The Pennsylvania State University

The Graduate School

Department of Industrial and Manufacturing Engineering

BAYESIAN METHODS FOR ROBUSTNESS IN PROCESS OPTIMIZATION

A Thesis in

Industrial Engineering and Operations Research

by

Ramkumar Rajagopal

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

August 2004

We approve the thesis of Ramkumar Rajagopal.

Date of Signature

Enrique del Castillo

Associate Professor of Industrial Engineering

Thesis Advisor

Chair of Committee

Tom M. Cavalier

Professor of Industrial Engineering

M. Jeya Chandra

Professor of Industrial Engineering

Steven F. Arnold

Professor of Statistics

Richard J. Koubek

Professor of Industrial Engineering

Head of the Department of Industrial and Manufacturing Engineering

ABSTRACT

The core objective of the research presented in this dissertation is to develop new methodologies based on Bayesian inference procedures for some problems occurring in manufacturing processes. The use of Bayesian methods of inference provides a natural framework to obtain solutions that are robust to various uncertainties in such processes as well as to assumptions made during the analysis. Specifically, the methods presented here aim to provide robust solutions to problems in process optimization, tolerance control and multiple criteria decision making.

Traditional approaches for process optimization start by fitting a model and then optimizing the model to obtain optimal operating settings. These methods do not account for any uncertainty in the parameters of the model or in the form of the model. Bayesian approaches have been proposed recently to account for the uncertainty on the parameters of the model, assuming the model form is known. This dissertation presents a Bayesian predictive approach to process optimization that accounts for the uncertainty in the model *form*, also accounting for the uncertainty of the parameters given each potential model. Both single response and multiple response systems are considered. The objective here is to optimize the model-averaged posterior predictive density (MAP) of the response where the weighted average is taken using the model posterior probabilities as weights. The MAP is thus used to maximize the posterior probability of obtaining the responses within given specification limits.

The Bayesian approach to model-robust process optimization is then extended to the

case where noise factors and non-normal error terms are present. Traditionally, in process optimization, methods such as the Dual Response Surface methodology are used in the presence of noise factors, and methods such as Robust Regression are used when the error terms are not normally distributed. In this dissertation, the idea of model-robustness using the Bayesian posterior predictive density is extended to cases where there is uncertainty due to noise factors and due to non-normal error terms.

The tolerance control problem is the inverse of the process optimization problem. Here, the objective is to find the specification or tolerance limits on the responses. We propose a Bayesian method to set tolerance or specification limits on one or more responses and obtain optimal values for a set of controllable factors. The dependence between the controllable factors and the responses is assumed to be captured by a regression model fit from experimental data, where the data is assumed to be available. The proposed method finds the optimal setting of the control factors (parameter design) and the corresponding specification limits for the responses (tolerance control) in order to achieve a desired posterior probability of conformance of the responses to their specifications.

In addition to process optimization and tolerance control, a new Bayesian method is presented for the multiple criteria decision making problem (MCDM). The usual approach to solving the MCDM problem is by either using a weighted objective function based on each individual objective or by optimizing one objective while setting constraints on the others. These approaches try to find a point on the efficient frontier or the Pareto optimal set based on the preferences of the decision maker. Here, a new algorithm is proposed to solve certain MCDM problems based on a Bayesian methodology. At a first stage, it is assumed that there are process responses that are functions of certain controllable factors or

regressors. At a second stage, the responses in turn influence the utility function of one or more decision makers. Both stages are modelled with Bayesian regression techniques. The methodology is applied to engineering design problems, providing a rigorous formulation to popular "Design for Six Sigma" approaches.

Although the research focusses on applications in process optimization, tolerance control and MCDM, some of the results can be directly applied to other applications such as process control. These and other ideas for further research are described in the concluding chapter of this dissertation.

Contents

List of Figures

List of Tables

List of Acronyms

RPD	Robust parameter design
DOE	Design of experiments
RSM	Response surface methodology
MAP	Model-averaged posterior predictive density
SMR	Standard Multivariate Regression
SUR	Seemingly Unrelated Regression
MCDM	Multiple criteria decision making
DM	Decision maker
CVD	Chemical vapor deposition
HPLC	High performance liquid chromatography

ACKNOWLEDGEMENTS

I take this opportunity to thank my advisor Dr. Enrique Del Castillo who guided me throughout the development of this thesis. His unique style of mentoring aided me in coming up with creative solutions to problems. My discussions with Dr. Del Castillo not only helped in solving current research problems but also in coming up with ideas for new problems. His commitment to research and attention to details are traits that I greatly admire in him and hope to emulate as I graduate.

It has been a tremendous honor to work with a Ph.D. committee comprising of Dr. Tom M. Cavalier, Dr. Jeya Chandra and Dr. Steven F. Arnold. I have learnt a great deal from them not only from their comments and critiques on my research proposal, but also from the classes I have taken with them. I thank them for their time and for being flexible in helping me schedule my proposal and defense. I also thank Dr. Natarajan Gautam for his comments on my thesis proposal.

I owe a debt of gratitude to my colleagues in the Industrial Engineering department, in particular, the Engineering Statistics Lab, with whom I have shared very useful and fruitful research discussions. My special thanks to Zilong Lian, Omer Arda Vanli, Guillermo Miro-Quesada and Nirmal Govind. I also convey my thanks to the faculty and staff of the Industrial Engineering department and the Statistics department at Penn State.

I thank Dr. John J. Peterson of GlaxoSmithKline Pharmaceuticals for sharing his ideas and co-authoring the paper on "*Model-Robust Process Optimization with Noise Factors and Non-normal Error Terms*". My thanks to Dr. R. Jean Ruth and Sharon Zielinski of

General Motors R&D Vehicle Development Research Lab for sharing their thoughts on the manuscript of the paper on "*A Bayesian Approach for Multiple Criteria Decision Making with application in Design for Six Sigma*" that was submitted to the European Journal of Operational Research.

Last but not the least, I thank my parents, R. Rajagopal and Sushila Rajagopal, and my sister, Anupama Rajagopal for their unconditional support and inspiration that keeps me constantly motivated to reach for the skies.

To My Parents

Chapter 1

Introduction

In any manufacturing process, there are typically one or more quality characteristics that are of interest to the process engineer or engineering designer. The quality characteristics that are the output of the process are referred to as *responses*. The responses are in general a function of a number of process parameters or factors that may or may not be under the control of the process engineer. The parameters that are under the control of the process engineer are called *control variables* or *control factors* and those that are not controllable are called *noise variables*. In addition to these, a third category called *noise factors* may also be present. These are noise variables that cannot be controlled at the "customer" level (whether this customer is the manufacturing plant or the end customer), but can be controlled under careful experimental conditions. For example, in the production of automotive tires, the type of driver and the driving conditions might be noise factors. Another example is in the polishing process in semiconductor manufacturing where the response might be the thickness of the silicon wafers. The control factors might be polishing pressure and rotation speed. A noise factor may be environmental temperature.

Statistical methods and other Operations Research techniques have been traditionally used in different stages of manufacturing processes to help improve quality, reduce costs, etc. The common applications for these methods in manufacturing include:

1. Statistical Process Modelling using design of experiments (DOE) and response surface methodology (RSM)
2. Process Optimization and Multiple Criteria Decision Making (MCDM)
3. Process Control
4. Tolerancing
5. Reliability and Fault Control
6. Scheduling

In this research, the contributions focus on applications in process optimization, MCDM and tolerancing, but some of the results presented may directly be applied to other applications. Although there are no new contributions presented in the statistical modelling and DOE/RSM area, traditional methods are extensively used throughout this dissertation in developing the new methods for process optimization, tolerancing and MCDM. Due to the stochastic uncertainties in the processes and the assumptions made during the analysis of any process, it is important that the solutions obtained using various statistical and operations research methods are *robust*. A robust solution is one which consistently produces desired responses that are insensitive to the assumptions made as well to the natural variabilities present in the system. There are different sources of uncertainty with respect

to which robustness can be achieved. These are discussed in the later chapters. The next section describes the problems in process optimization, tolerancing and MCDM that are addressed in the dissertation.

1.1 Dissertation Topics

1.1.1 Process Optimization

In manufacturing processes, it is usually desired to obtain responses that satisfy certain specifications, which are based on customer preferences. It is frequently the case that all the responses can be quantified and it will be assumed herein that they must lie between an upper and a lower specification limit. The objective of process optimization is then to find the settings of the control variables so that the process responses lie within these specification limits. It is possible to have responses for which the objective is to make them as large or as small as possible. This research will concentrate in the case when responses have specifications and the problem is process optimization. Other type of objectives in process optimization problems that are typically used include:

1. Target is best: The objective is to reduce the variability of the response around the given target which is a point rather than an interval

2. Smaller the better: The objective is to minimize the value of the response with a low variability around the minimum

3. Larger the better: The objective is to maximize the value of the response with a low variability around the maximum

Since not all combinations of settings of the control variables may be possible, the optimization should take into account any imposed constraints. In this research, the Bayesian predictive density is used to find the settings of the control factors that maximize the probability of conformance of the responses to their specification limits under the given constraints. Extensions are provided to handle noise factors and non-normal error terms.

1.1.2 Tolerancing

In process optimization, it is assumed that the specification limits are given, and the objective is to maximize the probability of conformance of the responses to the given limits. Tolerance control is the inverse problem. Here, the objective is to evaluate the specification or tolerance limits on the responses for managing their quality. The Bayesian method is used to find tolerance limits where the coverage or the probability of conformance of the response is at least equal to some given value. In addition, the problem we address finds the settings of the control factors that give the smallest such tolerance interval so that there is least variation among the conforming responses. Since the mean of the responses is also important, the optimization takes into account constraints on the possible region where the tolerance interval may lie. Based on these constraints the objective is to minimize the size of the tolerance interval with respect to the settings of the control factors such that the interval contains a given probability of conformance for the response. The problem is also solved for setting tolerance intervals simultaneously on multiple responses.

1.1.3 Multiple Criteria Decision Making

In practice, decision-making problems typically involve the consideration of two or more criteria that are often conflicting. These are referred to as Multiple Criteria Decision Making (MCDM) problems, where one has to take into account trade-offs between the conflicting criteria. The difference between MCDM problems and process optimization problems is that in the former, there is an additional function that is of interest, namely the utility function of the customer or the decision maker. The utility function can be thought of as a quantitative measure of the decision maker's satisfaction. As in process optimization, the responses are a function of the controllable factors. But in addition, the utility is a function of the outcome of the responses. Taking into account both these dependencies, the objective is to find the setting of the control factors that maximizes the utility of the decision maker. In this research, both the dependencies mentioned above are modelled using Bayesian regression. Since the utility function is obtained as a probability distribution, the MCDM optimization problem is formulated to find the settings of the control factors that maximizes the probability that the decision maker's utility function is at least equal to some given lower bound. The problem addressed in this research also extends the methodology to the case where there are multiple decision makers.

1.2 Research Objectives

The overall goal of the proposed research is to develop methodologies for process optimization, tolerance control and multiple criteria decision making (MCDM) that provide robust operating settings with respect to different sources of uncertainty. Sources of uncertainty

such as noise factors, non-normal error terms, uncertainty of parameters and uncertainty in the model form are considered. The specific objectives of the research presented in the dissertation are:

1. **Process Optimization for Single Response Systems:** In this case, it is assumed that there are multiple control factors in the process and a single response that is desired to lie within given specification limits. The goals here are to:

    - Develop a method for model-robust process optimization that evaluates the settings of the control factors, at which obtaining the response within the given specification limits is robust with respect to competing model forms that can be used to represent the process. In each of the competing model forms, the parameter estimates are also assumed to be uncertain.
    - Extend the model-robust Bayesian methodology for single response process optimization in the presence of noise factors.
    - Extend the model-robust Bayesian methodology for single response process optimization under the assumption of non-normal error terms. The method is extended for processes with t-distributed errors that have thicker tails than normally distributed errors.

2. **Process Optimization for Multiple Response Systems:** In this case, it is assumed that there are multiple control factors in the process and multiple responses that are desired to lie within each of their given specification limits. The goal is to extend the model-robust Bayesian methodology for process optimization to multiple response systems.

3. **Robust Tolerance Control and Parameter Design:** Develop a Bayesian approach for Robust Tolerance Control and Parameter Design.

4. **Multiple Criteria Decision Making:** Develop a Bayesian method for multiple criteria decision making (MCDM) problems with applications in "Design for Six Sigma".

1.3 Dissertation Outline

Chapter 2 reviews literature that is relevant to the topics covered in this research. The reviewed literature can be broadly classified into two categories. The first category includes current methods that are used in process optimization, in tolerancing and in multiple criteria decision making (MCDM). These methods are presented in order to demonstrate the contrast with the new methods presented in this research. The second category of literature reviewed includes published results that are used in developing some of the methodologies presented in this research.

In chapter 3, a new method is presented for process optimization using Bayesian model averaging techniques. The method provides a solution that is robust not only with respect to the uncertainty in the model parameters but also with respect to uncertainty in the model form. In this chapter, the methodology is presented for a single response system, with the assumption of normally distributed errors and the absence of noise factors.

Chapter 4 extends the model-robust Bayesian methodology for process optimization first to processes where there are noise factors in addition to control factors. The second extension is to processes where the error distribution is not normal. Instead, t-distributed errors that have thicker tails than the normal errors are assumed. The results developed

in this chapter are also for single response systems.

Chapter 5 extends the model-robust Bayesian methodology for process optimization of the previous chapters to the case of multiple response systems. Multiple response systems are classified into four categories depending on how the system is modelled. The chapter shows how, depending on the category, the complexity of the methodology for multiple response systems varies.

In chapter 6, a Bayesian method for tolerance control and parameter design is presented. This chapter presents a methodology to set tolerance or specification limits on one or more responses while at the same time identifying the value of the control factors that provide the desired limits.

Chapter 7 provides a new Bayesian approach for the multiple criteria decision making (MCDM) problem. In the problem considered here, there are process responses that are functions of certain control factors or regressors. In addition, the responses in turn influence the utility function of one or more decision makers. The method proposed in this chapter finds the optimal setting of the control factors by considering both the dependency of the responses on the control factors and that of the decision maker's utility on the responses, where both the dependencies are modelled using Bayesian regression.

All the methodologies presented in chapters 3-7 are illustrated by examples, the data for which come from published literature. Chapter 8 presents the summary of the research contributions as well as some ideas for future research.

Chapter 2

Literature Review

In this chapter, a review of the literature relevant to the topics in this dissertation is presented. A brief review of the Bayesian method of inference is provided followed by reviews of methods in process optimization, process tolerancing and multiple criteria decision making that are relevant to the remaining topics in this dissertation. A brief description of the algorithms for nonlinear optimization, which are used in the examples in the later chapters, is provided in appendix D.

2.1 Bayesian Methods of Inference

The overall idea of the Bayesian method of inference is to form posterior beliefs by updating prior beliefs based on the observed data [39, 58, 60]. The posterior beliefs are thus expressed as a conditional probability given the observed data. The Bayesian method of inference thus has two parts, the prior that depends upon subjective belief that the experimenter has before making any experimental observations, and the data that the experimenter observes

from the experiments. This method of inference follows from Bayes' Theorem given below.

Bayes' Theorem [8]: Suppose $\mathbf{y}$ is a vector of n observations of a response whose joint distribution $p(\mathbf{y}|\boldsymbol{\theta})$ depends on the value of k parameters $\boldsymbol{\theta}$. Suppose that $\boldsymbol{\theta}$ has a prior probability distribution $p(\boldsymbol{\theta})$. Then, given the observed data $\mathbf{y}$, the conditional distribution of $\boldsymbol{\theta}$ is given by Bayes' theorem:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})} \tag{2.1}$$

As $p(\mathbf{y}|\boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$ and not $\mathbf{y}$, it is called the likelihood function of $\boldsymbol{\theta}$ for given $\mathbf{y}$, and may be written as $l(\boldsymbol{\theta}|\mathbf{y})$. In other words, Bayes' theorem says that the distribution of $\boldsymbol{\theta}$ posterior to the data $\mathbf{y}$ is proportional to the product of the likelihood for $\boldsymbol{\theta}$ given $\mathbf{y}$ and the distribution of $\boldsymbol{\theta}$ prior to observing the data. Inferences about the unknown quantity $\boldsymbol{\theta}$ are then made from its posterior distribution.

The subjectivity involved in the Bayesian method of inference has been its biggest cause of criticism as opposed to non-Bayesian or *frequentist* methods of inference where the inference is based completely upon the observed data. However, a good argument of the natural inclusion of subjectivity in all scientific methods of inference is discussed in Press [58]. More recently, it is becoming increasingly common to accept both frequentist and Bayesian methods for their advantages rather than to choose one method over the other. Other advancements in the Bayesian method of inference includes the concept of *objective-Bayesian* methods. These are prior distributions that have been developed to reflect little or no prior information about the unknown parameters. Some of the results developed in this dissertation will be using the objective-Bayesian method. It is noted that sometimes mathematically identical results can be obtained when using objective priors as compared

to frequentist methods, however, the interpretation of the results is different based on the method of inference used [58]. Two other recent techniques to compute prior distributions are *empirical* and *hierarchical* Bayes techniques [12]. The empirical Bayes technique uses the observed data to estimate the parameters of the prior distribution in what may seem as a conflict to the Bayesian paradigm. The method hence draws a lot of criticism from staunch Bayesians as the prior in this case is in fact chosen posterior to observing the data. The hierarchical Bayesian approach models the lack of information on the parameters of the prior distribution through yet another prior distribution on the unknown parameters. The parameters of this distribution are called *hyperparameters*. This approach is generally more popular among Bayesians, but increases the analytical complexity of the methodology depending on how far down the hierarchy one goes.

Some other useful terms in the Bayesian methodology are given below [58]:

- **Improper priors**: Improper priors are priors that do not integrate to one over all values of the parameter. These priors are often used when there is little or no information a priori.

- **Vague priors**: Vague priors are improper priors that are uniform across all values of the unknown parameter. In other words, they express the prior belief that no value of the parameter $\boldsymbol{\theta}$ is more likely than any other value. For example, the vague prior on a parameter θ on $(-\infty, \infty)$ is

$$p(\theta) \propto \ constant,$$

and a vague prior on a parameter σ on $(0, \infty)$ is

$$p(\sigma) \propto \frac{1}{\sigma}.$$

- **Conjugate priors**: A family of distributions is said to make up a conjugate family of priors for a given likelihood, if the posterior also belongs to the same family of distributions.

- **Jeffrey's priors**: Jeffrey's priors are distributions such that the probabilities of the observable random variables are invariant to changes in the parametrization of the problem. Based on this requirement, it was shown by Jeffrey that the prior density for a parameter $\boldsymbol{\theta}$ must be of the form:

$$p(\boldsymbol{\theta}) \propto |\mathbf{J}|^{0.5},$$

where $\mathbf{J}$ is the Fisher information matrix associated with the likelihood function.

The Bayesian method facilitates a very natural way of making inferences on future values of the response using the posterior predictive distribution. Equation (2.1) gives the posterior distribution of the parameters $\boldsymbol{\theta}$ based on the prior distribution and the observed responses $\mathbf{y}$ in the form of the likelihood function. If y^* is a future value of the response that has not yet been observed, it is possible to make inferences on y^* from its posterior distribution given by

$$p(y^*|\mathbf{y}) = \int p(y^*|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}) \, d\boldsymbol{\theta}. \tag{2.2}$$

Thus, the posterior predictive distribution of the response also naturally accounts for the uncertainty in the true value of the parameters $\boldsymbol{\theta}$ as the expression uses the posterior distribution of $\boldsymbol{\theta}$ as opposed to point estimates that are used for prediction in frequentist approaches.

With this introduction, the next few sections describe both Bayesian and non-Bayesian

methods that exist in literature for process optimization, process tolerancing and multiple criteria decision making.

2.2 Review of Process Optimization Methods as Applied to Robustness Studies

In order to optimize a process, a transfer function model is required to relate the responses to the control variables. However, many advanced manufacturing processes are complex in nature and there is seldom a physical model to represent them. Traditionally (Figure 2.1), preliminary experiments are carried out using Design of Experiments (DOE) techniques, and a linear statistical model of the following form is fit to the data:

$$y_k = \mathbf{x_k}'\boldsymbol{\beta}_k + \boldsymbol{\epsilon}_k \tag{2.3}$$

where y_k is the k^{th} of a total of q responses, $\mathbf{x_k}$ are the control variables (regressors), $\boldsymbol{\beta}_k$ are the model parameters, and $\boldsymbol{\epsilon}_k$ are the error terms. Noise factors may be considered explicitly in the model, as described later in this dissertation. The settings for the control variables are then determined by optimizing the fitted model under the given constraints. A common methodology used in process optimization is Response Surface Methodology (RSM), where a sequence of first or second-order polynomial models (response surfaces) are fit to the data and optimized [49].

From an application point of view, process optimization is becoming increasingly important in several types of industries. The emphasis on improving quality while at the same time reducing costs has led to the adoption of statistical and optimization approaches re-

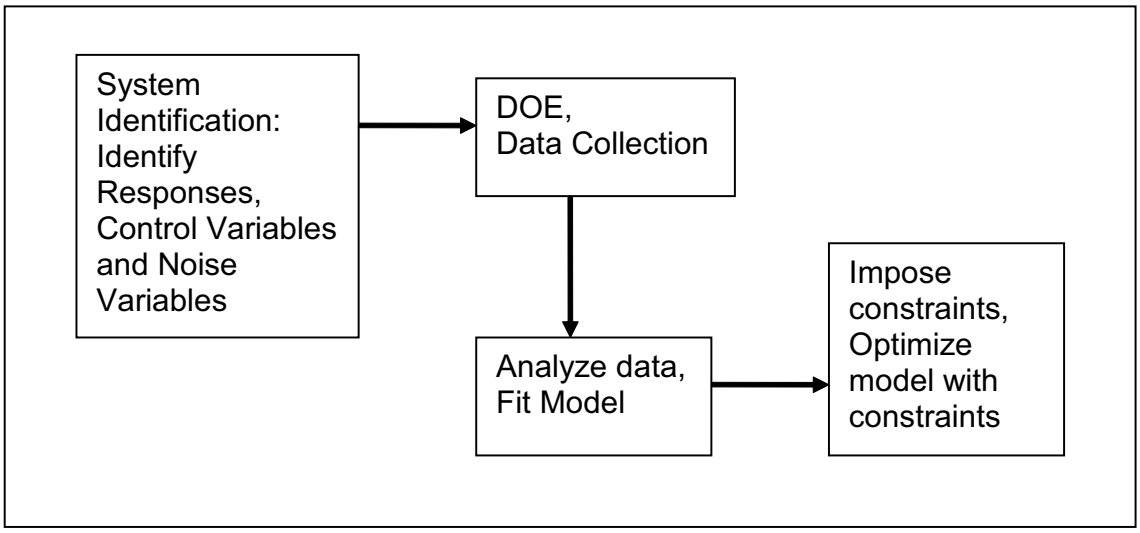

Figure 2.1: Traditional Process Optimization

ferred by the name of *Six Sigma* techniques [9]. In the automotive industry, for example, the design for six sigma (DFSS) methodology uses traditional process optimization techniques to identify set points for controllable factors earlier in the process or assembly line in order to meet quality requirements for the end customer.

There are shortcomings in the traditional process optimization approach mentioned in the previous section (Figure 2.1), many of which arise as a result of various underlying assumptions in the methodology. Due to the stochastic uncertainty involved in any manufacturing process, it is important that the optimal solutions be *robust*. A robust solution is one which consistently produces desired responses that are insensitive to the assumptions made as well to the natural variabilities present in the system. There are different sources of uncertainty with respect to which robustness can be achieved. Some of these have been addressed in literature and are discussed below.

1. **Robustness with respect to the optimal setting of the control variables**:

Instead of the traditional approach where the optimal solution is a single point estimate, this approach provides a confidence region around the estimated optimal settings ([10, 11, 16, 55]) in order to give an idea of how sensitive, or robust, the predicted optimal value of the response is around the optimal set point of the control variables. The confidence region provides information on how flat the response is around the optimal setting. Thus, as long as there is a good model fit, a larger confidence region around the estimated optimal setting implies that the solution is robust to changes around the optimal values of the control variables.

2. **Robustness with respect to non-normal error terms**: The traditional model-fitting approach uses ordinary least square (OLS) estimates for the parameters, $\boldsymbol{\beta}$, in the model shown in equation (2.3). However, when the standardized residuals from the fitted model have large magnitudes (large outliers), the estimates using OLS are poor. This happens when the noise term, $\boldsymbol{\epsilon}$, deviates from the assumed normal distribution (i.e., fatter tails in the distribution). In such cases, a *robust regression* approach is used [61], where the parameter estimates are obtained using methods such as Least Absolute Deviations (norm), M-Estimators, Least Median Squares, or Ranked Residuals ([61, 19]) which are less sensitive to non-normal errors than OLS. In statistics, this is the most common use of the word "robustness".

3. **Robustness with respect to noise factor variability**: In practice, there are some noise variables that cannot be controlled at the "customer" level (whether this customer is the manufacturing plant or the end customer), but can be controlled under careful experimental conditions. These are referred to as *noise factors*. For

example, in the production of automotive tires, the type of driver and the driving conditions might be noise factors. The objective then is to find a solution that is robust to the variation in the noise factors. This is the so-called Robust Parameter Design (RPD) problem and was first formulated by Genichi Taguchi ([66, 67, 68]). This method involves designing an experiment that involves varying both control and noise factors in a *crossed array*, with the control factors in the inner array and the noise factors in the outer array. The data is used to fit a model (such as equation 2.4 below) to a single response as a function of both the controllable and noise factors. This is called the *Dual Response Surface* approach (refer [7, 15, 49]), as this model is then used to get the mean (equation 2.5) and variance (equation 2.6) models (response surfaces) for the estimated response. These models follow from assuming that the noise factors, $\mathbf{z} \sim (0, \sigma^2\mathbf{I})$, and from taking the expected value and variance with respect to $\mathbf{z}$ in the fitted equation, that is:

$$\hat{y}(\mathbf{x}, \mathbf{z}) = b_o + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x} + \mathbf{z}'\mathbf{c} + \mathbf{x}'\mathbf{\Delta}\mathbf{z}, \tag{2.4}$$

so taking expected value and variance , we get

$$E_{\mathbf{z}}[\hat{y}(\mathbf{x}, \mathbf{z})] = b_o + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x} \tag{2.5}$$

and

$$Var_{\mathbf{z}}[\hat{y}(\mathbf{x}, \mathbf{z})] = \sigma_z^2(\mathbf{c}' + \mathbf{x}'\mathbf{\Delta})(\mathbf{c} + \mathbf{\Delta}'\mathbf{x}) + \hat{\sigma}^2, \tag{2.6}$$

where b_o, $\mathbf{b}$, $\mathbf{B}$, $\mathbf{c}$ and $\mathbf{\Delta}$ are estimated parameters, and $\hat{\sigma}^2$ is the estimated variance of the error terms (i.e., the mean squared error). The mean and variance models of the estimated response are used to arrive at the optimal solution by finding a region that

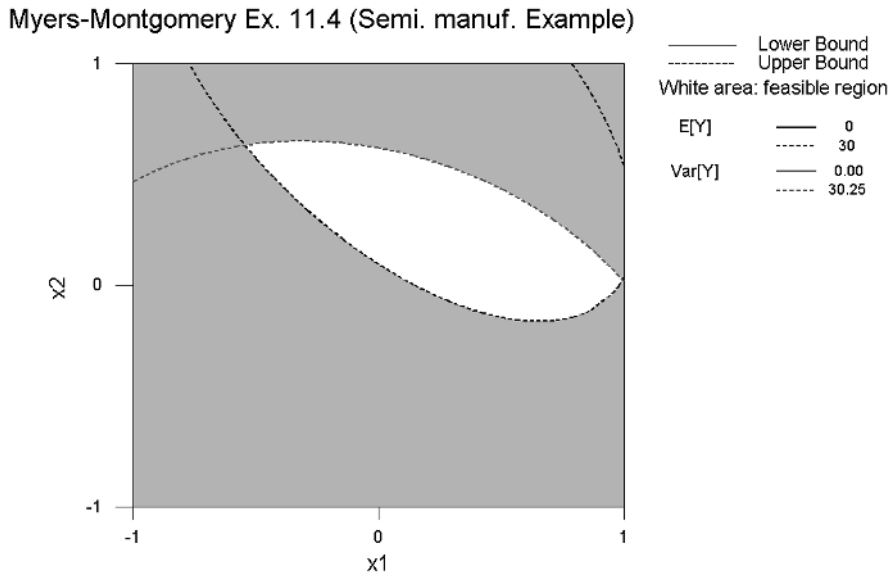

Figure 2.2: RPD example - Myers and Montgomery [49]

minimizes the variance and meets the specifications for the mean of the response. The example in figure 2.2, taken from Myers and Montgomery [49] shows a semiconductor manufacturing process where there is a single response with 2 control factors and 3 noise factors. The specification limits set by the manufacturer are $E_z[\hat{y}(\mathbf{x}, \mathbf{z})] < 30$, and $Var_z[\hat{y}(\mathbf{x}, \mathbf{z})] < 5.5$. The unshaded region in figure 2.2 gives the region where the manufacturer's specifications seem to be satisfied based on the fitted model.

In the case where the noise factors are correlated, i.e., $\mathbf{z} \sim (0, \mathbf{V_z})$, where $\mathbf{V_z}$ is the variance-covariance matrix, Myers and Montgomery [49] show that the variance model for the estimated response is

$$Var_{\mathbf{z}}[\hat{y}(\mathbf{x}, \mathbf{z})] = \hat{\mathbf{l}}'(\mathbf{x})\mathbf{V_z}\hat{\mathbf{l}}(\mathbf{x}) + \hat{\sigma}^2, \tag{2.7}$$

where $\hat{\mathbf{l}}(\mathbf{x}) = (\mathbf{c} + \boldsymbol{\Delta}'\mathbf{x})$. But, as the result in equation (2.7) is not an unbiased

estimator of $Var_{\mathbf{z}}[y(\mathbf{x}, \mathbf{z})]$, the authors recommend using instead an estimator that includes a bias-correction term:

$$Var_{\mathbf{z}}[\hat{y}(\mathbf{x}, \mathbf{z})] = \hat{\mathbf{l}}'(\mathbf{x})\mathbf{V}_{\mathbf{z}}\hat{\mathbf{l}}(\mathbf{x}) + \hat{\sigma}^2(1 - \text{tr } \mathbf{V}_{\mathbf{z}}\mathbf{C}), \tag{2.8}$$

where $\mathbf{C} = \text{Var}[\hat{\mathbf{l}}(\mathbf{x})]/\sigma^2$. Miro-Quesada and Del Castillo [47, 48] show that even though the variance estimate in equation (2.8) is unbiased, it can give a negative variance estimate, especially in regions where the coded controllable factors are far away from the origin. The authors reason that this is because the estimate in equation (2.8) considers only the variability in the noise factors, but not the variability in the parameter estimates in $\hat{\mathbf{l}}(\mathbf{x})$. They propose to use the estimate given by

$$\widehat{Var}_{\mathbf{z},\boldsymbol{\beta}^*}[\hat{y}(\mathbf{x}, \mathbf{z})] = \hat{\mathbf{l}}'(\mathbf{x})\mathbf{V}\hat{\mathbf{l}}(\mathbf{x}) + \hat{\sigma}^2(\mathbf{x}^{(\mathbf{m})\prime}\mathbf{C}_{\mathbf{x}^{(\mathbf{m})}}\mathbf{x}^{(\mathbf{m})}), \tag{2.9}$$

where $\boldsymbol{\beta}^*$ is the vector containing all the parameters in equation (2.4), $\mathbf{x}^{(\mathbf{m})}$ is the vector containing all the regressors associated with the control factors, and $\mathbf{C}_{\mathbf{x}^{(\mathbf{m})}}$ is the sub-matrix, corresponding to just the control factors, of the design matrix that includes both the control and the noise factors.

Need for a Bayesian predictive approach: In the example shown in figure 2.2, it is recommended, based on equations (2.5) and (2.6), to operate at any point in the unshaded region in order to produce responses that meet the specifications. However, the optimal setting $\mathbf{x}^*$ using equations (2.5) and (2.6) does not allow us to predict what fraction of future responses (e.g., proportion of products in a manufacturing process) will fall within the specifications as these equations give only the "mean models" (i.e., they give only point-estimate values for the mean and the variance of the response at $\mathbf{x}^*$). In other words, there can be no inference made about

the *reliability* of the process. A natural way to optimize any process from a quality and reliability standpoint is thus to maximize the probability of conformance of the predicted responses to their specification limits [56]. This can be achieved using a Bayesian predictive methodology. The benefits of using this methodology are that, (a) the posterior predictive density of the responses can be used to make inferences on their future values, thus giving us a means to calculate the probability of conformance of the future responses, (b) the methodology takes into account the mean and the variance of the responses including their correlation structure, and (c) the methodology takes into account uncertainty in the model parameters.

4. **Robustness with respect to parameter estimates**:

   Peterson [56] proposes a Bayesian approach for process optimization that considers uncertainty in the estimated model parameters. This technique involves obtaining the posterior predictive density of the responses based on the assumed model, and maximizing the probability of obtaining the predicted responses within specification limits.

   Peterson [56] assumes the following Standard Multivariate Regression (SMR) model [35]:

$$\mathbf{y} = \mathbf{x}'\mathbf{B} + \mathbf{u}, \tag{2.10}$$

   where $\mathbf{y}$ is a $q \times 1$ vector of the responses, $\mathbf{x}$ is a $p \times 1$ vector of the control factors and $\mathbf{u}$ is a $q \times 1$ vector of the error terms that may be correlated between the responses. In this context, the predictive density of a future response vector $\mathbf{y}^*$ at a given setting

of the control factors $\mathbf{x}^*$ is defined as [57]:

$$P(\mathbf{y}^*|\mathbf{x}^*,\ data) = \int\int L(\mathbf{y}^*|\mathbf{x}^*,\ data, \mathbf{B}, \mathbf{\Sigma}) P(\mathbf{B}, \mathbf{\Sigma}|data)\ d\mathbf{B}\ d\mathbf{\Sigma} \tag{2.11}$$

where L is the likelihood function, $P(\mathbf{B}, \mathbf{\Sigma}|data)$ is the posterior distribution of the model parameters, and $\mathbf{\Sigma}$ is the variance-covariance matrix of the error term. It is noted here that the uncertainty in the model parameters is considered by assuming that $\mathbf{B}$ and $\mathbf{\Sigma}$ are random variables and evaluating their posterior distributions using Bayes' theorem as follows:

$$P(\mathbf{B}, \mathbf{\Sigma}|data) \propto L(\mathbf{B}, \mathbf{\Sigma}|data)\ P(\mathbf{B}, \mathbf{\Sigma}) \tag{2.12}$$

where, $L(\mathbf{B}, \mathbf{\Sigma}|data)$ is the likelihood given the data, and $P(\mathbf{B}, \mathbf{\Sigma})$ is the prior distribution of the model parameters.

The objective function for the optimization problem used by Peterson [56] is:

$$\max_{\mathbf{x}^*}\ P(\mathbf{y}^* \in \mathbf{A}|\mathbf{x}^*,\ data) = \int_{\mathbf{A}} P(\mathbf{y}^*|\mathbf{x}^*,\ data)\ d\mathbf{y}^* \tag{2.13}$$

where $\mathbf{A}$ is a given specification region. This idea was extended by Miro-Quesada *et al.* ([46, 47]) for the case of uncertainty in model parameters in the presence of noise factors. Two cases of model classes are considered, the SMR case (as mentioned above) and the Seemingly Unrelated Regression (SUR) case [74]. The SUR model differs from the SMR model in that the former assumes that the vector of regressors $\mathbf{x_k}$ for response y_k, $k \in (1, ..., q)$ is different for different k, while the latter assumes that $\mathbf{x_k}$ is the same for all $k \in (1, ..., q)$. In case of the SMR model, the predictive density under a diffuse prior is shown by Press to have a closed form equal to a multivariate $\boldsymbol{t}$-distribution [57]. In case of the SUR model, Miro-Quesada *et al.* ([46, 47]) use a

Gibbs sampling procedure adapted from Percy [53] as there is no closed form available for the predictive density. We can use this methodology to calculate the posterior predictive density of the response in the example discussed earlier from Myers and Montgomery [49]. For example, the point (0,0) in the unshaded region in figure 2.2 gives a value of probability of conformance less than 0.5, whereas the probability of conformance is maximized at a value of 0.61 at the point (0.1,0.8) which is not in the unshaded region. This is an instance of specifications that are too demanding.

2.3 Review of Model Averaging Techniques

Though the area of model averaging is not new, it has not been applied in the context of process optimization. Meyers *et al.* ([43, 44, 45]) have derived the posterior probability of models for a class of priors for single response systems. They use a prior on the models based on the prior probability of a control factor being active. They use the calculated model posteriors for "factor-screening" in DOE. Chipman [13] presents a modified approach to this, where the prior probability of the model is chosen based on priors on the terms in the model rather than on the factors. Kass and Raftery [36] provide the Bayes factor interpretation of Bayesian model averaging, where the Bayes factor is the ratio of the posterior odds of the alternative hypothesis to the prior odds. In case of model averaging, each model that is considered is an alternative hypothesis to some base model which is the null hypothesis. Choosing the base model is not important as it is assumed that the posteriors of all considered models sum to 1. Berger and Pericchi ([4, 5]) compare different approaches for model selection. However, this comparison is made with respect to choosing

the best model, rather than to average across all models. The methods they consider include the prior distribution approach that is used here, the Bayes Information Criterion (BIC), the Intrinsic Bayes factor, and the Fractional Bayes factor. They conclude that the prior distribution approach is the most valuable when the sample size is small, and in this case all the other approaches become suspect. It is important to note here that the idea of model averaging for process optimization purposes is critical for experiments with small sample sizes, as the degree of uncertainty in the model in case of smaller samples is higher.

Hoeting *at al.* ([30, 31]) present a BIC approach for model averaging for calculating the predictive density for the purpose of "forecasting". Though the idea is similar to the posterior predictive density, there are serious drawbacks when considering small sample sizes because the BIC criteria is an asymptotic criteria and not as reliable for small samples. Draper [18] recommends to first find a good model and then average over a class of models expanded around this model. This is particularly useful when the number of models to be averaged over is very large.

