fractional Factorial Designs consists of k factors each at only two levels. However, the value of p specifies the degree to which the design is fractionated, determined by $1/2^p$. Generally, the first $k - p$ independent columns are generated by the runs in the 2^{k-p} design. In the 2^{k-p} design, the first $k - p$ columns are generated by the runs associated with the 2^{k-p} full factorial design. The remaining p columns can be generated as interactions of the first $k - p$ columns (Wu and Hamada, 2011). Because the design generators were determined by column interactions, the p factor effect estimates are aliased, meaning the factor effects on the system response can not be estimated separately from factor interactions. The degree to which the effects are aliased is given by the design resolution.

Plackett and Burman (1946) developed nonregular two-level fractional factorial designs which can study $k = N - 1$ variables in N runs, where N is a multiple of 4. If $N = 2^i$ for $i \geq 2$, PB designs are synonymous with 2^k factorial designs. The nonregular Plackett-Burman designs sacrifice a simple alias structure for better run economy and projectivity when compared to regular 2^{k-p} designs. Unfortunately, PB designs have complex alias structures. As a result, analysis of PB designs can become complex. Hamada and Wu (1992) discuss methods for analyzing designs with complex aliasing based upon the *sparsity of effect* and *effect heredity* principles.

Supersaturated designs are nonregular fractional factorial design where the number of factors k under investigation exceeds the number of available experimental runs N. Since $k > N - 1$, the degrees of freedom within the design are insufficient to estimate all the main effects and the design matrix cannot be orthogonal. Therefore in order for supersaturated designs to useful as screening designs only a few factors can be active. As such supersaturated designs are generally used when the number of potential factors is large but few are believed to have actual effects (effect sparsity) and either budget or time constraints limit the number of experimental runs. Some care must be exercised in the selection of a SSD. Since SSD can not obtain orthogonality, the SSD could produce misleading results if the design departs considerably from an orthogonal design. $E(s^2)$ gives an intuitive measure of nonorthogonality the smaller, the better.

When the response is believed to possess significant curvature, each factor needs at least three levels. If follow-on experiments are available, two-level regular designs can be augmented with follow-on design runs to accommodate curvature. However there are designs which are robust to the linear effect assumption such as the 3^k or 3^{k-p} fractional factorial design, the *Central Composite Design* (CCD), *Box-Behnken Design* (BBD), and saturated/near-saturated *Hoke*, *Hybrid*, and *Small Composite Designs* (SCD).

The 3^k Factorial Design, which consists of k factors each at only three levels, is a special case of the full factorial design with 3^k observations per replication. The addition of a third factor level over the 2^k design allows the response to be modeled as a quadratic function.

The 3^{k-p} Fractional Factorial Designs consists of k factors each at three levels. The value of p again specifies the degree to which the design is fractionated, determined by $1/3^p$. A general procedure for constructing a 3^{k-p} fractional factorial design is given by Montgomery (2013). Connor and Zelen (1959) and Xu (2005) provide an extensive list of 3^{k-p} designs. Unfortunately, especially as compared to 2^{k-p} designs, the aliasing structure for 3^{k-p} designs is very complex especially as the level of fractioning increases. If effect interactions are not negligible, design results can be difficult if not nearly impossible to interpret because of the partial aliasing of two-degree-of-freedom components (Montgomery, 2013).

Box and Wilson (1951) introduced an alternative class of designs to the 3^k factorial designs. The Central Composite Designs (CCD) contain a 2^k or 2_V^{k-p} design, axial/star runs, and center runs which are set at the middle of the factor range. The axial/star runs are selected so as to maintain a rotatable or near-rotatable design so that the variance of predicted response is constant (Montgomery, 2013). As such, the CCD typically involve k factors at 5 levels per factor. The CCD are popular design because of the sequential nature in which they can be implemented.

Box and Behnken (1960) developed a family of efficient rotatable/near-rotatable spherical three-level designs suitable for fitting second-order (quadratic) response models. In contrast to the CCD, the Box-Behnken design does not contain any points at the vertices or face-center of the design but rather at the center of the edges of the process space. As a result, the Box-Behnken designs avoid extreme values for factor-level combinations which may be impossible to test due to cost or physical process constraints (Montgomery, 2013). The BBD are formed by varying p parameters in a full factorial manner while the remaining $k - p$ parameters are kept steady at the center factor level setting. Additionally, the BBD uses three to five center runs to avoid singularity in the design matrix and to maintain favorable design qualities

(Myers and Anderson-Cook, 2009). Overall, the design run requirements for both the BBD and CCD are comparable. As a result, the benefit of employing a BBD design over a CCD is not necessarily due to run efficiency but rather the factor level combination location in the design space.

