

SPACE-FILLING DESIGNS FOR MIXTURE/PROCESS VARIABLE
EXPERIMENTS

by

Moses Yeboah Obiri

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Statistics

MONTANA STATE UNIVERSITY
Bozeman, Montana

December 2020

©COPYRIGHT

by

Moses Yeboah Obiri

2021

All Rights Reserved

DEDICATION

This dissertation is dedicated to my brother and his wife, Mr. and Mrs. K. Wiafe.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Prof. John Borkowski, for the invaluable guidance, assistance and advice he provided me throughout my time as a student in Montana State University. I would not be here without you, and I am incredibly grateful and lucky to have an advisor who cared so much about my work and welfare.

Special thanks to the former and current members of my committee: Dr. Nicole Carnegie, Prof. Mark Greenwood, Dr. Stacey Hancock, Dr. Megan Higgs, Dr. Laura Hildreth, and Dr. Andrew Hoegh. I sincerely appreciate your help, and the content of this dissertation has been enhanced by your helpful comments and suggestions.

I am also indebted to the staff and students in the Department of Mathematical Sciences, MT Peaks Program, and the Statistical Consulting and Research Services for their help and support. My education at Montana State University was made possible through their support.

Finally, I would like to thank my parents, siblings, and all those who have contributed greatly to my academic success.

TABLE OF CONTENTS

1. INTRODUCTION	1
1.1 Classical Designs	2
1.1.1 Full and Fractional Factorial Designs	4
1.1.2 Response Surface Designs	5
1.1.2.1 Designs for Fitting First-Order Models	6
1.1.2.2 Orthogonal Designs for First-Order Linear Models with Interactions	9
1.1.2.3 Designs for Fitting Second-Order Models	10
1.2 Response Surface Designs for Mixture Experiments	11
1.2.1 Simplex-Lattice Designs	15
1.2.2 Simplex Centroids Designs	18
1.2.3 Simplex Axial Designs	18
1.2.4 Minimum and Maximum Level Component Constraints	19
1.2.5 Mixture Experiments with Process Variables	24
1.3 Uniform Designs	26
1.3.1 Advantages of Uniform Designs	28
1.3.2 Methods of Generating Uniform Design Points	32
1.3.2.1 Good Lattice Point Sets (Fang and Wang, 1994)	32
1.3.2.2 Good Point Method (Fang and Wang, 1994)	35
1.3.2.3 Halton (H) Set Method (Fang and Wang, 1994)	38
1.3.2.4 Hammersley Method (Fang and Wang, 1994)	40
1.3.3 Measures of Uniformity	41
1.3.4 Objectives of the Study	47
2. UNIFORM DESIGNS FOR EXPERIMENTS WITH MIXTURES	50
2.1 Uniform Designs for Constrained and Unconstrained Mixture Experiments	51
2.2 New Proposed Uniform Designs for Mixture/Process Variable Experiments	55
2.3 New Measures of Uniformity for Mixture/Process Experiments	59
2.3.1 The Euclidean Minimum Spanning Tree (EMST)	60
2.3.2 New Uniformity Criteria based on EMST for Uni- form Mixture Designs in the Simplex	63
2.4 A New Conditional Approach For Finding Good Uniform Mixture/Process Designs Using EMST	71
2.4.1 Mixture/Process Design Example	75

TABLE OF CONTENTS – CONTINUED

3. EXAMPLE OF MIXTURE/PROCESS VARIABLE DESIGN AND DATA ANALYSIS	82
3.1 The Mixture/Process Variable Model.....	82
3.1.1 Mixture/Process Variable Example 1	84
3.1.2 Mixture/Process Variable Example 2.....	89
4. SPACE-FILLING DESIGN AUGMENTATION IN THE SIMPLEX.....	102
4.1 New Space-filling Design Augmentation Approach	104
4.1.1 Augmenting Boundary Mixture Design Points.....	104
4.1.2 Augmenting Boundary and Interior Mixture Design Points	108
4.1.3 Augmenting Mixture Designs with Existing Cluster of Points	109
5. CONCLUSIONS AND FUTURE RESEARCH	116
BIBLIOGRAPHY	121
APPENDIX: APPENDIX A.....	131

LIST OF TABLES

Table	Page
1.1 Number of Terms in Scheffe Mixture Models.....	15
1.2 Number of Points in $\{q, m\}$ Simplex-lattice Design.....	18
1.3 Example of discrepancy from uniform $(0, 2)$ distribution,	42
1.4 Star discrepancy ($D^*(n, \rho)$) and maximum distance (md) criteria for random sampling, square root sequence, and factorial designs in c^2	47
2.1 Steps for generating a 9-point design for a 3-component mixture experiment with 1 process variable using the new hybrid algorithm	58
2.2 9 mixture/process variable design points in grams and degrees Celsius.	58
2.3 9-point design for a 3-component mixture with temperature as a categorical process variable.	59
2.4 Demonstrating sensitivity property of the geometric mean to increasing variation.	67
2.5 New EMST-based measures of uniformity in differ- ent simplex regions	69
2.6 Relationship between $GMST$ and $adMST$, $sdMST$	70
2.7 34-point uniform design for 3-component mixture and 2-process variable experiment.	81
3.1 33-point D-optimal designs for L/I and Q/I models.	94
3.2 Summary table for fitted models for Example 1.	95
3.3 33-point Uniform design.	96
3.4 Summary of SC/I model using uniform design points.	97
3.5 34-point D-optimal design for Example 2.....	98
3.6 Summary of fitted models for Example 2.	99

LIST OF TABLES – CONTINUED

Table	Page
3.7 Simulated data using coefficients from quadratic mixture model and interaction process variable model.	100
3.8 Simulated data assuming special cubic mixture model and interaction process variable model.	101
A.1 Best design generators in S^3	132
A.2 Best design generators in S^4	133
A.3 Best design generators in S^5	134
A.4 Best design generators in S^3 and $K = 1$	135
A.5 Best design generators in S^3 and $K = 2$	136
A.6 Best design generators in S^4 and $K = 1$	137
A.7 Best design generators in S^4 and $K = 2$	138

LIST OF FIGURES

Figure	Page
1.1 Illustration of a process of an industrial experiment (Box and Draper, 2007).....	2
1.2 Simplex factor spaces for $q = 2$ and $q = 3$ components (Box and Draper, 2007)	16
1.3 $\{3, 3\}$ and $\{3, 2\}$ simplex-lattice designs.....	17
1.4 3-component simplex centroid and simplex axial designs.....	19
1.5 Example of a restricted design region from placing lower bounds on a 3-component mixture, $x_1 \geq 0.2$, $x_2 \geq 0.15$,and $x_3 \geq 0.12$	20
1.6 Example of a restricted design region from placing lower and upper bounds on a 3-component mixture, $0.1 \leq x_1 \leq 0.75$, $0.15 \leq x_2 \leq 0.8$, $0.1 \leq x_3 \leq 0.6$	21
1.7 A 9-point extreme vertices design for the experimen- tal region in Figure 1.6.	24
1.8 Example of a combined mixture and process variable design	25
1.9 21-point designs using uniform design and Monte Carlo approaches.	29
1.10 8-point uniform designs in C^2 using GLP method.	35
1.11 8-point design using CF method.....	36
1.12 8-point designs using SRS method.....	37
1.13 8-point designs using PP method.	38
1.14 8-point design using H-set method	39
1.15 8-point design using Hammersley method	41
1.16 Geometric illustration of how to calculate discrepancy.	44
1.17 From left to right: Random sampling, SRS, and Factor designs with $n = 36$ in C^2	48

LIST OF FIGURES – CONTINUED

Figure	Page
2.1 21-point designs in C^2 using four number theoretic methods.	53
2.2 Example of a restricted design region from placing lower and upper bounds on a 3-component mixture.	54
2.3 Illustration of the G function for $n = 21$ design points in the full simplex and the constrained simplex.	56
2.4 Spanning Trees.	61
2.5 Weighted Graph.	62
2.6 Cheapest edge out of every vertex highlighted with yellow line.	63
2.7 Example of a Minimum Spanning Tree.	63
2.8 Characterization of points in (μ, σ) plane (Franco et al., 2009).	64
2.9 11-point designs for 3-component mixture experi- ments with corresponding EMSTs and their $GMST$ values.	68
2.10 Effect plots of $GMST$ versus $adMST$ and $sdMST$ in the full simplex.	70
2.11 Effect plot of $GMST$ regressed against $adMST$ and $sdMST$ in the constrained simplex.	71
2.12 Mixture/process variable designs.	72
2.13 $GMST$ in the mixture space and in the combined space.	74
2.14 Good and bad uniform mixture/process variable design.	75
2.15 Best 15-point mixture and categorical process vari- able designs.	77
2.16 Conditional $GMST$ measure for 15-point mixture and categorical process variable designs.	78
2.17 The best to worst 29-point designs in the constrained simplex using the GLP method.	79

LIST OF FIGURES – CONTINUED

Figure	Page
2.18 The 10 best 3-component mixture and 1-process variable GLP designs.....	79
2.19 The best 3-component mixture and 2-process variable GLP designs.	80
3.1 LI-based D-optimal design for Example 1.	85
3.2 QI-based D-optimal design for Example 1.....	86
3.3 Uniform design design for Example 1.	87
3.4 Contour plots of predicted responses from SC/I model for Example 1 using space-filling mixture process variable design.	88
3.5 Constrained mixture design space for Example 2.....	89
3.6 Mixture/Process variable D-optimal design for Example 2.	90
3.7 Mixture/Process variable uniform design for Example 2.....	91
4.1 New design region (green) defined from existing design region (blue).....	103
4.2 Constrained design region for 3-component mixture experiment.	105
4.3 Results from adding 9 interior points to a $\{3, 2\}$ -simplex lattice design.	106
4.4 Results from adding 9 interior points to a 7-point extreme vertices design.....	107
4.5 Best augmented designs from adding 9 points to $\{3, 2\}$ -simplex lattice design.....	111
4.6 Best augmented designs from adding 9 points to 7-point extreme vertices design.	112
4.8 Best augmented designs from adding 9 points to a 6-point design and 8-point extreme vertices design.	114

LIST OF FIGURES – CONTINUED

Figure	Page
4.9 9-point mixture design for 3 component experiment with 3 clusters.	115
4.10 Augmenting 9-point design with 9 new runs.....	115
4.11 Augmented 18-point design consisting of 9 existing points and 9 new runs.	115

ABSTRACT

The ultimate objective of this dissertation was to present a statistical methodology and an algorithm for generating uniform designs for the combined mixture/process variable experiment. There are many methods available for constructing uniform designs and four of such methods have been used in this study. These are the Good Lattice Point (GLP) method, the cyclotomic field (CF) method, the square root sequence (SRS) method, and the power-of-a-prime (PP) method.

A new hybrid algorithm is presented for generating uniform designs for mixture/process variable experiments. The algorithm uses the G function introduced by Fang and Yang (2000), and adopted by Borkowski and Piepel (2009) to map $q - 1$ points from $q + k - 1$ points generated in the hypercube to the simplex.

Two new criteria based on the Euclidean Minimum Spanning Tree (EMST) which are more computationally efficient for assessing uniformity of mixture designs and mixture/process variable designs are presented. The two criteria were found to be interchangeable and the geometric mean of the edge lengths ($GMST$) criterion is preferred to the average and standard deviation of edge lengths ($adMST$, $sdMST$) criterion. The $GMST$ criterion uses only one statistic to quantify the uniformity properties of mixture and mixture/process variable designs. Tables of good uniform designs are provided for mixture experiments in the full simplex (S^q) for $q = 3, 4, 5$ and practical design sizes, $9 \leq n \leq 30$, using the four number theoretic methods in this study. A conditional approach based on the $GMST$ criterion for generating good uniform mixture/process variable designs is also introduced and tables of good uniform designs are given for the combined q -mixture and k process variable experiments for $q = 3, 4, 5$, $k = 1, 2$ and practical numbers of runs, $9 \leq n \leq 30$.

A new algorithm is provided to augment existing mixture design points with space-filling points including designs with existing clustered design points. In this algorithm, new design points are chosen from a candidate set of points such that the resulting augmented design has good space-filling properties. The SRS method is found to produce the best augmented space-filling mixture designs.

CHAPTER ONE

INTRODUCTION

Experimental design is a statistical methodology used to investigate various experimental inputs whose variation may impact one or more outputs or responses of interest (Myers et al., 2016). The experimental inputs are often referred to as independent variables or factors of the experiment, and are those that are assumed to potentially influence or affect the values of the experimental responses (also referred to as dependent variables). In all fields of inquiry, it is often necessary to change input variables in a system and to observe the changes in system output caused by changes in inputs. This process is called an experiment. An experiment can therefore be defined as a series of runs or trials in which we may observe changes in response outputs of a process or system as a result of changes made to the input variables of that process (Montgomery, 2017). An experimental run or trial is a set of experimental conditions to be used for collecting an experimental observation. The objects subjected to specific experimental conditions or experimental treatments are called experimental units.

The general process of performing an industrial experiment is illustrated in Figure 1.1. An industrial process usually involves a combination of machines, production, methods, operations, people and other resources that transforms inputs into one or more response outputs (Park, 2007). The process variables consists of controllable $x_1, x_2, \dots, x_p$ variables and uncontrollable $z_1, z_2, \dots, z_p$ variables that may have effect on the response output. A controllable variable is an input factor of interest which is manipulated in an experiment and whose effect on the response

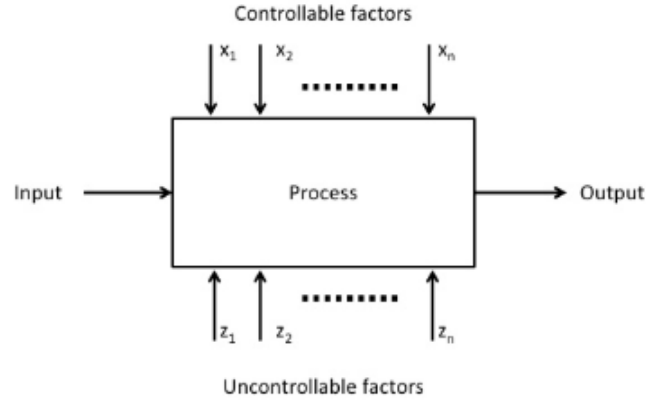


Figure 1.1: Illustration of a process of an industrial experiment (Box and Draper, 2007).

output is to be studied (Fang et al., 2006). An uncontrollable variable is a variable which is not of interest in an experiment but that has the potential to impact the relationship between the controllable variables or input factors and the response output. Thus, when a process is running, the uncontrollable variables vary and are not controlled. However, for experimental purposes, they will be controlled. Uncontrollable variables are sometimes called noise variables in the process.

1.1 Classical Designs

The development of the classic experimental design begins with deciding on the objectives of that experiment. For process optimization, key objectives are determining the most influential input variables on the response output and determining where to set the influential input variables so that the response output is maximized, minimized, or achieves a target value. Other objectives may include ascertaining where to set the influential inputs so that the variability in the response output is minimized or minimizing the effects of the uncontrollable process variables (Montgomery, 2017).

The basic principles of a classical experimental design are randomization, replication, and local control. Randomization is the use of some mechanism to randomly assign experimental units to treatments to create treatment groups that are similar before treatments are applied. The random assignment process implies that every possible allotment of units to treatments has the same probability and, when applicable, randomization also implies that the order of experimental runs is randomly determined. This helps prevent the introduction of systemic bias by evening out effects of extraneous factors, and provides the link between the actual experiment and the statistical model that underlies the data analysis (Winer, 1962).

The second principle is replication. Replication is independently repeating runs for one or more factor combinations in the experiment (Montgomery, 2017). Ideally, the experimenter would like to have multiple replications for every factor combination in the experiment. For example, if the objective of the experiment is to compare the yield of two varieties of rice, then replication implies multiple applications of each variety to experimental units. Replication is important because it allows for the estimation of experimental error. Experimental error is the difference between the true value of an output and its estimate (Dean et al., 2017). There are two kinds of experimental errors: errors that are inherent in the way an experiment is conducted and apply to all measurements equally (systematic error) and unforeseen variations in measurement that cannot always be identified (random errors). Systematic errors may be identified and corrected. For example, a faulty or a poorly calibrated weighing machine may produce results that are consistently 2 lbs higher or lower than actual weights. The machine can be re-calibrated to eliminate the errors if identified. Sources of random errors cannot always be identified. For example, results from a measuring instrument may be affected by changes in the surroundings. Knowledge of the degree of error from experimental results is essential if the results are to

be used appropriately. The last principle, local control, is an attempt to control extraneous sources of variation which are not removed by randomization. It typically refers to the use of blocking and balancing of experimental units. Blocking refers to the collection of similar experimental units together to form relatively homogeneous groups. Balancing refers to the balanced assignment of treatments to experimental units. The objective of local control is to improve precision or increase the efficiency of experimental design by decreasing the experimental error.

1.1.1 Full and Fractional Factorial Designs

Full factorial designs, fractional factorial designs, and response surface designs (such as optimal designs, orthogonal arrays, and central composite designs) are all classical statistical experimental designs that have been applied widely and extensively in industrial applications (Talke, 2012). A common approach when dealing with several experimental inputs or factors is to conduct a full or fractional factorial design. A full factorial design is an experimental design in which factors with discrete values (called levels) are varied, and all possible combinations of levels are studied (Kirk, 2012). It is denoted as an $\prod_{i=1}^k n_i$ design. That is, it is a design with k factors where the i^{th} factor has n_i levels and whose experimental units take on all $n_1 \times n_2 \times \dots \times n_k$ possible combinations across all k factors. For example, a 2^2 full factorial design is a design with two factors, each at two levels. A full factorial design allows the effect of a single factor to be investigated and to determine whether factor interactions exist. The main disadvantage of a full factorial design is the difficulty in conducting the experiment when there are three or more factors. The operational challenge of the three principles (randomization, replication, and local control) discussed earlier becomes too difficult when there are multiple factors for this type of design. For example, an unreplicated experiment with twelve factors at

two levels requires $2^{12} = 4096$ experimental runs. In such cases, fractional factorial designs may be used. A fractional factorial design is formed by taking a subset of experimental runs from a full factorial design. For example, a one-half fractional factorial design of a four-factor full factorial design with two levels for each factor (2^4 full factorial design) may require only 8 experimental runs instead of the original 16. The downside of using a fractional factorial design is that not all possible interaction effects can be investigated. Certain factors and factor interaction effects will be confounded with one another, and their effects cannot be separated.

1.1.2 Response Surface Designs

Response surface designs are often used when the objective of the experiment is to optimize some response or experimental output (Box and Draper, 2007). For example, suppose the goal is to investigate levels of some controllable factors, $x_1, x_2, \dots, x_k$, that maximize the yield of some process, and the process yield y is modeled as a function of levels of the factors

$$y = f(x_1, x_2, \dots, x_k) + \epsilon, \quad (1.1)$$

where f is the true unknown response function and ϵ represents other sources of variability in y unaccounted for in f . Often the mean of ϵ is assumed to be zero. Thus,

$$E(y) = f(x_1, x_2, \dots, x_k). \quad (1.2)$$

The first step in selecting a response surface design is to propose a suitable approximating model for the true unknown functional relationship between the response and the set of independent variables (Myers et al., 2016). In most response surface methods, using the first terms in a Taylor series expansion provides an

approximating function. A first-order approximating model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon \quad (1.3)$$

and a second-order approximating model if curvature is present in the system is

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i < j} \beta_{ij} x_i x_j + \epsilon. \quad (1.4)$$

1.1.2.1 Designs for Fitting First-Order Models If the goal is to fit a first-order model where the model parameters are estimated by the method of least squares, then orthogonality is a key concept in choosing a design that minimizes the variance of the regression coefficients (Box and Draper, 1987). Suppose a model matrix is denoted by X . Then a first-order orthogonal design is the one for which $X'X$ is a diagonal matrix (Myers et al., 2016). If there are p terms in a first-order model then for an experiment with n experimental runs, the model matrix X is an $n \times p$ matrix whose corresponding columns are the p terms in the first-order model (Borkowski, 2018). Orthogonality ensures that levels of two corresponding variables are linearly independent and the roles of the two variables can be assessed independently (Box and Draper, 1987).

Two classes of orthogonal first-order design are the 2^k full factorial designs and fractions of the 2^k designs. A 2^k factorial design is a design with k design variables x_j , each with two levels often referred to as low and high levels. The low and high levels of the k design variables are then coded to ± 1 so that -1 is the low level and 1 is the high level. The design matrix is an $n \times k$ matrix whose corresponding rows are the n experimental runs in an experiment (Borkowski, 2018). For example, for a

2^2 design, the design matrix is

$$\begin{array}{cc} x_1 & x_2 \\ \left[\begin{array}{cc} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{array} \right] \end{array}.$$

Orthogonality and factor levels of $x_j \in \{-1, +1\}$ in the design matrix ensures the variances of regression coefficients in the first-order model are minimized. That is, this design is variance-optimal (Box and Draper, 2007).

Fractional factorial designs of resolution *III* and higher are widely used first-order orthogonal designs in industry. A design is resolution *R* if no k factor effect is aliased with another effect containing less than $R - k$ factors (Myers et al., 2016). When two effects are aliased, both effects have identical columns in the model matrix X . For example, resolution *III* designs have no main effect aliased with any other main effect, but at least one main effect is aliased with a two-factor interaction.

If we consider a 2^{4-1} fractional factorial design of resolution *IV*, the model matrix

X for the first-order model, $y = \beta_0 + \sum_{i=1}^4 \beta_i x_i + \epsilon$ is given by

$$X = \begin{matrix} & x_1 & x_2 & x_3 & x_4 \\ \begin{bmatrix} 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 \end{bmatrix} \end{matrix}.$$

The matrix $(X'X)^{-1} = 1/8I_5$, and no other design with eight experimental runs in this region will have parameter estimates with variances smaller than $\sigma^2/8$ (Khuri and Mukhopadhyay, 2010). That is, this fractional factorial design is variance-optimal for a first-order model.

To permit estimation of experimental error for a 2^k design, replication of experimental runs is required. With quantitative factor levels, the design can also be augmented with several replications at the center $(0, 0, \dots, 0)$ (Montgomery, 2017). The addition of center runs does not change the orthogonality property of the design but the design is no longer variance-optimal as the requirement that all levels be at ± 1 is not met.

Simplex designs form another class of first-order orthogonal designs. These designs are saturated for first-order models in the sense that a first-order model can be fit, but for inference there are zero residual degrees of freedom (Montgomery, 2017). Center points can be added to provide residual degrees of freedom. The

simplex is a regular sided figure with $k + 1$ vertices in k dimensions (Khuri and Mukhopadhyay, 2010). For instance, the triangle is a regular sided figure with 3 vertices in 2 dimensions. Thus, it requires $N = k + 1$ observations to fit a first-order model with k variables. The design points in the design matrix X are arranged such that any two points make an angle of θ with the origin where

$$\cos\theta = \frac{-1}{k}.$$

For $k = 2$ variables, $\theta = 120^\circ$ and three experimental runs, the design matrix is given by

$$\begin{array}{cc} x_1 & x_2 \\ \left[\begin{array}{cc} \sqrt{3/2} & -1/\sqrt{2} \\ -\sqrt{3/2} & -1/\sqrt{2} \\ 0 & 2/\sqrt{2} \end{array} \right] \end{array}.$$

These points represent the vertices of an equilateral triangle.

1.1.2.2 Orthogonal Designs for First-Order Linear Models with Interactions

Some processes or systems require a first-order linear model with interactions involving the cross-products of the linear terms. This model has the form

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i < j=2} \sum \beta_{ij} x_i x_j + \epsilon. \quad (1.5)$$

The two-level full factorial designs are orthogonal and variance-optimal for this model (Montgomery, 2017). For example, if there are $k = 3$ design variables then a full 2^3 factorial design is orthogonal and variance-optimal. If a fractional factorial is to be used, then there must be sufficient resolution such that model terms are not aliased with each other. If two-factor interactions are included in the model but no higher-

order interactions, then a resolution of at least V will be required.

1.1.2.3 Designs for Fitting Second-Order Models If curvature is expected, then the following second-order model can be used,

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i < j} \beta_{ij} x_i x_j + \epsilon.$$

There must be at least $1 + 2k + k(k - 1)/2 = \binom{k+2}{2}$ distinct design points to equal the number of parameters in the model, and there must be at least three levels of each design variable to fit this model, otherwise $X'X$ is singular (Montgomery, 2017).

Central Composite Designs (Box and Wilson, 1951)

The class of central composite designs (CCD) is the most popular class of designs for fitting second-order models. A CCD consists of:

1. A 2^k full factorial or fraction of the 2^k design of at least resolution V . Each factorial point is coded to be of the form $(x_1, x_2, \dots, x_k) = (\pm 1, \pm 1, \dots, \pm 1)$. The factorial points, therefore, represent a variance-optimal design that allow for the estimation of first-order and interaction terms.
2. $2k$ axial points of the form $(\pm\alpha, 0, 0, \dots, 0), (0, \pm\alpha, 0, \dots, 0), \dots, (0, 0, \dots, 0, \pm\alpha)$. These axial points allow for the estimation of the pure quadratic terms in the model. The CCD design is referred to as a face-centered cube design when $\alpha = 1$, and it is called a spherical design when $\alpha > 1$.
3. n_c center runs of the form $(x_1, x_2, \dots, x_k) = (0, 0, \dots, 0)$. These center point replicates allow for the estimation of pure error and also provide information about the presence of curvature in the model.

Box-Behnken Designs

The class of Box-Behnken designs (BBD) were introduced by Box and Behnken (1960) and are also used for fitting second-order models. These designs are formed by combining two-level factorial or fractional factorial designs with incomplete block designs. The incomplete block design consists of k design variables, b blocks, t design variables in a block, and r is the number of times a design variable appears in the whole design. Most BBDs are arranged similarly as balanced incomplete block designs such that the number of times a pair of design variables appears in the same block is $\lambda = \frac{r(t-1)}{k-1}$ (Montgomery, 2017). Myers et al (2016) describe the construction of a BBD as follows:

1. Start with a 2^t full factorial or fractional factorial (for large t) arrangement of the specific t design variables appearing in a block.
2. Set the levels of the remaining $k - t$ design variables in a block to 0.
3. Add center points $(0, 0, \dots, 0)$.

The arrangement allows for the efficient estimation of first-order and second-order terms in the quadratic model.

1.2 Response Surface Designs for Mixture Experiments

In classical experiments, the response values depend on the amounts or levels of each independent variable, and the levels of the independent variables can be varied independently of the others. In some experiments, however, the levels of the input variables cannot be varied independently of the others and the response depends on the relative proportions of the components that comprise a mixture (Box and Draper, 2007; Cornell, 2011). This type of response surface experiment is called a mixture

experiment. For example, the taste of an orange and pineapple juice blend depends on the proportion of each fruit in the blended juice. The taste of the blended juice does not change if the amount of orange and pineapple are doubled or tripled. That is, the response is driven by the proportion of the input variables (ingredients or components) rather than the amounts of the components in the mixture. A second example is the strength of a stainless steel alloy which depends on the proportions of iron, carbon, nickel, and chromium used.

Many end products or materials in industry are formed by mixing two or more components (ingredients). The aim of many mixture experiments is to search for a blended product or mixture that has better product properties than a single component product (Cornell, 2011). For example, in building construction, the goal is to mix certain proportions of sand, water, and cement to create a concrete with optimum physical properties to boost demand and profit for the manufacturer. The component proportions are nonnegative and the sum of the nonnegative component proportions is one (unity). If a mixture is assumed to have q components, and x_i represents the i^{th} component proportion in the mixture, then

$$x_i \geq 0, i = 1, 2, \dots, q \quad \text{and} \quad \sum_{i=1}^q x_i = 1.$$

As a result of these fundamental constraints in mixture experiments, the levels of each component cannot be selected independently. Hence, the level of any component x_i is completely determined by the other $q - 1$ component levels. For example,

$$x_q = 1 - \sum_{i=1}^{q-1} x_i.$$

That is, if we know the proportion of sand and water in a concrete then the proportion of cement is also known.

Despite the dependency between levels of components, the analytical tools in response surface methodology can still be used in mixture experiments. If we consider a $(q = 2)$ -component mixture experiment and assume a second-order model to be a good approximating representation of the mixture surface, then the fitted second-order model is

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2,$$

where the b_i are the least-square estimates of model coefficients. Using the facts that

$$1 = x_1 + x_2,$$

$$x_1^2 = x_1x_1 = x_1(1 - x_2) = x_1 - x_1x_2,$$

and

$$x_2^2 = x_2x_2 = x_2(1 - x_1) = x_2 - x_1x_2.$$

Then, after substituting, the fitted second-order model becomes

$$\begin{aligned}\hat{y} &= b_0(x_1 + x_2) + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}(x_1 - x_1x_2) + b_{22}(x_2 - x_1x_2) \\ &= (b_0 + b_1 + b_{11})x_1 + (b_0 + b_2 + b_{22})x_2 + (b_{12} - b_{11} - b_{22})x_1x_2,\end{aligned}$$

which can be simplified to a three-parameter model:

$$\hat{y} = a_1x_1 + a_2x_2 + a_{12}x_1x_2.$$

The resulting three-parameter mixture model is a reduction of the corresponding over-parameterized six-parameter second-order response surface model. In general, every x_i^2 term in a q -factor second-order model can be replaced with $x_i^2 = x_i(1 - \sum_{j \neq i}^q x_j)$,

and the intercept for a q -factor response surface model can be removed (Smith, 2005).

In terms of model parameters, a full cubic mixture model is

$$\hat{y} = \sum_{i=1}^q a_i x_i + \sum_{i<j}^{q-1} \sum_j^q a_{ij} x_i x_j + \sum_{i<j}^{q-1} \sum_j^q \delta_{ij} x_i x_j (x_i - x_j) + \sum_{i<j}^{q-2} \sum_{j<k}^{q-1} \sum_k^q a_{ijk} x_i x_j x_k.$$

This third-degree polynomial is also derived by removing the intercept and replacing every x_i^2 term with $x_i(1 - \sum_{j \neq i}^q x_j)$. The special cubic mixture model removes the $\delta_{ij} x_i x_j (x_i - x_j)$ terms from the full cubic mixture model, leaving

$$\hat{y} = \sum_{i=1}^q a_i x_i + \sum_{i<j}^{q-1} \sum_j^q a_{ij} x_i x_j + \sum_{i<j}^{q-2} \sum_{j<k}^{q-1} \sum_k^q a_{ijk} x_i x_j x_k.$$

The reduced mixture models were suggested by Scheffe (1958, 1963) and are called Scheffe models or canonical mixture models because they retain the symmetry in the x 's. Table 1.1 summarizes the number of terms for q -component mixture experiments for linear, quadratic and full cubic canonical mixture models.

Because of the fundamental constraints on the components in a mixture experiment, the standard design region containing the q components is a regular $(q - 1)$ -dimensional simplex. The simplex is the generalization of a triangular or tetrahedral region of space to arbitrary dimensions (Pawlowsky-Glahn and Egozcue, 2001). For $q = 2$ components, the simplex design region is a line segment between $(0, 1)$ and $(1, 0)$ shown in Figure 1.2a. For $q = 3$ components, the simplex design region shown in Figure 1.2b is an equilateral triangle with vertices $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$.

The coordinate system used for the components is called a simplex coordinate system. For example, with a 3-component mixture, the pure mixtures or single-component blends occur at vertices of the triangle $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. The

Table 1.1: Number of Terms in Scheffe Mixture Models

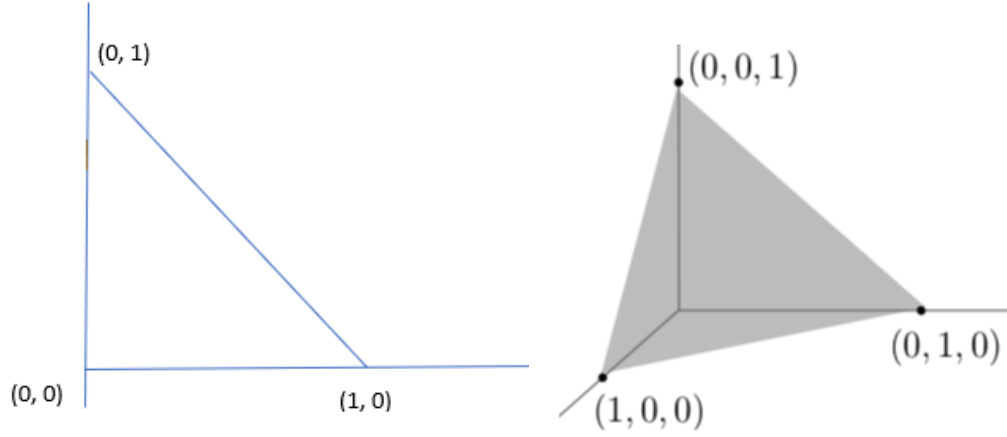
Number of components (q)	Linear	Quadratic	Full cubic
2	2	3	-
3	3	6	10
4	4	10	20
5	5	15	35
6	6	21	56
.	.	.	.
.	.	.	.
.	.	.	.
q	q	$q(q+1)/2$	$q(q+1)(q+2)/6$

interior points of the equilateral triangle represent nonzero component proportions ($0 < x_i < 1, i = 1, 2, 3$). The centroid of the triangle represents equal proportions ($1/3, 1/3, 1/3$) of the three components. The binary blends or two-component mixtures are the edge points $(a, 1-a, 0)$, $(a, 0, 1-a)$, and $(0, a, 1-a)$, for $0 < a < 1$.

When there are q components in a mixture experiment, there are infinitely many possible mixtures in the $(q-1)$ -dimensional simplex design region. The simplex-lattice, simplex centroid, and the axial designs have been used to address this issue (Cornell, 2011; Scheffe, 1958).

1.2.1 Simplex-Lattice Designs

Scheffe (1958) introduced the simplex-lattice designs. A simplex lattice is a certain type of structural arrangement of points on a simplex (Cornell, 2011). For simplex-lattice designs, a lattice is defined by the assumed polynomial model for the



(a) 2-component space

(b) 3-component space

Figure 1.2: Simplex factor spaces for $q = 2$ and $q = 3$ components (Box and Draper, 2007)

mixture response surface. The number of design points depend on the number of mixture components q , and the order of the assumed polynomial model m . A $\{q, m\}$ simplex-lattice design for q component mixtures consists of combinations of all valid design points formed from $(m + 1)$ equally spaced levels of component proportions. That is, $x_i \in \{0, \frac{1}{m}, \frac{2}{m}, \dots, \frac{m-1}{m}, 1\}$ for $i = 1, 2, \dots, q$, and all points formed must satisfy the mixture constraint $\sum_{i=1}^q x_i = 1$. The $\{q, m\}$ simplex-lattice design contains $\binom{q+m-1}{m}$ design points and the design can be used to fit a Scheffe polynomial model up to an order of m . For example, a $\{3, 3\}$ simplex-lattice design can be used to fit a Scheffe polynomial model up to an order of 3 and consists of 10 design points whose component levels are $0, 1/3, 2/3, 1$ such that

$$(x_1, x_2, x_3) = (1, 0, 0), (0, 1, 0), (0, 0, 1), (2/3, 1/3, 0), (2/3, 0, 1/3), (1/3, 2/3, 0),$$

$(1/3, 0, 2/3)$, $(0, 2/3, 1/3)$, $(0, 1/3, 2/3)$, and $(1/3, 1/3, 1/3)$. The $\{3, 2\}$ simplex-lattice design can be used to fit a Scheffe model up to an order of 2 and the design consists of the following 6 points:

$$(x_1, x_2, x_3) = (1, 0, 0), (0, 1, 0), (0, 0, 1), (1/2, 1/2, 0), (1/2, 0, 1/2), (0, 1/2, 1/2).$$

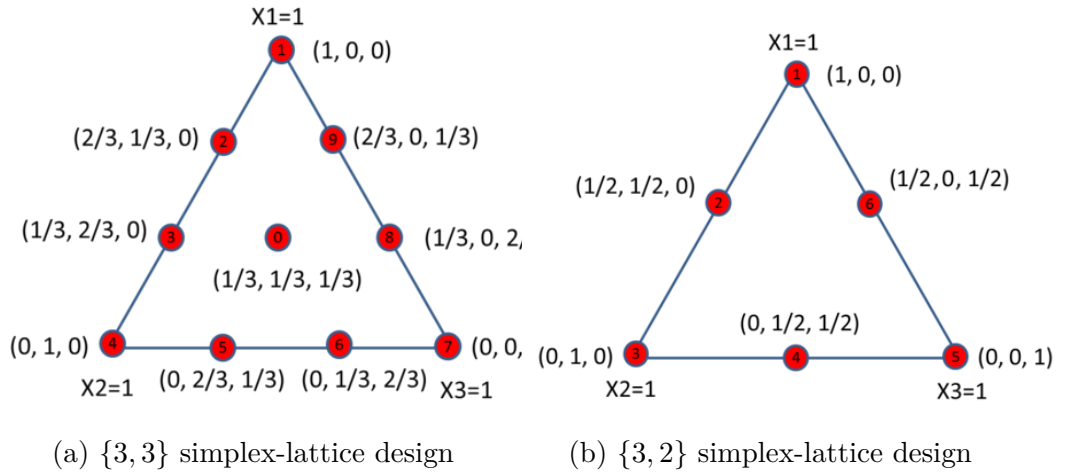


Figure 1.3: $\{3, 3\}$ and $\{3, 2\}$ simplex-lattice designs

The $\{3, 2\}$ simplex-lattice design in Figure 1.3b does not have any interior points, but it can be augmented with the centroid $(1/3, 1/3, 1/3)$. In general, $\{q, 2\}$ simplex-lattice designs do not contain any interior points but can be supplemented with centroids of each 2-dimensional face. Such augmented $\{q, 2\}$ designs are called special cubic simplex lattice designs. A special cubic simplex lattice design is therefore a $\{q, 2\}$ simplex-lattice design supplemented with $\binom{q}{3}$ centroids.