Press [58] shows that the posterior predictive density for a diffuse prior in the parameters is a t-distribution in the case of single response, and a multivariate $\boldsymbol{t}$-distribution in the case of multiple responses under the SMR model. The diffuse prior is, however, not suitable for model-averaging. The reason for this and the form of some possible priors are discussed in section 3.4. Fernandez *et al.* [21] use Bayes factors to evaluate model posteriors using the similar priors proposed here. The authors recommend choosing the priors of the parameters based on their predictive ability in single response systems obtained using simulation. They apply the results to "forecasting" problems. Apley and Kim [2] present a Bayesian approach to process control by minimizing the mean square error for a single response system using

the posterior mean and covariance of the parameters, instead of the posterior predictive density of the response. They note that this can be extended to averaging over models, but do not give any results, and do not use a predictive approach.

2.4 Review of Tolerancing Methods

Tolerancing is an important step in the designing and manufacturing of a product. In practice, there are variations associated with products and processes at every step in a manufacturing plant. In addition, there is also variation associated with measurements. Therefore, from a cost and quality perspective it is important to quantify the variations at different stages of a manufacturing process. There are different problems related to tolerancing that have been studied in the literature [32]. In most of the cases, tolerancing is not considered in conjunction with a regression model to describe the relationship between the factors and the responses. Most of the literature in tolerancing comes from mechanical tolerancing where tolerances are typically expressed using conventional plus/minus tolerances or using geometric dimensioning. In this context, some of the typical problems addressed are:

1. **Tolerance Analysis**: Tolerance analysis involves identifying the variation in the response by taking into account individual tolerances on the factors. This is done at the design stage in manufacturing to ensure that the tolerance on the response meets the requirements. A review of statistical approaches to tolerance analysis is presented in [51].

2. **Tolerance Synthesis**: As opposed to tolerance analysis, here, the tolerance on the response is first quantified and this tolerance is used to arrive at the tolerance limits for the factors.

3. **Tolerance Transfer**: Tolerance transfer methods are used to connect the design stage with the manufacturing stage by a suitable tolerance scheme. Here, tolerance charting is a common method used [40].

In the standard approaches to statistical tolerancing [26], if Y is a quality characteristic with probability distribution P_Y^θ, where θ are the parameters, and if a sample $(y_1...y_n)$ of n independent observations is available, two common statistical methods have been used for constructing a tolerance region A. These are defined as,

1. the α-expectation tolerance region, given by:

$$E[C(A)] = \alpha, \tag{2.14}$$

2. the α-content tolerance region at confidence level γ, given by:

$$p[C(A) \geq \alpha] = \gamma, \tag{2.15}$$

where $C(A)$ is the coverage of the region A. These definitions are applicable for both classical or frequentist and Bayesian approaches. Both approaches are discussed in Guttman [25, 26]. However, they do not consider the problem of tolerance control in conjunction with regression, where the response depends on the settings of control factors. Also, the methods in the literature do not address the problem of finding the smallest tolerance region that satisfies one of the two criteria shown in equations (2.14) and (2.15).

The idea of a robust tolerance design was originally proposed by Taguchi [67]. Taguchi recommended tolerance design as the stage in quality control that follows parameter design. Parameter design is used to fit regression models to data and identify levels of the controllable factors that give the required mean and variation of the fitted response models. This methodology is also called "robust engineering" as it aims to provide the desired quality that is robust to the presence of noise factors. Taguchi's robust tolerance design is used to adjust the tolerances of the controllable factors that have a large influence on the response(s). Taguchi's idea is related to what is called "transmission of errors", where variation in the controllable factors causes additional variation in the responses. In some sense, the idea of setting tolerances on control factors is paradoxical to their very definition by which it is assumed that control factors can be set to desired values by the manufacturer. But in practice, it is arguable if control factors in the strictest sense of the definition exist. Thus it is also important to consider the transmission of errors due to inherent variation in the setting of the control factors while setting tolerances on the responses.

There have been other contributions in the tolerancing literature based on Bayesian methods. Singpurwalla [64] provides a Bayesian framework to approaching Taguchi's idea of parameter and tolerance design. Wolfinger [71] uses a Bayesian simulation based approach in order to set tolerance intervals for variance component models. Hamada [28] uses a Bayesian method to set tolerance interval control limits for control charts used in process monitoring.

2.5 Review of Multiple Criteria Decision Making

MCDM problems in the presence of uncertainty are traditionally solved using stochastic programming or statistical methods. The statistical approach to MCDM differs from stochastic programming methods in that the former uses information or data collected about the system in the analysis of solutions while the latter uses only a *priori* information about the random variables in the analysis (see [70]). In this research we focus on statistical methods using a Bayesian method of inference. Other non-Bayesian statistical MCDM approaches that use data collected on the system, typically fit expected value models that are in turn used to obtain the efficient frontier or Pareto set [65]. An example of the Pareto set is given in figure 2.3. In the figure there are two responses f_1 and f_2 that are to be minimized. The feasible range shown in the figure is plotted in the space of the responses and is the set of all possible values of the responses that can be obtained based on their expected values. In the figure, the dark line in the feasible shows the pareto set or the efficient frontier. Note that points A and B shown in the figure lie in the efficient frontier. The efficient frontier is thus made up of all such points in the feasible space where it is not possible to minimize one of the functions f_1 or f_2 any further without compromising on the other function. As opposed to these points point C shown in the figure is not on the efficient frontier as it is possible to move to another point in the feasible space while at the same time minimizing both f_1 and f_2. Once the efficient frontier is obtained using expected value models, there are different methods suggested in the literature that can be used to identify the operating point on the efficient frontier. Some of the methods used are:

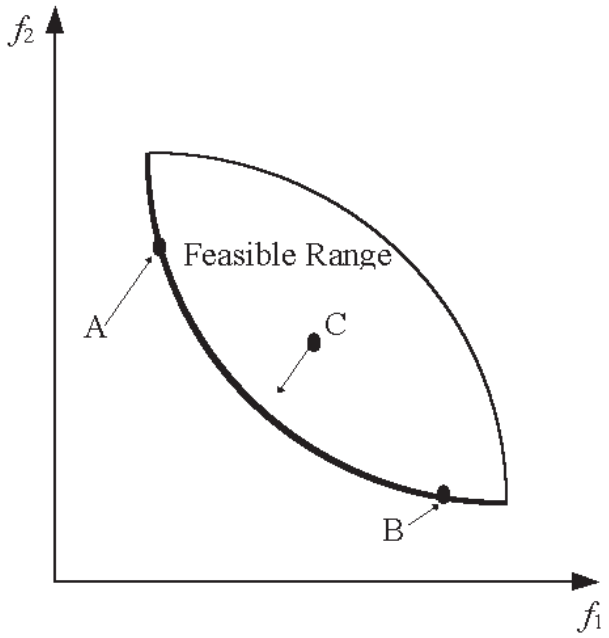

Figure 2.3: Traditional MCDM methods using the Pareto Set

1. Weighted Objective Function: Here, the decision maker (DM) sets weight w_i on criterion f_i for all the K criteria, such that $\sum w_i = 1$. The weighted sum $\sum w_i f_i$ is then used to form a single objective optimization problem.

2. Method of Global Criterion [72, 73]: Here, no input from the DM is taken. The optimization problem is formulated to minimize the deviation from the ideal solution, where the ideal solution f_i^* for criteria f_i is the optimum for f_i ignoring all other criteria. The objective function here for K criteria is given by

$$Z = \sum_{i=1}^{K} \left(\frac{f_i^* - f_i}{f_i^*} \right)^p,$$

where a value of $p = 1$ or $p = 2$ is commonly used.

3. Compromise Programming [72, 73]: Compromise programming is similar to the method of global criterion, except that here L_p metrics are used to measure the

deviation. The objective function to be minimized is given by

$$L_p = \left[\sum_{i=1}^{K} w_i^p (f_i^* - f_i)^p\right]^{1/p}.$$

Typically, values of $p = 1$ (rectilinear distance), $p = 2$ (Euclidean distance) or $p = \infty$ (Tchebychev norm) are used. The w_i are weights for each criterion that are assigned by the decision maker.

4. Interactive Methods: These are methods where a single, a pair or a cluster of solution points from the efficient frontier is/are shown to the DM to get a utility measure. Based on the DM's answer, a new solution or a new pair or cluster of solutions is/are shown to the DM. This is repeated until the DM is "satisfied" with the solution. There are different interactive methods in the literature based on this logic [63].

More recently, there have been developments in the application of Bayesian methods to certain MCDM problems, though these do not fall in the category of the MCDM problems that are discussed in this research. One such example is by Hahn [27] who uses Bayesian inferencing to derive priorities in Saaty's Analytic Hierarchy Process (AHP) [62].

Chapter 3

Model-Robust Process Optimization using Bayesian Model Averaging

3.1 Introduction: Process Optimization

In the "end game" of Response Surface Methodology (RSM, see [6, 49]), optimization of a process traditionally consists of two steps as shown in figure 2.1. The first step is to design the experiment, collect data and fit a polynomial model, usually of second order or higher to allow for curvature. Once the model is fitted, the next step is to optimize the response based on the fitted model and obtain estimated optimal operating settings. The second step in this process strongly depends on the assumption that the fitted model is the correct representation of this process. It is possible that a second different model, which arguably fits the data as well as the first model, provides considerably different optimal operating conditions (see the example section in this chapter for cases when this occurs). A frequentist approach, common in RSM practice, is to assess the effect of the

uncertainty of the parameter estimates on the optimum by computing a confidence region on the location of the optimum (see [55]). A more recent Bayesian approach implicitly considers the uncertainty of the parameters given the model form [56, 46]. The technique used by these authors involves obtaining the posterior predictive density of the response based on the assumed model, and maximizing the probability of obtaining the predicted response to lie within certain limits or specifications.

In this chapter, the Bayesian predictive approach is taken one step further by averaging over possible competing models. Here, no single model is assumed. Instead, as a first step, the Bayesian posterior probabilities for all possible models (belonging to a class, or classes, of models that are appropriate for the process) are calculated. Once the model posteriors are determined, the next step is to determine the posterior predictive density of the response for each of the competing models. The model-averaged posterior predictive density (MAP) is then computed by taking the weighted average of the densities over all competing models. The model posteriors computed earlier are used as the weights. The MAP is then used to maximize the probability of obtaining a response value within the given specification limits. The proposed method is illustrated by means of a block diagram in figure 3.1.

As the uncertainty in the model is more acute in cases where there are fewer runs, the examples provided in the later sections will focus on smaller designs. However, the main idea can be applied to any design where the form of the best model is in question.

In the next section, the technical details about the application of Bayesian model averaging to process optimization are discussed. The predictive approach we adopt focuses on making inferences on future values of the "observable" y [23]. For doing this, the posterior

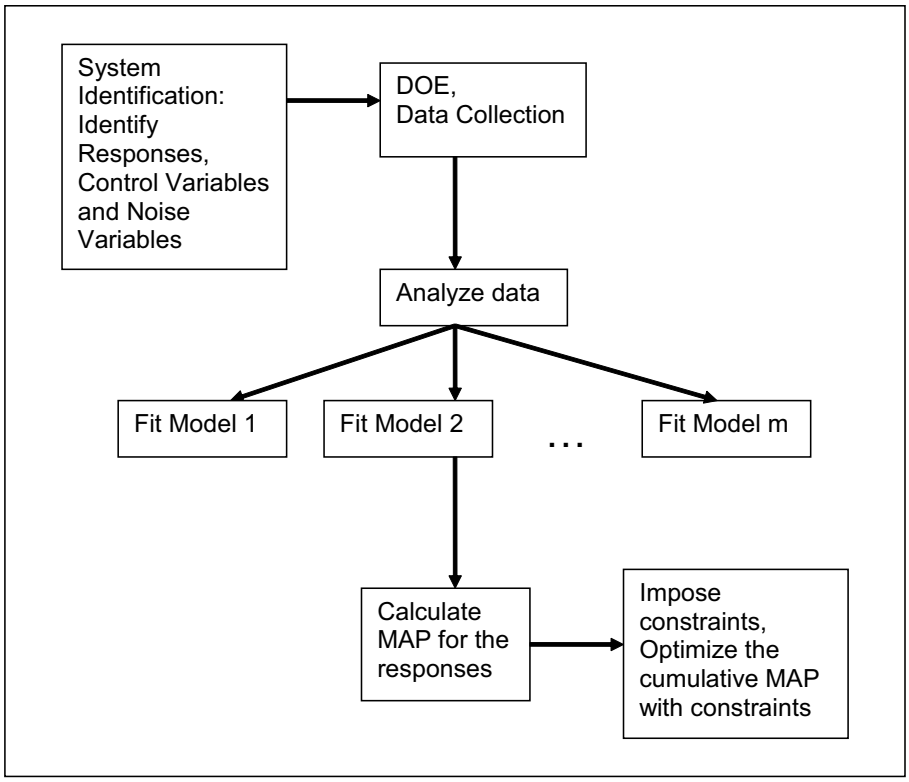

Figure 3.1: Proposed Method for Process Optimization

predictive density of the response under a particular choice of priors and the assumed likelihood needs to be derived. This is discussed in section 3.2.2 and the details are shown in appendix B. This is followed by two examples, one of which is a mixture experiment and the other a small composite design.

3.2 Bayesian Model Averaging

Consider a process with a single response variable y which is dependent on a $(p \times 1)$ vector of regressors $\mathbf{x}$ that are in turn functions of k controllable factors. It is assumed that a suitable experiment with n runs has been designed and carried out and the data from the experiment is available. The vector of responses from the experiment is given by the $(n \times 1)$ vector $\mathbf{y}$. Each observation of the model is assumed to be generated from a model linear in the parameters of the form

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \tag{3.1}$$

where ϵ is the error term, and $\boldsymbol{\beta}$ is the vector of process parameters (i.e., $\mathbf{x}$ is in model form). The particular functional form of the $\mathbf{x}'\boldsymbol{\beta}$ term (a function of the k controllable factors) is not known with certainty. In this chapter, we focus on the case where ϵ is normally distributed.

It is assumed that the goal of the optimization is to identify the values of the controllable factors that result in a response y such that $L < y < U$ where L denotes a given lower bound (or specification) and U denotes a given upper bound. The approach adopted here maximizes the posterior predictive probability of obtaining the response y within these bounds. In this procedure, the potential models that are under consideration based on

a family or families of models are listed first. Next, the posterior probability of each of these models given the experimental data $P(M_i|\mathbf{y})$ is calculated. The posterior predictive density of the response is then calculated for each model M_i as a function of the controllable variables. This is denoted by $P(y^*|M_i, \mathbf{x}^*, \mathbf{y})$, where y^* is the posterior value of the response at a new set of observed regressors $\mathbf{x}^*$. In order to average the predictive density of the response over all competing models, the weighted average of $P(y^*|M_i, \mathbf{x}^*, \mathbf{y})$ is taken over all i, using the model posteriors $P(M_i|\mathbf{y})$ as the weights. The model-averaged posterior predictive density (MAP) is thus of the form of a mixture of distributions over all competing models, namely:

$$MAP = P(y^*|\mathbf{x}^*, \mathbf{y}) = \sum_i P(y^*|\mathbf{x}^*, \mathbf{y}, M_i)P(M_i|\mathbf{y}). \tag{3.2}$$

The optimal control variables are then determined by maximizing the probability that the predicted response lies within the target bounds, i.e.,

$$\max_{x_1^*,...,x_k^*} P(L \leq Y^* \leq U) = \sum_i \left[\int_L^U P(y^*|M_i, \mathbf{x}^*, \mathbf{y})dy^*\right] P(M_i|\mathbf{y}), \tag{3.3}$$

where the maximization is over the k control factors $(x_1, x_2, ..., x_k)$ that $\mathbf{x}^*$ depends on. It should be pointed out that this approach does not average the optimal levels of the controllable factors for each model. Instead, the optimal levels of the controllable factors are prescribed by averaging the predictive density of the response over all models. Constraints on the controllable factors x_i can be included in (3.3) if desired.

3.2.1 Calculating model posteriors

There is considerable literature on the calculation of model posterior probabilities (see [43], [58] and the references therein). The most common approach is to assume a candidate

list of models based on a class (or classes) of models with f_i out of the total k factors present in model M_i. A useful method to determine model priors, proposed by Meyer et al. [43, 44, 45] is to choose the model priors based on the active factors (i.e., the factors present in each model). Denote the probability of factor j to be active as π_j, $j \in \{1, ..., k\}$. Assuming that the prior probabilities of active factors are independent, the model prior is given by

$$P(M_i) = \prod_{j \in M_i} (\pi_j) \prod_{j' \notin M_i} (1 - \pi_{j'}). \tag{3.4}$$

Note that this is an improper prior, that is the sum of the probabilities of the models a *priori* need not add up to 1. If $\pi_j = \pi$, $\forall j \in \{1, ..., k\}$, then $P(M_i) = \pi^{f_i}(1-\pi)^{k-f_i}$. Let r_i be the number of terms in model M_i and t_i be the number of terms in model M_i excluding the constant term. Thus, if the model includes a constant term, we have that $t_i = r_i - 1$, otherwise $t_i = r_i$. Let $\mathbf{X}_i$ be the $(n \times r_i)$ design matrix corresponding to M_i. The posterior probability of M_i is given by Bayes' theorem,

$$P(M_i|\mathbf{y}) = \frac{P(\mathbf{y}|M_i)P(M_i)}{\sum_i P(\mathbf{y}|M_i)P(M_i)}, \tag{3.5}$$

where $P(\mathbf{y}|M_i)$ is the marginal likelihood of the model given the data. This marginal is defined as:

$$P(\mathbf{y}|M_i) = \int_{\sigma^2} \int_{\boldsymbol{\beta}_i} P(\mathbf{y}|M_i, \sigma^2, \boldsymbol{\beta}_i) P(\sigma^2, \boldsymbol{\beta}_i|M_i)\, d\boldsymbol{\beta}_i\, d\sigma^2, \tag{3.6}$$

where $P(\mathbf{y}|M_i, \sigma^2, \boldsymbol{\beta}_i)$ is the likelihood function. Under the assumption of normality of the error terms, the likelihood is given by:

$$P(\mathbf{y}|M_i, \sigma^2, \boldsymbol{\beta}_i) \propto \sigma^{-n} exp\left[\frac{-1}{2\sigma^2}(\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta}_i)'(\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta}_i)\right]. \tag{3.7}$$

$P(\sigma^2, \boldsymbol{\beta}_i|M_i)$ is the joint prior of the model parameters for model M_i. The parameters $\boldsymbol{\beta}_i$ and σ^2 are assumed to be independent a priori. The priors on the parameters are chosen

as

$$\boldsymbol{\beta}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i \sigma^2), \tag{3.8}$$

$$P(\sigma^2) \propto \frac{1}{\sigma^2}, \tag{3.9}$$

and

$$P(\boldsymbol{\beta}_i, \sigma^2) = P(\sigma^2) P(\boldsymbol{\beta}_i). \tag{3.10}$$

Here, it is assumed that $\boldsymbol{\Sigma}^{-1} = (\mathbf{X}_i' \mathbf{X}_i) \mathbf{V_i}$, where $\mathbf{V_i} = \frac{1}{g} \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{t_i} \end{pmatrix}$, and g is a parameter whose value is to be chosen. Thus, the priors on all the β_i's except for the constant term are assumed to be normally distributed using Zellner's g-prior [74]. The priors on the β_i for the constant term and on $log(\sigma^2)$ are assumed to be non-informative. A discussion on the choice of priors is included in section 3.4.

From the assumed priors and from equation (3.7), the integral in equation (3.6) can be computed and yields (see [45], [58]):

$$P(\mathbf{y}|M_i) \propto \gamma^{-t_i} |\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i' \mathbf{X}_i|^{-\frac{1}{2}} S_i^{-\frac{(n-1)}{2}}, \tag{3.11}$$

where γ is such that

$$\frac{g}{\gamma^2} \mathbf{V}_i = \boldsymbol{\Sigma}_i^{-1}. \tag{3.12}$$

Then, by omitting the constant denominator in equation (3.5),

$$P(M_i|\mathbf{y}) \propto \pi^{f_i} (1-\pi)^{k-f_i} \gamma^{-t_i} |\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i' \mathbf{X}_i|^{-\frac{1}{2}} S_i^{-\frac{(n-1)}{2}}, \tag{3.13}$$

where

$$S_i = (\mathbf{y} - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i)'(\mathbf{y} - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i{}' \boldsymbol{\Sigma}_i^{-1} \hat{\boldsymbol{\beta}}_i \tag{3.14}$$

$$= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}_i (\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{y}, \tag{3.15}$$

and

$$\hat{\boldsymbol{\beta}}_i = (\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}. \tag{3.16}$$

Here, S_i is the Bayesian analog to the residual sum of squares and $\hat{\boldsymbol{\beta}}_i$ gives the parameter estimates for model M_i. The probabilities for each of the models computed from the above equations are scaled by dividing each one of them by the sum of all the probabilities in order to obtain the model posterior probabilities for the models that sum to 1.

It should be noted that not all models in the original candidate list will have significant posterior probabilities. Hence, it is recommended to choose a subset of m models from the original list based on the calculated posteriors. Methods for choosing this subset are discussed in appendix A.

3.2.2 Calculating the predictive density

The predictive density for the new response y^* at a new set of regressors $\mathbf{x}^*$ for a given model M_i is given by:

$$P(y^*|M_i, \mathbf{x}^*, \mathbf{y}) = \int_{\sigma^2}\int_{\boldsymbol{\beta}_i} P(y^*|M_i, \mathbf{x}^*, \mathbf{y}, \sigma^2, \boldsymbol{\beta}_i)P(\boldsymbol{\beta}_i, \sigma^2|\mathbf{y}, M_i)\, d\boldsymbol{\beta}_i\, d\sigma^2, \tag{3.17}$$

where $P(y^*|M_i, \mathbf{x}^*, \mathbf{y}, \sigma^2, \boldsymbol{\beta}_i)$ is the likelihood function, and $P(\boldsymbol{\beta}_i, \sigma^2|\mathbf{y}, M_i)$ is the joint posterior of the model parameters [57]. Based on the observed data, the probability that the predicted response lies between the lower and upper bounds for a given model at a given set of regressors $\mathbf{x}^*$ is obtained using the cumulative posterior predictive density, that is:

$$P(L \leq Y^* \leq U|M_i, \mathbf{x}^*, \mathbf{y}) = \int_L^U P(y^*|M_i, \mathbf{x}^*, \mathbf{y})dy^*. \tag{3.18}$$

If the model parameters $log(\sigma^2)$ and $\boldsymbol{\beta}_i$ are assumed to have non-informative priors, then it has been shown by Press [58] that the predictive density obeys a t_{n-r_i} distribution. For the priors assumed here, the predictive density is shown in appendix B to follow a t_{n-1} distribution. The cumulative posterior predictive density can thus be obtained from the c.d.f. of a t-distribution that is very easy to compute using the incomplete beta function. This avoids the use of any numerical methods for the integration in equation (3.18). The cumulative predictive density is computed by:

$$P\left(\frac{y^* - \mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}_i}{\hat{\sigma}_i\sqrt{1+\mathbf{x}^{*\prime}(\boldsymbol{\Sigma}_i^{-1}+\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^*}} < t|M_i,\mathbf{x}^*,\mathbf{Y}\right) = \frac{1}{2}\left[1+I_{\frac{t^2}{\nu+t^2}}\left(\frac{1}{2},\frac{\nu}{2}\right)\right], \qquad (3.19)$$

where $I_z(a,b)$ is the incomplete beta function, $\nu = n-1$, and $\hat{\sigma}_i^2 = S_i/(n-1)$. The objective function in equation (3.3) is thus the cumulative model averaged posterior predictive density, and can be calculated at a given observation $\mathbf{x}^*$ using equations (3.13) and (3.19).

3.3 Examples

There are two hyper-parameters to be chosen in the priors, namely π and g. In both of the examples below, we choose a value of $\pi = 0.5$, which implies equal prior chances of a factor being active or inactive. The parameter g is chosen based on the value of γ that gives the lowest posterior probability for the null model (model with just the constant term). This choice is suggested by Meyer *at al.* [43], and is discussed further in section 3.4. A study of the sensitivity of the solution to the choice of priors is also done by solving the optimization problem for various values of these parameters.

The optimizations were carried out using MATLAB's *fmincon* routine. This function uses a sequential quadratic programming method. This is used to maximize the cumulative

MAP between the upper and lower bounds, over all feasible values of $x_1, ..., x_k$. As with most nonlinear programming algorithms, this method requires an initial starting point $x_1, ..., x_k$. In order to avoid local optimums, different random starting values arranged in a latin hypercube were utilized (see [59]) to better cover the feasible region. In the two examples that follow, convergence to the same point was always achieved, so the optimality of the solutions obtained seems to be well established.

3.3.1 Example 1: Mixture Experiment

This example, taken from Frisbee et al. [22], shows a mixture experiment where the response is glass transition temperature of films cast from poly(DL-lactide) (PLA), and the controllable variables are amounts of non-ionic surfactants, namely, Polaxamer 188 NF (Pluronic® F68), Ployoxyethylene 40 monostearate (Myrj® 52-S) and Polyoxyethylene sorbitan fatty acid ester NF (Tween® 60). The authors are interested in finding the composition of the controllable factors that minimize the glass transition temperature. The data from [22] is given in Table 3.1. The authors fit a regression equation that is given by:

$$y = 18.50x_1 + 13.88x_2 + 36.06x_3 - 35.21x_1x_3 + 19.55x_2x_3. \tag{3.20}$$

Based on the fitted equation, Frisbee *et al.* [22] use contour plots to determine the minimal plateau region for glass transition temperature. However, as the experiment consisted of only 11 runs, the accuracy of the model used is suspect. There are a some other regression models that provide a reasonable fit to the data, and each of these would result in a different optimal solution. Cahya [10] suggested that a different class of models, namely a Becker

x_1	x_2	x_3	y
1.000	0.000	0.000	18.90
0.000	1.000	0.000	15.20
0.000	0.000	1.000	35.00
0.500	0.500	0.000	16.10
0.500	0.000	0.500	18.90
0.000	0.500	0.500	31.20
0.333	0.333	0.333	19.30
0.666	0.167	0.167	18.20
0.167	0.666	0.167	17.70
0.167	0.167	0.666	30.10
0.333	0.333	0.333	19.00

Table 3.1: Mixture data from [22] where the response is Glass Transition temperature

model [3], also adequately represents this process. The Becker model is of the form

$$y = b_1x_1 + b_2x_2 + b_3x_3 + b_4\, min(x_1, x_2) + b_5\, min(x_1, x_3) + b_6\, min(x_2, x_3). \qquad (3.21)$$

The ordinary least square (OLS) regression statistics for the models in equation (3.20) and (3.21), as well as for all other models belonging to these two classes of models are shown in Table 3.2. Higher order terms in each model were considered only if the corresponding lower order terms were present. In the table, each row represents a competing model and under the columns containing the model terms (effects), a '1' indicates that the term is present in the model and a '0' indicates otherwise. The OLS statistics shown in the table are based on the sum of squares of the residuals (SSE), the total sum of squares (SST), and the standard error ($S.E.$). It is noted that since the mixture models are fitted without the constant term, the SSE/SST ratio is greater than 1 for some models. This means that the model $y = \bar{y}$, where $\bar{y}$ is the mean of the observed responses, fits the data better than the models for which the SSE/SST ratio is greater than 1. Even based on the OLS statistics, there are many possible models that can be used to represent the process.

$Model\ No.$	const	A	B	C	AB	AC	BC	$min(A,B)$	$min(A,C)$	$min(B,C)$	$\frac{SSE}{SST}$	$\frac{SSE/n-r_i}{SST/n-1}$	$S.E.$	$P(M_i\|data)$
1	0	1	1	1	0	0	0	0	1	1	0.0223	0.0372	1.32	0.3557
2	0	1	1	1	0	0	0	0	1	0	0.0696	0.0994	2.1572	0.143
3	0	1	1	1	0	1	1	0	0	0	0.0478	0.0797	1.9323	0.138
4	0	1	1	1	0	0	0	1	1	1	0.0182	0.0365	1.3071	0.091
5	0	1	1	1	0	1	0	0	0	0	0.091	0.13	2.4675	0.0803
6	0	1	1	1	1	1	1	0	0	0	0.0384	0.0769	1.8972	0.0759
7	0	1	1	1	1	1	0	0	0	0	0.083	0.1383	2.5451	0.0392
8	0	1	1	1	0	0	0	1	1	0	0.0692	0.1153	2.3232	0.0279
9	0	1	1	1	0	0	0	0	0	0	0.2252	0.2815	3.6304	0.0146
10	0	1	1	1	0	0	1	0	0	0	0.1875	0.2678	3.5413	0.0119
11	0	1	1	1	1	0	0	0	0	0	0.2146	0.3066	3.7886	0.0068
12	0	1	1	1	1	0	1	0	0	0	0.1753	0.2922	3.6991	0.0058
13	0	1	1	1	0	0	0	0	0	1	0.212	0.3028	3.7654	0.0037
14	0	1	1	1	0	0	0	1	0	0	0.2119	0.3027	3.7646	0.0034
15	1	0	0	0	0	0	0	0	0	0	1	1	6.8426	0.0015
16	0	1	1	1	0	0	0	1	0	1	0.1901	0.3169	3.8519	0.0011
17	0	0	1	1	0	0	0	0	0	0	1.0888	1.2098	7.5262	0.0001
18	0	1	0	1	0	0	0	0	0	0	1.2697	1.4107	8.1273	0.0001
19	0	0	1	1	0	0	1	0	0	0	1.0671	1.3338	7.9026	0.0001
20	0	1	0	1	0	1	0	0	0	0	1.0958	1.3697	8.0083	0
21	0	0	1	1	0	0	0	0	0	1	1.044	1.305	7.8169	0
22	0	0	0	1	0	0	0	0	0	0	2.7513	2.7513	11.35	0
23	0	1	0	1	0	0	0	0	1	0	1.1881	1.4852	8.339	0
24	0	0	1	0	0	0	0	0	0	0	7.2363	7.2363	18.407	0
25	0	1	0	0	0	0	0	0	0	0	7.5341	7.5341	18.7818	0
26	0	1	1	0	0	0	0	0	0	0	4.9586	5.5096	16.0613	0
27	0	1	1	0	1	0	0	0	0	0	4.9138	6.1422	16.9584	0
28	0	1	1	0	0	0	0	1	0	0	4.9459	6.1824	17.0138	0

Table 3.2: Least square regression statistics and posterior probabilities for competing models for example 1

In order to perform Bayesian model-robust optimization, the first step was to define the prior parameters required for model averaging. Using the method proposed by Meyer et al. [43], a value of $\gamma = 10$ was chosen. The parameter π was then set at 0.5 for all the factors. The sensitivity of the optimal solution with respect to the chosen parameters is discussed later. Model posteriors were calculated for all models discussed earlier. The resulting posterior probabilities are shown in the last column of Table 3.2 for each model.

Since there are no constant terms in the mixture models considered, Zellner's g-prior is used for all the $\boldsymbol{\beta}_i$, i.e., $\boldsymbol{\Sigma}_i^{-1} = (1/g)(\mathbf{X}_i'\mathbf{X}_i)$ in equation (3.8). In this case, the cumulative posterior predictive density is given by (see appendix B),

$$P\left(\frac{y^* - \mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}_i}{\hat{\sigma}_i\sqrt{1+\mathbf{x}^{*\prime}(\boldsymbol{\Sigma}_i^{-1}+\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^*}} < t|M_i, \mathbf{x}^*, \mathbf{Y}\right) = \frac{1}{2}\left[1 + I_{\frac{t^2}{\nu+t^2}}\left(\frac{1}{2}, \frac{\nu}{2}\right)\right], \qquad (3.22)$$

where $I_z(a,b)$ is the incomplete beta function, $\nu = n$, and $\hat{\sigma}_i^2 = S_i/n$.

Based on the model posteriors, only models with $P(M_i|data) > 0.0279$ were considered for averaging as they accounted for 95% of the probability. Table 3.3 shows these models. The model numbers correspond to the respective models in Table 3.2. As the objective is to minimize the response, the lower bound L was set at ∞ and the upper bound U was set at 18 for illustration purposes. The optimization of the MAP resulted in point$(0.133, 0.867, 0)$ as the optimum levels of the controllable factors, where the probability of obtaining $Y^* \in (-\infty, 18)$ was 0.9388. Table 3.3 also shows the optimal values of the controllable factors at which the posterior predictive densities of each individual model is maximized for $Y^* \in (-\infty, 18)$. The maximum value of the posterior predictive density is given in the column labeled 'z^*'. It can be seen that the optimal solution can vary drastically based on the model chosen.

$ModelNo.$	$P(M_i\|data)$	x_1^*	x_2^*	x_3^*	z^*
1	0.3557	0	1	0	0.9998
2	0.1430	0	1	0	0.9969
3	0.1380	0	1	0	0.8933
4	0.0910	0.2756	0.7244	0	0.9998
5	0.0803	0	1	0	0.7358
6	0.0759	0.2249	0.7751	0	0.8993
7	0.0392	0.3243	0.6757	0	0.7695
8	0.0279	0.2469	0.7531	0	0.9960

Table 3.3: Optimum for individual models for example 1

Figure 3.2 shows the cumulative MAP plotted on a 2-D simplex as well as a 3-D plot. The 2-D plot shows the points at which $P(-\infty < Y^* < 18|\mathbf{y}, x_1^*...x_k^*)$ was evaluated, with the squares representing points where $P(-\infty < Y^* < 18\mathbf{y}, x_1^*...x_k^*) > 0.7$. The 3-D plot shows the same points with cumulative MAP plotted on the vertical axis. Figure 3.3 shows $P(-\infty < Y^* < 18|M_i, \mathbf{y}, x_1^*...x_k^*)$, plotted at the same points for the eight competing models with the squares representing points where the individual predictive density is greater than 0.7. In order to better understand the importance of maximizing the MAP, table 3.4 shows the probabilities of conformance for various cases of the true model and the assumed model. The table shows the value of $P(L \leq Y^* \leq U|M_i, \mathbf{y}, x_1^*, ..., x_k^*)$ where M_i is the true model and control factors $x_1^*, ..., x_k^*$ are set at their optimal values obtained from solving from maximizing this probability using the *assumed* model. Thus, for example, if the

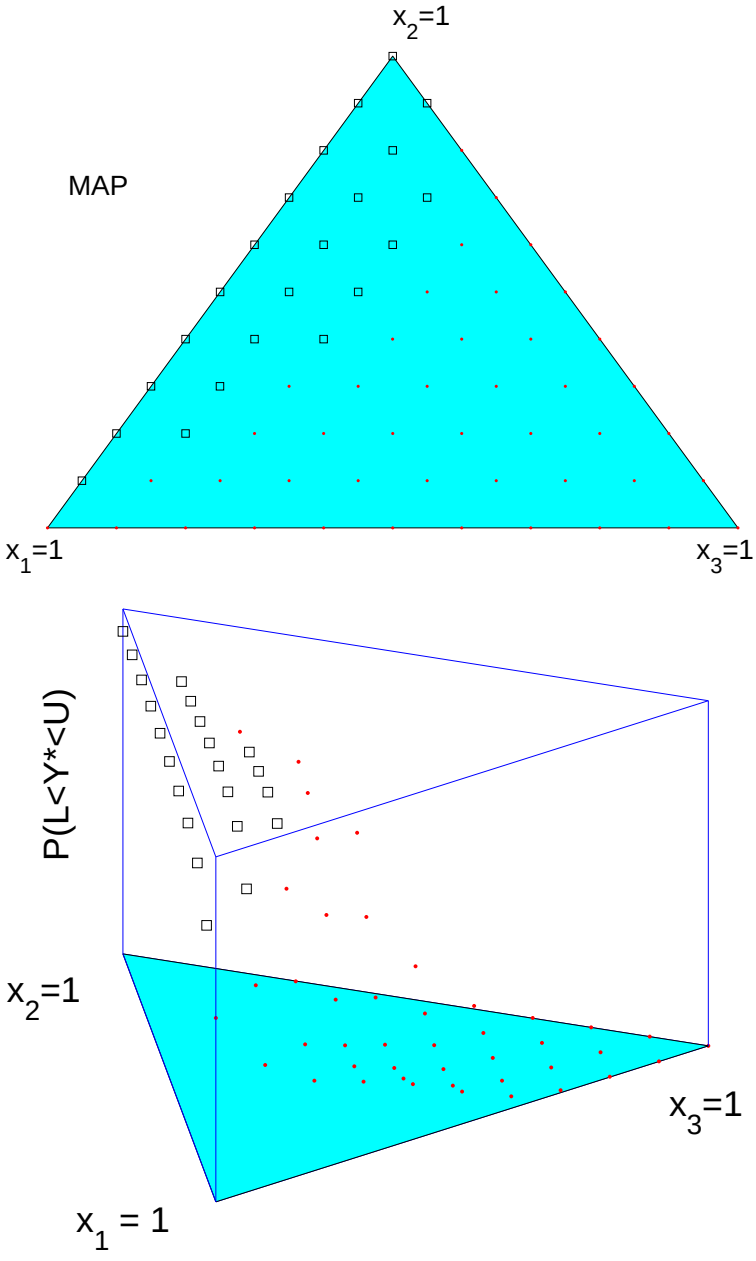

Figure 3.2: Simplex and 3-D plot of cumulative model-averaged posterior predictive density for example 1

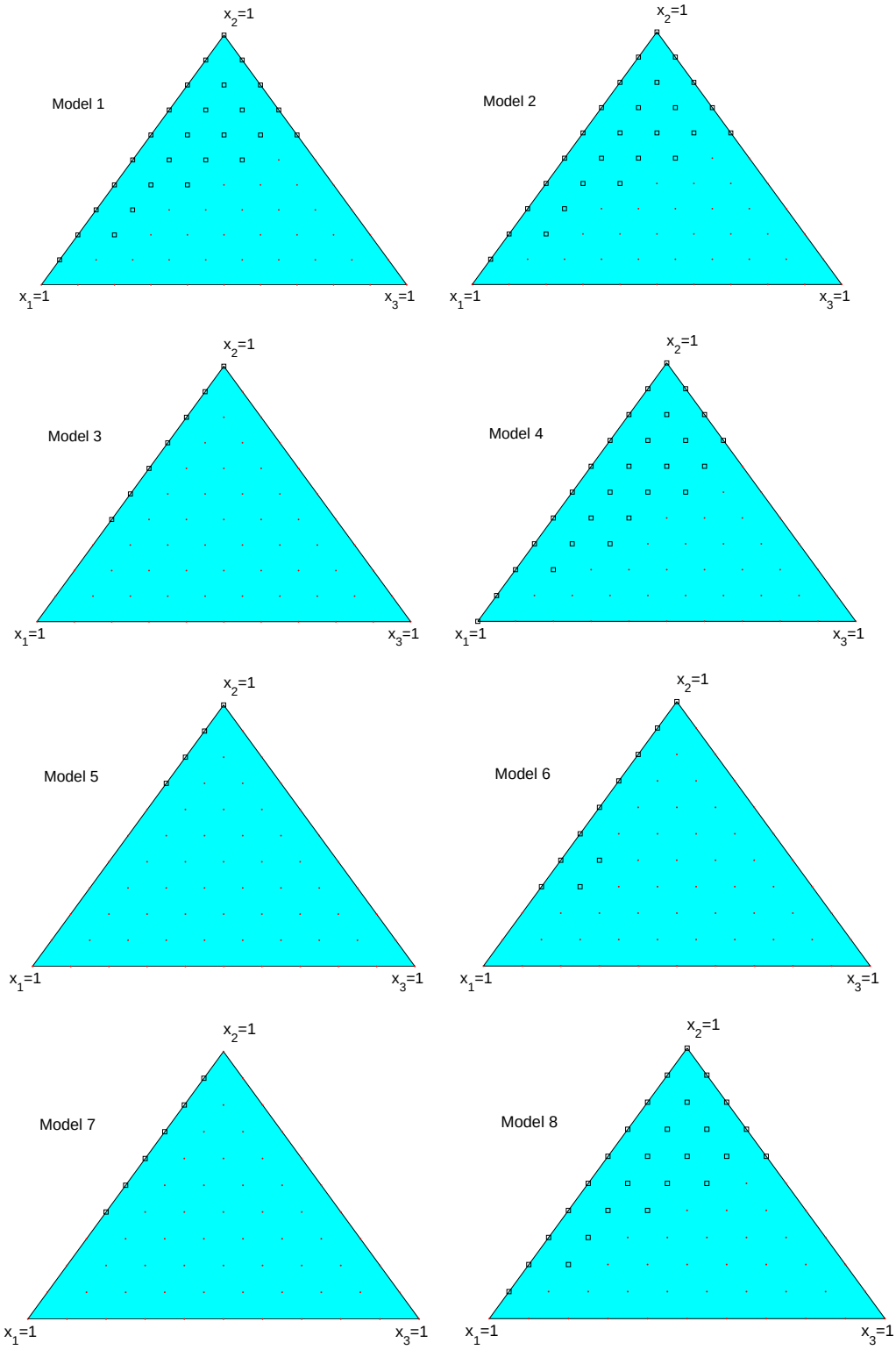

Figure 3.3: Simplex plot of cumulative posterior predictive densities for individual models for example 1

assumed model is model 1, then the probability of conformance is maximized at the point $(0, 1, 0)$, as shown in table 3.3, yielding a probability of 0.9998. However, this is actually the probability of conformance only if the true model is also model 1. If, for example, it so happens that the true model is model 7, then the probability of having $Y^* \in (-\infty, 18)$ is actually 0.6584 when using the solution point $(0, 1, 0)$, obtained with the wrong model. Similarly, the last column on the table shows $P(L \leq Y^* \leq U|M_i, \mathbf{y}, x_1^*, ..., x_k^*)$ for the true model, evaluated at the solution $x_1^*, ..., x_k^*$ obtained from maximizing the MAP. Based on the column statistics, it can be seen that operating at the point which maximizes the MAP has highest average probability of conformance (and among lowest std. deviation of this probabilities) compared to probabilities provided by solutions obtained by assuming single one of the competing models. The MAP also has higher minimum probability of conformance, thus it improves the worst-case scenario (worst true model). Therefore, it is seen that regardless the true process model (within the assumed family of models), the solution obtained using the model-average approach provides an operating point that gives relative high probabilities of conformance. It is in this sense that the solutions obtained are *robust* to the uncertainty in the form of the true model.