Through the years some of the original BBD have been improved upon in terms of rotatability, average prediction variance, $D-$ and $G-$efficiency (Nguyen and Borkowski, 2008). In addition, new Box-Behnken type designs with larger k (Mee, 2000) and differing orthogonally blocked solutions (Nguyen and Borkowski, 2008) than the original BBD have been proposed. Most recently *small* Box-Behnken Designs (SBBD) have been proposed which reduce the run size requirement of the original BBD by replacing the full 2^k factorial designs partly by 2^{3-1}_{III} designs and partly by full factorial designs (Zhang et al., 2011). When compared to the original BBD, the SBBD possess smaller $D-$efficiency values but the values are still relatively high ($> 70\%$) for $k \leq 11$ while requiring fewer runs.

While reduced run designs like the CCD and the BBD provide more efficient designs than the full model estimable designs 2^k and 3^k, these designs still can possess far more design points than needed to estimate the second-order response effects. As a result, the class of saturated or near-saturated designs have been developed. Saturated or near-saturated designs are designs such that the number of design points are equal to or near, but not less than, the number of terms in the design model.

Hoke (1974) presented a class of second-order designs for $k = 3$ to 6 factors at 3 levels based on saturated and near-saturated irregular fractions of the 3^k factorial. For each number of factors k, several versions of the Hoke designs exist consisting of a mixture of factorial, axial, and edge points making the Hoke designs suitable for a cuboidal region of interest (Myers and Anderson-Cook, 2009).

Roquemore (1976) presented a set of saturated or near-saturated second-order designs for $k = 3$ to 6 factors which are rotatable or near-rotatable while achieving the same degree of orthogonality as a CCD. The hybrid designs for k variables is constructed by first augmenting a $k - 1$ variable central composite design with an additional column for variable k. The design is then augmented with additional runs for variable k at different levels to create desirable design properties.

In contrast to the CCD, which contain a 2^k or 2_V^{k-p} factorial design, Hartley (1959) suggested replacing the factorial design with a special resolution III factorial design, where two-factor interactions are not aliased with other two-factor interactions. As a result, the number of design runs is decreased resulting in *Small Composite Designs (SCD).* The SCD sacrifices good prediction variance properties with the reduction in run size because main effects could be aliased with two-factor interactions. However, the SCD design still allows for the estimation of all main-effect because the star portion of the design provides additional information.

Unfortunately, while the 3^k or 3^{k-p} fractional factorial design, the *Central Composite Design* (CCD), *Box-Behnken Design* (BBD), and saturated/near-saturated *Hoke*, *Hybrid*, and *Small Composite Designs* (SCD) are robust to the linear effect assumption, these designs are not very run size efficient in terms of screening designs as they are built to accommodate curvature for all factors under consideration.

As a result, recent literature has proposed employing a single experimental design capable of preforming both factor screening and response surface exploration when conducting multiple experiments is unrealistic due to time, budget, or other constraints. Initial attempts to use designs capable of performing both factor screening and response surface exploration with a single design relied upon the designs projection capacity.

Cheng and Wu (2001), hereafter referred to as CW, introduced a two-stage analysis method where the key linkage between stages was the ability to project the initial larger factor space onto a smaller factor space capable of fitting a second-order model. Because a design can project onto many different combinations of factors, a projection-efficiency criterion was developed to compare orthogonal designs based upon (1) the number of eligible projected designs (designs which can fit a second-order model) and (2) the estimation efficiency for eligible projected designs determined by the ratio of each designs $D-$ and $G-$efficiences (Cheng and Wu, 2001).

CW studied three orthogonal array (OA) designs which demonstrated desirable projection properties. In contrast to 3^{k-p} designs which have defining contrast subgroups to describe the design structure, the designs studied by CW required computer search to classify the possible projected designs. Fortunately, while more complex, the overall projection properties are better and generally required less runs. When compared to CCDs, the designs studied exhibited good $D-$efficiences but poor $G-$efficiences as the number of projected factors, increases.

Improving on the designs of CW, Xu et al. (2004), hereafter referred to by XCW, proposed a combinatorial method for constructing new and efficient OA designs and a design selection approach based upon a *projection aberration criterion* (Xu and Wu, 2001) for factor screening and the projection-efficiency criteria (Cheng and Wu, 2001) for interaction detection. XCW's three-step approach involves: (1) screening out poor orthogonal arrays (OA) for factor screening using the generalized word-length pattern, (2) applying the projection aberration criterion to select a best design from step 1, and (3) determining the best level permutations of the design from step 2 to improve design projection eligibility and estimation efficiency under the second-order model.

Ye et al. (2007), hereafter referred to as YTL, also examined 3-level 18-run and 27-run orthogonal designs; however, in addition to considering the projection properties of designs, their design choices were based on both model estimation and model discrimination criteria.

