Characterizing Uncertainty in Correlated Response Variables for Pareto Front Optimization

THESIS

Peter A. Calhoun, 1st Lt, USAF

AFIT-ENS-MS-20-M-136

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-20-M-136

CHARACTERIZING UNCERTAINTY IN CORRELATED RESPONSE VARIABLES FOR PARETO FRONT OPTIMIZATION

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Peter A. Calhoun, B.S.

1st Lt, USAF

26 March 2020

AFIT-ENS-MS-20-M-136

CHARACTERIZING UNCERTAINTY IN CORRELATED RESPONSE VARIABLES FOR PARETO FRONT OPTIMIZATION

THESIS

Peter A. Calhoun, B.S.
1st Lt, USAF

Committee Membership:

Lt Col Beau A. Nunnaly, Ph.D.
Chair

Lance E. Champagne, Ph.D.
Reader

AFIT-ENS-MS-20-M-136

Abstract

Current research provides a method to incorporate uncertainty into Pareto front optimization by simulating additional response surface model parameters according to a Multivariate Normal Distribution (MVN). This research shows that analogous to the univariate case, the MVN understates uncertainty, leading to overconfident conclusions when variance is not known and there are few observations (less than 25-30 per response). This research builds upon current methods using simulated response surface model parameters that are distributed according to an Multivariate t-Distribution (MVT), which can be shown to produce a more accurate inference when variance is not known. The MVT better addresses uncertainty in the parameters which can affect the frequency of treatments appearing on the Pareto front resulting in potentially different proposed solution spaces from that of the MVN.

Acknowledgements

I would like to thank my wife and daughter for being incredibly supportive and understanding through the process of this research. Additionally, I am greatly appreciative of the guidance provided to me by my advisor, Dr. Beau Nunnally, as well as the instruction from both the ENC and ENS departments at AFIT as it was pivotal in conducting this research.

Peter A. Calhoun

Table of Contents

Page

List of Figures

List of Tables

CHARACTERIZING UNCERTAINTY IN CORRELATED RESPONSE VARIABLES FOR PARETO FRONT OPTIMIZATION

I. Introduction

1.1 Background

Optimization is a core technique of operations research, one that relies on applying certain mathematical concepts with the objective of making some product or process the best that it can be. One particular case where optimization is used is response surfaces. A response surface is defined by a dependent variable that is explained by two or more independent variables. These response surfaces may be optimized using an array of techniques such as gradient search, overlaying contour plots, constrained optimization, desirability functions and Pareto front optimization (PFO) to name a few. The first three techniques will be reviewed within Chapter II but the focus of this research is PFO and its modifications to incorporate uncertainty to better provide optimization results.

1.2 Problem Statement

PFO is a two sequential step deterministic method to locate optimal points when multiple objectives or goals are present. When considering models that have random error present, the Pareto front will only capture one snapshot of a surface to be optimized when there will actually be many. A key field where this is used is Response Surface Methodology (RSM). A model designed with RSM in mind may have multiple correlated objectives that Pareto frontiers are commonly used for handling [1]; how-

ever, RSM is based on the analysis of variance (ANOVA) which assumes error in order to represent the response surface as a random variable. The assumed error is necessary within the parameter estimates as they are calculated using a sample. Failing to account for uncertainty in parameter estimates may have impacts on PFO causing it not to be perceptive to the potential solutions resulting in missed opportunities for optimization [2].

Chapman et al. recognizes the need to incorporate uncertainty into PFO and develops a method to do so by sampling parameters from a Multivariate Normal Distribution (MVN) to generate additional response surfaces that are within stochastically appropriate intervals of the original parameter estimates [3]. Research suggests that the Multivariate Normal Distribution (MVN) understates uncertainty, leading to overconfident conclusions when variance is not known and there are few observations (less than 25-30 per response). This methodology builds upon Chapman's research using simulated response surface model parameters that are sampled from a Multivariate t-Distribution (MVt), which provides a more conservative inference when covariance is not known due to its heavier "tails" [4].

1.3 Conclusion

This research will have the following structure and content. Chapter II presents relevant prior research of the subject matter and develops an extension from univariate to multivariate methods. In Chapter III, the methodology introduced by Chapman et al. is expanded. Chapter IV formally presents the statistical inference and implications of altering the sampling distribution to include uncertainty using the methodology introduced in Chapter III with a chemical process optimization problem. Chapter V outlines the research conducted as well as discusses further areas of study that may be derived from this research.

II. Literature Review

2.1 Overview

This chapter reviews the statistical foundation of RSM to include parameter estimation and designed experiments for second-order models, multi-objective optimization methods, uncertainty, and probability distributions. The literature reviewed here will form the basis for topics in Chapter III.

2.2 Statistical Foundation of RSM

The general methodology of RSM is to sequentially use Design of Experiments (DOE) with gradient search to obtain an optimal response. Typically, designed experiments are generated and then a model fit to recorded data using a linear regression model. The two most common methods of estimation are Ordinary Least Squares (OLS) and Maximum Likelihood [5]. Often curvature in a response surface is significant enough that second-order models are required [1]. Because curvature is generally present, this research focuses on the standard second-order model designs. The designs discussed in Section 2.2.2 are the Central Composite Design (CCD), Box-Behnken Design (BBD), and Definitive Screening Design (DSD).

2.2.1 Parameter Estimation

RSM uses regression models to approximate a response surface given responses, $\boldsymbol{y}$, in terms of factors, $\boldsymbol{X}$. The theoretical linear model is of the form:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{y}$ is $n \times 1$ vector of responses, $\boldsymbol{X}$ is $n \times p$ model matrix of fixed values, $\boldsymbol{\beta}$ is $p \times 1$ vector of model parameters, and $\boldsymbol{\varepsilon}$ is $n \times 1$ vector of error terms that are assumed to be independently and identically distributed random variables with mean 0 and variance σ^2. The error term is most commonly distributed normally for convenience of testing; however, other distributions can be used depending on the needs of the analysis. Both OLS and Maximum Likelihood have been shown to arrive at the same minimum variance unbiased estimators with the exception of the Maximum Likelihood estimate for variance which is biased [5].

2.2.1.1 Ordinary Least Squares

OLS, which assumes no distribution restriction, searches for the equations that minimize the sum of squared error. Equation 1 can be rearranged in terms of error.

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \tag{2}$$

Minimizing the sum of squared error can be accomplished by the following:

$$\begin{aligned} L &= \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \\ &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \end{aligned} \tag{3}$$

where the addition in between the last two lines can be done since $(\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y})' = \boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta}$ and is scalar. The least squares estimator, L, must satisfy

$$\left.\frac{\partial L}{\partial \beta}\right|_{\boldsymbol{b}} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = 0$$

Finally, dividing each term by 2 and rearranging the equation gives

$$\begin{aligned} \boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} &= \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{b} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \end{aligned} \tag{4}$$

Equation 4 are named the normal equations and they are used to calculate unbiased estimators for the β coefficients of the regression model. The sampling distribution of b is Multivariate Normal (MVN) with mean $\boldsymbol{\beta}$ and covariance $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ [1]. Using these normal equations with the original linear model, the expected value of the estimated regression equation can be found

$$\begin{aligned} E[\boldsymbol{y}] &= E[\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}] \\ &= E[\boldsymbol{X}\boldsymbol{\beta}] + E[\boldsymbol{\varepsilon}] \\ &= E[\boldsymbol{X}\boldsymbol{\beta}] && \text{since } E[\boldsymbol{\varepsilon}] = 0 \\ &= \boldsymbol{X}\boldsymbol{b} && \text{since } \boldsymbol{X} \text{ is a constant matrix} \end{aligned} \tag{5}$$

$E[\boldsymbol{y}]$ can be rewritten as $\hat{\boldsymbol{y}}$ so the equation becomes $\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{b}$.

2.2.1.2 Maximum Likelihood

Maximum Likelihood assumes a distribution which is then used to determine the most likely parameter estimates. Using the density of a probability distribution, this method is able to return different probabilities based on the given parameters. More specifically, the likelihood function for n observations $y_1, y_2, \ldots, y_n$ is the product of the individual densities [5, 6]. Under the assumption of normality the normal density function for the errors is

$$f(\varepsilon_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon_i}{2\sigma^2}\right) \tag{6}$$

The likelihood function is then defined as the product of equation 6 for n observations:

$$L(\boldsymbol{\varepsilon}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} f(\varepsilon_i) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{2\sigma^2}\right)$$

which can then be written as

$$L(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^2}\right) \tag{7}$$

Taking the natural log of the likelihood function provides and simpler equation to work with

$$ln(L(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2)) = -\frac{n}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \tag{8}$$

Therefore, for a fixed value of σ^2, the log-likelihood function is maximized when $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$ is minimized which is equivalent to the OLS estimators under the normal error assumption [6].

2.2.2 Standard Designed Experiments for Second Order Models

Generally, a second order model will be designed under the assumption that the sequential movement required to move into a region of interest has already been accomplished and curvature has been detected through the inclusion of center points [1]. As designed experiments have become increasingly popular due to their methodological approach to conducting experiments, Anderson-Cook discusses criteria commonly used to determine a recommended designs using optimality criterion and graphical methods [7]. The following subsections will focus on the classical designs such as CCD, BBD, and DSD.

2.2.2.1 Central Composite Design

Introduced by George Box and K.B. Wilson in 1951, the CCD is the most popular second order design [1]. A CCD is a designed experiment that has been augmented from a standard 2^k Factorial design to allow for estimation of quadratic effects. The design has 2^k base runs, where k is the number of factors of interest, typically 3-5 center points to capture curvature, and $2k$ axial points to model the pure quadratic effects [1, 8]. The axial point distance from the center of the design can be determined in different ways, Montgomery suggests using $\alpha = \sqrt[4]{2^k}$ to obtain rotatability. Rotatability provides the benefit of constant variance for the predicted response at all points of $\boldsymbol{X}$ that are the same distance from the center of the design [1, 8]. Figure 1 depicts the graphical representation of a CCD with two factors.

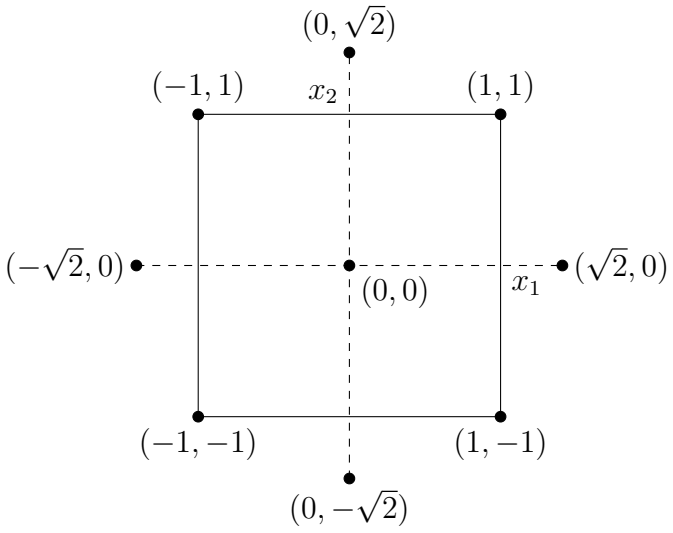


Figure 1. Central Composite Design with 2 Factors and $\alpha = \sqrt{2}$

2.2.2.2 Box-Behnken Design

The BBD was developed in 1960 by its namesakes, George Box and Donald Behnken. It is a family of three-level designs used to fit second order response surfaces and requires a minimum of three factors. The inspiration for BBD stems from the

construction of balanced incomplete block designs. In other words, all combinations of two factors are paired together resembling standard 2^2 factorial designs while the other $k-2$ factors are fixed at 0. The number of runs, N, required for k factors is given by $N = 2^2 \cdot \binom{k}{2} + n_c = 4 \cdot \binom{k}{2} + n_c$, where n_c is the number of center runs. This design is considered a spherical design and does not provide adequate coverage of the extreme points i.e., the corners [1, 8]. Figure 2 shows this as the design points are either in the center or on the edge between each corner for $k = 3$ factors. If a three-level design is desired with the extremes captured, a modification of the CCD where $\alpha = 1$ which is a cuboidal design called a Face-Centered Cube design.