Assumed Model→	1	2	3	4	5	6	7	8	MAP
True Model↓									
1	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9997	0.9998	0.9998
2	0.9969	0.9969	0.9969	0.9962	0.9969	0.9965	0.9958	0.9964	0.9967
3	0.8933	0.8933	0.8933	0.8517	0.8933	0.8623	0.8399	0.8579	0.8778
4	0.9997	0.9997	0.9997	0.9998	0.9997	0.9998	0.9998	0.9998	0.9998
5	0.7358	0.7358	0.7358	0.7057	0.7358	0.7129	0.6980	0.7099	0.7240
6	0.8434	0.8434	0.8434	0.8976	0.8434	0.8993	0.8933	0.8990	0.8920
7	0.6584	0.6584	0.6584	0.7680	0.6584	0.7623	0.7695	0.7653	0.7373
8	0.9949	0.9949	0.9949	0.9960	0.9949	0.9960	0.9958	0.9960	0.9957
Min	0.6584	0.6584	0.6584	0.7057	0.6584	0.7129	0.6980	0.7099	0.7240
Max	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
Mean	0.8903	0.8903	0.8903	0.9019	0.8903	0.9036	0.8990	0.9030	0.9029
Std. Dev.	0.1342	0.1342	0.1342	0.1170	0.1342	0.1157	0.1194	0.1161	0.1173

Table 3.4: Model-Robustness analysis for example 1. Table gives the probability $P(L \leq Y^* \leq U|M_i, \mathbf{y}, x_1, ..., x_k)$ where M_i is the true model, evaluated at the settings $x_1, ..., x_k$ obtained from maximizing the probability of conformance using the assumed model.

Table 3.5 shows the sensitivity of the solution with respect to the chosen parameters γ and π. It is seen that the sensitivity of the solution to π is dependent on γ. At the value of γ chosen, the optimal controllable variables as well as the optimal predictive density are insensitive to the choice of π.

3.3.2 Example 2: Small-composite Design

The second example uses data from Czitrom and Spagon [14] for a chemical vapor deposition (CVD) process. The goal of the experiment was to investigate the Uniformity and Stress responses. This example illustrates the model-averaging approach on the first response. The central composite inscribed (CCI) design that was used and the experimental data are shown in Table 3.6. There are two controllable factors: Pressure and ratio of the gaseous reactants H_2 and WF_6 (denoted by H_2/WF_6). The goal was to minimize the response, as a smaller value of "Uniformity" indicates a more uniform layer being deposited on a wafer. The models considered included combinations of main effects, two-way interactions and quadratic effects. In all the models higher order effects were included only if the corresponding main effect(s) is(are) present in the model. Table 3.7 lists these models along with their least square regression statistics and posterior probabilities. The prior on the factors, π, was set at 0.5 and a value of $\gamma = 2$ was chosen using the method described in section 3.4.

Models with $P(M_i|data) > 0.0254$ were used for model averaging as they accounted for 95% of the probability. Based on these models and within the region $\{-1 \leq x_1 \leq 1, -1 \leq x_2 \leq 1\}$, the MAP was maximized for $Y^* \in (-\infty, 5)$ at the point $(1.0000, -0.9198)$ yielding

γ	π	x_1^*	x_2^*	x_3^*	z^*
0.5	0.25	0.6841	0.3159	0.0000	0.7213
0.5	0.50	0.6703	0.3297	0.0000	0.6993
0.5	0.75	0.7433	0.2567	0.0000	0.6891
2	0.25	0.6167	0.3833	0.0000	0.6906
2	0.50	0.5043	0.4957	0.0000	0.6759
2	0.75	0.5043	0.4957	0.0000	0.6759
5	0.25	0.3233	0.6767	0.0000	0.8252
5	0.50	0.3232	0.6768	0.0000	0.8252
5	0.75	0.3233	0.6767	0.0000	0.8252
10	0.25	0.1330	0.8670	0.0000	0.9388
10	0.50	0.1330	0.8670	0.0000	0.9388
10	0.75	0.1330	0.8670	0.0000	0.9388
30	0.25	0.0213	0.9787	0.0000	0.9856
30	0.50	0.0213	0.9787	0.0000	0.9856
30	0.75	0.0213	0.9787	0.0000	0.9856
100	0.25	0.0000	1.0000	0.0000	0.9767
100	0.50	0.0000	1.0000	0.0000	0.9825
100	0.75	0.0000	1.0000	0.0000	0.9825

Table 3.5: Sensitivity of solution with respect to the parameters γ and π for example 1

$Coded\ Pressure$	$Coded\ H_2/WF_6$	$Uniformity$
1	0	4.6
0	0	6.2
0.71	-0.71	3.4
-0.71	0.71	6.9
-1	0	7.3
0	0	6.4
-0.71	-0.71	8.6
0	-1	6.3
0.71	0.71	5.1
0	1	5.4
0	0	5

Table 3.6: Design and experimental data for CVD process [14]

$Model\ no.$	$constant$	A	B	AB	A^2	B^2	R^2	R^2_{Adj}	$S.E.$	$P(M_i\|data)$
1	1	1	1	1	0	0	0.8703	0.8148	0.6145	0.2827
2	1	1	0	0	0	0	0.7186	0.6874	0.7982	0.2396
3	1	1	1	1	1	0	0.8715	0.7858	0.6607	0.108
4	1	1	1	1	0	1	0.8703	0.7839	0.6637	0.1053
5	1	1	0	0	1	0	0.7198	0.6498	0.8449	0.0907
6	1	1	1	0	0	0	0.7285	0.6607	0.8316	0.0671
7	1	1	1	1	1	1	0.8716	0.7431	0.7235	0.0416
8	1	1	1	0	1	0	0.7297	0.6139	0.8871	0.0254
9	1	1	1	0	0	1	0.7285	0.6122	0.8891	0.025
10	1	1	1	0	1	1	0.7298	0.5496	0.9581	0.0098
11	1	0	0	0	0	0	0	0	1.4276	0.0035
12	1	0	1	0	0	0	0.0099	-0.1001	1.4974	0.0009
13	1	0	1	0	0	1	0.0099	-0.2376	1.5882	0.0003

Table 3.7: Least square regression statistics and posterior probabilities for competing models for example 2

Model no.	$P(M_i\|data)$	x_1^*	x_2^*	z^*
1	0.2827	1	-1	0.9665
2	0.2396	1	N/A	0.8132
3	0.1080	1	-1	0.9569
4	0.1053	1	-0.9017	0.9618
5	0.0907	1	N/A	0.7776
6	0.0671	1	1	0.8477
7	0.0416	1	-0.9018	0.9464
8	0.0254	1	1	0.8178

Table 3.8: Optimum for individual models for example 2

a maximum probability of conformance of 0.8851. The optimum values of the controllable factors obtained by maximizing the individual predictive densities, and the maximum value of the predictive density for the individual models for $Y^* \in (-\infty, 5)$ are given in table 3.8. It can be seen that for all the models the optimum value of x_1 is 1, but the optimum setting for x_2 can vary anywhere from -1 to 1. Figure 3.4 shows the surface plot of the cumulative posterior predictive density of the response $Y^* \in (-\infty, 5)$ for different possible values of the control factors. Model-robustness analysis for the competing models is given in Table 3.9. (Note that models 2 and 5 are independent of the second factor, x_2 (H_2/WF_6). In the table, for the columns associated with these two models, the probabilities of conformance were evaluated at the point $(1, 0)$). Similarly as in the previous example, it can be seen that the solution obtained by maximizing the MAP is robust to the uncertainty in the true

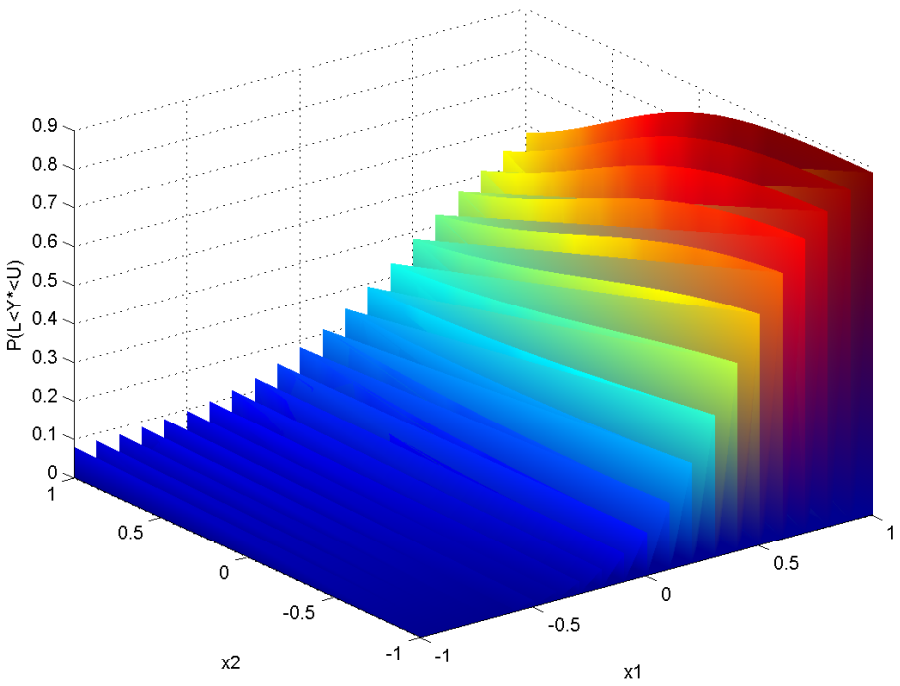

Figure 3.4: Surface plot of the $P(Y^* \in (-\infty, 5))$

model of the process. Table 3.10 shows results of sensitivity analysis to the solution with respect to the parameters π and γ. The sensitivity of the solution to π is dependent on the value of the γ chosen. Here, also, it is seen that at a given value of γ, the solution is insensitive to the selection of the π parameter.

Assumed Model→	1	2	3	4	5	6	7	8	MAP
True Model↓									
1.00	0.9665	0.8680	0.9665	0.9648	0.8680	0.3935	0.9648	0.3935	0.9651
2.00	0.8132	0.8132	0.8132	0.8132	0.8132	0.8132	0.8132	0.8132	0.8132
3.00	0.9569	0.8323	0.9569	0.9540	0.8323	0.3733	0.9540	0.3733	0.9546
4.00	0.9612	0.8614	0.9612	0.9618	0.8614	0.3991	0.9618	0.3991	0.9618
5.00	0.7776	0.7776	0.7776	0.7776	0.7776	0.7776	0.7776	0.7776	0.7776
6.00	0.7379	0.8165	0.7379	0.7472	0.8165	0.8477	0.7472	0.8477	0.7455
7.00	0.9456	0.8318	0.9456	0.9464	0.8318	0.3749	0.9464	0.3749	0.9463
8.00	0.7043	0.7809	0.7043	0.7130	0.7809	0.8178	0.7130	0.8178	0.7114
Min	0.7043	0.7776	0.7043	0.7130	0.7776	0.3733	0.7130	0.3733	0.7114
Max	0.9665	0.8680	0.9665	0.9648	0.8680	0.8477	0.9648	0.8477	0.9651
Mean	0.8579	0.8227	0.8579	0.8598	0.8227	0.5996	0.8598	0.5996	0.8594
Std. Dev.	0.1111	0.0330	0.1111	0.1075	0.0330	0.2302	0.1075	0.2302	0.1082

Table 3.9: Model-robustness analysis for example 2. Table gives the probability $P(L \leq Y^* \leq U|M_i, \mathbf{y}, x_1, ..., x_k)$ where M_i is the true model, evaluated at the settings $x_1, ..., x_k$ obtained from maximizing the probability of conformance using the assumed model.

3.3.3 Pre-Posterior Analysis

In the above example the maximum model averaged posterior probability of conformance to the specifications $(-\infty, 5)$ was 0.8851. In practice, a process engineer may feel that such probability of conformance is too small. There are two possible reasons for a relative low probability of conformance. The first reason is that the data is limited, and so given the available data, this is the highest probability of conformance that can be obtained. In this case, running more experiments and using the additional data could give a higher value of posterior probability of conformance, especially when the repeatability of the observed measures is high. The second reason is that the specification limits set by the process engineer are unrealistic. In such case there is no point in running more experiments as the additional data will not increase the probability of conformance. These two situations can be discerned by using a pre-posterior approach, as suggested by Peterson [56]. Table 3.11 shows the posterior probability of conformance, z^*, as well as the optimal levels of the control factors, (x_1^*, x_2^*), and the mean and standard deviation estimates of the posterior response at (x_1^*, x_2^*) for two cases. Both cases use the same values for the hyper-parameters as before with $\pi = 0.5$, and $\gamma = 2$. The first case (labelled "data" in table 3.11) uses the original data that is shown in table 3.6, and the second case (labelled "data+replicate" in table 3.11) uses the original data along with a replicate of the original data appended to the data. The data used in the second case would be valid if the experimental observations are completely repeatable. The way to mimic more data is simply based on replicating the X_i matrices and changing the corresponding degrees of freedom in the MAP computations [56]. For each of these two cases, the results are shown for various values of specification limits.

γ	π	x_1^*	x_2^*	z^*
0.5	0.25	1	-1	0.4878
0.5	0.50	1	-1	0.5191
0.5	0.75	1	-1	0.5218
1	0.25	1	-0.9151	0.7149
1	0.50	1	-0.9154	0.7487
1	0.75	1	-0.9154	0.7621
2	0.25	1	-0.9684	0.8428
2	0.50	1	-0.9198	0.8851
2	0.75	1	-0.9198	0.9094
5	0.25	1	-0.7760	0.8438
5	0.50	1	-0.7928	0.8992
5	0.75	1	-0.7930	0.9305
10	0.25	1	-1	0.8090
10	0.50	1	-0.5696	0.8557
10	0.75	1	-0.5995	0.8967
100	0.25	1	-0.9072	0.5563
100	0.50	1	0.5146	0.7038
100	0.75	1	0.7584	0.7944

Table 3.10: Sensitivity of solution with respect to parameters π and γ for example 2

	data					data+replicate				
	z^*	x_1^*	x_2^*	mean	std. dev.	z^*	x_1^*	x_2^*	mean	std. dev.
$Y^* < 5$	0.8851	1	-0.9198	3.5792	0.9670	0.9955	1	-1	2.8083	0.7599
$Y^* < 4$	0.6494	1	-1	3.5296	0.9896	0.9338	1	-1	2.8083	0.7599
$Y^* < 3$	0.3329	1	-1	3.5296	0.9896	0.5986	1	-1	2.8083	0.7599
$Y^* < 2$	0.1057	1	-1	3.5296	0.9896	0.1499	1	-1	2.8083	0.7599

Table 3.11: Pre-posterior analysis for example 2

It can be seen from the table that for the specification limits used earlier ($-\infty < Y^* < 5$), the posterior probability of conformance increases from 0.8851 to 0.9955 when one more replicate is used. Therefore, this is evidence that in this case it is worth considering running additional experiments in order to obtain a higher posterior probability of conformance given the data. However, when the specification limits are set as ($-\infty < Y^* < 2$), the posterior probability of conformance increases from 0.1057 to only about 0.1499. Thus, even when the repeatability of the process is high, the highest possible posterior probability of conformance is still very low. In this case, this is evidence that there is a need for re-designing the specification limits on the response.

3.4 Choice of Priors and Hyper-parameters

The previous sections were based on the assumption of an non-informative prior for $\log(\sigma^2)$, a non-informative prior on the β for the constant term, and a normally distributed g-prior for the remaining β's. Other choices of prior that are typically considered in the literature

are the use of a non-informative prior for all the β's, or the use of a normally distributed g-prior for all the β's.

A non-informative prior for all the β's is the same as the prior we use earlier when $g \to \infty$. The non-informative prior is convenient for the calculation of the posterior predictive density since the resulting distribution is a t-distribution [58]. However, in the calculation of the model posteriors, this prior tends to favor the null model (i.e., a model with just the constant term). This can be explained based on Bayes' factors, since the model posteriors can be used for model selection in a Bayesian hypothesis testing for the true model [36]. Fernandez *et al.* [21] use Bayes factors to recommend using a non-informative prior just for the constant term β rather than using the g-prior for all the β's. They make the recommendation based on the ease of choosing the hyper-parameters when computing the Bayes' factors for the model posteriors.

For the priors chosen, there are only two hyper-parameters to be chosen, namely π and γ. In all cases here, we have assumed $\pi = 0.5$, so that the model posteriors are proportional to the marginal likelihood of the data. To choose γ, Meyer *et al.* [43] recommend using the value that minimizes the posterior probability of the null model. They use an empirical Bayesian approach to show that this value of gamma also maximizes the posterior density, $p(\gamma|\mathbf{y})$. As was done in the examples of section 3, a sensitivity analysis of the solutions with respect to variations on these two parameters should be conducted. Further justification for (essentially equivalent) priors as used here can be found in Meyer et al. [43].

3.5 Discussion

A Bayesian methodology for process optimization is proposed that prescribes operating points that are robust to uncertainty in the response model. Analytical results have been derived to obtain closed form expressions for the cumulative model-averaged posterior predictive density. The results have been applied to two examples that demonstrate the advantages of model-averaging using the Bayesian predictive approach. For cases where the model-averaged posterior probability of conformance to the specifications is small, a pre-posterior analysis is recommended. As shown in example 2, this analysis could be used to determine if additional experiments could result in a higher probability of conformance or if the specifications were too demanding to start with.

All the results in this chapter were based on the assumption that there are no noise factors present and that the error terms are normally distributed. In the next chapter, the methodology for model-robust process optimization is extended in the presence of noise factors and when the error terms follow the t-distribution which has a thicker tail than the normal distribution.

Chapter 4

Model-Robust Process Optimization with Noise Factors and Non-normal Error Terms

4.1 Introduction

As seen in the previous chapter, a natural way to optimize any process from a quality and reliability standpoint is to maximize the probability of conformance of the predicted responses to their specification limits (see [56]), and this can be achieved using a Bayesian predictive approach. The benefits of using this methodology are that, (a) the posterior predictive density of the responses can be used to make inferences on their future values, thus providing a mechanism to calculate the probability of conformance of the future responses, (b) the methodology takes into account the mean and the variance of the response, and (c) the methodology takes into account uncertainty in the model parameters. Peterson

[56] uses a Bayesian approach that involves obtaining the posterior predictive density of the response based on an assumed model, and maximizing the probability of obtaining the predicted response within certain limits or specifications. Miro-Quesada, Del Castillo and Peterson [46] extended this approach to include the presence of noise factors. In the previous chapter, this idea was taken one step ahead by using Bayesian model averaging to compute the model-averaged posterior predictive density (MAP) of the response in a single response process. The MAP is used for optimization with respect to the control factors in order to obtain the levels of the control factors that maximize the posterior probability of obtaining the response within some given specification limits. The solution presented is thus robust to the uncertainty in the true process model as well as to the model parameters for each competing model that is considered to represent the process. It was assumed in the previous chapter that there are no noise factors present in the system, and that the error terms in all the competing models are normally distributed. It is the purpose of this chapter to extend the MAP approach to cases where there are noise factors and when errors are t-distributed.

4.2 The Model-robust Approach for Process Optimization

In the previous chapter the process model was assumed to be of the form

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \tag{4.1}$$

where the scalar response variable y is dependent on a vector of regressors $\mathbf{x}$ given by a $(p \times 1)$ vector that are in turn functions of k controllable factors (i.e., $\mathbf{x}$ is in model form), ϵ is the error term, and $\boldsymbol{\beta}$ is the vector of process parameters. Given specification limits L and U for the response, the optimization problem was formulated as

$$\max_{x_1^*,\ldots,x_k^*} \quad P(L \leq Y^* \leq U) = \sum_i \left[\int_L^U P(y^*|M_i, \mathbf{x}^*, \mathbf{y}) dy^* \right] P(M_i|\mathbf{y}), \tag{4.2}$$

where y^* is the predicted value of the response at a new set of observed regressors $\mathbf{x}^*$ and $\mathbf{y}$ is the $(n \times 1)$ vector of observed responses from the experiment. The optimization was carried out with respect to the k control factors $(x_1, x_2, ..., x_k)$. $P(M_i|\mathbf{y})$ is the posterior probability of model M_i given the data, and $P(y^*|M_i, \mathbf{x}^*, \mathbf{y})$ is the posterior predictive density of the response for the model M_i, given the data at a new set of observed regressors, $\mathbf{x}^*$. For the priors given by equations (3.4), (3.8), (3.9) and (3.10), the model posteriors were given by

$$P(M_i|\mathbf{y}) \propto \pi^{f_i}(1-\pi)^{k-f_i}\gamma^{-t_i}|\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i|^{-\frac{1}{2}} S_i^{-\frac{(n-1)}{2}}, \tag{4.3}$$

where

$$\frac{g}{\gamma^2}\mathbf{V}_i = \boldsymbol{\Sigma}^{-1}, \tag{4.4}$$

$$\begin{aligned} S_i &= (\mathbf{y} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i)'(\mathbf{y} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i{}'\boldsymbol{\Sigma}_i^{-1}\hat{\boldsymbol{\beta}}_i & (4.5) \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}_i(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}, & (4.6) \end{aligned}$$

and

$$\hat{\boldsymbol{\beta}}_i = (\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}. \tag{4.7}$$

It was also shown that the posterior predictive density for model M_i follows a t-distribution,

$$y^*|M_i, \mathbf{x}^*, \mathbf{y} \ \propto \ t_\nu \left(\mathbf{x}^*\hat{\boldsymbol{\beta}}_i, \hat{\sigma}_i^2 \left[1 + \mathbf{x}^{*\prime}(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^* \right] \right). \tag{4.8}$$

In the next section these results are modified to account for the presence of noise factors.

4.3 Extension to Noise Factors

In practice, there are some factors that cannot be controlled at the "customer" level (whether this customer is the manufacturing plant or the end customer), but can be controlled under careful experimental conditions. These are referred to as *noise factors*. For example, in the production of automotive tires, the type of driver and the driving conditions might be noise factors. The objective in process optimization then is to find a solution, given by the optimal levels of the control factors, that is also robust to the variation in the noise factors. This is the so-called Robust Parameter Design (RPD) problem and was first formulated by Genichi Taguchi (see [66, 67, 68]). The traditional Taguchi experimental design involves varying both the control and the noise factors in a crossed array, with the control factors in the inner array and the noise factors in the outer array. More recently, the analysis of RPD problems is performed using the Dual Response approach [7, 49]. In this approach the mean and the variance of the response are modelled independently as functions of the control factors from a replicated experiment. Alternatively, the data from an unreplicated experiment may be used to fit a model of the form

$$\hat{y}(\mathbf{x_c}, \mathbf{x_n}) = b_o + \mathbf{x_c'b} + \mathbf{x_c'Bx_c} + \mathbf{x_n'c} + \mathbf{x_c'\Delta x_n}, \tag{4.9}$$

to the response as a function of both the controllable and noise factors. In equation (4.9), $\mathbf{x_c}$ is the vector of control factors, $\mathbf{x_n}$ is the vector of noise factors, and b_o, $\mathbf{b}$, $\mathbf{B}$, $\mathbf{c}$ and $\mathbf{\Delta}$ are the estimated parameters. This model is then used to get the mean and variance models (response surfaces) from assuming that the noise factors vary according to some known distribution, e.g., $\mathbf{x_n} \sim N(0, \mathbf{V_n})$. Here, the mean model is given by $E_{\mathbf{x_n}}[\hat{y}(\mathbf{x_c}, \mathbf{x_n})]$, and the variance model is given by $Var_{\mathbf{x_n}}[\hat{y}(\mathbf{x_c}, \mathbf{x_n})] + Var[\epsilon]$, where ϵ is the error term.

The mean and variance models are used to formulate an optimization problem such as finding a solution that minimizes the variance model subject to given bounds on the mean of the response [15]. However, this approach does not allow us to predict what fraction of future responses (e.g., proportion of products in a manufacturing process) will fall within the specifications at the optimal setting $\mathbf{x_c^*}$ as the dual response surfaces give only the "mean models" (i.e., they give only point-estimate values for the mean and the variance of the response at the optimal setting). In other words, there can be no inference made about the reliability or conformance of the process.

Miro-Quesada, Del Castillo and Peterson [46] present a Bayesian predictive approach for process optimization in the presence of noise factors for a multiple response process assuming a known Standard Multivariate Regression model. Their approach addresses the uncertainty in the parameter estimates of a given model, but does not address the uncertainty in the true model of the process. Here, the Bayesian predictive approach is extended to also account for uncertainty in the true process model using the MAP approach reviewed in section 4.2.

Consider a process with a single response variable y which is dependent on a vector of regressors $\mathbf{x}$ that are in turn functions of k factors. It is assumed that k_c out of the k factors are control factors, and the rest are noise factors. It is assumed that a suitable experiment with n runs has been designed and carried out and the data from the experiment is available. The $(n \times k)$ design matrix used is denoted by $\mathbf{X}$, that includes treatment combinations of both the control and the noise factors. The $(n \times 1)$ vector of responses from the experiment

is denoted by $\mathbf{y}$. Each of the potential process models is assumed to be of the form

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \tag{4.10}$$

where ϵ is the normally distributed error term, and $\boldsymbol{\beta}$ is the vector of process parameters. Let r_i be the number of terms in model M_i and t_i be the number of terms in the model excluding the constant term. Let $\mathbf{X}_i$ be the $(n \times r_i)$ design matrix corresponding to M_i. Here, it is noted that the terms in the models could contain functions of both the control and the noise factors, and $\mathbf{X}_i$ contains a column for each of these terms. Given the data, the posterior probability of the competing models in the presence of noise factors is also given by equation (4.3).

For each model M_i it is necessary to compute the cumulative posterior predictive density of the future value of the response y^* at some future level of the factors $\mathbf{x}^*$. In order to do this, we partition the vector $\mathbf{x}^*$ as $[\mathbf{x}^*_\mathbf{c}, \mathbf{x}^*_\mathbf{n}]$, where $\mathbf{x}^*_\mathbf{c}$ is the future level of the control factors, and $\mathbf{x}^*_\mathbf{n}$ is the future level of the noise factors. The cumulative posterior predictive density $P(Y^* < y|M_i, \mathbf{x}^*, \mathbf{y})$ is given by equation (3.19). However, because of the presence of noise factors whose future value at the "customer" level cannot be controlled, it is of interest to compute the expected posterior probability at a given level of the control factors $\mathbf{x}^*_\mathbf{c}$ with respect to all possible values of the noise factors $\mathbf{x}^*_\mathbf{n}$. Just as in the Dual Response approach, it will be assumed that the noise factors at the "customer" level are distributed with known p.d.f. $f_{\mathbf{x_n}}$ according to $\mathbf{x_n} \sim N(0, \mathbf{V_n})$. Then, the cumulative posterior predictive density is given by,

$$P(L \leq Y^* \leq U|M_i, \mathbf{x}^*, \mathbf{y}) = \int_{\mathbf{x}^*_\mathbf{n}} P(L \leq Y^* \leq U|M_i, \mathbf{x}^*_\mathbf{c}, \mathbf{x}^*_\mathbf{n}, \mathbf{y}) f_{\mathbf{x_n}}(\mathbf{x}^*_\mathbf{n})\, d\mathbf{x}^*_\mathbf{n} \tag{4.11}$$

$$= E_{\mathbf{x}^*_\mathbf{n}}\left[P(L \leq Y^* \leq U|M_i, \mathbf{x}^*_\mathbf{c}, \mathbf{x}^*_\mathbf{n}, \mathbf{y})\right]. \tag{4.12}$$

The optimization problem in the presence of noise factors is then formulated as

$$\max_{\mathbf{x_c^*}} \quad E_{\mathbf{x_n^*}}\left[P(L \leq Y^* \leq U)\right] = E_{\mathbf{x_n^*}}\left[\sum_i \left(\int_L^U P(y^*|M_i, \mathbf{x}^*, \mathbf{y})dy^*\right) P(M_i|\mathbf{y})\right] \quad (4.13)$$

$$= \sum_i E_{\mathbf{x_n^*}}\left[\left(\int_L^U P(y^*|M_i, \mathbf{x}^*, \mathbf{y})dy^*\right)\right] P(M_i|\mathbf{y}) \quad (4.14)$$

The objective function above can be computed using equations (4.3) and (3.18). The expected value with respect to the noise factors can be computed by simulation, using the steps below:

1. Set $count = 1$

2. Generate a sample $\mathbf{x_n}(count)$ from its assumed distribution $N(0, \mathbf{V_n})$

3. Compute $P(L \leq Y^* \leq U|M_i, \mathbf{x_c^*}, \mathbf{x_n^*}(count), \mathbf{y})$ for the sample using equation (3.18).

4. Set $count = count + 1$. Repeat steps 2 and 3 until $count > N$.

5. Estimate the expected value using the Weak Law of Large Numbers (WLLN),

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left[P(L \leq Y^* \leq U|M_i, \mathbf{x_c^*}, \mathbf{x_n^*}(i)\mathbf{y})\right] = E_{\mathbf{x_n^*}}\left[P(L \leq Y^* \leq U|M_i, \mathbf{x_c^*}, \mathbf{x_n^*}\mathbf{y})\right]. \quad (4.15)$$

The example below illustrates the proposed method.

4.3.1 Example

The data for this example is taken from Derringer and Suich [17], and is given in table 4.1. There are three factors, x_1(hydrated silica level), x_2 (silane coupling agent level) and x_3 (sulfur level), and four responses, y_1 (PICO Abrasion index), y_2 (200% modulus), y_3 (Elongation at break), and y_4 (Hardness). Here, only consider response y_3 is considered. It is

assumed that it is desired to obtain y_3 within the specification limits $[400, 600]$. It is further assumed that factor x_1 is a noise factor. Using the data in table 4.1, and the priors mentioned earlier, the model posteriors for all subsets belonging to the model class shown in equation (4.9) are computed. Here, the hyper-parameters are set at $\pi = 0.5$, and $\gamma = 0.6$. The choice of these hyper-parameters is as discussed in the previous chapter. In the models considered, higher order effects are included only if the corresponding main effect is present. The model posterior probabilities and Ordinary Least Square (OLS) statistics for the 20 models with the highest posteriors are shown in table 4.2. In the table, each row represents a competing model and under the columns containing the model terms (effects), a '1' indicates that the term is present in the model and a '0' indicates otherwise. The OLS statistics shown in the table are the R-square, the Adjusted R-square, and the standard error (S.E.). It can be seen based on both the model posteriors and the OLS statistics that there are multiple competing models for the process. For simplicity when averaging over the models, only model numbers 1-7 from table 4.2 are considered as theyt account for over 95% of the total probability (see appendix A).

For the optimization, it is assumed that the coded control factors are constrained to lie in the interval $[-1, 1]$. It is also assumed that the coded noise factor has a $N(0, 1/3^2)$ distribution so that its corner points in the design in table 4.1 are set at a value equal to 3 times the standard deviation and its center point is at the mean. If all the factors are assumed to be controllable, then the cumulative model averaged posterior predictive density for the given specification limits computed using equations (4.3) and (3.18) is maximized at the setting $x_1 = -0.7490$, $x_2 = -0.5294$ and $x_3 = -0.2568$ giving a probability of conformance of 0.7968.

If instead x_1 is considered to be a noise factor and the optimization is performed only with respect to x_2 and x_3 using equation (4.14), then the optimal settings are $x_2 = -0.9876$ and $x_3 = -1.0000$ giving a probability of conformance of 0.7275. The presence of the noise factor affects not only the variance of the posterior predictive distribution of the response at a given $\mathbf{x}$, but also the mean because of the presence of potential models containing interaction terms between the control and the noise factors. Thus, when x_1 is considered as a noise factor in the optimization, there is not only a decrease in the posterior probability of conformance but also a shift in the optimal set point of the control factors.

In this example, the expected value in equation (4.12) was computed within the optimization routine by simulating over a total of 2000 runs. Ten such replicates at the optimal setting of the control factors give an estimated posterior probability of conformance with mean 0.7275 and standard error 0.0021.