While previous work focused primarily on the designs projection capacity, Edwards and Truong (2011) applied Jones and Nachtsheim (2011b) method for finding efficient designs, deemed MA designs, with minimal aliasing between main effects and two-factor interactions. Edwards and Truong (2011) compared the 27-run orthogonal arrays of XCW and YTL with MA designs in terms of D-efficiency of projection and, via a simulation study, the proportion of active factors declared significant (Power 1) and the proportion of simulations in which only the true active factors are declared significant (Power 2). Although ranked last in terms of D-efficiency, the MA designs showed superior performance with their ability to detect active factors (Edwards and Truong, 2011).

A common thread connecting all CW, XCW, YTL, and MA designs is the use of a linear and quadratic main-effects only analysis for factor screening. Unfortunately, if the *strong effect heredity* principle fails to hold important interactions can be missed leading to a misspecified response surface model. However, if the concern exists where a factor's significance is only present in interactions with other factors, the authors proposed either the Bayesian approaches of Box and Meyer (1993) or Chipman et al. (1997) to account for significant factors outside of main effects when the *strong effect heredity* principle fails to hold (Cheng and Wu, 2001). Unfortunately, these methods are not readily available to practitioners in statistical software packages and are computationally intensive procedures, thus likely making their use impractical (Edwards and Truong, 2011). Another area of concern for the CW, XCW, YTL, and MA designs is the projection of main and/or quadratic effects deemed significant during the first stage analysis does not always yield a second-order design.

Edwards and Mee (2011) introduced new spherical Fractional Box-Behnken designs (FBBD) aimed at overcoming the projection deficiencies and main/quadratic effect only analysis issues found in the CW/XCW/YTL/MA designs. The FBBD provide the ability to explore interactions during the screening stage and to fit second-order models via a backward elimination analysis strategy to each of the $(k-1)$-factor projections.

The FBBDs are developed by taking subsets of the two-level fractional factorial designs which compose a BBD (Edwards and Mee, 2011). The number of runs associated with the FBBD vary depending upon the number of factors involved. While FBBDs require more runs than CW/XCW/YTL/MA designs, their ease of construction and aliasing structure facilitate an analysis strategy which cannot be applied to the CW/XCW/YTL/MA designs. Additionally, as k increases, the FBBD designs require less runs than CCD/BBD.

Jones and Nachtsheim (2011a) introduced a class of three-level designs referred to as "definitive screening designs" where main effects are not biased by second-order effects and all quadratic effects are estimable. Consisting of $2k+1$ runs for k factors, these designs were constructing using the same Jones and Nachtsheim (2011b) method used by Edwards and Truong (2011).

Dougherty et al. (2013a) describes a computer generated $D-$optimality design augmentation technique which uses a $k-$factor Definitive Screening Design (DSD) as a baseline fixed design and augments the design with $k-1$ additional runs. The DSD+ focus on improving the robustness of the DSD to the assumptions of heredity and sparsity and significant second-order factor identification.

5.4 Case Study

Arnold Engineering and Development Center (AEDC) provided Hill et al. (2011) the legacy wind tunnel test data set for a 21% scale model of a system used to simulate a supersonic, expendable, low-altitude, anti-ship missile. The data set consisted of approximately 9000 design points within both the transonic and supersonic subspace test regions. Six design variables (Angle of Attack, Roll Angle, Elevator Deflection, Aileron Deflection, Rudder Deflection, and Mach Number) were used via an OFAT testing methodology to aid in the characterization of the overall aerodynamic performance of the missile in both test regions.

Hill et al. (2011) used the 9000+ design points and multiple linear regression to develop "ground truth" response surface models of the missile system for the two partitioned design regions. Equations 5.1 and 5.2 are the fitted regression models representing this "ground truth" for the transonic and supersonic design regions, respectively.

$$\begin{aligned} Y_{T(GT)} = \; & 0.7276645 - 0.008916X_1 + 0.0052574X_2 - 0.020997X_3 - 0.010612X_4 \\ & - 0.035216X_5 + 0.6167071X_6 + 0.0104301X_1X_2 + 0.0877043X_1X_3 \\ & - 0.011519X_1X_4 - 0.018356X_1X_5 + 0.0176079X_1X_6 - 0.017622X_2X_4 \\ & + 0.0199821X_3X_4 - 0.137442X_4X_5 - 0.007419X_5X_6 + 0.0036692X_1^2 \\ & + 0.1299117X_3^2 + 0.027947X_4^2 + 0.2434671X_5^2 - 0.05499X_6^2 \\ & - 0.370656X_6^3 \end{aligned} \quad (5.1)$$

$$
\begin{aligned}
Y_{S(GT)} \quad = \quad & -0.126954 + 0.0591609X_1 + 0.006896X_2 - 0.012877X_3 - 0.006569X_4 \\
& - 0.015172X_5 - 0.075103X_6 + 0.0065374X_1X_2 + 0.0561804X_1X_3 \\
& - 0.00621X_1X_4 - 0.012867X_1X_5 + 0.0200284X_1X_6 - 0.008101X_2X_4 \\
& + 0.0159264X_3X_4 - 0.01437X_3X_5 - 0.0037195X_3X_6 + 0.009108X_1^2 \\
& + 0.0779585X_3^2 + 0.0174788X_4^2 + 0.1448347X_5^2 - 0.004695X_6^2 \\
& - 0.037185X_1^3 \qquad (5.2)
\end{aligned}
$$

The partitioning of the data into transonic and supersonic design regions enabled the fitting of low-order polynomial models. As such, Hill et al. (2011) limited the models to full quadratic models and pure cubic terms.