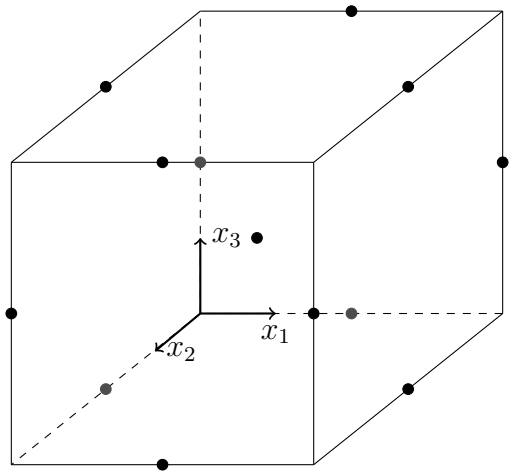


Figure 2. Box-Behnken Design with 3 Factors and center point

2.2.2.3 Definitive Screening Design

This class of design was introduced by Bradley Jones and Christopher Nachtsheim. The DSD is generally suited for fitting first-order models with interactions but can be adapted to be used in fitting second order models through a heavy reliance on the sparsity of effects principle. The sparsity of effects principle simply states that most systems are dominated by main effects and low order interactions with higher order interaction terms considered insignificant [8]. When too many higher order terms are

significant, it results in aliasing which causes difficulty in determining which effects have a true impact. They use three levels of each factor and use nonregular structures. Nonregular structures allow for flexible aliasing structures allowing for more terms to be estimated with some correlation between factors present. A DSD for second-order models is desirable when there is reason to believe that the optimal location is within the current design region [1].

2.3 Multi-Objective Optimization

2.3.1 Overlaying Contour Plots

A simple technique of overlaying contour plots of the response can be done to get a general idea of which solution sets will be near optimal for all responses. This optimization method is more suited for two process variables as it requires the response variables to be graphed in each dimension. Myers et al. states that this method can be done with more than two process variables but it becomes awkward as the additional $k-2$ process variables must be held constant which requires a lot of trial and error to determine preferred operating locations [1]. More formal techniques are desired, however, overlaying contour plots can be very useful in determining initial operating conditions and achieving a fairly correct solution in a short amount of time.

2.3.2 Constrained Optimization

Constrained Optimization is a familiar practice in Operations Research where some function $f(x)$ is either minimized or maximized subject to constraints of the form

$$\begin{aligned} &\text{Max } f(x) \\ &\text{s.t. } g_i(x) \leq c_i \\ &h_j(x) = d_j \end{aligned}$$

where i is the number of inequality constraints and j is the number of equality constraints. c_i, d_j are constants and $g_i(x), h_j(x)$ are functions of the same x in the objective function. In conjunction with RSM, the specific field of constrained optimization is nonlinear programming. This method focuses on one response as the primary goal with the other responses set as constraints. Both direct search and numerical optimization algorithms can be used to find optimum locations [1]. Carlyle et al. overviews the methods for finding general solutions with constrained optimization for these problems [9].

2.3.3 Desirability Functions

The desirability function was originally introduced by Harrington (1965) and later popularized by Derringer and Suich (1980) in the application of simultaneously optimizing multiple responses with desirability functions [1, 10, 11]. This method standardizes each response by converting them using desirability functions, d_i, over the range

$$0 \leq d_i \leq 1$$

and then maximizes the overall desirability. Combining the individual desirability functions into one is accomplished with either a weighted sum called the additive form or a weighted product called the multiplicative form. For m objectives, the additive form is

$$D = \sum_{i=1}^{m} w_i d_i \tag{9}$$

which allows for superior performance in one or more objectives to compensate for poor performance of another [1]. Similarly, the multiplicative form is

$$D = \prod_{i=1}^{m} d_i^{w_i} \tag{10}$$

which penalizes poor performance of a single objective regardless of other objectives performance [1]. For both forms, the weights must sum to 1 to restrict D between 0 and 1. Depending on the objective, the individual desirability functions, d_i, take on slightly different forms. When the target, T, is to maximize y, the form is

$$d_i = \begin{cases} 0, & y < L \\ \left(\frac{y-L}{T-L}\right)^r, & L \leq y \leq U \\ 1, & y > T \end{cases} \tag{11}$$

where r is a weight that places more emphasis on being close to T when $r > 1$ and less emphasis when $0 < r < 1$ but is linear when $r = 1$. L and U are the lower and upper limits that are user defined. If the target for the response y is to be minimized, the form is

$$d_i = \begin{cases} 1, & y < T \\ \left(\frac{U-y}{U-T}\right)^r, & L \leq y \leq U \\ 0, & y > U \end{cases} \tag{12}$$

There is also a two-sided desirability function which assumes the target is between L and U. The two-sided desirability function is a combination of the two aforementioned functions. Maximizing D can be done through direct search methods to find a preferred solution for given preferences [1]. Del Castillo, Montgomery, and Mc-

Carville discuss modified desirability functions that use polynomial approximations for the individual desirability functions that use other search methods in the event that direct search becomes too computationally expensive [1, 12].

2.3.4 Pareto Front Optimization

PFO is a method of optimizing multiple objectives through two sequential steps: Step one is an objective step that builds a Pareto front for a given solution set to remove poor candidate solutions and step two is a subjective step that investigates trade-offs of different weighting schemes by examining the Pareto front using specified weights based on the goals of a given study. Chapman et al. details the procedure of this method and compares it with constrained optimization and desirability functions [13]. A Pareto set is found by identifying solutions that *Pareto dominate* other points and are not dominated themselves. In other words, a Pareto dominant point must be at least as good in all of the criteria values and strictly better in at least one of the criteria. The *Pareto set* is the collection of all solutions that are not Pareto dominated by others [1, 13]. While the first step removes noncontenders from the solution space, there might still be some data reduction required to acquire more manageable results.

Multiple authors discuss a method that uses an adapted 'Utopia Point' method to determine a smaller set of more promising solutions from the Pareto Front by selecting the solutions that are "closest" in distance to the "ideal" Utopia point solution [3, 13, 14]. The Utopia point is a solution that performs the best with respect to all responses but is unobtainable. The distance to the Utopia point is dependent upon a selected distance metric where Lu et al. proved that using the Utopia point approach with an L_1-norm distance metric selects the same optimal solutions as the additive desirability function [3, 14]. Additionally, the multiplicative desirability function is equivalent to the Utopia point approach with the same L_1-

norm distance metric on the log scale.

2.4 Uncertainty

Chapman et al. recognize the need to account for uncertainty as response variability will affect the reliability of operating conditions, since future observed responses are not guaranteed to perform identically [13]. This can be combated using worst-case prediction intervals to capture the worst-case estimates of the responses and determine if the preferred solutions unacceptably differ. Their later article, which inspired this research, proposes an alternate method of incorporating uncertainty by sampling vectors of $\boldsymbol{\beta}$ parameters distributed MVN to simulate new response surfaces which can then be summarized for analysis. Their research is important as multi-objective optimization generally assumes deterministic solutions [2]. This deterministic assumption leads to an under defined solution space which can result in sub-optimal solutions.

2.5 Probability Distributions

As noted in Section 2.2.1, standard linear regression relies on the assumption that errors are independent and identically distributed (iid) random variables. Most commonly, the iid errors will be distributed normally. Additionally, the unbiased estimator, $\boldsymbol{b}$, can be used in place of $\boldsymbol{\beta}$ as the mean vector in the MVN sampling distribution. The normal distribution is used frequently in linear regression as it is a well-behaved distribution such that properly standardized sums of iid random variables will converge in distribution to the standard normal distribution as the sample size increases due to the Central Limit Theorem (CLT) [15]. The normal distribution also has a useful pivotal quantity as it is a part of the location-scale family. The pivotal quantity can be used to show that the t-distribution is appropriate to use when approximating samples from a normal distribution when variance is not known

[15]. This pivot is derived in subsection 2.5.1 and the subsequent generalization to the multivariate case is expanded upon in section 2.5.3. The multivariate distributions used in this research have been defined below as there are multiple generalizations.

2.5.1 Normal Pivotal Quantity

Let $X_1, \ldots, X_n$, be independent random samples from a normal distribution with mean μ and variance σ^2, i.e. $X_i \sim N(\mu, \sigma^2)$. The random variable X_i can be standardized by subtracting its mean and dividing by the standard deviation. This random variable is defined by

$$Z = \frac{X_i - \mu}{\sigma} \tag{13}$$

this Z statistic is distributed Standard Normal which is $N(0, 1)$ and is a pivotal quantity for X_i as it no longer relies on the given parameters, μ and σ^2 [15].

The sample mean and sample variance are defined as $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, respectively [15]. $\bar{X}$ and S^2 are independent random variables that are functions of the random sample, $X_1, \ldots, X_n$. They each come from a respective probability distribution called a sampling distribution. $\bar{X}$ is the sum of normally distributed random variables scaled by n which means it is also a normally distributed random variable where the sampling distribution is $N(\mu, \sigma^2/n)$. The sampling distribution for S^2 is $GAMMA(\frac{n-1}{2}, 2\sigma^2/(n-1))$. S^2 can be transformed to $(n-1)S^2/\sigma^2$ which is distributed as a chi-squared with $n-1$ degrees of freedom [15]. As was done with X_i, standardizing $\bar{X}$ provides the following pivot

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{14}$$

which is distributed $N(0, 1)$. Therefore, when σ^2 is known, inference about μ can be made using Z-score tables. In situations where inference about μ is desired but σ^2 is

unknown, S^2 is used as an unbiased estimator in Equation 14

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \tag{15}$$

Equation 15 is distributed as a t-Distribution with $\nu = n - 1$ degrees of freedom. Standardizing normally distributed random variables causes them to no longer rely on μ or σ^2 which makes it a pivotal quantity.