Table 1.2 summarizes the number of design points for $\{q, m\}$ simplex-lattice designs as well as the model that each design can fit.

Table 1.2: Number of Points in $\{q, m\}$ Simplex-lattice Design

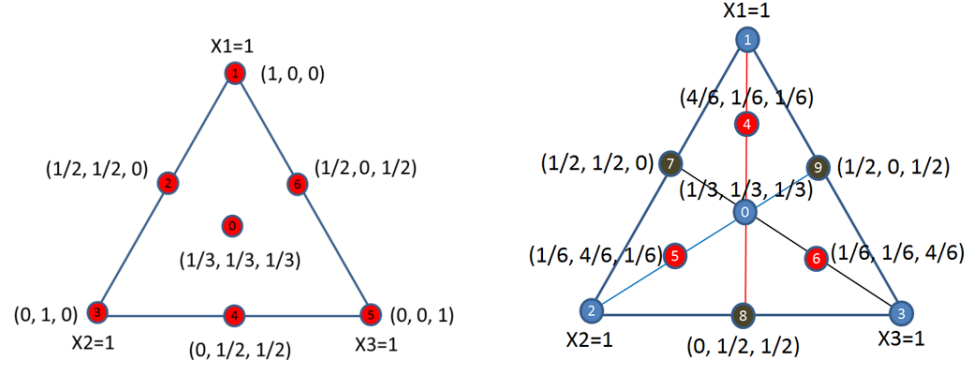
Model	Number of components (q)					
	3	4	5	6	7	8
First-order ($m = 1$)	3	4	5	6	7	8
Quadratic ($m = 2$)	6	10	15	21	28	36
Special cubic ($m = 2$)	7	14	25	41	63	92
Cubic ($m = 3$)	10	20	35	56	84	120
Quartic ($m = 4$)	15	35	70	126	210	330

1.2.2 Simplex Centroids Designs

A q -component simplex centroid design has $2^q - 1$ points and provides information at the centroids of the $(q - 1)$ -dimensional simplex design region. The design consists of q pure blends, $\binom{q}{2}$ binary blends, $\binom{q}{3}$ ternary blends and continuing until the last point is the overall centroid, $(1/q, \dots, 1/q)$. The model that can be fitted depends on how many of the subsets of points are included. For the full q -component simplex centroid design, a Scheffe polynomial model up to an order of q can be fitted. A 3-component simplex centroid design is the same as the special cubic simplex lattice design. An example of a 3-component simplex centroid design is shown in Figure 1.4a.

1.2.3 Simplex Axial Designs

An axis is a line segment joining a vertex of a simplex S_q with its centroid. A simplex axial design consists of q design points on q axes such that the distance d ($0 < d < (q - 1)/q$) of each design point to the centroid is the same, q pure blends, q binary blends and one overall center point. A 3-component simplex axial design with $d = \sqrt{\frac{1}{6}}$ is shown in Figure 1.4b. Points 4, 5, and 6 are the axial blends, points 1, 2, and 3 are pure blends, points 7, 8, and 9 are binary blends, and point 0 is the overall



(a) 3-component simplex centroid design (b) 3-component simplex axial design

Figure 1.4: 3-component simplex centroid and simplex axial designs

center point.

1.2.4 Minimum and Maximum Level Component Constraints

In many situations, the entire simplex cannot be explored due to restrictions placed on component proportions. For example, one cannot have coffee if the proportion of coffee is zero. In that instance, it may be reasonable to consider mixtures with at most 70% coffee, and with sugar and milk making up at least 30% of the mixture. The restrictions would then limit the desired mixtures to a subspace of the simplex. A mixture design that has lower and/or upper limit constraints for each component is called a single component constraint (SCC) design.

When only lower bounds L_i are placed on each of the components ($x_i \geq L_i$) in a q -component mixture, the resulting restricted design region of the $(q - 1)$ -dimensional simplex is itself a simplex (Cornell, 2011). For a 3-component mixture, placing minimum restrictions $x_1 \geq 0.2$, $x_2 \geq 0.15$, and $x_3 \geq 0.12$ on the three components yields the blue shaded triangle in Figure 1.5.

Since the minimum restrictions on the component proportions produce a simplex-shaped subspace of the original, coordinates of the subspace can be redefined

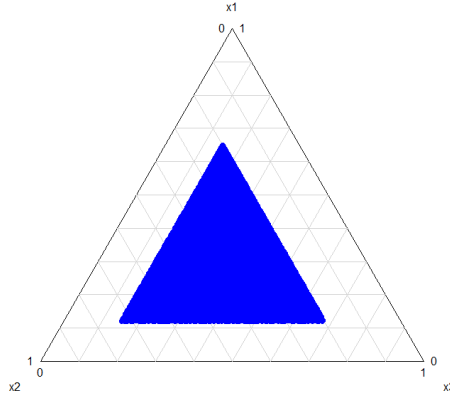


Figure 1.5: Example of a restricted design region from placing lower bounds on a 3-component mixture, $x_1 \geq 0.2$, $x_2 \geq 0.15$, and $x_3 \geq 0.12$.

in terms of an alternative system called pseudocomponents. The pseudocomponents x'_i are defined as

$$x'_i = \frac{x_i - L_i}{1 - L}, \text{ where } L = \sum_{i=1}^q L_i < 1 \text{ and } x'_i \geq 0, \sum_{i=1}^q x'_i = 1.$$

The shape and orientation of the pseudocomponents x'_i design region (simplex) is the same as the shape and orientation of the original components x_i design region. Hence, an appropriate design with levels corresponding to the pseudocomponents can be chosen and these levels can be transformed back to the original setting using

$$x_i = L_i + (1 - L)x'_i.$$

The experiment can then be conducted using the original component levels.

In certain mixture experiments, lower restrictions L_i and upper restrictions U_i are placed on component proportions x_i . Such restrictions can be written as

$$0 \leq L_i \leq x_i \leq U_i \leq 1$$

where all or some of the L_i may be zero and some of the U_i may be unity. The resulting design space is the set of mixture points in the polytope defined by the lower and upper restrictions. For example, the blue shaded polygon in Figure 1.6 is a result of using a 3-component mixture with $0.1 \leq x_1 \leq 0.75$, $0.15 \leq x_2 \leq 0.8$, $0.1 \leq x_3 \leq 0.6$.

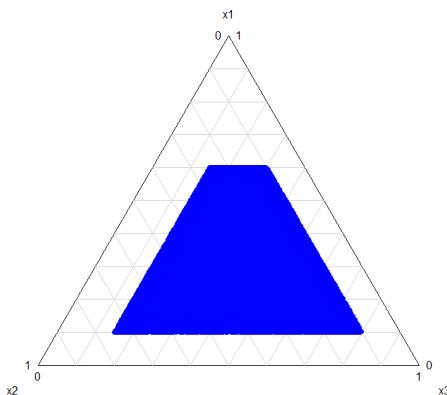


Figure 1.6: Example of a restricted design region from placing lower and upper bounds on a 3-component mixture, $0.1 \leq x_1 \leq 0.75$, $0.15 \leq x_2 \leq 0.8$, $0.1 \leq x_3 \leq 0.6$.

Generally, the vertices of the constrained region are used as design points when a first-order model is considered. The coordinates of the centroids of the faces, sides or edges may be added when a higher order model is considered. The extreme vertices (McLean and Anderson, 1966) and XVERT (Snee and Marquardt, 1974) are two algorithms used to locate vertices and centroids of the constrained region. To generate an extreme vertices design for a q -component mixture with lower (L_i) and upper (U_i) constraints, the steps are:

1. All ordered $q2^{q-1}$ combinations of (L_i) and (U_i) are written down leaving one component blank.
2. All admissible combinations (that is, $\sum_{i=1}^q x_i = 1$ and all constraints are satisfied) are filled in. Each admissible combination is a vertex of the

constrained region.

3. Set apart all vertices for which $q - r$ component levels remain constant and take the average of the remaining r components levels to form r -dimensional centroids ($r \geq 2$).

The construction of an extreme vertices design for the restricted region in Figure 1.6 is as follows.

1. All ordered combinations of (L_i) and (U_i) are written down leaving one component blank.

x_1	x_2	x_3
0.1	0.15	
0.1	0.8	
0.75	0.15	
0.75	0.8	
0.1		0.1
0.1		0.6
0.75		0.1
0.75		0.6
	0.15	0.1
	0.15	0.6
	0.8	0.1
	0.8	0.6

2. All admissible combinations are filled in (denoted in bold).

x_1	x_2	x_3
0.1	0.15	
0.1	0.8	0.1
0.75	0.15	0.1
0.75	0.8	
0.1	0.8	0.1
0.1	0.3	0.6
0.75	0.15	0.1
0.75		0.6
0.75	0.15	0.1
0.25	0.15	0.6
0.1	0.8	0.1
	0.8	0.6

3. Four centers of edges points are formed by fixing

$x_1 = 0.1, 0.75$, $x_2 = 0.15, 0.8$, and $x_3 = 0.1, 0.6$ resulting in four feasible edge

Fixed value	Vertex values	New centroid
$x_1 = 0.1$	(0.1 0.8 0.1)	(0.1 0.55 0.35)
	(0.1 0.3 0.6)	
$x_2 = 0.15$	(0.75 0.15 0.1)	(0.45 0.15 0.35)
	(0.25 0.15 0.6)	
$x_3 = 0.1$	(0.1 0.8 0.1)	0.425 0.475 0.1)
	(0.75 0.15 0.1)	
$x_3 = 0.6$	0.1 0.3 0.6)	(0.175 0.225 0.6)
	(0.25 0.15 0.6)	

centroids. The overall centroid is $(0.3, 35, 0.35)$, calculated by

$$x_1 = (0.1 + 0.1 + 0.75 + 0.25)/4 = 0.3,$$

$$x_2 = (0.8 + 0.3 + 0.15 + 0.15)/4 = 0.35,$$

$$x_3 = (0.1 + 0.6 + 0.1 + 0.6)/4 = 0.35.$$

The resulting 9 points of the extreme vertices design are shown in Figure 1.7.

When there are many vertices, the XVERT algorithm can be used for selecting a subset of extreme vertices based on some optimality criterion. A comprehensive review of this algorithm can be found in (Cornell, 2011). Optimal computer generated designs (Coetzer and Haines, 2017; Goos et al., 2016; Limmun et al., 2019) based on different optimality criteria are also used to generate mixture experiments with constraints.

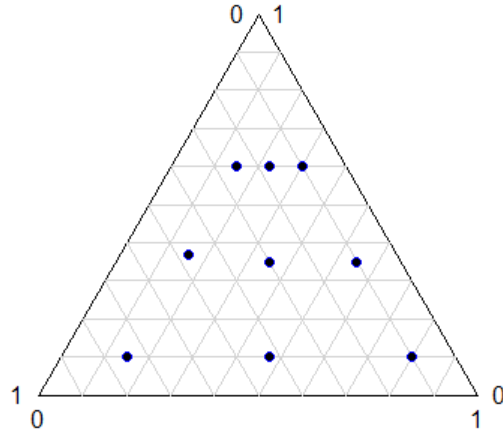


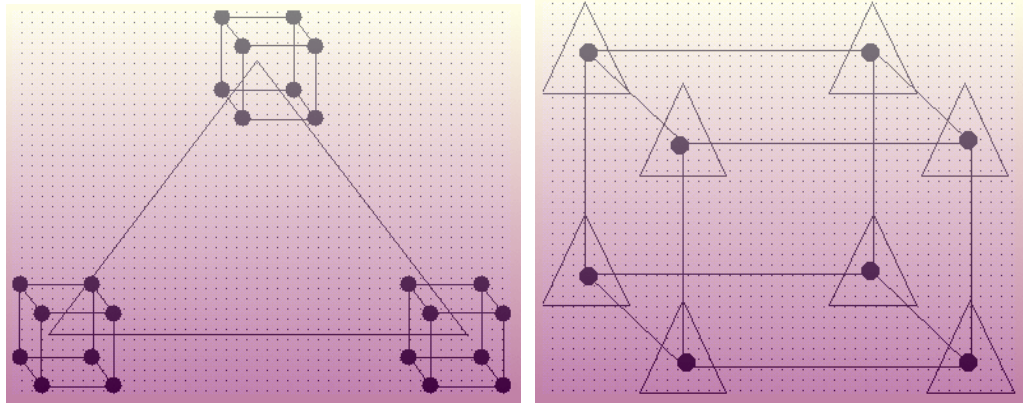
Figure 1.7: A 9-point extreme vertices design for the experimental region in Figure 1.6.

1.2.5 Mixture Experiments with Process Variables

Suppose we consider an orange and pineapple juice blend. The taste of the blend may also depend on how cold or warm the juice is. Hence, temperature may have an influence on the blending properties of the juice. That is, in some mixture experiments, the response of interest may also depend on process conditions. Process variables $(z_1, \dots, z_k)$ in mixture experiments are experimental factors that do not form any portion of the mixture but can potentially affect or influence the response output of the mixture (Cornell, 2011).

Suppose a mixture experiment consists of q components and k process variables. Then, the design space for the q components is the $(q-1)$ -dimensional simplex and the region of interest for the k process variables may be a k -dimensional hypercube. The combined design region of interest is then of dimensionality $q-1+k$ (Pradhan et al., 2017). Cornell (1971) introduced graphical displays of design point configuration for mixture experiments with process variables. The process variable settings are displayed in a response surface design arrangement at each experimental mixture in

the simplex region or the experimental mixtures in the mixture simplex are displayed at each combination of the process variables.



(a) 2^3 full factorial arrangement at the vertices of a 3-component mixture simplex region
(b) 3-component mixture simplex design at each point of 2^3 full factorial run

Figure 1.8: Example of a combined mixture and process variable design

Figure 1.8 shows the combined space for a 3-component mixture experiment with three process variables in a 2^3 design configuration. The combined design can be presented as either a full factorial run at each vertex of the simplex region (Figure 1.8a) or a 3-component simplex design at each point of the 2^3 full factorial design (Figure 1.8b). To minimize the number of experimental runs, a fractional factorial arrangement in the settings of the process variables may be considered instead of a full factorial arrangement.

A common model for a mixture experiment with process variables is the product of a Scheffe mixture model in q components and a response surface model in k process variables (Cornell, 2011). For example, the combined model for a q component quadratic mixture model and an interaction response surface model is the product model

$$y = \left(\sum_{i=1}^q \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j \right) \left(\alpha_0 + \sum_{n=1}^k \alpha_n z_n + \sum_{n < l} \alpha_{nl} z_n z_l \right) + \epsilon,$$

which after expanding and substituting products $\alpha_i\beta_j$ with a single γ -parameters, becomes the simplified model

$$y = \left(\sum_{i=1}^q \gamma_i^0 + \sum_{n=1}^k \gamma_i^n z_n + \sum_{n<l} \gamma_i^{nl} z_n z_l \right) x_i + \left(\sum_{i<j}^q \gamma_{ij}^0 + \sum_{n=1}^k \gamma_i^n z_n + \sum_{n<l} \gamma_i^{nl} z_n z_l \right) x_i x_j + \epsilon.$$

1.3 Uniform Designs

Full factorial designs, fractional factorial designs, response surface designs (such as optimal designs, orthogonal arrays, and central composite designs) are all classical statistical experimental designs that have played a significant role in the industrial applications including optimizing responses such as the yield of a product produced from a chemical process under different reaction conditions, tensile strength of material produced from different raw materials, or the mean time to failure of an electrical component manufactured under different settings (Fang and Lin, 2003).

One main goal of designing an experiment is to explore how the response output is affected by relevant contributing factors (Zhang et al., 1998). That is, we seek to gain an insight into the relationship between the inputs driving the output. Such an understanding is crucial, especially in the manufacturing industries, for optimizing product yield and maximizing profit. This relationship may be formulated as

$$output = f(inputs) + \epsilon,$$

where ϵ is the random error associated with the response not accounted for by the assumed function $f(\cdot)$. Most classical designs are associated with a particular functional form of $f(\cdot)$ (Li et al., 2004). One common functional form used assumes

that the response output Y is a linear combination of the input variables such that,

$$Y = \beta X + \epsilon,$$

where β is the vector of parameter coefficients associated with an assumed linear model, and X is the model matrix. Homoscedasticity, independence and normal distribution of errors are all common assumptions that aid in data analysis. One key property of a good experimental design is efficiency (Myers et al., 2016). That is, a good design helps obtain sufficient information for parameter estimation in a small number of experimental runs. When the number of input factors and/or factor levels increases, the number of parameters to estimate increases exponentially. Accordingly, the number of experimental runs required also increases exponentially. A full factorial design is used when there are enough resources to run all combinations of factor levels. However, there are often not enough resources, and a fractional factorial design may be an alternative. A fractional factorial design allows for estimation of lower-order polynomial terms in a model, thereby reducing the number of experimental runs. The disadvantages in using fractional factorial designs come from the fact that most commonly used fractional factorial designs are unreplicated and each model effect is aliased with one or more other effects (Dean et al., 2017).

Fractional factorial and optimal designs may also be used to reduce the number of experimental runs. These designs, however, depend on the form of an assumed model that describes the relationship between factors and the mean of the response. In reality, this relationship is unknown and the assumptions imposed on the error distribution may not hold. In such situations, a uniform design (UD), which spreads the design points uniformly over the entire design region may be ideal for exploring the relationship between the response and input factors.

Uniform space-filling designs are based on number theory and quasi-Monte Carlo Methods (Fang and Wang, 1994). Quasi-Monte Carlo methods are methods used for numerical analyses and solving various problems using quasi-random sequences or deterministic inputs (Gerber and Chopin, 2015). Methods used for examining uniform designs are called number theoretic methods (NTM). Fang and Wang (1994) showed that the rate of convergence of the mean response using quasi-Monte Carlo methods is

$$|E(f(x)) - \bar{f}_N| \leq O[N^{-1}(\log(N))^s] \quad (1.6)$$

for a design region that is an s -dimensional unit cube C^s , where $E(f(x))$ is the expected value of the mean response at input x and $\bar{f}_N = \frac{1}{N} \sum_{i=1}^N Y$ is the sample mean. On the other hand, by the Central Limit Theorem, the rate of convergence of the Monte Carlo method is $N^{-1/2}$. That is, uniform designs are more efficient than Monte Carlo methods because uniform designs choose design points in a deterministically uniform manner, while Monte Carlo methods choose points in a completely random manner (Fang and Wang, 1994). This is visually evident in Figure 1.9. The 21 design points using a uniform design are more uniformly scattered over the 2-dimensional unit cube C^2 than a Monte Carlo approach over the same region.

1.3.1 Advantages of Uniform Designs

For most industrial applications, uniform designs have been used in situations where the relationship between the mean responses of interest and the experimental factors are unknown. A primary goal of most classical designs is to estimate some unknown model parameters under some assumed model forms. In such situations, optimal designs that give high-precision parameter estimates are chosen. Uniform

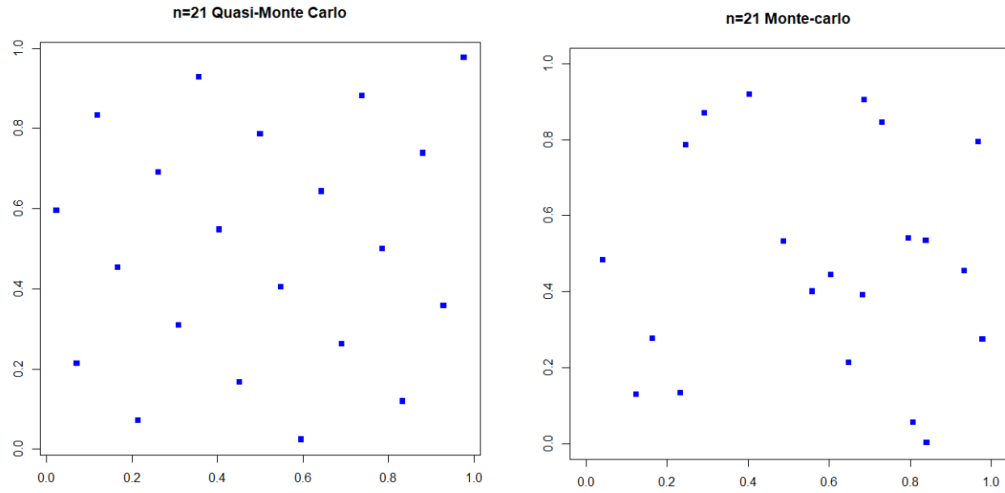


Figure 1.9: 21-point designs using uniform design and Monte Carlo approaches.

designs are valuable in situations where the underlying model form is unknown. Due to its space-filling properties, a uniform design allows the researcher to estimate the mean response of interest without knowing the functional form of the response surface model prior to data collection. This gives the researcher the flexibility of fitting a wide variety of models so the researcher chooses the *best* among the set of models for the experimental data collected (Fang and Lin, 2003).

Another advantage of uniform designs is that they are not defined in terms of combinatorial structure like the classical designs, but rather, in terms of the spread of the design points over the entire experimental region. Uniform designs are therefore valuable for experiments with a large number of factors and a large number of levels of the factors. For example, in an experiment with 5 factors and 5 levels on each factor, a full factorial design requires that the number of runs should be $5^5 = 3125$. With uniform designs, it is possible to complete the experiment with a relatively small number of runs. Therefore, the number of runs required for uniform designs is smaller than orthogonal arrays, saving valuable resources and time (Pham, 2006).

The usefulness of many classical and optimal designs depends on the accuracy of the assumed model form (Wiens, 1991). To obtain optimal designs, an assumption is made that the proposed model is exactly correct. In many instances, however, the true model form deviates from the assumed model form and the use of these classical and optimal designs will be less than ideal. Another advantage of uniform designs is that they are flexible to departures from an assumed model type, as they allow several models to be fitted that are not possible with classical designs.

A model lack-of-fit test is any statistical test that addresses the fit of an assumed model. The relationship between uniform designs and model lack-of-fit (LOF) was addressed by Wiens (1991). Suppose, a “full” response model is written as

$$Y = \beta'g(x) + h(x) + \epsilon, \quad (1.7)$$

where β is the vector of unknown parameters, $g(x)$ is a vector of specified functions, and $h(x)$ is the unknown bias. If the bias $h(x) = 0$, for all points in the experimental region, then the “reduced” model is

$$Y = \beta'g(x) + \epsilon. \quad (1.8)$$

The test statistic for model lack-of-fit of the “reduced” model within the “full” model for a design with replication is

$$F = \frac{(n - p)S^2 - (n - c)S_{PE}^2}{(c - p)S_{PE}^2}, \quad (1.9)$$

where c is the number of distinct replications, p is the number of unknown parameters, S^2 is the least square estimate of σ^2 , and S_{PE}^2 is the pure error estimate. Wiens (1991) showed that if the “full” model is true and the following two conditions are satisfied:

1. $\int_{C^s} h^2(x)dx$ is bounded below and
2. $\int_{C^s} \beta' g(x)h(x)dx = 0$,

then the continuous uniform design maximizes the power of the lack-of-fit test. A continuous uniform design is a design which establishes a continuous uniform probability measure over the design region such that all intervals of the same length over the design space are equally probable. For unreplicated designs, if the “full” model is true and the following two conditions hold:

1. $\int_{C^s} h^2(x)dx$ is bounded above and
2. $\int_{C^s} \beta' g(x)h(x)dx = 0$,

then the maximum bias in the estimation of σ^2 is minimized by continuous uniform designs. Wiens (1991) asserts that continuous uniform designs are impractical in most instances. However, continuous uniform designs may be approximated by discrete uniform designs with evenly-spaced design points throughout the experimental region. This goes to show that, unlike many classical designs, we do not have to know or assume the true model form before data collection in order to select a good uniform design.

Another significant property of a good uniform design is its coverage to lower-dimensional subspace projections over a design region (Damblin et al., 2013). Suppose the design region is an s -dimensional unit cube experimental region (C^s), then a good uniform design for the s -dimensional region ensures that any k -dimensional, $k < s$, subspace of C^s has good space-filling properties. This property is really important because a good coverage on the subspace formed by these experimental inputs is necessary to capture an interaction effect between certain experimental inputs (Damblin et al., 2013; Joseph et al., 2015). Also, when a selection of

experimental inputs is made during factor screening, good space-filling projections onto the smaller subsets of experimental inputs provide valuable information of the sub-region spanned by the selected active experimental inputs (Cheng, 2006; Joseph et al., 2015).

1.3.2 Methods of Generating Uniform Design Points

Fang (1980) and Wang and Fang (1981) first applied Number Theoretic Methods (NTMs) to develop Uniform Designs (UDs) by generating a set of points called a set of quasi-random numbers or a number-theoretic net (NT-net). A NTM is a quasi-Monte Carlo method which is a combination of number theory and numerical analysis used to solve a variety of statistical problems such as numerical evaluation of probabilities and moments of continuous multivariate distributions in different experimental regions such as the ball, sphere, cube, and the simplex (Fang and Wang, 1994). If the goal is to design an experiment for s variables over an s -dimensional unit cube experimental region (C^s), then a set of n design points $P_n = \{X_1, \dots, X_n | X_i \in C^s\}$ is chosen such that these points are “uniformly” scattered in C^s . An n -point UD is denoted by $U_n(q^s)$, where q is the same number of levels for each of the s factors. If s_1 and s_2 factors have q_1 and q_2 levels, respectively, then an n -point UD is denoted by $U_n(q_1^{s_1} \times q_2^{s_2})$. This can be generalized to three or more factors.

Different ways of generating uniform space-filling design points are examined in Fang and Wang (1994) and Hua and Wang (2012). Four of these methods are now presented.

1.3.2.1 Good Lattice Point Sets (Fang and Wang, 1994) As an approximation method for evaluating multiple integrals, Korobov (1959) proposed the good lattice point (GLP) method. This quasi Monte-Carlo method involves calculating average

function values in the integration region over a uniformly scattered set of points. Therefore, it is different from the ordinary Monte-Carlo method, which evaluates multiple integrals based on averaging function values over a random point selection in the integration region (Warnock, 1972).

This method's calculation starts with finding a candidate set of positive integers called a generating vector $h_n = (h_1, \dots, h_s)$ where s is the number of factors in an n -run experiment and the greatest common divisor, $\gcd(h_i, n) = 1$, for $i = 1, 2, \dots, s$. The i^{th} design row is calculated using $ih_n = (ih_1, \dots, ih_s) \bmod n$, for $i = 1, 2, \dots, n$ so that each column is a permutation of $1, 2, \dots, n$ (Fang and Wang, 1994). This is accomplished by replacing the 0 in each column with n . Fang et al. (2006) showed that Euler's function value $\phi(n)$ equals the maximum number of h_i 's for a given n and is defined as

$$\phi(n) = n \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \dots \left(1 - \frac{1}{p_r}\right),$$

where $p_1^{t_1} \times \dots \times p_r^{t_r}$ is the prime decomposition of n and $t_1, \dots, t_r$ are positive integers. An $n \times s$ matrix $Q_s = (q_{ij})$ is formed from the vector h_n , such that $q_{ij} = ih_j \bmod n$, for $i = 1, \dots, n$ and $j = 1, \dots, s$. Each column of Q_k is a permutation of $(1, 2, \dots, n)'$ by replacing each 0 with n . From the matrix Q , an $n \times s$ design matrix X_s in C^s is formed such that

$$x_{ij} = \frac{2q_{ij} - 1}{2n}$$

for $i = 1, \dots, n$ and $j = 1, \dots, s$. Each column of X_s is a permutation of $(1/2n, 3/2n, \dots, 2n-1/2n)'$ and each X_s subset of k columns creates a k -factor uniform design with n runs. For instance, consider a two-factor uniform design with 8 runs in $C^2 = [0, 1]^2$, where the levels of the two factors have been scaled to the unit interval. Given $n = 2^3$ then $p_1^3 = 2^3$ is the prime number decomposition of n . The

Euler function $\phi(8) = 8(1 - \frac{1}{2}) = 4$. That is, there are 4 integers less than 8 that are relatively prime with 8 (or have a $gcd = 1$). The resulting candidate generating vector $h_n = (1, 3, 5, 7)$ for $s = 4$ and the results of modular arithmetic (mod 8) would be

$$Q_4 = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 6 & 2 & 6 \\ 3 & 1 & 7 & 5 \\ 4 & 4 & 4 & 4 \\ 5 & 7 & 1 & 3 \\ 6 & 2 & 6 & 2 \\ 7 & 5 & 3 & 1 \\ 8 & 8 & 8 & 8 \end{bmatrix} \quad X_4 = \begin{bmatrix} 1/16 & 5/16 & 9/16 & 13/16 \\ 3/16 & 11/16 & 3/16 & 11/16 \\ 5/16 & 1/16 & 13/16 & 9/16 \\ 7/16 & 7/16 & 7/16 & 7/16 \\ 9/16 & 13/16 & 1/16 & 5/16 \\ 11/16 & 3/16 & 11/16 & 3/16 \\ 13/16 & 9/16 & 5/16 & 1/16 \\ 15/16 & 15/16 & 15/16 & 15/16 \end{bmatrix}.$$

Fang and Wang (1994) showed that the first column of X_s should always be selected due to the equivalence of generator subsets to reduce the number of uniform designs to be generated. Thus, any matrix formed from column 1 and one of X_4 's three other columns would be a two-factor uniform design. The points in the three such possible 2-factor uniform designs are shown in Figure 1.10. Generators $h = (1, 3)$ and $h = (1, 5)$ appear to produce better uniformly scattered design points than $h = (1, 7)$. Different uniformity measures will be addressed in a later section to determine which of these uniform designs has the “best” space-filling properties.

For a prime p , p -run experiments with a relatively large number of experimental factors, Fang and Wang (1994) and Fang et al. (2006) suggested the power generator method using primitive roots to reduce the large number of potential generating vectors. They proposed using the vector $h_p = (1, a, a^2, \dots, a^{s-1})(mod(p))$ to generate the design matrices, where a is the primitive root of p and $a^i \neq a^j$, $1 \leq i < j <$

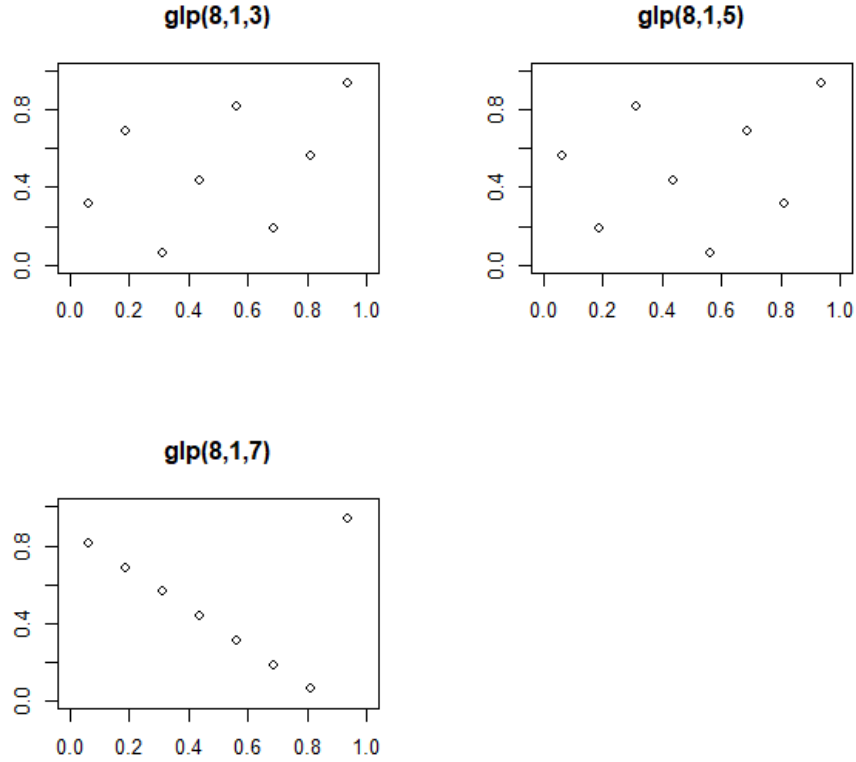


Figure 1.10: 8-point uniform designs in C^2 using GLP method.

p . Talke (2012) found that it would be possible to exclude generating vectors that produce "best" uniform designs by limiting the search to designs generated by this method. Fang and Wang (1994) recommended using the GLP method for $s \leq 10$. Fang and Lin (2003), Fang and Wang (1994), Fang et al. (2006), Li et al. (2004), and Zhang et al. (1998) have extensive details about constructing uniform designs using the good lattice method.

1.3.2.2 Good Point Method (Fang and Wang, 1994) The good point method introduced by Fang and Wang (1994) and Keng and Yuan (1983) can be applied to generate a set of space-filling design points. A set of the form $\{\lfloor k\lambda_1 \rfloor, \dots, \lfloor k\lambda_s \rfloor\}$ for $k = 1, \dots, n$ is a set of n design points in C^s obtained through the good point (gp)

method. $\lambda = (\lambda_1, \dots, \lambda_s)$ is the generating vector and $\lfloor \cdot \rfloor$ is the function that returns the decimal part of $k\lambda_i$ for $i = 1, \dots, s$. There are three common good point methods: cyclotomic field, square root sequence and prime root. The main difference between these methods comes from the function that returns the decimal part of $k\lambda_i$.

For the cyclotomic field (CF) method, the set is of the form

$$\left\{ \left\lfloor k2\cos\left(\frac{2\pi}{p}\right) \right\rfloor, \dots, \left\lfloor k2\cos\left(\frac{2\pi s}{p}\right) \right\rfloor \right\},$$

where p is a prime number such that $n \geq p \geq 2s + 3$. A generating vector $\lambda = (2\cos(\frac{2\pi}{p}), \dots, 2\cos(\frac{2\pi s}{p}))$ that generates good space-filling design points is first required. For example, for a two-factor uniform design with 8 runs, the only candidate prime is $p = 2s + 3 = 7$ with a generating vector $\lambda = (0.2470, 0.4450)$. The design matrix and a plot of the design points are shown in Figure 1.11 .