Hill et al. (2011) proceeded to generate response surface models for both the transonic and supersonic design regions using approximately 900 experimental design points of two alternative designs (Nested Face-Centered Design (NFCD) and an I-optimal computer generated design) by sampling from the "ground truth" model with an $N(0, 0.0125)$ error added. They then compared the corresponding surfaces using a Monte Carlo sampling methodology coupled with a statistical comparison to determine the functional equivalency of the surfaces. Hill et al. (2011) demonstrated the ability to generate equivalent response surfaces at a 90% reduction in experimental effort.

Whereas Hill et al. (2011) were focused on the ability to generate equivalent response surfaces with a fraction of the the original experimental runs, we are interested in screening the nonlinear "ground truth" models for significant effects with a single 18-run six-factor augmented Definitive Screening Design (DSD+) experiment, see Table 56.

Table 56. Six-Factor Augmented Definitive Screening Design (DSD+)

X_1	X_2	X_3	X_4	X_5	X_6
0	1	1	1	1	1
0	-1	-1	-1	-1	-1
1	0	-1	1	1	-1
-1	0	1	-1	-1	1
1	-1	0	-1	1	1
-1	1	0	1	-1	-1
1	1	-1	0	-1	1
-1	-1	1	0	1	-1
1	1	1	-1	0	-1
-1	-1	-1	1	0	1
1	-1	1	1	-1	0
-1	1	-1	-1	1	0
0	0	0	0	0	0
1	-1	1	1	1	1
1	-1	-1	-1	-1	-1
-1	1	1	1	1	1
-1	1	1	-1	-1	-1
-1	-1	-1	1	-1	-1

Table 56 was generated by adding $k - 1 = 5$ runs to a $k = 6$-factor DSD via a computerized search algorithm. Instead of the information matrix being a main effects only model, the information matrix contains the main effects and the five two-factor interactions involving a particular factor, X_1 in this instance. The DSD+ design was constructed using a variant of the coordinate exchange algorithm of Meyer and Nachtsheim (1995) to maximize the determinant of the updated information matrix. Details on the computer search algorithm employed for the DSD and DSD+ can be found in Jones and Nachtsheim (2011a) and Dougherty et al. (2013a), respectively. To guard against local maxima, 10000 random starting designs, which included the baseline six-factor DSD, were explored. As a result, twelve equivalent designs were generated based upon both *D-optimal* and *I-efficient* criteria. For this case study the first of the designs (Table 56) is used; however, in practice other criteria could be used to further differentiate between the twelve designs.

5.5 Discussion

Simulated responses where generated by sampling from both "ground truth" models with an $N(0, 0.0125)$ error included for the design points specified by Table 56. Five sets of data were simulated for each design region (transonic and supersonic). Second-order empirical models are then constructed through forward stepwise regression. With a p-value of 0.1 to enter, effects are added into the second-order model while forcing a strong heredity model. As such, when either two-factor interactions or pure-quadratic effects are included in the model, the lower order terms must also be included.

Since the "ground truth" models contain all six design variables at various signal levels, screening design success is based upon the ability to determine active effects based upon the signal-to-noise ratio. For instance with a 0.0125 noise factor, factors X_2 and X_6 have 0.42 and 49.34 signal to noise ratios, respectively, within the transonic region. As such, failing to identify X_6 vice X_2 would be a more egregious error. Table 57 displays the average percentage (across five replications) of effects identified at four different signal-to-noise ratios.

Table 57. Second Order Screening Design Results for Case Study : 5 Replication Average

Scenario		**Transonic**				**Supersonic**			
	δ/ε	>0	>1	>2	>3	>0	>1	>2	>3
$\varepsilon \sim N(0, 0.0125)$	**ME**	90%	100%	100%	100%	90%	100%	100%	100%
	2FI	44%	57%	50%	50%	27%	36%	100%	100%
	- with X_1	36%	40%	0%	0%	36%	60%	100%	100%
	PQ	32%	40%	40%	47%	48%	67%	100%	100%
	Total	55%	62%	58%	57%	51%	65%	100%	100%

Note: Identified percentages correspond to percentage of Main Effects (ME), Two-Factor Interaction (2FI) effects, and Pure Quadratic (PQ) effects.