The t-Distribution converges in distribution to the standard normal distribution which allows it to be used as a pivotal quantity for the normal distribution. To show this, Equation 15 can be multiplied by σ/σ where, by definition, $\sigma^2 = \sum_{i=1}^{n}(X_i - \mu)^2/n$ and $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$ to get

$$\begin{aligned}
t &= \frac{\bar{X} - \mu}{S/\sqrt{n}} \cdot \frac{\sigma}{\sigma} \\
&= \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \cdot \sqrt{\frac{\sigma^2}{S^2}} \\
&= \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \cdot \sqrt{\frac{\sum_{i=1}^{n}(X_i - \mu)^2/n}{\sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)}} \\
&= \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \cdot \sqrt{\frac{(n-1)}{n-1+1} \cdot \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}} \\
&= \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \cdot \sqrt{\frac{\nu}{\nu+1} \cdot \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}
\end{aligned}$$

Taking the limit as $\nu \to \infty$, $\bar{X}$ converges in probability to μ which gives the following

$$\begin{aligned}
&\Rightarrow \lim_{\nu\to\infty} \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \cdot \sqrt{\frac{\nu}{\nu+1} \cdot \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}} \\
&= \lim_{\nu\to\infty} \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \cdot \sqrt{1 \cdot \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}} \\
&= \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \cdot \sqrt{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sum_{i=1}^{n}(X_i-\mu)^2}} \\
&= \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \qquad (16)
\end{aligned}$$

where $\frac{\nu}{\nu+1}$ tends to 1 by applying L'Hopital's Rule. Equation 16 is the pivot to the Standard Normal Distribution for the sample mean from Equation 14. Thus, it can be said that the t-distribution converges in distribution to the standard normal distribution as n approaches infinity. This allows normally distributed mean values to be estimated and tested with small sample sizes and variance is unknown. It also allows the t-distribution to be used as the sampling distribution to approximate a normal population when variance is not known. It is possible to show graphically that as ν increases, the random samples from the t-distribution approach that of a normal distribution as can be seen in Figures 3(a-d).

In Figure 3, the squares are 50 observations sampled from a $N(0,1)$ distribution and the crosses are 50 observations sampled from a t-distribution with 3, 5, 10, and 20, degrees of freedom for a, b, c, and d, respectively. The t-Distribution has larger tails, which is caused by the larger variation as shown in the definitions of σ^2 and S^2. This prevents overconfidence when using it as an alternative to the normal distribution.

Clearly, as the t-Distribution has more degrees of freedom, it better approximates the normal distribution eventually converging in distribution.

2.5.2 Multivariate

The MVN and MVt are multivariate generalizations of the Normal and t-Distribution, respectively. They incorporate covariance, $\mathbf{\Sigma}$, and p, which allows for sampling p correlated variables at once.

Let $\boldsymbol{X}$ be a vector of p random variables with mean vector, $\boldsymbol{\mu}$, and covariance matrix, $\mathbf{\Sigma}$, then the pdf of the MVN is defined as

$$f(\boldsymbol{x}|\boldsymbol{\mu}, p, \mathbf{\Sigma}) = \frac{1}{(\sqrt{2\pi})^p |\mathbf{\Sigma}|^{1/2}} e^{-(\boldsymbol{x}-\boldsymbol{\mu})' \mathbf{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})/2} \tag{17}$$

where $\boldsymbol{x} \in \mathbb{R}$, $\boldsymbol{\mu} \in \mathbb{R}$, $p \in \mathbb{N}$, and $\mathbf{\Sigma}$ is positive definite [16]. If $\mathbf{\Sigma}$ is a diagonal matrix, the MVN can be factored such that

$$f(\boldsymbol{x}|\boldsymbol{\mu}, p, \mathbf{\Sigma}) = f(x_1|\mu_1, \sigma_1^2) \cdot f(x_2|\mu_2, \sigma_2^2) \cdot \cdots \cdot f(x_n|\mu_n, \sigma_n^2)$$

which shows that the MVN becomes the product of independent Normal Distributions when there is no covariance between variables [15].

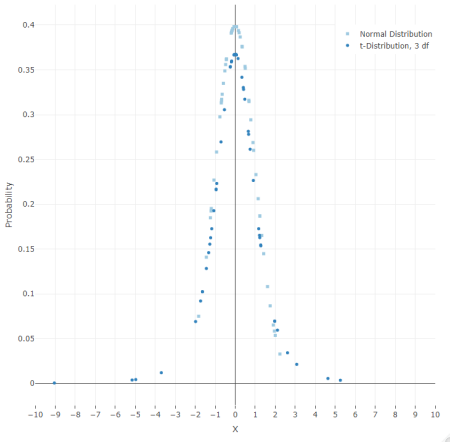
Let $\boldsymbol{Y}$ be a vector of p random variables with shift vector, $\boldsymbol{\mu}$, and correlation matrix, $\mathbf{R}$, then the pdf of the MVt used in this research is defined as

$$f(\boldsymbol{y}|\boldsymbol{\mu}, \nu, p, \mathbf{R}) = \frac{\Gamma(\frac{\nu+p}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{(\sqrt{\nu\pi})^p |\mathbf{R}|^{1/2}} (1 + \frac{1}{\nu}(\boldsymbol{y}-\boldsymbol{\mu})' \mathbf{R}^{-1} (\boldsymbol{y}-\boldsymbol{\mu}))^{-(\nu+p)/2}$$

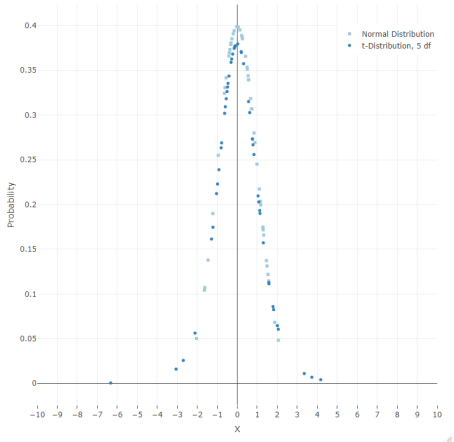
more commonly written as

$$f(\boldsymbol{y}|\boldsymbol{\mu}, \nu, p, \mathbf{R}) = \frac{\Gamma((\nu+p)/2)}{(\nu\pi)^{p/2} \Gamma(\nu/2) |\mathbf{R}|^{1/2}} \left[1 + \frac{1}{\nu}(\boldsymbol{y}-\boldsymbol{\mu})' \mathbf{R}^{-1} (\boldsymbol{y}-\boldsymbol{\mu})\right]^{-(\nu+p)/2} \tag{18}$$

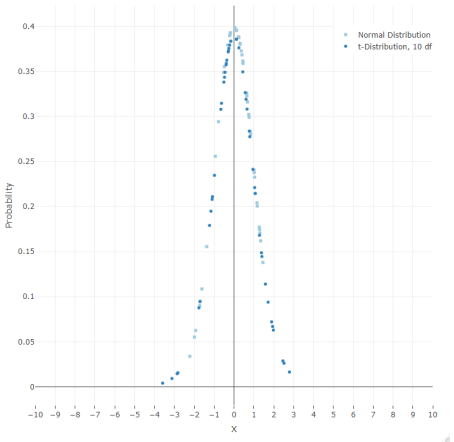
where $\boldsymbol{y} \in \mathbb{R}$, $\boldsymbol{\mu} \in \mathbb{R}$, $\nu \in \mathbb{N}$, $p \in \mathbb{N}$, and r_{ij} are entries of $\mathbf{R}$ where $-1 \leq r_{ij} \leq 1$.



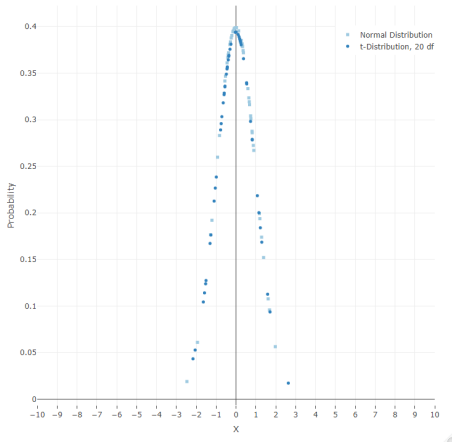
(a) 3 Degrees of Freedom



(b) 5 Degrees of Freedom



(c) 10 Degrees of Freedom



(d) 20 Degrees of Freedom

Figure 3. Normal Distribution Samples with t-Distribution Samples Overlay

$\mathbf{R}$ has corresponding $\mathbf{\Sigma}$ as the covariance matrix [4]. There are multiple derivations of the MVt described in Kotz' Multivariate t book, however, the distribution seen in Equation 18 was chosen for a few reasons. It is the most common and natural form of the MVt distributions and similar to the MVN, it is a direct generalization of the univariate t-Distribution. Furthermore, if $\boldsymbol{X} \sim MVN(\boldsymbol{\mu} = \mathbf{0}, p, \mathbf{\Sigma})$ and if $\nu S^2/\sigma^2 \sim \chi^2_\nu$, independent of $\boldsymbol{X}$, then $\boldsymbol{Y} = S^{-1}\boldsymbol{X} + \mu$, which implies $X|S = s \sim MVN(\boldsymbol{\mu}, p, (1/s^2)\mathbf{\Sigma})$ [4, 17]. As s^2 is known to be $\frac{\nu}{\nu-2}$, it is trivial to see that the numerator is larger which causes the covariance also to be larger than that of the MVN

which is desired to ensure there is enough variability when covariance is unknown and sample covariance is used as an estimator.

2.5.3 Multivariate Generalization

Estimating a single random variable or multiple independent random variables that are normally distributed was shown in section 2.5.1. This section recalls the Normal Distributions pivot to the t-Distribution and expands into the multivariate case. The Hotelling T^2 test was developed by Harold Hotelling. The intuition behind this test is to take Equation 15 and square it to get

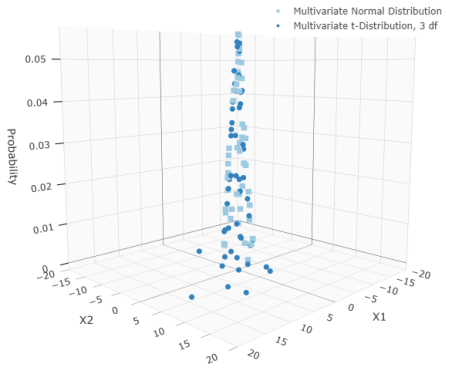
$$t^2 = n(\bar{X} - \mu)(S^2)^{-1}(\bar{X} - \mu)$$

Then by replacing $\bar{x}$ and μ with a vector of sample means, $\bar{\boldsymbol{x}}$, and expected values, $\boldsymbol{\mu}$, as well as replacing the sample variance, S^2, with the sample covariance matrix, C, the Hotelling T^2 test is derived as

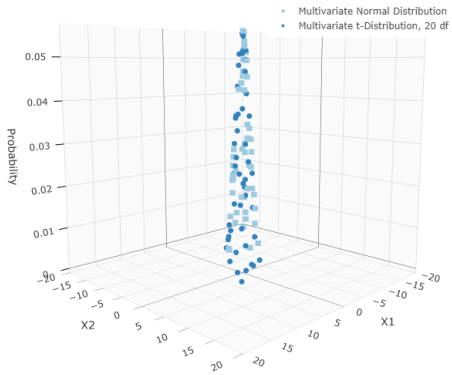
$$T^2 = n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{C}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) \tag{19}$$

which can be used to test the mean vector of a MVN in the absence of a known covariance matrix. This multivariate test implies that by being able to test means with a sample covariance matrix, samples from a MVt should be able to approximate samples from a MVN similar to the univariate case. Kotz explains that the MVt is better equipped to handle real-world data as it has larger tails which captures additional variability in the absence of the population covariance matrix [4]. He also explains that the limiting distribution of the MVt as $\nu \to \infty$ is the MVN with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ [4]. This asymptotic relationship has been demonstrated in a similar way for the multivariate case in Figure 4 as it was for the