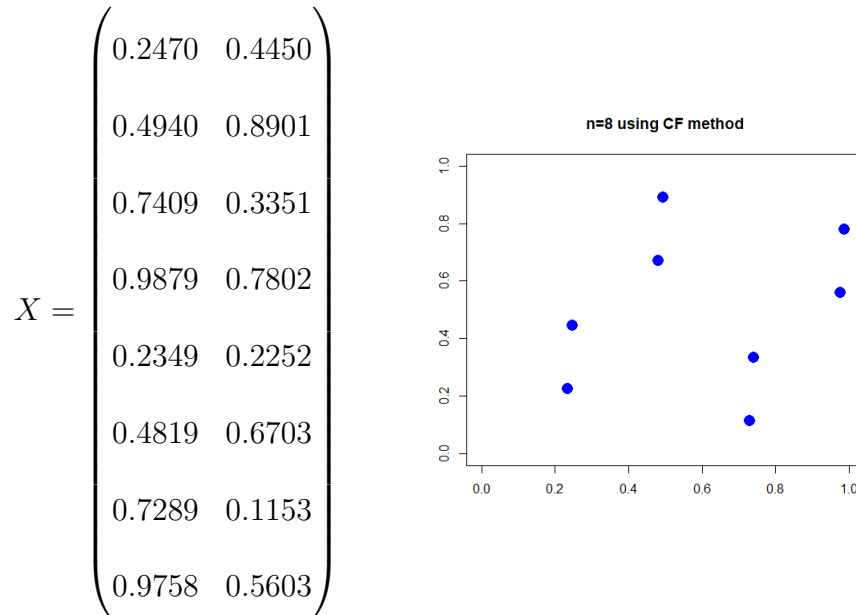


Figure 1.11: 8-point design using CF method.

The best 8-point designs by the GLP method seem to have better space-filling

properties than the one produced by the CF method. Talke (2012) found that the best designs with the GLP method generally have better space-filling properties than those obtained with the gp methods. The CF method is typically used when the number of factors is large (Fang and Lin, 2003; Fang and Wang, 1994; Fang et al., 2006).

The good point set by the square root sequence (SRS) method is formed by taking s primes $p_1, p_2, \dots, p_s$ with $\lambda = (\sqrt{p_1}, \dots, \sqrt{p_s})$, and the design set is of the form $\{\lfloor k\sqrt{p_1} \rfloor, \dots, \lfloor k\sqrt{p_s} \rfloor\}$, $k = 1, \dots, n$. For example, for $n = 8$ and $s = 2$, the candidate primes can be $p = 2, 3, 5, 7$ such that $p_i \leq n = 8$ to limit the computational effort. The design matrix X formed by using the first two primes ($p_1 = 2, p_2 = 3$) and plots of some designs generated from using six prime generators from the set of candidate primes are shown in Figure 1.12.

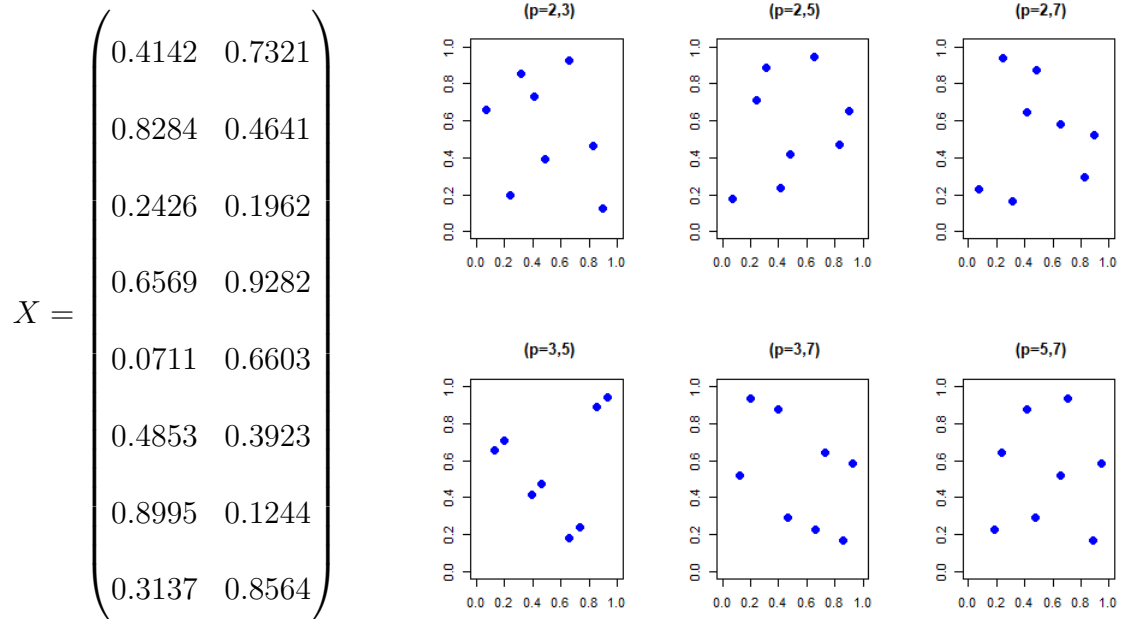


Figure 1.12: 8-point designs using SRS method.

The 8-point best designs by the SRS method appear to be more uniformly

scattered than the one generated using the CF method. The SRS method generally produces best uniform designs for $s \leq 10$ than the other two good point methods (Fang and Lin, 2003; Fang and Wang, 1994; Fang et al., 2006; Zhang et al., 1998).

The last good point method is the powers-of-a-prime (PP) method. A set of n points is generated using the PP method by taking $\lambda = (q, q^2, \dots, q^s)$, where $q = p^{1/s+1}$ for any prime p . To limit computational burden, p can be taken from the set of primes less than n . For example, for a two-factor design with $n = 8$ runs, the candidate primes may be $p = 2, 3, 5, 7$. The design matrix X using $p = 2$ and candidate designs using $p = 2, 3, 5, 7$ are shown in Figure 1.13. The design generated using $p = 2$ appears to have more uniformly scattered points than the other primes.

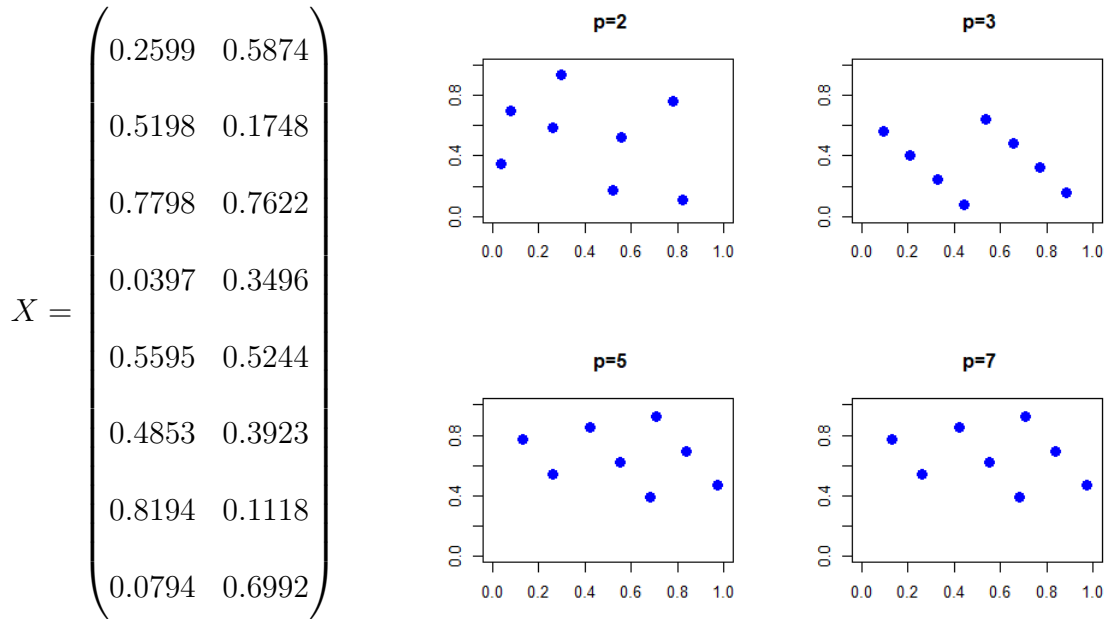


Figure 1.13: 8-point designs using PP method.

1.3.2.3 Halton (H) Set Method (Fang and Wang, 1994) Another method used to construct uniformly scattered set of points is the Halton (H) method. The

Halton set is based on a deterministic method that uses co-prime numbers as its foundation to generate points in the interval $[0, 1]$ (Koblitz, 1984). For example, when we have $n = 8$ runs and $s = 2$ factors, the first two prime numbers ($p_1 = 2, p_2 = 3$) can be used to generate the H-set points. To generate the set based on $p_1 = 2$, we start by dividing the interval $[0, 1]$ in half, then in fourths, eighths, etc., which generates $(1/2, 1/4, 3/4, 1/8, 5/8, 3/8, 7/8, 1/16)$. For $p_2 = 3$, We then divide the interval $[0, 1]$ into thirds, then ninths, twenty-sevenths, etc. to generate $(1/3, 2/3, 1/9, 4/9, 7/9, 2/9, 5/9, 8/9)$. Pairing them up yields the following 8-point H-set design X (Figure 1.14). This can be generalized to other primes and any number of experimental runs.

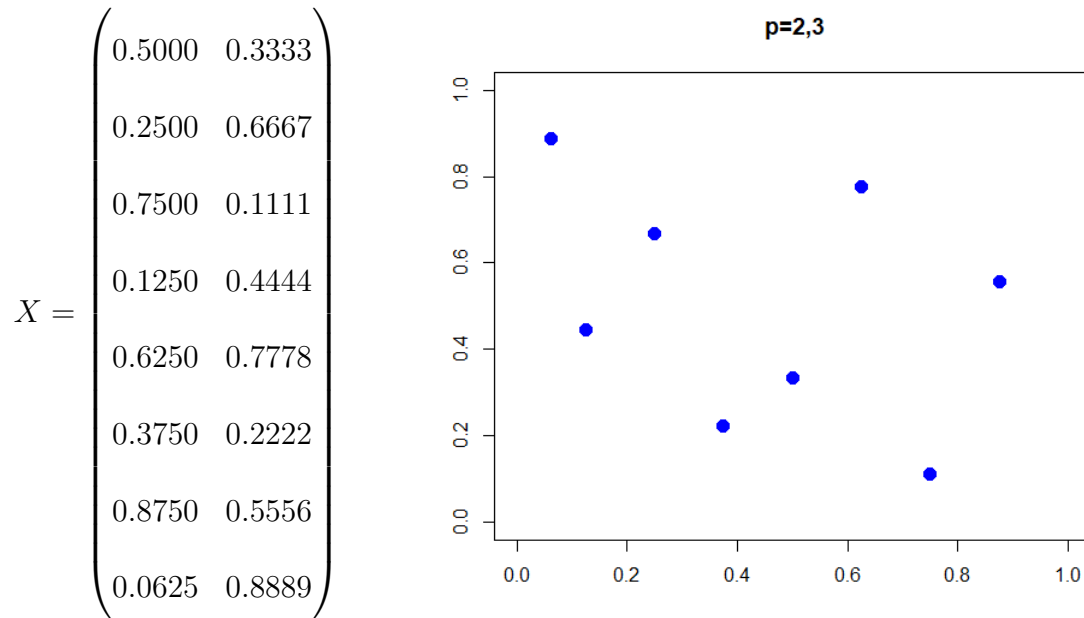


Figure 1.14: 8-point design using H-set method

Fang and Wang (1994) discuss the mathematics behind this method in detail as follows:

- Any positive integer k can be expressed in a unique manner using a prime base

p as

$$k = b_0 + b_1p + b_2p^2 + \dots + b_mp^m, \quad 0 \leq b_i \leq p-1, \quad p^m \leq k \leq p^{m+1}.$$

- In the interval $[0, 1]$, any rational number c can be represented uniquely as

$$c = c_0p^{-1} + c_1p^{-2} + \dots + c_mp^{-m-1}, \quad 0 \leq c_i \leq p-1.$$

- There is a 1-1 correspondence between the set of rational numbers in $[0, 1]$ and the set of positive integers. This relationship enables us to establish that, for any integer k ,

$$y_p(k) = b_0p^{-1} + b_1p^{-2} + \dots + b_mp^{-m-1}.$$

The function $y_p(k)$ is called the radical inverse of k base p and the H-set is the set of points x_k such that

$$x_k = (y_{p_1}(k), \dots, y_{p_s}(k)), \quad k = 1, \dots, n.$$

When the number of primes and experimental factors s are large, the computation burden is high. Fang and Wang (1994) recommend using the H-set when s is small. Fang (1980) and Fang and Wang (1994) have extensive details on the Halton method for generating uniformly scattered set of points.

1.3.2.4 Hammersley Method (Fang and Wang, 1994) The Hammersley set is given by

$$x_k = \left(\frac{2k-1}{2n}, y_{p_1}(k), \dots, y_{p_{s-1}}(k) \right) \quad k = 1, \dots, n.$$

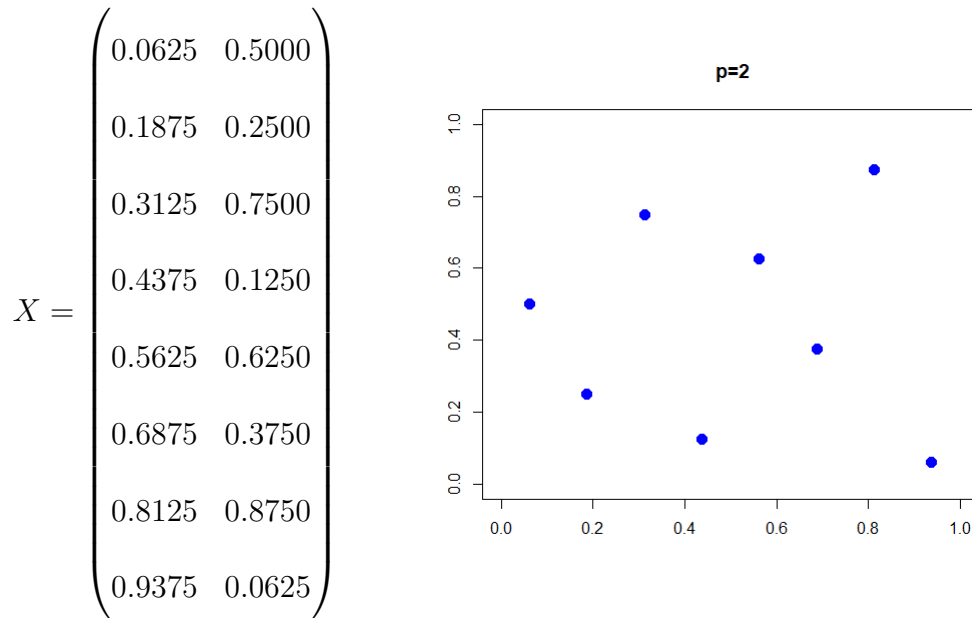


Figure 1.15: 8-point design using Hammersley method

This is just a modification of the Halton method. That is, $s - 1$ primes are used to generate a Hammersley set instead of s number of primes used by the Halton method. For example, for $n = 8$ runs and $s = 2$ experimental factors, and using $p_1 = 2$, the Hammersley set design in Figure 1.15 shows that the design points generated by the Hammersley method are scattered more uniformly than those generated by the Halton method.

1.3.3 Measures of Uniformity

Uniform design methods are aimed at producing sets of design points that are distributed evenly across the experimental region. However, these techniques alone do not provide a quantifiable measure of uniformity and do not yield outcomes of equal uniformity. Various uniformity measures have been introduced by various authors. The two primary ones are measures of discrepancy and distance metric methods.

Conceptually, discrepancy measures how far a given distribution deviates from

an ideal one (Shirley, 1991). It measures how well a set of n points, $\rho = (x_1, \dots, x_n)$ represent some multivariate distribution $F(x)$ in R^n . For instance, if the given distribution is a uniform distribution, then discrepancy is the measure of how the empirical distribution of a point set deviates from that uniform distribution. The empirical distribution of the point set is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_i(x_i \leq x), \quad (1.10)$$

where $I_i(x_i \leq x) = 1$ if $x_i \leq x$ and 0 otherwise. Fang and Wang (1994) define discrepancy or F-discrepancy to be

$$D_F(n) = \sup_x |F_n(x) - F(x)|. \quad (1.11)$$

For example, the discrepancy of 8 points (1.41, 0.26, 1.97, 0.33, 0.55, 0.77, 1.46, 1.18) from uniform (0, 2) distribution can be calculated as shown in Table 1.3. The

Table 1.3: Example of discrepancy from uniform (0, 2) distribution,

x	0.26	0.33	0.55	0.77	1.18	1.41	1.46	1.97
$F_n(x)$	0.125	0.250	0.375	0.500	0.625	0.750	0.875	1
$F(x)$	0.130	0.165	0.275	0.385	0.590	0.705	0.730	0.985
$ F_n(x) - F(x) $	0.005	0.085	0.100	0.115	0.035	0.045	0.145	0.015

discrepancy is $D_F(n) = 0.145$, which is the maximum of the absolute difference between the empirical distribution of these 8 points and the given uniform (0, 2) distribution.

In terms of space-filling properties, the discrepancy measure for a set of n points

$\rho = \{x_1, \dots, x_n\}$ is

$$D(n, \rho) = \sup \left| \frac{N(\gamma, \rho)}{n} - v([0, \gamma]) \right|,$$

where the vector $\gamma = (\gamma_1, \dots, \gamma_s) \in C^s$, $N(\gamma, \rho)$ is the number of points that satisfy $x_k \leq \gamma$ for $k = 1, \dots, n$, and $v([0, \gamma]) = \prod_{i=1}^s \gamma_i$ is the volume of the hyper-rectangle formed by γ and the origin. The star discrepancy $D^*(n, \rho)$, first used by Fang and Wang (1994) in measuring uniformity expands discrepancy to more general rectangular regions. The star discrepancy $D^*(n, \rho)$ is defined as

$$D^*(n, \rho) = \sup \left| \frac{N([a, b], \rho)}{n} - v([a, b]) \right|$$

for the set of hyper-rectangles $[a, b] \in C^s$, $0 \leq a \leq b \leq 1$ and $v([a, b]) = \prod_{i=1}^s (b_i - a_i)$ is the volume of $[a, b]$. That is, $D^*(n, \rho)$ represents the supremum of the absolute difference over all possible hyper-rectangles $[a, b]$ (Talke, 2012). Fang and Wang (1994) assert that a small $D(n, \rho)$ or $D^*(n, \rho)$ is an indication that the proportion of points within each hyper-rectangle is nearly proportional to the volume of the hyper-rectangle signaling a good space-filling design.

The geometric illustration of star discrepancy is shown in Figure 1.16a. The experimental region C^2 has $n = 24$ design points, and the internal rectangle $\gamma = (0.5, 0.75)$ has 10 design points, hence $N(\gamma, \rho)/n = 10/24 = 0.4167$ and $v([0, \gamma]) = 0.3750$. Therefore, $|\frac{N(\gamma, \rho)}{n} - v([0, \gamma])| = |0.4167 - 0.3750| = 0.0417$ is the discrepancy for this $\gamma = (0.5, 0.75)$. The discrepancy would be the maximum of absolute deviations over all $\gamma \in C^2$.

Talke (2012) proposed a modified star discrepancy measure where the experimental region C^s is partitioned into 2^s hyper-rectangles formed by $[a, b]$ and the 2^s

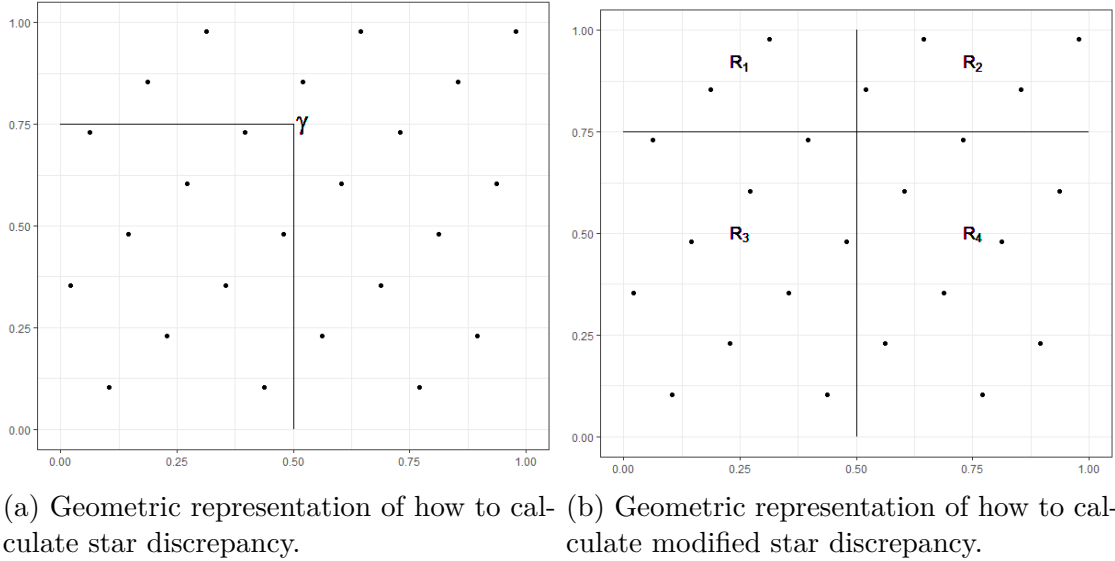


Figure 1.16: Geometric illustration of how to calculate discrepancy.

vertices of C^s . The proposed modified star discrepancy (MSTRD) is defined as

$$MSTRD(n, \rho) = \sup \left(\max \left| \frac{N([a, b], \rho, R_j)}{n} - v(R_j) \right| \right),$$

where R_j is the j th hyper-rectangle formed by $[a, b)$ and the vertex v_j of C^s for $j = 1, \dots, 2^s$. $N([a, b], \rho, R_j)$ is the number of design points in R_j , and $v(R_j)$ is the volume of R_j . Thus, MSTRD is the maximum of the 2^s discrepancies associated with the set of 2^s hyper-rectangles.

Figure 1.16b illustrates the calculation of MSTRD for a two factor ($s = 2$) 24-point design in C^2 . For any given $[a, b)$, C^2 is partitioned into 4 rectangles. In this case, the discrepancies for the 4 rectangles is $|\pm 1/24| = 0.0417$ and consequently, the MSTRD for this $[a, b) = (0.5, 0.75)$ is 0.0417. MSTRD is a better measure of uniformity than the original star discrepancy because it incorporates more subspaces of the experimental region C^s in its calculation. For extensive details on discrepancy measures and other forms of discrepancy, see Fang and Wang (1994), Fang et al.

(2006), Hickernell (1998), Morokoff and Caflisch (1994), and Zhang et al. (1998).

Distance-based metrics are usually more desirable measures of uniformity because discrepancy measures are difficult to compute as the number of experimental factors or dimension of the design space increases. Johnson et al. (1990) introduced minimax and maximin principles that do not depend on the geometry of the experimental region (Trosset, 1999). The minimax and maximin criteria aim at minimizing maximum distances between design points and/or maximizing minimum distances between design points. As space-filling measures, Borkowski and Piepel (2009) presented the following two distance-based criteria and an alternate version of the maximin criterion.

Assume $X'_1, \dots, X'_n$ forms the rows of an $n \times s$ design matrix X . Borkowski and Piepel (2009) define the distance between any point w in the region C^s and the design matrix X as

$$d(w, X) = \min_j \sqrt[2]{(w - x_j)(w - x_j)^T},$$

for $j = 1, 2, \dots, n$. In other words, n distances are calculated between a point w in C^s and each of the n points in X and the minimum of the n distances is kept. Specifically, $d(w, X)$ is the distance of a point w in experimental region C^s from the closest design point in the design matrix X . Borkowski and Piepel (2009) used the following criteria based on $d(w, X)$ to assess the space-filling properties of a design:

- The maximum distance ($MD(X)$), which is defined as $MD(X) = \max_w(d(w, X))$, the maximum of $d(w, X)$ over X in the design space.
- The root mean square distance ($RMSD(X)$), defined as $RMSD = \sqrt[2]{E[d(w, X)^2]}$. This is the square root of the expected squared distance between a point w in C^s and the design point of X closest to w .

- The average distance ($AD(X)$), defined as $AD(X) = E[d(w, X)]$, which is the expected value of the distance $d(w, X)$.

For $RMSD(X)$ and $AD(X)$, the expectation assumes X is uniformly distributed over the design space. The exact $RMSD(X)$ and $AD(X)$ are based on expectations that have no closed form and are hard to calculate as they involve integration across the experimental region C^s . Borkowski and Piepel (2009) recommend the following approximations based on a large random sample of n points $w_1, w_2, \dots, w_n$ in C^s :

- $md(X) = \max(d(w_i, X))$,
- $rmsd(X) = \sqrt[2]{\frac{\sum_{i=1}^n (d(w_i, X))^2}{n}}$, and
- $ad(X) = \frac{\sum_{i=1}^n |d(w_i, X)|}{n}$,

for $i = 1, 2, \dots, n$.

Small values of $md(X)$, $rmsd(X)$, and $ad(X)$ indicate that a design X has good space-filling properties because each of the points in C^s would be close to at least one design point from X and no subspace is denser than the others in the design region (Borkowski and Piepel, 2009).

Importantly, discrepancy and distance metric measures cannot be used individually to assess all of the space-filling properties of a design (Franco et al., 2009). Essentially, discrepancy assesses the non-uniformity of a design, but we do not just want to measure non-uniformity, but also evaluate how evenly the design points are spread throughout the region (Damblin et al., 2013; Franco et al., 2009). For example, discrepancies were calculated (Table 1.4) for three designs: random sampling, square root sequence, and factorial designs in C^2 . The results indicated that designs generated by random sampling and square root sequence techniques were “better” uniform designs than the factorial designs, yet design points for factorial

designs were spread more evenly than design points for random sampling and square root sequence designs (Figure 1.17). On the other hand, the “best” uniformity values were obtained by factorial designs using the maximum distance (md) criterion (Table 1.4). Franco et al. (2009) has shown that, in general, distance-based criteria tend to favor designs with regular grid-like structures even if the design points of those designs do not adequately fill up the design space.

In other words, discrepancy and distance-based criteria should be used specifically for comparing designs and not for validating designs (Damblin et al., 2013; Franco et al., 2009; Gunzburger and Burkardt, 2004). In chapter Two, a criterion that takes into account filling a design space and the distribution of design points to qualify designs will be introduced.

Table 1.4: Star discrepancy ($D^*(n, \rho)$) and maximum distance (md) criteria for random sampling, square root sequence, and factorial designs in c^2 .

Random Sampling			Square Root Sequence (SRS)		Factorial Design	
	$D^*(36, \rho)$	md	$D^*(36, \rho)$	md	$D^*(36, \rho)$	md
$n = 16$	0.326	0.431	0.170	0.303	0.389	0.234
$n = 25$	0.274	0.259	0.108	0.229	0.305	0.176
$n = 36$	0.223	0.227	0.084	0.212	0.250	0.140

1.3.4 Objectives of the Study

Existing research has explored space-filling designs for process variable experiments in hypercuboidal and hyperspherical design regions, and for mixture experiments in the simplex or constrained subspaces of the simplex. To date, no research has been published for space-filling designs for the combined mixture/process variable experiment.

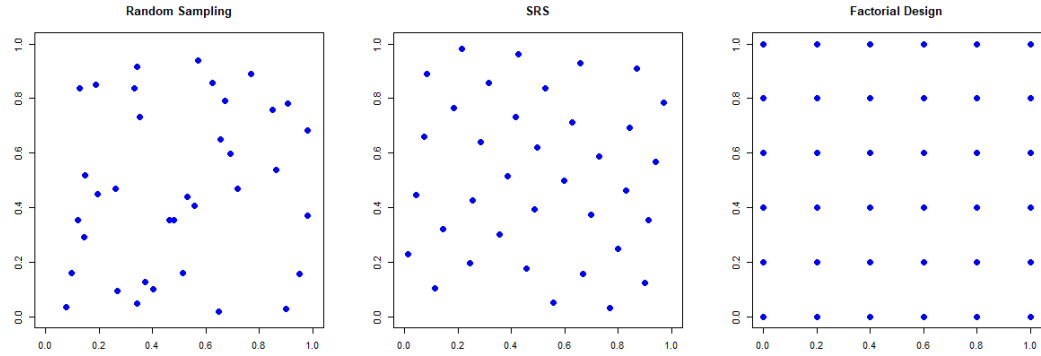


Figure 1.17: From left to right: Random sampling, SRS, and Factor designs with $n = 36$ in C^2 .

Development of the statistical methodology and design generating algorithms are the primary research goals of this dissertation. Specifically, the objectives of this study are to:

1. Explore and implement existing algorithms for the generation of space-filling designs in the simplex and subspaces of the simplex. This will necessarily include consideration of distance-based criteria for assessing the uniformity of scatter.
2. Propose a hybrid algorithm for constructing uniform designs for a mixture/process variable experiment.
3. Develop new measures for assessing uniformity of scatter in the full simplex, sub-space of the simplex, and the combined mixture/process variable design regions.
4. Provide detailed tables of good uniform mixture designs based on the new measures in S^q for $q = 3, 4, 5$ mixture components designs of size $9 \leq n \leq 30$.
5. Provide detailed tables of good uniform designs based on the new measures for the combined q -mixture and k -process variable experiments for $q = 3, 4, 5$,

$k = 1, 2$ and design sizes, $9 \leq n \leq 30$.

6. Present an algorithm for design augmentation in the simplex.

The first three objectives of this dissertation are presented in Chapter Two. Objectives four and five are be found in APPENDIX A, and Chapter Four discusses objective six.

CHAPTER TWO

UNIFORM DESIGNS FOR EXPERIMENTS WITH MIXTURES

Scheffe type and optimal computer generated designs are popular designs for experiments with mixtures (Box and Draper, 2007; Coetzer and Haines, 2017; Cornell, 2011; Fang and Wang, 1994; Goos et al., 2016; Limmun et al., 2019; Smith, 2005). Unfortunately, these model-dependent designs also place most design points on or near the boundary of the experiment regions, leaving the interior largely unexplored (Borkowski and Piepel, 2009; Fang and Wang, 1994; Talke, 2012). Maximin or minimax distance designs (Johnson et al., 1990; Trosset, 1999) have good space-filling properties; however, they do not have good projection properties, as a large proportion of design points are placed on or near the boundaries of moderate to higher dimensional experimental regions (Joseph et al., 2015; Piepel et al., 1993). Latin hypercube sampling designs and randomized orthogonal arrays (Bingham et al., 2009; Butler, 2001; Husslage et al., 2011; Owen, 1992; Park, 1994) may be used to overcome the bad projection properties of maximin or minimax designs in higher dimensions. However, the projection properties of Latin hypercube sampling designs and randomized orthogonal arrays in lower dimensions have not been established yet, and they have also not been extended to simplex experimental regions or highly constrained irregularly shaped regions like sub-spaces of the simplex (Borkowski and Piepel, 2009; Joseph et al., 2015).

Fast Flexible Filling designs can be used for irregularly shaped experimental regions formed by constraints on mixture components. The design points of Fast Flexible Filling designs are obtained by an optimality criterion to select from clusters formed by a clustering algorithm after generating large a number of random points

within the design region (Joseph et al., 2015; Lekivetz and Jones, 2015). The efficiency of Fast Flexible Filling designs depends on the clustering algorithm used, and, generally, Monte Carlo-based space-filling designs are less efficient than uniform designs, which are based on quasi Monte Carlo methods (Fang and Wang, 1994).

2.1 Uniform Designs for Constrained and Unconstrained Mixture Experiments

Many authors have used numerous methods to distribute design points uniformly throughout the simplex. For example, Fang and Wang (1994) used a Scheffe type design and contracted the design points to the inner part of the simplex, Piepel et al. (1993) proposed a layered design approach for mixture regions with irregular shapes, Gomes et al. (2018) summarized various selection-based Monte Carlo algorithms, Cooley and Piepel (2003) proposed two-stage layered mixture designs for experiments with a large number of mixture components, Liu and Liu (2016) proposed new discrepancy-based methods, and Linse (2008) also used a mesh approach to construct space-filling mixture designs. However, most of these methods focus on constructing design points for the full simplex space. In this study, the G -mapping function introduced by Fang and Yang (2000) and adopted by Borkowski and Piepel (2009) will be used.

The G function uses a conditional distribution approach and can be used to construct design points in the full simplex (S^q) and sub-spaces of the simplex ($S_{A,B}^q$). The method and steps proposed by Borkowski and Piepel (2009) for constructing designs for q -component mixture experiments are described as follows:

1. To generate n design points in the full simplex (S^q) or sub-space of the simplex ($S_{A,B}^q$), first generate n points in the hypercube C^{q-1} using a number theoretic method. Minimum and maximum constraints may be placed on each mixture

component, $A_i \leq y_i \leq B_i$ such that $0 \leq A_i \leq B_i \leq 1$ for $i = 1, 2, \dots, s$. Some of the constraints may be redundant. Hence, to get rid of such redundancies, $A_i \leq y_i \leq B_i$ are replaced with $A_{ci} \leq y_i \leq B_{ci}$, where $A_{ci} = \max(A_i, B_i + 1 - B)$ and $B_{ci} = \min(B_i, A_i + 1 - A)$ (Cornell, 2011), where A and B are the sum of all minimum and maximum constraints respectively, placed on the mixture components.

2. The G function is then used to map the points $x_1, \dots, x_{q-1}$ from C^{q-1} to the design points $y_1, \dots, y_q$ in the full simplex or sub-space of the simplex. The G function is defined as:

$$y_k = G(x_{k-1}, d_k, \phi_k, \Delta_k, k-1) = \Delta_k (1 - [x_{k-1}(1 - \phi_k)^{q-1} + (1 - x_{k-1})(1 - d_k)^{q-1}]^{1/q-1}), \quad (2.1)$$

for $k = 2, \dots, q$, and $y_1 = 1 - (y_2 + y_3 + \dots + y_q)$. The following quantities in the G function are calculated as follows:

$$\Delta_q = 1,$$

$$\Delta_k = 1 - (x_k + \dots + x_{q-1}),$$

$$d_k = \max\left(\frac{A_{ck}}{\Delta_k}, 1 - \frac{B_{c1} + B_{c2} + \dots + B_{c(k-1)}}{\Delta_k}\right),$$

and

$$\phi_k = \max\left(\frac{B_{ck}}{\Delta_k}, 1 - \frac{A_{c1} + A_{c2} + \dots + A_{c(k-1)}}{\Delta_k}\right),$$

for $k = 2, 3, \dots, q$.

The following shows how the G function can be applied to generate 21-point designs of a 3-component mixture experiment in the full simplex (S^3) or the

constrained simplex ($S_{A,B}^3$), with restrictions $0.1 \leq x_1 \leq 0.75$, $0.15 \leq x_2 \leq 0.8$, and $0.1 \leq x_3 \leq 0.6$) shown in Figure 2.2.

1. Generate 21 design points in C^2 using any number theoretic method. Figure 2.1 shows four different 21-point designs generated using the good lattice point (GLP), square root sequence (SRS), power-of-prime (PP) and cyclotomic field (CF) methods.

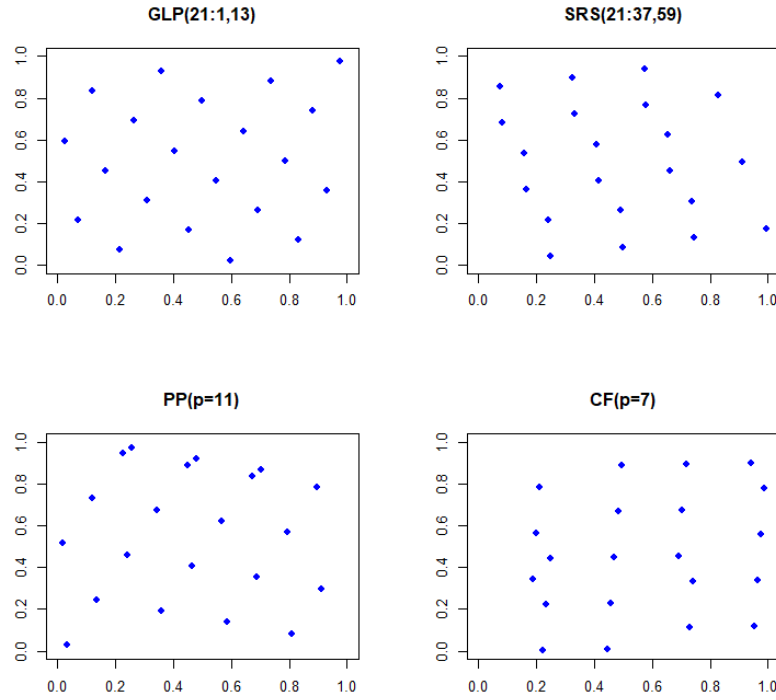


Figure 2.1: 21-point designs in C^2 using four number theoretic methods.