Run	x_1	x_2	x_3	y_1	y_2	y_3	y_4
1	-1	-1	-1	102	900	470	67.5
2	1	-1	-1	120	860	410	65
3	-1	1	-1	117	800	570	77.5
4	1	1	-1	198	2294	240	74.5
5	-1	-1	1	103	490	640	62.5
6	1	-1	1	132	1289	270	67
7	-1	1	1	132	1270	410	78
8	1	1	1	139	1090	380	70
9	-1.63	0	0	102	770	590	76
10	1.63	0	0	154	1690	260	70
11	0	-1.63	0	96	700	520	63
12	0	1.63	0	163	1540	380	75
13	0	0	-1.63	116	2184	520	65
14	0	0	1.63	153	1784	290	71
15	0	0	0	133	1300	380	70
16	0	0	0	133	1300	380	68.5
17	0	0	0	140	1145	430	68
18	0	0	0	142	1090	430	68
19	0	0	0	145	1260	390	69
20	0	0	0	142	1344	390	70

Table 4.1: Data for example in section 4.3.1 from Derringer and Suich [17]

Model No.	const.	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x_2^2	x_3^2	R^2	R^2_{Adj}	*S.E.*	$P(M_i\|data)$
1	1	1	1	1	0	1	0	1	0	0.7009	0.5941	69.8170	0.3063
2	1	1	1	1	1	1	0	1	0	0.7023	0.5649	72.2860	0.1591
3	1	1	1	1	0	1	1	1	0	0.7010	0.5630	72.4450	0.1557
4	1	1	1	1	0	1	0	1	1	0.7009	0.5629	72.4520	0.1277
5	1	1	1	1	1	1	1	1	0	0.7024	0.5287	75.2310	0.0809
6	1	1	1	1	1	1	0	1	1	0.7023	0.5286	75.2370	0.0663
7	1	1	1	1	0	1	1	1	1	0.7010	0.5266	75.4030	0.0649
8	1	1	1	1	1	1	1	1	1	0.7024	0.4859	78.5760	0.0337
9	1	0	0	0	0	0	0	0	0	0.0000	0.0000	109.5870	0.0054
10	1	1	0	0	0	0	0	0	0	0.0576	0.0052	109.3010	0.0000
11	1	0	1	0	0	0	0	0	0	0.0438	-0.0093	110.0940	0.0000
12	1	1	1	0	0	0	0	0	0	0.1014	-0.0043	109.8230	0.0000
13	1	1	0	1	0	0	0	0	0	0.6380	0.5954	69.7040	0.0000
14	1	1	0	0	0	0	0	1	0	0.0740	-0.0350	111.4870	0.0000
15	1	0	1	1	0	0	0	0	0	0.6243	0.5801	71.0140	0.0000
16	1	0	1	0	0	0	0	0	1	0.0439	-0.0686	113.2810	0.0000
17	1	1	1	1	0	0	0	0	0	0.6818	0.6222	67.3590	0.0000
18	1	1	1	0	1	0	0	0	0	0.1028	-0.0655	113.1170	0.0000
19	1	1	1	0	0	0	0	1	0	0.1178	-0.0476	112.1650	0.0000
20	1	1	1	0	0	0	0	0	1	0.1015	-0.0670	113.1970	0.0000

Table 4.2: Least square regression statistics and posterior probabilities for competing models in example 2.1

4.4 Extension to Non-normal Error Terms

Traditional model-fitting approaches in regression analysis use ordinary least square (OLS) estimates for the parameters $\boldsymbol{\beta}$ in a model of the form given by equation (4.1). However, when the standardized residuals from the fitted model have large magnitudes (large outliers), the estimates using OLS are poor. This can happen when the noise term $\boldsymbol{\epsilon}$ deviates from the assumed normal distribution (i.e., thicker tails in the distribution). In such cases, a robust regression approach is used, where the parameter estimates are obtained using methods such as Least Absolute Deviations (norm), M-Estimators, Least Median Squares, or Ranked Residuals [61, 19], which are less sensitive to non-normal errors than OLS. However, these are non-Bayesian and cannot be incorporated in our approach. Robust regression techniques do not consider uncertainties in the model or in the model parameters. Here, the idea of Bayesian model-averaged process optimization is extended to obtain a solution that is also robust to t_ν-distributed error terms. Here also, the objective is to maximize the posterior predictive probability that the response lies within the specification limits.

Consider here a process with a single response variable y which is dependent on regressors $\mathbf{x}$ given by a $(p \times 1)$ column vector which are functions of k controllable factors. It is assumed that there are no noise factors present for simplicity, although the results from the previous section can be applied here. It is assumed that a suitable experiment with n runs has been designed and carried out and the data from the experiment is available. The $(n \times k)$ design matrix used is denoted by $\mathbf{X}$, and the $(n \times 1)$ vector of responses from the

experiment is denoted by $\mathbf{y}$. Each of the potential models is assumed to be of the form

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \tag{4.16}$$

where ϵ is the error term that is assumed in this section to follow a t-distribution with ν degrees of freedom. As is well known, if $e \sim N(0,1)$, and $u \sim \chi^2(\nu)$, then $e/\sqrt{w}$ has a t-distribution with ν degrees of freedom, where $w = u/\nu$. Thus, conditional on w, equation (4.16) can be considered as a weighted regression model,

$$y = \mathbf{x}'\boldsymbol{\beta} + e/\sqrt{w}. \tag{4.17}$$

The transformations $\tilde{y} = \sqrt{w}y$ and $\tilde{\mathbf{x}} = \sqrt{w}\mathbf{x}$ conditional on w, give the non-weighted regression model,

$$\tilde{y} = \tilde{\mathbf{x}}'\boldsymbol{\beta} + e. \tag{4.18}$$

Since the error term in equation (4.18) has a standard normal distribution, the model posteriors in this case can be obtained from equation (4.3). If $\mathbf{X}_i$ is the design matrix corresponding to model M_i, then using the notations and the priors in equations (3.4), (3.8), (3.9), and (3.10), the model posteriors conditional on $\mathbf{W}_i$ are given by,

$$\begin{aligned} P(M_i|\mathbf{y}, w_1, ...w_n) &\propto \pi^{f_i}(1-\pi)^{k-f_i}\gamma^{-t_i}|\boldsymbol{\Sigma}_i^{-1} + \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i|^{-\frac{1}{2}}\tilde{S}_i^{-\frac{(n-1)}{2}} && (4.19)\\ &= \pi^{f_i}(1-\pi)^{k-f_i}\gamma^{-t_i}|\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i|^{-\frac{1}{2}}\tilde{S}_i^{-\frac{(n-1)}{2}}, && (4.20) \end{aligned}$$

where, $\tilde{\mathbf{X}}_i = \begin{bmatrix} \sqrt{w_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{w_n} \end{bmatrix} \mathbf{X}_i$, diagonal matrix $\mathbf{W}_i = \begin{bmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{bmatrix}$,

$$\begin{aligned} \tilde{S}_i &= (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}}_i)'(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i{}'\boldsymbol{\Sigma}_i^{-1}\hat{\boldsymbol{\beta}}_i && (4.21)\\ &= (\mathbf{y} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i)'\mathbf{W}(\mathbf{y} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i{}'\boldsymbol{\Sigma}_i^{-1}\hat{\boldsymbol{\beta}}_i, && (4.22) \end{aligned}$$

and,

$$\hat{\boldsymbol{\beta}}_i = (\boldsymbol{\Sigma}_i^{-1} + \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i)^{-1}\tilde{\mathbf{X}}_i'\tilde{\mathbf{y}} \tag{4.23}$$

$$= (\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{W}_i\mathbf{y}. \tag{4.24}$$

The model posteriors, $P(M_i|\mathbf{y})$ are then obtained by taking the expected value in equation (4.20) with respect to $(w_1, ..., w_n)$, that is

$$P(M_i|\mathbf{y}) = E_{w_1,...,w_n}\left[P(M_i|\mathbf{y}, w_1, ..., w_n)\right]. \tag{4.25}$$

The expected value with respect to $(w_1, ..., w_n)$ in the above equation is computed by sampling from a chi-square distribution according to the following numerical procedure:

1. Set $count = 1$
2. Generate samples $w_1(count), ...w_n(count)$ from their assumed distribution $\sim \chi^2(\nu)$
3. Compute $P(M_i|\mathbf{y}, w_1, ...w_n)$ for the sample using equation (4.20).
4. Set $count = count + 1$. Repeat steps 2 and 3 until $count > N$.
5. Estimate the desired expected value using the Weak Law of Large Numbers (WLLN),

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left[P(M_i|\mathbf{y}, w_1, ...w_n)\right] = E_{w_1,...w_n}\left[P(M_i|\mathbf{y}, w_1, ...w_n)\right]. \tag{4.26}$$

Given a future setting of the control factors $\mathbf{x}^*$ and $w^* = u^*/\nu$, where $u^* \sim \chi^2{}_\nu$, and using the transformation $\tilde{\mathbf{x}^*} = \sqrt{w^*}\mathbf{x}^*$, we get the posterior predictive density of $\tilde{y^*} = \sqrt{w^*}y^*$ from equation (4.8) as,

$$\tilde{y^*}|M_i, \mathbf{x}^*, w^*, w_1, ..., w_n, \mathbf{y} \;\propto\; t_\nu\left(\tilde{\mathbf{x}^*}\hat{\boldsymbol{\beta}}_i, \hat{\sigma}_i^2\left[1 + \tilde{\mathbf{x}^*}'(\boldsymbol{\Sigma}_i^{-1} + \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i)^{-1}\tilde{\mathbf{x}^*}\right]\right), \tag{4.27}$$

where $\nu = n - 1$ and $\hat{\sigma}_i^2 = \tilde{S}_i/(n-1)$ for models that include a constant term, and $\nu = n$ and $\hat{\sigma}_i^2 = \tilde{S}_i/n$ otherwise. Thus,

$$y^*|M_i, \mathbf{x}^*, w^*, w_1, ..., w_n, \mathbf{y} \ \propto \ t_\nu \left(\sqrt{w^*}\tilde{\mathbf{x}^*}\hat{\boldsymbol{\beta}}_i, w^*\hat{\sigma}_i^2 \left[1 + \tilde{\mathbf{x}^*}'(\boldsymbol{\Sigma}_i^{-1} + \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i)^{-1}\tilde{\mathbf{x}^*} \right] \right). \quad (4.28)$$

The cumulative posterior predictive density, $P(L < y^* < U|M_i, \mathbf{x}^*, \mathbf{Y})$, is thus calculated from the c.d.f. of the t-distribution taking the expected value with respect to w^* and $(w_1, ..., w_n)$:

$$P(L < y^* < U|M_i, \mathbf{x}^*, \mathbf{Y}) = E_{w^*} \left[E_{w_1,...,w_n} \left\{ P(L < Y^* < U|M_i, \mathbf{x}^*, w^*, w_1, ..., w_n, \mathbf{Y}) \right\} \right]. \quad (4.29)$$

The expected values in equation (4.29) can also be computed using simulation. The objective function in equation (4.2) can be computed using equations (4.25) and (4.29). The examples below illustrate the method. The first example illustrates the interplay of the different models under consideration and the effect of a thick-tail distribution. The second example is an application to a real experiment.

4.4.1 Example 1

To illustrate the methodology, consider the simulated data given in table 4.3, where there is a single response y and a single controllable factor x. Assuming normally distributed errors, the parameters for the linear and quadratic model can be estimated using equation (4.7). The fitted models are plotted in figure 4.1, using a value of $\pi = 0.5$ and $\gamma = 2.4$. The choice of these hyper-parameters is as discussed in the previous chapter. The posterior probabilities for the models under consideration are computed by using equation (4.8) for the case of normal errors and by using equation (4.25) for the case of t-distributed errors

x	y
-1.0	0.8551
-0.5	-1.8702
0.0	1.3421
0.5	3.7778
1.0	8.2322

Table 4.3: Sample data for example in section 4.4.1

model form	$P(M_i\|y)$	
	normal errors	t-distributed errors (5 d.o.f.)
$y = \beta_0$	0.108	0.127
$y = \beta_0 + \beta_1 x$	0.263	0.298
$y = \beta_0 + \beta_1 x + \beta_2 x^2$	0.629	0.575

Table 4.4: Model posteriors for example in section 4.4.1

with 5 degrees of freedom. These are shown in table 4.4. From table 4.4, it can be seen that the ratio of the posterior probability of the quadratic model to that of the linear model is higher when we assume normal errors as opposed to t-distributed errors. Figure 4.2 shows the cumulative model averaged posterior predictive density $P(L < y^* < U|x^*)$ over values of x^* in the range $[-1, 1]$, under both normal and t-distributed errors, for different choices of the specification limits. In this example, the optimal x^* in all the cases in the figure does not change much between normal errors and t-distributed errors with 5 degrees of freedom. However, it can be seen that depending on the specification limits,

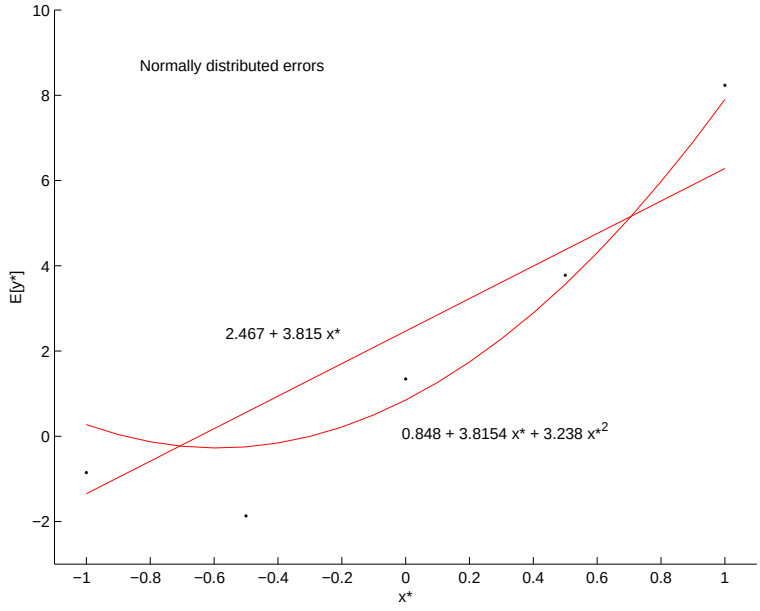

Figure 4.1: Parameter estimates for linear and quadratic model under normal errors assumption

the probability $P(L < y^* < U|x^*)$ can be much different depending on the the distribution of the error term. It is also seen that large differences occur at those values of x^* where the linear and quadratic models shown in figure 4.1 are wider apart. As the assumption of normally distributed errors relatively favors the quadratic model more as compared to the assumption of t-distributed errors, it can be seen that $P(L < y^* < U|x^*)$ at the optimal x^* is higher for the normally distributed errors in the cases where $L = 5, U = \infty$ and $L = 1, U = 3$, and higher for the t-distributed errors in the case where $L = -\infty, U = 2$.

4.4.2 Example 2: A mixture experiment

This example applies the methodology to real data. The data for this example is a mixture experiment taken from Frisbee et al. [22], and is given in table 4.5. The response is glass

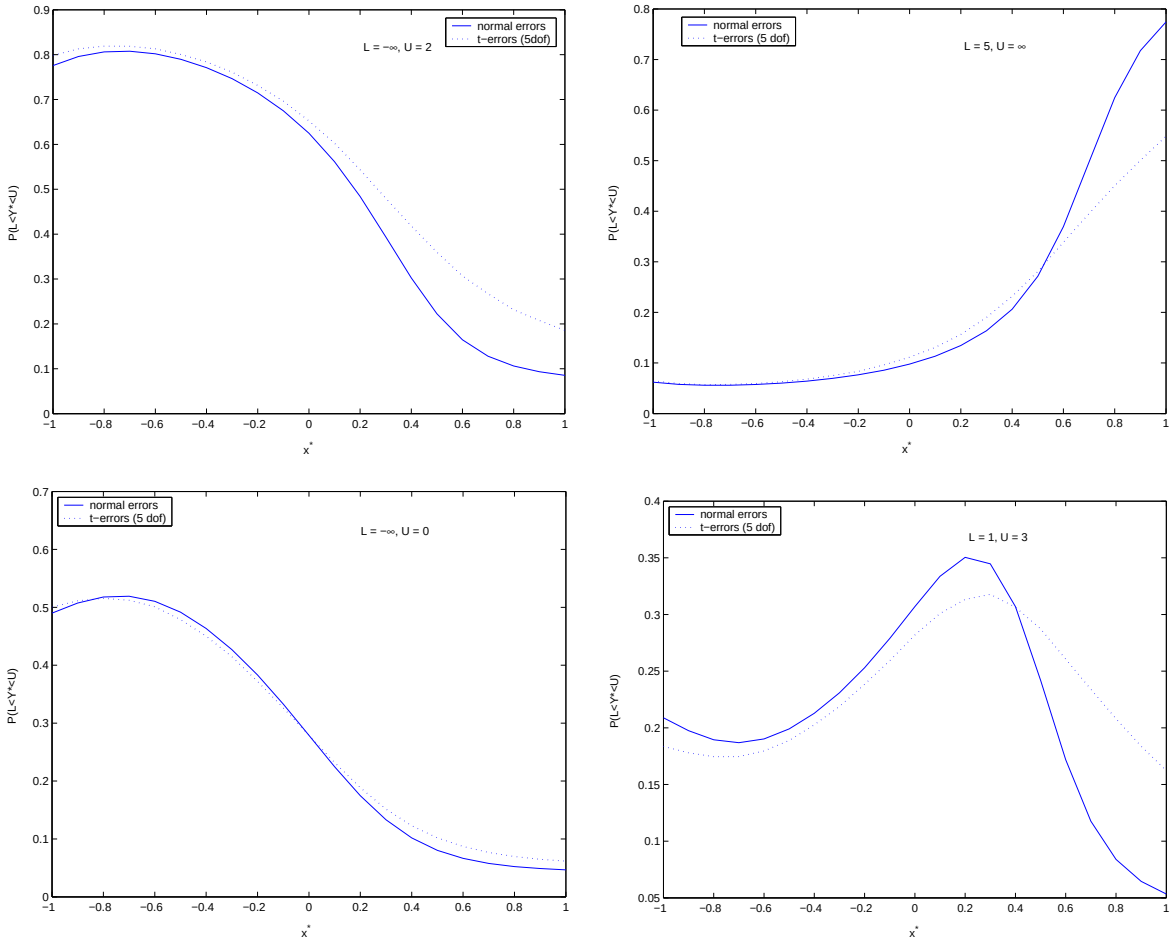

Figure 4.2: Cumulative model-averaged posterior probabilities for different specifications L, U

x_1	x_2	x_3	y
1.000	0.000	0.000	18.90
0.000	1.000	0.000	15.20
0.000	0.000	1.000	35.00
0.500	0.500	0.000	16.10
0.500	0.000	0.500	18.90
0.000	0.500	0.500	31.20
0.333	0.333	0.333	19.30
0.666	0.167	0.167	18.20
0.167	0.666	0.167	17.70
0.167	0.167	0.666	30.10
0.333	0.333	0.333	19.00

Table 4.5: Mixture data from [22] where the response is Glass Transition temperature

transition temperature of films cast from poly(DL-lactide) (PLA), and the controllable variables are amounts of non-ionic surfactants, namely, Polaxamer 188 NF (Pluronic® F68), Ployoxyethylene 40 monostearate (Myrj® 52-S) and Polyoxyethylene sorbitan fatty acid ester NF (Tween® 60). The authors are interested in finding the composition of the controllable factors that minimizes the glass transition temperature. Models belonging to the following two classes of models are considered:

$$y = b_1x_1 + b_2x_2 + b_3x_3 + b_4(x_1x_2) + b_5(x_1x_3) + b_6(x_2x_3), \tag{4.30}$$

$$y = b_1x_1 + b_2x_2 + b_3x_3 + b_4\, min(x_1, x_2) + b_5\, min(x_1, x_3) + b_6\, min(x_2, x_3). \tag{4.31}$$

The results for model-robust process optimization for these classes of models under normal errors were seen in the previous chapter. However, as there are only 11 runs, it is difficult to verify the distribution of the error terms. Tables 4.6, 4.7 and 4.8 give the model posteriors for the cases where the error distribution is assumed to be normal, t-distributed with 100 degrees of freedom (d.o.f.) and t-distributed with 10 d.o.f., respectively. These are obtained using hyper-parameters $\pi = 0.5$ and $\gamma = 10$ as discussed in the previous chapter. The tables also show the OLS statistics for the models that are based on the sum of squares of the residuals (SSE), the total sum of squares (SST), and the standard error ($S.E.$). As seen in the previous example, the posterior probabilities of the competing models for the t-distributed errors is different from those obtained using normally distributed errors, especially at lower degrees of freedom. It is also noted that the ordering of models according to the model posteriors is different depending on the error distribution. Thus, it is expected that the optimal solution will differ depending on the error distribution.

Table 4.9 shows the results of the optimization under the different error distributions. For each type of distribution, the MAP is used to maximize the posterior probability of obtaining a glass transition temperature lesser than 18, i.e., $P(Y^* < 18)$. The table shows the optimal setting of the controllable factors (x_1^*, x_2^*, x_3^*) obtained for each case of the noise distribution. As observed in the previous example, although there is a difference in the probability of conformance of the response for t-distributed errors as compared to normally distributed errors, especially as the degrees of freedom is lesser, there is no drastic shift in the optimal setting of the control factors. A possible explanation is that although the model posteriors differ depending on the chosen error distribution, the shape of the surface of the model-averaged posterior probability of conformance as a function of the

control factors is very close for all the error distributions considered, as can be seen in figure 4.2 for the first example. In other words, model averaging under normal errors will "robustify" the optimal solution if models that better explain abnormal observations with respect to some other model are included in the analysis. Because of this, the differences between normal MAP and t-MAP optimization will not be much if a "rich enough" set of models is included.

Model No.	const.	A	B	C	AB	AC	BC	$min(A,B)$	$min(A,C)$	$min(B,C)$	$\frac{SSE}{SST}$	$\frac{SSE/n-r_i}{SST/n-1}$	$S.E.$	$P(M_i\|data)$
1	0	1	1	1	0	0	0	0	1	1	0.0223	0.0372	1.3200	0.3557
2	0	1	1	1	0	0	0	0	1	0	0.0696	0.0994	2.1572	0.1430
3	0	1	1	1	0	1	1	0	0	0	0.0478	0.0797	1.9323	0.1380
4	0	1	1	1	0	0	0	1	1	1	0.0182	0.0365	1.3071	0.0910
5	0	1	1	1	0	1	0	0	0	0	0.0910	0.1300	2.4675	0.0803
6	0	1	1	1	1	1	1	0	0	0	0.0384	0.0769	1.8972	0.0759
7	0	1	1	1	1	1	0	0	0	0	0.0830	0.1383	2.5451	0.0392
8	0	1	1	1	0	0	0	1	1	0	0.0692	0.1153	2.3232	0.0279
9	0	1	1	1	0	0	0	0	0	0	0.2252	0.2815	3.6304	0.0146
10	0	1	1	1	0	0	1	0	0	0	0.1875	0.2678	3.5413	0.0119
11	0	1	1	1	1	0	0	0	0	0	0.2146	0.3066	3.7886	0.0068
12	0	1	1	1	1	0	1	0	0	0	0.1753	0.2922	3.6991	0.0058
13	0	1	1	1	0	0	0	0	0	1	0.2120	0.3028	3.7654	0.0037
14	0	1	1	1	0	0	0	1	0	0	0.2119	0.3027	3.7646	0.0034
15	1	0	0	0	0	0	0	0	0	0	1.0000	1.0000	6.8426	0.0015
16	0	1	1	1	0	0	0	1	0	1	0.1901	0.3169	3.8519	0.0011
17	0	0	1	1	0	0	0	0	0	0	1.0888	1.2098	7.5262	0.0001
18	0	1	0	1	0	0	0	0	0	0	1.2697	1.4107	8.1273	0.0001
19	0	0	1	1	0	0	1	0	0	0	1.0671	1.3338	7.9026	0.0001
20	0	1	0	1	0	1	0	0	0	0	1.0958	1.3697	8.0083	0.0000
21	0	0	1	1	0	0	0	0	0	1	1.0440	1.3050	7.8169	0.0000
22	0	0	0	1	0	0	0	0	0	0	2.7513	2.7513	11.3500	0.0000
23	0	1	0	1	0	0	0	0	1	0	1.1881	1.4852	8.3390	0.0000
24	0	0	1	0	0	0	0	0	0	0	7.2363	7.2363	18.4070	0.0000
25	0	1	0	0	0	0	0	0	0	0	7.5341	7.5341	18.7818	0.0000
26	0	1	1	0	0	0	0	0	0	0	4.9586	5.5096	16.0613	0.0000
27	0	1	1	0	1	0	0	0	0	0	4.9138	6.1422	16.9584	0.0000
28	0	1	1	0	0	0	0	1	0	0	4.9459	6.1824	17.0138	0.0000

Table 4.6: Least square regression statistics and posterior probabilities for competing models for example 3.3 - normal errors

Model No.	const.	A	B	C	AB	AC	BC	$min(A,B)$	$min(A,C)$	$min(B,C)$	$\frac{SSE}{SST}$	$\frac{SSE/n-r_i}{SST/n-1}$	$S.E.$	$P(M_i\|data)$
1	0	1	1	1	0	0	0	0	1	1	0.0223	0.0372	1.3200	0.3514
2	0	1	1	1	0	0	0	0	1	0	0.0696	0.0994	2.1572	0.1453
3	0	1	1	1	0	1	1	0	0	0	0.0478	0.0797	1.9323	0.1365
4	0	1	1	1	0	0	0	1	1	1	0.0182	0.0365	1.3071	0.0903
5	0	1	1	1	0	1	0	0	0	0	0.0910	0.1300	2.4675	0.0816
6	0	1	1	1	1	1	1	0	0	0	0.0384	0.0769	1.8972	0.0755
7	0	1	1	1	1	1	0	0	0	0	0.0830	0.1383	2.5451	0.0400
8	0	1	1	1	0	0	0	1	1	0	0.0692	0.1153	2.3232	0.0285
9	0	1	1	1	0	0	0	0	0	0	0.2252	0.2815	3.6304	0.0150
10	0	1	1	1	0	0	1	0	0	0	0.1875	0.2678	3.5413	0.0123
11	0	1	1	1	1	0	0	0	0	0	0.2146	0.3066	3.7886	0.0071
12	0	1	1	1	1	0	1	0	0	0	0.1753	0.2922	3.6991	0.0062
13	0	1	1	1	0	0	0	0	0	1	0.2120	0.3028	3.7654	0.0038
14	0	1	1	1	0	0	0	1	0	0	0.2119	0.3027	3.7646	0.0035
15	1	0	0	0	0	0	0	0	0	0	1.0000	1.0000	6.8426	0.0015
16	0	1	1	1	0	0	0	1	0	1	0.1901	0.3169	3.8519	0.0011
17	0	0	1	1	0	0	0	0	0	0	1.0888	1.2098	7.5262	0.0001
18	0	1	0	1	0	0	0	0	0	0	1.2697	1.4107	8.1273	0.0001
19	0	0	1	1	0	0	1	0	0	0	1.0671	1.3338	7.9026	0.0001
20	0	1	0	1	0	1	0	0	0	0	1.0958	1.3697	8.0083	0.0000
21	0	0	1	1	0	0	0	0	0	1	1.0440	1.3050	7.8169	0.0000
22	0	0	0	1	0	0	0	0	0	0	2.7513	2.7513	11.3500	0.0000
23	0	1	0	1	0	0	0	0	1	0	1.1881	1.4852	8.3390	0.0000
24	0	0	1	0	0	0	0	0	0	0	7.2363	7.2363	18.4070	0.0000
25	0	1	0	0	0	0	0	0	0	0	7.5341	7.5341	18.7818	0.0000
26	0	1	1	0	0	0	0	0	0	0	4.9586	5.5096	16.0613	0.0000
27	0	1	1	0	1	0	0	0	0	0	4.9138	6.1422	16.9584	0.0000
28	0	1	1	0	0	0	0	1	0	0	4.9459	6.1824	17.0138	0.0000

Table 4.7: Least square regression statistics and posterior probabilities for competing models for example 3.3 - t_{100} errors

Model No.	const.	A	B	C	AB	AC	BC	$min(A,B)$	$min(A,C)$	$min(B,C)$	$\frac{SSE}{SST}$	$\frac{SSE/n-r_i}{SST/n-1}$	$S.E.$	$P(M_i\|data)$
1	0	1	1	1	0	0	0	0	1	1	0.0223	0.0372	1.3200	0.3021
2	0	1	1	1	0	0	0	0	1	0	0.0696	0.0994	2.1572	0.1697
3	0	1	1	1	0	1	1	0	0	0	0.0478	0.0797	1.9323	0.1256
4	0	1	1	1	0	1	0	0	0	0	0.0910	0.1300	2.4675	0.0915
5	0	1	1	1	0	0	0	1	1	1	0.0182	0.0365	1.3071	0.0825
6	0	1	1	1	1	1	1	0	0	0	0.0384	0.0769	1.8972	0.0728
7	0	1	1	1	1	1	0	0	0	0	0.0830	0.1383	2.5451	0.0464
8	0	1	1	1	0	0	0	1	1	0	0.0692	0.1153	2.3232	0.0365
9	0	1	1	1	0	0	0	0	0	0	0.2252	0.2815	3.6304	0.0211
10	0	1	1	1	0	0	1	0	0	0	0.1875	0.2678	3.5413	0.0161
11	0	1	1	1	1	0	0	0	0	0	0.2146	0.3066	3.7886	0.0108
12	0	1	1	1	1	0	1	0	0	0	0.1753	0.2922	3.6991	0.0089
13	0	1	1	1	0	0	0	0	0	1	0.2120	0.3028	3.7654	0.0057
14	0	1	1	1	0	0	0	1	0	0	0.2119	0.3027	3.7646	0.0055
15	1	0	0	0	0	0	0	0	0	0	1.0000	1.0000	6.8426	0.0021
16	0	1	1	1	0	0	0	1	0	1	0.1901	0.3169	3.8519	0.0019
17	0	0	1	1	0	0	0	0	0	0	1.0888	1.2098	7.5262	0.0003
18	0	0	1	1	0	0	1	0	0	0	1.0671	1.3338	7.9026	0.0002
19	0	0	1	1	0	0	0	0	0	1	1.0440	1.3050	7.8169	0.0001
20	0	1	0	1	0	0	0	0	0	0	1.2697	1.4107	8.1273	0.0001
21	0	1	0	1	0	1	0	0	0	0	1.0958	1.3697	8.0083	0.0001
22	0	1	0	1	0	0	0	0	1	0	1.1881	1.4852	8.3390	0.0000
23	0	0	0	1	0	0	0	0	0	0	2.7513	2.7513	11.3500	0.0000
24	0	0	1	0	0	0	0	0	0	0	7.2363	7.2363	18.4070	0.0000
25	0	1	0	0	0	0	0	0	0	0	7.5341	7.5341	18.7818	0.0000
26	0	1	1	0	0	0	0	0	0	0	4.9586	5.5096	16.0613	0.0000
27	0	1	1	0	1	0	0	0	0	0	4.9138	6.1422	16.9584	0.0000
28	0	1	1	0	0	0	0	1	0	0	4.9459	6.1824	17.0138	0.0000

Table 4.8: Least square regression statistics and posterior probabilities for competing models for example 3.3 - t_{10} errors

Error Distribution	Optimal Setting			$P(Y^* < 18)$
	x_1	x_2	x_3	
normal	0.1330	0.8670	0.0000	0.9388
t-100	0.0000	1.0000	0.0000	0.9189
t-10	0.0000	1.0000	0.0000	0.8792

Table 4.9: Optimization results for example 3.2

4.5 Discussion

An extension to the Bayesian method for model-robust optimization was presented which includes robustness to the presence of noise factors and to the case of non-normal error distribution. In the presence of noise factors, the model-averaged posterior predictive density was used to maximize the probability of conformance by optimizing over possible values of the controllable factors, while simulating the noise factors from their assumed distribution. The resulting optimal solution thus provided a setting of the controllable factors that is not only robust to the form of the true model, but also to the variation in the noise factors. When t-distributed error terms are assumed instead of normal errors, it was observed that the posterior probabilities of the models changed, as demonstrated using two examples. The posterior predictive density of the response given a model naturally decreases as the tail of the assumed error distribution gets thicker. The optimization thus gives a different solution, both in terms of the settings of the controllable factors as well as the probability of conformance, although there is no drastic shift in the former. It is recommended that the optimization be carried out under different assumptions of the error distribution, espe-

cially when the number of runs in the original design is small, so that the resulting solution is robust to the assumed distribution. One way to ensure robustness is to evaluate the probability of conformance at the optimal setting given by normally distributed errors, by assuming t-distributed errors with varying degrees of freedom. If the probability of conformance at this setting does not vary much, then the solution is robust to the distribution of the error terms.

The results in this and in the previous chapter were restricted to single response systems. The next chapter extends the Bayesian methodology for model-robust process optimization to multiple responses systems.

Chapter 5

Model-Robust Process Optimization - Extension to Multiple Responses

5.1 Introduction

In practice, most processes have multiple responses, and it is of interest to obtain each of these responses within their individual specification limits. As two or more responses may depend on the same control factors, it is not sufficient to consider the responses individually for optimization. Since it is very unlikely that the optimal setting of the control factor for one response is also the optimal for the other responses that depend on the same control factor, the methodology must consider all the responses simultaneously. In this chapter, the Bayesian method for model-robust process optimization is extended to multiple response systems.

Here, it is assumed that there are $q > 1$ responses. The joint model-averaged posterior predictive density for the q responses is calculated and is then used to find the setting of

the control factors that maximizes the joint posterior probability of conformance of the responses to their respective specification limits.

5.2 Bayesian Model Averaging

For multiple responses, each of the q responses is modelled as

$$y_j = \mathbf{x}_j'\boldsymbol{\beta}_j + \epsilon_j, \qquad j = \{1...q\}. \tag{5.1}$$

In the case of multiple responses, the methodology depends on:

1. The correlation between the ϵ_j for the different responses, and

2. The regressors $\mathbf{x}_j$ in the model for each of the response.

We denote the $(q \times q)$ variance-covariance matrix of the error terms by $\boldsymbol{\Sigma}_e$, or sometimes for convenience by $\boldsymbol{\Lambda}^{-1}$. Note that if all the responses have identical regressors, then $\mathbf{x}_j = \mathbf{x}$ for all j, and if the error terms ϵ_j are uncorrelated between the responses, then $\boldsymbol{\Sigma}_e$ is a diagonal matrix. Thus there are four different cases of multiple response systems based on the regressors present in the models for the individual responses and the correlation of the error terms between the responses as shown in table 5.1. The case when there are identical regressors for all the responses and correlated errors is the Standard Multivariate Regression (SMR) case. The case when there are unrelated regressors for the responses and correlated errors is called Seemingly Unrelated Regression (SUR) because though the regressors may be completely different for different responses, they cannot be modelled independently of each other because of the correlated error terms [74].

	identical regressors	unrelated regressors
uncorrelated errors	All the y_j have the same set of regressors, $\mathbf{x}_j = \mathbf{x}$, and the error terms ϵ_j are uncorrelated between the responses	Different y_j may have a different set of regressors, i.e., $\mathbf{x}_j$ is different for different j, and the error terms ϵ_j are uncorrelated between the responses
correlated errors	All the y_j have the same set of regressors, $\mathbf{x}_j = \mathbf{x}$, and the error terms ϵ_j are correlated between the responses	Different y_j may have a different set of regressors, i.e., $\mathbf{x}_j$ is different for different j, and the error terms ϵ_j are correlated between the responses

Table 5.1: Categories of models for multiple response systems

In this chapter, the Bayesian methodology for model-robust process optimization is extended to three of the four above mentioned cases. The methodology is extended for the two cases where the error terms are correlated, and for the SMR case. Recommendations for extending the method to the SUR case are provided in the summary in chapter 8. In the single response case, the number of models considered for averaging is relatively small, although it depends on the number of control factors in any particular application. For example, if all the models with just the first-order terms are considered, then are a total of $2^k - 1$ models that could potentially be considered, where k is the number of control factors. Thus, as seen in chapter 3, given m models $M_i, i = 1...m$, the posterior probability for each individual model could be calculated. For multiple response systems each $M_i, i = 1...m$, refers to a set of q models, one for each response. However, in the

cases where the responses have identical regressors, the number of models that could be potentially considered is the same as that of the single-response case, since all the q models in the set M_i must have the same regressors. However, when we assume that the responses can have different regressors, the number of potential models can be extremely large, even for cases with very few control factors. This is because the q models in the set M_i need not have the same regressors. Hence, multiple combinations of the vector of regressors are possible in each of the q models. Therefore, in this case, it is important to choose a small number of subsets of models (for e.g., one plausible model per response in the subset) for averaging rather than using an exhaustive list of models.

As in the single response case, it is assumed that there is data from an experiment with n runs. We denote the $(n \times p)$ design matrix by $\mathbf{X}$, and the $(n \times q)$ matrix of responses from the experiment by $\mathbf{Y}$. The model-averaged posterior predictive density (MAP) is calculated by taking the weighted average of the posterior predictive density $P(\mathbf{y}^*|x_1^*...x_k^*, \mathbf{Y}, M_i)$ of a future vector of responses $\mathbf{y}^*$ at a future setting of the control factors $(x_1^*...x_k^*)$ over each model set M_i, using the posterior probabilities of the model set $P(M_i|\mathbf{Y})$ as the weights. The optimization problem is formulated as:

$$\max_{x_1^*,...,x_k^*} P(\mathbf{y}^* \in \mathbf{R}) = \sum_i \left[\int_{\mathbf{R}} P(\mathbf{y}^*|M_i, x_1^*...x_k^*, \mathbf{Y}) d\mathbf{y}^* \right] P(M_i|\mathbf{Y}), \tag{5.2}$$

where $\mathbf{R}$ denotes the region of interest formed by the specification limits $[l_j, u_j]$ on each individual response y_j. The optimization thus requires two parts, namely, the posterior probability of M_i and the posterior predictive density of the response given M_i. The results for model-robust process optimization for the different cases are derived below.

Uncorrelated Errors:

In this case, since there is no correlation between the error terms of the responses, the joint posterior density for the vector of responses is equal to the product of the marginal posterior densities of the individual responses under the normality assumption. This is true no matter if the responses have identical regressors or not. So, in this case, the model posteriors and the posterior predictive density can be obtained directly from the results for the single response case.

Thus, if model set M_i contains models $M_{i1}...M_{iq}$ corresponding to responses $y_1...y_q$ respectively, the posterior probability is given by

$$P(M_i|\mathbf{Y}) = \prod_{j=1}^{q} P(M_{ij}|\mathbf{Y}), \tag{5.3}$$

where each of the $P(M_{ij}|\mathbf{Y})$ is the model posterior probability for the single response y_j that can be computed using equation (3.13). Note that the priors used are the same as in chapter 3. Using equations (3.13) and (5.3), the posterior probability $P(M_i|\mathbf{Y})$ can be computed for all $i = \{1...m\}$ candidate sets of models.

Now for a given model set M_i, we also need the joint posterior probability of conformance of the responses given by $P(\mathbf{y}^* \in \mathbf{R}|x_1^*...x_k^*, \mathbf{Y}, M_i)$. Since the errors are uncorrelated, this is simply the product of the marginal posterior probability of conformance of the individual responses, i.e.,

$$P(\mathbf{y}^* \in \mathbf{R}|x_1^*...x_k^*, \mathbf{Y}, M_i) = \prod_{j=1}^{q} P({y_j}^* \in [l_j, u_j]|x_1^*...x_k^*, \mathbf{Y}, M_{ij}), \tag{5.4}$$

where each of the $P({y_j}^* \in [l_j, u_j]|x_1^*...x_k^*, \mathbf{Y}, M_{ij})$ is the posterior probability of conformance of a single response y_j to its specification limits and can be evaluated using equation (3.19).