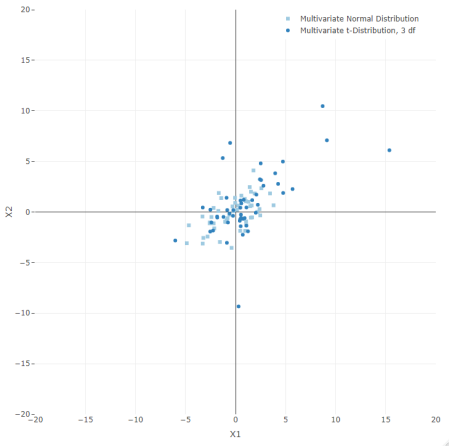
univariate case where the squares are 50 samples distributed MVN and the crosses are 50 samples distributed MVt with 3 and 20 degrees of freedom for Figure 4a and b, respectively. Figure 4c and d present a two-dimensional top-down view of the 3 and 20 degrees of freedom cases to show how the two distributions have very similar appearances in spread with the MVt generating more values in its tails which



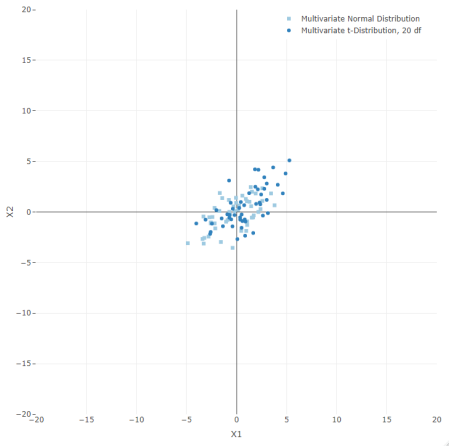
(a) 3 Degrees of Freedom



(b) 20 Degrees of Freedom



(c) 3 Degrees of Freedom Top-Down View



(d) 20 Degrees of Freedom Top-Down View

Figure 4. Multivatiate Normal Distribution Samples with Multivatiate t-Distribution Samples Overlay

become tighter as ν increases. Using the MVt when variance is unknown has also been explored previously for simultaneous confidence bounds on regression coefficients that are distributed MVN with a common variance-covariance matrix [18]. This exemplifies the necessity for the additional variance which accompanies the MVt when sampling mean vectors that are distributed MVN with a common unknown variance. Therefore, for a vector distributed MVN with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, the appropriate sampling distribution when covariance is unknown and sample size is small is MVt with shift vector $\boldsymbol{\mu}$ and scale matrix $\frac{\nu}{\nu-2}\boldsymbol{C}$ where $\boldsymbol{C}$ is the sample covariance matrix.

2.6 Summary

RSM and DOE are an excellent framework to use when optimizing products and processes. With the appropriate assumptions, linear regression can approximate response surfaces based on collected data to provide inference on the effects of significant process variables. The primary concern with optimizing these approximated response surfaces is that current techniques generally assume a deterministic surface and do not account for uncertainty. The lack of uncertainty may cause the optimization to give sub-optimal solutions which performed well for that single approximated surface from the sample but behaves poorly overall. Chapman et al. propose methods to incorporate uncertainty using a combination of PFO and the MVN distribution. It is argued that using the MVN distribution still results in overconfidence of solution spaces as it assumes a known variance which is seldom available. Instead, a MVt distribution should be used to ensure enough variability is in each random sample. Chapter IV will examine the differences in solutions between using a MVN and MVt distribution as the sampling distribution for $\boldsymbol{\beta}$.

III. Methodology

3.1 Overview

This chapter details an updated method to incorporate uncertainty into PFO. Chapman et al. partitions the process into steps 0 through 2c. As this research expands on Chapman's, it is beneficial to outline a similar process in order to appropriately compare sampling distributions. The topics covered are model construction, simulating response surfaces, Pareto fronts, desirability functions, and sampling distribution comparisons [3].

3.2 Mean Model Construction

An experiment with k process variables is designed and conducted to collect data for m responses. The results from the experiment is used to fit m response surface models of the form $\boldsymbol{y}_r = \boldsymbol{X}_r\boldsymbol{\beta}_r+\boldsymbol{\varepsilon}_r, r = 1, \ldots, m$ using linear regression which are then checked for normality and constant variance assumptions. The $\hat{\boldsymbol{\beta}}_r$ vector and mean square error ($\hat{\sigma_r^2}/MSE_r$) from each response surface model are recorded in addition to the $(\boldsymbol{X}_r'\boldsymbol{X}_r)^{-1}$ matrix. To remain consistent with Chapman et al. this original model will be called the 'mean model' (MM).

3.3 Simulating Response Surfaces

As mentioned in section 2.2.1, the unbiased parameters calculated using OLS, $\boldsymbol{b}$, are sampled from a MVN$(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$. Two sets of s vectors for each of the m responses are randomly sampled and denoted as $\boldsymbol{b}_r^*$. The first set will use Chapman's proposal and be sampled from a MVN distribution with mean vector, $\boldsymbol{b}_r$, and covariance matrix, $\hat{\sigma_r^2}(\boldsymbol{X}_r'\boldsymbol{X}_r)^{-1}$. The second set will be sampled from the MVt distribution with shift vector, $\boldsymbol{b}_r$, and scale matrix, $\frac{\nu_r}{\nu_r-2} \cdot \hat{\sigma_r^2}(\boldsymbol{X}_r'\boldsymbol{X}_r)^{-1}$ where $\nu_r = n - p_r$. These

random samples will be consistent with the model parameter estimates by using the appropriate mean/shift vectors and covariance/scale matrices [3]. Clearly, the MVN and MVt distributions use the same mean vector and only differ in their scale by a covariance matrix that is scaled by the degrees of freedom from error in the respective regression models. The randomly sampled coefficients are then used to simulate the m response surfaces for each response using a grid of solution locations in $\boldsymbol{X}$-space within the boundaries of the designed experiment.

3.4 Pareto Front

Once several surfaces are simulated, Pareto fronts are found to remove dominated points from each of the s surfaces. This step greatly reduces the sample space as it removes any combinations in the $\boldsymbol{X}$-space that should not be selected in a particular response surface as they under perform compared to other contenders. The Pareto fronts can then be summarized across all s surfaces by calculating the frequency of occurrence for each solution within the $\boldsymbol{X}$-space. The combinations that appear with high frequency, $\geq 90\%$ are considered promising locations. The solutions considered promising at this step are recorded and maintained for comparison.

3.5 Desirability

The next step requires finding more general solutions with respect to a decision makers needs. More specifically, this step determines the frequency of solutions that appear for a given set of weights. The Pareto front removes several solutions that should not be considered at all but the preferred solution may change based on which response is considered more important than the others. To conduct this, desirability functions can be used. Section 2.3.3 discussed the two different desirability functions that combine multiple objectives into one tractable function. The multiplicative de-

sirability function is used in this research to place penalties on the poor performance of a single objective. The desirability functions generally use a given upper, lower, and target value from the response to determine the 'best' and 'worst' of each response for scaling. Due to the estimation uncertainty in this situation, the bounds need to be larger to account for the extra variability introduced with the random sampling [3]. Chapman et al. uses the 95% prediction bounds characterized by

$$\boldsymbol{x}_0'\boldsymbol{b} \pm t_{0.975,n-p}\sqrt{MSE(1+\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0)} \tag{20}$$

where $\boldsymbol{x}_0$ is a single solution vector and $\boldsymbol{b}$ are the original OLS estimators for $\boldsymbol{\beta}$. The individual desirability of each solution is calculated according to the proper desirability function outlined in section 2.3.3. For an objective being maximized, the target, T, is determined by the largest upper bound of the 95% prediction intervals and the low, L, is the smallest lower bound. An objective trying to match a specific value will have a T that is a deviation of 0 and the upper, U, is the largest absolute deviation from the 95% prediction interval. An objective that is being minimized, T is the smallest lower bound of the 95% prediction interval and U is the largest upper bound. As these are only 95% prediction intervals, it is possible for there to be values that are outside of the aforementioned ranges. By definition of desirability functions, the values that are better than T receive a value of 1 and the values that are worst than L for maximizing and U for minimizing receive a value of 0. A combination of weight values for the m responses is compiled where each weight combination must sum to 1 and the change in each weight is $\frac{1}{30}$ to create a fine mesh that will allow higher resolution of changes in preference. After the multi-objective desirability function is calculated for each solution in all s response surfaces, the solution with the highest desirability is paired with a weight combination. This step is more subjective as it relies on different weighting schemes depending on the goal of the experiment. Using

the pair of solutions and weights, the area that a particular solution is desirable in can be found and used to provide guidance on which operating conditions should be used.

3.6 Comparison

There will be data on number of solutions as well as which solutions are still contenders at each sequential step of the PFO for the MM, the simulated MVN surfaces and the simulated MVt surfaces. This data will be compared to see which solutions appear after each step to see if there is a drastic difference between methods. After the desirability function step, solutions can be recommended, both by the preference that is placed on them and by hypothesized operating costs for that particular level.

- Fit m response surface models, $r = 1, \ldots, m$ using OLS,
- For each r, record vector of $\boldsymbol{b}_r$ coefficients and calculate sample covariance matrix $\boldsymbol{C}_r$,
- For each r, sample n random vectors from multivariate distribution using $\boldsymbol{b}_r$ as mean/shift and $\boldsymbol{C}_r$ as covariance/scale,
- Generate evenly spaced hyper-grid of $\boldsymbol{X}^*$ confined by region of designed experiment,
- For each n, for each r, calculate response surfaces, $\hat{y}_r = \boldsymbol{X}^* \boldsymbol{b}_r$,
- For each n, create Pareto front with m response surfaces to remove dominated points
- Calculate frequency of occurrence for each combination within $\boldsymbol{X}^*$.

Figure 5. Algorithmic steps to generate simulated response surfaces

3.7 Summary

This methodology is for the comparison of chosen multivariate distributions with respect to PFO, however, it should be easy to see that each step can be implemented with a single distribution. Figure 5 shows the algorithmic steps for generating the simulated response surfaces assuming a designed experiment has already been conducted and data is collected as the simulation of randomly sampled surfaces is the primary step where uncertainty is incorporated. Chapman et al. has already shown that the MM results in a sub-optimal solution space, however, utilizing a MVN distribution in this methodology results in too many solutions after the Pareto front step due to lack of enough variability which causes tighter samples. The MVt will shrink the number of solutions to be considered and provide more robust solutions by causing the frequency of solutions appearing on the Pareto front to become more dispersed.

IV. Analysis

4.1 Overview

This chapter presents analysis on an example problem used in Myers's RSM book and Chapman et al's research. This analysis is to support the research and methodology in previous chapters by examining a real problem that is impacted by lack of estimation error. The analysis begins with the construction of adequate response surface models. PFO is then executed sequentially to see how the solution space changes with the MM coefficients, the randomly sampled MVN coefficients, and the randomly sampled MVt coefficients. The analysis was conducted using R v3.5.3 with R Studio v1.2.1335.