2. Then $\Delta_3 = 1$, $d_3 = \max(\frac{A_{c3}}{\Delta_3}, 1 - \frac{B_{c1}+B_{c2}}{\Delta_3})$, and $\phi_3 = \min(\frac{B_{c3}}{\Delta_3}, 1 - \frac{A_{c1}+A_{c2}}{\Delta_3})$ are calculated and substituted into the G function to find

$$y_3 = G(x_2, d_3, \phi_3, \Delta_3, 2) = \Delta_3(1 - [x_2(1 - \phi_3)^2 + (1 - x_2)(1 - d_3)^2]^{1/2},$$

y_2 is then generated given that y_3 is known by calculating

$$\Delta_2 = \Delta_3 - y_3, \quad d_2 = \max(\frac{\Delta_2}{B_{c2}}, 1 - \frac{\Delta_2}{B_{c1}}), \quad \text{and } \phi_2 = \min(\frac{\Delta_2}{B_{c2}}, 1 - \frac{\Delta_2}{A_{c1}}). \quad \text{The } G$$

function is then applied to find

$$y_2 = G(x_1, d_2, \phi_2, \Delta_2, 1) = \Delta_2(1 - [x_1(1 - \phi_2) + (1 - x_1)(1 - d_2)]),$$

$y_1 = 1 - (y_2 + y_3)$ is then found using the fact that sum of the components must

be 1.

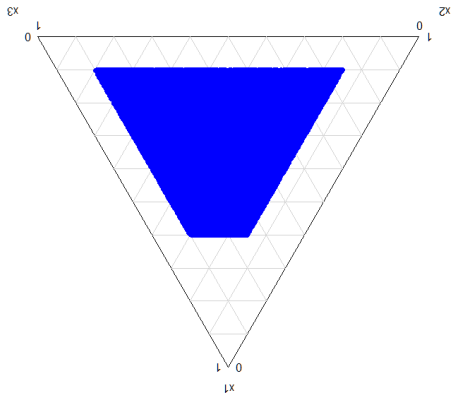


Figure 2.2: Example of a restricted design region from placing lower and upper bounds on a 3-component mixture.

The generated 21 design points in the full simplex (S^3) and the restricted simplex ($S_{A,B}^3$) using the four number theoretic methods are shown in Figure 2.3. The plots do show that design points generated by the GLP method are graphically more uniformly scattered across the two design regions (full simplex (S^3), constrained simplex ($S_{A,B}^3$)) than the other three number theoretic methods. Figure 2.3 also indicates that in the full simplex (S^3) or the restricted simplex ($S_{A,B}^3$), the straight-line dispersion of design points in the unit square (C^2) becomes curvilinear, even though the G function generally maintains the structure of the design patterns from C^2 . However, there is no

preservation of the systematic pattern of uniformity across the two design regions due to the nonlinear nature of the G transformation. For instance, design points obtained from the powers of prime (PP) method appear to be more uniformly scattered in C^2 than those obtained from the cyclotomic fields (CF) method, but the mapped design points obtained using the PP method in the full simplex and the constrained simplex regions seem to be less uniformly scattered than those of the CF method. That is, the best space-filling design in C^2 may not necessarily be the best space-filling design in S^3 or $S_{A,B}^3$ and at the same time, the best space-filling design in S^3 may not actually be the best space-filling design in $S_{A,B}^3$. Consequently, Borkowski and Piepel (2009) advised that space-filling properties of a design should only be assessed after the transformation. Borkowski and Piepel (2009) and Talke (2012) presented extensive details and explanation of these results. Details of the G function transformation and its derivation can be found in Fang and Yang (2000).

2.2 New Proposed Uniform Designs for Mixture/Process Variable Experiments

In practice, the blending properties of the response output of most industrial mixture experiments depend on process conditions. Cornell (1971) first introduced graphical displays of designs for mixture experiments with process variables. The configuration involves arranging combinations of all experimental mixture proportions at certain levels of the process variables or arranging the process variable combinations at each experimental mixture. Goos and Donev (2007) proposed split plot designs for mixture/process variable experiments with process variables as whole plot factors and experimental mixtures as the sub-plot factors. Other authors have presented various optimal designs for mixture experiments, with process variables (Anderson-Cook et al., 2004; Chung et al., 2007; Czitrom, 1988; De Ketelaere et al., 2011; Goldfarb et al., 2003) . However, these model-dependent designs also have bad space-

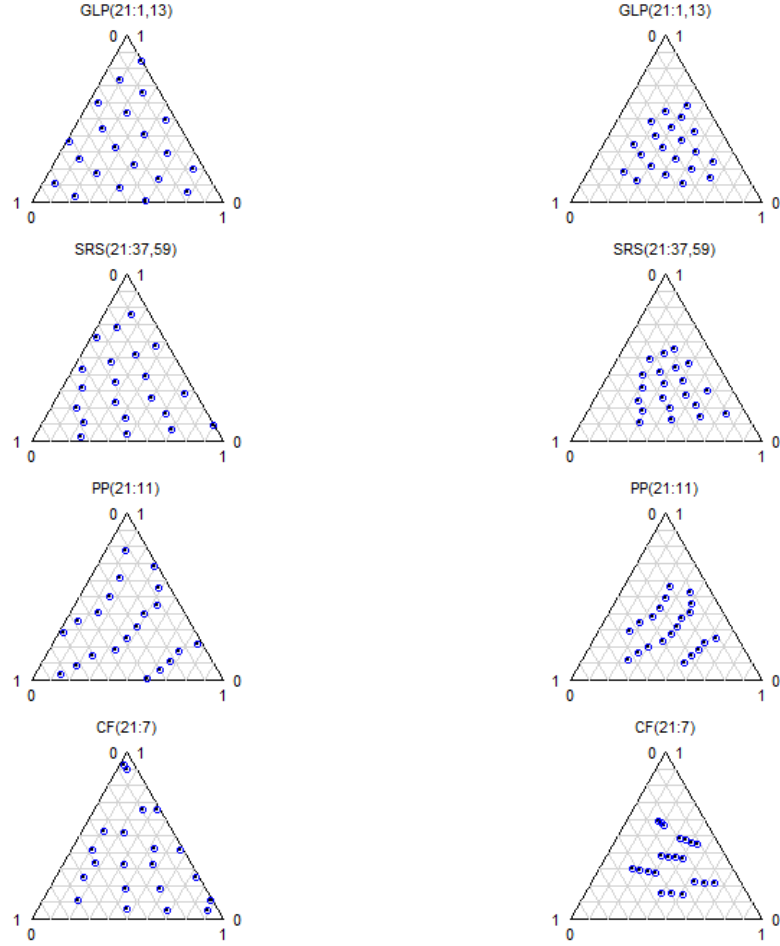


Figure 2.3: Illustration of the G function for $n = 21$ design points in the full simplex and the constrained simplex.

filling properties and require a large number of experimental runs. For instance, using the Cornell configuration, the combined unreplicated design of a full factorial run at each vertex of the simplex region for a 3-component mixture experiment with 3 process variables at 2 levels requires 24 experimental runs.

Existing research has explored space-filling designs for process variable experiments in hypercuboidal and hyperspherical design regions and for mixture experiments in the simplex or constrained subspaces of the simplex (Borkowski and Piepel, 2009; Cornell, 2011; Fang and Lin, 2003; Fang et al., 2006; Lekivetz and

Jones, 2015; Linse, 2008). To date, no research has been published for space-filling designs for the combined mixture/process variable experiment. In this study, a hybrid algorithm is proposed for generating uniform designs for the combined q -mixture and k -process variable experiments. The algorithm can be summarized in the following steps.

1. Generate n points $(x_1, \dots, x_{q-1}, x_q, \dots, x_{q+k-1})$ in the hypercube C^{q+k-1} using a number theoretic method.
2. For each of the n points:
 - Use the G function to map the first $q - 1$ coordinates $(x_1, \dots, x_{q-1})$ from C^{q+k-1} to the mixtures $(y_1, \dots, y_q)$ in S^q or subspace of S^q .
 - Leave the last k coordinates $(x_q, \dots, x_{q+k-1})$ unchanged to represent the coded levels of the k process variables. These will be relabeled $(z_1, \dots, z_k)$.
3. Combine each mixture $(y_1, \dots, y_q)$ with the process variable levels $(z_1, \dots, z_k)$ to form a point of the uniform design in the combined mixture and process variable design region.
 - For statistical application, convert the k process variable levels $(z_1, \dots, z_k)$ in each of the n design points to its original experimental scale (e.g., time, temperature,...).

This new algorithm was used to generate a 9-point design for an experiment conducted to investigate the effect of vanadium (V), molybdenum (Mo), and strontium (Sr) mixture ratios and process temperature on the electrical and ionic conductivity of a solid fuel cell. There were no restrictions on the 3 mixture components. Temperature (T) between $950^\circ C$ and $1250^\circ C$, and a 10 gram batch for

each mixture run was considered. The 9-point design (Table 2.1) for this experiment was generated as follows.

- **Step 1:** 9 points in C^3 are generated using the GLP method ($h = (1, 4, 7)$).
- **Steps 2 and 3:** The G function is then used to map (x_1, x_2) to the full simplex (S^3) and x_3 is used for the design point levels for the process variable.

Step 1			Steps 2 and 3			
x_1	x_2	x_3	y_1	y_2	y_3	z_1
0.0556	0.3889	0.7222	0.7383	0.0434	0.2183	0.7222
0.1667	0.8333	0.5000	0.3402	0.0680	0.5918	0.5000
0.2778	0.2778	0.2778	0.6138	0.2361	0.1502	0.2778
0.3889	0.7222	0.0556	0.3221	0.2050	0.4730	0.0556
0.5000	0.1667	0.8333	0.4564	0.4564	0.0871	0.8333
0.6111	0.6111	0.6111	0.2425	0.3811	0.3764	0.6111
0.7222	0.0556	0.3889	0.2700	0.7019	0.0282	0.3889
0.8333	0.5000	0.1667	0.1179	0.5893	0.2929	0.1667
0.9444	0.9444	0.9444	0.0131	0.2226	0.7643	0.9444

Table 2.1: Steps for generating a 9-point design for a 3-component mixture experiment with 1 process variable using the new hybrid algorithm

- For statistical application, the mixture and process variable design points are then converted to grams and degrees Celsius, respectively, shown in Table 2.2 .

Step 3			
V/g	Mo/g	Sr/g	$T/^{\circ}C$
7.383	0.434	2.183	1116.66
3.402	0.680	5.918	1100.00
6.138	2.361	1.502	1033.34
3.221	2.050	4.730	966.68
4.564	4.564	0.871	1200.00
2.425	3.811	3.764	1133.30
2.700	7.019	0.282	1066.67
1.179	5.893	2.929	1000.00
0.131	2.226	7.643	1233.32

Table 2.2: 9 mixture/process variable design points in grams and degrees Celsius.

In this example, the values of the settings of the process temperature are between 950°C and 1250°C . This algorithm can also be used when settings of the process variables are categories or a fixed set of levels. For instance, if we reconsider this example and suppose the process temperature is a categorical variable with 3 levels (950°C , 1100°C , 1250°C). Then, values of z_1 in Table 2.1 may be binned into 3 levels such that values 0.0556, 0.2778, 0.1667 are combined to give new level 1 (950°C), values 0.3889, 0.5000, 0.6111 are combined to give new level 2 (1100°C), and values 0.722, 0.8333, 0.9444 are combined to give level 3 (1250°C). The resulting design points in grams and Celsius are shown in Table 2.3.

Step 3			
V/g	Mo/g	Sr/g	$T/^{\circ}\text{C}$
7.383	0.434	2.183	1250
3.402	0.680	5.918	1100
6.138	2.361	1.502	950
3.221	2.050	4.730	950
4.564	4.564	0.871	1250
2.425	3.811	3.764	1100
2.700	7.019	0.282	1100
1.179	5.893	2.929	950
0.131	2.226	7.643	1250

Table 2.3: 9-point design for a 3-component mixture with temperature as a categorical process variable.

2.3 New Measures of Uniformity for Mixture/Process Experiments

Distance-based metrics are largely used to evaluate space-filling qualities of regularly-shaped and irregularly-shaped constrained mixture regions. Gomes et al. (2018) summarized various distance-based metrics commonly used. Borkowski and Piepel (2009), however, noted that these criteria should be calculated by first normalizing the components with respect to their range of variation ($B - A$), where A and B are lower and upper component constraints, in order to ensure

that each component of the mixture is treated equally in terms of importance. Furthermore, as discussed previously, Franco et al. (2009) outlined the drawback in using only distance-based criteria or discrepancy-based measures to assess the space-filling properties of designs. For this study, I will extend the Euclidean Minimum Spanning Tree (EMST) criterion introduced for assessing space-filling properties of designs (Damblin et al., 2013; Dussert et al., 1986; Franco et al., 2009) to unconstrained regularly-shaped and single component constrained (*SCC*) irregularly-shaped mixture/process regions.

2.3.1 The Euclidean Minimum Spanning Tree (EMST)

In graph theory, a graph consists of vertices or nodes connected by edges with a number called “weight” assigned to each edge (West, 2001). A spanning tree is a subset of a graph which connects all vertices but without a cycle, and a Minimum Spanning Tree is a spanning tree where the sum of the weights of the edges is less than or equal to all other spanning trees of the same graph (Graham and Hell, 1985; McDonald et al., 2005; Singh, 2009). In this dissertation, the weights assigned to the edges of a Euclidean Minimum Spanning Tree (EMST) are the edge lengths, which are the Euclidean distances between the vertices or design points.

Figure 2.4a is a connected graph with 3 vertices and 3 edges and has 3 possible spanning trees (Figures 2.4b-d). Because the sum of the weights of the spanning tree in Figure 2.4b is smaller than the sum of the weights of the other two spanning trees, it is the Minimum Spanning Tree (MST) for the graph in Figure 2.4a. Figure 2.4b is the Euclidean Minimum Spanning Tree (EMST) for this graph if the weights are the euclidean distances between the vertices.

Given a graph has n vertices, a spanning tree has $n - 1$ edges and the maximum number of spanning trees of a complete graph is n^{n-2} . There may be more than one

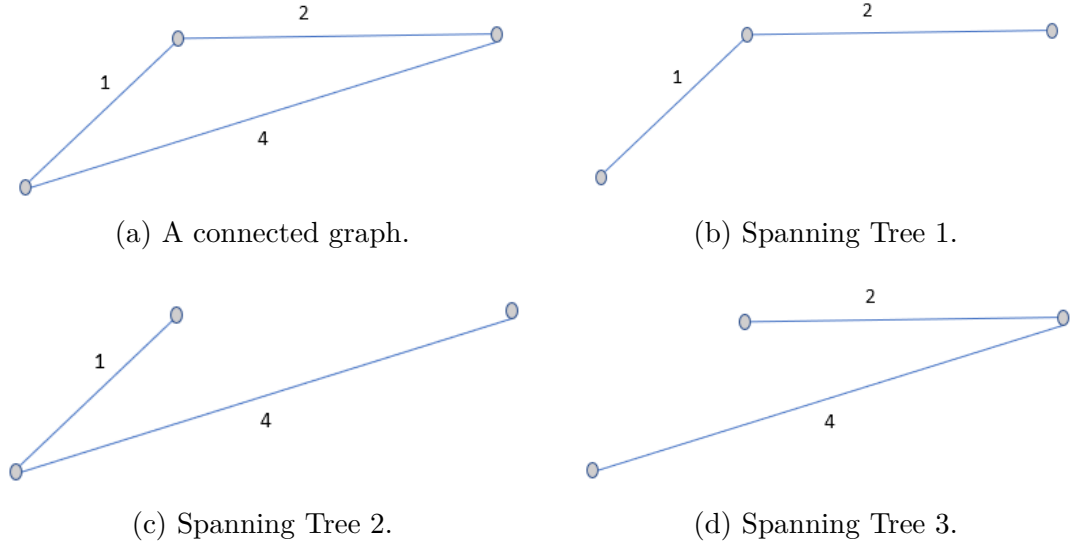


Figure 2.4: Spanning Trees.

EMST for a given graph or set of points, depending on the starting point, but the frequency distributions of edge lengths of all EMSTs will be the same (Dussert et al., 1986; Franco et al., 2009; Zahn, 1971).

Traditional algorithms, such as those by Kruskal (1956) and Prim (1957), allow MST or EMST to grow from a single vertex by adding the closest vertex to the current tree at each stage. The running time and efficiency of these conventional algorithms depend on the method of finding each component's nearest neighbor pair. In this study, a fast EMST algorithm introduced by March et al. (2010) will be used. This algorithm uses a dual-tree method to implement the Boruvka procedure (Nešetřil et al., 2001) to construct EMST. Dual-tree algorithms are the fastest known computational framework used to solve nearest neighbors problems (Gray and Moore, 2001) and an extensive overview of these methods can be found in Curtin et al. (2013) and Mou et al. (2019). For a graph with n vertices and m edges, the Boruvka minimum spanning tree algorithm takes at most $\log n$ steps and a total running time of $O(m \log n)$ to find the minimum-weight edge incident to each vertex of the

graph and add all of those edges together until every vertex can be reached. The construction of a Minimum Spanning Tree (MST) using the Boruvka algorithm is illustrated in the following steps.

1. Initially, every vertex is a component and weights are assigned to edges as shown in Figure 2.5. The MST is said to be empty at this point. Where EMST is concerned, the weights are the Euclidean distances between the vertices.

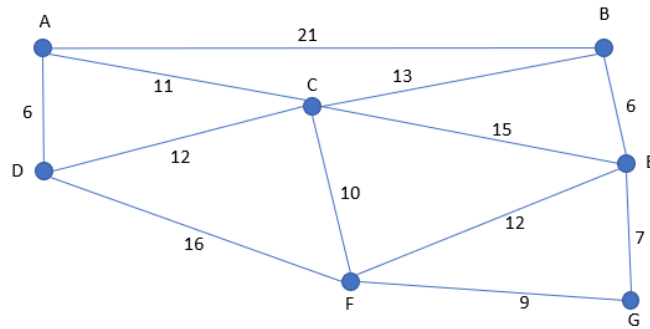


Figure 2.5: Weighted Graph.

2. The cheapest (i.e., shortest) outgoing edge is highlighted for every vertex in the first iteration of the algorithm. Figure 2.6 shows the cheapest edge out of every vertex highlighted with a yellow line. Two components or subset of vertices are formed from this first iteration (AD and $CFGE$). In some cases, the cheapest weighted edge incident on two vertices may be the same. Such edges are highlighted twice (AD , BE). The cheapest edges when constructing EMST are the shortest edges.
3. In the next iteration, the components formed from the first iteration are then connected with each other through the cheapest edges. This step is repeated until all vertices are included in the tree. For this example, and as shown in Figure 2.7, the two components (AD and $CFGE$) formed from the first

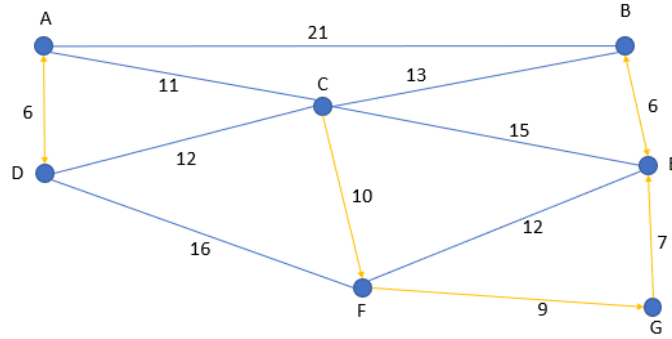


Figure 2.6: Cheapest edge out of every vertex highlighted with yellow line.

iteration are connected through edge AC . Every vertex in the graph can be reached and we are done.

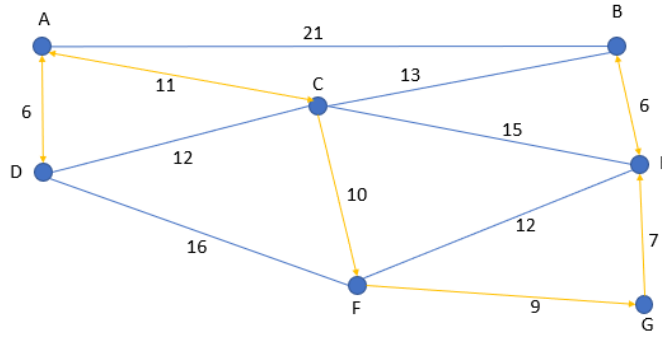


Figure 2.7: Example of a Minimum Spanning Tree.

2.3.2 New Uniformity Criteria based on EMST for Uniform Mixture Designs in the Simplex

Dussert et al. (1986) were the first to present the concept of using EMST to characterize order and disorder of particles using the mean, μ , and standard deviation, σ , of edge lengths. Particles which exhibited partial order were found to have large mean values and small standard deviation values. Wallet and Dussert (1998) showed the discriminatory and stability power of using the mean, μ , and standard deviation,

σ , of edge lengths to characterize a set of points as opposed to using *mindist* criterion and other distance-based criteria. Franco et al. (2009) later extended the concept to characterizing computer designs in the s -dimensional hypercube, C^s . Designs that are described as quasi-periodic are good uniform designs with large mean μ and small standard deviation σ values compared to random and cluster designs (Figure 2.8). Large μ is related to adequately filling or covering the design space and small σ is associated with similar distances between design points corresponding to an even or uniform distribution of design points over the design space.

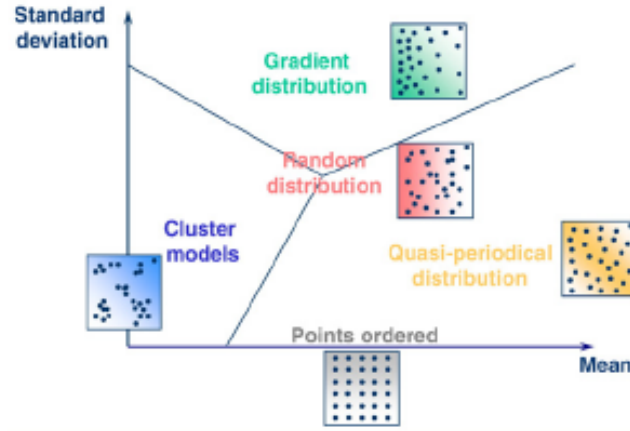


Figure 2.8: Characterization of points in (μ, σ) plane (Franco et al., 2009).

In this study, the EMST criterion is extended to characterizing mixture designs in the q -dimensional simplex, S^q . Two new criteria for assessing uniformity in mixture experiments are proposed: a measure based on the average (*adMST*) and standard deviation (*sdMST*) of the edge lengths, and a measure based on the geometric mean (*GMST*) of the edge lengths.

(i) **Average (*adMST*) and Standard Deviation (*sdMST*) Criteria**

This first approach is adopted from the method presented by Franco et al. (2009). However, for mixture designs, edge lengths in a Euclidean Minimum

Spanning Tree (EMST) will be replaced by scaled edge lengths. This criterion is summarized as follows.

- The design points in the design X in the simplex, $S_{A,B}^q$ are scaled with respect to their range of variation $(B - A)$, where A and B are non-redundant lower and upper component constraints placed on the q mixture component proportions, $(x_1, x_2, \dots, x_q) \in X$. Specifically, $(x_1, x_2, \dots, x_q) \in X$ are replaced with $(\frac{x_1}{a_1-b_1}, \frac{x_2}{a_2-b_2}, \dots, \frac{x_q}{a_q-b_q})$. The purpose behind this scaling is to ensure that each component is treated as equally important before EMST is constructed. That is, components with large range and small range of variations contribute equally to the EMST construction .
- EMST is constructed for the rescaled design points, and the average (*adMST*) and standard deviation (*sdMST*) of the edge lengths are used to assess the space-filling properties of the design. Larger *adMST* corresponds to better filling or covering the design space, and smaller *sdMST* relates to more similar edge lengths and more even spacing of design points across the design region. Consequently, a partial order is introduced, and design X has better space-filling properties than design Y if $adMST(X) > adMST(Y)$ and $sdMST(X) < sdMST(Y)$.

(ii) **Geometric Mean (*GMST*) Criterion**

The second approach uses the geometric mean (*GMST*) of edge lengths from an EMST to characterize uniformity of a design. The rationale behind this new approach lies in the fundamental theorem of the arithmetic mean-geometric mean inequality, and the sensitivity of the geometric mean to variability. The arithmetic mean-geometric mean inequality states that, for any set of nonnegative real numbers, the geometric mean of the set is less than or equal

to the arithmetic mean of the set (Vamanamurthy and Vuorinen, 1994). That is, the geometric mean and arithmetic mean of $n - 1$ edge lengths of an EMST, $d_1, d_2, \dots, d_{n-1}$ are related algebraically as

$$\left(\prod_{i=1}^{n-1} d_i \right)^{1/n-1} \leq \sum_{i=1}^{n-1} \frac{d_i}{n-1},$$

with equality if and only if $d_1 = d_2 = \dots = d_{n-1}$. With regard to the space-filling property, the goal is to generate an EMST with maximum and identical edge lengths. Let $D = \{d_1, \dots, d_{n-1}\}$, in terms of arithmetic mean-geometric mean inequality, this can be expressed as

$$\left(\prod_{i=1}^{n-1} d_i \right)^{1/n-1} \leq \left(\prod_{i=1}^{n-1} \max(D) \right)^{1/n-1} \leq \sum_{i=1}^{n-1} \frac{\max(D)}{n-1} = \max(D),$$

with equality if and only if $d_i = \max(D)$ for all i . Therefore, the geometric mean of the $n - 1$ edge lengths of an EMST is maximized when the maximum edge lengths are identical. Consequently, for a convex bounded region such as the simplex, large *GMST* corresponds to large identical or near identical edge lengths of an EMST. As a result, once an EMST is built for a mixture design, large *GMST* relates to large but similar interpoint-distances between all design points.

The sensitivity of the geometric mean to unequal values is illustrated in Table 2.4. The geometric mean is largest when the numbers in the set are equal and the standard deviation is zero. Thus, as previously discussed, large *GMST* is equivalent to large *adMST* and small *sdMST* values for a bounded convex region such as the simplex or sub-space of the simplex. However, the geometric mean produces a lower score for a fixed arithmetic mean as the variability in the

number set increases, i.e., the standard deviation increases. That is, in terms of EMST, $GMST$ decreases with decreasing edge lengths and approaches zero as one or more edge length tends to zero. This can be expressed mathematically as

$$\lim_{d_1 \rightarrow 0} \left(\prod_{i=1}^{n-1} d_i \right)^{1/n-1} \leq \lim_{d_1 \rightarrow 0} d_1^{1/n-1} \left(\prod_{i=2}^{n-1} d_i \right)^{1/n-1} \leq$$

$$\lim_{d_1 \rightarrow 0} \max(d_1)^{1/n-1} \left(\prod_{i=2}^{n-1} \max(D') \right)^{1/n-1} = 0,$$

where $D' = \{d_2, d_3, \dots, d_{n-1}\}$. Hence, once EMST is built for a mixture design, small $GMST$ is associated with a design with poor space-filling properties.

Number Set	Arithmetic mean	Standard deviation	Geometric mean
(3,3,3,3,3)	3.00	0.00	3.00
(4,2,4,2,4,2)	3.00	1.10	2.83
(5,1,5,1,5,1)	3.00	2.19	2.24
(6,0,6,0,6,0)	3.00	3.29	0.00

Table 2.4: Demonstrating sensitivity property of the geometric mean to increasing variation.

For example, EMSTs for two 11-point designs constructed using GLP method for a 3-component mixture experiment are shown in Figure 2.9. The better uniform design (Figure 2.9a) has larger $GMST$ value and fills the design space better than the design with smaller $GMST$ (Figure 2.9b). In addition, a partial order can be introduced: If $GMST(X) > GMST(Y)$, then the mixture design X has better space-filling properties than the mixture design Y .

The two new measures of uniformity for a 3-component mixture experiment with

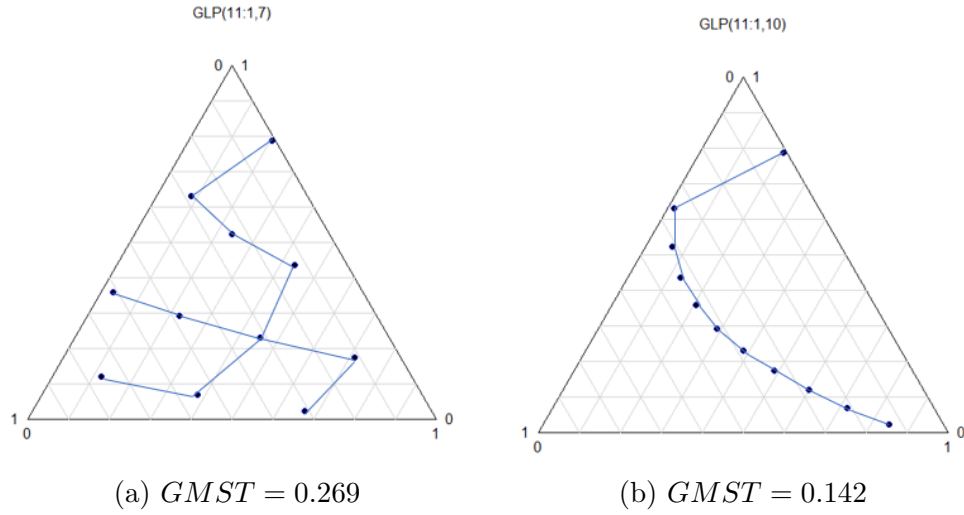


Figure 2.9: 11-point designs for 3-component mixture experiments with corresponding EMSTs and their $GMST$ values

$n = 15$ design points in the full simplex (S^3) and the constrained region $S_{A,B}^3$ in Figure 2.2 using the GLP method are presented in Table 2.5. The designs are arranged from best to worst based on the two new criteria. Thus, in S^3 and $S_{A,B}^3$, both criteria pick designs obtained by the generator (1, 11) as the best space-filling designs. As mentioned earlier, while the G -function transformation conserves the spatial structure of C^2 designs in both S^3 and $S_{A,B}^3$, the systematic spacing uniformity is not strictly preserved. Therefore, the order of best and worst design generators may differ among the two design regions and what is best in S^3 may not be best in $S_{A,B}^3$, and vice versa. For example, (1, 2) is the third best design generator in $S_{A,B}^3$ but it is the fifth best generator in S^3 , and the order of the last five worst generators in S^3 and $S_{A,B}^3$ are not the same.

The relationship between the two new criteria is shown in Table 2.6. $GMST$ values from the eleven design generators in both S^3 and $S_{A,B}^3$ are regressed against $adMST$ and $sdMST$ values. The *Est*, *Std.Error*, *t-stat* and *p* columns are the respective estimated regression coefficients, standard errors of the regression

Design Region	$adMST$	$sdMST$	$GMST$	Generator
Full Simplex (S^3)	0.2272	0.0325	0.2250	(1 11)
	0.2195	0.0283	0.2177	(1 4)
	0.1758	0.0714	0.1654	(1 7)
	0.1659	0.0504	0.1603	(1 8)
	0.1642	0.0476	0.1587	(1 2)
	0.1603	0.0488	0.1545	(1 13)
	0.1151	0.0739	0.1044	(1 14)
Design Region	$adMST$	$sdMST$	$GMST$	Generator
Constrained Simplex ($S^3_{A,B}$)	0.2492	0.0358	0.2464	(1 11)
	0.2404	0.0446	0.2358	(1 4)
	0.1745	0.0290	0.1723	(1 2)
	0.1749	0.0397	0.1714	(1 13)
	0.1859	0.0925	0.1707	(1 7)
	0.1686	0.0710	0.1591	(1 8)
	0.1237	0.0864	0.1110	(1 14)

Table 2.5: New EMST-based measures of uniformity in different simplex regions

coefficients, test statistics, and corresponding p-values. The last row shows R^2 values for the two models. Increasing $GMST$ is associated with increasing $adMST$ and/or decreasing $sdMST$. This relationship is shown in the effect plots in Figures 2.10 and 2.11 for the full simplex and the constrained simplex experimental regions respectively. Also note R^2 value, which shows that approximately 100% of total variation in $GMST$ can be explained by changes in the linear combination of $adMST$ and $sdMST$. Hence, no information is lost in using either criteria to compare designs.

These are justification that the two criteria are interchangeable, and designs can be classified by just considering ($adMST$, $sdMST$) or $GMST$ values. Nevertheless, there is computational advantage in using only one criterion to compare designs, rather than two criteria. Therefore, in this study, the $GMST$ criterion is the preferred criterion. Thus, the tables of good uniform mixture/ process designs given in Appendix A for $q = 3, 4, 5$ and $k = 0, 1, 2$ have been obtained using the $GMST$ criterion.

<i>Predictors</i>	GMST_Full				GMST_Constrained			
	<i>Est.</i>	<i>Std. Error</i>	<i>t-stat</i>	<i>p</i>	<i>Est.</i>	<i>Std. Error</i>	<i>t-stat</i>	<i>p</i>
(Intercept)	0.0047	0.0014	3.4220	0.027	0.0049	0.0018	2.7523	0.051
adMST	0.9980	0.0050	200.2890	<0.001	0.9977	0.0070	142.9209	<0.001
sdMST	-0.2046	0.0109	-18.7503	<0.001	-0.2048	0.0118	-17.4128	<0.001
R ²	1.000				1.000			

Table 2.6: Relationship between $GMST$ and $adMST$, $sdMST$.

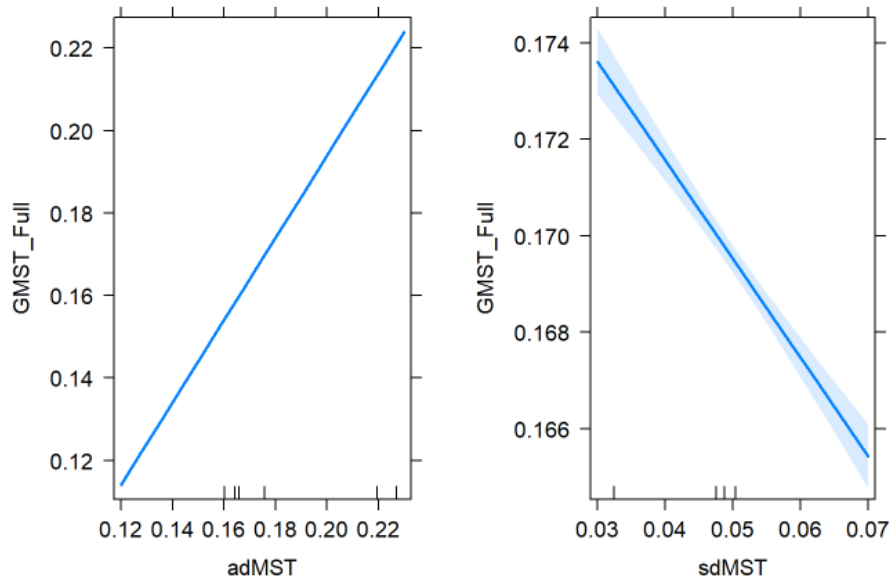


Figure 2.10: Effect plots of $GMST$ versus $adMST$ and $sdMST$ in the full simplex.