From equations (3.13), (3.19), (5.3) and (5.4), the objective function in equation (5.2) can be computed for multiple response systems with uncorrelated errors.

Correlated Errors and Identical Regressors:

In this case, the process follows a Standard Multivariate Regression (SMR) model. To distinguish between the variables for the multiple competing models, the subscript i will be used for variables that are specific to model set M_i. The model set M_i is given by

$$\mathbf{y} = \boldsymbol{\beta}_i' \mathbf{x}_i + \epsilon, \tag{5.5}$$

where $\mathbf{y}$ is the $(q \times 1)$ vector of responses, $\boldsymbol{\beta}_i$ is the $(p_i \times q)$ matrix of model parameters, $\mathbf{x}_i$ is the $(p_i \times 1)$ vector of regressors and $\boldsymbol{\epsilon}$ is the correlated error term with distribution $N(\mathbf{0}, \boldsymbol{\Sigma}_e)$. Assuming $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_e^{-1}$, we have $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Lambda}^{-1})$. As in the single response case it is assumed that there is data from an experiment with n runs. Denote the $(n \times p_i)$ design matrix by $\mathbf{X}_i$, and the $(n \times q)$ matrix of responses from the experiment by $\mathbf{Y}$.

The priors used on the parameters are similar to those used in the single response case. Vague priors are assumed for $\boldsymbol{\Lambda}$ and for the $\boldsymbol{\beta}_i$ for the constant term, and the Zellner's g-prior [74] is assumed for the remaining $\boldsymbol{\beta}_i$. Therefore, the prior probabilities are given by

$$P(\boldsymbol{\Lambda}) \propto \frac{1}{|\boldsymbol{\Lambda}|^{(q+1)/2}}, \tag{5.6}$$

$$P(\boldsymbol{\beta}_i) \propto \gamma^{-qt_i} |\boldsymbol{\Lambda}|^{t_i/2} \mathrm{e}xp \left[-\frac{1}{2} \mathrm{t}r \, \boldsymbol{\Lambda} (\boldsymbol{\beta}_i' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\beta}_i) \right], \tag{5.7}$$

and

$$P(\boldsymbol{\beta}_i, \boldsymbol{\Lambda}) \propto P(\boldsymbol{\beta}_i) P(\boldsymbol{\Lambda}), \tag{5.8}$$

where t_i is the number of terms in the model excluding the constant term (i.e., if the model includes a constant term, then $(t_i = p_i - 1)$ and if the model excludes a constant term then $(t_i = p_i)$, $\mathbf{\Sigma}_i^{-1} = (\mathbf{X}_i'\mathbf{X}_i)\mathbf{V}_i$, where $\mathbf{V}_i = \frac{1}{g}\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{t_i} \end{pmatrix}$, and γ is such that

$$\frac{g}{\gamma^2}\mathbf{V}_i = \mathbf{\Sigma}_i^{-1}. \tag{5.9}$$

Since, the responses have identical regressors, the prior on M_i is the same as that in the single response case given by

$$P(M_i) = \pi^{f_i}(1-\pi)^{k-f_i}, \tag{5.10}$$

where f_i out of the k control factors are present in M_i, and π is the prior probability of an active factor, assumed equal for all factors. The posterior probability of M_i is given by the Bayes' theorem:

$$P(M_i|\mathbf{Y}) = \frac{P(\mathbf{Y}|M_i)P(M_i)}{\sum_i P(\mathbf{Y}|M_i)P(M_i)}, \tag{5.11}$$

where $P(\mathbf{Y}|M_i)$ is the marginal likelihood, defined as

$$P(\mathbf{Y}|M_i) = \int_{\mathbf{\Lambda}}\int_{\boldsymbol{\beta}_i} P(\mathbf{Y}|M_i, \mathbf{\Lambda}, \boldsymbol{\beta}_i)P(\mathbf{\Lambda}, \boldsymbol{\beta}_i|M_i)\, d\boldsymbol{\beta}_i\, d\mathbf{\Lambda}. \tag{5.12}$$

$P(\mathbf{Y}|M_i, \mathbf{\Lambda}, \boldsymbol{\beta}_i)$ is the likelihood function, given by:

$$P(\mathbf{Y}|M_i, \mathbf{\Lambda}, \boldsymbol{\beta}_i) \propto |\mathbf{\Lambda}|^{n/2} exp\left[-\frac{1}{2}\mathrm{t}r\, \mathbf{\Lambda}(\mathbf{Y} - \mathbf{X}_i\boldsymbol{\beta}_i)'(\mathbf{Y} - \mathbf{X}_i\boldsymbol{\beta}_i)\right]. \tag{5.13}$$

Based on the assumed priors given in equations (5.6), (5.7),(5.8) and (5.10), the integral in equation (5.12) can be computed (see [45], [58]) and is given by,

$$P(\mathbf{Y}|M_i) \propto \gamma^{-qt_i}|\mathbf{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i|^{-\frac{1}{2}}|\mathbf{S}_i|^{-\frac{(n-1)}{2}}. \tag{5.14}$$

From equations (5.11) and (5.14), the model posterior probability is given by,

$$P(M_i|\mathbf{Y}) \propto \pi^{f_i}(1-\pi)^{k-f_i}\gamma^{-qt_i}|\mathbf{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i|^{-\frac{1}{2}}|\mathbf{S}_i|^{-\frac{(n-1)}{2}}, \tag{5.15}$$

where

$$\mathbf{S}_i = (\mathbf{Y} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i)'(\mathbf{Y} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i) + (\hat{\boldsymbol{\beta}}_i'\mathbf{\Sigma}_i^{-1}\hat{\boldsymbol{\beta}}_i), \tag{5.16}$$

and

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{Y}. \tag{5.17}$$

The posterior predictive density for the new vector of responses $\mathbf{y}^*$ at a new set of regressors $\mathbf{x}^*$ for a given model M_i under the priors on the parameters given by equations (5.6), (5.7) and (5.8) is derived in appendix C and is given by:

$$\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}, M_i \propto \mathbf{T}_q(\hat{\boldsymbol{\beta}}_i'\mathbf{x}^*, \mathbf{H}_i^{-1}) \tag{5.18}$$

where,

$$\mathbf{H}_i = \frac{\nu_i \mathbf{S_i}^{-1}}{1 + \mathbf{x}^{*\prime}(\mathbf{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^*},$$

and

$$\nu_i = n + t_i + 1 - p_i - q.$$

As t_i is the number of terms in model M_i excluding the constant term and p_i is the number of regressors in model M_i, the degrees of freedom for models having a constant term is $\nu_i = n - q$, and for models without constant terms is $\nu_i = n + 1 - q$.

Using equations (5.15) and (5.18), the objective function in equation (5.2) can be evaluated for the SMR case. The example below demonstrates the proposed methodology for the SMR case.

5.3 Example

This example uses data from a high performance liquid chromatography (HPLC) process [54] and is given in table 5.2. There are four responses, namely, y_1: the critical resolution (Rs), y_2: total run time, y_3: signal-to-noise ratio of the last peak and y_4: the tailing factor of the major peak. There are three controllable factors, x_1: %IPA, x_2: temperature and x_3: pH. It is assumed that the model is of the SMR form, i.e., all the responses have the same regressors and the error terms are correlated between the responses. Therefore, the joint posterior distribution of the responses for a given model is given by a multivariate **T**-distribution as shown in equation (5.18).

All models from the class of models including main effects, two-way interactions and quadratic effects are taken into consideration for model averaging. Models with higher order terms are considered only if the corresponding main effects are also present. There are 95 such models including the null model (model with only the constant term). The hyperparameters π and γ are chosen as described in chapter 3, i.e., the value of π is set at 0.5 and the value of γ is chosen such that the posterior probability of the null model is minimized. Here, a value of 5.6 is obtained for γ. Table 5.3 shows the posterior probabilities for 20 of these models sorted in the descending order of the model posteriors. In the table, each row represents a competing model and under the columns containing the model terms (effects), a '1' indicates that the term is present in the model and a '0' indicates otherwise. The table also gives the corresponding ordinary least square statistics. The columns $\prod R^2$ and $\prod R^2_{Adj}$ give the product of the R^2 and R^2_{Adj} for the individual responses, respectively. As in the case of single response systems, models that account for

over 95% of the total probability are taken into consideration for computing the model-averaged posterior predictive density.

The specification limits for optimization are chosen as $\{\mathbf{R} : 2.0 \leq y_1, y_2 \leq 15,300 \leq y_3, 0.75 \leq y_4 \leq 0.85\}$. For these specifications, the model-averaged posterior probability was maximized at a value of 0.8588 at the setting of the control factors given by $\mathbf{x}^* = (0.9951, 0.6091, -0.3485)$ in coded units. In order to compute the posterior probability of conformance for a given model, the following steps were used:

1. Set $count = 1$ and $iter = 1$.

2. Generate sample $\mathbf{y}^*(iter)$ from the $\mathbf{T}$-distribution shown in equation (5.18).

3. If $\mathbf{y}^*(iter) \in \mathbf{R}$, where $\mathbf{R}$ is the specification region, then increment $count = count+1$.

4. Set $iter = iter + 1$. Repeat steps 2-4 until $iter > N$.

5. An estimate of the posterior probability is given by $count/N$.

In this optimization, the posterior probabilities were estimated with a value of $N = 10,000$. Figure 5.1 shows the scatter plot of the model-averaged posterior probability of conformance $P(\mathbf{y}^* \in \mathbf{R}|\mathbf{x}^*)$ for this example. The legend shows the color code that corresponds to the value of $P(\mathbf{y}^* \in \mathbf{R}|\mathbf{x}^*)$. In addition, the size of the dots in the figure is proportional to the value of $P(\mathbf{y}^* \in \mathbf{R}|\mathbf{x}^*)$.

Run	%IPA, x_1	Temp., x_2	pH, x_3	coded x_1	coded x_2	coded x_3	Rs, y_1	Run time, y_2	S/N, y_3	Tailing, y_4
1	65	30	0.175	-1	-1	0	2.14	22	172	0.76
2	65	50	0.175	-1	1	0	1.73	12	311	0.88
3	65	40	0.05	-1	0	-1	1.93	16	251	0.8
4	65	40	0.3	-1	0	1	1.95	16	241	0.8
5	70	40	0.175	0	0	0	2.17	14	278	0.79
6	70	50	0.05	0	1	-1	1.97	11	371	0.86
7	70	30	0.3	0	-1	1	2.38	19	194	0.74
8	70	50	0.3	0	1	1	1.98	11	360	0.86
9	70	30	0.05	0	-1	-1	2.37	18	204	0.74
10	70	40	0.175	0	0	0	2.2	14	280	0.78
11	75	40	0.3	1	0	1	2.42	13	314	0.78
12	75	30	0.175	1	-1	0	2.61	17	223	0.73
13	75	50	0.175	1	1	0	2.14	10	410	0.85
14	75	40	0.05	1	0	-1	2.42	12	324	0.78
15	70	40	0.175	0	0	0	2.2	14	281	0.79

Table 5.2: Data for HPLC process example [54]

Model No.	const	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x_1^2	x_2^2	x_3^2	$\prod R^2$	$\prod R^2_{Adj}$	$P(M_i\|\mathbf{Y})$
1	1	1	1	0	1	0	0	0	1	0	0.9727	0.9619	0.8707
2	1	1	1	0	1	0	0	1	1	0	0.9316	0.9134	0.0559
3	1	1	1	1	1	0	0	0	1	0	0.9793	0.968	0.0436
4	1	1	1	1	1	0	0	1	1	0	0.9792	0.9677	0.0234
5	1	1	1	1	1	1	0	0	1	0	0.9493	0.9358	0.0035
6	1	1	1	1	1	0	1	0	1	0	0.9858	0.9753	0.0014
7	1	1	1	1	1	0	0	0	1	1	0.9378	0.9136	0.0010
8	1	1	1	0	1	0	0	0	0	0	0.9089	0.8943	0.0003
9	1	1	1	1	1	1	0	1	1	0	0.9360	0.9111	0.0001
10	1	1	1	1	1	0	1	1	1	0	0.9558	0.9385	0.0001
11	1	1	1	1	1	0	0	1	1	1	0.9806	0.9661	0.0001
12	1	1	1	1	1	1	1	0	1	0	0.9808	0.9666	0
13	1	1	1	1	1	1	0	0	1	1	0.9807	0.9664	0
14	1	1	1	1	1	0	0	0	0	0	0.9557	0.9383	0
15	1	1	1	1	1	0	1	0	1	1	0.9875	0.9751	0
16	1	1	1	0	0	0	0	0	1	0	0.9874	0.9749	0
17	1	1	1	1	1	1	1	1	1	0	0.9869	0.9739	0
18	1	1	1	0	1	0	0	1	0	0	0.9402	0.9078	0
19	1	1	1	1	1	1	0	1	1	1	0.9395	0.9068	0
20	1	1	1	1	1	0	1	1	1	1	0.9422	0.9110	0

Table 5.3: Least square regression statistics and posterior probabilities for competing models for HPLC example

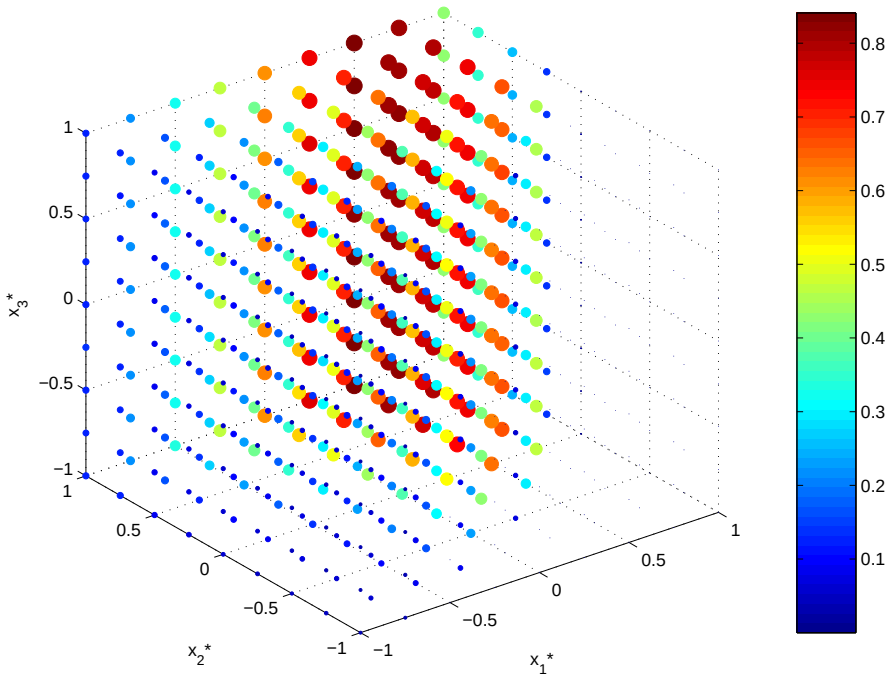

Figure 5.1: Scatter plot of $P(\mathbf{y}^* \in \mathbf{R}|\mathbf{x}^*)$ for HPLC example

5.4 Discussion

In this chapter the results of the model-robust process optimization using Bayesian model averaging were extended to multiple response systems. It was seen that for cases where there is no correlation in the error terms between the responses, the joint posterior probability of conformance for a given model set M_i as well as the joint model posteriors for all M_i could be obtained simply by taking the product of the marginal probabilities for the individual responses, where each of the marginal probabilities is obtained from the results for single response systems presented in chapter 3. For the SMR case, where the error terms

between the responses are correlated and the responses have identical regressors within a given model set M_i, the model posteriors are given by equation (5.15) and the posterior predictive distribution for a given M_i is given by a multivariate **T**-distribution as shown in equation (5.18). For the SUR case, suggestions are given in chapter 8 to implement the model-averaged process optimization.

In the last three chapters on process optimization, it was assumed that the specification limits for the response(s) are given, and the setting of the control factors that maximizes the posterior probability of conformance of the response(s) to these specification limits is obtained. In the next chapter, we solve the inverse problem, one of tolerance control, where a desired minimum posterior probability of conformance for each of the responses is given, and the optimization is formulated to find the setting of the control factors that gives the smallest tolerance or specification region while meeting the minimum posterior probability of conformance for the responses.

Chapter 6

A Bayesian Method for Robust Tolerance Control and Parameter Design

6.1 Introduction: Tolerance Control

In engineering design, the limits defining the acceptable quality of a product are called tolerance or specification limits. The problem of setting these limits based on different criteria is known as tolerance control. This chapter addresses the problem of setting tolerance limits on one or more quality characteristics that depend on controllable factors, $(x_1...x_k)$. The approach presented is Bayesian. Bayesian methods have been used in the literature [56, 46] to identify the settings of $(x_1...x_k)$ that maximize the probability of conformance of the responses or quality characteristics to a pre-defined given tolerance region. However, in this chapter we address the inverse problem: that of identifying a tolerance region such

that the probability of conformance of the response(s) to the region is at least equal to a user-defined value ϕ.

In tolerance control problems, it is also of interest to identify the setting of the control factors that gives the smallest such tolerance region. For example, suppose the design engineers in a company design a part that should be machined with thickness between *3mm* and *5mm*, and suppose that there are two controllable factors in the machine, cutting speed and pressure, that the operator can adjust to get the required thickness. Using the Bayesian optimization approaches in the literature, it is possible to find the settings of these two controllable factors that maximize the posterior probability that the part will have a thickness that is within the specification (tolerance) limits [56]. However, it is possible that at these settings the value of the posterior probability is quite low, say 0.6, which means that even at the best operating setting only 60% of the manufactured parts will meet the tolerances that have been set. This is a common problem in tolerancing and can be overcome if the designer transfers tolerance requirements to the manufacturing plant, while keeping in mind the limitations of the machinery and making the most of the design flexibility [32]. Thus in the previous example, the designer could prefer to adjust the design such that it is possible to set a different tolerance limit on the part that gives a higher posterior probability of obtaining conforming parts. It may be of interest, for example, to determine if there is another setting of the controllable factors where there is a high posterior probability of conformance to thickness between *5mm* and *6mm*, say a 90% probability. If this is true, at this new setting not only is the probability of conformance higher, but the tolerance region is also smaller, thus giving a lower variation in the conforming parts.

In the standard approaches to statistical tolerancing [26], if Y is a quality characteristic

with probability distribution P_Y^{θ}, where θ are the parameters, and if a sample $(y_1...y_n)$ of n independent observations is available, two common statistical methods have been used for constructing a tolerance region A. These are defined as,

1. the α-expectation tolerance region, given by:

$$E[C(A)] = \alpha, \tag{6.1}$$

2. the α-content tolerance region at confidence level γ, given by:

$$p[C(A) \geq \alpha] = \gamma, \tag{6.2}$$

where $C(A)$ is the coverage of the region A. These definitions are applicable for both classical or frequentist and Bayesian approaches. Both approaches are discussed in Guttman [25, 26]. However, they do not consider the problem of tolerance control in conjunction with regression, where the response depends on the settings of control factors. Also, the methods in the literature do not address the problem of finding the smallest tolerance region that satisfies one of the two criteria shown in equations (6.1) and (6.2).

The idea of a robust tolerance design was originally proposed by Taguchi [67]. Taguchi recommended tolerance design as the stage in quality control that follows parameter design. Parameter design is used to fit regression models to data and identify levels of the controllable factors that give the required mean and variation of the fitted response models. Singpurwalla [64] provides a Bayesian framework to approaching Taguchi's idea of parameter and tolerance design. Taguchi's robust tolerance design is used to adjust the tolerances of the controllable factors that have a large influence on the response(s). However, this does not address the problem of setting tolerance or specification limits on the

responses themselves. Taguchi's idea is related to what is called "transmission of errors", where variation in the controllable factors causes additional variation in the responses. In this chapter, we first address the problem of setting tolerances on the responses assuming that the controllable factors can be set to fixed desired settings. The effect of transmission of errors because of variation in the controllable factors is addressed in the discussion in section 6.4

The remainder of this chapter is organized as follows. The next section describes the proposed method for constructing tolerance limits for systems with a single response or quality characteristic. Section 6.3 discusses the multiple response case. A summary of the approach is given in the discussion section.

6.2 Single Response Systems

This section focusses on processes with a single response or quality characteristic of interest, y. It is assumed that this response depends on k controllable factors, $x_1 ... x_k$. It is also assumed that we have data from an experiment with n runs from which we can fit a model to the response of the form

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \tag{6.3}$$

where $\mathbf{x}$ is the $(p \times 1)$ vector of regressors that are functions of the k controllable factors, $\boldsymbol{\beta}$ is the $(p \times 1)$ vector of model parameters and ϵ is the error term which is assumed to be normally distributed, $N(0, \sigma^2)$. Denote the design matrix from the experiment by an $(n \times p)$ matrix $\mathbf{X}$ and the vector of observed responses from the experiment by an $(n \times 1)$ vector $\mathbf{y}$.

6.2.1 Bayesian Predictive Density

The posterior predictive density of a future response vector y^* at a given setting of the model regressors $\mathbf{x}^*$ for the given data $\mathbf{y}$ is defined as [57]:

$$p(y^*|\mathbf{x}^*, \mathbf{y}) = \int_{\sigma^2} \int_{\boldsymbol{\beta}} p(y^*|\mathbf{x}^*, \mathbf{y}, \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) \, d\boldsymbol{\beta} \, d\sigma^2, \tag{6.4}$$

where $p(y^*|\mathbf{x}^*, \mathbf{y}, \boldsymbol{\beta}, \sigma^2)$ is the likelihood function, and $p(\boldsymbol{\beta}, \sigma^2|\mathbf{y})$ is the posterior distribution of the model parameters. It is noted that the uncertainty in the model parameters is naturally accounted for by considering $\boldsymbol{\beta}$ and σ^2 to be random variables and evaluating their posterior distributions using Bayes' theorem:

$$p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}, \sigma^2), \tag{6.5}$$

where $p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2)$ is the likelihood function, and $p(\boldsymbol{\beta}, \sigma^2)$ is the joint prior distribution of the model parameters. For the system described earlier, under a diffuse prior given by,

$$p(\boldsymbol{\beta}) \propto constant, \tag{6.6}$$

$$p(\sigma^2) \propto \frac{1}{\sigma^2}, \tag{6.7}$$

and

$$p(\boldsymbol{\beta}, \sigma^2) = p(\boldsymbol{\beta}) p(\sigma^2), \tag{6.8}$$

the posterior predictive density is given by a t-distribution (see Press [57]). That is,

$$y^*|\mathbf{x}^*, \mathbf{y} \sim t_\nu(\mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}, \hat{\sigma}^2(1 + \mathbf{x}^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^*)), \tag{6.9}$$

where $\nu = n - p$,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{6.10}$$

and

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{n - p}. \tag{6.11}$$

As the posterior distribution of the response is a t-distribution, the posterior mean of the response given $\mathbf{x}^*$ is

$$E[y^*|\mathbf{x}^*, \mathbf{y}] = \mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}, \tag{6.12}$$

and the posterior variance of the response given $\mathbf{x}^*$ is

$$Var[y^*|\mathbf{x}^*, \mathbf{y}] = \frac{\nu}{\nu - 2}\hat{\sigma}^2(1 + \mathbf{x}^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^*). \tag{6.13}$$

6.2.2 Optimization for Tolerance Control

The objective of the tolerance control problem is to find the setting of the controllable factors, $x_1^*...x_k^*$, that gives the smallest interval $[l, u]$ such that the posterior probability of conformance, $p(y^* \in [l, u]|\mathbf{x}^*, \mathbf{y})$, is at least ϕ, where ϕ is decided by the process engineer or the designer. In addition, there may be constraints imposed on the ranges of $\mathbf{x}^*$, l, and u. In mathematical notation, the problem is formulated as:

$$\begin{aligned}
&\min_{x_1^*...x_k^*} \; u - l \\
&s.t., \\
&p(l \le y^* \le u|\mathbf{x}^*, \mathbf{y}) \;\; &\ge \;\; &\phi \\
&l \;\; &\ge \;\; &B_l \\
&u \;\; &\le \;\; &B_u \\
&x_1^*...x_k^* \;\; &\in \;\; &\Re,
\end{aligned}$$

where B_l and B_u are, respectively, a lower bound on l and an upper bound on u, determined by the user. For a given $\mathbf{x}^*$, we have from equation (6.9) that $y^*|\mathbf{x}^*, \mathbf{y} \sim t_\nu(\mu_{y^*}, \sigma^2_{y^*})$, where $\mu_{y^*} = \mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}$, and $\sigma^2_{y^*} = \hat{\sigma}^2(1 + \mathbf{x}^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^*)$. Let F_t and f_t be, respectively, the c.d.f. and the p.d.f. of this posterior t-distribution (note that F_t and f_t depend on $\mathbf{x}^*$). Based on the constraints on the bounds, B_l and B_u, a given setting $\mathbf{x}^*$ is infeasible if

$$F_t(B_u) < \phi, \tag{6.14}$$

or if

$$F_t(B_l) > 1 - \phi. \tag{6.15}$$

If the above inequalities are not true, then $P(l \leq y^* \leq u|\mathbf{x}^*, \mathbf{y}) \geq \phi$ can be satisfied, and from figure 6.1, it is evident that the smallest interval $[l, u]$ that encloses an area at least equal to ϕ should be centered about the mean μ_{y^*} and can be computed by

$$l = F_t^{-1}\left(\frac{1-\phi}{2}\right), \tag{6.16}$$

$$u = 2\mu_{y^*} - l. \tag{6.17}$$

However, the values of l and u computed using equations (6.16) and (6.17) need to satisfy also the constraints $l \geq B_l$ and $u \leq B_u$. Thus, when using equations (6.16) and (6.17) to obtain l and u, there are four cases as shown in figure 6.2. These are:

1. $\underline{l \geq B_l, u \leq B_u}$: This case satisfies all the constraints. Here, the values of l and U from equations (6.16) and (6.17) give the smallest interval $[l, u]$ for the given $\mathbf{x}^*$.

2. $\underline{l \geq B_l, u > B_u}$: Here, the value of u from (6.17) does not satisfy the constraint $u \leq B_u$. So to find the smallest interval, we set $u = B_u$ and $l = F_t^{-1}[F_t(B_u) - \phi]$. Note

that because of the feasibility check in equation (6.14), $F_t(B_u) - \phi$ is never less than zero. Now if the new value of $l \geq B_l$, then $[l = F_t^{-1}(F(B_u) - \phi), u = B_u]$ is the smallest interval satisfying all constraints. Otherwise, if the new $l < B_l$, then the problem for the given $\mathbf{x}^*$ is infeasible.

3. $\underline{l < B_l, u \leq B_u}$: Here, the value of l from (6.16) does not satisfy the constraint $l \geq B_l$. So, to get the smallest interval, we set $l = B_l$ and $u = F_t^{-1}[F_t(B_l) + \phi]$. Here, because of the feasibility check in equation (6.15), $F_t(B_l) + \phi$ is never greater than one. Now if the new value of $u \leq B_u$, then $[l = B_l, u = F_t^{-1}(F_t(B_l) + \phi)]$ is the smallest interval satisfying all constraints. Otherwise, if the new $u > B_u$, then the problem for the given $\mathbf{x}^*$ is infeasible.

4. $\underline{l < B_l, u > B_u}$: In this case, the problem for the given $\mathbf{x}^*$ is infeasible, as no interval $[l, u]$ that satisfies the constraints $l \geq B_l$ and $u \leq B_u$ will contain an area under the p.d.f. curve at least equal to ϕ.

The analysis of these cases provides a solution algorithm for the tolerance control problem. This is summarized in the algorithm in figure 6.3 to compute the values of l and u that give the minimum bound, $u - l$, for a given $\mathbf{x}^*$. The algorithm is used within a nonlinear optimization program that searches in the space of $(x_1^*...x_k^*)$ to find the setting that gives the smallest $(u - l)$ under the given constraints. Note that if there are no constraints on l and u, then the optimization problem reduces to finding the value of $\mathbf{x}^*$ that gives the minimum posterior variance. The optimization problem then can be reformulated as:

$$\min_{x_1^*...x_k^*} \sigma_{y^*}^2 \qquad s.t., \qquad x_1^*...x_k^* \in \Re.$$

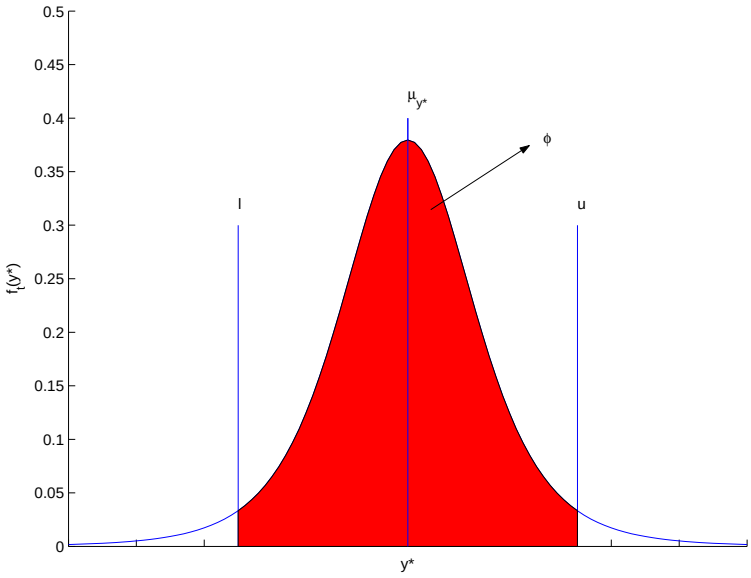

Figure 6.1: Sample t-distribution

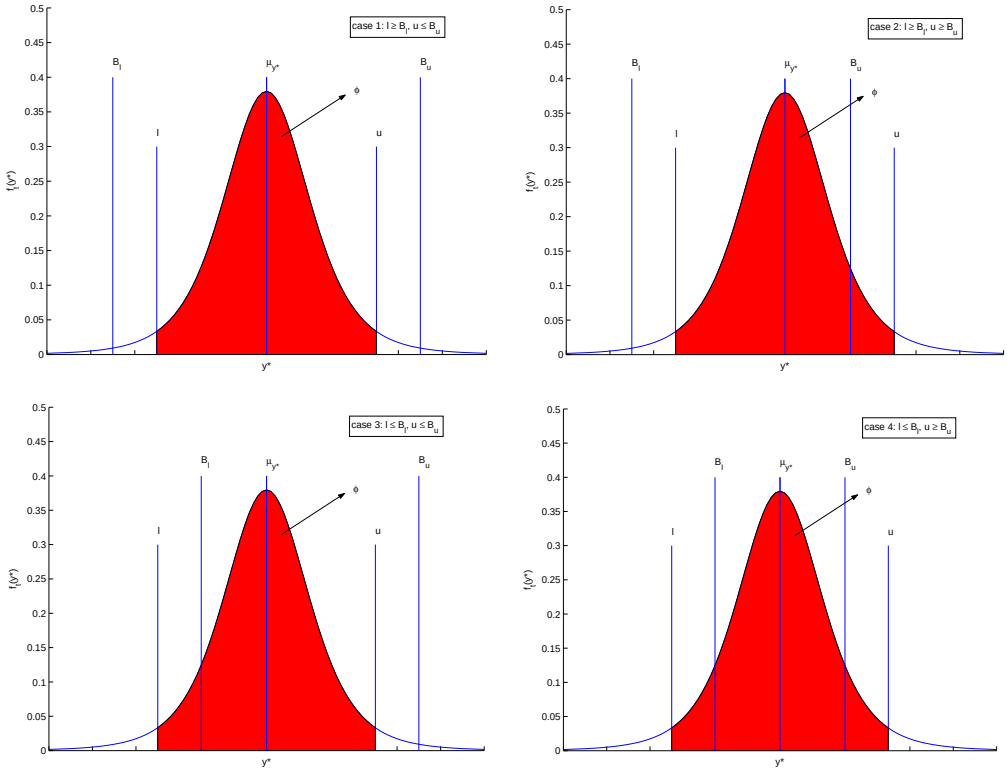

Figure 6.2: Four possible cases for constraints on the bounds

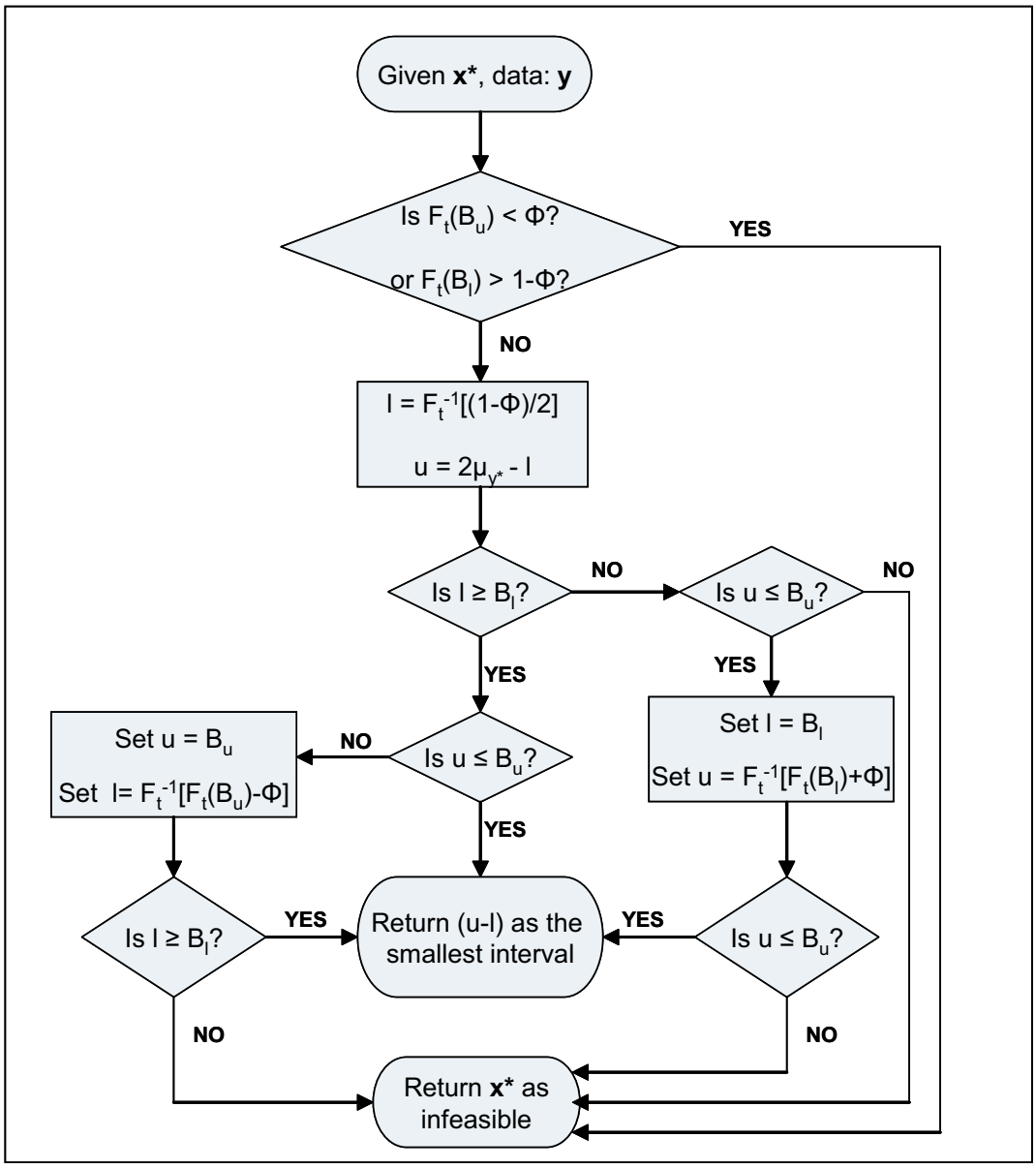

Figure 6.3: Algorithm to determine smallest interval $[l, u]$ for a given $\mathbf{x}^*$ under given constraints

In this case, the smallest interval $[l, u]$ can be found at any $\mathbf{x}^*$ using equations (6.16) and (6.17). Note that it is frequently the case where several settings $(x_1^*...x_k^*)$ satisfy the constraints of the tolerance problem, and this set of feasible solutions need not be convex or even connected. As the feasible regions maybe non-convex and even disconnected, it is recommended to run the nonlinear search algorithm in the space of $(x_1^*...x_k^*)$ using multiple starting points to avoid local optimums.

The examples below illustrate the method. The first example illustrates how the set of feasible solutions $(x_1^*...x_k^*)$ may be formed by disconnected subsets. The second example in section 6.2.4 presents a real manufacturing experiment.

6.2.3 Example 1: Two controllable factors

In this example, the data used is taken from Khuri and Cornell [38], and is shown in table 6.1. The goal of the experiment was to investigate the effect of two controllable factors on a single response. As the response in this example is yield, a higher value of the response is desired. In the table, the factors are given in coded variables determined by a rotatable central composite design (CCD). The data shows observed responses for two replicates of each treatment combination. Based on the two replicates of the experiment, the model fitted to the data using the parameter estimates shown in equation (6.10) is given by,

$$\hat{y} = 16.3647 + 1.6753x_1 + 2.7651x_2 - 0.3337x_1x_2 - 2.4637x_1^2 - 1.9310x_2^2. \qquad (6.18)$$

Control Factors		Response	
x_1	x_2	Replicate 1	Replicate 2
-1	-1	7.52	8.12
1	-1	12.37	11.84
-1	1	13.55	12.35
1	1	16.48	15.32
-1.414	0	8.63	9.44
1.414	0	14.22	12.57
0	-1.414	7.90	7.33
0	1.414	16.49	17.40
0	0	15.73	17.00

Table 6.1: Design and experimental data [38] for example in section 6.2.3

Note that the models are different when only one of the replicates is used. The model using only the first replicate is given by

$$\hat{y} = 15.7296 + 1.9608x_1 + 2.7862x_2 - 0.4800x_1x_2 - 1.9851x_1^2 - 1.6000x_2^2, \tag{6.19}$$

while the model using only the second replicate is given by

$$\hat{y} = 16.9999 + 1.3897x_1 + 2.7440x_2 - 0.1875x_1x_2 - 2.9423x_1^2 - 2.2621x_2^2. \tag{6.20}$$

As the treatment combinations for both the replicates are identical, it is noted that the coefficients in equation (6.18) are the average of the respective coefficients in equations (6.19) and (6.20). It is assumed that the goal of the experiment is to find the value of the controllable factors in the interval $[-1, 1]$, while at the same time setting tolerance or specification limits $[l, u]$ on the response with the desired probability of conformance. As higher values of the response are desirable in this example, it is important to set a lower bound B_l on the value of l. In addition, since the optimization finds the smallest interval $u - l$, operating at the optimal set point $(x_1^*...x_k^*)$ not only gives the desired probability of conformance, but also the least variation in the response under the given constraints.