4.2 Chemical Process Optimization Problem

The chemical process optimization problem provided in Table 1 was retrieved from Myers et al. RSM book and was used in the analysis of Chapter IV [1]. This Chemical Process Optimization Problem is comprised of a 13-run rotatable CCD in 2 process variables, time and temperature as well as 3 response variables, yield, viscosity, and number-average molecular weight. x_i are the coded process variables which were scaled using

$$x_{i,j} = \frac{\xi_{i,j} - [max(\xi_i) + min(\xi_i)]/2}{[max(\xi_i) - min(\xi_i)]/2} \tag{21}$$

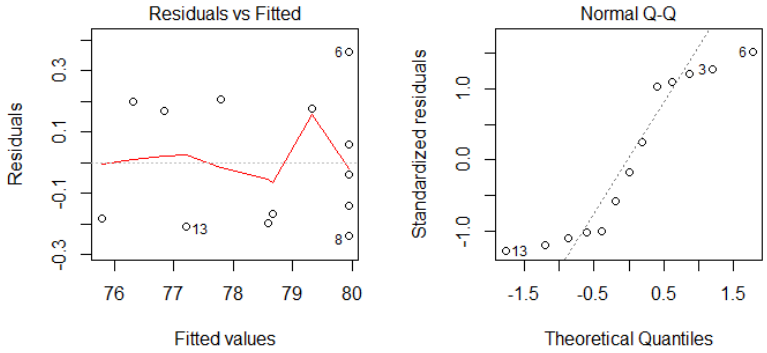
where the maximum and minimum of ξ_i are with respect to the 2^k factorial design rather than the entire CCD. The goals of this problem are to maximize yield, match a target of 65 for viscosity, and minimize molecular weight [1].

Table 1. Chemical Process Optimization Problem

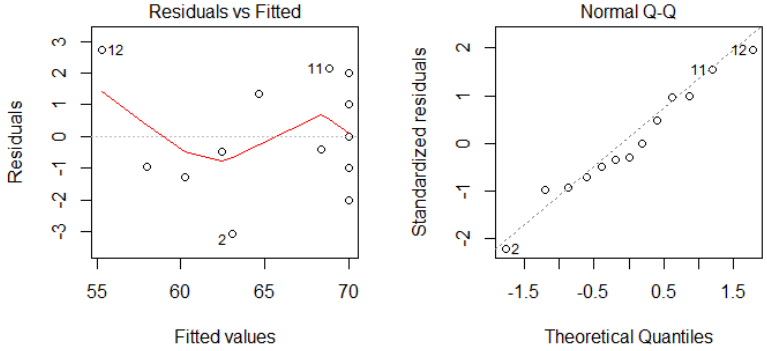
	Natural Variables		Coded Variables		Responses		
i	ξ_1(**Time**)	ξ_2(**Temp**)	x_1	x_2	y_1(**Yield**)	y_2(**Viscosity**)	y_3(**Molecular Weight**)
1	80	170	-1	-1	76.5	62	2940
2	80	180	-1	1	77.0	60	3470
3	90	170	1	-1	78.0	66	3680
4	90	180	1	1	79.5	59	3890
5	85	175	0	0	79.9	72	3480
6	85	175	0	0	80.3	69	3200
7	85	175	0	0	80.0	68	3410
8	85	175	0	0	79.7	70	3290
9	85	175	0	0	79.8	71	3500
10	92.07	175	1.414	0	78.4	68	3360
11	77.93	175	-1.414	0	75.6	71	3020
12	85	182.07	0	1.414	78.5	58	3630
13	85	167.93	0	-1.414	77.0	57	3150

4.3 Mean Model Construction

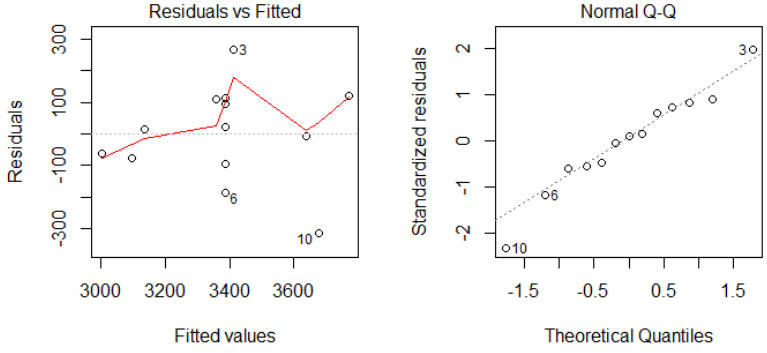
Three response surface models were fit using the sample data from table 1 using OLS to obtain the parameter estimate vector, $\boldsymbol{b}$, to be used in the multivariate sampling. Figure 6 shows the residual by fitted value plots and normal probability plots for all three models to check for the required assumptions to make inference on the parameters. While Figure 6a shows slightly heavy tails on the normal probability plots, there is no concern regarding the assumptions. The yield model was found to have significant first and second order terms with an insignificant interaction term that was maintained to preserve the form of a response surface. The viscosity model was found to have only significant second order terms, the first order terms were maintained due to model hierarchy with the interaction kept for the same reasoning as the yield model. The molecular weight model had only significant first order terms, thus the second order and interaction terms were removed to achieve a parsimonious model which is now a plane in the X-space. The three models obtained were



(a) Yield Response Surface Model



(b) Viscosity Response Surface Model



(c) Molecular Weight Response Surface Model

Figure 6. Residual by Predicted y and Normal Probability Plots for Response Surface Models

$$\hat{y}_1 = 79.9400 + 0.9951x_1 + 0.5152x_2 + 0.2500x_1x_2 - 1.3764x_1^2 - 1.0013x_2^2$$

$$\hat{y}_2 = 70.0002 - 0.1553x_1 - 0.9484x_2 - 1.2500x_1x_2 - 0.6873x_1^2 - 6.6891x_2^2$$

$$\hat{y}_3 = 3386.2 + 205.1x_1 + 177.4x_2$$

Figure 7 shows each of the models respective contour and perspective plots. Revisiting the objectives, this problem aims to maximize yield, meet a target value of 65 for viscosity, and minimize molecular weight. The contour plots can be used to observe that yield is maximized when (x_1, x_2) is around $(0.5, 0.5)$ and decreases in every direction from there, the viscosity target is met when (x_1, x_2) is around $(x_1, 0.75)$ or $(x_1, -0.75)$, and molecular weight is minimized when (x_1, x_2) are at their lowest values. This shows that none of these objectives can be optimized without at least one other objective being negatively affected. These correlated response variables provide an ideal example to demonstrate how the solution space is affected by estimation error.

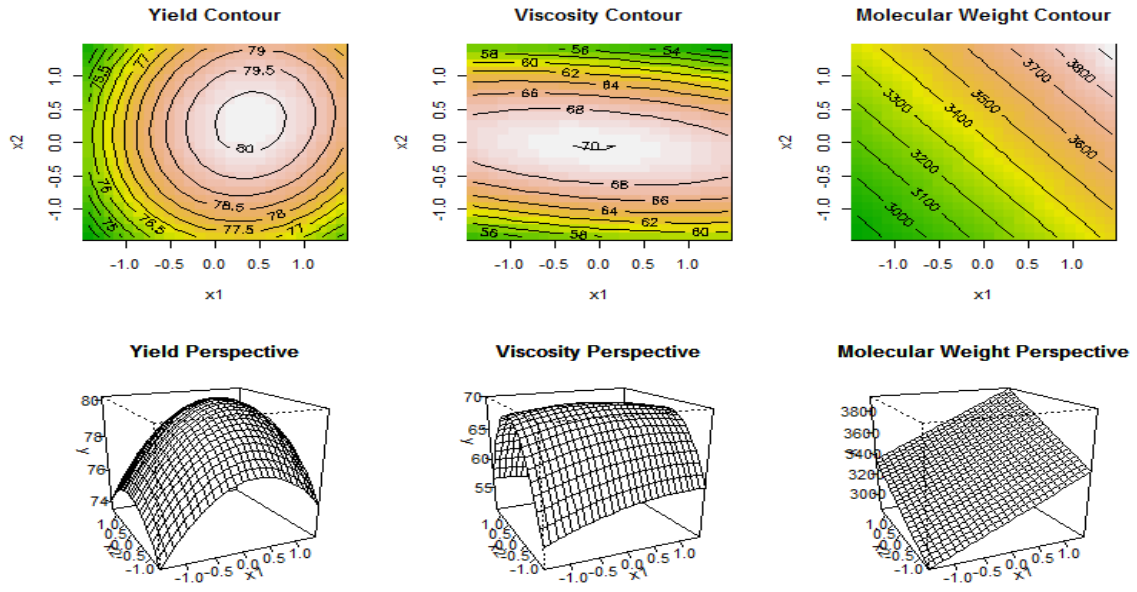


Figure 7. Contour and Perspective Plots of Responses

4.4 Simulating Response Surfaces

500 random samples for each response, $\boldsymbol{b}_r^*$, were generated using both the MVN and MVt distributions with $\boldsymbol{b}_r$ as the mean/shift vector, $\hat{\sigma}_r^2(\boldsymbol{X}_r'\boldsymbol{X}_r)$ for the MVN variance-covariance matrix and $\frac{\nu_r}{\nu_r-2}\hat{\sigma}_r^2(\boldsymbol{X}_r'\boldsymbol{X}_r)$ for the MVt scale matrix. The function in R for sampling from the MVN, 'rmvnorm', is relatively straight forward, however, the function for sampling from the MVt, rmvt, is a bit more nuanced where the guide by Hofert is helpful in avoiding common fallacies [19]. Using $\boldsymbol{b}_r^*$, 500 response surfaces are simulated using the grid of points that are evenly spaced by 0.1 units and confined by the CCD boundaries seen in figure 8 which totals 633 solutions. The numbers in parentheses on the left are the designated solution number for the leftmost solution which is then incremented as the x_1 value is increased by each unit of 0.1. In addition

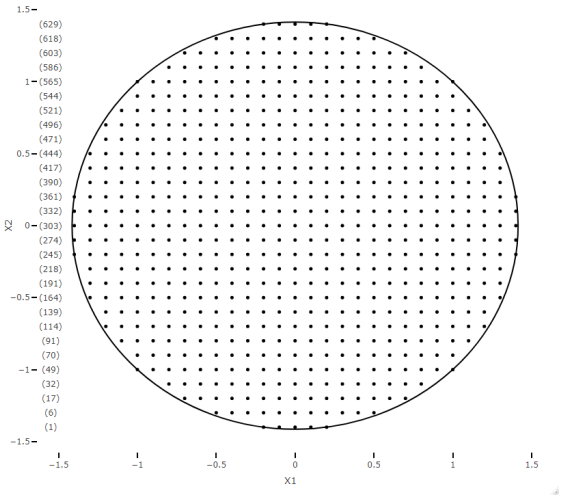


Figure 8. Grid of Possible Solutions in X-Space

to all of the simulated response surfaces, the response surface for the MM coefficients was calculated. Something to note when comparing this research with Chapman's is that, most likely due to round-off differences, this research uses 633 points where Chapman uses 630 observations. This causes an offset by 1 for locations 70-565 where Chapman would have 69-564 and then an offset by 3 for locations 586-633

where Chapman would have 583-630.

4.5 Pareto Front

Starting with 633 solutions, each solution was compared to the others to find Pareto dominate points. Recall from section 2.3.4 that a Pareto dominate solution is one that is at least better for all responses and strictly better for at least one solution. This was first done with the MM response surface to find the deterministic solution space and then with each of the multivariate simulated response surfaces.