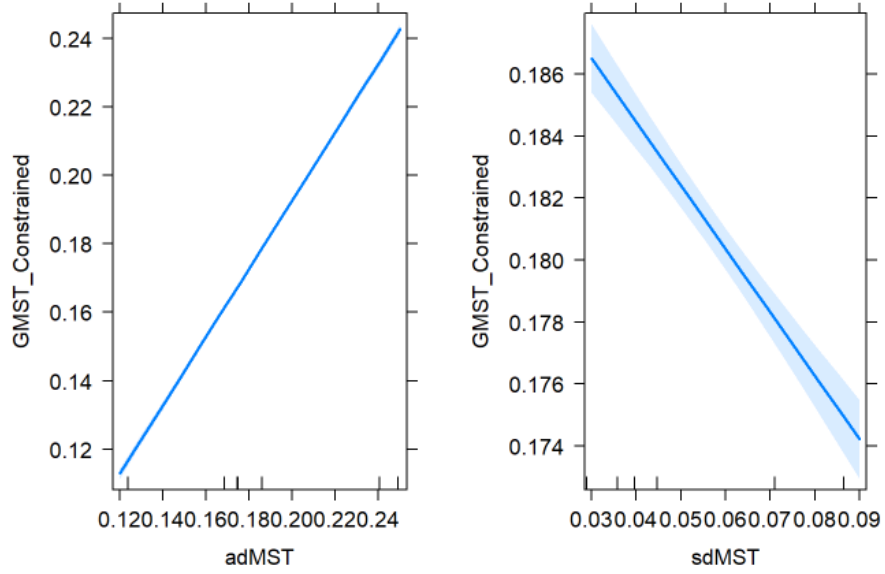


Figure 2.11: Effect plot of $GMST$ regressed against $adMST$ and $sdMST$ in the constrained simplex.

2.4 A New Conditional Approach For Finding Good Uniform Mixture/Process Designs Using EMST

The purpose of a good uniform design for a q -component mixture experiment with k process variables is to provide design points which have good space-filling properties across the combined $q - 1 + k$ -dimensional experimental region. In Section 2.2, a new hybrid algorithm for generating uniform designs for such experiments was introduced. However, there are challenges in quantifying space-filling properties of designs in the combined $q - 1 + k$ -dimensional experimental region. For instance, suppose we consider a 3-component mixture/process experiment with one process variable and a mixture design region similar to the design space in Figure 2.2. The challenges in assessing the degree of uniformity of designs in this combined design space are illustrated in Figure 2.12. The two designs shown are generated via the GLP method using design generators $(1, 11, 4)$ and $(1, 7, 4)$. Each point in the simplex

is a mixture design point, and each color denotes the corresponding process variable value or level. Similar colors that are close in the mixture space indicate clustering of design points at the similar process variable levels while arrangement with dissimilar colors of neighboring points suggests that design points at the process variable level are more uniformly spread. The process variable levels from design points generated by $(1, 7, 4)$ are more evenly spread than process variable values from design points generated by $(1, 11, 4)$. Consequently, the *GMST* criterion picks $(1, 7, 4)$ as the better uniform mixture/process design generator although this design has poor space-filling properties when projected down to the simplex sub-space as shown in Figure 2.12 and Table 2.5. A good uniform mixture/process design should, however, have good space-filling properties in the combined region and also have good space-filling properties when projected down to the mixture space or at any lower dimensional projection.

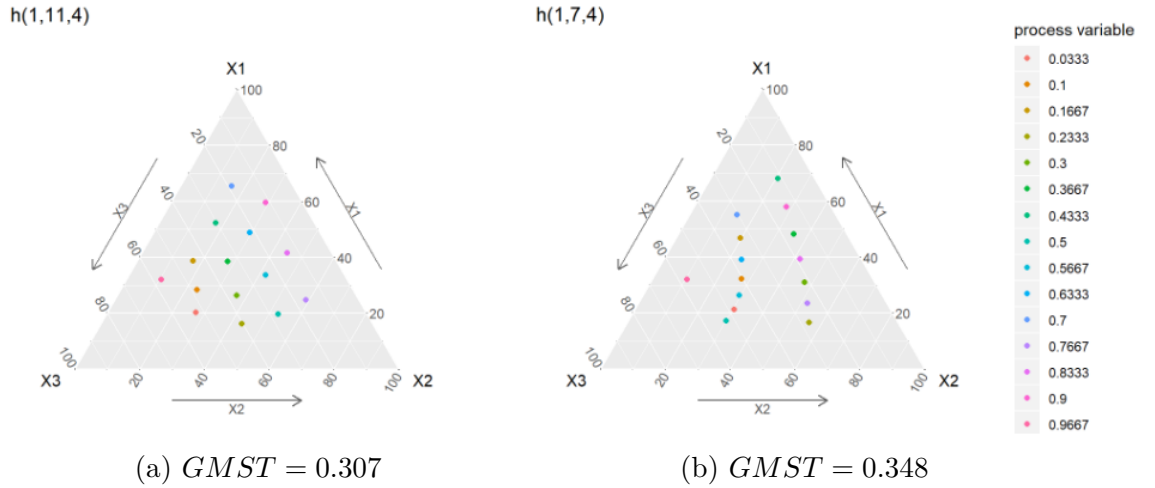


Figure 2.12: Mixture/process variable designs

A conditional approach that captures the space-filling properties of mixture/process variable designs for q -component mixture experiments with k process variables in the combined $q - 1 + k$ -dimensional design region and lower projected sub-regions down to $q - 1$ simplex regions will now be presented.

- (i) Find the best mixture designs in S^q using the *GMST* criterion. That is, select designs with high *GMST* values.
- (ii) Add one process variable, and choose the best combined mixture and process variable designs created by using EMSTs conditioned on a subset containing the best mixture designs in Step (i).
- (iii) Retain the best designs having both large *GMST* in S^q and large *GMST* for the combined space.
- (iv) Repeat steps (ii) and (iii) to sequentially add $2, \dots, k$ process variables, and keep the best uniform mixture/process designs which are designs with large *GMST* in S^q and large *GMST* in each $(q - 1 + j)$ -dimensional region for $j = 1, 2, \dots, k$.

We will continue the example, where the goal is to generate good uniform designs for a 3-component mixture experiment with one process variable and restrictions $0.1 \leq x_1 \leq 0.75$, $0.15 \leq x_2 \leq 0.8$, and $0.1 \leq x_3 \leq 0.6$ on the mixture components. From Table 2.5, the best three design generators in $S_{A,B}^3$ based on to the *GMST* criterion are $(1, 11)$, $(1, 4)$, and $(1, 2)$.

The results from adding one process variable to these three design generators are shown in Figure 2.13. Note, for visual clarity, 1 is removed from each generator, e.g. $(1, 11, 2) \rightarrow (11, 2)$ and $(1, 2, 7) \rightarrow (2, 7)$. The overall best design generator is $(1, 11, 2)$, and the plot of this design is shown in Figure 2.14a. One can see the benefits of using the proposed conditional method in Figure 2.14. While the two design generators $(1, 11, 2)$ and $(1, 2, 7)$ in the combined region have identical *GMST* values, the $(1, 11, 2)$ design has better space properties in the sub-space of the simplex. Thus, this generator is correctly selected as the better generator.

This conditional approach can also be applied when the settings of the process variables are categories. For example, if we reconsider this example and assume

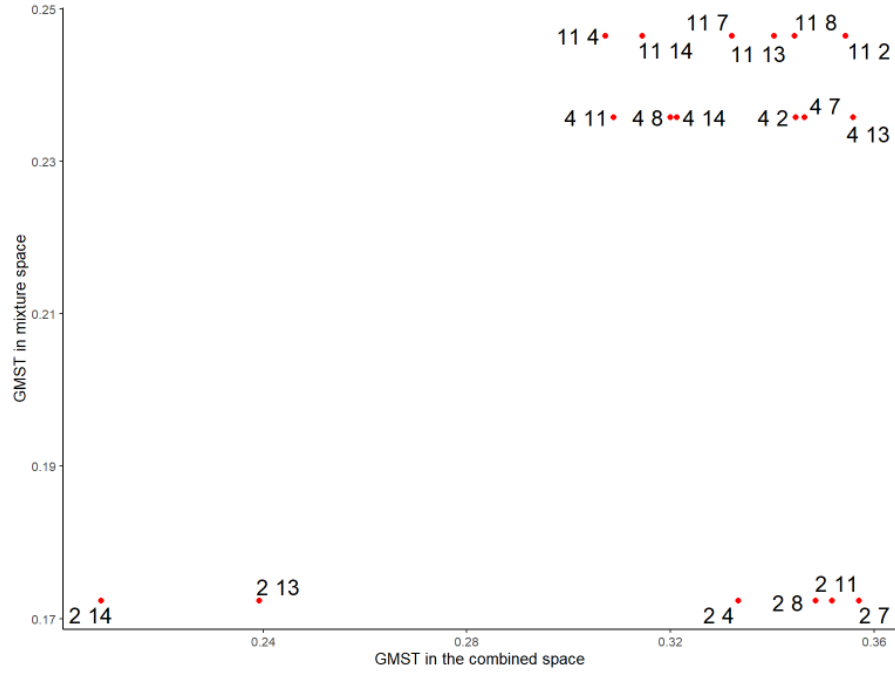
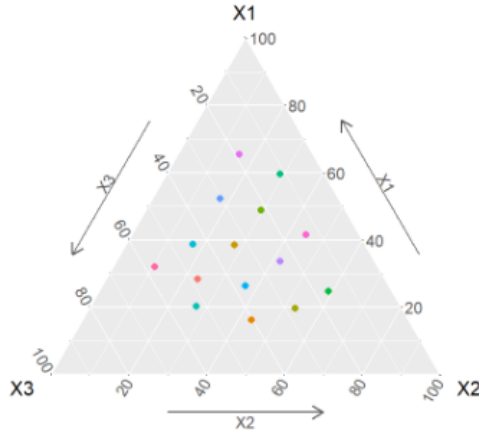


Figure 2.13: *GMST* in the mixture space and in the combined space.

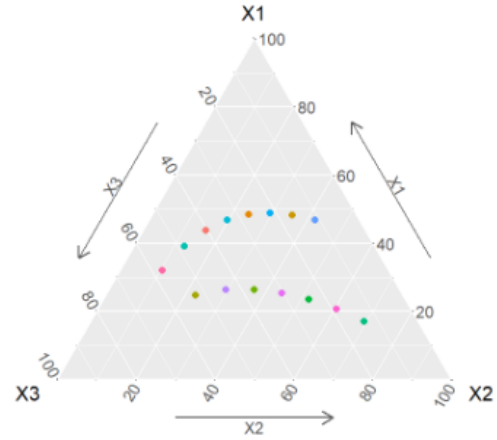
that the process variable is a categorical variable with 3 levels, then the initial 15 values of the process variable can be grouped into 3 equally spaced levels (0, 0.5, 1). Research goals, costs, and other considerations may drive how the initial 15 points are distributed into the 3 categories. For example, suppose the goal is to learn more about the process variable's low and high regions, then 6 points can be grouped into the low and high (0 and 1) levels, with 3 points falling into the middle level (0.5). To achieve this, the 15 points are binned into 3 categories, thus grouping points between 0 and 0.4 into level 1, points between 0.41 and 0.6 into level 2, and points between 0.61 and 1 into level 3 ($[0, 0.4] \rightarrow 0$, $(0.4, 0.6] \rightarrow 0.5$, and $(0.6, 1] \rightarrow 1$). Figures 2.16a and 2.16b show the results from this scenario, as well when the 15 points are distributed equally across the 3 categories. The order of the best and worst designs for the two scenarios is slightly different. However, the best design is obtained in both scenarios with generator (1, 11, 2), and these two designs are shown in Figure 2.15.

h(1,11,2)



(a) Good design.

h(1,2,7)



(b) Bad design.

Figure 2.14: Good and bad uniform mixture/process variable design.

2.4.1 Mixture/Process Design Example

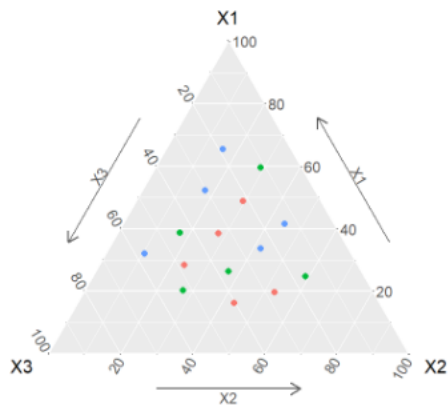
This example is taken from *Response Surfaces: Designs and Analyses* by Myers et al. (2016), pages 766-769. Adhesives are complex formulations for binding substrates or sealing joints or gaps. In this example, an adhesive is being formulated for use in an aerospace application. The adhesive which consists of resin and two cross-linkers is applied to the component, and then the entire assembly is cured for 12 hr at controlled temperature and humidity. There are three mixture component proportions: resin (x_1), cross-link 1 (x_2), and cross-link 2 (x_3), and two process variables: temperature (z_1) and relative humidity (z_2). The following constraints apply:

mixture components: $0.7 \leq x_1 \leq 0.9$, $0.05 \leq x_2 \leq 0.10$, and $0.05 \leq x_3 \leq 0.20$,

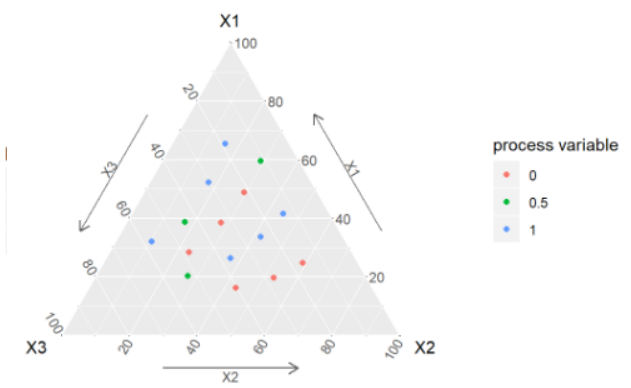
process variables: $40 \leq z_1 \leq 100$ (degrees F) and $15 \leq z_2 \leq 85$ (percent).

The experimenters chose to use a computer generated optimal design with 34 runs consisting of 29 distinct points and 5 replicates. The shortcomings of model-based optimal computer designs were highlighted in Chapter One. Hence, the

proposed conditional algorithm is used to generate a 34-run uniform design. The uniform design also consists of 29 distinct design points of which 5 runs at the extreme ends of the minimum spanning tree were replicated. The *GMST* values for 29-point mixture designs using the GLP method are shown in Figure 2.17. The two best designs are attained by generators (1, 5) and (1, 24). These are the designs with the two highest *GMST* values. Design points for temperature are then conditionally added to these two best designs in the mixture space. Note that there is little difference in *GMST* among the first five designs, so the temperature design points could be added to these five best designs in the mixture space. The two best designs in the mixture space were used in this study for computational efficiency. The two temperature levels ($40^{\circ}F$, $100^{\circ}F$) are coded as -1 and 1 respectively. The 10 best mixture/process variable design generators after adding the temperature design points are shown in Figure 2.18. The final step involves adding the relative humidity design points conditioned on the 10 best designs in Figure 2.18. The relative humidity levels (15%, 85%) are also coded as -1 and 1 respectively. The best mixture/process variable designs resulting from adding the second process variable points to the 10 best designs in Figure 2.18 are shown in Figure 2.19. The overall best mixture/process variable design generator is (1, 5, 12, 14), and the corresponding design points are shown in Table 2.7.

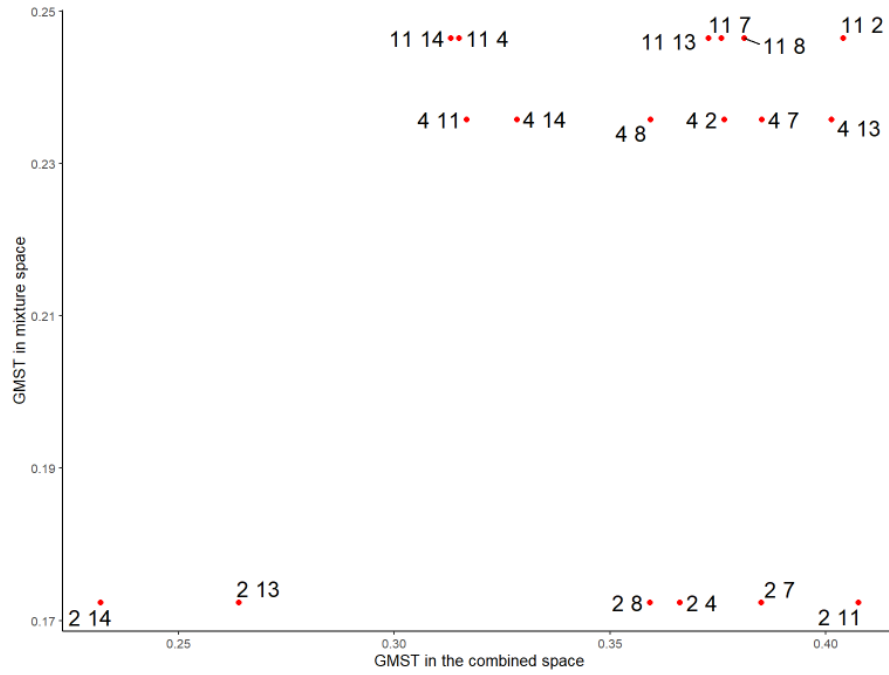


(a) Equal points per category.

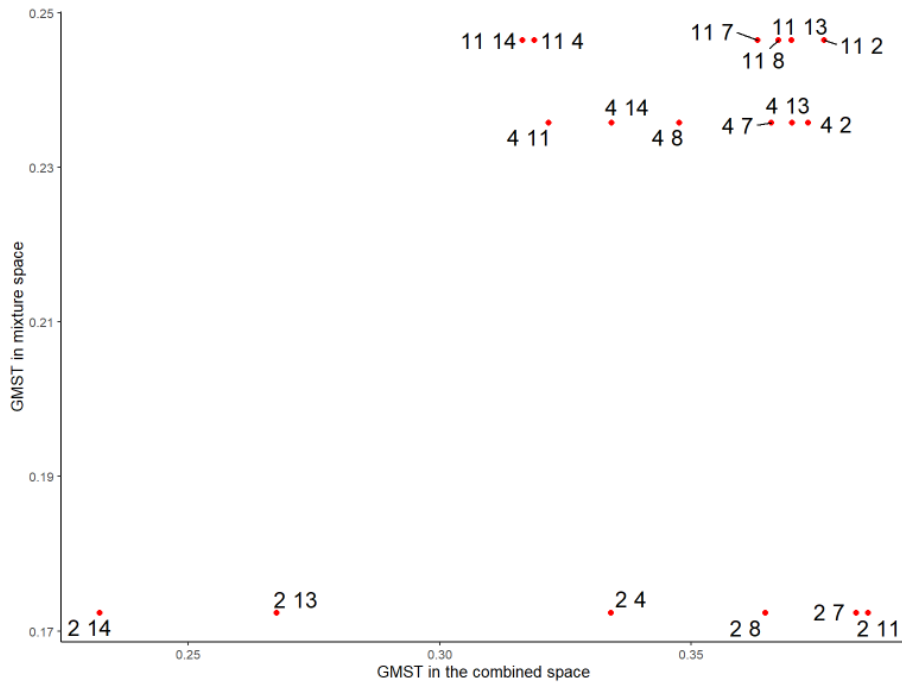


(b) More points at low and high levels.

Figure 2.15: Best 15-point mixture and categorical process variable designs.



(a) Equal distribution of points into the 3 levels.



(b) Unequal distribution of points into the 3 levels.

Figure 2.16: Conditional $GMST$ measure for 15-point mixture and categorical process variable designs.

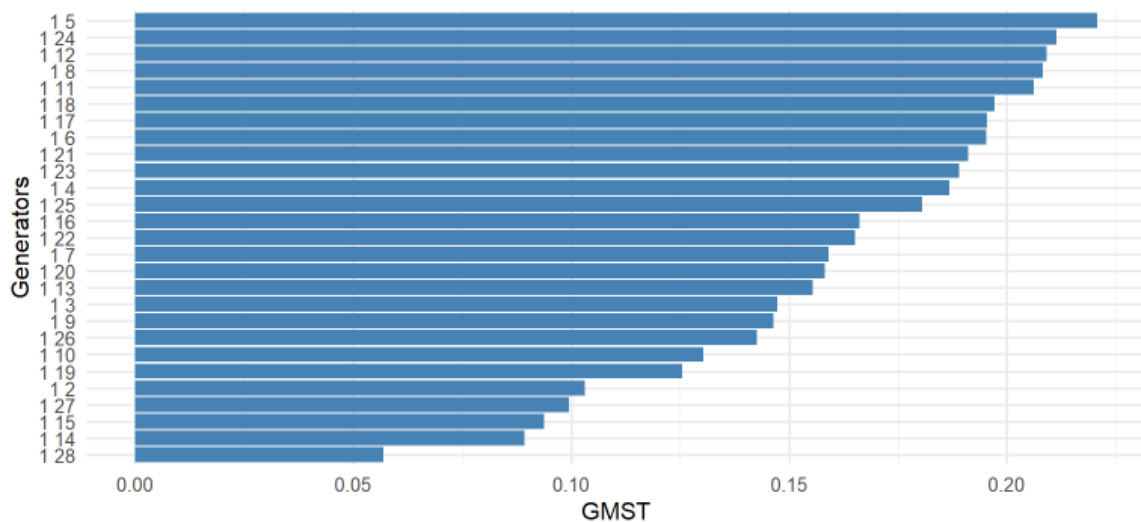


Figure 2.17: The best to worst 29-point designs in the constrained simplex using the GLP method.

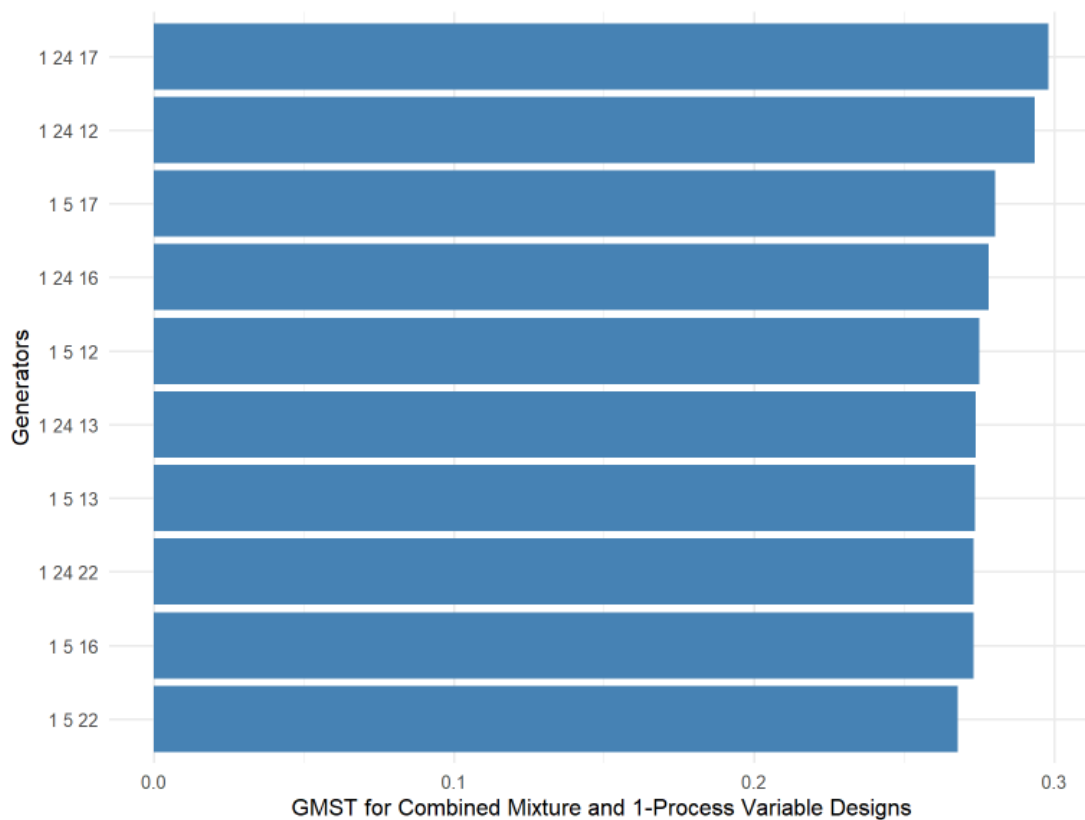


Figure 2.18: The 10 best 3-component mixture and 1-process variable GLP designs.

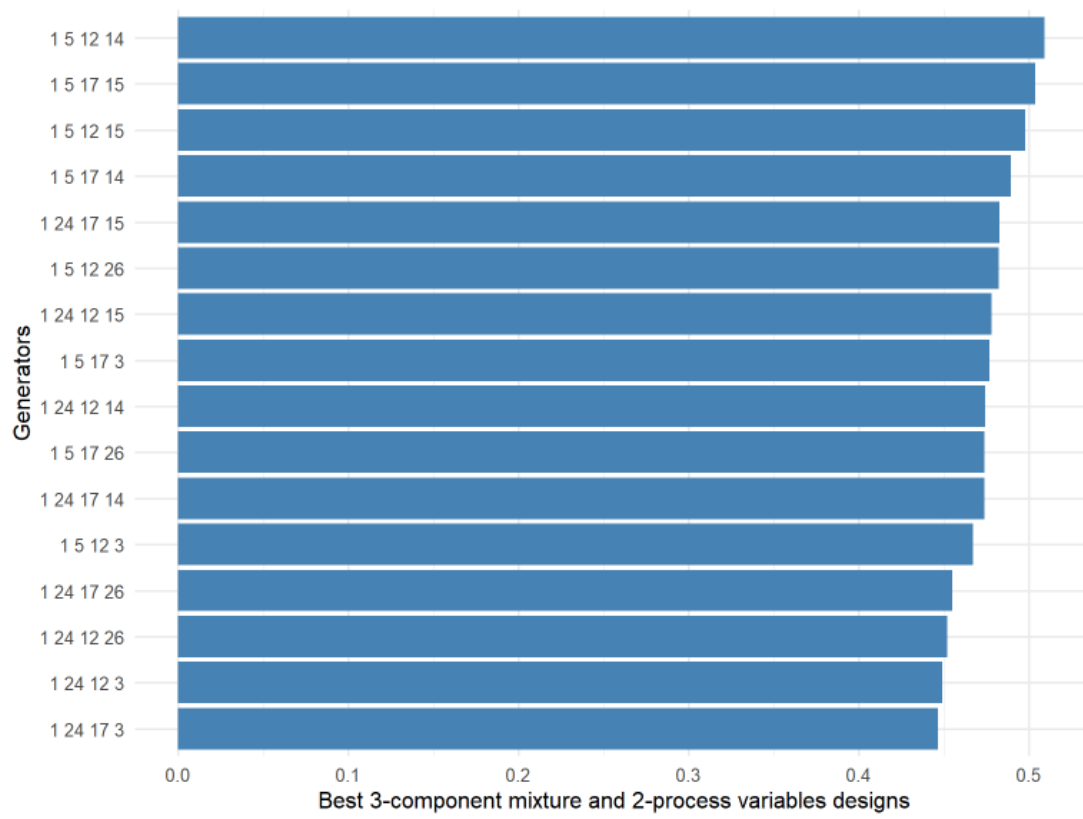


Figure 2.19: The best 3-component mixture and 2-process variable GLP designs.

	Mixture Variables			Coded Process		Actual Process	
Runs	x_1	x_2	x_3	z_1	z_2	Temp ($^{\circ}F$)	RH (%)
1	0.8775	0.0509	0.0717	-1	-1	40	15
2	0.851	0.0526	0.0964	1	1	100	85
3	0.8239	0.0543	0.1218	-1	-1	40	15
4	0.796	0.056	0.1479	1	1	100	85
5	0.7673	0.0578	0.1749	-1	-1	40	15
6	0.8881	0.0595	0.0524	-1	1	40	85
7	0.8622	0.0612	0.0766	1	-1	100	15
8	0.8357	0.0629	0.1014	-1	1	40	85
9	0.8084	0.0647	0.127	1	-1	100	15
10	0.7803	0.0664	0.1533	-1	1	40	85
11	0.7515	0.0681	0.1804	1	-1	100	15
12	0.873	0.0698	0.0572	1	1	100	85
13	0.847	0.0716	0.0815	-1	-1	40	15
14	0.8203	0.0733	0.1065	1	1	100	85
15	0.7928	0.075	0.1322	-1	-1	40	15
16	0.7646	0.0767	0.1586	1	1	100	85
17	0.7356	0.0784	0.186	-1	-1	40	15
18	0.8578	0.0802	0.062	-1	1	40	85
19	0.8317	0.0819	0.0864	1	-1	100	15
20	0.8048	0.0836	0.1115	-1	1	40	85
21	0.7773	0.0853	0.1374	1	-1	100	15
22	0.7489	0.0871	0.164	-1	1	40	85
23	0.7196	0.0888	0.1916	-1	-1	40	15
24	0.8427	0.0905	0.0668	1	1	100	85
25	0.8164	0.0922	0.0914	-1	-1	40	15
26	0.7894	0.094	0.1167	1	1	100	85
27	0.7617	0.0957	0.1427	-1	-1	40	15
28	0.7331	0.0974	0.1695	1	-1	100	15
29	0.7037	0.0991	0.1972	1	1	100	85
Replicates							
30	0.8775	0.0509	0.0717	-1	-1	40	15
31	0.8881	0.0595	0.0524	-1	1	40	85
32	0.7515	0.0681	0.1804	1	-1	100	15
33	0.8048	0.0836	0.1115	-1	-1	40	15
34	0.7037	0.0991	0.1972	1	-1	100	15

Table 2.7: 34-point uniform design for 3-component mixture and 2-process variable experiment.

CHAPTER THREE

EXAMPLE OF MIXTURE/PROCESS VARIABLE DESIGN AND DATA
ANALYSIS

New methods for generating uniform designs (UDs) for mixture/process variable experiments and new methods for assessing the space-filling properties of such designs were discussed and described in detail in the Chapter Two of this dissertation. This chapter will include a presentation and discussion of the layout and analysis of experimental data for UD generated by the new proposed conditional mixture/process space-filling algorithm. This chapter will also include examples using a simulation study to demonstrate the advantages and robustness of mixture/process variable UD in data analysis. As used in this dissertation, robustness refers to the flexibility for choice of a potential model.

3.1 The Mixture/Process Variable Model

Consider an q -component quadratic mixture model:

$$y = \sum_{i=1}^q \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j + \epsilon,$$

and the interaction response surface model for k process variables:

$$y = \alpha_0 + \sum_{t=1}^t \alpha_t z_t + \sum_{t < r} \alpha_{tr} z_t z_r + \epsilon.$$

Look at the general **product model**:

$$y = \left[\sum_{i=1}^q \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j \right] \left[\alpha_0 + \sum_{t=1}^t \alpha_t z_t + \sum_{t < r} \alpha_{tr} z_t z_r \right] + \epsilon.$$

Expanding the product yields

$$\begin{aligned}
 y &= \sum_{i=1}^q \left[\alpha_0 \beta_i + \sum_{t=1}^t \beta_i \alpha_t z_t + \sum_{t < r} \beta_i \alpha_{tr} z_t z_r \right] x_i \\
 &+ \sum_{i < j} \left[\alpha_0 \beta_{ij} + \sum_{t=1}^t \beta_{ij} \alpha_t z_t + \sum_{t < r} \beta_{ij} \alpha_{tr} z_t z_r \right] x_i x_j + \epsilon.
 \end{aligned}$$

This model is over-parameterized. Therefore, let

$$\gamma_i^0 = \alpha_0 \beta_i, \quad \gamma_i^t = \alpha_k \beta_i, \quad \gamma_i^{tr} = \alpha_{kl} \beta_i,$$

$$\gamma_{ij}^0 = \alpha_0 \beta_{ij}, \quad \gamma_{ij}^t = \alpha_k \beta_{ij}, \quad \gamma_{ij}^{tr} z_t z_r = \alpha_{kl} \beta_{ij}.$$

Then, after substitution, we get

$$y = \sum_{i=1}^q \left[\gamma_i^0 + \sum_{t=1}^t \gamma_i^t z_t + \sum_{t < r} \gamma_i^{tr} z_t z_r \right] x_i + \sum_{i < j} \left[\gamma_{ij}^0 + \sum_{t=1}^t \gamma_{ij}^t z_t + \sum_{t < r} \gamma_{ij}^{tr} z_t z_r \right] x_i x_j + \epsilon,$$

$$\epsilon \sim N(0, \sigma^2).$$

This is an example of a *mixture/process variable model*. In general, a **mixture/process variable model**, will be the reduced model resulting from the product of a Scheffe mixture model in q components and a response surface design in k process variables. For example, when there are $q = 3$ mixture components and $k = 2$ process variables, the 24-term product model is

$$\begin{aligned}
y = & \sum_{i=1}^3 \gamma_i^0 x_i + \sum_{i=1}^3 \sum_{t=1}^2 \gamma_i^t x_i z_t + \sum_{i=1}^3 \gamma_i^{12} x_i z_1 z_2 \\
& + \sum_{i=1}^2 \sum_{j=i+1}^3 \gamma_{ij}^0 x_i x_j + \sum_{i=1}^2 \sum_{j=i+1}^3 \sum_{t=1}^2 \gamma_{ij}^t x_i x_j z_t + \sum_{i=1}^2 \sum_{j=i+1}^3 \gamma_{ij}^{12} x_i x_j z_1 z_2 + \epsilon \quad (3.1)
\end{aligned}$$

3.1.1 Mixture/Process Variable Example 1

For this example, we consider the deterministic model

$$y = 10 + 200 \sum_{i=1}^3 (x_i + 0.001) * \exp(-3x_1 - 2x_2 + 2x_3 + z_1 + z_2 + 4x_1 z_1 - 2z_1 z_2) \quad (3.2)$$

and a three-component (x_1, x_2, x_3) mixture experiment with two process variables (z_1, z_2) . Suppose there are resources to run a 33-point design and the goal is to find an approximating model and input combinations that optimize the response variable y . We will compare the performance of a computer generated D-optimal design and a uniform design for this example.

A D-optimal design is a model-dependent computer design which maximizes the determinant of the information matrix $X'X$ of a model matrix X , so that the determinant of the covariance matrix of the parameter estimates may be minimized (Aguiar et al., 1995). Two D-optimal designs are considered: 1) a design that assumes a linear model for the mixture space and an interaction model to represent the process variables (L/I model), and 2) a design that assumes a quadratic model for the mixture space and an interaction model to represent the process variables space (Q/I model). The seven candidate set of points for the mixture design are taken to be the three vertices, the three midpoint binary blends, and the overall centroid. A 2^2 factorial arrangement was also used for the process setting, where low and high levels are coded

as -1 and 1 . The design configurations of the L/I based D-optimal design and the Q/I based D-optimal design are shown in Figure 3.1 and Figure 3.2 respectively. These configurations consist of arranging the distinct design points in the mixture space at each of the four combinations of the two process variables levels $(-1, 1)$. Table 3.1 shows the two D-optimal designs along with simulated response values generated using equation 3.2 and random noise from independent normal distributions with mean 0 and standard deviation 1. The L/I based D-optimal design consists of 12 distinct points and 21 replicates while the Q/I based D-optimal design has 24 distinct points and 9 replicates. These design points were obtained using the *optFederov()* function from the *AlgDesign* package (Wheeler and Braun, 2019) in the R statistical package version 4.0.3 (Team, 2020).