The dots in figure 6.4 represent all feasible $\mathbf{x}^*$ computed at points on a grid spaced 0.05 apart in the region $\{x_1^* \in [-1, 1], x_2^* \in [-1, 1]\}$ using the algorithm shown in figure 6.3. Figure 6.4 shows four cases based on different values of B_l, B_u and ϕ. The figure is plotted using the data from both replicates and the corresponding model in equation (6.18). Figure 6.5 shows the feasible $\mathbf{x}^*$ for two of those cases using just the data from replicate 1 and the corresponding model in equation (6.19), and figure 6.6 shows the feasible $\mathbf{x}^*$ for the same two cases as in figure 6.5 using just the data from replicate 2 and the corresponding model in equation (6.20). It can be seen from these figures that:

1. The feasible region need not be convex as shown in figure 6.4 for the case where $B_l = 13$, $B_u = 20$ and $\phi = 0.99$. It is noted that depending on the data and the constraints, the feasible region may even be disconnected.

2. For the case where $B_l = 13$, $B_u = 20$ and $\phi = 0.99$, the feasible set is empty when only replicate 1 (figure 6.5) is used or when only replicate 2 (figure 6.6) is used. However, when both replicates are used, although the posterior mean of y^* at a given $\mathbf{x}^*$ is the average of the posterior means of the two replicates, the posterior standard deviation of y^* is less than that of either of the replicates. Therefore, the feasible region in figure 6.4 is not empty when both replicates are used for $B_l = 13$, $B_u = 20$ and $\phi = 0.99$.

3. As expected in all cases, for the same values of B_l and B_u the feasible region is larger as the constraint on the value of the probability of conformance ϕ decreases.

The optimization is performed using both replicates of the data and is presented in table 6.2 for different combinations of B_l, B_u and ϕ. Note that the value of $\mathbf{x}^*$ that gives the minimum posterior variance for the response is obtained by solving the optimization problem without any constraints on l and u, i.e., by setting $B_l = -\infty$ and $B_u = \infty$. In this example, when the posterior variance of the response is minimized, the smallest interval $(u - l)$ obtained is 6.8967 for $\phi = 0.99$, 4.9194 for $\phi = 0.95$, and 4.0241 for $\phi = 0.90$. It can be seen from table 6.2 that the smallest interval obtained in all the cases is equal to or very close to the smallest possible interval without any constraints on l and u, for the respective values of ϕ. This is because the region with the desired high posterior mean also has the lowest posterior variance for the response, as can be seen in the plots shown in figure 6.7.

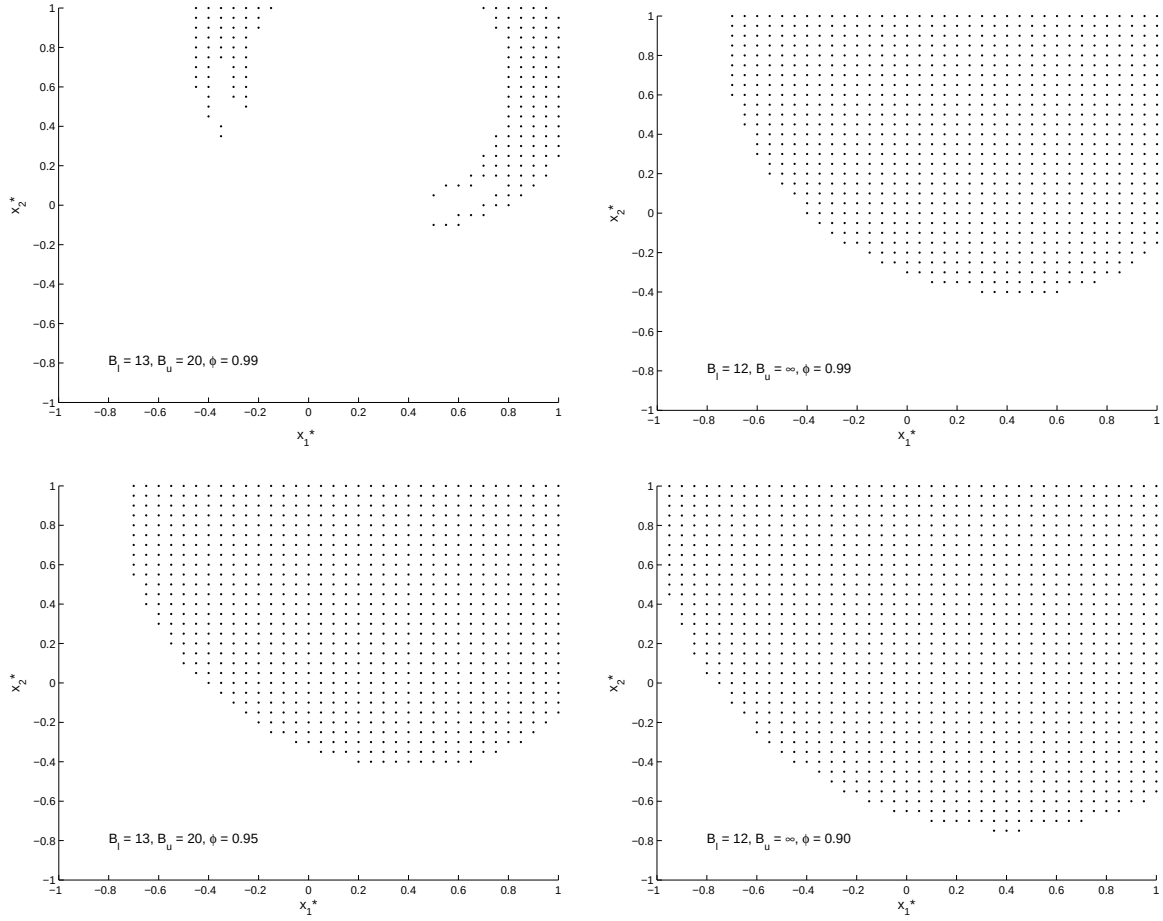

Figure 6.4: Feasible region for different constraints using both replicates

However, this need not always be the case as demonstrated in the next example.

6.2.4 Example 2: Three controllable factors

This example uses machining data from Taraman [69] and is presented in table 6.3. There are 3 controllable factors - cutting speed (x_1), feed (x_2) and depth of cut (x_3), and three responses - surface roughness (R), tool life (T) and cutting force (F). Table 6.3 gives the values of the controllable factors in $[-1, 1]$ coded form based on a central composite design. The table also shows the logarithm of the observed responses, which are used for modelling

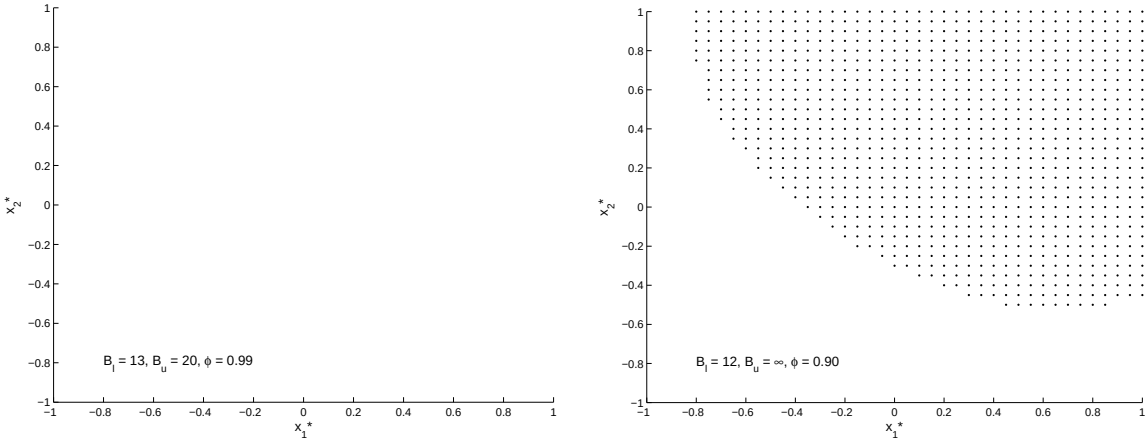

Figure 6.5: Feasible region under different constraints using replicate 1

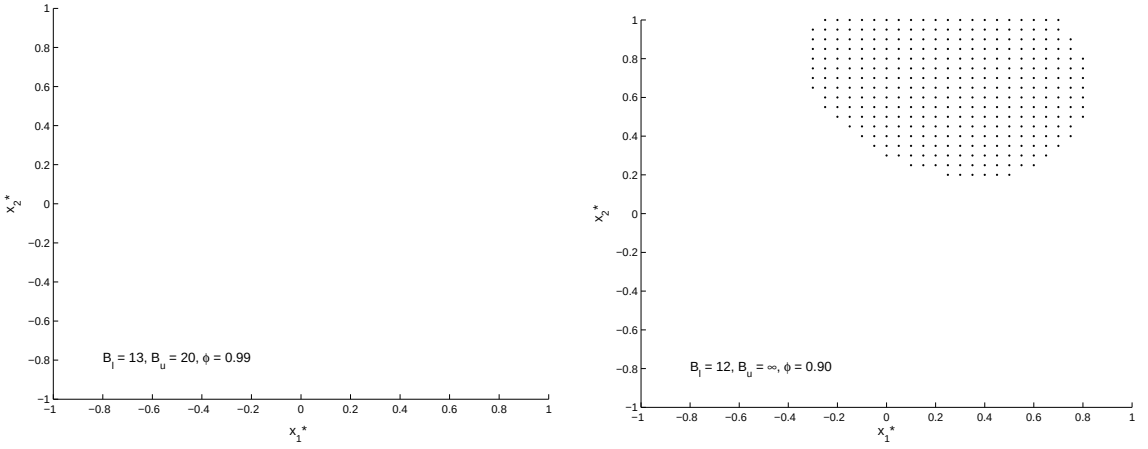

Figure 6.6: Feasible region under different constraints using replicate 2

B_l	B_u	ϕ	optimal x_1^*	optimal x_2^*	optimal $(u-l)$	l	u
13	20	0.99	0.9365	0.6291	6.8967	13.1032	19.9999
12	∞	0.99	0.7661	0.8281	6.8967	13.5077	20.4044
14	22	0.99	0.3596	0.9878	6.9218	14.0000	20.9218
13	20	0.95	0.7722	0.8224	4.9194	14.4857	19.4051
12	∞	0.95	-0.8022	0.7931	4.9194	12.1663	17.0857
14	22	0.95	0.8308	0.7631	4.9194	14.3702	19.2896
13	20	0.90	0.8269	0.7674	4.0241	14.8264	18.8505
12	∞	0.90	-0.8048	0.7904	4.0241	12.6002	16.6243
14	22	0.90	0.6204	0.9422	4.0241	15.1397	19.1639

Table 6.2: Optimization results for constructing tolerances for example in section 6.2.3

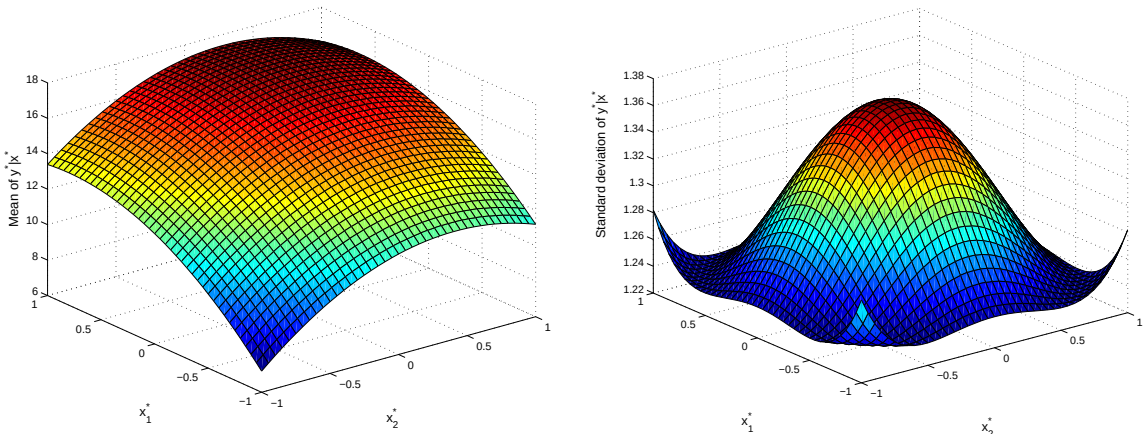

Figure 6.7: Posterior mean and standard deviation of the response for example in section 6.2.3

because it is expected, based on prior knowledge of the process [69], that the log scaled response is better suited to fit a linear statistical model of the form shown in equation (6.3). In this section the proposed approach for tolerance control is demonstrated for the response, tool life (T). Based on the given data, the model fit to the log of the response, $y = log(T)$, is

$$\hat{y} = 3.5009 - 0.3031x_1 - 0.0922x_2 - 0.0915x_3 + 0.0483x_1^2 + 0.0416x_2^2 + 0.0682x_3^2. \quad (6.21)$$

Based on the above model, the optimal tolerance interval can be computed for any given value of B_l , B_u and ϕ. In this example, when the posterior variance of the response is minimized, the smallest interval $(u - l)$ obtained is 0.8297 for $\phi = 0.99$, 0.6040 for $\phi = 0.95$, and 0.4980 for $\phi = 0.90$. The optimization results for a few combinations of these constraints are given in table 6.4. As the optimization is performed using the log of the tool life as the response, the table also shows the results transformed back into the original variable. Thus, for example, with a constraint on l of $B_l = 40$ and on u of $B_u = 100$, the smallest tolerance interval that can be set is a tool life of $[40, 73.46]$ minutes with a 95% conformance. It can also be seen from table 6.4 that not all combinations are feasible. For example, with a value of $B_l = 45$ and no constraint on u for tool life, there is no feasible solution at 99% conformance, but there are feasible solutions at 95% or lower conformance. It is noted that unlike in the previous example the smallest interval $(u - l)$ here is different for different values of B_l and B_u, for the same value of ϕ. This is because the region with the desired high posterior mean does not provide low posterior variance. Hence, as the constraints on l and u get tighter, the optimal interval $(u - l)$ is larger. This can be seen in the scatter plots of the posterior mean and variance shown in figure 6.8.

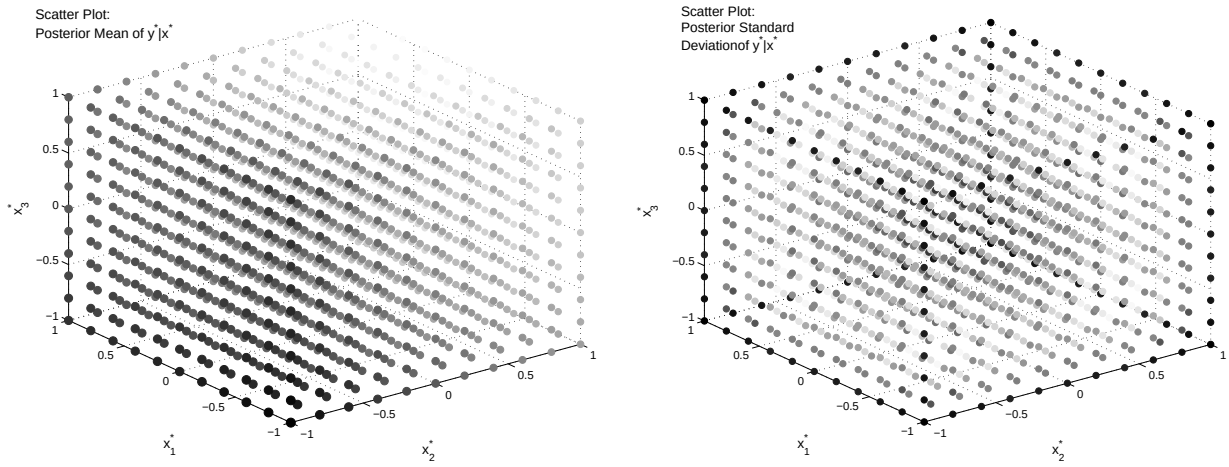

Figure 6.8: Posterior mean and standard deviation of the response for example in section 6.2.4

In the figure, the darker circles show regions with high magnitudes for both the posterior mean and standard deviation. In the plot the posterior mean is high at the corner point $(-1, -1, -1)$, but the posterior standard deviation is also high in this region. Thus as the constraint B_l on the lower bound is increased, the optimal solution moves closer to the corner point, causing an increase in the posterior standard deviation and consequently resulting in a wider tolerance interval. Using the results from this example, it is possible to choose a setting (x_1^*, x_2^*, x_3^*) where the tool life is obtained within the optimal tolerance intervals at the required probability of conformance. However, the performance of the other two responses, surface finish (R) and cutting force (F), at this setting is not known. The next section extends the Bayesian method for constructing simultaneous tolerance intervals to multiple response systems.

x_1	x_2	x_3	$R(CLA\,\mu in.)$	$T(min)$	$F(lbs)$	$log(R)$	$log(T)$	$log(F)$
-1	-1	-1	88	70	53	4.4773	4.2485	3.9703
1	-1	-1	76	29	48	4.3307	3.3673	3.8712
-1	1	-1	259	60	100	5.5568	4.0943	4.6052
1	1	-1	194	28	92	5.2679	3.3322	4.5218
-1	-1	1	105	64	76	4.6540	4.1589	4.3307
1	-1	1	82	32	74	4.4067	3.4657	4.3041
-1	1	1	270	44	155	5.5984	3.7842	5.0434
1	1	1	250	24	150	5.5215	3.1781	5.0106
0	0	0	123	35	82	4.8122	3.5553	4.4067
0	0	0	136	31	85	4.9127	3.4340	4.4427
0	0	0	130	38	83	4.8675	3.6376	4.4188
0	0	0	121	35	85	4.7958	3.5553	4.4427
-1.414	0	0	159	52	88	5.0689	3.9512	4.4773
1.414	0	0	115	23	80	4.7449	3.1355	4.3820
0	-1.414	0	77	40	50	4.3438	3.6889	3.9120
0	1.414	0	324	28	129	5.7807	3.3322	4.8598
0	0	-1.414	114	46	68	4.7362	3.8286	4.2195
0	0	1.414	215	33	124	5.3706	3.4965	4.8203
-1.414	0	0	139	46	87	4.9345	3.8286	4.4659
1.414	0	0	111	27	78	4.7095	3.2958	4.3567
0	-1.414	0	61	37	49	4.1109	3.6109	3.8918
0	1.414	0	340	34	130	5.8289	3.5264	4.8675
0	0	-1.414	128	41	71	4.8520	3.7136	4.2627
0	0	1.414	232	28	123	5.4467	3.3322	4.8122

Table 6.3: Design and experimental data for machining example in section 6.2.4 [69]

B_l	B_u	ϕ	optimal x_1^*	optimal x_2^*	optimal x_3^*	optimal $(u-l)$	l	u
Transformed Response: log (Tool Life in min.)								
$log(45)$	∞	0.99	Infeasible					
$log(40)$	log(100)	0.99	-1.0000	-0.8471	-0.9385	0.8598	3.6889	4.5486
$log(45)$	∞	0.95	-1.0000	-0.8533	-0.9465	0.6264	3.8067	4.4331
$log(40)$	log(100)	0.95	-0.8687	-0.6983	-0.7361	0.6079	3.6889	4.2968
$log(45)$	∞	0.90	-0.9845	-0.7429	-0.8020	0.5071	3.8067	4.3138
$log(40)$	log(100)	0.90	-0.7669	-0.6668	-0.6874	0.4988	3.6889	4.1876
Original Response: Tool Life in min.								
45	∞	0.99	Infeasible					
40	100	0.99	-1.0000	-0.8471	-0.9385	54.4992	40.0008	94.5000
45	∞	0.95	-1.0000	-0.8533	-0.9465	39.1903	45.0017	84.1920
40	100	0.95	-0.8687	-0.6983	-0.7361	33.4635	40.0008	73.4643
45	∞	0.90	-0.9845	-0.7429	-0.8020	29.7222	45.0017	74.7239
40	100	0.90	-0.7669	-0.6668	-0.6874	25.8637	40.0008	65.8645

Table 6.4: Optimization results for constructing tolerances for example in section 6.2.4

6.3 Multiple Response Systems

Here, it is assumed that there are q responses or quality characteristics of interest that depend on one or more of k controllable factors. It is assumed that each of the q responses is of the form

$$y_j = \mathbf{x}_j'\boldsymbol{\beta}_j + \epsilon_j, \tag{6.22}$$

where $\mathbf{x}_j$ is a $(p_j \times 1)$ vector of regressors, $\boldsymbol{\beta}_j$ is a $(p_j \times 1)$ vector of model parameters and ϵ_j is the error term for response y_j. Denote by $\boldsymbol{\Sigma}$ the $(q \times q)$ variance-covariance matrix of the error terms. Note that if all the responses have identical regressors, then $\mathbf{x}_j = \mathbf{x}$ for all j, and if the error terms ϵ_j are uncorrelated between the responses, then $\boldsymbol{\Sigma}$ is a diagonal matrix. There are four different ways to model multiple response systems based on the regressors present in the models for the individual responses and the correlation of the error terms between the responses, as summarized in table 5.1. The Bayesian posterior predictive density depends on how the multiple responses are modelled and is discussed below.

6.3.1 Bayesian Predictive Density

For the cases where the error term ϵ_j is uncorrelated between the responses, i.e., $\boldsymbol{\Sigma}$ is diagonal, each of the q responses can be modelled independently from the data, regardless of whether the responses have identical regressors or not. As in the single response case, it is assumed that there is data from an experiment with n runs. The observed responses from the experiment are denoted by $(n \times 1)$ vectors $\mathbf{y}_j$, where $j = \{1...q\}$, and the corresponding design matrices are denoted by $\mathbf{X}_j$. It is noted that $\mathbf{X}_j = \mathbf{X}$ for all j if the responses

are modelled with identical regressors. When $\mathbf{\Sigma}$ is diagonal, the joint posterior probability of conformance at a new setting $\{x_1^*...x_k^*\}$ for the q responses $\mathbf{y}^* = (y_1^*...y_q^*)$ is simply the product of the marginal posterior probabilities of conformance of the individual responses. Thus, given the data $\mathbf{Y} = (\mathbf{y}_1...\mathbf{y}_q)$,

$$p(\mathbf{y}^* \in V|\mathbf{x}_j^* \forall j, \mathbf{Y}) \equiv p(y_1^* \in [l_1, u_1], y_2^* \in [l_2, u_2]...y_q^* \in [l_q, u_q]|\mathbf{x}_j^* \forall j, \mathbf{Y}) \quad (6.23)$$

$$= \prod_{j=1}^{q} p(y_j^* \in [l_j, u_j]|\mathbf{x}_j^*, \mathbf{y}_j), \quad (6.24)$$

where V is the region enclosed by $[l_i, u_i] \forall i$. Therefore, for the diffuse priors described by equations (6.6), (6.7) and (6.8), each of the $p(y_j \in [l_j, u_j]|\mathbf{x}_j^*, \mathbf{y}_j)$ is obtained from the c.d.f. of the t-distribution shown in equation (6.9).

Equations (6.23) and (6.24) do not hold if the error terms are correlated, i.e., $\mathbf{\Sigma}$ is non-diagonal. In such cases, the responses can be modelled as either Standard Multivariate Regression (SMR) or Seemingly Unrelated Regression (SUR), where the former assumes that all the response models have the same set of regressors, i.e., $\mathbf{X}_j = \mathbf{X} \, \forall j$ and the latter assumes that each response model may have different regressors. For the SMR case, the joint posterior probability distribution under a diffuse prior is given by a multivariate $\mathbf{T}$-distribution [57]. That is, given the $(n \times p)$ design matrix $\mathbf{X}$, and the $(n \times q)$ response data matrix $\mathbf{Y}$, the posterior density at a future set of observations given by $(p \times 1)$ vector $\mathbf{x}^*$ is

$$\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y} \sim T_\nu^q(\mathbf{B}'\mathbf{x}^*, \mathbf{H}^{-1}), \quad (6.25)$$

where $\nu = n - p - q + 1$,

$$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \quad (6.26)$$

$$\mathbf{H} = \frac{\nu \mathbf{S^{-1}}}{1 + \mathbf{x}^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^{*}}, \tag{6.27}$$

and

$$\mathbf{S} = (\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB}). \tag{6.28}$$

For the SUR case, the posterior predictive density has to be computed by Gibbs sampling. Percy [53] shows how a sample of the posterior observation $\mathbf{y}^*$ can be obtained from its posterior distribution using Gibbs sampling.

6.3.2 Optimization

It is assumed that the objective is to find the setting $(x_1^*...x_k^*)$ that minimizes some given function, A, such that the posterior probability of conformance of each response, $p(y_i^* \in [l_i, u_i]|\mathbf{x}_i^*, \mathbf{y}_i)$ is at least ϕ_i, where the ϕ_i's are decided by the plant engineer or designer. Here also, each of the l_i's and u_i's may be constrained to lie within given bounds. For example, the objective function $A = \prod_i (u_i - l_i)$ finds the smallest (in terms of volume) q-dimensional cuboid that satisfies the given constraints. The function A could also be chosen such that different weights are given to the bounds on different responses. The

objective function in this case is formulated as,

$$\min_{x_1^*...x_k^*} A = \prod_{i=1}^{q}(u_i - l_i)$$

$$s.t.,$$

$$\begin{aligned}
p(l_1 \leq y_1^* \leq u_1 | \mathbf{x}_1^*, \mathbf{y}_1) &\geq \phi_1 \\
p(l_2 \leq y_2^* \leq u_2 | \mathbf{x}_2^*, \mathbf{y}_2) &\geq \phi_2 \\
\vdots &\geq \vdots \\
p(l_q \leq y_q^* \leq u_q | \mathbf{x}_q^*, \mathbf{y}_q) &\geq \phi_q \\
l_i &\geq B_{l_i}, \ \forall i = \{1...q\} \\
u_i &\leq B_{u_i}, \ \forall i = \{1...q\} \\
x_1^*...x_k^* &\in \Re.
\end{aligned}$$

For a given $\mathbf{x}_i^*$, if the error terms are uncorrelated, the smallest interval $[l_i, u_i]$ for each response y_i can be found using the marginal posterior distribution $p(y_i^*|\mathbf{x}_i^*, \mathbf{y}_i)$ and the algorithm previously shown in figure 6.3. As all the $(u_i - l_i) > 0$, this also gives the smallest value of $A = \prod_i (u_i - l_i)$ for that $\mathbf{x}^*$. This is true for any A that is an increasing function of each $(u_i - l_i)$. The algorithm for finding the smallest A for a given $\mathbf{x}_i^*$ is used within a nonlinear optimization program that searches within the space of the feasible $(x_1^*...x_k^*)$ to find the setting that gives the smallest value of A. The methodology is illustrated by an example in the next section.

If the error terms are correlated, then a nonlinear optimization program that searches in the feasible space of both $(x_1^*...x_k^*)$ and $[l, u]$ must be used to solve the optimization problem for both the SMR and the SUR models. It is noted that depending on the size of

the problem, the optimization could be tedious especially in the SUR case.

6.3.3 Multiple response example

This example uses the machining data [69] shown in table 6.3 to simultaneously set tolerances on all the three responses. Once again, the log of the responses are used for modelling, where $y_1 = log(R)$, $y_2 = log(T)$, and $y_3 = log(F)$. The response models obtained are,

$$\hat{y}_1 = 4.8773 - 0.0960x_1 + 0.5336x_2 + 0.1429x_3 - 0.0216x_1^2 + 0.0543x_2^2 + 0.0969x_3^2, \quad (6.29)$$

$$\hat{y}_2 = 3.5009 - 0.3031x_1 - 0.0922x_2 - 0.0915x_3 + 0.0483x_1^2 + 0.0416x_2^2 + 0.0682x_3^2, \quad (6.30)$$

$$\hat{y}_3 = 4.4260 - 0.0332x_1 + 0.3391x_2 + 0.2092x_3 - 0.0019x_1^2 - 0.0208x_2^2 + 0.0522x_3^2. \quad (6.31)$$

Figure 6.9 shows the feasible $\mathbf{x}^*$ plotted on a grid spaced 0.1 apart in the region $\{x_1^* \in [-1, 1], x_2^* \in [-1, 1], x_3^* \in [-1, 1]\}$, assuming a desired probability of conformance $\phi_i = 0.80 \, \forall \, i$ for four different cases:

1. All the three responses have constraints on l and u. Here, surface roughness has a constraint $B_u = 120$, tool life has a constraint $B_l = 35$, and cutting force has a constraint $B_u = 60$. Note that all the constraints are one-sided as it is desired that the surface roughness be as low as possible, tool life be as high as possible and cutting force be as low as possible.

2. Only the surface roughness response is constrained with a value of $B_u = 120$.

3. Only the tool life response is constrained with a value of $B_l = 35$.

4. Only the cutting force response is constrained with a value of $B_u = 60$.

As expected, from figure 6.9, it is seen that the feasible region itself is much smaller when constraints are imposed on all the three responses simultaneously. Table 6.5 shows the results of the optimization to set simultaneous tolerance limits on all the responses for different combinations of the constraints. Here the objective function used is $A = \prod_i (u_i - l_i)$, where $[l_i, u_i]$ is the tolerance limit on response i in the logarithmic scale. Thus for example, in the table, at a desired value of probability of conformance $\phi_i = 0.9\ \forall i$, and with constraints $B_u = 110$ on surface roughness, $B_l = 45$ on tool life and $B_u = 90$ on cutting force, the optimal setting of the controllable factors obtained is $[-0.9309, -0.8317, -0.8001]$ where the tolerance limit on surface roughness is $[71.8, 110.0]$, on tool life is $[45.0, 74.8]$, and on cutting force is $[53.7, 58.4]$. Thus, using the methodology proposed it is possible to set simultaneous tolerances on multiple responses with a desired probability of conformance.

For the multiple response case, the solution to the optimization problem also depends on the choice of the user-defined function A. For example, suppose there are two responses. Suppose response 1 gives the smallest interval $(u_1^a - l_1^a) = 1$ at the setting $\mathbf{x}_a^*$ and the smallest interval $(u_1^b - l_1^b) = 2$ at the setting $\mathbf{x}_b^*$, and response 2 gives the smallest interval $(u_2^a - l_2^a) = 1.2$ at the setting $\mathbf{x}_a^*$ and the smallest interval $(u_2^b - l_2^b) = 0.5$ at the setting $\mathbf{x}_b^*$. Assuming $A = \prod_i (u_i - l_i)$, the value of A at $\mathbf{x}_a^*$ is 1.2 and at $\mathbf{x}_b^*$ is 1.0. In other words the solution at $\mathbf{x}_a^*$ is sub-optimal to the solution at $\mathbf{x}_b^*$. If instead, it is assumed that $A = \sum_i (u_i - l_i)$, then the value of A at $\mathbf{x}_a^*$ is 2.2 and at $\mathbf{x}_b^*$ is 2.5. Therefore, the solution at $\mathbf{x}_b^*$ is now sub-optimal to the solution at $\mathbf{x}_a^*$.

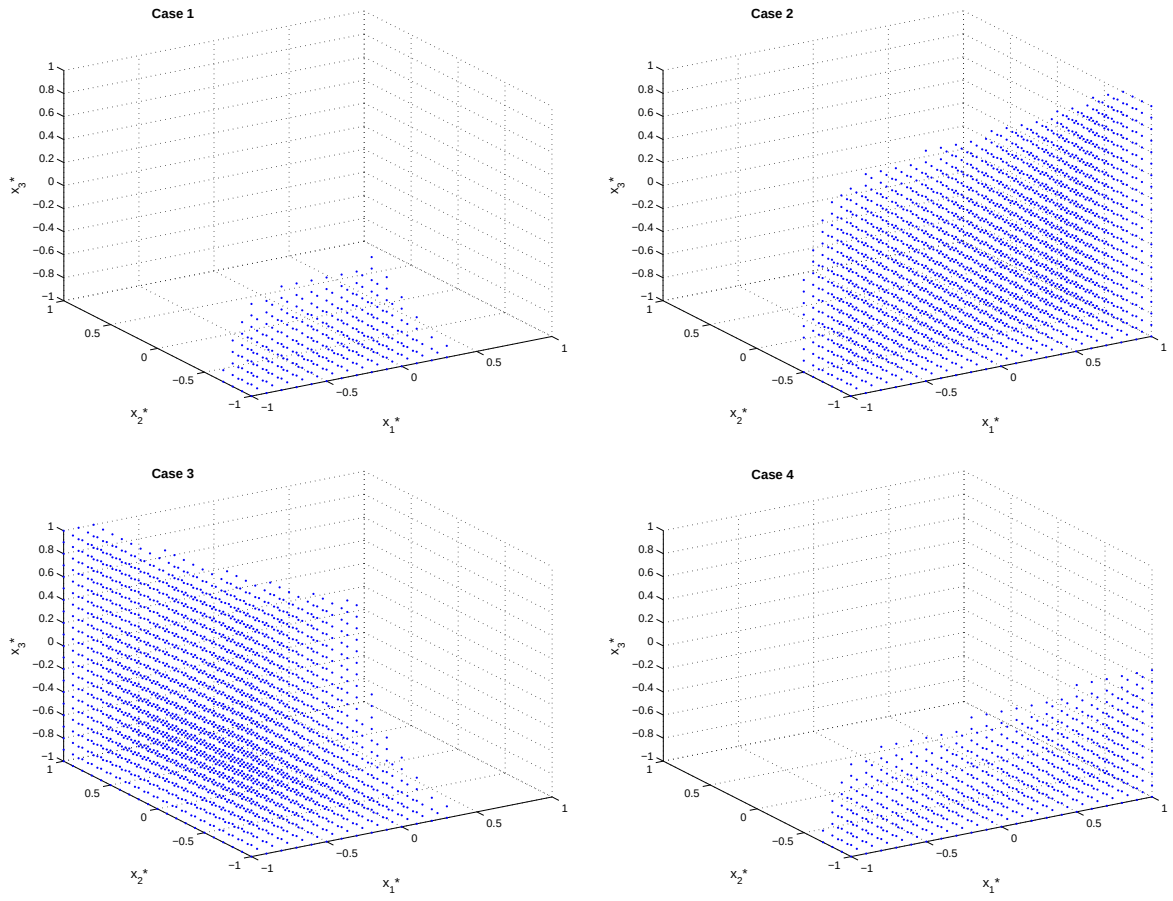

Figure 6.9: Feasible Regions for $\phi_i = 0.8 \ \forall \, i$ for the multiple response example in section 6.3.3

		R		T		F		optimal values				R		T		F	
	$\phi_i\ \forall i$	B_l	B_u	B_l	B_u	B_l	B_u	x_1^*	x_2^*	x_3^*	A	l_1	u_1	l_2	u_2	l_3	u_3
Transformed Responses: logarithmic scale																	
1	0.90	$-\infty$	$log(110)$	$log(45)$	∞	$-\infty$	$log(90)$	-0.931	-0.832	-0.800	0.0183	4.274	4.701	3.807	4.315	3.983	4.067
2	0.90	$-\infty$	$log(110)$	$log(45)$	∞	$-\infty$	∞	-0.945	-0.833	-0.773	0.0183	4.274	4.701	3.807	4.315	3.986	4.071
3	0.90	$-\infty$	$log(110)$	$-\infty$	∞	$-\infty$	$log(90)$	0.648	-0.647	-0.646	0.0174	4.222	4.641	3.241	3.739	4.021	4.104
4	0.90	$-\infty$	∞	$log(45)$	∞	$-\infty$	$log(90)$	-0.960	-0.732	-0.820	0.0183	4.323	4.749	3.807	4.315	4.018	4.103
5	0.75	$-\infty$	$log(110)$	$log(45)$	∞	$-\infty$	$log(90)$	-0.777	-0.729	-0.803	0.0057	4.382	4.671	3.807	4.150	4.029	4.087
6	0.75	$-\infty$	$log(100)$	$log(50)$	∞	$-\infty$	$log(60)$	-0.978	-0.900	-0.776	0.0060	4.311	4.605	3.912	4.263	3.975	4.033
7	0.75	$-\infty$	$log(100)$	$log(55)$	∞	$-\infty$	$log(60)$	infeasible									
8	0.75	$-\infty$	$log(90)$	$log(50)$	∞	$-\infty$	$log(60)$	infeasible									
Original Responses																	
1	0.90	0	110	45	∞	0	90	-0.931	-0.832	-0.800	N/A	71.8	110.0	45.0	74.8	53.7	58.4
2	0.90	0	110	45	∞	0	∞	-0.945	-0.833	-0.773	N/A	71.8	110.0	45.0	74.8	53.8	58.6
3	0.90	0	110	0	∞	0	90	0.648	-0.647	-0.646	N/A	68.2	103.6	25.5	42.0	55.7	60.6
4	0.90	0	∞	45	∞	0	90	-0.960	-0.732	-0.820	N/A	75.4	115.5	45.0	74.8	55.6	60.5
5	0.75	0	110	45	∞	0	90	-0.777	-0.729	-0.803	N/A	80.0	106.8	45.0	63.4	56.2	59.5
6	0.75	0	100	50	∞	0	60	-0.978	-0.900	-0.776	N/A	74.5	100.0	50.0	71.0	53.2	56.4
7	0.75	0	100	55	∞	0	60	infeasible									
8	0.75	0	90	50	∞	0	60	infeasible									

Table 6.5: Optimization results for setting tolerance regions for the multiple response example in section 6.3.3 where $A = \prod_i (u_i - l_i)$

6.4 Discussion

A Bayesian method was proposed to set tolerance limits on one or more responses to provide a given desired probability of conformance, and to determine at the same time the optimal settings of the control factors that the response(s) depend on. The method not only gives the tolerance interval that satisfies the constraints on the mean and the probability of conformance, but also finds the smallest such interval in case of single response systems, and the smallest value of a given function of the intervals for multiple response systems. This ensures that the quality characteristics or responses that adhere to the specification also have the smallest variation between themselves. The proposed method was illustrated by two examples for single responses systems and an example for a multiple response system. Some additional comments on the method are given below:

1. As the posterior predictive distribution of the response depends on the observed data, the solution to the tolerance control problem is also dependant on it. As seen in the example in section 6.2.3, the optimization problem had a non-empty feasible region when data from both the replicates was used, but had empty feasible regions when only one of the replicates was used. Thus, when additional experimental data are used in any of the examples presented, it is possible to get smaller tolerance regions that satisfy the given constraints.