4.5.1 Mean Model Response Surface

The Pareto front found for the MM resulted in 181 out of 633 solutions. Figure 9 shows the shape within the $\boldsymbol{X}$-space where the black points are the solutions that are within the Pareto set and the gray points are dominated solutions. Within this Pareto front, the solution that performs best for yield is at solution 407 when $(x_1, x_2) = (0.4, 0.3)$ with a value of 80.21218, for viscosity it is at solution 515 when $(x_1, x_2) =$

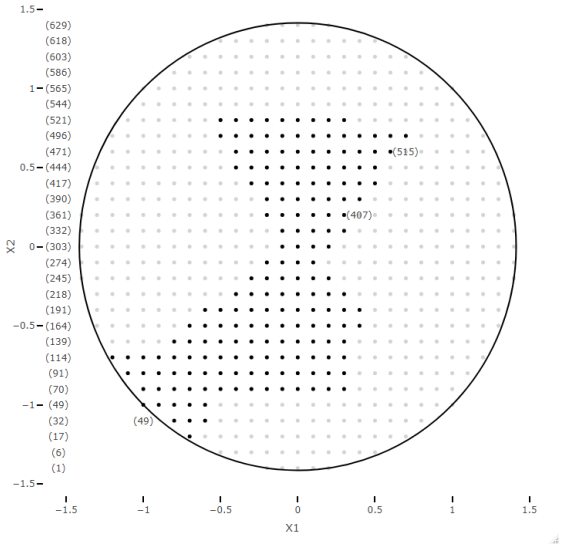
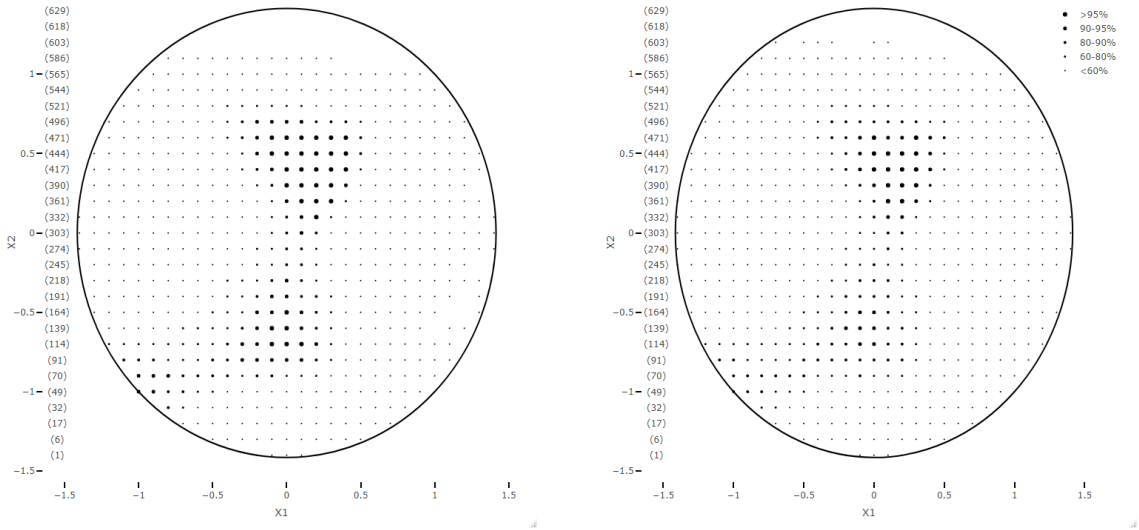


Figure 9. Mean Model Pareto Front

$(0.7, 0.7)$ with a value of 65.00068, and for molecular weight at solution 49 when $(x_1, x_2) = (-1.0, -1.0)$ with a value of 3003.661 which have all been labeled on figure 9.

4.5.2 Simulated Response Surfaces

Pareto fronts were found for all 500 surfaces for both the MVN and MVt simulated surfaces using the same seed sequence. The frequency of each solution was calculated by summing the total number of times a solution appeared on Pareto front out of the 500 surfaces. Figure 10 shows the frequency of times a solution appeared on the Pareto front for both (a) the MVN surfaces and (b) the MVt surfaces. The size of the point is characterized by the frequency of the point where a larger point indicates more appearances which can be seen in the legend. Solutions that did not appear on any of the Pareto fronts were omitted from these figures such as those on the upper boundary. The Pareto fronts for both the MVN and MVt have very similar shapes to that of the MM indicating that the response surfaces were indeed randomly sampled



(a) MVN Pareto Front Frequency (b) MVt Pareto Front Frequency

Figure 10. Frequency of Solution in Pareto Front Solution Space

to be consistent with the MM parameter estimates.

As expected, the MVt surfaces have more of a spread resulting in fewer observations appearing in the higher frequency categories. The total count of each category as well as the difference in count between the two distribution samples can be seen in table 2. Of the 633 solutions, 589 appeared on the Pareto front at all with only 63 appearing on 90% or more of the surfaces with the MVN samples. The MVt maintained 601 solutions, however, only 34 appeared on 90% or more of the surfaces. Something to note is that the 34 solutions from the MVt simulated surfaces are a subset of the 63 solutions that appear on 90% or more of the MVN simulated surfaces. This analysis finds that the range of values for each response for the MVN simulated surfaces are similar to that of Chapman's research when considering solutions appearing on at least one Pareto front. For yield the estimated value range is 75 to 81, 54 to 73 for viscosity, and 2721 to 3946 for molecular weight. The MVt simulated surfaces range from 75 to 81 for yield, 51 to 77 for viscosity, and 2557 to 4029 for molecular weight. As was shown in Chapman's research, the MVN simulated surfaces reduces

Table 2. Number of Solutions Appearing in Each Frequency Bracket for Both MVN and MVt Samples

Frequency Range (%)	**MVN Count**	**MVt Count**	**Difference**
(0-60)	426	441	15 (+)
[60-80)	60	69	9 (+)
[80-90)	40	57	17 (+)
[90-95)	32	19	13 (-)
[95,100]	31	15	16 (-)

the number of solutions to consider compared to the MM and prevents overconfidence of selecting operating conditions that may not be truly optimal. Similarly, using the MVt simulated surfaces performs the same task with respect to the MVN simulated surfaces as there are even fewer solutions to compare that can be considered as more robust.

4.6 Desirability Functions

This step of PFO was completed in the same manner as the Pareto front where the MM was completed first followed by the simulated response surfaces. The multi-objective desirability was calculated for each of the 181 solutions in the MM as well as the 63 and 34 solutions found on 90% or more of the Pareto fronts from the respective MVN and MVt simulated surfaces. The desirability functions outlined in section 2.3.3 were used where $r = 1$ for all three surfaces to place no particular emphasis. D was calculated using the multiplicative desirability function to penalize the poor performance of a single objective. Using the 95% prediction intervals defined in Chapter III the target T, upper U, and lower L bounds were chosen for each response. For yield, $T = 80.98343$ was the largest upper bound and $L = 75.39837$ was the smallest lower bound. For viscosity, $T = 0$ was the smallest absolute deviation from 65 and $U = 11.6295$ was the largest absolute deviation from 65. For molecular weight, $T = 2578.559$ was the smallest lower bound and $U = 4047.819$ was the largest upper bound.

496 weights combinations across the three responses were considered where each weight, w_i, is between 0 and 1 and are incremented by $\frac{1}{30}$. w_1 is the weight associated with yield, w_2 is the weight associated with viscosity, and w_3 is the weight associated with molecular weight. The most desirable solution for each weight combination was chosen in each response surface and then classified into a section. Figure 11 is a ternary plot that shows how the weight combinations were classified. Solutions selected as best when yield was emphasized, $w_1 > 0.5$, were labeled Section 1, those that were selected best for viscosity emphasis, $w_2 > 0.5$ were labeled Section 3, those selected as best for molecular weight emphasis $w_3 > 0.5$ were labeled Section 4, and those the solutions that were selected as best for equal emphasis, $w_1 \leq 0.5, w_2 \leq 0.5, w_3 \leq 0.5$, were labeled Section 2.

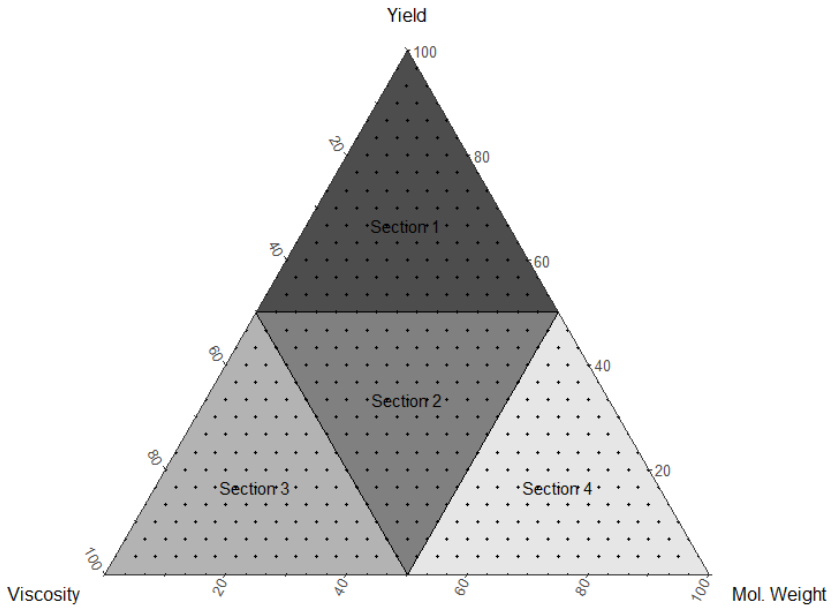


Figure 11. Classification of Desirability Weights

4.6.1 Mean Model Response Surface

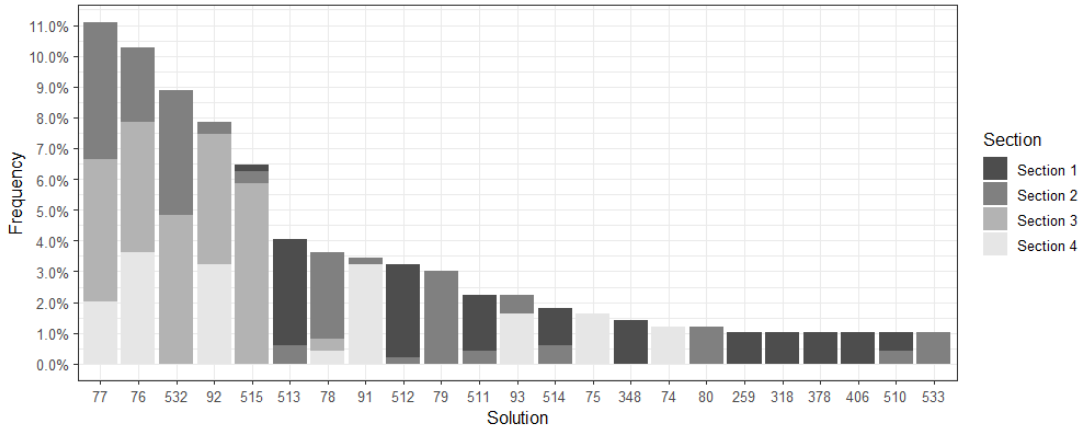
The MM response surface solution space was considerably reduced from 633 solutions to 181 solutions with the implementation of the Pareto front. After considering the 496 possible weight combinations, there were 78 solutions that was considered 'best' for at least one combination. The 78 were further reduced to 23 by only considering solutions that had been labeled best for 1% or more of the total weight combinations. Figure 12a shows the percentage of weight combinations that a solution, labeled on the x-axis, had the highest desirability in a specific section. For example, solution 77 had the highest desirability for 55 of the 496 weight combinations which is approximately 11%. Of those 55, 0 (0%) were in section 1, 22 (40%) were in section 2, 23 (42%) were in section 3, and 10 (18%) were in section 4.