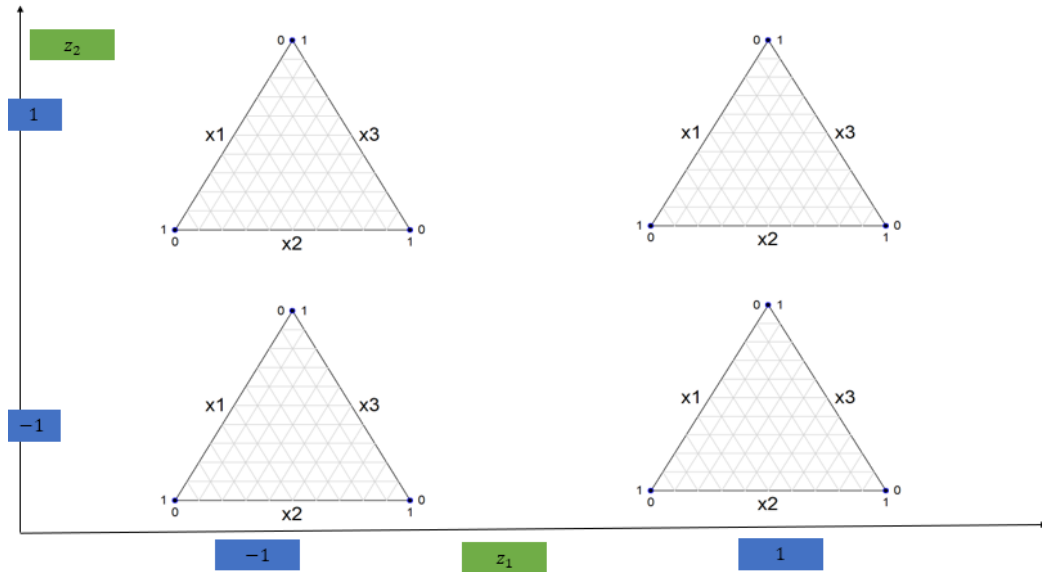


Figure 3.1: LI-based D-optimal design for Example 1.

The third design considered is a 33-point uniform. The design configuration is shown in Figure 3.3. Table 3.3 shows the uniform design points generated using the GLP method and the generator $(1, 5, 13, 3)$, along with simulated response values also

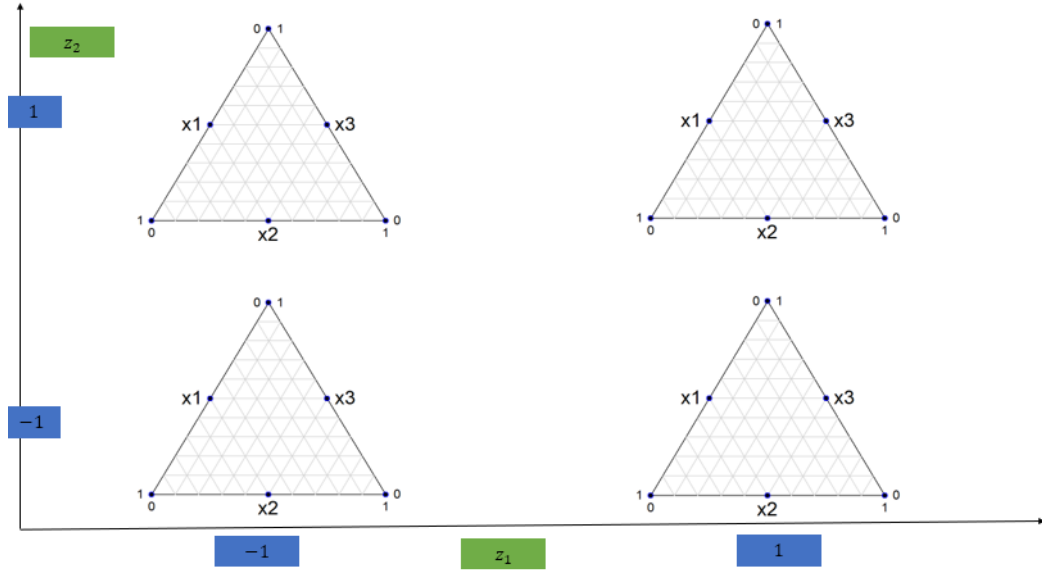


Figure 3.2: QI-based D-optimal design for Example 1.

produced using equation 3.2 and random noise from a normal distribution with mean 0 and standard deviation 1. There are 28 distinct design points with 5 replicated runs selected from the extreme ends of the minimum spanning tree constructed for this design.

Model summaries from fitting different models to the optimal and uniform design points are presented in Table 3.2. Note that, R^2 values from models without intercepts like mixture models may be inflated due to the fact that the total sum of squares is not adjusted for the overall mean, and an impression may be given that the models fit better than they do (Cornell, 2011). The mixture models in this dissertation were fitted to address this issue by leaving out one linear term so that the correct R^2 values are produced and intercepts reported in the models summary are replaced with the linear terms that were left out (Lawson and Willden, 2016).

The best approximating model is the special cubic $\times$ interaction product model (SC/I model) fitted to the uniform design points. This SC/I model has the smallest

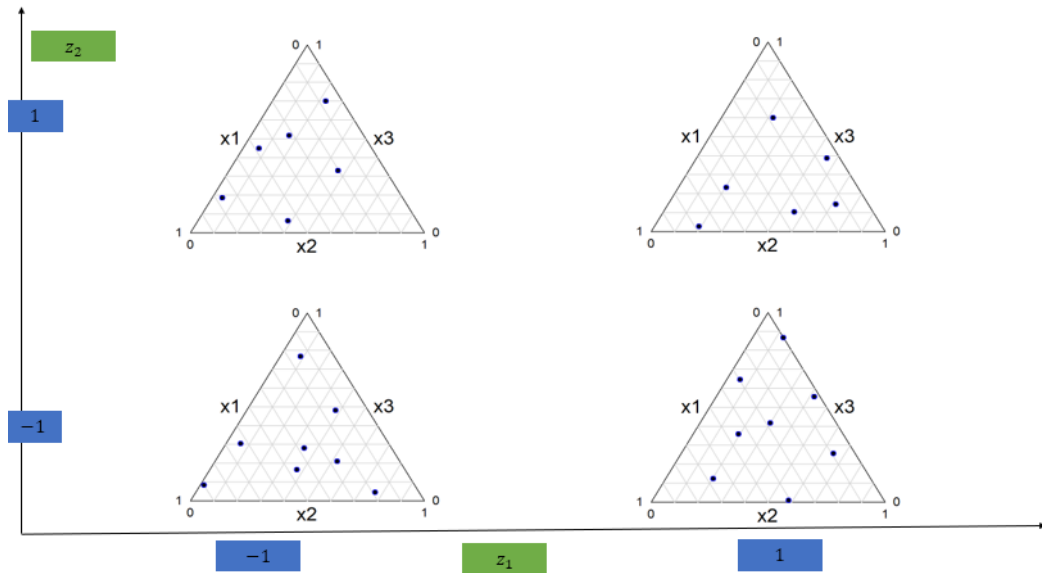


Figure 3.3: Uniform design design for Example 1.

residual error of 0.5329, smallest AIC of 47.835, smallest BIC of 91.233, largest R^2 of 0.9999, and model p-value less than 0.001. The summary for this model is presented in Table 3.4, and Figure 3.4 presents the predicted response values at each of the four combinations of the process variables. The highest values of the response variable are obtained at high and low levels of the first and second process variables, respectively. On the other hand, if the goal is to minimize the response variable, then low levels of both process variables may be considered. However, depending on the available resources and operational costs, other combinations of process variables other than the cube corners can also be considered for optimal response values.

Note that the two optimal designs cannot fit the SC/I model as this model would require at least 28 distinct points. This is one major advantage of uniform designs, the flexibility to fit alternative models that may not be possible with model-dependent optimal and Scheffe-type designs. That is, unlike model-dependent computer generated designs, uniform designs provide the flexibility in the number of

distinct and replicated runs used, and consequently, the number of potential models that can be considered. Also, from Table 3.2, the Q/I model fitted to the uniform design points was a better fit ($AIC=73.510$, $BIC=102.923$) than the Q/I model fitted to the D-optimal design points ($AIC=92.089$, $BIC=129.502$). This further illustrates the drawbacks of Scheffe-type and optimal designs for mixture experiments. Their usefulness depends on how well the assumed model can approximate the true model form, and, therefore, an assumption that the fitted model will provide a reasonable approximation to the truth. In certain instances, the true model form may deviate from this assumed model form, and the use of such classical and optimal designs will be less than ideal. In such instances, a good uniform design that spreads design points evenly throughout the design region, ensuring that the information obtained is representative of the design region, is closer to ideal.

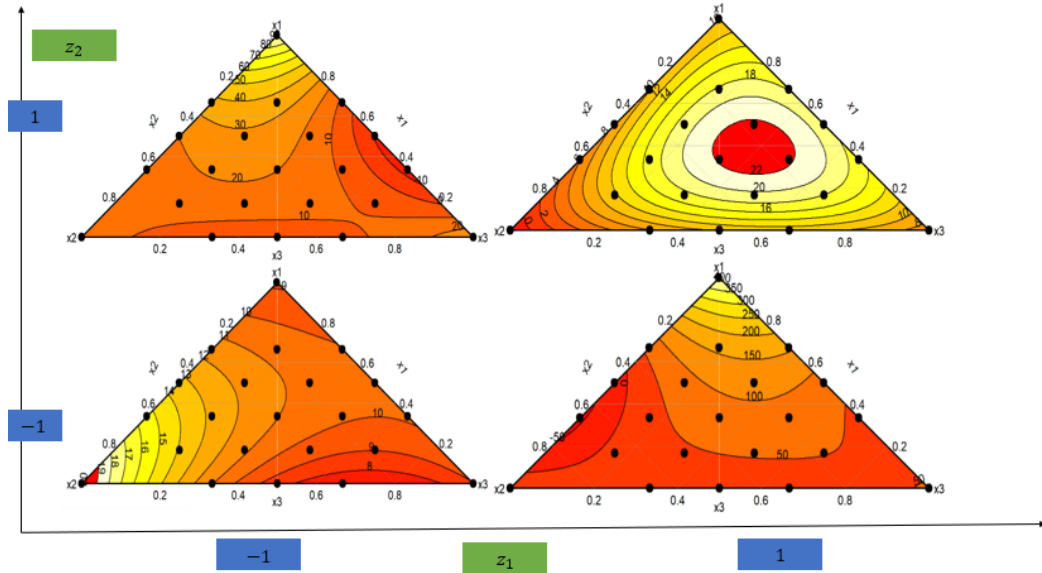


Figure 3.4: Contour plots of predicted responses from SC/I model for Example 1 using space-filling mixture process variable design.

3.1.2 Mixture/Process Variable Example 2

We will revisit the adhesives formulations experiment example introduced in Section 2.4.1. In this example, an adhesive which consists of resin and two cross-linkers is being formulated for use in an aerospace application. There are three mixture component proportions: resin (x_1), cross-link 1 (x_2), and cross-link 2 (x_3), and two process variables: temperature (z_1) and relative humidity (z_2). The following constraints apply:

mixture components: $0.7 \leq x_1 \leq 0.9$, $0.05 \leq x_2 \leq 0.10$, and $0.05 \leq x_3 \leq 0.20$,

process variables: $40 \leq z_1 \leq 100$ (degrees F) and $15 \leq z_2 \leq 85$ (percent). The mixture design space is shown in Figure 3.5.

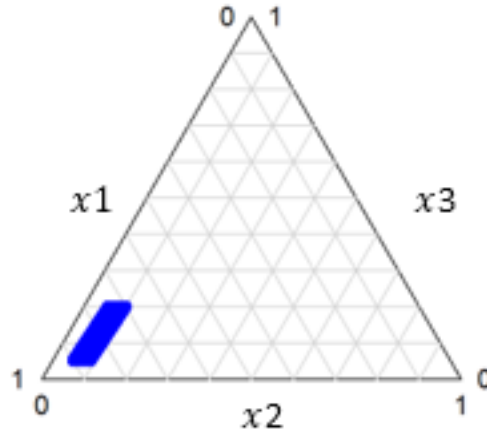


Figure 3.5: Constrained mixture design space for Example 2.

The goal of the experiment is to find combinations of the three mixture components and two process variables which maximize the adhesive's pull-off force (lbs). The process variables are viewed as fixed factors: temperature at 40°F and 100°F, and relative humidity at 15% and 85%. The researchers used the quadratic mixture and the interaction process variable product model in Equation 3.1 to generate a D-optimal design with 34 design points selected from 52 candidate points

consisting of four vertices, four edge centers, an overall centroid and four axial check blends for the constrained mixture at each of the four process variables combinations. The 34-point D-optimal design consists of 29 distinct points and 5 replicates. The design configuration is shown in Figure 3.6 and the design points with corresponding response values are shown in Table 3.5. Note that for design generation, the coded levels $(-1, 1)$ of the process variables are used instead of the actual values.

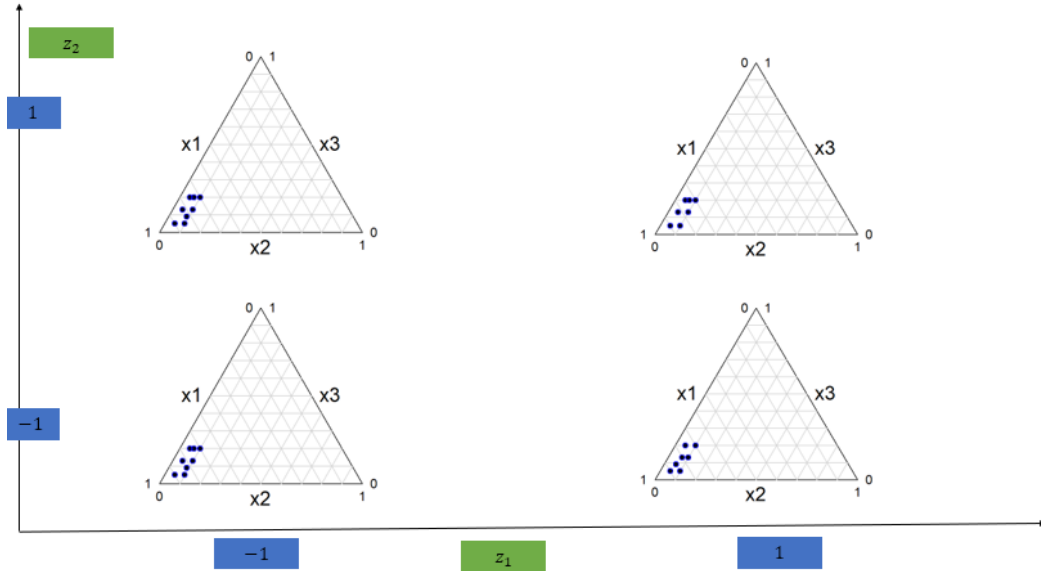


Figure 3.6: Mixture/Process variable D-optimal design for Example 2.

Next, we consider the scenario where a uniform design is used for this experiment. As discussed earlier, one advantage of uniform design is that we do not have to assume any model form prior to data collection. For this example, we will use the 34-point uniform design generated in Section 2.4.1 and presented in Table 2.7. Recall that this design consists of 29 distinct design points and 5 replicated runs. Also note that the coded levels $(-1, 1)$ of the process variable levels are used instead of the actual values. The design configuration is shown in Figure 3.7.

We are going to consider two data generation scenarios:

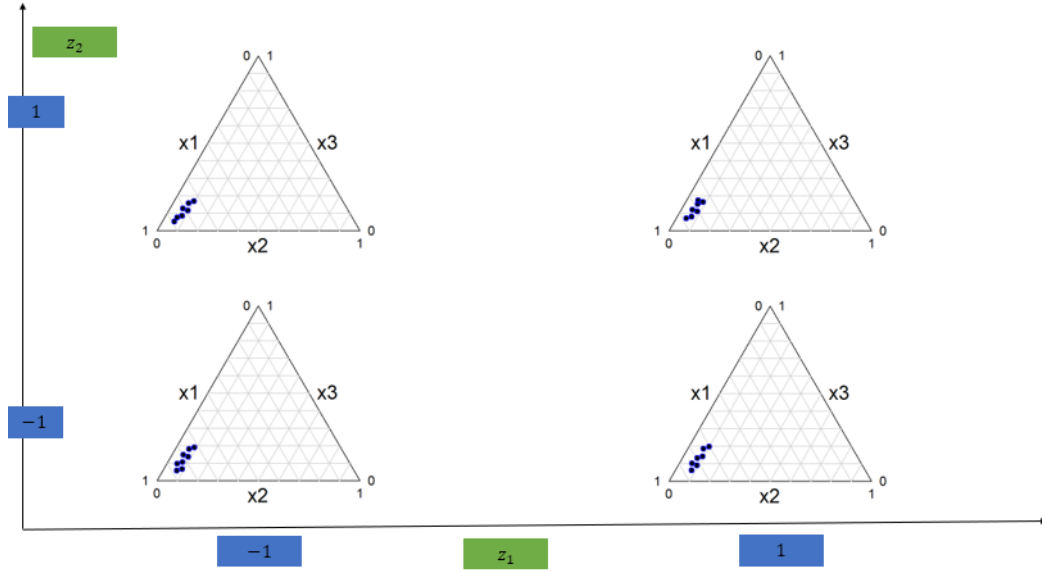


Figure 3.7: Mixture/Process variable uniform design for Example 2.

1. Random datasets using the quadratic mixture and interaction process variable product model in Equation 3.1. The following specified parameter values are used:

$$\sigma = 4.598, \gamma_1^o = 10.37, \gamma_2^o = -6127.15, \gamma_3^o = 466.08, \gamma_1^1 = -25.94, \gamma_2^1 = -11.94,$$

$$\gamma_1^2 = -4177.41, \gamma_2^2 = 397.35, \gamma_1^3 = -358.61, \gamma_2^3 = 187.22, \gamma_1^{12} = -12.40,$$

$$\gamma_2^{12} = -1898.31, \gamma_3^{12} = 353.59, \gamma_{12}^o = 7292.79, \gamma_{13}^o = -732.53, \gamma_{23}^o = 7860.39,$$

$$\gamma_{12}^1 = 5009.38, \gamma_{12}^2 = -519.78, \gamma_{13}^1 = 462.39, \gamma_{13}^2 = -204.36, \gamma_{23}^1 = 5295.43,$$

$$\gamma_{23}^2 = -555.07, \gamma_{12}^{12} = 2294.33, \gamma_{13}^{12} = 2294.33, \gamma_{13}^{12} = -470.53, \gamma_{23}^{12} = 1642.42.$$

These parameter values are obtained from fitting the quadratic mixture and interaction process variable product model to the original D-optimal design points and the actual response values in Table 3.5 using the *lm* function in the

R statistical package version 4.0.3 (Team, 2020). The corresponding simulated response values using these parameter values for the D-optimal and uniform designs are presented in Table 3.7.

2. Random datasets using the special cubic mixture model and interaction process variable product model:

$$y = \text{Equation (3.1)} + \gamma_{123}^0 x_1 x_2 x_3 + \sum_{k=1}^2 \gamma_{123}^k x_1 x_2 x_3 z_k + \gamma_{123}^{12} x_1 x_2 x_3 z_1 z_2. \quad (3.3)$$

The parameter values for the first 24 terms and the noise parameter (σ) are the same as those used for the quadratic mixture and process variable interaction product model. The four additional terms contained in the model are as follows:

$$\gamma_{123}^0 = -5000, \gamma_{123}^1 = -2000, \gamma_{123}^2 = -4000, \gamma_{123}^{12} = -1000.$$

The simulated response values using these parameter values and their respective design points are presented in Table 3.8.

We will consider fitting two product models:

- Quadratic mixture model $\times$ Interaction process variable model.
- Special cubic mixture model $\times$ Interaction process variable model.

Under each data generation scenario and model combination, the results of the data analyses for the two design types are presented in Table 3.6. Both designs were able to identify the Q/I model as the best fitting model for the data generated using the Q/I model. Both designs also identified the SC/I model as the best fitting model for the data generated assuming the SC/I model. Mixture formulations

of $(0.8048, 0.0836, 0.115)$ and $(0.85, 0.1, 0.05)$ using uniform and D-optimal designs, respectively, produced the highest predicted force values (80 and above) when temperature is high (+1) and humidity is low (-1). In this case, the true models were known and the data were generated using parameter estimates from models fitted to the original D-optimal design points. So, in this case, the D-optimal designs are known to be suitable. Nonetheless, the uniform design correctly identified these true models. This highlights the robustness and an advantage of a good uniform design. That is, a good uniform design has nearly evenly-spaced design points throughout the experimental region, and this allows a researcher to fit and compare multiple models to identify the ones that best represent the mean response over the design region.

D-optimal Design for L/I model						D-optimal Design for Q/I model					
x_1	x_2	x_3	z_1	z_2	$y + \epsilon$	x_1	x_2	x_3	z_1	z_2	$y + \epsilon$
1	0	0	-1	-1	9.4978	1	0	0	-1	-1	10.3871
1	0	0	-1	-1	10.1315	0	1	0	-1	-1	11.4356
1	0	0	-1	-1	9.9211	0	0	1	-1	-1	10.8701
0	1	0	-1	-1	10.8868	0.5	0.5	0	-1	-1	10.0234
0	1	0	-1	-1	10.117	0.5	0.5	0	-1	-1	9.0781
0	1	0	-1	-1	10.3186	0.5	0	0.5	-1	-1	10.3187
0	0	1	-1	-1	9.4182	0	0.5	0.5	-1	-1	9.4191
0	0	1	-1	-1	10.7146	0	0.5	0.5	-1	-1	10.7155
0	0	1	-1	-1	9.1748	1	0	0	1	-1	9.1788
1	0	0	1	-1	9.6442	0	1	0	1	-1	9.6403
1	0	0	1	-1	10.0939	0	1	0	1	-1	10.0901
1	0	0	1	-1	10.1003	0	0	1	1	-1	10.1072
0	1	0	1	-1	9.7986	0.5	0.5	0	1	-1	10.0233
0	1	0	1	-1	10.74	0.5	0	0.5	1	-1	12.4022
0	1	0	1	-1	10.1236	0.5	0	0.5	1	-1	11.7858
0	0	1	1	-1	9.9816	0	0.5	0.5	1	-1	10.3416
0	0	1	1	-1	9.6221	0	0.5	0.5	1	-1	9.9821
0	0	1	1	-1	10.5218	1	0	0	-1	1	10.5109
1	0	0	-1	1	9.0862	0	1	0	-1	1	9.0864
1	0	0	-1	1	12.3103	0	1	0	-1	1	12.3105
1	0	0	-1	1	9.5619	0	0	1	-1	1	9.5728
0	1	0	-1	1	10.7643	0.5	0.5	0	-1	1	10.7682
0	1	0	-1	1	10.2622	0.5	0	0.5	-1	1	10.2924
0	1	0	-1	1	10.7736	0.5	0	0.5	-1	1	10.8039
0	0	1	-1	1	9.1966	0	0.5	0.5	-1	1	9.5566
0	0	1	-1	1	9.5725	1	0	0	1	1	9.5621
0	0	1	-1	1	9.2907	1	0	0	1	1	9.2803
1	0	0	1	1	10.2315	0	1	0	1	1	10.231
1	0	0	1	1	8.8428	0	0	1	1	1	8.8437
0	1	0	1	1	10.2471	0.5	0.5	0	1	1	10.2775
0	1	0	1	1	9.9089	0.5	0.5	0	1	1	9.9393
0	0	1	1	1	11.7589	0.5	0	0.5	1	1	11.9824
0	0	1	1	1	9.8635	0	0.5	0.5	1	1	9.9123

Table 3.1: 33-point D-optimal designs for L/I and Q/I models.

	Space-filling Design		D-optimal Design for L/I Model	D-optimal Design for Q/I Model
Model	Q/I	SC/I	L/I	Q/I
Multiple R-squared	0.9744	0.9999	0.3012	0.7174
Adjusted R-squared	0.9090	0.9995	-0.0649	-0.0047
Residual Error	0.6720	0.5329	0.7567	0.8767
AIC	73.510	47.835	86.337	92.089
BIC	102.923	91.233	105.791	129.502
P-value	0.001	<0.001	0.62	0.5372

Table 3.2: Summary table for fitted models for Example 1.

x_1	x_2	x_3	z_1	z_2	$y + \epsilon$
0.8998	0.0164	0.0839	-1	-1	9.4978
0.7693	0.0435	0.1872	1	-1	37.5585
0.6324	0.062	0.3056	-1	-1	9.922
0.4821	0.0689	0.449	1	-1	88.6671
0.2967	0.0568	0.6464	-1	1	16.8256
0.7818	0.1911	0.0272	1	1	11.6541
0.6729	0.2034	0.1237	-1	1	9.613
0.562	0.2056	0.2324	1	1	20.7635
0.4463	0.1945	0.3591	-1	1	12.0191
0.3183	0.1635	0.5182	1	-1	122.2672
0.1447	0.0868	0.7685	-1	-1	10.1411
0.5512	0.3842	0.0646	1	-1	28.98
0.462	0.3726	0.1655	-1	-1	9.8011
0.3727	0.347	0.2804	-1	-1	10.7485
0.2808	0.3016	0.4175	1	1	22.04
0.179	0.2219	0.5991	-1	1	31.5822
0.407	0.584	0.009	1	1	9.8382
0.3362	0.5603	0.1036	-1	1	11.6264
0.2682	0.5223	0.2094	1	-1	39.719
0.2028	0.4653	0.3318	-1	-1	12.3317
0.1386	0.3789	0.4825	1	-1	63.1106
0.0694	0.2294	0.7012	-1	-1	10.83
0.1875	0.7669	0.0457	1	-1	13.1101
0.1375	0.7181	0.1443	-1	1	13.3659
0.093	0.651	0.256	1	1	10.7547
0.0547	0.5577	0.3876	-1	1	18.2335
0.0237	0.4195	0.5568	1	1	10.8412
0.0024	0.1312	0.8664	1	1	10.571
0.8998	0.0164	0.0839	-1	-1	8.8423
0.7818	0.1911	0.0272	1	1	11.5825
0.1447	0.0868	0.7685	-1	-1	9.9601
0.1875	0.7669	0.0457	1	-1	14.6055
0.0024	0.1312	0.8664	1	1	10.2022

Table 3.3: 33-point Uniform design.

Term	Estimates	Std.Error	t value	Pr(> t)
x1	25.3427	1.1181	22.665	3.11e-06
x2	19.2781	2.4044	8.018	0.000488
x3	132.3976	2.0466	64.691	1.67e-08
x1:x2	-63.7579	6.3880	-9.981	0.000173
x1:x3	-217.9770	8.3621	-26.067	1.55e-06
x2:x3	-285.8027	7.7819	-36.727	2.82e-07
x1:z1	6.4313	1.1181	5.752	0.002228
x2:z1	0.2902	2.4044	0.121	0.908629
x3:z1	81.0739	2.0466	39.614	1.93e-07
x1:z2	-7.5112	1.1181	-6.718	0.001107
x2:z2	-11.4176	2.4044	-4.749	0.005111
x3:z2	-80.5702	2.0466	-39.368	1.99e-07
x1:x2:z1	-10.6622	6.3880	-1.669	0.155970
x1:x3:z1	-75.3139	8.3621	-9.007	0.000282
x2:x3:z1	-212.5340	7.7819	-27.311	1.23e-06
x1:x2:z2	38.7043	6.3880	6.059	0.001767
x1:x3:z2	92.8374	8.3621	11.102	0.000103
x2:x3:z2	223.7691	7.7819	28.755	9.53e-07
x1:z1:z2	-17.5107	1.1181	-15.661	1.93e-05
x2:z1:z2	-10.0223	2.4044	-4.168	0.008753
x3:z1:z2	-123.1029	2.0466	-60.150	2.40e-08
x1:x2:x3	1003.6975	34.2919	29.269	8.73e-07
x1:x2:z1:z2	66.9753	6.3880	10.485	0.000136
x1:x3:z1:z2	241.0381	8.3621	28.825	9.42e-07
x2:x3:z1:z2	291.9707	7.7819	37.519	2.53e-07
x1:x2:x3:z1	579.8874	34.2919	16.910	1.32e-05
x1:x2:x3:z2	-494.5808	34.2919	-14.423	2.89e-05
x1:x2:x3:z1:z2	-890.4322	34.2919	-25.966	1.58e-06

Table 3.4: Summary of SC/I model using uniform design points.

x_1	x_2	x_3	z_1	z_2	Force(lb)
0.9	0.05	0.05	-1	-1	44
0.9	0.05	0.05	1	-1	70
0.9	0.05	0.05	-1	1	19
0.9	0.05	0.05	1	1	33
0.7	0.1	0.2	-1	-1	48
0.7	0.1	0.2	1	-1	72
0.7	0.1	0.2	-1	1	32
0.7	0.1	0.2	1	1	59
0.75	0.05	0.2	-1	-1	32
0.75	0.05	0.2	1	-1	58
0.75	0.05	0.2	-1	1	21
0.75	0.05	0.2	1	1	38
0.85	0.1	0.05	-1	-1	51
0.85	0.1	0.05	1	-1	76
0.85	0.1	0.05	-1	1	22
0.85	0.1	0.05	1	1	49
0.8	0.07	0.13	-1	1	17
0.73	0.07	0.2	1	-1	69
0.73	0.07	0.2	-1	-1	40
0.77	0.1	0.13	-1	-1	37
0.8	0.07	0.13	1	1	46
0.77	0.1	0.13	-1	1	21
0.77	0.1	0.13	1	-1	82
0.77	0.1	0.13	1	1	43
0.82	0.05	0.13	-1	-1	32
0.82	0.05	0.13	1	-1	60
0.85	0.06	0.09	-1	1	14
0.85	0.06	0.09	1	1	38
0.82	0.09	0.09	-1	-1	45
0.77	0.1	0.13	-1	1	18
0.73	0.07	0.2	1	-1	70
0.85	0.1	0.05	-1	1	10
0.77	0.1	0.13	1	1	52
0.85	0.1	0.05	1	1	42

Table 3.5: 34-point D-optimal design for Example 2.

	Data Generation Model: Q/I Model			
	Space-filling Design		D-optimal Design	
Fitted Model	Q/I Model	SC/I Model	Q/I Model	SC/I Model
Adjusted R^2	0.9709	0.823	0.9681	0.8888
Multiple R^2	0.9912	0.8740	0.9903	0.9209
Residual Error	3.697	5.363	3.509	4.827
Maximum Predicted Force(lbs.)	84.776	80.03	80.948	74
Optimum Design points	(0.8048,0.0836, 0.115,1, -1)	(0.8048,0.0836, 0.115,1, -1)	(0.85,0.1, 0.05,1, -1)	(0.85,0.1, 0.05,1, -1)

	Data Generation Model: SC/I Model			
	Space-filling Design		D-optimal Design	
Fitted Model	Q/I Model	SC/I Model	Q/I Model	SC/I Model
Adjusted R^2	0.8549	0.9541	0.871	0.9853
Multiple R^2	0.9010	0.9918	0.9145	0.9973
Residual Error	5.603	5.095	3.346	5.527
Maximum Predicted Force(lbs.)	75	84	80	83
Optimum Design points	(0.8048,0.0836, 0.115,1, -1)	(0.8048,0.0836, 0.115,1, -1)	(0.85,0.1, 0.05,1, -1)	(0.85,0.1, 0.05,1, -1)

Table 3.6: Summary of fitted models for Example 2.

D-optimal Design						Uniform Design					
x_1	x_2	x_3	z_1	z_2	Force	x_1	x_2	x_3	z_1	z_2	Force
0.9	0.05	0.05	-1	-1	43	0.877	0.051	0.072	1	-1	66
0.9	0.05	0.05	1	-1	69	0.851	0.053	0.096	-1	-1	36
0.9	0.05	0.05	-1	1	18	0.824	0.054	0.122	1	-1	67
0.9	0.05	0.05	1	1	37	0.796	0.056	0.148	-1	-1	36
0.7	0.1	0.2	-1	-1	47	0.767	0.058	0.175	1	-1	66
0.7	0.1	0.2	1	-1	76	0.888	0.059	0.052	-1	1	19
0.7	0.1	0.2	-1	1	29	0.862	0.061	0.077	-1	1	13
0.7	0.1	0.2	1	1	62	0.836	0.063	0.101	1	1	44
0.75	0.05	0.2	-1	-1	29	0.808	0.065	0.127	-1	1	12
0.75	0.05	0.2	1	-1	54	0.780	0.066	0.153	1	-1	71
0.75	0.05	0.2	-1	1	21	0.751	0.068	0.180	-1	-1	38
0.75	0.05	0.2	1	1	38	0.873	0.070	0.057	-1	-1	49
0.85	0.1	0.05	-1	-1	50	0.847	0.072	0.081	1	-1	78
0.85	0.1	0.05	1	-1	81	0.820	0.073	0.106	-1	-1	43
0.85	0.1	0.05	-1	1	17	0.793	0.075	0.132	1	1	48
0.85	0.1	0.05	1	1	45	0.765	0.077	0.159	-1	1	20
0.8	0.07	0.13	-1	1	14	0.736	0.078	0.186	1	1	54
0.73	0.07	0.2	1	-1	72	0.858	0.080	0.062	1	1	50
0.73	0.07	0.2	-1	-1	37	0.832	0.082	0.086	-1	1	11
0.77	0.1	0.13	-1	-1	51	0.805	0.084	0.112	1	-1	91
0.8	0.07	0.13	1	1	44	0.777	0.085	0.137	-1	-1	38
0.77	0.1	0.13	-1	1	23	0.749	0.087	0.164	1	-1	81
0.77	0.1	0.13	1	-1	79	0.720	0.089	0.192	-1	-1	45
0.77	0.1	0.13	1	1	51	0.843	0.091	0.067	-1	-1	51
0.82	0.05	0.13	-1	-1	26	0.816	0.092	0.091	1	1	44
0.82	0.05	0.13	1	-1	62	0.789	0.094	0.117	-1	1	15
0.85	0.06	0.09	-1	1	12	0.762	0.096	0.143	1	1	47
0.85	0.06	0.09	1	1	40	0.733	0.097	0.169	-1	1	26
0.82	0.09	0.09	-1	-1	38	0.704	0.099	0.197	1	1	53
0.77	0.1	0.13	-1	1	21	0.877	0.051	0.072	1	-1	69
0.73	0.07	0.2	1	-1	69	0.888	0.059	0.052	-1	1	17
0.85	0.1	0.05	-1	1	24	0.751	0.068	0.180	-1	-1	46
0.77	0.1	0.13	1	1	47	0.805	0.084	0.112	1	-1	79
0.85	0.1	0.05	1	1	45	0.704	0.099	0.197	1	1	58

Table 3.7: Simulated data using coefficients from quadratic mixture model and interaction process variable model.

D-optimal Design						Uniform Design					
x_1	x_2	x_3	z_1	z_2	Force	x_1	x_2	x_3	z_1	z_2	Force
0.9	0.05	0.05	-1	-1	42	0.8775	0.0509	0.0717	1	-1	64
0.9	0.05	0.05	1	-1	67	0.8510	0.0526	0.0964	-1	-1	34
0.9	0.05	0.05	-1	1	18	0.8239	0.0543	0.1218	1	-1	68
0.9	0.05	0.05	1	1	32	0.7960	0.0560	0.1479	-1	-1	34
0.7	0.1	0.2	-1	-1	51	0.7673	0.0578	0.1749	1	-1	69
0.7	0.1	0.2	1	-1	76	0.8881	0.0595	0.0524	-1	1	20
0.7	0.1	0.2	-1	1	23	0.8622	0.0612	0.0766	-1	1	13
0.7	0.1	0.2	1	1	49	0.8357	0.0629	0.1014	1	1	41
0.75	0.05	0.2	-1	-1	24	0.8084	0.0647	0.1270	-1	1	3
0.75	0.05	0.2	1	-1	57	0.7803	0.0664	0.1533	1	-1	74
0.75	0.05	0.2	-1	1	14	0.7515	0.0681	0.1804	-1	-1	36
0.75	0.05	0.2	1	1	35	0.8730	0.0698	0.0572	-1	-1	54
0.85	0.1	0.05	-1	-1	49	0.8470	0.0716	0.0815	1	-1	76
0.85	0.1	0.05	1	-1	83	0.8203	0.0733	0.1065	-1	-1	46
0.85	0.1	0.05	-1	1	14	0.7928	0.0750	0.1322	1	1	39
0.85	0.1	0.05	1	1	32	0.7646	0.0767	0.1586	-1	1	6
0.8	0.07	0.13	-1	1	10	0.7356	0.0784	0.1860	1	1	41
0.73	0.07	0.2	1	-1	59	0.8578	0.0802	0.0620	1	1	35
0.73	0.07	0.2	-1	-1	42	0.8317	0.0819	0.0864	-1	1	13
0.77	0.1	0.13	-1	-1	49	0.8048	0.0836	0.1115	1	-1	87
0.8	0.07	0.13	1	1	26	0.7773	0.0853	0.1374	-1	-1	29
0.77	0.1	0.13	-1	1	18	0.7489	0.0871	0.1640	1	-1	80
0.77	0.1	0.13	1	-1	70	0.7196	0.0888	0.1916	-1	-1	38
0.77	0.1	0.13	1	1	44	0.8427	0.0905	0.0668	-1	-1	56
0.82	0.05	0.13	-1	-1	37	0.8164	0.0922	0.0914	1	1	45
0.82	0.05	0.13	1	-1	59	0.7894	0.0940	0.1167	-1	1	8
0.85	0.06	0.09	-1	1	11	0.7617	0.0957	0.1427	1	1	36
0.85	0.06	0.09	1	1	33	0.7331	0.0974	0.1695	-1	1	17
0.82	0.09	0.09	-1	-1	42	0.7037	0.0991	0.1972	1	1	40
0.77	0.1	0.13	-1	1	25	0.8775	0.0509	0.0717	1	-1	79
0.73	0.07	0.2	1	-1	68	0.8881	0.0595	0.0524	-1	1	17
0.85	0.1	0.05	-1	1	10	0.7515	0.0681	0.1804	-1	-1	35
0.77	0.1	0.13	1	1	38	0.8048	0.0836	0.1115	1	-1	81
0.85	0.1	0.05	1	1	41	0.7037	0.0991	0.1972	1	1	43

Table 3.8: Simulated data assuming special cubic mixture model and interaction process variable model.