2. In the previous sections, it was assumed that the controllable factors can be set to desired values by the user. However, in practice, there are errors associated with these settings. This error is also transmitted to the response which could result in a lower probability of conformance of the response ϕ_r to the calculated tolerance region,

than what was originally constrained by ϕ. The reduced probability of conformance ϕ_r can be estimated if the distribution of the errors in the control factors is known.

Suppose l^o and u^o are the calculated tolerance limits at the desired setting of the controllable factors $\mathbf{x^o} = (x_1^o ... x_k^o)$ as a result of the optimization, then assuming that the j^{th} controllable factor is normally distributed with a variance of σ_j^2 about the setting x_j^o, it is possible to represent the actual value of the controllable factor as $\tilde{x}_j = x_j^o + z_j$, where $z_j \sim N(0, \sigma_j^2)$. Thus, the actual settings of the controllable factors can be written as $\tilde{\mathbf{x}} = \mathbf{x^o} + \mathbf{z}$, where $\tilde{\mathbf{x}} = (\tilde{x}_1 ... \tilde{x}_k)$ and $\mathbf{z} = (z_1 ... z_k)$. The posterior distribution of the response for a single response system given the actual setting of the controllable factors $p(y^*|\tilde{\mathbf{x}}, \mathbf{y})$ is given by equation (6.9). If $\tilde{\mathbf{x}}$ is known then the reduced probability of conformance ϕ_r is given by $p(y^* \in [l^o, u^o]|\tilde{\mathbf{x}}, \mathbf{y})$, and can be computed from the c.d.f. of the t-distribution. However, as the actual value of $\tilde{\mathbf{x}}$ is not known because of the random component $\mathbf{z}$, the value of ϕ_r is computed by taking the expected value with respect to $\mathbf{z}$:

$$\phi_r = E_{\mathbf{z}} \left[p(y^* \in [l^o, u^o]|\tilde{\mathbf{x}}, \mathbf{y}) \right]. \tag{6.32}$$

In the above equation, the expected value can be estimated by simulation using the steps below:

(a) Choose sufficiently large N. Set $count = 1$.

(b) Generate $\mathbf{z}(count)$ by sampling from the distribution $z_j \sim N(0, \sigma_j^2)\ \forall j$.

(c) Set $\tilde{\mathbf{x}}(count) = \mathbf{x^o} + \mathbf{z}(count)$.

(d) Compute $p(y^* \in [l^o, u^o]|\tilde{\mathbf{x}}(count), \mathbf{y})$.

(e) Set $count = count + 1$. Repeat steps a, b and c until $count > N$.

(f) Estimate the expected value using the Weak Law of Large Numbers (WLLN):

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} [p(y^* \in [l^o, u^o] | \tilde{\mathbf{x}}(i), \mathbf{y})] = E_{\mathbf{z}} [p(y^* \in [l^o, u^o] | \tilde{\mathbf{x}}, \mathbf{y})] . \tag{6.33}$$

It is noted that if large variations in the controllable factors are expected, then the reduced probability of conformance ϕ_r may be considerably less than ϕ. In such cases, it is recommended that the user is conservative in choosing ϕ during the optimization to obtain the tolerance limits. For multiple response systems, the reduced probability of conformance can similarly be obtained for each response by taking the expected value of the marginal posterior probability of conformance $p(y_i^* \in [l_i^o, u_i^o] | \tilde{\mathbf{x}}_i, \mathbf{y}_i)$ with respect to the corresponding $\mathbf{z}_i$.

In all the chapters thus far in this dissertation, the setting of the control factors was obtained either by maximizing the posterior probability of conformance of the responses to their specification (or tolerance) limits or by minimizing some function of the tolerance intervals where the responses meet a given posterior probability of conformance. In the next chapter, a Bayesian method is presented where the setting of the control factors depends not only on the posterior distribution of the responses, but also on the posterior distribution of the utility function or satisfaction of the customer or decision maker, where the latter distribution depends on the outcome of the responses.

Chapter 7

A Bayesian Approach for Multiple Criteria Decision Making with application in "Design for Six Sigma"

7.1 Introduction

In practice, decision-making problems typically involve the consideration of two or more criteria that are often conflicting. These are referred to as Multiple Criteria Decision Making (MCDM) problems, where one has to take into account trade-offs between the conflicting criteria. An example from the manufacturing industry is in automobile design. For instance, the suspension in a sports car has to be designed considering trade-offs between a "sporty" feel for the driver and a ride that may be too bumpy. In the MCDM problem that is considered here, it is of interest to maximize the utility of the customer, which is an unknown function u of the future outcomes or responses, $y = (y_1 ... y_q)$. It is assumed

that these responses are in turn functions of variables $x = (x_1...x_p)$ that are under the control of the decision-maker and can be set by him/her to desired values subject to given constraints.

This chapter proposes a new algorithm to solve MCDM problems based on a Bayesian methodology that adopts a probabilistic approach to solving the MCDM problem. Section 7.3 presents MCDM methodology for the case of a single decision maker. This section discusses the application of the methodology to "design for six sigma" problems. Section 7.4 then extends the methodology to the case of multiple decision makers. Examples are provided in both section 7.3 and 7.4. A summary of the approach is discussed in section 7.5.

7.2 Bayesian Predictive Density

Single Response System

Consider a single response of interest y that depends on k controllable factors, $x_1...x_k$. Assume that we have data from an experiment with n runs from which the following model can be fitted to the response:

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \tag{7.1}$$

where $\mathbf{x}$ is the $(p \times 1)$ vector of regressors that are functions of the k controllable factors (i.e., $\mathbf{x}$ is in model form), $\boldsymbol{\beta}$ is the $(p \times 1)$ vector of model parameters and ϵ is the error term which is assumed to be normally distributed, $N(0, \sigma^2)$. Denote the design matrix from the experiment by the $(n \times p)$ matrix $\mathbf{X}$ and the vector of observed responses from the experiment by the $(n \times 1)$ vector $\mathbf{y}$.

The posterior predictive density of a future response vector, y^*, at a given setting of the model regressors, $\mathbf{x}^*$, for the given data, $\mathbf{y}$, is defined as (see [57]) given in equation (6.4). It is noted that the uncertainty in the model parameters is naturally accounted for in the expression by considering $\boldsymbol{\beta}$ and σ^2 to be random variables and evaluating their posterior distributions using Bayes' theorem. Under a diffuse prior given by equations (6.6), (6.7) and (6.8), the posterior predictive density is given by a t-distribution (see [57]). That is,

$$y^*|\mathbf{x}^*, \mathbf{y} \sim t_\nu(\mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}, \hat{\sigma}\sqrt{1 + \mathbf{x}^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^*}), \tag{7.2}$$

where $\nu = n - p$,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{7.3}$$

and,

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{n - p}. \tag{7.4}$$

From equation (7.2), if $[l, u]$ is the desired specification for the response it is possible to compute the posterior probability of conformance, $p(y^* \in [l, u]|\mathbf{x}^*)$, by using the c.d.f. of the t-distribution. The posterior mean and the posterior variance of the response at $\mathbf{x}^*$ are given by equations (6.12) and (6.13), respectively.

Multiple Response Systems

For multiple responses, each of the q responses is modelled as

$$y_i = \mathbf{x}_i'\boldsymbol{\beta}_i + \epsilon_i. \tag{7.5}$$

If the error term, ϵ_i is uncorrelated between the responses, then the responses can be modelled independently. In this case, the joint posterior probability of conformance for the

q responses is simply the product of the marginal posterior probabilities of conformance of the individual responses, as shown in equation (6.24), where each of the marginal posteriors is given by the t-distribution shown in equation (7.2). If the error terms are correlated, then the responses can be modelled as either a Standard Multivariate Regression (SMR) or a Seemingly Unrelated Regression (SUR) model, where the former assumes that all the response models have the same set of regressors and the latter assumes that each response model may have different regressors. For the SMR case, the joint posterior probability distribution under a diffuse prior is given by a multivariate **T**-distribution [57], as shown in equation (6.25). For the SUR case, the posterior predictive density has no closed form but can be computed numerically. In particular, Percy [53] shows how a the posterior predictive distribution of a new observation $\mathbf{y}^*$ can be approximated using Gibbs sampling.

7.3 Bayesian Method for MCDM

Suppose that there are q responses $(y_1...y_q)$, that depend on k controllable factors, $x_1...x_k$. It is assumed that data from an experiment with n runs is available from which a model can be fit to the responses of the form,

$$y_j = \mathbf{x}_j'\boldsymbol{\beta}_j + \epsilon_j, \quad \forall j \in \{1, ..., q\} \tag{7.6}$$

where for each response y_j, $\mathbf{x}_j$ is the $(p_j \times 1)$ vector of regressors that are functions of the k controllable factors, $\boldsymbol{\beta}_j$ is the $(p \times 1)$ vector of model parameters and ϵ_j is the error term. Depending on the model used for the multiple responses, it is possible to sample from the posterior distribution of the response as discussed in the previous section.

Suppose initially that there is a single customer whose utility function u depends on the

q responses $(y_1...y_q)$. It is assumed that data from a survey with m questions is available, where each question is a different combination of values of the q responses, and each answer is a score given by the customer on a numerical scale (e.g. 0 to 10) that indicates his/her preference to that combination. Based on the survey, it is now possible to fit a model:

$$u = \mathbf{y}_f'\boldsymbol{\gamma} + \epsilon, \tag{7.7}$$

where u is the customer's score or utility, $\mathbf{y}_f$ is the $(p_f \times 1)$ vector of regressors where each regressor is a function of the q responses $(y_1...y_q)$ (i.e., $\mathbf{y}_f$ is written in model form, thus its subscript is used to distinguish this from the data vector $\mathbf{y}$ of section 7.2), $\boldsymbol{\gamma}$ is the $(p_f \times 1)$ vector of model parameters, and ϵ is the error term, assumed $N(0, \sigma_f^2)$. Note that the responses $(y_1...y_q)$ are treated as regressors in (7.7). The combinations of the responses in the survey can be chosen based on DOE (design of experiments) techniques in order to get a good fit for the statistical model shown in equation (7.7). If $\mathbf{Y}_f$ is the $(m \times p_f)$ design matrix of the survey, and $\mathbf{u}$ is the vector of answers from the survey, then for a diffuse prior on the model parameters in equation (7.7) given by,

$$p(\boldsymbol{\gamma}) \propto constant, \tag{7.8}$$

$$p(\sigma_f^2) \propto \frac{1}{\sigma_f^2}, \tag{7.9}$$

and,

$$p(\boldsymbol{\gamma}, \sigma_f^2) = p(\boldsymbol{\gamma})p(\sigma_f^2), \tag{7.10}$$

the posterior distribution of the customer's utility u^* at a given value of the responses $(y_1^*...y_q^*)$ can be obtained from equation (7.2) and is given by:

$$u^*|\mathbf{y}_f^*, \mathbf{u} \sim t_{\nu_f}(\mu_{u^*}, \sigma_{u^*}^2), \tag{7.11}$$

where $\nu_f = m - p_f$, $\mu_{u^*} = \mathbf{y}_f^{*\prime}\hat{\boldsymbol{\gamma}}$, and $\sigma_{u^*}^2 = \hat{\sigma}_f^2(1 + \mathbf{y}_f^{*\prime}(\mathbf{Y}_f'\mathbf{Y}_f)^{-1}\mathbf{y}_f^*)$. Here,

$$\hat{\boldsymbol{\gamma}} = (\mathbf{Y}_f'\mathbf{Y}_f)^{-1}\mathbf{Y}_f'\mathbf{u}, \tag{7.12}$$

and,

$$\hat{\sigma}_f^2 = \frac{(\mathbf{u} - \mathbf{Y}_f\boldsymbol{\gamma})'(\mathbf{u} - \mathbf{Y}_f\boldsymbol{\gamma})}{m - p_f}. \tag{7.13}$$

Note that the posterior mean of u^* at $\mathbf{y}_f^*$ is given by,

$$E(u^*|\mathbf{y}_f^*, \mathbf{u}) = \mu_{u^*} = \mathbf{y}_f^{*\prime}\hat{\boldsymbol{\gamma}}, \tag{7.14}$$

and the posterior variance of u^* at $\mathbf{y}_f^*$ is given by

$$Var(u^*|\mathbf{y}_f^*, \mathbf{u}) = \frac{\nu_f}{\nu_f - 2}\sigma_{u^*}^2 = \frac{\nu_f}{\nu_f - 2}\hat{\sigma_f}^2\left(1 + \mathbf{y}_f^{*\prime}(\mathbf{Y}_f'\mathbf{Y}_f)^{-1}\mathbf{y}_f^*\right). \tag{7.15}$$

Thus, as it can be seen, the proposed Bayesian MCDM approach consists of linking two levels or stages, with each stage modelled via Bayesian regression (see figure 7.1).

7.3.1 Optimization

It is assumed that the objective of the MCDM problem is to find the values of the controllable factors $(x_1^*...x_k^*)$ that maximizes the probability that the customer's utility is at least l_u. In mathematical notations, the objective function is written as

$$\begin{aligned}
\max_{x_1^*...x_k^*} \quad & p(u^* \geq l_u|x_1^*...x_k^*, \mathbf{u}, \mathbf{Y}) \\
= \quad & \int_{\mathbf{y}^*=(y_1^*...y_q^*)} \left[p(u^* \geq l_u|\mathbf{y}^*, \mathbf{u})p(\mathbf{y}^*|x_1^*...x_k^*, \mathbf{Y})\right]\, d\mathbf{y}^* \\
= \quad & E_{\mathbf{y}^*}\left[p(u^* \geq l_u|\mathbf{y}^*, \mathbf{u})\right].
\end{aligned}$$

It is noted that for any given setting $(x_1^*...x_k^*)$, the outcome of the responses $(y_1^*...y_q^*)$ follows one of the distributions discussed earlier, based on the model used for the multiple

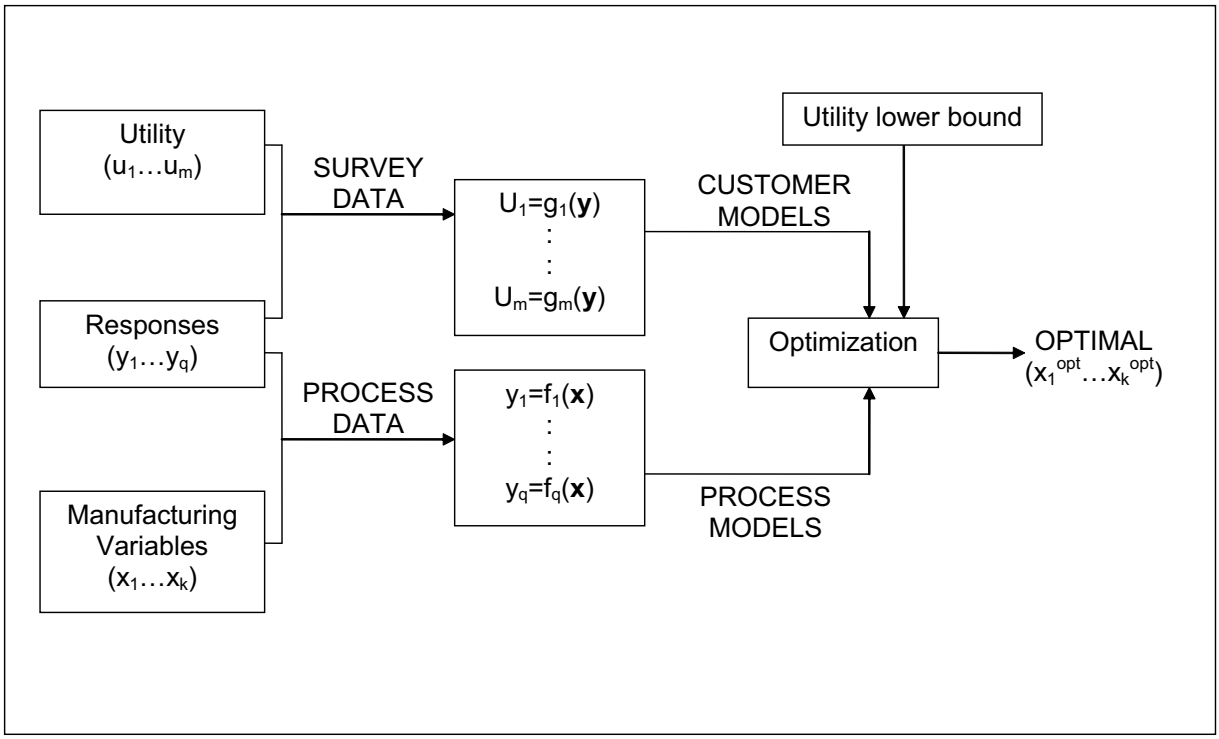

Figure 7.1: Block diagram of the proposed Bayesian MCDM method

responses. For each possible outcome of the responses, the distribution of the customer's utility u follows the distribution shown in equation (7.11). Thus given $(x_1^*...x_k^*)$, the probability that $u^* > l_u$ at this setting is determined by taking the expected value over the distribution of $(y_1^*...y_q^*)$ at that setting. The expected value in the objective function can be found by Monte Carlo simulation as shown in the steps below:

1. Set $count = 1$

2. Generate $\mathbf{y}^*(count) = \{y_1^*(count)...y_q^*(count)\}$ by sampling from the posterior distribution of the responses.

3. Compute $p(u^* \geq l_u|\mathbf{y}^*(count), \mathbf{u})$ for the sample using the c.d.f. (cumulative distribution function) of the distribution given in equation (7.11).

4. Set $count = count + 1$. Repeat from step 2 until $count > N$.

5. Estimate the expected value using the Weak Law of Large Numbers (WLLN), that is

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[p(u^* \geq l_u|\mathbf{y}^*(count), \mathbf{u})\right] = E_{\mathbf{y}^*}\left[p(u^* \geq l_u|\mathbf{y}^*, \mathbf{u})\right]. \tag{7.16}$$

The optimization problem can be solved with any constraints imposed on the feasible region of $(x_1^*...x_k^*)$ using nonlinear search algorithms. The example in section 7.3.3 illustrates the proposed methodology.

7.3.2 Application to Six Sigma Manufacturing

Six Sigma methodologies for managing quality are rapidly gaining popularity in industry [9]. A six sigma quality level performance corresponds to about 3.4 defects per million for

a normally distributed process that is off-target by 1.5σ, where σ is the standard deviation of the process. The Design for Six Sigma (DFSS) approach is used to achieve six sigma quality levels for the customer from the ground up by adjusting the manufacturing variables or control variables in the process. For example, in the manufacturing of light bulbs, the DFSS approach would be to adjust the manufacturing variables, say melting point of the filament and refractive index of the glass, so that the distribution of the lifetime of the bulb meets customer satisfaction at the six sigma quality level. These measures are often termed "CTQ's" in industry, that is Critical-To-Quality metrics. Thus in the light bulb example, the DFSS approach is to adjust the manufacturing variables (melting point of the filament and refractive index of the glass) so that the distribution of the system CTQ's (lifetime of the bulb) meets the customer CTQ (customer satisfaction or utility) at the six sigma quality level.

The conventional non-Bayesian approach to this problem is to fit a customer-utility model to the customer CTQ's as functions of the system CTQ's and to fit a process model to the system CTQ's as functions of the manufacturing variables. Using the customer-utility model, the desired values of the system CTQ's that maximize the expected value of the the customer CTQ are identified. The process model is then used to identify the values of the manufacturing variables such that the expected value of the system CTQ's from the model is equal to the desired value of the system CTQ's obtained using the customer-utility model. The reliability is measured using process capability indices that, however, do not give a probability of conformance and do not account for the uncertainty in the parameters of the models. In other words, the distribution of the products meeting the customer's utility score is not known.

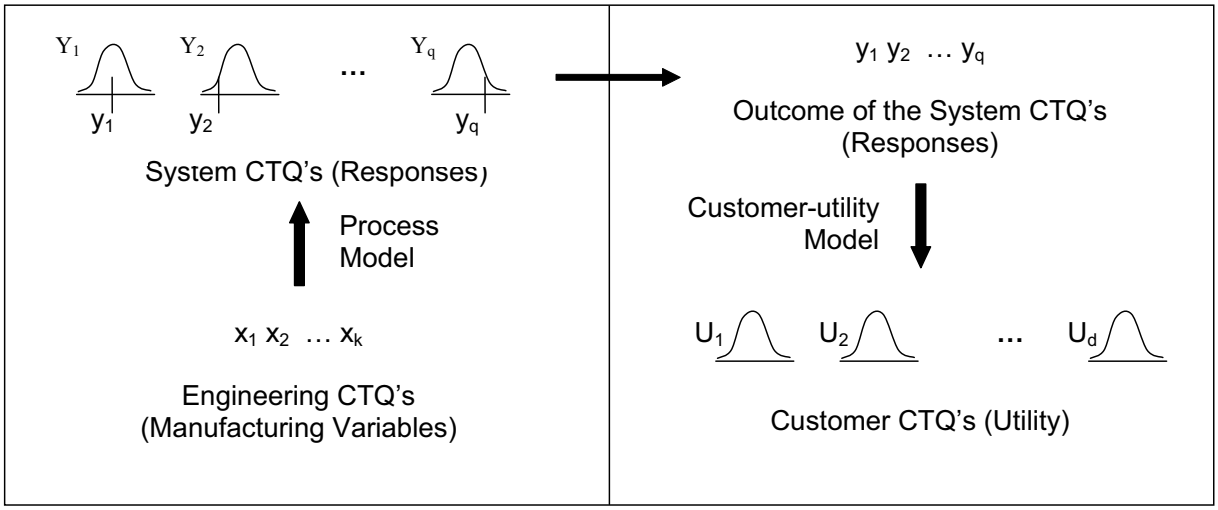

Figure 7.2: Block diagram of the Design for Six Sigma approach

In the proposed Bayesian approach, it is also true that customer-utility models are fit to the customer CTQ's as functions of the system CTQ's, and process models are fit to the system CTQ's as functions of the manufacturing variables (see figure 7.1). However, here the reliability is measured in terms of the probability that the customer's satisfaction or utility is above a given lower bound. Figure 7.2 shows the Bayesian approach to DFSS, where the posterior distribution of the responses are a function of the the setting of the manufacturing variables and the posterior distribution of the utility is a function of realized values of the responses. The values of the manufacturing variables are identified by finding those settings that result in the distribution (as opposed to the expected value) of the responses such that the probability that the customer's satisfaction or utility score is above a given lower bound. This posterior predictive distribution implicitly models the uncertainty of the parameters.

The examples below illustrate the proposed Bayesian methodology for MCDM.

7.3.3 Example: CVD Process

This example uses data from Czitrom and Spagon [14] for a chemical vapor deposition (CVD) process, an important step in the manufacture of semiconductors. The goal of the experiment was to model the deposition layer uniformity and deposition layer stress responses. The central composite inscribed (CCI) design that was used and the experimental data are shown in Table 7.1. There are two controllable factors: Pressure measured in torr, and ratio of the gaseous reactants H_2 and WF_6 (denoted by H_2/WF_6). The response "Uniformity" indicates the variation in the layer being deposited on a wafer. Therefore, smaller Uniformity values are preferred. A smaller value of the second response "Stress" is also desirable. Note that the controllable factors are provided in the $[-1, 1]$ coded variable range which is preferable for modelling. Based on the coded data, the model obtained for the responses using the ordinary least square estimates shown in equation (7.3) are given by:

$$\hat{y}_1 = 5.8661 - 1.9097x_1 - 0.2241x_2 + 1.6862x_1x_2 + 0.1337x_1^2 + 0.0337x_2^2, \tag{7.17}$$

$$\hat{y}_2 = 7.7900 + 0.7359x_1 + 0.4969x_2 + 0.0694x_1x_2 - 0.5287x_1^2 - 0.1187x_2^2. \tag{7.18}$$

Table 7.2 gives the results from the survey based on different combinations of the responses. Each combination of the responses, y_1 and y_2, in the survey is based on runs from a central composite design as can be seen in the coded variables. The coding is based on setting the smallest observed value of the response in table 7.1 as -1, and the largest observed value as 1. Table 7.2 shows sample results from 4 different surveys, which would typically be filled out by the plant engineer. More desirable responses correspond to higher u_i values. It is of interest to observe how the optimization results vary according to the answers to the

Run	Pressure, x_1	H_2/WF_6, x_2	coded x_1	coded x_2	Uniformity, y_1	Stress, y_2
1	80	6	1	0	4.6	8.04
2	42	6	0	0	6.2	7.78
3	68.87	3.17	0.71	-0.71	3.4	7.58
4	15.13	8.83	-0.71	0.71	6.9	7.27
5	4	6	-1	0	7.3	6.49
6	42	6	0	0	6.4	7.69
7	15.13	3.17	-0.71	-0.71	8.6	6.66
8	42	2	0	-1	6.3	7.16
9	68.87	8.83	0.71	0.71	5.1	8.33
10	42	10	0	1	5.4	8.19
11	42	6	0	0	5	7.9

Table 7.1: Data for CVD process example [14]

survey by comparing the results between the different surveys.

combination	y_1	y_2	coded y_1	coded y_2	Survey 1, u_1	Survey 2, u_2	Survey 3, u_3	Survey 4, u_4
1	3.40	6.49	-1	-1	10	8	9	8
2	8.60	6.49	1	-1	4	3	7	7
3	3.40	8.33	-1	1	6	6	2	1
4	8.60	8.33	1	1	1	2	0	0
5	2.32	7.41	-1.4142	0	9	8	5	5
6	9.68	7.41	1.4142	0	2	1	3	4
7	6.00	6.11	0	-1.4142	7	5	9	10
8	6.00	8.71	0	1.4142	3	3	3	0
9	6.00	7.41	0	0	6	4	2	2

Table 7.2: Data for 4 different surveys for CVD example

For each survey, i, a quadratic model of the form shown in equation (7.7) is fitted to the utility u_i from the data:

$$\hat{u}_1 = 6.00 - 2.61y_1 - 1.58y_2 + 0.25y_1y_2 - 0.25y_1^2 - 0.50y_2^2, \tag{7.19}$$

$$\hat{u}_2 = 4.00 - 2.36y_1 - 0.72y_2 + 0.25y_1y_2 + 0.37y_1^2 + 0.12y_2^2, \tag{7.20}$$

$$\hat{u}_3 = 2.00 - 0.85y_1 - 2.81y_2 + 0.87y_1^2 + 1.87y_2^2, \tag{7.21}$$

$$\hat{u}_4 = 2.00 - 0.42y_1 - 3.51y_2 + 1.06y_1^2 + 1.31y_2^2. \tag{7.22}$$

where the y_j are in coded form. For the optimization, suppose that it is desired to find the settings of Pressure and H_2/WF_6, that maximize the probability that the customer's (in this case, the plant engineer's) utility is at least 8 on a $0-10$ scale. Figure 7.3 shows the value of the probability $p(u^* \geq 8|x_1^*, x_2^*, \mathbf{u}, \mathbf{Y})$ over all the values of Pressure and H_2/WF_6 shown in coded variables for each survey. Note that the profile of the surface is different based on each survey. The optimization results are shown in table 7.3. Thus, if the plant engineer had filled out the survey as given in survey 1 or 2, then the best setting of the controllable factors is a pressure of 80 torr and H_2/WF_6 ratio of 2. If instead the plant engineer had filled out the survey as given in survey 3 or 4, then the best setting of the controllable factors is pressure of 4 torr and H_2/WF_6 ratio of 2. Note that the maximum value of the probability $p(u^* \geq 8|x_1^*, x_2^*, \mathbf{u}, \mathbf{Y})$ is also different for each survey.

7.3.4 Example: HPLC Process

The data for this example is taken from [54] and is presented in table 7.4. There are four responses in the high performance liquid chromatography (HPLC) system namely, the

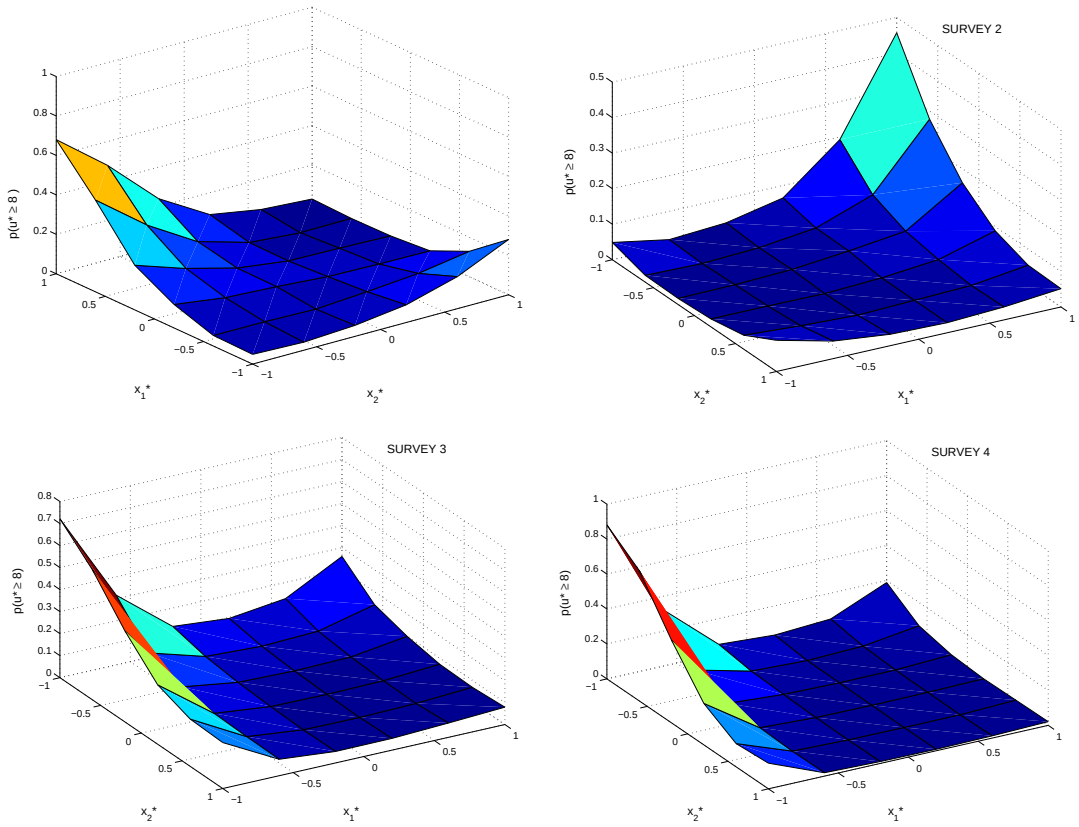

Figure 7.3: Surface plot of $p(u^* \geq 8)$ for different values of $\mathbf{x}^*$ for CVD example

Optimization Results					
	$p(u^* > 8)$	coded x_1^*	coded x_2^*	Pressure	H_2/WF_6
Survey 1	0.7102	1	-1	80	2
Survey 2	0.4436	1	-1	80	2
Survey 3	0.7271	-1	-1	4	2
Survey 4	0.8828	-1	-1	4	2

Table 7.3: Optimization results for the 4 surveys

critical resolution (Rs), total run time, signal-to-noise ratio of the last peak and the tailing factor of the major peak. There are three controllable factors: %IPA, temperature and pH. Here, it is assumed that the model is of the SMR form, i.e., all the responses have the same regressors and the error terms are correlated between the responses. The vector of regressors used for the process model is $(1, x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2)'$. The parameter estimates are obtained using equations (6.26 - 6.28). As the model is SMR, the joint posterior distribution of the responses is given by a multivariate **T**-distribution as shown in equation (6.25), from which the responses are sampled in step 3 of the optimization process described in section 7.3.1.

Table 7.5 gives data from a single sample survey based on different combinations of the responses. Each combination of the responses (y_1, y_2, y_3, y_4) in the survey is based on runs from a small composite design as can be seen in the coded variables. As in the previous example, the coding is based on setting the smallest observed value of the response in table 7.4 as -1, and the largest observed value as 1. It is noted here that the utility in the survey is a score in the range 0-20. Based on the survey data, the model fitted to the utility is given by,

$$\hat{u} = 14 + y_1 - 4.25y_2 + 6.50y_3 + 0.75y_4 - y_2y_3 - 2.75y_1^2 + 1.25y_2^2 - 0.25y_3^2 - 6.25y_4^2. \quad (7.23)$$

Figure 7.4 shows the scatter plot of the posterior probability $p(u^* \geq 15|x_1^*, x_2^*, x_3^*, \mathbf{u}, \mathbf{Y})$ over all combinations of the control factors $\mathbf{x}^*$, plotted on a grid 0.4 apart in the space $\{x_1^* \in [-1, 1], x_2^* \in [-1, 1], x_3^* \in [-1, 1]\}$. In the plot, the larger and darker circles indicate a higher value of this posterior probability. In this example, the posterior probability $p(u^* \geq 15|x_1^*, x_2^*, x_3^*, \mathbf{u}, \mathbf{Y})$ is maximized at a value of 0.787 at the setting $\mathbf{x}^* =$

$[-0.015, 0.653, -0.366]$. In the original (uncoded) units, this corresponds to setting %IPA at 69.9, temperature at 46.5 and pH at 0.129.

Run	%IPA, x_1	Temp., x_2	pH, x_3	coded x_1	coded x_2	coded x_3	Rs, y_1	Run time, y_2	S/N, y_3	Tailing, y_4
1	65	30	0.175	-1	-1	0	2.14	22	172	0.76
2	65	50	0.175	-1	1	0	1.73	12	311	0.88
3	65	40	0.05	-1	0	-1	1.93	16	251	0.8
4	65	40	0.3	-1	0	1	1.95	16	241	0.8
5	70	40	0.175	0	0	0	2.17	14	278	0.79
6	70	50	0.05	0	1	-1	1.97	11	371	0.86
7	70	30	0.3	0	-1	1	2.38	19	194	0.74
8	70	50	0.3	0	1	1	1.98	11	360	0.86
9	70	30	0.05	0	-1	-1	2.37	18	204	0.74
10	70	40	0.175	0	0	0	2.2	14	280	0.78
11	75	40	0.3	1	0	1	2.42	13	314	0.78
12	75	30	0.175	1	-1	0	2.61	17	223	0.73
13	75	50	0.175	1	1	0	2.14	10	410	0.85
14	75	40	0.05	1	0	-1	2.42	12	324	0.78
15	70	40	0.175	0	0	0	2.2	14	281	0.79

Table 7.4: Data for HPLC process example [54]

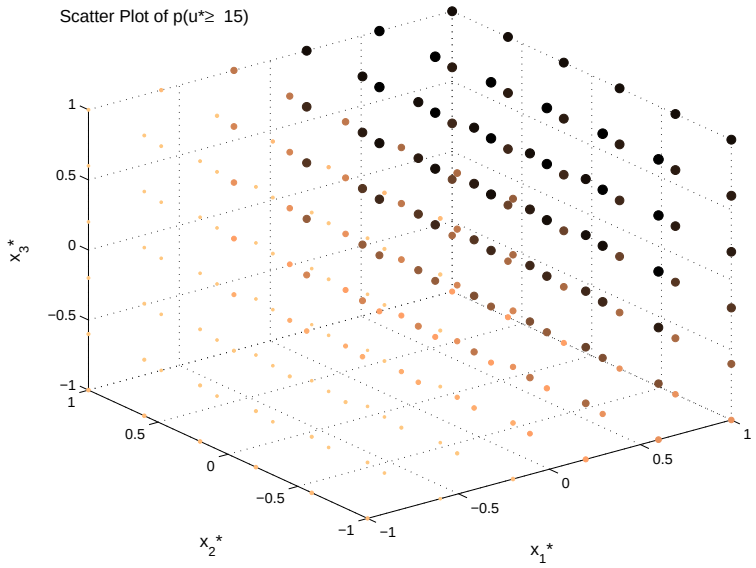

Figure 7.4: Scatter plot of $p(u^* \geq 15)$ for different values of $\mathbf{x}^*$ for HPLC example

7.4 Extension to Multiple Decision Makers

In the previous sections, the discussion of the proposed methodology was restricted to the case of a single customer or decision maker (DM). In this section, the methodology is extended to cases where there are two or more DM's. Here, each of the d DM's fills out a survey, and the optimization is carried out in order to maximize the probability that the utility of the i^{th} DM is at least equal to l_{u_i}.