This indicates that had the analysis chosen to only use the MM response surface with PFO, solution 77, $(x_1, x_2) = (-0.3, -0.9)$, would be a fine operating condition recommendation when a decision maker places more importance on viscosity, molecular weight, or an approximately equal emphasis on all three. If yield is of utmost importance, than perhaps solution 513, $(x_1, x_2) = (0.5, 0.7)$, would be preferred as it performs best for the most weight combinations in section 1. These are of course without considering the possible costs required to operate at those specific solutions.

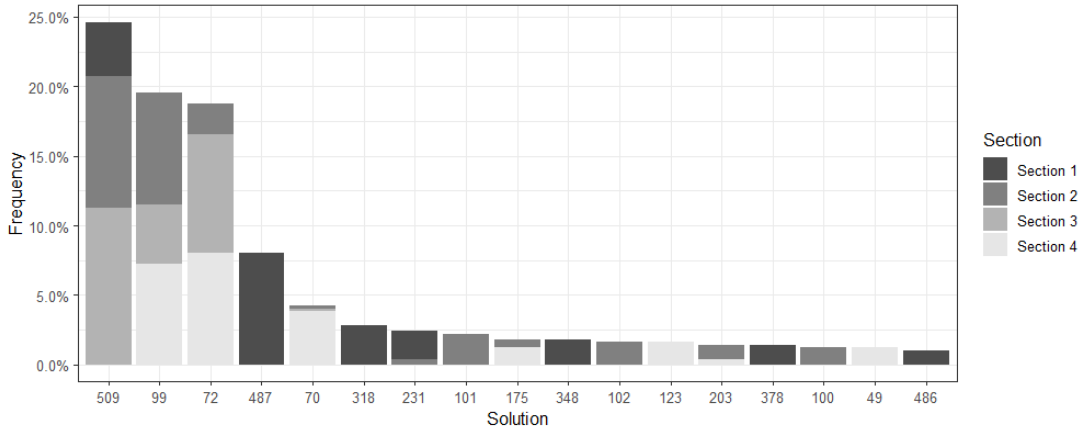
4.6.2 Simulated Response Surfaces

The multi-objective desirability for all 496 weight combinations was calculated for the subset of 63 and 34 solutions found on 90% or more of surfaces generated from the MVN and MVt, respectively. For each simulated surface, the solution with the highest desirability was recorded as the best solution. The best solutions were then summarized across all simulated surfaces for each weight combination by selecting the solution that appeared the most frequently for a particular weight as the overall best for that weight combination. Figure 12b,c portray identical information to figure 12a for the simulated response surfaces. Similar to the MM, the solutions labeled best for fewer than 1% of the total weight combinations were truncated from the right side of these figures as they would not be considered favorable solutions.

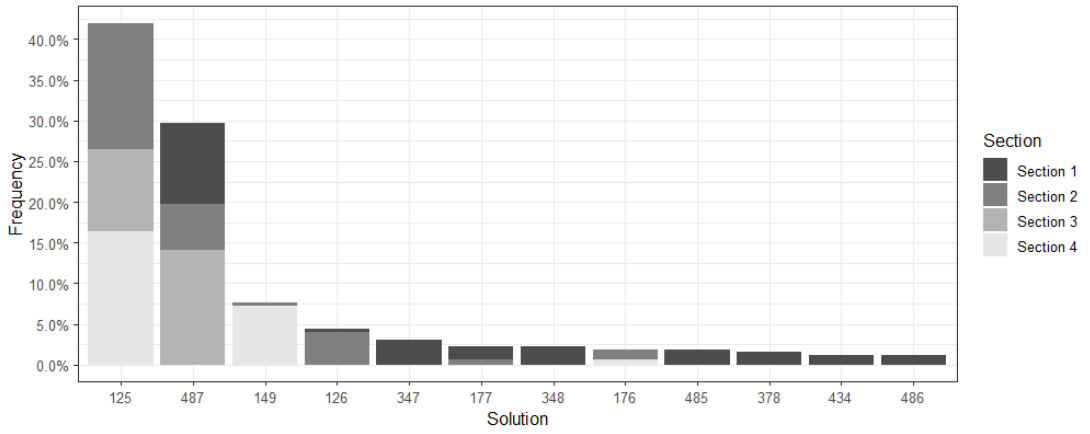
Figure 12b shows that with the variability included by sampling from a MVN distribution, solution 509, $(x_1, x_2) = (0.1, 0.7)$, is now the one that performs best for 122 of the 496 weight combinations, or a little less than 25%. 19 (16%) of those weight combinations were in section 1, 47 (39%) were in section 2, and 56 (46%) were in section 3 with none in section 4 which would require a shift of focus to solution 99, $(x_1, x_2) = (-0.3, -0.8)$, or 72, $(x_1, x_2) = (-0.8, -0.9)$. The variability included by sampling from a MVt distribution seen in figure 12c changes the preferred result



(a) MM Desirability



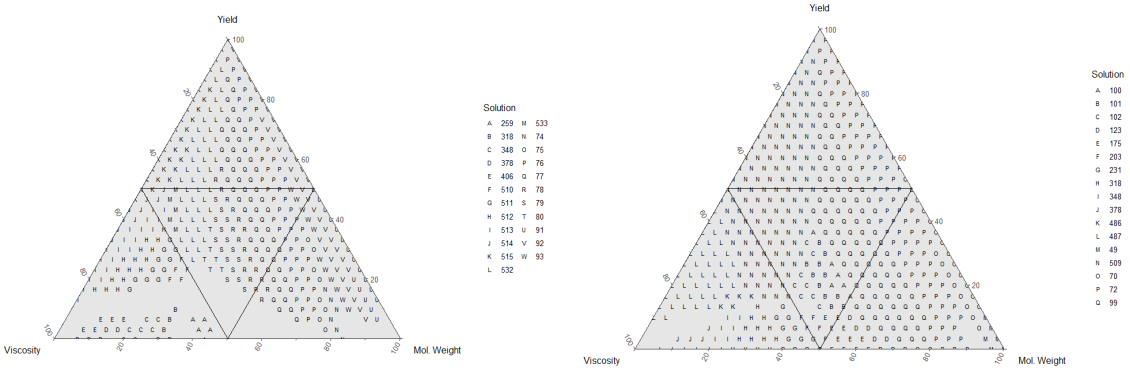
(b) MVN Desirability



(c) MVt Desirability

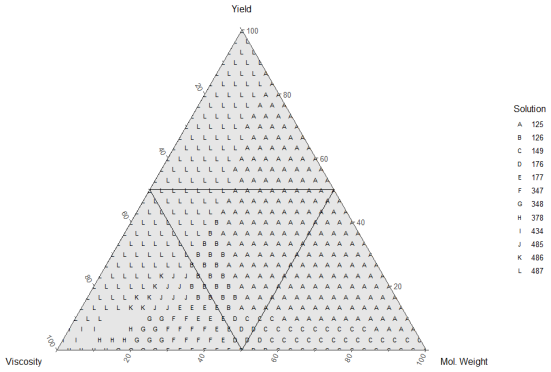
Figure 12. Percentage of Total Weight Combinations a Solution is Labeled 'Best'

once again to solution 125, $(x_1, x_2) = (-0.1, 0.7)$. Solution 125 is labeled best for 208 of the 496 weight combinations which is around 42%. 77 (37%) are in section 2, 50 (24%) are in section 3, 81 (39%) are in section 4, and none are in section 1 but solution 487, $(x_1, x_2) = (0.4, 0.6)$ can be recommended if yield is higher priority.



(a) MM Desirability Ternary

(b) MVN Desirability Ternary



(c) MVt Desirability Ternary

Figure 13. Area of Desirability Weights Solutions Considered 'Best' Most Frequently

The ternary plots in Figure 13 show the desirability weight combinations where a solution was most desirable for the MM, MVN surfaces, and MVt surfaces. The percentages on these plots directly correspond to the frequencies from figure 12 to provide better clarity of the solution space. This is shown in the MVt Desirability Ternary plot with solution 125. As previously noted, this solution accounts for roughly 42% of the 496 solutions and is denoted by 'A' in the ternary plot which appears to cover a little less than half. The desirability weight combinations on Figure 13 that appear to be empty are the solutions that obtained the best desirability for less than 1% of the weight combinations.

Figure 14 shows the location of the solutions with the highest percentage of weight combinations from the MM, MVN, and MVt data in $\boldsymbol{X}$-space, the solutions with labels above them are those with high frequency from Figure 12. The points appear on either the MM, MVN, MVt, or on some combination of the three each with their own respective shape. Each of these solutions agree with the shape found in the Pareto fronts. The top 6 solutions from the MM, 76, 77, and 92 in the the lower left

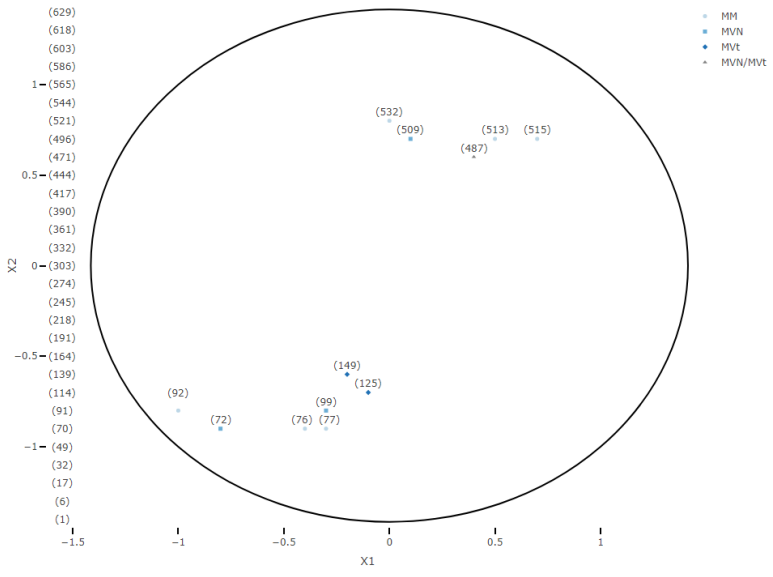


Figure 14. Solutions Labeled as Best for at least 1% of Weight Combinations from MM, MVN, and MVt Data

quadrant and 513, 515, 532 in the upper right quadrant, are labeled best for about 49% of all the weight combinations. The top 4 solutions, 72 and 99 in the lower left quadrant as well as 487 and 509 in the upper right quadrant, from the MVN simulated surfaced are labeled best for about 71% of the weight combinations. The top 3 solutions from the MVt simulated surfaced, 125 and 149 in the lower middle as well as 487 in the upper middle portion, account for about 79% of all the weight combinations.