CHAPTER FOUR

SPACE-FILLING DESIGN AUGMENTATION IN THE SIMPLEX

In this chapter, we will consider the situation in which a researcher has run an initial mixture experiment, perhaps based on an educated guess or on historical data from a similar system, and it is felt that the design properties can be improved by adding space-filling runs to the initial design. Box and Wilson (1951) are credited with the notion that experimentation should preferably be a sequential process. Current methods (Hinrichs et al., 2020; Nachtsheim and Jones, 2018; Piepel et al., 2002, 2005; Sall et al., 2017) of augmenting a mixture experiment with new runs are based on certain optimality criteria. Such model-dependent designs have drawbacks, as discussed in Chapter 2, so this Chapter introduces a space-filling approach that spreads the new runs evenly across the design region, independent of any model assumption. The idea is that the gathering of information through successive experiments and the resulting space-filling design help in exploring alternative models that may better approximate the true response surface or discovering new regions in which it is possible to enhance or refine the process or response output.

Therefore, the impetus behind a space-filling design augmentation approach largely falls into three key categories:

- Adding new runs to explore alternative models that may not be possible with the existing design points: The current design region is not altered in this situation, but additional runs are added to explore the design region better and seek alternative models that are better approximated over the design region. For instance, 4 or 5 interior points may be added to the 6 boundary points of a $\{3, 2\}$ -simplex lattice design so that additional models such as a special cubic

model or cubic model can be fit.

- Exploring new regions based on current run knowledge to enhance the process or optimize the response output: In this scenario, the analysis based on the current information is used to identify a new region in which the process or response can be improved. This involves shrinking the existing design region and concentrating on a smaller region in which the response may be improved, expanding the existing design region to include new regions in which the process or output can be improved, or moving along the direction of steepest ascent to establish a new region in which the process is enhanced. The existing design region is altered in this case, and an illustration is shown in Figure 4.1 with a 3-component mixture experiment where a new region (green-shaded) is identified from the default full simplex region (blue-shaded) and new runs are added to the existing design points (red points) in this new region.

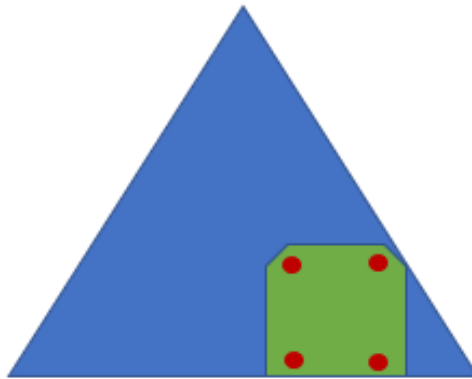


Figure 4.1: New design region (green) defined from existing design region (blue).

- As part of the model building process to assess the predictive power of a model via model validation: Model validation tests the accuracy of a model's predictions relative to independent sets of data called validation sets that were

not used in the construction of the model (Anderson and Bates, 2001). New space-filling design points can be used to form the validation sets to ensure that the prediction prowess of a model throughout the design region can be properly assessed.

4.1 New Space-filling Design Augmentation Approach

The goal is to add n new runs to k existing design points in the full simplex or sub-space of the simplex such that the resulting design with $n + k$ runs has good space-filling properties. The proposed method utilizes the procedure for constructing uniform designs for a q -component mixture experiment described in Section 2.1 and is summarized as follows:

1. Generate $n + m$ candidate points in the $(q - 1)$ -dimensional simplex or sub-space of the simplex which has k existing design points using the method presented by Borkowski and Piepel (2009) and described in Section 2.1.
2. Delete the new m points nearest to the existing k design points. $1 \leq m \leq k$ is recommended for computational efficiency, and the idea is that the deleted m points are replaced with the nearest existing k points such that the resulting design becomes a uniform design with $n + k$ points.
3. An Euclidean Minimum Spanning Tree (EMST) for the $n + k$ design points is then constructed, and the best augmented space-filling designs are designs with large $GMST$ values.

4.1.1 Augmenting Boundary Mixture Design Points

This new space-filling approach is illustrated for a 3-component mixture experiment where the goal is to augment a $\{3, 2\}$ -simplex lattice design and a 7-

point extreme vertices design for the experimental region shown in Figure 4.2 with 9 interior points.

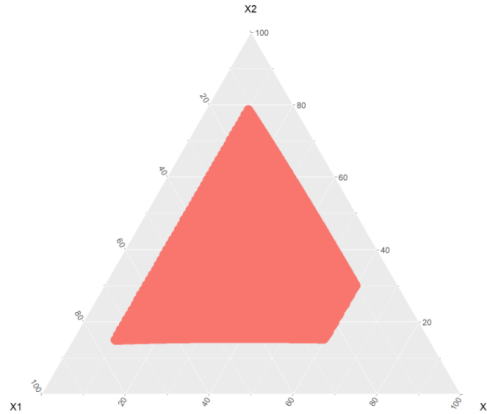


Figure 4.2: Constrained design region for 3-component mixture experiment.

Figure 4.3 shows the results from using this new algorithm to add 9 new interior points to the 6 boundary points of a $\{3, 2\}$ -simplex lattice design using three number theoretic methods (Good Lattice Point (GLP), Square Root Sequence (SRS), and Powers of a Prime (PP)). *GMST* values are shown on the y-axis and the different generators used are shown on the x-axis. The 10 best generators for each number of candidate points ($9 + m$) were selected using the SRS and PP methods, while all possible generators were used using the GLP method for each number of candidate points ($9 + m$). There is a lot of variability in the space-filling properties of augmented designs generated by the GLP and the PP methods. The generators using the SRS method tend to produce augmented designs with the largest *GMST* values for any given number of candidate points and consequently the better space-filling augmented designs. For all three methods, the space-filling properties of the augmented designs do not seem to decrease or increase with increasing or decreasing number of candidate points $9 + m$, that is, there does not appear to be any discernible relationship between *GMST* values and the number of candidate points $9 + m$.

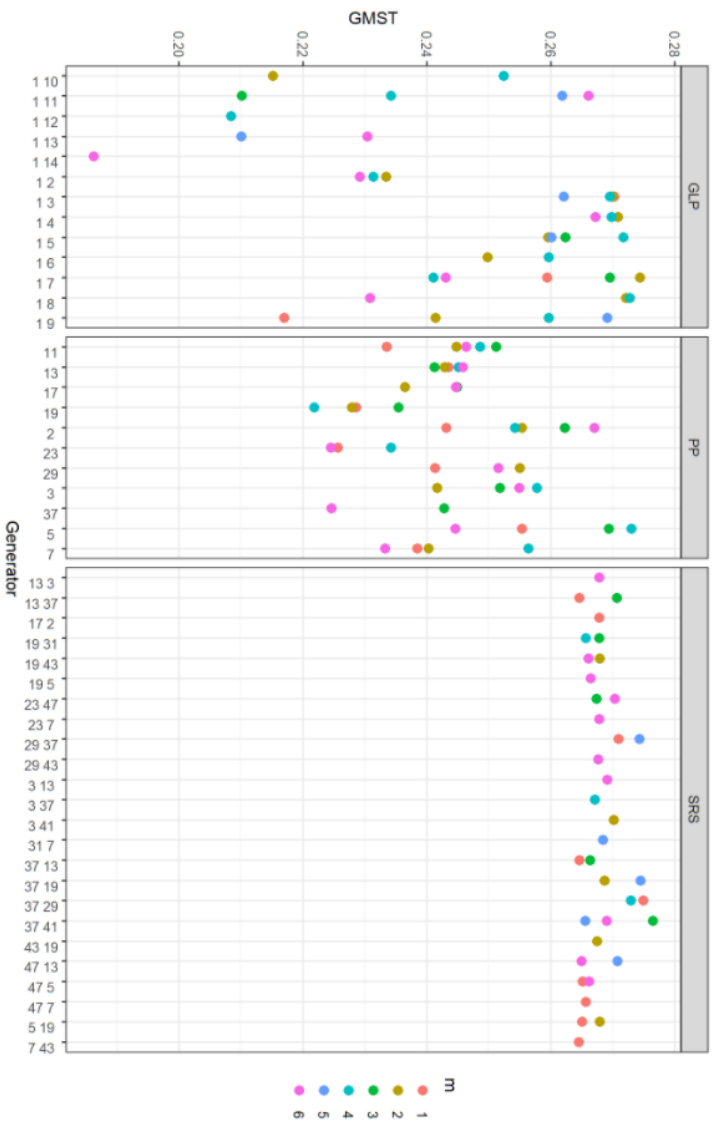


Figure 4.3: Results from adding 9 interior points to a $\{3, 2\}$ -simplex lattice design.

Figure 4.5 outlines the new proposed procedure as well as the 3 best augmented 15-point designs generated by each number theoretical method (GLP, SRS, PP). Initially, 11 candidate points are generated using the GLP method with generator (1, 7), 12 candidate points are generated using the SRS method with generator (37, 41), and 13 candidate points are generated using the PP method with generator ($p = 5$). The nearest m points to the existing points are then discarded. In Figure 4.5, these are the circled points. For instance, from the initial 12 points created by the SRS method, the $m = 3$ points closest to the existing points are removed. The resulting design is a 15-point good space-filling augmented design with 6 existing points and 9 added new runs. The existing and added runs are labeled as old and new points, respectively.

The results of adding 9 new runs to the 7-point extreme vertices design are also

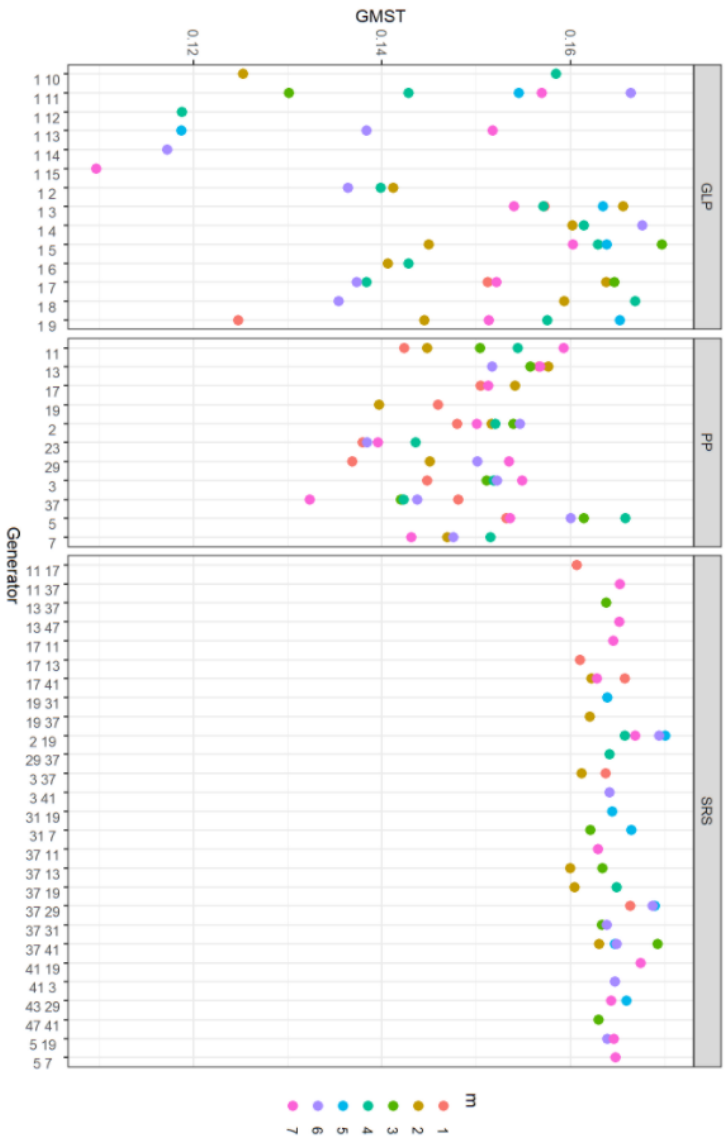


Figure 4.4: Results from adding 9 interior points to a 7-point extreme vertices design.

presented in Figure 4.4. The results are similar to the previous example. Augmented designs generated using the SRS method have consistently high $GMST$ values, and there is a lot of variability in the space-filling quality of designs generated using the GLP and PP methods. The design generation process and the best augmented design from each method are shown in Figure 4.6. For the GLP method, the augmented design from adding 9 points obtained by the generator (12; 1, 5) is the best space-filling design, whereas generators (14; 2, 19) and (13; 5) produced the best augmented designs for the SRS and PP methods, respectively. The circled points are the nearest points to existing points that are removed, resulting in 9 new runs. Visually, for these 3 designs, the design points are evenly spread across the constrained simplex region, signaling that these are good space-filling designs.

4.1.2 Augmenting Boundary and Interior Mixture Design Points

We will now consider the scenario where there are existing interior points or existing interior and boundary points and the goal is to augment these existing points with new runs such that the resulting augmented design has good space-filling properties. We will look at the following situations: generating an augmented 15-point space-filling design by adding 9 new runs to an existing 6-point interior design, and generating an augmented 17-point design by also adding 9 new runs to an extreme vertices design with 8 existing points for the design region in Figure 4.2.

The results from augmenting the 6-point design and the 7-point design with 9 new points are presented in Figure 4.7a, and Figure 4.7b, respectively. There is a lot of variability in the space-filling quality of designs produced by the GLP and PP methods, and the space-filling quality of designs by these two methods also do not increase or decrease with increasing the number of candidate points $(9 + m)$, as seen from the previous examples. Using the GLP method, the best 15-point and the best 17-point augmented designs are generated by generators $(13; 1, 8)$ and $(15; 1, 4)$, respectively (4.8a). For the PP method, generators $(15; 13)$ and $(13; 5)$ respectively produce the best 15-point and best 17-point augmented designs (Figure 4.8c). The SRS method appears to generate consistently better space-filling designs than the other two methods, and the space-filling quality of the augmented designs seem to improve as the number of candidate points $(9 + m)$ increase. In the first scenario in Figure 4.7a for example, augmented designs obtained by adding 9 points out of 13 ($m = 4$), 14 ($m = 5$), and 15 ($m = 6$) candidate points have higher *GMST* values than designs generated by adding 9 points out of 10 ($m = 1$), 11 ($m = 2$), and 12 ($m = 3$) candidate points. The best overall augmented 15-point space-filling design (Figure 4.8b) is obtained by adding 9 new points out of 13 ($m = 4$) candidate points using the SRS method with generator $(29, 37)$. In the second scenario also,

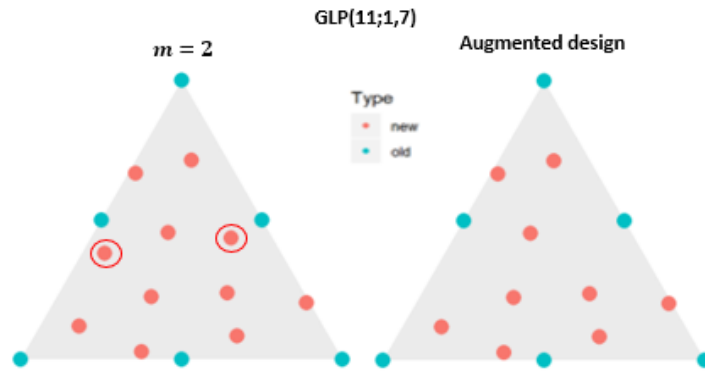
augmented designs obtained by adding 9 new points out of a greater number of candidate points are better space-filling designs than designs obtained from a smaller number of candidate points using the SRS method (Figure 4.7b). The best overall 17-point augmented design is obtained using the generator $(14; 2, 19)$ using the SRS method (Figure 4.8b).

4.1.3 Augmenting Mixture Designs with Existing Cluster of Points

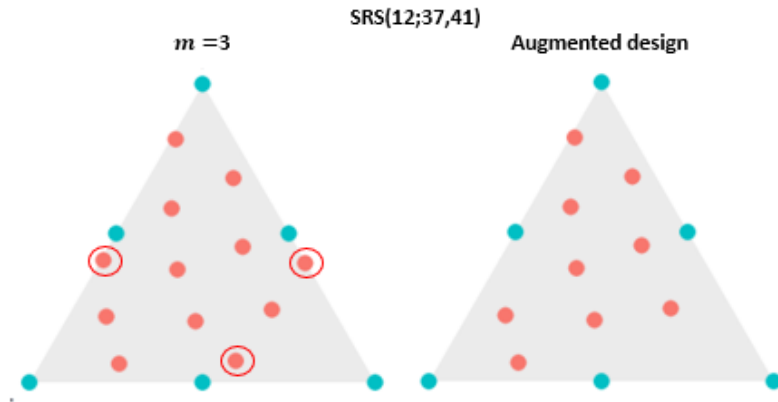
In certain situations, the existing design points may be clustered, and the points in a cluster may share the same nearest neighbor such that a small *GMST* would automatically be obtained by keeping all points in a cluster, making it difficult to differentiate between good and bad space-filling designs. The following method is a modified version of the proposed algorithm to account for the existing clusters so that it is possible to properly measure the space-filling properties of the resulting augmented designs. If the existing k points have t distinct points outside the clusters and i clusters of points that may share the same nearest neighbor, then the centroids of these clusters are used such that the m ($1 \leq m \leq t + i$) new points nearest to the resulting $i + t$ existing points are removed. The Euclidean Minimum Spanning Tree is then constructed for the $(n + i + t)$ points to pick the best space-filling design, but the actual statistical analysis may be done using the $n + k$ points. What constitutes a cluster is dependent on the research and design, but as a rule of thumb, a cluster may be described as a group of points with maximum pairwise distances less than 5% of the design region.

For example, the 9-point design in Figure 4.9a for a 3-component mixture experiment in the full simplex has $i = 3$ clusters and 0 distinct points outside the clusters. Using the proposed algorithm, the centroids of these 3 clusters are used, resulting in the 3-point design in Figure 4.9b. Suppose there are resources to augment

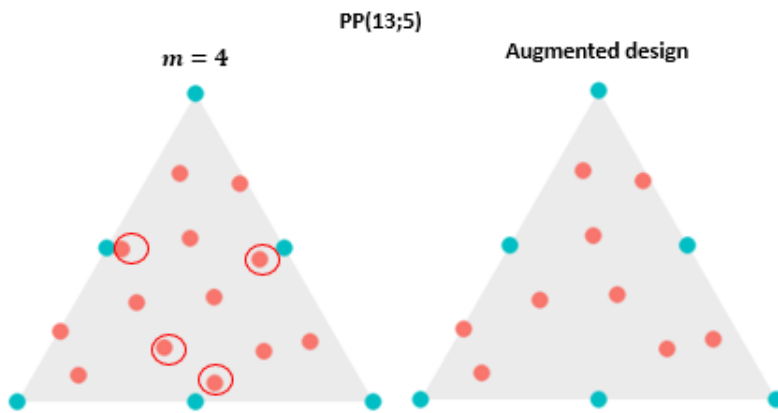
the existing design with 9 new points; then 12 candidate points are added to the 3 centroids using the SRS method with generator (43, 29) in Figure 4.10a. The 3 closest points (the circled points in 4.10a) to the 3 centroids are removed, and the ensuing design is a 12-point design with 9 new runs (Figure 4.10b). The final space-filling augmented design with 9 existing points and 9 added points is shown in Figure 4.11.



(a) Best design using the GLP method.

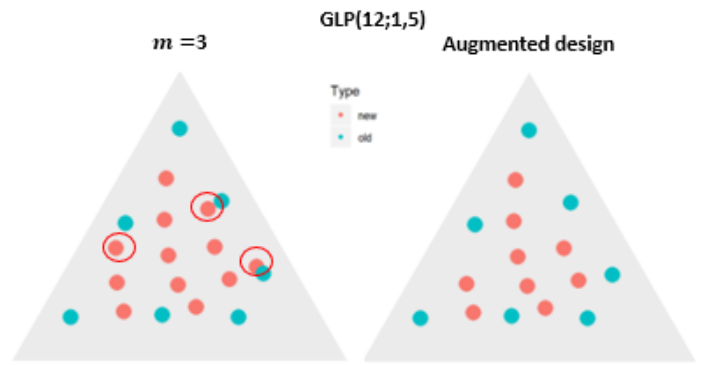


(b) Best design using the SRS method.

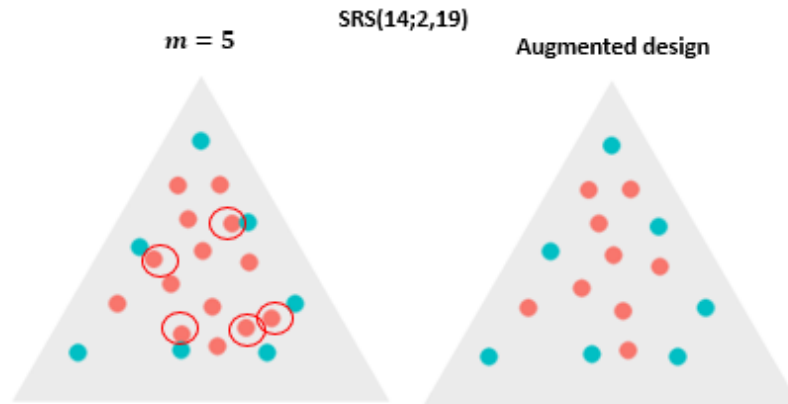


(c) Best design using the PP method.

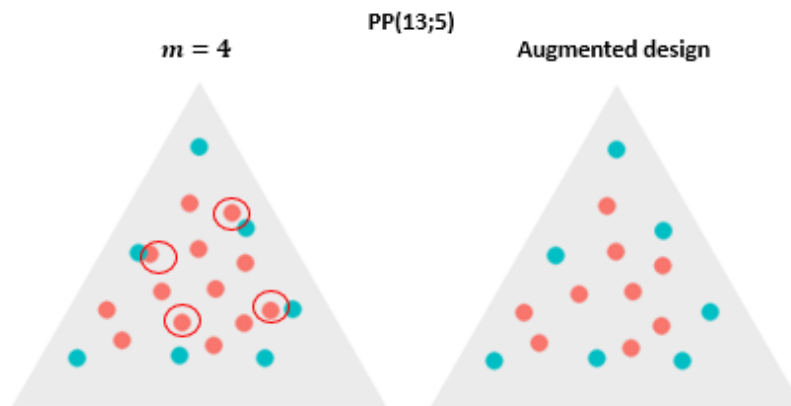
Figure 4.5: Best augmented designs from adding 9 points to $\{3, 2\}$ -simplex lattice design.



(a) Best design using the GLP method.

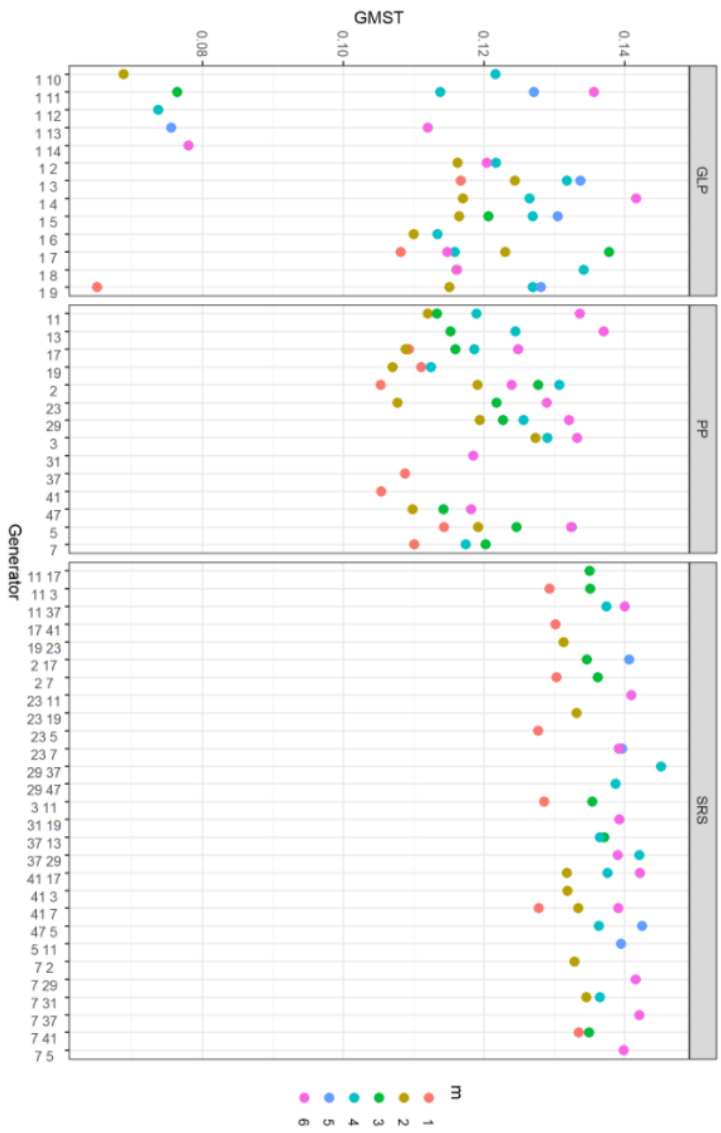


(b) Best design using the SRS method.

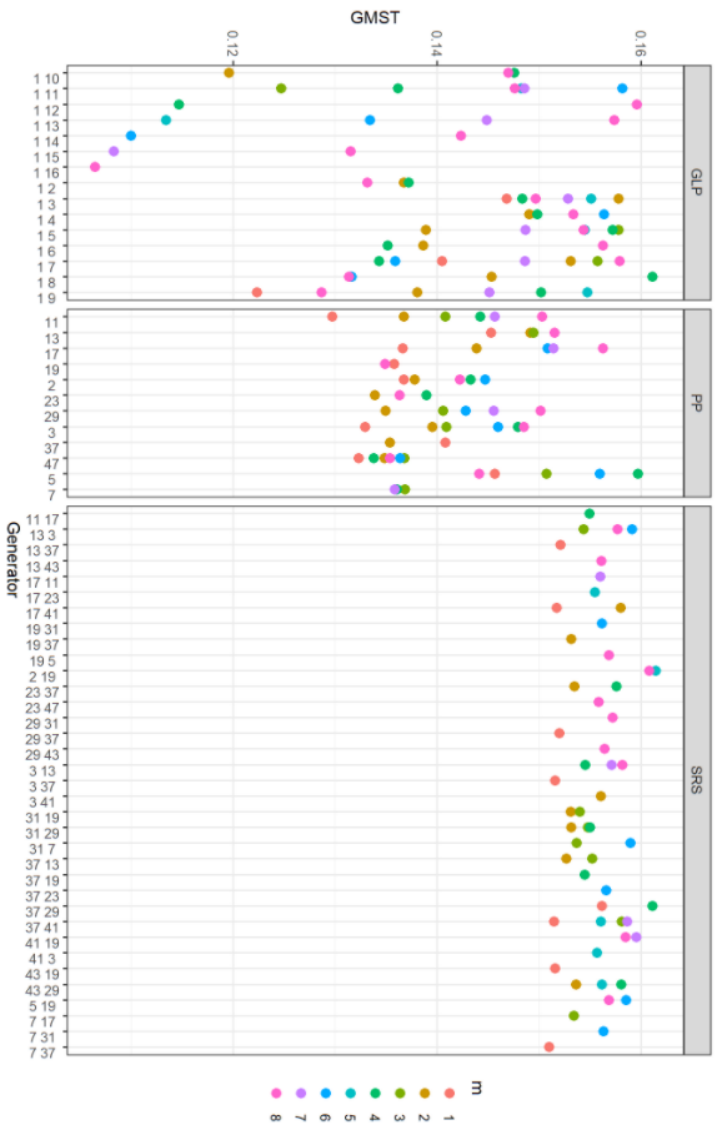


(c) Best design using the PP method.

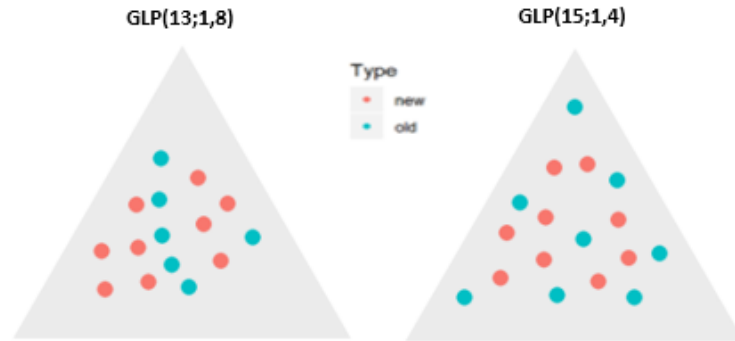
Figure 4.6: Best augmented designs from adding 9 points to 7-point extreme vertices design.



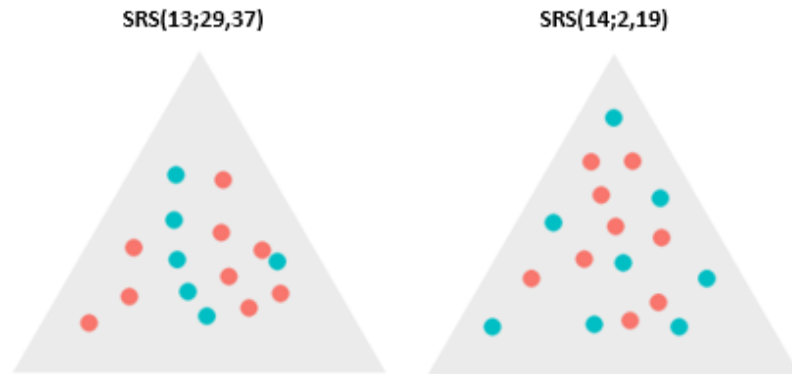
(a) Results from adding 9 points to a 6-point design.



(b) Results from adding 9 points to an 8-point extreme vertices design.



(a) Best augmented designs by the GLP method.



(b) Best augmented designs by the SRS method.



(c) Best augmented designs by the PP method.

Figure 4.8: Best augmented designs from adding 9 points to a 6-point design and 8-point extreme vertices design.

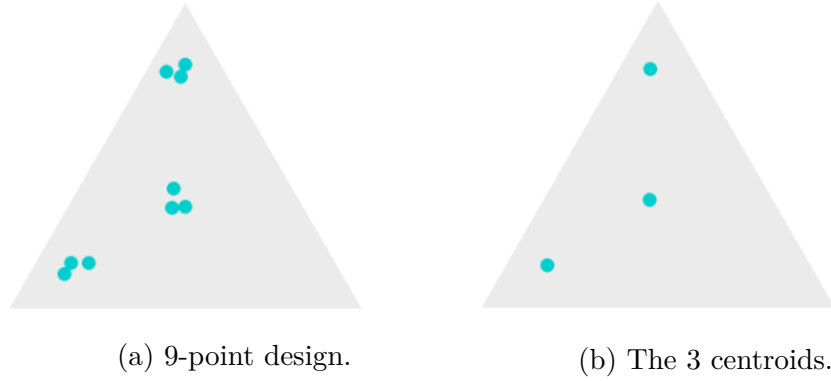


Figure 4.9: 9-point mixture design for 3 component experiment with 3 clusters.

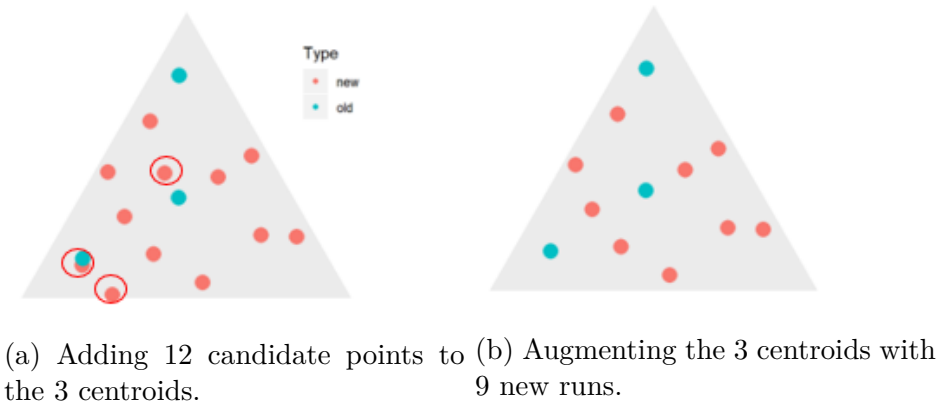


Figure 4.10: Augmenting 9-point design with 9 new runs.



Figure 4.11: Augmented 18-point design consisting of 9 existing points and 9 new runs.

CHAPTER FIVE

CONCLUSIONS AND FUTURE RESEARCH

The cost of classical designs, such as orthogonal arrays, factorial and fractional designs, which are defined in terms of the combinatorial structure of the number of experimental factors and levels, may surpass the available resources when dealing with experiments with a large number of factors and levels. In such instances, particularly when the underlying model form is unknown, an excellent alternative is to use model-independent uniform designs. There are many methods available for constructing uniform designs, and four of such methods have been used in this study. These are the Good Lattice Point (GLP) method, the cyclotomic field (CF) method, the square root sequence (SRS) method, and the power-of-a-prime (PP) method, discussed in Chapter One.

In Chapter Two, an existing algorithm which uses the G function introduced by Fang and Yang (2000) and adopted by Borkowski and Piepel (2009) to map points from the hypercube to the simplex or sub-space of the space is explored and implemented. While the G function usually preserves the design patterns from the hypercube, it has been found that, due to the nonlinear transformation of the G function, there is no preservation of systematic patterns of uniformity across regions in the simplex. Graphically, mixture design points for the best designs generated by the GLP method were found to be more uniformly scattered than the points in the best designs generated by the other three methods.

The ultimate objective of this dissertation was to present a statistical methodology and a design algorithm for uniform designs for the combined mixture/process variable experiment. Current research has explored uniform designs for process

variables in hypercuboidal and hyperspherical designs, and for mixture experiments in the simplex or constrained subspaces of the simplex. To date, no research has been published for space-filling designs for the combined mixture/process variable experiment. A new hybrid algorithm was presented for generating uniform designs for mixture/process variable experiments. For the q -component mixture experiment with k process variable experiment, the algorithm uses the G function introduced by Fang and Yang (2000), and adopted by Borkowski and Piepel (2009) to map q points from $q + k$ points generated in the hypercube to the simplex or sub-space of the simplex as the mixture design points, leaving the k points as the coded design points for the process variables.