For a signal-to-noise (δ/ε) ratio ≥ 0, the screening design is trying to identify 20 effects as being significant within both Equations 5.1 and 5.2, excluding only the cubic term. In contrast, for a signal-to-noise (δ/ε) ratio ≥ 2, the screening design is trying to identify 8 and 5 effects within both Equations 5.1 and 5.2, respectively. Specifically for Equation 5.2, the screening design is looking to identify X_1, X_6, X_1X_3, X_3^2 and X_5^2.

Overall, with only 18 design runs, the DSD+ was able to identify over half of the effects for both design regions. As the signal-to-noise ratio increases, so does the percentage of effects identified, particularly with regards to the supersonic region. Thereby signifying the design is identifying the larger more significant terms for the response surface. Most importantly the design is nearly perfect, even at lower signal-to-noise ratios, at identifying main effect (ME) factors.

The DSD+ struggled at identifying active second-order effects, regardless of the signal-to-noise ratio, within the transonic design region and at the lowest signal-to-noise ratio for the supersonic region. This performance is not surprising giving the large number of "active" effects at smaller signal-to-noise ratios and the level of confounding between pure quadratic and two-factor effects. Screening designs stem from the Pareto principle which states that most of the variability in a system or process output is due to a small number of inputs. This is not the case for the transonic region where 20 out of 28 effects for a full second-order empirical model using 6 factors are deemed active or significant.

Additionally, the DSD+ analysis methodology does not allow for the inclusion or estimation of cubic terms. While both the transonic and supersonic "ground truth" models contain only a single cubic term, the cubic term in the transonic model is ten times larger and far more significant than the cubic term in the supersonic model. As such, the impact of excluding cubic terms in the empirical model causes biasing in the remaining model terms because the cubic terms effect will be attributed to either model error or other effects, even potentially insignificant effects.

5.6 Conclusions

In budget limitations, testers need to carefully control run size. Hill et al. (2011) succeeded in generating equivalent response surface models for both the transonic and supersonic design regions at a 90% reduction in experimental effort when compared to traditional wind tunnel testing methodology. But neither Hill et al. (2011) nor the original wind tunnel testers were restricted in the number of available experimental runs. While screening designs are never perfect, they offer a mechanism to determine those factors likely most active in defining system response when resources are restricted. At a 99.8% and 98% reduction in experimental effort when compared to traditional wind tunnel testing and Hill et al. (2011), respectively, the DSD+ was able to identify the majority of the significant second-order effects and all but the smallest main effect, particularly for the supersonic test region.

Systems and processes continue to become more complex, as a result the number of factors being considered on the system or process response grows. In a time when resources are shrinking, the increase in factors of interest requires additional experimental runs if more efficient design methodologies are not employed. This insight leads to customization of the full design in order to yield better run efficiencies; however, there is always a chance of misleading results. Such is the case with most any statistical analysis. Thus, there will always be the need for system-specific expertise as a complement to the system experimental data analysis. The analysis of screening designs may not be an easy task. Statistical proficiency and capable analytical packages will be required.

VI. Summary and Recommendations

6.1 Summary of Work

Screening designs are a category of experimental designs, usually performed during the early stages of a process or system study, used to determine which of the many factors (if any) have a significant effect on the system or process. Selecting which screening design to use from the multitude of available designs is not always straight forward. Assumptions related to the principles of sparsity and heredity and the form of the empirical model used to represent the process or system response should help determine selection of an appropriate screening design. Factors such as whether or not follow-up experiments are available and the ease of which statistical analysis of data can be done can also effect design selection. As a result, with the help of subject matter experts, research and development of specialized computer generated designs which exhibit desirable design parameters has increased.

Second-order screening designs are a relatively new focus in statistical research and largely unknown to the defense test community. Second-order screening designs are single experimental designs capable of preforming both factor screening and response surface estimation when conducting multiple experiments is unrealistic due to time, budget, or other constraints. This dissertation explored the robustness of leading designs, developed an augmentation strategy to improve one of the leading designs, and provided a case study application of such designs. Chapter II contains a detailed literature review of screening and response surface designs, partitioned by sequential and single phase methods for fitting first order and second order response surfaces. Chapters III, IV, and V are self-contained research articles on second-order screening designs. Each contains a literature review of the research relevant to that chapter.

Two important principles used in developing successful screening designs are sparsity and heredity. However, the degree to which factor sparsity holds as the number of factors grows has resulted in a debate between *effect sparsity* and factor sparsity. Heredity, either strong or weak, is the second screening principle commonly used when considering model selection. Strong heredity implies that if a model includes a two-factor interaction, then its constituent main effects are included in the model. Conversely, weak heredity requires only one of the two constituent main effects be included in the model. To date, evaluation of screening design performance has assumed both factor sparsity and strong effect heredity. Chapter III formally examines the robustness of the two arguably best second-order screening designs with respect to the assumptions of both sparsity (factor or effect) and heredity (strong or weak).