4.7 Discussion

As the MVt distribution is more appropriate to be used in this analysis as covariance is unknown, the fact that there are only low frequency solutions in the lower left quadrant of the Pareto front of figure 10b that result in no solutions after the desirability function step implies it is not an ideal zone for operating conditions. The solutions stemming from the MM and MVN simulated surfaces in the lower left quadrant perform well for three of the four sections of weight combinations and they are also associated with lower time and temperature, which is conjectured to be associated with lower cost resulting in overconfidence of the region with one of these solutions being more likely to be selected for sustained operations. As this is a rotatable design, the prediction variance is the same for all points of $\boldsymbol{X}$ that are the same distance from the design center. The prediction variance will also be the lowest near the center of the design due to the 5 center points stabilizing the variance in the center which increases towards the boundaries. The increased prediction variance as the boundaries are approached make solution 72, 92, and 515 particularly less appealing. As the prediction variance is essentially a function of Euclidean distance from the center of the design, the MVt surfaces are able to capture the inflation of variability moving towards the edges and restricts the preferred solutions to be those

that are closer to the center. The MVt surfaces are also maintaining solutions similar to that of the MM and MVN which is desired as it supports that the MVt should be used as a sampling distribution when covariance is unknown. Figure 15 shows

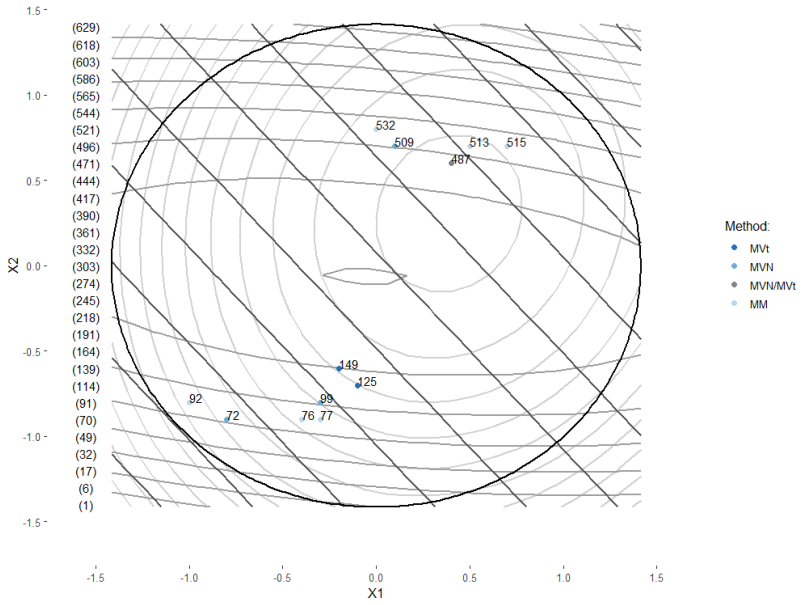


Figure 15. Overlaid Contour Plots for Yield, Viscosity, and Molecular Weight with Solutions

the overlaid contour plots which can be compared to figure 14 to see that the two clusters primarily remain near the optimal regions for viscosity where their position is dictated by whether yield (top cluster) or molecular weight (bottom cluster) are more important.

4.8 Summary

The three responses from the chemical process problem were correlated with conflicting objectives which required mathematical techniques, such as PFO, to optimize them concurrently. The MM, MVN simulated surfaces, and MVt simulated surfaces coupled with PFO all demonstrated similar appearances after the Pareto front step. The MM required a traditional Pareto front process, which maintained 181 out of

633 solutions. The MVN and MVt simulated surfaces used the traditional Pareto front process for each of their 500 surfaces, but were then summarized across 500 to find which solutions appeared most frequently. Of the solutions on the summarized Pareto fronts of the MVN and MVt, only those that appeared on the Pareto fronts 90% or more of the time were taken forward to the desirability function step. This resulted in 63 solutions from the MVN, and 34 solutions from the MVt.

The desirability functions removed all but 78 solutions, which was further reduced to 23 by only looking at solutions with 1% or more of weight combinations of the MM response surface. Performing a similar process as that of the Pareto front where each simulated response surface had desirability calculated and then summarized over the 500 with solutions having less than 1% of the total weight combinations removed resulted in 17 possible solutions from the MVN and 12 possible solutions from the MVt. The location of these solutions indicate that the MVt provides slightly more sound solutions that are less likely to result in a overconfident answer.

V. Conclusion

5.1 Conclusion

Prior research has shown that the stochastic nature of response surface models characterized by ANOVA requires a stochastic approach to handling optimization. Chapman et al's research seeks to handle the issue of uncertainty through the introduction of a methodology which incorporates uncertainty through the simulation of 500 response surfaces where the parameters are distributed MVN. This research has expanded upon that methodology by arguing for the sampling distribution of the 500 response surfaces to be distributed MVt. In reviewing the statistical theory driving the use of a t-distribution as a sampling distribution for normally distributed samples when variance is unknown, an analogous argument was made for the multivariate case.

The MVN distribution is a direct generalization of the normal distribution and the MVt distribution used in this research is a direct generalization of the t-distribution. The multivariate generalizations coupled with the Hotelling T^2 statistic, which is derived from a t-statistic, as well as literature suggesting that the MVt is better suited for 'real-world' problems that also converges to a MVN as the sample size increases provides a sufficient foundation to suggest that the use of a MVN is not strictly appropriate in Chapman et al's methodology.

Applying Chapman et al's methodology to the same chemical process optimization problem with the MM, MVN simulated surfaces, and the MVt simulated surfaces confirmed the hypothesis that the solution space would differ. The three avenues resulted in some similar solutions with an overall Pareto front that have shapes resembling a bowtie, but none of the final preferred solutions to be considered for selection were exactly the same. The MVN simulated surfaces was able to account for some of the

variability within the response surfaces model but allowed for solutions to appear more frequently on the Pareto front. Without the known covariance, it is impossible to determine whether the amount of MVN simulated surfaces solutions on the Pareto front is appropriate. The MVt simulated surfaces generates similar surfaces as the MM and MVN model with increased variability resulting in fewer solutions appearing frequently on the Pareto front. This indicates that by using the MVt distribution as the sampling distribution in place of the MVN distribution for parameter estimation in PFO when sample size is small and covariance is unknown, the experimenter can be sure that the analysis is more conservative and less likely to lead to overconfident solutions.

5.2 Future Research

Future research can be done in a few avenues.

- Additional Multivariate Distribution Theory

  More research into the theory of multivariate distributions, specifically, finding more of a connection between the MVN and MVt distributions. This was outside the scope of this research.

- Higher Resolution Solution Space

  The resolution of the solution space is lacking in the desirability function step of the PFO. Similar to Chapman's research, analysis into specific pockets of the weighting combinations can be accomplished as well as looking into more DOE/RSM concepts with fraction of design plots.

- Evaluate Other Second Order Designs/Problems

  This example case was on a CCD with a classic textbook problem. Applying this methodology to a BBD designed problem or a problem in industry may

be beneficial. The exploration of additional factors should be accomplished as they prevent the easy use of contour plots for confirmation as well as rotatability and excess center runs should be altered as they make the solution space rather simple as the prediction variance simply increases as the edges are approached in this problem.

- Desirability Function Confidence Intervals

  Research methods to create efficient confidence intervals on desirability, either through a known distribution and parametric methods or using nonparametric statistics. Finding a confidence interval on a desirability function would allow the analyst to determine if a certain solution is truly more desirable or not. If solution A is said to be the most desirable but costly and has an overlapping confidence interval with solution B that is not as costly, then the two solutions are not statistically different and solution B can be used.

Other topics within this research can likely be improved upon or found as well.

Bibliography

1. R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook, *Response Surface Methodology*. Hoboken, NJ: Wiley, 2016.

2. L. Costa, I. A.C.P. Espírito Santo, and P. Oliveira, "Uncertainty on Multi-Objective Optimization Problems," in *AIP Conference Proceedings*, pp. 775–778, 2011.

3. J. L. Chapman, L. Lu, and C. M. Anderson-Cook, "Incorporating response variability and estimation uncertainty into Pareto front optimization," *Computers and Industrial Engineering*, 2014.

4. S. Kotz and S. Nadarajah, *Multivariate t-distributions and their applications*. Cambridge University Press, 2004.

5. M. H. Kutner, C. J. Nachtsheim, and J. Neter, *Applied Linear Regression Models*. New York, NY: McGraw-Hill/Irwin, 2004.

6. D. Montgomery, E. Peck, and G. Vining, *Introduction to Linear Regression Analysis*. Wiley Series in Probability and Statistics, Wiley, 2015.

7. C. M. Anderson-Cook, C. M. Borror, and D. C. Montgomery, "Response surface design evaluation and comparison," 2009.

8. D. C. Montgomery, *Design and Analysis of Experiments*. Hoboken, NJ: Wiley, 2017.

9. W. M. Carlyle, D. C. Montgomery, and G. C. Runger, "Optimization problems and methods in quality control and improvement," *Journal of Quality Technology*, vol. 32, no. 1, pp. 1–17, 2000.

10. E. C. Harrington, "The desirability function," *Industrial quality control*, vol. 21, no. 10, pp. 494–498, 1965.

11. G. Derringer and R. Suich, "Simultaneous optimization of several response variables," *Journal of quality technology*, vol. 12, no. 4, pp. 214–219, 1980.

12. E. D. Castillo, D. C. Montgomery, and D. R. McCarville, "Modified desirability functions for multiple response optimization," *Journal of Quality Technology*, vol. 28, no. 3, pp. 337–345, 1996.

13. J. L. Chapman, L. Lu, and C. M. Anderson-Cook, "Process optimization for multiple responses utilizing the pareto front approach," *Quality Engineering*, vol. 26, no. 3, pp. 253–268, 2014.

14. L. Lu, C. M. Anderson-Cook, and T. J. Robinson, "Optimization of Designed Experiments Based on Multiple Criteria Utilizing a Pareto Frontier," *Technometrics*, 2011.

15. G. Casella and R. L. Berger, *Statistical Inference*. Belmont, CA: Duxbury, 2002.

16. A. C. Rencher, *Methods of Multivariate Analysis*. New York, NY: Wiley, 2017.

17. P.-E. Lin, "Some characterizations of the multivariate t distribution," *Journal of Multivariate Analysis*, vol. 2, no. 3, pp. 339 – 344, 1972.

18. S. John, "On the evaluation of the probability integral of the multivariate t-distribution," *Biometrika*, vol. 48, no. 3/4, pp. 409–417, 1961.

19. M. Hofert, "On Sampling from the Multivariate t Distribution," *The R Journal*, vol. 5, no. 2, pp. 129–136, 2013.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
26–03–2020	Master's Thesis	Sept 2018 — Mar 2020

4. TITLE AND SUBTITLE

CHARACTERIZING UNCERTAINTY IN CORRELATED RESPONSE VARIABLES FOR PARETO FRONT OPTIMIZATION

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Calhoun, Peter A, 1st Lt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

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

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-MS-20-M-136

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

Intentionally Left Blank

10. SPONSOR/MONITOR'S ACRONYM(S)

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

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Current research provides a method to incorporate uncertainty into Pareto front optimization by simulating additional response surface model parameters according to a Multivariate Normal Distribution (MVN). This research shows that analogous to the univariate case, the MVN understates uncertainty, leading to overconfident conclusions when variance is not known and there are few observations (less than 25-30 per response). This research builds upon current methods using simulated response surface model parameters that are distributed according to an Multivariate t-Distribution (MVT), which can be shown to produce a more accurate inference when variance is not known. The MVT better addresses uncertainty in the parameters which can affect the frequency of treatments appearing on the Pareto front resulting in potentially different proposed solution spaces from that of the MVN.

15. SUBJECT TERMS

Response Surface Methodology, Design of Experiments, Statistics, Multivariate Statistics, Optimization

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Lt Col Beau A. Nunnally, Ph.D, AFIT/ENC
U	U	U	U	59	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636, x4394; beau.nunnally@afit.edu