Uniformity measures, including the three distance-based measures for assessing uniformity in the simplex used by Borkowski and Piepel (2009) were discussed in Chapter One. Two new criteria based on the Euclidean Minimum Spanning Tree (EMST) which are more computationally efficient for assessing uniformity of mixture designs and mixture/process variable designs were presented in Chapter Two. The two criteria are (i) the arithmetic mean and standard deviation of edge lengths criterion ($adMST, sdMST$), and (ii) the geometric mean of edge lengths criterion ($GMST$). The two criteria were found to be interchangeable, and the preferred criterion in this study was the $GMST$ criterion, which uses only one statistic to quantify the uniformity properties of mixture and mixture/process variable designs. A large $GMST$ value relates to large similar interpoint-distances between all design points, which means that design is a good uniform design, and the design points adequately fill the design space. Tables of good uniform designs are provided for mixture experiments in the full simplex (S^q) for $q = 3, 4, 5$ and practical design sizes, $9 \leq n \leq 30$, using the four number theoretic methods used in this study. It was found that the GLP and SRS methods generate mixture designs with superior $GMST$ values

in comparison to the other methods.

A conditional approach based on the *GMST* criterion for generating good uniform designs for mixture/process experiments was presented in Chapter Two. Tables of good uniform designs produced by the four number theoretic methods were produced for the combined q mixture and k process variable experiments for $q = 3, 4, 5$, $k = 1, 2$ and practical numbers of runs, $9 \leq n \leq 30$. The GLP and SRS were consistently found to be the best methods for generating mixture/process variable designs.

In Chapter Three, examples comparing data analyses for data from uniform designs and computer generated optimal designs was provided. The data analyses were done using the *lm* function in R. There are two simulated data examples highlighting the advantages and robustness of uniform designs, particularly in mixture and mixture/process variable experiments.

Finally, a new algorithm is provided in Chapter Four to augment existing mixture design points with space-filling points, including designs with existing clustered design points. The SRS method was found to produce the best augmented space-filling mixture designs, and, in some design regions, it was found that the space-filling properties of augmented designs may depend on the number of candidate points initially generated.

The best uniform design for an experiment, as seen in this study, depends on the design region, design size, the number theoretic method used, and the available design generators. An area of future research is finding a local optimization algorithm that may be used to improve the best possible uniform design generated for an experiment. That is, a local search method that may be used to obtain a better design from the best possible uniform design generated by searching and moving one or more design points around the existing design points in the design region until the optimal space-

filling design is found.

The G function works by mapping points in the hypercube to the simplex for a single component restriction mixture experiment. Borkowski and Piepel (2009) presented an inclusion-exclusion approach that deals with multiple constraint mixture experiments by selecting the design points from a set of candidate points that fall into the restricted design area. There is no function such as the G function, however, that transforms design points from the hypercube to the simplex for multiple constraint mixture experiments. Therefore, finding an equivalent function that can be used for multiple constraint mixture experiments to map points to the simplex from the hypercube is a good problem for future study.

The design augmentation method presented in this study works for the constrained and unconstrained mixture experiments. A future research area is extending this algorithm to mixture/process variable experiments.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Aguiar, P. F. de, Bourguignon, B, Khots, M., Massart, D., and Phan-Than-Luu, R (1995). “D-optimal designs”. *Chemometrics and Intelligent Laboratory Systems* 30.2, pp. 199–210.
- Anderson, M. and Bates, P. (2001). “Model Validation”. *Perspectives in Hydrological Science. John Wiley & Sons*.
- Anderson-Cook, C. M., Goldfarb, H. B., Borror, C. M., Montgomery, D. C., Canter, K. G., and Twist, J. N. (2004). “Mixture and mixture–process variable experiments for pharmaceutical applications”. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry* 3.4, pp. 247–260.
- Bingham, D., Sitter, R. R., and Tang, B. (2009). “Orthogonal and nearly orthogonal designs for computer experiments”. *Biometrika* 96.1, pp. 51–65.
- Borkowski, J. J. (2018). *Response Surface Lecture Notes (Unpublished Lecture Notes), Spring 2018*.
- Borkowski, J. J. and Piepel, G. F. (2009). “Uniform Designs for Highly Constrained Mixture Experiments”. *Journal of Quality Technology* 1, pp. 35–47.
- Box, G. E. and Behnken, D. W. (1960). “Some new three level designs for the study of quantitative variables”. *Technometrics* 2.4, pp. 455–475.
- Box, G. E. and Draper, N. R. (1987). *Empirical Model-Building and Response Surfaces*. John Wiley & Sons, New York.
- Box, G. E. and Wilson, K. B. (1951). “On the experimental attainment of optimum conditions”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 13.1, pp. 1–38.
- Box, G. E. and Draper, N. R. (2007). *Response Surfaces, Mixtures, and Ridge Analyses*. Vol. 649. John Wiley & Sons, New York.

- Butler, N. A. (2001). “Optimal and orthogonal Latin hypercube designs for computer experiments”. *Biometrika* 88.3, pp. 847–857.
- Cheng, C.-S. (2006). *Projection Properties of Factorial Designs for Factor Screening*. Springer, New York, pp. 156–168.
- Chung, P. J., Goldfarb, H. B., and Montgomery, D. C. (2007). “Optimal designs for mixture-process experiments with control and noise variables”. *Journal of Quality Technology* 39.3, pp. 179–190.
- Coetzer, R. and Haines, L. M. (2017). “The Construction of D - and I -Optimal Designs for Mixture Experiments with Linear Constraints on the Components”. *Chemometrics and Intelligent Laboratory Systems*, pp. 112–124.
- Cooley, S. K. and Piepel, G. F. (2003). “A Two-Stage Layered Mixture Experiment Design for a Nuclear Waste Glass Application-Parts 1 and 2”. In: *Proceedings of the American Statistical Association*, pp. 1044–1051.
- Cornell, J. A. (2011). *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*. Vol. 403. John Wiley & Sons, New York.
- Cornell, J. A. (1971). “Process variables in the mixture problem for categorized components”. *Journal of the American Statistical Association* 66.333, pp. 42–48.
- Curtin, R. R., Cline, J. R., Slagle, N. P., March, W. B., Ram, P., Mehta, N. A., and Gray, A. G. (2013). “MLPACK: A scalable C++ machine learning library”. *Journal of Machine Learning Research* 14.Mar, pp. 801–805.
- Czitrom, V. (1988). “Mixture experiments with process variables: D-optimal orthogonal experimental designs”. *Communications in Statistics-Theory and Methods* 17.1, pp. 105–121.

- Damblin, G., Couplet, M., and Iooss, B. (2013). “Numerical studies of space-filling designs: optimization of Latin Hypercube Samples and subprojection properties”. *Journal of Simulation* 7.4, pp. 276–289.
- De Ketelaere, B., Goos, P., and Brijs, K. (2011). “Prespecified factor level combinations in the optimal design of mixture-process variable experiments”. *Food Quality and Preference* 22.7, pp. 661–670.
- Dean, A., Voss, D., and Draguljić, D. (2017). *Response Surface Methodology*. Springer, New York, pp. 565–614.
- Dussert, C, Rasigni, G, Rasigni, M, Palmari, J, and Llebaria, A (1986). “Minimal spanning tree: A new approach for studying order and disorder”. *Physical Review B* 34.5, p. 3528.
- Fang, K.-T. (1980). “Uniform design : application of number-theoretic methods in experimental design”. *Acta Mathematicae Applicatae Sinka* 3, pp. 363–372.
- Fang, K.-T. and Lin, D. K. (2003). “Ch. 4. Uniform experimental designs and their applications in industry”. *Handbook of Statistics*, pp. 131–170.
- Fang, K.-T. and Yang, Z.-H. (2000). “On uniform design of experiments with restricted mixtures and generation of uniform distribution on some domains”. *Statistics & Probability Letters* 46.2, pp. 113–120.
- Fang, K. and Wang, Y. (1994). *Number-Theoretic Methods in Statistics*. Chapman & Hall, New York, p. 340.
- Fang, K., Li, R.-z., and Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC, New York, p. 290.
- Franco, J, Vasseur, O, Corre, B, and Sergent, M (2009). “Minimum Spanning Tree: A new approach to assess the quality of the design of computer experiments”. *Chemometrics and Intelligent Laboratory Systems* 97.2, pp. 164–169.

- Gerber, M. and Chopin, N. (2015). “Sequential quasi monte carlo”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77.3, pp. 509–579.
- Goldfarb, H. B., Borror, C. M., and Montgomery, D. C. (2003). “Mixture-process variable experiments with noise variables”. *Journal of Quality Technology* 35.4, pp. 393–405.
- Gomes, C, Claeys-Bruno, M, and Sergent, M (2018). “Space-filling designs for mixtures”. *Chemometrics and Intelligent Laboratory Systems* 174, pp. 111–127.
- Goos, P. and Donev, A. N. (2007). “Tailor-made split-plot designs for mixture and process variables”. *Journal of Quality Technology* 39.4, pp. 326–339.
- Goos, P., Jones, B., and Syafitri, U. (2016). “I-Optimal design of mixture experiments”. *Journal of the American Statistical Association* 111.514, pp. 899–911.
- Graham, R. L. and Hell, P. (1985). “On the history of the minimum spanning tree problem”. *Annals of the History of Computing* 7.1, pp. 43–57.
- Gray, A. G. and Moore, A. W. (2001). “N-body’problems in statistical learning”. In: *Advances in Neural Information Processing Systems*, pp. 521–527.
- Gunzburger, M. and Burkardt, J. (2004). “Uniformity measures for point sample in hypercubes”. *Rapp. Tech. Florida State University (cf. p. 73)*.
- Hickernell, F. (1998). “A generalized discrepancy and quadrature error bound”. *Mathematics of Computation of the American Mathematical Society* 67.221, pp. 299–322.
- Hinrichs, C., Boiler, C., and Walsh, S. (2020). *JMP Essentials: An Illustrated Guide for New Users*. SAS Institute.
- Hua, L.-K. and Wang, Y. (2012). *Applications of Number Theory to Numerical Analysis*. Springer Science & Business Media, New York.

- Husslage, B. G., Rennen, G., Van Dam, E. R., and Den Hertog, D. (2011). “Space-filling Latin hypercube designs for computer experiments”. *Optimization and Engineering* 12.4, pp. 611–630.
- Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990). “Minimax and maximin distance designs”. *Journal of Statistical Planning and Inference* 26.2, pp. 131–148.
- Joseph, V. R., Gul, E., and Ba, S. (2015). “Maximum projection designs for computer experiments”. *Biometrika* 102.2, pp. 371–380.
- Keng, H. L. and Yuan, W. (1983). “Applications of number theory to numerical analysis”. *Bulletin (New Series) of the American Mathematical Society* 8.3, pp. 489–496.
- Khuri, A. I. and Mukhopadhyay, S. (2010). “Response surface methodology”. *Wiley Interdisciplinary Reviews: Computational Statistics* 2.2, pp. 128–149.
- Kirk, R. E. (2012). *Experimental Design*. Vol. 2. Wiley Online Library.
- Koblitz, N. (1984). *P-adic Numbers, P-adic Analysis, and Zeta-Functions*. Springer, New York, pp. 1–20.
- Korobov, N. (1959). “Approximate evaluation of repeated integrals”. *Doklady Akademii Nauk SSSR* 124.6, pp. 1207–1210.
- Kruskal, J. B. (1956). “On the shortest spanning subtree of a graph and the traveling salesman problem”. *Proceedings of the American Mathematical society* 7.1, pp. 48–50.
- Lawson, J., Willden, C., et al. (2016). “Mixture experiments in R using mixexp”. *Journal of Statistical Software* 72.Code Snippet, 2, pp. 1–20.
- Lekivetz, R. and Jones, B. (2015). “Fast flexible space-filling designs for nonrectangular regions”. *Quality and Reliability Engineering International* 31.5, pp. 829–837.

- Li, R., Lin, D. K., and Chen, Y. (2004). “Uniform design: design, analysis and applications”. *International Journal of Materials and Product Technology* 20.1/2/3, p. 101. ISSN: 0268-1900.
- Limmun, W., Chomtee, B., and Borkowski, J. J. (2019). “The construction of robust mixture-process experimental designs via genetic algorithm”. *Quality and Reliability Engineering International*.
- Linse, G. (2008). “Generating Space-Filling Designs Using Meshes”. *Department of Mathematical Sciences, Montana State University, Masters Degree Writing Project*.
- Liu, Y. and Liu, M.-Q. (2016). “Construction of uniform designs for mixture experiments with complex constraints”. *Communications in Statistics-Theory and Methods* 45.8, pp. 2172–2180.
- March, W. B., Ram, P., and Gray, A. G. (2010). “Fast euclidean minimum spanning tree: algorithm, analysis, and applications”. In: *Proceedings of the 16th ACM SIGKDD international conference on Knowledge Discovery and Data Mining*, pp. 603–612.
- McDonald, R., Pereira, F., Ribarov, K., and Hajič, J. (2005). “Non-projective dependency parsing using spanning tree algorithms”. In: *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 523–530.
- McLean, R. A. and Anderson, V. L. (1966). “Extreme Vertices Design of Mixture Experiments”. *Technometrics* 8.3, p. 447. DOI: 10.2307/1266691.
- Montgomery, D. C. (2017). *Design and Analysis of Experiments*. John Wiley & Sons, New York.
- Morokoff, W. J. and Caflisch, R. E. (1994). “Quasi-random sequences and their discrepancies”. *SIAM Journal on Scientific Computing* 15.6, pp. 1251–1279.

- Mou, C., Chen, S., and Tu, Y.-C. (2019). “A comparative study of dual-tree algorithms for computing spatial distance histograms”. *The Computer Journal* 62.1, pp. 42–62.
- Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M. (2016). *Response Surface Methodology: Process and Product Optimization using Designed Experiments*. John Wiley & Sons, New York.
- Nachtsheim, C. J. and Jones, B. (2018). “Design augmentation for response optimization and model estimation”. *Quality Engineering* 30.1, pp. 38–51.
- Nešetřil, J., Milková, E., and Nešetřilová, H. (2001). “Otakar Boruvka on minimum spanning tree problem translation of both the 1926 papers, comments, history”. *Discrete Mathematics* 233, pp. 3–36.
- Owen, A. B. (1992). “Orthogonal arrays for computer experiments, integration and visualization”. *Statistica Sinica* 2.2, pp. 439–452.
- Park, G.-J. (2007). *Design of experiments*. Springer, London, pp. 309–391.
- Park, J.-S. (1994). “Optimal Latin-hypercube designs for computer experiments”. *Journal of Statistical Planning and Inference* 39.1, pp. 95–111.
- Pawlowsky-Glahn, V. and Egozcue, J. J. (2001). “Geometric approach to statistical analysis on the simplex”. *Stochastic Environmental Research and Risk Assessment* 15.5, pp. 384–398.
- Pham, H., ed. (2006). *Springer Handbook of Engineering Statistics*. London: Springer London.
- Piepel, G., Anderson, C., and Redgate, P. (1993). “Response surface designs for irregularly-shaped regions (Parts 1, 2, and 3)”. In: *1993 Proceedings of the Section on Physical and Engineering Sciences*. American Statistical Association Alexandria, VA, pp. 205–227.

- Piepel, G. F., Cooley, S. K., Peeler, D. K., Vienna, J. D., and Edwards, T. B. (2002). “Augmenting a waste glass mixture experiment study with additional glass components and experimental runs”. *Quality Engineering* 15.1, pp. 91–111.
- Piepel, G. F., Cooley, S. K., and Jones, B. (2005). “Construction of a 21-component layered mixture experiment design using a new mixture coordinate-exchange algorithm”. *Quality Engineering* 17.4, pp. 579–594.
- Pradhan, U. K., Lal, K., Dash, S., and Singh, K. N. (2017). “Design and analysis of mixture experiments with process variable”. *Communications in Statistics - Theory and Methods* 46.1, pp. 259–270.
- Prim, R. C. (1957). “Shortest connection networks and some generalizations”. *The Bell System Technical Journal* 36.6, pp. 1389–1401.
- Sall, J., Stephens, M. L., Lehman, A., and Loring, S. (2017). *JMP start statistics: a guide to statistics and data analysis using JMP*. Sas Institute.
- Scheffe, H. (1958). “Experiments with mixtures”. *Journal of the Royal Statistical Society: Series B (Methodological)* 20.2, pp. 344–360.
- (1963). “The simplex-centroid design for experiments with mixtures”. *Journal of the Royal Statistical Society: Series B (Methodological)* 25.2, pp. 235–251.
- Shirley, P. et al. (1991). “Discrepancy as a quality measure for sample distributions”. In: *Proceedings Eurographics*. Vol. 91, pp. 183–194.
- Singh, A. (2009). “An artificial bee colony algorithm for the leaf-constrained minimum spanning tree problem”. *Applied Soft Computing* 9.2, pp. 625–631.
- Smith, W. F. (2005). *Experimental Design for Formulation*. Vol. 15. ASA-SIAM, New York.
- Snee, R. D. and Marquardt, D. W. (1974). “Extreme vertices designs for linear mixture models”. *Technometrics* 16.3, pp. 399–408.

- Talke, I. S. (2012). “Number Theoretic Methods in Designing Experiments”. Statistics PhD Dissertation. Montana State University.
- Team, R. C. (2020). “R: A Language and Environment for Statistical Computing”.
- Trosset, M. W. (1999). “Approximate maximin distance designs”. In: *Proceedings of the Section on Physical and Engineering Sciences*, pp. 223–227.
- Vamanamurthy, M. K. and Vuorinen, M. (1994). “Inequalities for means”. *Journal of Mathematical Analysis and Applications* 183.1, pp. 155–166.
- Wallet, F and Dussert, C (1998). “Comparison of spatial point patterns and processes characterization methods”. *EPL (Europhysics Letters)* 42.5, p. 493.
- Wang, Y. and Fang, K. T. (1981). “A note on uniform distribution and experimental design”. *Kexue Tongbao (English Ed.)* 26.6, pp. 485–489. ISSN: 0250-7862.
- Warnock, T. T. (1972). *Applications of Number Theory to Numerical Analysis*. Elsevier. Chap. Computational Investigations of Low-discrepancy Point Sets, pp. 319–343.
- West, D. B. et al. (2001). *Introduction to Graph Theory*. Vol. 2. Prentice Hall, Upper Saddle River.
- Wheeler, B. and Braun, M. J. (2019). *Package ‘AlgDesign’*.
- Wiens, D. P. (1991). “Designs for approximately linear regression: two optimality properties of uniform designs”. *Statistics & Probability Letters* 12.3, pp. 217–221.
- Winer, B. J. (1962). *Statistical Principles in Experimental Design*. McGraw-Hill , New York.
- Zahn, C. T. (1971). “Graph-theoretical methods for detecting and describing gestalt clusters”. *IEEE Transactions on Computers* 100.1, pp. 68–86.

Zhang, L., Liang, Y.-Z., Jiang, J.-H., Yu, R.-Q., and Fang, K.-T. (1998). “Uniform design applied to nonlinear multivariate calibration by ANN”. *Analytica Chimica Acta* 370.1, pp. 65–77.

APPENDIX ATABLES IN S^q FOR $q = 3, 4, 5$

	GLP		SRS		PP		CF	
n	Gen	GMST	Gen	GMST	Gen	GMST	Gen	GMST
9	(1 4)	0.2691	(5 47)	0.3381	19	0.2970	NA	NA
10	(1 3)	0.2768	(17 41)	0.3208	5	0.2509	NA	NA
11	(1 7)	0.2685	(37 29)	0.2978	5	0.2471	11	0.2013
12	(1 7)	0.2502	(29 37)	0.2935	5	0.2440	11	0.1805
13	(1 5)	0.2516	(29 37)	0.2919	5	0.2368	13	0.2520
14	(1 9)	0.2266	(7 37)	0.2658	5	0.2179	13	0.2372
15	(1 11)	0.2250	(41 3)	0.2496	13	0.2076	13	0.2309
16	(1 5)	0.2030	(11 37)	0.2373	29	0.2023	13	0.2220
17	(1 12)	0.2149	(7 5)	0.2368	13	0.2040	13	0.2148
18	(1 5)	0.2111	(29 43)	0.2251	11	0.1977	13	0.2101
19	(1 14)	0.1959	(11 2)	0.2174	3	0.1922	19	0.2070
20	(1 13)	0.1663	(11 37)	0.2060	3	0.1844	13	0.1995
21	(1 13)	0.1917	(5 31)	0.2004	3	0.1831	13	0.1997
22	(1 17)	0.1873	(11 2)	0.1919	3	0.1799	13	0.1995
23	(1 5)	0.1826	(23 37)	0.1901	3	0.1779	17	0.1890
24	(1 17)	0.1770	(3 29)	0.1860	3	0.1717	17	0.1819
25	(1 7)	0.1777	(13 19)	0.1804	3	0.1735	17	0.1683
26	(1 7)	0.1697	(3 29)	0.1800	3	0.1668	17	0.1638
27	(1 22)	0.1630	(47 13)	0.1753	3	0.1615	17	0.1547
28	(1 5)	0.1591	(13 19)	0.1730	3	0.1547	17	0.1505
29	(1 21)	0.1634	(29 7)	0.1675	17	0.1489	17	0.1458
30	(1 23)	0.1530	(29 7)	0.1666	17	0.1450	17	0.1438

Table A.1: Best design generators in S^3

	GLP		SRS		PP		CF	
n	Gen	GMST	Gen	GMST	Gen	GMST	Gen	GMST
9	(1 2 4)	0.3342	(5 31 17)	0.4150	31	0.3297	NA	NA
10	(1 7 3)	0.2816	(23 17 13)	0.3752	13	0.3079	NA	NA
11	(1 6 8)	0.3163	(37 31 3)	0.3629	41	0.3018	11	0.2988
12	(1 5 7)	0.2467	(5 37 31)	0.3490	13	0.2906	11	0.2839
13	(1 3 5)	0.3025	(29 47 37)	0.3440	13	0.2917	11	0.2626
14	(1 9 11)	0.2893	(7 23 37)	0.3346	13	0.2675	11	0.2551
15	(1 8 11)	0.2562	(7 23 37)	0.3084	13	0.2662	11	0.2453
16	(1 5 3)	0.2701	(17 29 37)	0.2983	13	0.2514	11	0.2382
17	(1 5 3)	0.2688	(2 17 37)	0.2869	13	0.2476	17	0.2400
18	(1 7 5)	0.2479	(2 17 37)	0.2826	13	0.2475	17	0.2385
19	(1 13 15)	0.2558	(29 31 37)	0.2714	13	0.2488	17	0.2373
20	(1 13 17)	0.2459	(13 31 37)	0.2643	13	0.2476	17	0.2313
21	(1 16 13)	0.2406	(17 2 37)	0.2606	13	0.2368	17	0.2253
22	(1 13 3)	0.2318	(17 37 2)	0.2603	2	0.2259	11	0.2221
23	(1 8 3)	0.2402	(17 2 37)	0.2555	2	0.2182	17	0.2214
24	(1 19 11)	0.2293	(37 29 31)	0.2487	2	0.2109	17	0.2134
25	(1 16 3)	0.2279	(37 19 13)	0.2464	2	0.2057	11	0.2122
26	(1 19 23)	0.2200	(37 29 31)	0.2383	2	0.1987	11	0.2089
27	(1 4 10)	0.2287	(37 7 41)	0.2307	2	0.1977	11	0.2041
28	(1 19 13)	0.2175	(29 43 19)	0.2291	2	0.1954	11	0.2035
29	(1 24 16)	0.2214	(29 43 19)	0.2269	2	0.1957	11	0.2017
30	(1 13 7)	0.2099	(7 43 2)	0.2203	2	0.1912	17	0.1997

Table A.2: Best design generators in S^4

	GLP		SRS		PP		CF	
n	Gen	GMST	Gen	GMST	Gen	GMST	Gen	GMST
9	(1 5 7 2)	0.3370	(11 17 31 23)	0.4346	23	0.3292	NA	NA
10	(1 7 9 3)	0.2649	(41 11 5 17)	0.4017	41	0.3233	NA	NA
11	(1 7 9 8)	0.3264	(31 19 3 17)	0.3856	19	0.3145	13	0.3085
12	(1 11 5 7)	0.2338	(7 3 17 2)	0.3700	19	0.3006	13	0.2917
13	(1 3 8 6)	0.3088	(29 5 47 37)	0.3675	19	0.3087	13	0.2906
14	(1 9 3 11)	0.2791	(29 47 7 37)	0.3543	19	0.2975	13	0.2825
15	(1 7 11 2)	0.2938	(11 23 37 3)	0.3389	19	0.2922	13	0.2706
16	(1 13 5 9)	0.3043	(3 31 29 37)	0.3212	43	0.2794	13	0.2607
17	(1 4 7 2)	0.2841	(31 41 5 37)	0.3123	3	0.2698	17	0.2896
18	(1 7 5 13)	0.2453	(31 41 5 37)	0.3052	3	0.2642	17	0.2867
19	(1 13 8 15)	0.2835	(13 17 37 7)	0.2957	29	0.2587	17	0.2834
20	(1 7 17 9)	0.2638	(13 47 37 7)	0.2926	19	0.2541	17	0.2712
21	(1 2 5 8)	0.2589	(29 37 5 31)	0.2836	29	0.2509	17	0.2639
22	(1 7 9 19)	0.2620	(47 19 2 11)	0.2821	29	0.2463	17	0.2608
23	(1 4 16 10)	0.2639	(23 19 37 11)	0.2758	29	0.2458	17	0.2621
24	(1 19 13 7)	0.2528	(7 37 11 23)	0.2704	29	0.2467	17	0.2514
25	(1 19 16 21)	0.2559	(13 43 37 17)	0.2725	29	0.2482	17	0.2453
26	(1 3 15 21)	0.2497	(29 19 37 17)	0.2641	29	0.2420	17	0.2437
27	(1 10 8 23)	0.2505	(41 37 3 31)	0.2593	29	0.2357	17	0.2427
28	(1 13 5 3)	0.2430	(29 43 3 2)	0.2554	29	0.2327	17	0.2410
29	(1 11 24 26)	0.2462	(2 19 37 17)	0.2539	3	0.2340	17	0.2373
30	(1 7 19 13)	0.2348	(3 5 31 13)	0.2520	3	0.2247	17	0.2367

Table A.3: Best design generators in S^5

	GLP		SRS		PP		CF	
n	Gen	GMST	Gen	GMST	Gen	GMST	Gen	GMST
9	(1 4 7)	0.4682	(5 47 43)	0.4851	2	0.4452	NA	NA
10	(1 3 7)	0.3757	(17 41 23)	0.4529	2	0.4345	NA	NA
11	(1 7 2)	0.4178	(37 29 3)	0.4442	2	0.4001	11	0.3993
12	(1 7 5)	0.3324	(29 37 3)	0.4279	2	0.3952	11	0.3774
13	(1 5 3)	0.4068	(29 37 47)	0.4307	2	0.3850	11	0.3613
14	(1 9 11)	0.3992	(7 37 23)	0.4172	13	0.3650	11	0.3487
15	(1 11 8)	0.3389	(41 3 37)	0.3911	13	0.3582	11	0.3408
16	(1 5 9)	0.3717	(11 37 31)	0.3778	47	0.3564	11	0.3392
17	(1 12 3)	0.3512	(7 5 37)	0.3632	47	0.3520	11	0.3344
18	(1 5 11)	0.3342	(29 43 3)	0.3485	47	0.3536	11	0.3307
19	(1 14 8)	0.3592	(11 2 17)	0.3332	47	0.3552	11	0.3203
20	(1 13 9)	0.3291	(11 37 3)	0.3426	47	0.3525	11	0.3123
21	(1 13 16)	0.3277	(5 31 19)	0.3212	47	0.3454	11	0.3106
22	(1 17 15)	0.3231	(11 2 17)	0.3187	47	0.3168	11	0.3120
23	(1 5 8)	0.3252	(23 37 5)	0.3298	13	0.2995	11	0.3099
24	(1 17 19)	0.3041	(3 29 2)	0.3152	11	0.2903	11	0.3041
25	(1 7 4)	0.2949	(13 19 37)	0.3302	11	0.2821	11	0.3054
26	(1 7 15)	0.3024	(3 29 2)	0.3113	11	0.2744	11	0.3018
27	(1 22 8)	0.3208	(47 13 11)	0.2902	2	0.2685	11	0.2975
28	(1 5 13)	0.2933	(13 19 37)	0.2973	2	0.2628	11	0.2940
29	(1 21 24)	0.2955	(29 7 43)	0.2974	2	0.2632	11	0.2909
30	(1 23 13)	0.2876	(29 7 43)	0.2954	2	0.2571	11	0.2849

Table A.4: Best design generators in S^3 and $K = 1$

	GLP		SRS		PP		CF	
n	Gen	GMST	Gen	GMST	Gen	GMST	Gen	GMST
9	(1 4 7 5)	0.5646	(5 47 43 7)	0.6246	17	0.5825	NA	NA
10	(1 3 7 9)	0.4578	(17 41 23 3)	0.5742	41	0.5443	NA	NA
11	(1 7 2 5)	0.5527	(37 29 3 47)	0.5714	43	0.5301	11	0.4507
12	(1 7 5 11)	0.3936	(29 37 3 43)	0.5705	19	0.5123	11	0.4333
13	(1 5 3 6)	0.5300	(29 37 47 11)	0.5635	19	0.5112	13	0.4820
14	(1 9 11 5)	0.4740	(7 37 23 31)	0.5288	19	0.4989	13	0.4685
15	(1 11 8 2)	0.5207	(41 3 37 23)	0.5136	19	0.4918	13	0.4482
16	(1 5 9 3)	0.5227	(11 37 31 23)	0.5201	19	0.4780	13	0.4340
17	(1 12 3 10)	0.4867	(7 5 37 17)	0.4684	3	0.4607	17	0.4916
18	(1 5 11 13)	0.3950	(29 43 3 5)	0.4738	3	0.4568	17	0.4765
19	(1 14 8 17)	0.4826	(11 2 17 37)	0.4686	19	0.4515	17	0.4706
20	(1 13 9 17)	0.4389	(11 37 3 43)	0.4680	19	0.4512	17	0.4595
21	(1 13 16 19)	0.4449	(5 31 19 3)	0.4449	29	0.4354	17	0.4592
22	(1 17 15 9)	0.4549	(11 2 17 37)	0.4461	3	0.4317	17	0.4582
23	(1 5 8 6)	0.4386	(23 37 5 19)	0.4510	29	0.4284	17	0.4528
24	(1 17 19 11)	0.4077	(3 29 2 37)	0.4316	29	0.4357	17	0.4414
25	(1 7 4 9)	0.4293	(13 19 37 17)	0.4516	29	0.4339	17	0.4335
26	(1 7 15 23)	0.4163	(3 29 2 37)	0.4346	29	0.4245	17	0.4253
27	(1 22 8 10)	0.4180	(47 13 11 7)	0.4208	29	0.4116	17	0.4263
28	(1 5 13 3)	0.4147	(13 19 37 47)	0.4231	29	0.4014	17	0.4227
29	(1 21 24 11)	0.4104	(29 7 43 47)	0.4103	3	0.4018	17	0.4147
30	(1 23 13 11)	0.4001	(29 7 43 5)	0.4052	3	0.3925	17	0.4144

Table A.5: Best design generators in S^3 and $K = 2$

	GLP		SRS		PP		CF	
n	Gen	GMST	Gen	GMST	Gen	GMST	Gen	GMST
9	(1 2 4 5)	0.4341	(5 31 17 7)	0.5271	43	0.4662	NA	NA
10	(1 7 3 9)	0.3535	(23 17 13 11)	0.4953	19	0.4360	NA	NA
11	(1 6 8 9)	0.4440	(37 31 3 47)	0.4705	19	0.4375	11	0.3584
12	(1 5 7 11)	0.3142	(5 37 31 11)	0.4626	19	0.4310	11	0.3385
13	(1 3 5 2)	0.4036	(29 47 37 19)	0.4686	19	0.4328	13	0.4063
14	(1 9 11 3)	0.3760	(7 23 37 31)	0.4488	19	0.4107	13	0.3918
15	(1 8 11 2)	0.4077	(7 23 37 3)	0.4363	19	0.3961	13	0.3661
16	(1 5 3 7)	0.3937	(17 29 37 11)	0.4157	19	0.3802	13	0.3477
17	(1 5 3 7)	0.3747	(2 17 37 11)	0.4098	3	0.3637	17	0.3955
18	(1 7 5 13)	0.3201	(2 17 37 11)	0.3993	3	0.3588	17	0.3871
19	(1 13 15 8)	0.3786	(29 31 37 5)	0.3841	29	0.3590	17	0.3850
20	(1 13 17 9)	0.3437	(13 31 37 3)	0.3772	29	0.3535	17	0.3655
21	(1 16 13 19)	0.3567	(17 2 37 11)	0.3484	29	0.3484	17	0.3603
22	(1 13 3 7)	0.3519	(17 37 2 11)	0.3491	29	0.3457	17	0.3583
23	(1 8 3 17)	0.3432	(17 2 37 11)	0.3440	29	0.3425	17	0.3577
24	(1 19 11 17)	0.3226	(37 29 31 17)	0.3500	29	0.3419	17	0.3418
25	(1 16 3 7)	0.3275	(37 19 13 17)	0.3503	29	0.3429	17	0.3358
26	(1 19 23 21)	0.3243	(37 29 31 3)	0.3356	29	0.3386	17	0.3291
27	(1 4 10 8)	0.3297	(37 7 41 47)	0.3250	29	0.3268	17	0.3304
28	(1 19 13 17)	0.3291	(29 43 19 47)	0.3299	29	0.3231	17	0.3302
29	(1 24 16 7)	0.3209	(29 43 19 47)	0.3310	3	0.3178	17	0.3235
30	(1 13 7 11)	0.3123	(7 43 2 3)	0.3243	3	0.3059	17	0.3215

Table A.6: Best design generators in S^4 and $K = 1$

	GLP		SRS		PP		CF	
n	Gen	GMST	Gen	GMST	Gen	GMST	Gen	GMST
9	(1 2 4 57)	0.5346	(5 31 17 7 11)	0.6248	23	0.5256	NA	NA
10	NA	NA	(23 17 13 11 3)	0.5913	23	0.5241	NA	NA
11	(1 6 8 9 7)	0.5562	(37 31 3 47 19)	0.5978	5	0.5112	11	0.5130
12	NA	NA	(5 37 31 11 47)	0.5739	5	0.5057	11	0.5107
13	(1 3 5 2 9)	0.5356	(29 47 37 19 5)	0.5707	5	0.5088	11	0.4998
14	(1 9 11 3 5)	0.4525	(7 23 37 31 5)	0.5615	5	0.5009	11	0.4867
15	(1 8 11 2 4)	0.4814	(7 23 37 3 47)	0.5527	23	0.4802	11	0.4598
16	(1 5 3 7 13)	0.4842	(17 29 37 11 43)	0.5216	23	0.4757	11	0.4702
17	(1 5 3 7 13)	0.4787	(2 17 37 11 3)	0.5063	23	0.4689	17	0.4963
18	(1 7 5 13 11)	0.3880	(2 17 37 11 3)	0.4957	11	0.4670	17	0.4953
19	(1 13 15 8 3)	0.4652	(29 31 37 5 3)	0.4809	23	0.4708	17	0.4877
20	(1 13 17 9 7)	0.4221	(13 31 37 3 5)	0.4704	23	0.4709	17	0.4837
21	(1 16 13 19 4)	0.4461	(17 2 37 11 19)	0.4474	23	0.4589	17	0.4776
22	(1 13 3 7 5)	0.4743	(17 37 2 11 7)	0.4466	23	0.4529	17	0.4675
23	(1 8 3 17 10)	0.4604	(17 2 37 11 19)	0.4467	23	0.4487	17	0.4672
24	(1 19 11 17 13)	0.3866	(37 29 31 17 7)	0.4513	23	0.4454	17	0.4536
25	(1 16 3 7 4)	0.4235	(37 19 13 17 11)	0.4604	23	0.4440	17	0.4502
26	(1 19 23 21 11)	0.4432	(37 29 31 3 5)	0.4622	23	0.4259	17	0.4507
27	(1 4 10 8 7)	0.4251	(37 7 41 47 11)	0.4410	41	0.4215	17	0.4490
28	(1 19 13 17 5)	0.4450	(29 43 19 47 3)	0.4250	41	0.4103	17	0.4431
29	(1 24 16 7 10)	0.4298	(29 43 19 47 3)	0.4258	41	0.4060	17	0.4400
30	(1 13 7 11 23)	0.3800	(7 43 2 3 5)	0.4334	41	0.3947	17	0.4340

Table A.7: Best design generators in S^4 and $K = 2$