Whenever a screening design is employed, analytical tradeoffs must be accepted. Definitive Screening Designs are run size efficient when strong heredity and factor sparsity are present and when few second-order effects are active. Chapter IV describes a computer generated $D-$optimality design augmentation technique which uses a $k-$factor Definitive Screening Design (DSD) as a baseline fixed design and augments the design with $k-1$ additional runs. In a simulation study, the proposed augmented Definitive Screening Design (DSD+) was able to increase the robustness of the original DSD to the principles of heredity and sparsity while also increasing the detection rate of two-order effects when both two-factor interactions and pure-quadratic effects are active.

Chapter V presents the use of design of experiments and response surface designs in the area of second-order screening designs, particularly as applied to defense testing, through demonstrating the viable use of second-order screening designs in a wind tunnel case study.

6.2 Recommendations for Future Research

This work focused on the robustness and augmentation of existing second-order screening designs. Evaluation of design performance was based upon the original design authors' recommended analysis methodology. It would be interesting to study alternative analysis methodology to see if the designs ability to identify active effects can be attributed to design structure or analysis methodologies.

Supersaturated designs can be used in large screening experiments when the number of factors exceeds the number of available run. While research on improving supersaturated designs construction and analysis continues, unfortunately, very little work is being done on constructing supersaturated designs which are capable of response surface exploration. It would be interesting to study construction of designs which are supersaturated in terms of total number of effects vice number of factors.

Bibliography

Angelopoulos, P., H. Evangelaras, and C. Koukouvinos (2009). Small, balanced, efficient and near rotatable central composite designs. *Journal of Statistical Planning and Inference 139*(6), 2010–2013.

Beattie, S., D. Fong, and D. Lin (2002). A two-stage Bayesian model selection strategy for supersaturated designs. *Technometrics 44*(1), 55–63.

Booth, K. and D. Cox (1962). Some systematic supersaturated designs. *Technometrics 4*(4), 489–495.

Box, G. and D. Behnken (1960). Some new three level designs for the study of quantitative variables. *Technometrics 2*(4), 455–475.

Box, G. and P. Liu (1999). Statistics as a catalyst to learning by scientific method part I-an example. *Journal of Quality Technology 31*(1), 1–15.

Box, G. and R. Meyer (1993). Finding the active factors in fractionated screening experiments. *Journal of Quality Technology 25*(2), 94–94.

Box, G. and K. Wilson (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society. Series B (Methodological) 13*(1), 1–45.

Box, M. and N. Draper (1974). On minimum-point second-order designs. *Technometrics 16*(4), 613–616.

Candes, E. and T. Tao (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics 35*(6), 2313–2351.

Cheng, S. and C. Wu (2001). Factor screening and response surface exploration. *Statistica Sinica 11*(3), 553–579.

Chipman, H., M. Hamada, and C. Wu (1997). A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics 39*(4), 372–381.

Connor, W. and M. Zelen (1959). *Fractional factorial experiment designs for factors at three levels*, Volume 54. US Government Printing Office Washington, DC.

Dougherty, S., R. Hill, J. Pignatiello, J. Simpson, and E. White (2013a). Augmentation of definitive screening designs (DSD+).

Dougherty, S., R. Hill, J. Pignatiello, J. Simpson, and E. White (2013b). An Empirical study of the effect of heredity and sparsity on second-order screening design performance.

Draper, N. (1985). Small composite designs. *Technometrics 27*(2), 173–180.

Draper, N. and D. Lin (1990). Small response-surface designs. *Technometrics 32*(2), 187–194.

Dubova, I. and V. Federov (1972). Tables of optimum designs II (saturated D-optimal designs on a cube). *Preprint* (40).

DuMouchel, W. and B. Jones (1994). A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model. *Technometrics 36*(1), 37–47.

Edwards, D. and R. Mee (2011). Fractional BoxBehnken designs for one-step response surface methodology. *Journal of Quality Technology 43*(4), 288–306.

Edwards, D. and D. Truong (2011). A comparison of designs for one-step screening and response surface estimation. *Quality and Reliability Engineering International 27*(8), 1009–1024.

Fang, K., D. Lin, and C. Ma (2000). On the construction of multi-level supersaturated designs. *Journal of Statistical Planning and Inference 86*(1), 239–252.

Fries, A. and W. Hunter (1980). Minimum aberration 2^{k-p} designs. *Technometrics 22*(4), 601–608.

Gilmore, J. (2010). Guidance on the use of design of experiments (DOE) in operational test and evaluation.

Gilmore, J. (2013). Best practices for assessing the statistical adequacy of experimental designs used in operational test and evaluation.

Gilmour, S. (2006). Response surface designs for experiments in bioprocessing. *Biometrics 62*(2), 323–331.

Guo, Y., J. Simpson, and J. Pignatiello (2009). Deciphering all those minimum aberration criteria for experimental designs. *Quality Engineering 21*(4), 432–445.

Hamada, M. and C. Wu (1992). Analysis of designed experiments with complex aliasing. *Journal of Quality Technology 24*(3), 130–137.

Hartley, H. (1959). Smallest composite designs for quadratic response surfaces. *Biometrics 15*(4), 611–624.

Hill, R., D. Leggio, S. Capehart, and A. Roesener (2011). Examining improved experimental designs for wind tunnel testing using monte carlo sampling methods. *Quality and Reliability Engineering International 27*(6), 795–803.

Hill, W. and W. Hunter (1966). A review of response surface methodology: a literature survey. *Technometrics 8*(4), 571–590.

Hoke, A. (1974). Economical second-order designs based on irregular fractions of the 3^n factorial. *Technometrics 16*(3), 375–384.

Jones, B., W. Li, C. Nachtsheim, and K. Ye (2007). Model discriminationanother perspective on model-robust designs. *Journal of Statistical Planning and Inference 137*(5), 1576–1583.

Jones, B., D. Lin, and C. Nachtsheim (2008). Bayesian D-optimal supersaturated designs. *Journal of Statistical Planning and Inference 138*(1), 86–92.

Jones, B. and C. Nachtsheim (2011a). A class of three-level designs for definitive screening in the presence of second-order effects. *Journal of Quality Technology 43*(1), 1–15.

Jones, B. and C. Nachtsheim (2011b). Efficient designs with minimal aliasing. *Technometrics 53*(1), 62–71.

Khuri, A. and J. Cornell (1996). *Response surfaces: designs and analyses*, Volume 152. CRC press.

Lawson, J. (2003). One-step screening and process optimization experiments. *The American Statistician 57*(1), 15–20.

Li, W. and C. Wu (1997). Columnwise-pairwise algorithms with applications to the construction of supersaturated designs. *Technometrics 39*(2), 171–179.

Lin, D. (1993). A new class of supersaturated designs. *Technometrics 35*(1), 28–31.

Lin, D. (1995). Generating systematic supersaturated designs. *Technometrics 37*(2), 213–225.

Lin, D. and N. Draper (1992). Projection properties of Plackett and Burman designs. *Technometrics 34*(4), 423–428.

Marley, C. and D. Woods (2010). A comparison of design and model selection methods for supersaturated experiments. *Computational Statistics & Data Analysis 54*(12), 3158–3167.

Matsuura, S., H. Suzuki, T. Iida, H. Kure, and H. Mori (2011). Robust parameter design using a supersaturated design for a response surface model. *Quality and Reliability Engineering International 27*(4), 541–554.

Mead, R. and D. Pike (1975). A biometrics invited paper. a review of response surface methodology from a biometric viewpoint. *Biometrics 31*(4), 803–851.

Mee, R. (2000). New Box−Behnken designs. *University of Tennessee, Department of Statistics Technical Report 4*.

Meyer, R. and C. Nachtsheim (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics 37*(1), 60–69.

Montgomery, D. (2013). *Design and analysis of experiments*, Volume 8. Wiley New York.

Morris, M. (2000). A class of three-level experimental designs for response surface modeling. *Technometrics 42*(2), 111–121.

Myers, R. and C. Anderson-Cook (2009). *Response surface methodology: process and product optimization using designed experiments*, Volume 705. Wiley. com.

Myers, R., A. Khuri, and W. Carter (1989). Response surface methodology: 1966–l988. *Technometrics 31*(2), 137–157.

Myers, R., D. Montgomery, G. Vining, C. Borror, and S. Kowalski (2004). Response surface methodology: a retrospective and literature survey. *Journal of quality technology 36*(1), 53–77.

Nguyen, N. (1996). An algorithmic approach to constructing supersaturated designs. *Technometrics 38*(1), 69–73.

Nguyen, N. and J. Borkowski (2008). New 3-level response surface designs constructed from incomplete block designs. *Journal of Statistical Planning and Inference 138*(1), 294–305.

Notz, W. (1982). Minimal point second-order designs. *Journal of Statistical Planning and Inference 6*(1), 47–58.

Oehlert, G. and P. Whitcomb (2002). Small, efficient, equireplicated resolution V fractions of 2^k designs and their application to central composite designs. In *Proceedings of 46th Fall Technical Conference*.

Phoa, F., Y. Pan, and H. Xu (2009). Analysis of supersaturated designs via the Dantzig selector. *Journal of Statistical Planning and Inference 139*(7), 2362–2372.

Plackett, R. and J. Burman (1946). The design of optimum multifactorial experiments. *Biometrika 33*(4), 305–325.

Rechtschaffner, R. (1967). Saturated fractions of 2^n and 3^n factorial designs. *Technometrics 9*(4), 569–575.

Roquemore, K. (1976). Hybrid designs for quadratic response surfaces. *Technometrics 18*(4), 419–423.

Satterthwaite, F. (1959). Random balance experimentation. *Technometrics 1*(2), 111–137.

Truong, D. (2010). *Single-Step Factor Screening and Response Surface Optimization Using Optimal Designs with Minimal Aliasing*. Ph. D. thesis, Virginia Commonwealth University.

Wang, J. and C. Wu (1995). A hidden projection property of Plackett-Burman and related designs. *Statistica Sinica 5*(1), 235–250.

Warner, C. (2011). Continuing the emphasis on scientific rigor in test and evaluation. *ITEA Journal 32*(1), 15–17.

Westlake, W. (1965). Composite designs based on irregular fractions of factorials. *Biometrics 21*(2), 324–336.

Wu, C. J. and M. S. Hamada (2011). *Experiments: planning, analysis, and optimization*, Volume 552. John Wiley & Sons.

Xiao, L., D. Lin, and F. Bai (2012). Constructing definitive screening designs using conference matrices. *Journal of Quality Technology 44*(1), 2–8.

Xu, H. (2005). A catalogue of three-level regular fractional factorial designs. *Metrika 62*(2-3), 259–281.

Xu, H., S. Cheng, and C. Wu (2004). Optimal projective three-level designs for factor screening and interaction detection. *Technometrics 46*(3), 280–292.

Xu, H. and C. Wu (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *Annals of statistics 29*(2), 549–560.

Yamada, S., Y. Ikebe, H. Hashiguchi, and N. Niki (1999). Construction of three-level supersaturated design. *Journal of Statistical Planning and Inference 81*(1), 183–193.

Yamada, S. and D. Lin (1999). Three-level supersaturated designs. *Statistics & Probability Letters 45*(1), 31–39.

Yamada, S., M. Matsui, T. Matsui, D. Lin, and T. Takahashi (2006). A general construction method for mixed-level supersaturated design. *Computational Statistics & Data Analysis 50*(1), 254–265.

Ye, K., K. Tsai, and W. Li (2007). Optimal orthogonal three-level factorial designs for factor screening and response surface exploration. In *mODa 8-Advances in Model-Oriented Design and Analysis*, pp. 221–228. Springer.

Zhang, T., J. Yang, and D. Lin (2011). Small Box-Behnken design. *Statistics & Probability Letters 81*(8), 1027–1033.

Vita

Lt Col Shane A. Dougherty is a 1990 graduate from Highlands High School in Fort Thomas, Kentucky. He attended the United States Air Force Academy in Colorado Springs, Colorado and was a Distinguished Graduate in 1994 with a Bachelor of Science degree in Aernautical Engineering. He attended the Air Force Institute of Technology on Wright-Patterson Air Force Base in Dayton, Ohio and was a Distinguished Graduate in 2002 with a Masters of Science degree in Operations Research. In the years following, he worked as the F-22 Operations Research Analyst for the 28th Test Squadron Detachment 1, Nellis AFB, NV and Chief, Scientist Assignments Officer at the Air Force Personnel Center, Randolph AFB, TX. He attended the College of Naval Command and Staff in Newport, Rhode Island and graduated in 2007 with a Masters of Arts Degree in National Security and Strategic Studies after which he served as the Chief, Operations Analyst at NATO Component Command - Air HQ, Ramstein AB, Germany.

Lt Col Dougherty began his doctoral studies at AFIT in 2010. He conducted his dissertation research in the area of experimental design under the guidance of Dr. Raymond R. Hill. Upon graduation, he will transfer to the Headquarters Air Force Force Management Division (AF/A1PF) where he will act as Chief, Special Programs Branch.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
12–12–2013	PhD Dissertation	Aug 2010 – Dec 2013

4. TITLE AND SUBTITLE

A Comparison Study of Second-Order Screening Designs and Their Extension

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dougherty, Shane A., LtCol, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-DS-13-D-01

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Dr. Catherine Warner
Office of the Secretary of Defense
1700 Defense Pentagon
Washington, D.C. 20301
(703) 697-7247, catherine.warner@osd.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

OSD

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A: APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT Recent literature has proposed employing a single experimental design capable of preforming both factor screening and response surface estimation when conducting sequential experiments is unrealistic due to time, budget, or other constraints. Military systems, particularly aerodynamic systems, are complex. It is not unusual for these systems to exhibit nonlinear response behavior. Developmental testing may be tasked to characterize the nonlinear behavior of such systems while being restricted in how much testing can be accomplished. Second-order screening designs provide a means in a single design experiment to effectively focus test resources onto those factors driving system performance. Sponsored by the Office of the Secretary of Defense (ODS) in support of the Science of Test initiative, this research characterizes and adds to the area of second-order screening designs, particularly as applied to defense testing. Existing design methods are empirically tested and examined for robustness. The leading design method, a method that is very run efficient, is extended to overcome limitations when screening for non-linear effects. A case study and screening design guidance for defense testers is also provided.

15. SUBJECT TERMS

Design of Experiments, Optimal Designs, Definitive Screening

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Raymond R. Hill
					19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	139	(937) 255–3636 ext 7469; raymond.hill@afit.edu