Once again, it is assumed that there are k controllable factors, which can be set to desired values, and that there are q responses that depend on these controllable factors. It is assumed that data from an experiment with n runs is available and the responses can be modelled independently using linear regression as functions of the controllable factors if the error terms are correlated, or using SMR or SUR if the error terms are correlated. Based on the chosen model, it is possible to obtain a sample $\mathbf{y}^*$ from the corresponding

combination	y_1	y_2	y_3	y_4	coded y_1	coded y_2	coded y_3	coded y_4	Survey, u
1	1.73	16	291	0.805	-1	0	0	0	8
2	2.61	16	291	0.805	1	0	0	0	12
3	2.17	10	291	0.805	0	-1	0	0	17
4	2.17	22	291	0.805	0	1	0	0	11
5	2.17	16	172	0.805	0	0	-1	0	7
6	2.17	16	410	0.805	0	0	1	0	18
7	2.17	16	291	0.73	0	0	0	-1	5
8	2.17	16	291	0.88	0	0	0	1	8
9	2.17	16	291	0.805	0	0	0	0	14
10	1.95	13	231.5	0.7675	-0.5	-0.5	-0.5	-0.5	12
11	2.39	13	231.5	0.8425	0.5	-0.5	-0.5	0.5	12
12	1.95	19	231.5	0.8425	-0.5	0.5	-0.5	0.5	7
13	2.39	19	231.5	0.7675	0.5	0.5	-0.5	-0.5	7
14	1.95	13	350.5	0.8425	-0.5	-0.5	0.5	0.5	20
15	2.39	13	350.5	0.7675	0.5	-0.5	0.5	-0.5	20
16	1.95	19	350.5	0.7675	-0.5	0.5	0.5	-0.5	14
17	2.39	19	350.5	0.8425	0.5	0.5	0.5	0.5	14

Table 7.5: Data for 4 different surveys for HPLC example

posterior predictive density of the responses, as described in section 7.3. As in the case with a single DM, it is assumed that data from a survey with m questions is available, where each question presents a different combination of values of the q responses to the DM's who give a score on a numerical scale (e.g. 0-10) giving his/her preferences to that combination. Assuming that the error terms in the models of the utility function are uncorrelated between the DM's, it is now possible to model the utility of each of the DM's based on the survey as functions of the q responses as shown in equation (7.7). The model for the utility function of the i^{th} DM is given by:

$$u_i = \mathbf{y}'_{f_i}\boldsymbol{\gamma}_i + \epsilon_i, \tag{7.24}$$

where u_i is the customer's score or utility, $\mathbf{y}_{f_i}$ is the $(p_{f_i} \times 1)$ vector of regressors where each regressor is a function of the q responses $(y_1...y_q)$, $\boldsymbol{\gamma}_i$ is the $(p_{f_i} \times 1)$ vector of model parameters, and ϵ_i is the error term, assumed $N(0, \sigma^2_{f_i})$. Suppose $\mathbf{Y}_{f_i}$ is the $(m \times p_{f_i})$ design matrix of the survey for the i^{th} DM, and $\mathbf{u}_i$ is the corresponding vector of answers from the survey. Following the assumption that the error terms ϵ_i are uncorrelated for all i, for a diffuse prior on the model parameters in equation (7.7) for all $i \in \{1...d\}$ as shown in equations (7.8-7.10), the posterior predictive density of the i^{th} DM follows a t-distribution, i.e.,

$$u^*_i|\mathbf{y}^*_{f_i}, \mathbf{u}_i \sim t_{\nu_{f_i}}(\mu_{u^*_i}, \sigma^2_{u^*_i}), \tag{7.25}$$

where $\nu_{f_i} = m - p_{f_i}$, $\mu_{u^*_i} = \mathbf{y}^{*\prime}_{f_i}\hat{\boldsymbol{\gamma}}_i$, and $\sigma^2_{u^*_i} = \hat{\sigma}^2_{f_i}(1 + \mathbf{y}^{*\prime}_{f_i}(\mathbf{Y}'_{f_i}\mathbf{Y}_{f_i})^{-1}\mathbf{y}^*_{f_i})$. Here,

$$\hat{\boldsymbol{\gamma}}_i = (\mathbf{Y}'_{f_i}\mathbf{Y}_{f_i})^{-1}\mathbf{Y}'_{f_i}\mathbf{u}_i, \tag{7.26}$$

and,

$$\hat{\sigma}^2_{f_i} = \frac{(\mathbf{u}_i - \mathbf{Y}_{f_i}\boldsymbol{\gamma}_i)'(\mathbf{u}_i - \mathbf{Y}_{f_i}\boldsymbol{\gamma}_i)}{m - p_{f_i}}. \tag{7.27}$$

Therefore, the joint posterior probability that the utility of the i^{th} DM is at least l_{u_i} is simply the product of the marginals, i.e.,

$$p(u_1^* \geq l_{u_1}, u_2^* \geq l_{u_2} ... u_d^* \geq l_{u_d} | \mathbf{y}_{f1}^* ... \mathbf{y}_{fd}^*, \mathbf{u}_1 ... \mathbf{u}_d) = \prod_{i=1}^{d} p(u_i^* \geq l_{u_i} | \mathbf{y}_{f_i}^*, \mathbf{u}_i). \tag{7.28}$$

7.4.1 Optimization

In the case of multiple DM's, a reasonable objective in the MCDM problem is to find the values of the controllable factors $(x_1^* ... x_k^*)$ that maximize the probability that the utility of the i^{th} DM is at least l_{u_i} for all $i = \{1...d\}$. The objective function is written as

$$\begin{aligned} \max_{x_1^* ... x_k^*} \quad & \prod_{i=1}^{d} p(u_i^* \geq l_{u_i} | x_1^* ... x_k^*, \mathbf{y}_{f_i}^*, \mathbf{u}_i) \\ = \quad & \int_{\mathbf{y}^* = (y_1^* ... y_q^*)} \left[\prod_{i=1}^{d} p(u_i^* \geq l_{u_i} | \mathbf{y}_{f_i}^*, \mathbf{u}_i) p(\mathbf{y}^* | x_1^* ... x_k^*, \mathbf{Y}) \right] d\mathbf{y}^* \\ = \quad & E_{\mathbf{y}^*} \left[\prod_{i=1}^{d} p(u_i^* \geq l_{u_i} | \mathbf{y}_{f_i}^*, \mathbf{u}_i) \right]. \end{aligned}$$

Here, for each possible outcome of the responses, the distribution of the i^{th} DM's utility u_i follows the distribution shown in equation (7.25). Thus given $(x_1^* ... x_k^*)$, the probability that $u_i^* > l_{u_i} \, \forall \, i$ at this setting is determined by taking the expected value over the distribution of $(y_1^* ... y_q^*)$ at that setting. The expected value in the objective function can be found by Montecarlo simulation as shown in the steps below:

1. Set $count = 1$

2. Generate $\mathbf{y}^*(count) = \{y_1^*(count) ... y_q^*(count)\}$ by sampling from the posterior distribution of the responses. Note that regressors $\mathbf{y}_{f_i}^*(count)$ can be obtained from the sample for all i.

3. Compute $\prod_{i=1}^{d} p(u_i^* \geq l_{u_i}|\mathbf{y}^*(count), \mathbf{u}_i)$ for the sample using the c.d.f. (cumulative distribution function) of the distribution given in equation (7.25).

4. Set $count = count + 1$. Repeat from step 2 until $count > N$.

5. Estimate the expected value using the Weak Law of Large Numbers (WLLN),

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[\prod_{i=1}^{d} p(u_i^* \geq l_{u_i}|\mathbf{y}^*(count), \mathbf{u}_i) \right] = E_{\mathbf{y}^*} \left[\prod_{i=1}^{d} p(u_i^* \geq l_{u_i}|\mathbf{y}^*, \mathbf{u}_i) \right]. \quad (7.29)$$

The optimization problem can be solved with any constraints imposed on the feasible region of $(x_1^*...x_k^*)$ using a nonlinear search algorithm. The example below illustrates the proposed methodology.

7.4.2 Example: Multiple DM's

To illustrate the method for multiple DM's, consider the data from the example in section 7.3.3. In that section, the MCDM problem was solved for each DM individually. Therefore, for each survey shown in table 7.2, the optimization problem gives a different solution as shown in table 7.3. Here also, the same data as shown in tables 7.1 and 7.2 is used. However, now all the 4 surveys are considered simultaneously. The models for the responses and the utility are the same as in section 7.3.3 as given by equations (7.17-7.18) and (7.19-7.22).

Figure 7.5 shows the surface plot of the joint posterior predictive density of the 4 DM's utility as functions of the manufacturing variables. The figure shows two cases, the first one where $p(u_i^* \geq 8|x_1^*, x_2^*, \mathbf{u}_i, \mathbf{Y}) \,\forall\, i$ and the second where $p(u_i^* \geq 5|x_1^*, x_2^*, \mathbf{u}_i, \mathbf{Y}) \,\forall\, i$. In both cases the posterior probability is maximized at the setting $\mathbf{x}^* = [1, -1]$, where the value of

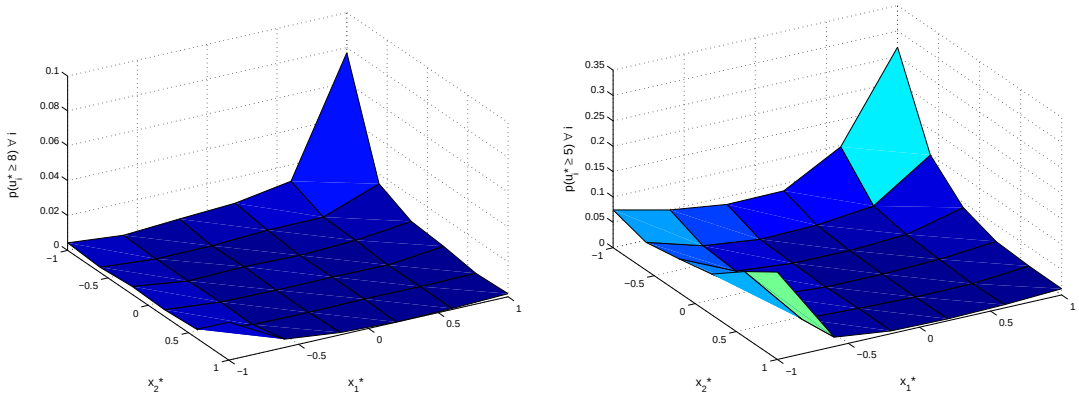

Figure 7.5: Surface plot of $p(u_i^* \geq 8|x_1^*, x_2^*, \mathbf{u}_i, \mathbf{Y})$ and $p(u_i^* \geq 5|x_1^*, x_2^*, \mathbf{u}_i, \mathbf{Y})$ for all i, for different values of coded Pressure and H_2/WF_6

$p(u_i^* \geq 8|x_1^*, x_2^*, \mathbf{u}_i, \mathbf{Y}) \,\forall\, i = 0.07$, and the value of $p(u_i^* \geq 5|x_1^*, x_2^*, \mathbf{u}_i, \mathbf{Y}) \,\forall\, i = 0.28$.

7.5 Discussion

A new algorithm to solve the MCDM problem was presented. The methodology maximizes the probability that the DM's utility function is greater than some user-defined value. As opposed to traditional methods that use expected value models for optimization, the Bayesian methodology takes into account the uncertainties in the model parameters. The examples provided demonstrate how the solution to the MCDM varies with differences in the preferences of the decision maker. The methodology was also extended to the case of multiple decision makers.

In the examples shown, the survey was designed by coding the maximum observed value of the responses at 1 and the minimum at -1, and using a central composite design. It should be noted that other designs such as space-filling or D-optimal designs could also be used depending on the type of the customer utility model. In the examples, the regressors

for both the process and the customer models essentially included main effects, two-way interactions and quadratic effects. For a general case, it is recommended to choose the regressors based on any prior knowledge of the response surface, especially for the customer utility model. It should be pointed out that diffuse or non-informative priors were used throughout, thus the resulting approach can be classified as "objective-Bayesian".

Chapter 8

Summary of Contributions and Directions of Future research

The ideas presented in this dissertation addressed three main applications in manufacturing, namely, process optimization, tolerance control and multiple criteria decision making. For each of these applications, new methodologies were presented based on the Bayesian framework of inference in order to obtain solutions that are robust with respect to different sources of uncertainty. Examples were provided that use real application data to demonstrate the proposed methodologies. In this chapter, we summarize the contributions made on each of the three main topics and discuss some directions for future research.

8.1 Contributions in Process optimization

In process optimization, the proposed methodology finds the optimal setting of the control factors that maximizes the posterior probability that the response lies within given specifica-

tion limits. The posterior probability is calculated by averaging across multiple competing models that may be used to represent the process, using the Bayesian model-averaged posterior predictive density (MAP). The MAP is calculated by taking the weighted average of the posterior predictive density for each individual model, using the posterior probability of the models as the weights. The solution to the optimization problem is therefore robust to the uncertainty in the true model form. In addition, the posterior predictive density for each individual model is obtained by integrating over the posterior distribution of the parameters in the model. The solution is thus not only robust with respect to the model form, but also with respect to the uncertainty in the value of the parameters for each competing model. The significant results of the model-robust optimization method proposed are listed below:

1. The calculation of the MAP requires the computation of the posterior probability of all the competing models, as well as the posterior predictive density of the response for each of these models. The computation of the model posteriors is a Bayesian hypothesis testing problem while that of the posterior predictive density is a Bayesian estimation problem. Although an objective-Bayesian approach using the uninformative prior can be used for Bayesian estimation, it is not suitable for Bayesian hypothesis testing. Hence, the prior chosen here is a combination of the uninformative prior and the Zellner's g-prior [74] as shown in equations (3.4), (3.8), (3.9) and (3.10). However, this introduces hyperparameters π and γ that have to be estimated. We use a value of π chosen by the user, and estimate γ using an empirical Bayesian approach proposed by [43]. The sensitivity analysis on the choice of these

hyperparameters performed for the examples in chapter 3 revealed that the solution is relatively less sensitive to the choice of π at the value of γ chosen using the above method.

2. For the priors chosen, the posterior predictive density of the response for a given model, M_i, at a given setting of the control factors, $\mathbf{x}^*$ was derived to follow the t-distribution.

3. The advantage of using the MAP for robust process optimization was demonstrated in the examples in chapter 3. It was seen that the optimal solution obtained using the MAP provided a response with the highest mean probability of conformance (with lowest standard deviation) across different cases of the true model, in comparison to using any single model.

4. The Bayesian method for model-robust process optimization was extended in chapter 4 to the case where there are noise factors present. In case of processes with noise factors, the method computes the posterior predictive density of the response for a given model at a given setting of the control factors by taking the expected value over the distribution of the noise factors. This density is then used to calculate the MAP used in the optimization. The example provided showed that the presence of noise factors can not only decrease the probability of conformance of the response to the specification limits, but can also cause a shift in the optimal setting of the control factors because of the models containing interaction terms between the control and the noise factors.

5. In chapter 4, the Bayesian method for model-robust process optimization was also extended to the case where the error terms in the model do not follow a normal distribution. Instead it was assumed that the errors follow the t-distribution that has thicker tails than the normal distribution. The method proposed uses the fact that a t-distributed random variable can be expressed as a function of a standard normal random variable and a chi-square distributed random variable. The results for the normally distributed case as derived in chapter 3 are then used by taking the expected value with respect to the chi-square distributed variables.

6. In chapter 5, the results for model-robust process optimization were extended to systems with multiple responses. Different cases are analyzed based on the correlation between the error terms of the responses and the regressors present in the models for the multiple responses.

8.1.1 Application to Process Control

Although the results presented in this dissertation for robustness with respect to the model form focussed on application in process optimization, the results can be directly applied to process control, specifically, to what is called "single-period control" [57]. Consider the single-period control of a response y which can be controlled by adjusting the inputs $\mathbf{x} = (x_1 ... x_k)$. As in process optimization, the process model is assumed to be of the form

$$y_t = \boldsymbol{\beta}' \mathbf{x}_t + \epsilon_t, \tag{8.1}$$

where the subscript t here refers to the time. In the process control problem, the objective is to find the optimal setting $\mathbf{x}_{T+1}$ for time $(T+1)$, in order to obtain a desirable response

y_{T+1}. Observations of $\mathbf{x}_t$ and y_t from time $t = 1$ to $t = T$ are used as the data to fit the process model. As seen in chapter 3, the data can be used to fit multiple competing models of the form shown in equation (8.1). The optimization problem is then formulated to find the value of $\mathbf{x}_{T+1}$ that maximizes the model-averaged posterior probability that the future response y_{T+1} lies within the desired specification limits. The proposed methodology for non-normal error terms and for multiple response systems can also be applied similarly to single-period control.

8.2 Contributions in Tolerance Control

A Bayesian method was proposed to set tolerance or specification limits on responses. The methodology uses a Bayesian approach for robust parameter design and tolerance control. The predictive density of the response is used to find the limits defining the tolerance interval that contains a given value of probability of conformance of the response. Some of the significant results in tolerance control from the dissertation are listed below:

1. The optimization problem is formulated to also find the smallest tolerance interval where the probability of conformance of the response is at least equal to some given value. Therefore, there is also minimum variance among the conforming responses. In addition, constraints are set on the bounds for the lower and upper tolerance limits. Therefore, the mean of the conforming responses are constrained to lie within a given region of interest.

2. The methodology was also extended to multiple response systems in order to simultaneously set tolerance limits on all responses. In this case, the objective in the

optimization is formulated as a function of the tolerance limits on the individual responses.

3. Uninformative priors are used in this approach. Hence, the proposed method is objective-Bayesian.

4. The analysis also included an approach to account for the reduction in the probability of conformance of the response to the calculated tolerance interval due to any variation in the setting of the control factors. It is assumed that the control factors vary with a known distribution about the calculated optimal setting. The reduced probability of conformance is calculated by taking the expected value of the probability of conformance with respect to the distribution of the control factors about the optimal setting. This is the Bayesian analogy to account for the "transmission of errors" from the control factors to the responses.

8.3 Contributions in Multiple Criteria Decision Making

A Bayesian approach was presented in chapter 7 for the MCDM problem. The dependence of the responses on the control factors and that of the decision maker's utility on the responses are modelled using Bayesian regression. Therefore, the methodology naturally takes into account uncertainty in the model parameters. The optimization problem is formulated to find the setting of the control factors that maximizes the posterior probability that the utility of the decision maker is greater than a given lower bound. Some of the

significant results on the proposed method are listed below:

1. Uninformative priors are used in the Bayesian regression. Hence, the proposed method is objective-Bayesian.

2. Application of the proposed method to the popular "Design for Six Sigma" quality technique was demonstrated.

3. Using the Bayesian approach, the distribution of the future responses $(y_1^*...y_q^*)$ is obtained for a given setting of the control factors $(x_1^*...x_k^*)$. In addition, for given values of the future responses, the posterior distribution of the customer's utility u^* is also obtained. The probability that u^* is greater than a given lower bound at a given setting of the control factors is thus determined by taking the expected value over the distribution of $(y_1^*...y_q^*)$ at that setting. The method uses a nonlinear program to find the setting of the control factors that maximizes this posterior probability.

4. The method was presented for the case where there is a single decision maker as well for the case where there are multiple decision makers.

8.4 Robustness with respect to Optimal Setting of the Control Factors

Control factors, by definition, are factors that can be set to desired values by the user. However, in practice, there are rarely any factors that do not have any variation about their setting. In some cases, this variation may be very small to cause any significant effect

on the response. But in general, it is of interest to ensure that the optimal solution is robust to the variation in the setting of the control factors for all the applications discussed in this dissertation, namely, process optimization, tolerance control and multiple criteria decision making.

In optimization problems, it is a common practice to study the sensitivity of the value of the objective function with respect to the optimal value of the control or decision variables, typically by computing the gradient of the objective function with respect to each control variable at the optimal point. However, this does not use any information on the distribution of the control factors about the optimal setting.

The Bayesian approach used here lends a convenient way of addressing robustness with respect to the optimal setting of the control factors. Consider an objective function Z, for example, the posterior probability of conformance of the response to its specification limits, whose optimal value Z^o is obtained at the setting $\mathbf{x}^o$. If the actual value of the control factors is $\tilde{\mathbf{x}} = \mathbf{x}^o + \boldsymbol{\delta}$, where the distribution of $\boldsymbol{\delta}$ is assumed to be known, then the expected value of the objective function due to the variation in the control factors can be calculated with respect to this distribution. The expected change Δ in the value of the objective function due to the variation in the control factors can be quantified and is given by

$$\Delta = Z^o - E_\delta[Z]. \tag{8.2}$$

8.5 Recommendations for Future Research

Some recommendations for future research are listed below:

1. In the results for Bayesian model-robust process optimization for multiple responses, the case of seemingly unrelated regression (SUR) was not considered in this dissertation. In case of SUR, both the model posterior probabilities as well as the posterior predictive density of the responses for the given model are not available in closed form. Percy [53] proposes a Gibbs' sampling procedure to generate a sample of responses $\mathbf{y}$ from the posterior distribution using the conditional distributions of $\mathbf{y}$, the parameters $\boldsymbol{\beta}$ and the precision matrix $\boldsymbol{\phi}$. However, the method assumes a vague prior for all the parameters which is not suitable for model-averaging. In addition, the computation time for calculating the probability of conformance is higher in the SUR case even for a single model because the computation involves a Monte Carlo simulation where each iteration within the simulation requires a sample $\mathbf{y}$ generated by Gibbs sampling. Considering all the limitations, a recommendation for future study is to use the method proposed by Percy to sample from the posterior distribution of $\mathbf{y}$ assuming vague priors for the parameters in the candidate list of models, and approximate the model posteriors $P(M_i|\mathbf{Y})$ using criteria such as Akaike or Bayesian information criteria (AIC and BIC, respectively) in order to approximate the model-averaged posterior probability of conformance as suggested in the review in section 2.3.

2. In this dissertation, the uncertainty in the model was taken into account during the model-fitting and process optimization stage. It was assumed here that data from an experiment already exists. However, usually optimal designs such as those based on D-optimality and G-optimality depend on the model form or the columns of the design matrix $\mathbf{X}$. Thus, given the parameters to be included in the model,

these optimality criteria choose a subset of runs (or rows of the **X** matrix) from all possible combinations of levels of the factors based on some condition to maximize the information about the model parameters, such as minimizing the determinant of the covariance matrix. DuMouchel and Jones [20] present a Bayesian D-optimality approach that is less dependent on the assumed model by splitting the model terms into "primary" terms, which are present in the model, and "potential" terms, which are typically the higher order terms. Here, the determinant of the posterior covariance matrix of the parameters is minimized. Neff [50] provides a Bayesian two-stage design for designing experiments under model uncertainty. A genetic algorithm approach to model-robust optimal designs is presented in [29]. A problem for future study is to combine these model-robust Bayesian approaches for experimental design and that of process optimization presented here.

3. In the Bayesian methods for robust tolerance control and for multiple criteria decision making, the approach used was objective-Bayesian. In both these methods, it was possible to use vague priors because a single model was assumed in the Bayesian regression. Thus, the results obtained are for the given model form. An interesting extension to these methods would be to use the idea of model-averaging in combination with robust tolerancing and MCDM. Thus, instead of using the posterior predictive density based on a single model as in chapters 6 and 7, one could potentially consider multiple competing models to calculate the model averaged posterior predictive density (MAP). The MAP could be used in the optimization formulations for tolerance control and MCDM.

4. In the Bayesian method for MCDM presented in chapter 7, there are two stages of Bayesian regression. In the first stage the responses are modelled as functions of the control factors and in the second the decision makers'(DM's) utilities are modelled as functions of the responses. Some potential problems of investigation here are:

    - In the second stage of the regression, the data used for modelling comes from a survey. In the examples provided, the survey was designed by coding the maximum observed value of the responses at 1 and the minimum at -1, and using a central composite design. However, in case the consistency of the DM is high, then the model will be deterministic. In such cases, a space filling design might be more suitable. A potential problem that can be investigated is the effect of the survey design on the results of the optimization.
    - Another potential problem here is the effect of the size of the survey (i.e., total number of combinations presented to the DM) on the optimization results. In this study, the cost of asking too many combinations to the DM should be compared to the cost of not getting adequate information from the DM.
    - In chapter 7, the utility models essentially included main effects, two-way interactions and quadratic effects. A possible extension is to study the effect of choosing different kinds of regressors for the utility model.

Bibliography

[1] Abramowitz, M. and Stegun, I. A. (eds), *Handbook of Mathematical Functions*, Dover Publications, New York, (1965).

[2] Apley, D., and Kim J., "A Cautious Approach to Robust Design with Model Uncertainty", *Proceedings of the 2002 IIE Annual Conference*, Orlando, FL, May 19-21, (2002).

[3] Becker, N.G., "Models for the response of a mixture", *Journal of the Royal Statistical Society, Series B (Methodological)*, 30, (1968).

[4] Berger, J.O., and Pericchi, L.R., "The Intrinsic Bayes factor for Model Selection and Prediction", *'Journal of American Statistical Association*, 91, 109-122, (1996).

[5] Berger, J.O., and Pericchi, L.R., "Objective Bayesian Methods for Model Selection: Introduction and Comparison", *Model Selection*, (P.Lahiri, editor), Institute of Mathematical Statistics Lecture Notes, Monograph Series vol. 38, (2001).

[6] Box, G.E.P., and Draper, N.R., *Empirical Model-building and Response Surfaces*, New York: Wiley (1987).

[7] Box, G.E.P., and Jones, S., "Designing Products that are Robust to the Environment", *CQPI Report No. 56*, University of Wisconsin, (1990).

[8] Box, G.E.P. and Tiao, G. C., *Bayesian Inference in Statistical Analysis*, John Wiley & Sons, New York, NY, (1992).

[9] Breyfogle III, F.W., *Implementing Six Sigma*, 2nd ed., John Wiley & Sons, New York, NY, (2003).

[10] Cahya, S., *Sampling Properties of Optimal Operating Conditions of Single and Multiple Response Surface Systems*, Doctoral Dissertation, IME department, Penn state University, (2001).

[11] Cahya, S., Del Castillo, E., and Peterson, J.J., "Computation of Confidence Regions for Optimal Factor Levels in Constrained Response Surface Problems," to appear in *Journal of Computational and Graphical Statistics*, (2004).

[12] Carlin, B.P., and Louis, T.A., *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed., Chapman and Hall/CRC Press, Boca Raton, FL, (2000).

[13] Chipman, H., "Handling Uncertainty in Analysis of Robust Design Experiments", *Journal of Quality Technology*, 30, 11-17, (1998).

[14] Czitrom, V., and Spagon, P.D., *Statistical Case Studies for Industrial Process Improvement*, American Statistical Association and the Society for Industrial and Applied Mathematics, (1997).

[15] Del Castillo, E., and Montgomery, D.C., "A Nonlinear Programming Solution to the Dual Response Problem", *Journal of Quality Technology*, 25(4), 199-204, (1993).

[16] Del Castillo, E., and Cahya, S., "A Tool for Computing Confidence Regions on the Stationary Point of a Response Surface", *The American Statistician*, 55(4), 358-365, (2001).

[17] Derringer, G., and Suich, R., "Simultaneous Optimization of Several Response Variables," *Journal of Quality Technology*, 12(4), 214-219, (1980).

[18] Draper, D., "Assessment and Propagation of Model Uncertainty", *Journal of the Royal Statistical Society Series B*, 57, 45-97, (1995).

[19] Draper, N.R., and Smith, H., *Applied Regression Analysis*, 3rd ed., John Wiley & Sons, New York, (1998).

[20] DuMouchel, M., and Jones, B., "A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model", *Technometrics*, 36(1), 37-47, (1994).

[21] Fernandez, C., Ley, E., and Steel, M.F.J., "Benchmark Priors for Bayesian Model Averaging", *Journal of Econometrics*,100(2), 381-427, (2001).

[22] Frisbee, S. E., and McGinity, J.W., "Influence of Nonionic Surfactants on the Physical and Chemical Properties of a Biodegradable Pseudolatex," *Eur. J. Pharm Biopharm.*, 40(6), 355-363, (1994).

[23] Geisser, S., *Predictive Inference: An Introduction*, Chapman and Hall, CRC Press, New York, (1993).

[24] Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B., *Bayesian Data Analysis*, Chapman and Hall, CRC Press, New York, (1995).

[25] Guttman, I., *Statistical Tolerance Regions: Classical and Bayesian*, Griffin, London (1981).

[26] Guttman, I., *Tolerance regions, statistical. Vol. 9. Encyclopedia of Statistical Sciences, Eds.: Kotz, S., and Johnson, N. L.*, John Wiley & Sons, New York, (1988).

[27] Hahn, E.D., "Decision Making with Uncertain Judgments: A Stochastic Formulation of the Analytic Hierarchy Process", *Decision Sciences*, 34(3), 443-466, (2003).

[28] Hamada, M., "Bayesian Tolerance Interval Control Limits for Attributes", *Quality and Reliability Engineering International*, 18, 45-52, (2002).

[29] Heredia-Langer, A., Montgomery, D.C., Carlyle, W.M., and Borror, C.M., "Model-Robust Optimal Designs: A Genetic Algorithm Approach", to appear in Journal of Quality Technology, (2004).

[30] Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T., "Bayesian Model Averaging", *Technical report 9814*, Department of Statistics, Colorado State University, (1998).

[31] Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T., "Bayesian Model Averaging - A Tutorial", *Statistical Science*, 14(4), 382-417, (1999).

[32] Hong, Y.S., and Chang, T.C., "A Comprehensive Review of Tolerancing Research", *International Journal of Production Research*, 40(11), 2425-2459, (2002).

[33] Johnson, N.L., Kotz, S., and Balakrishnan, N., *Continuous Univariate Distributions, Volume 1*, 2nd ed., John Wiley & Sons, New York, (1994).

[34] Johnson, N.L., Kotz, S., and Balakrishnan, N., *Continuous Univariate Distributions, Volume 2*, 2nd ed., John Wiley & Sons, New York, (1995).

[35] Johnson, R.A., and Wichern, D.W., *Applied Multivariate Statistical Analysis*, 2nd ed., Prentice Hall, New Jersey, (1998).

[36] Kass, R.E., and Raftery, A.E., "Bayes Factors", *Journal of the American Statistical Association*, 90(430), 773-795, (1995).

[37] Kass, R.E., and Wasserman, L., "The Selection of Prior Distributions by Formal Rules", *Journal of the American Statistical Association*, 91(435), 1343-1369, (1996).

[38] Khuri, A.I., and Cornell, J.A., *Response Surfaces*, Marcel Dekker Inc., New York, NY (1987).

[39] Lee, P.M., *Bayesian Statistics - An Introduction*, Arnold Publishing, London (1997).

[40] Lehtihet, E.A., Ranade, S., and Dewan, P., "Comparative Evaluation of Tolerance Control Chart Models", *International Journal of Production Research*, 38, 1539-1556, (2000).

[41] Madigan, D., and Raftery, A.E., "Model Selection and Accounting for Model Uncertainty in Graphical Models using Occam's Window", *Journal of the American Statistical Association*, 89 (428), (1994).

[42] "Matlab Optimization Toolbox 2.0, User Guide and Theory Document", Available at http://www.mathworks.com/access/helpdesk/help/pdf_doc/optim/optim_tb.pdf.

[43] Meyer, R.D., and Box, G., "Finding the Active Factors in Fractionated Screening Experiments", *CQPI Report No. 80*, University of Wisconsin, (1992).

[44] Meyer, R.D., and Box, G., "Finding the Active Factors in Fractionated Screening Experiments", *Journal of Quality Technology*, 25(2), 94-105, (1993).

[45] Meyer, R.D., Steinberg, D.M., and Box, G., "Follow-up Designs to Resolve Confounding in Multifactor Experiments", *Technometrics*, 38(4), 303-313, (1996).

[46] Miro-Quesada, G., Del Castillo, E., and Peterson, J. P., "A Bayesian Approach for Multiple Response Surface Optimization in the Presence of Noise Variables", *Journal of Applied Statistics*, 31(3), 251-270, (2004).

[47] Miro-Quesada, G., "Topics in the Optimization of Response Surface and Parameter Design Problems", Doctoral Dissertation, IME department, Penn state University, (2003).

[48] Miro-Quesada, G., and Del Castillo, E., "Two approaches for enhancing the dual response approach to Robust Parameter Design", *Journal of Quality Technology*, 36(2), 154-168, (2004).

[49] Myers, R.H., and Montgomery, D.C., *Response Surface Methodology*, 2nd ed., John Wiley & Sons, New York, NY, (2002).

[50] Neff, A.R., Myers, R.H., and Ye, K., "Bayesian Two Stage Design Under Model Uncertainty", *Technical Report No. 96-7*, Department of Statistics, Virginia Polytechnic Institute and State University, (1996).

[51] Nigam, S.D., and Turner, J.U., "Review of Statistical Approaches to Tolerance Analysis", *Computer-Aided Design*, 27(1), 6-15, (1995).

[52] Nocedal, J., and Wright, S.J., *Numerical Optimization*, Springer-Verlag New York Inc., New York, NY, (1999).

[53] Percy, D.F., "Predictions for Seemingly Unrelated regressions", *Journal of the Royal Statistical Society Series B*, 54, 243-252, (1992).

[54] Peterson, J.J., A Probability-based Desirability Function for Multiresponse Optimization, Proceedings of the Section on Quality and Productivity, Annual Meeting of the American Statistical Association, Indianapolis, IN, (2000).

[55] Peterson, J.J., Cahya, S., and Del Castillo, E., "A General Approach to Confidence Regions for Optimal Factor Levels of Response Surfaces", *Biometrics*, 58, 422-431, (2002).

[56] Peterson, J.P., "A Posterior Predictive Approach to Multiple Response Surface Optimization ", *Journal of Quality Technology*, 36(2), (2004).

[57] Press, S.J., *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*, 2nd ed., Robert E. Krieger Publishing, Malabar, FL, (1982).

[58] Press, S.J., *Subjective and Objective Bayesian Statistics*, 2nd ed., John Wiley & Sons, New York, NY, (2003).

[59] Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P., *Numerical Recipes in C*, New York: Cambridge University Press, (1992).

[60] Robert, C.P., *The Bayesian Choice*, Springer-Verlag New York Inc., New York, NY, (2001).

[61] Rousseeuw, P.J., and Leroy, A.M., *Robust Regression and Outlier Detection*, John Wiley & Sons, New York, NY, (1987).

[62] Saaty, T.L., *The Analytic Hierarchy Process*, McGraw-Hill International Book Company, New York, (1980).

[63] Shin W.S., and Ravindran A., "Interactive Multiple Objective Optimization: Survey 1− Continuous Case", *Computers and Operations Research*, 18(1), 97-114, (1991).

[64] Singpurwalla, N.D., "A Bayesian Perspective on Taguchi's Approach to Quality Engineering and Tolerance Design", *IIE Transactions*, 24(5), 18-27, (1992).

[65] Steuer R.E., *Multiple Criteria Optimization: Theory, Computation and Application*, Robert E. Krieger Publishing, Malabar, FL, (1989).

[66] Taguchi, G., and Wu, Y., *Introduction to Off-line Quality Control*, Central Japan Quality Control Association, Nagoya, Japan, (1980).

[67] Taguchi, G., *Introduction to Quality Engineering*, Asian Productivity Organization, UNIPUB, White Plains, New York, (1986).

[68] Taguchi, G., *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost*, UNIPUB/Kraus International, White Plains, New York, (1987).

[69] Taraman, K., "Multi Machining Output - Multi Independent Variable Turning Research by Response Surface Methodology", *International Journal of Production Research*, 12(2), 233-245, (1974).

[70] Wets, R.J.B., "Stochastic Programming", in *Handbooks in Operations Research and Management Science Vol 1: Optimization*, G.L. Nemhauser, A.H.G. Rinnooy Kan, and M.J. Todd (eds.), Elsevier Science Publishers B.V., Amserdam, The Netherlands, (1989).

[71] Wolfinger, R.D., "Tolerance Intervals for Variance Component Models using Bayesian Simulation", *Journal of Quality Technology*, 30(1), 18-32, (1998).

[72] Yu, P.L., "A Class of Solutions for Group Decision Problems", *Management Science*, 19(8), 936-946, (1973).

[73] Zeleny, M., "Compromise Programming", in *Multiple Criteria Decision Making*, J. L. Cochrane and M. Zeleny (eds.), Columbia: South Carolina: University of South Carolina Press, 262–301, (1998).

[74] Zellner, A., "On Assessing Prior Distributions and Bayesian Regression Analysis with g-prior Distributions", *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti*, eds. Goel, P.K., and Zellner, A., Amsterdam, North-Holland, (1986).

Appendix A

Choosing Model Subset for Averaging

In the method presented in section 3.2, it is usually not necessary to average over all m models considered originally. There are different methods proposed in literature to choose a subset of these, either based on the marginal likelihood $P(\mathbf{y}|M_i)$, or on the model posterior probabilities $P(M_i|\mathbf{y})$. Madigan *et al.* [41] propose an algorithm based on what they termed "Occam's window" (that comes from the well known concept of Occam's razor) to choose the subset of models. A simple criteria is to choose the subset of models, M_j, such that

$$\frac{\max_i P(M_i|\mathbf{y})}{P(M_j|\mathbf{y})} \leq c' \tag{A.1}$$

Here, c' is a constant whose value remains to be chosen. The above criteria are useful when the number of candidate models is very large. In the examples we discuss, there are fewer candidate models. Also, as the focus of this paper is on process optimization, we use a

simpler criteria to choose the subset of models from the original candidate list. Here, we order the model posteriors in descending order and include only the top m models, the sum of whose posteriors account for at least 95% of the total probability.

Appendix B

Derivation of Posterior Predictive Density: Single Response Case

Theorem: For a single response process with k controllable factors, under the process model of the form given in equation (3.1) with normally distributed error terms, and under the priors on the parameters given by equations (3.8), (3.9) and (3.10) with $\mathbf{\Sigma}_i^{-1} = (\mathbf{X}_i'\mathbf{X}_i)\mathbf{V}_i$, where $\mathbf{V}_i = \frac{1}{g}\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{t_i} \end{pmatrix}$, the cumulative posterior predictive density for the new response y^* at a new set of regressors $\mathbf{x}^*$ for a given model M_i, is given by:

$$P\left(\frac{y^* - \mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}_i}{\hat{\sigma}_i\sqrt{1+\mathbf{x}^{*\prime}(\mathbf{\Sigma}_i^{-1}+\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^*}} < t|M_i,\mathbf{x}^*,\mathbf{Y}\right) = \frac{1}{2}\left[1 + I_{\frac{t^2}{\nu+t^2}}\left(\frac{1}{2},\frac{\nu}{2}\right)\right], \qquad \text{(B.1)}$$

where $I_z(a,b)$ is the incomplete beta function, $\nu = n-1$, and $\hat{\sigma}_i^2 = S_i/(n-1)$.

Proof: The posterior predictive density is given by,

$$P(y^*|M_i,\mathbf{x}^*,\mathbf{y}) = \int_{\sigma^2}\int_{\boldsymbol{\beta}_i} P(y^*|M_i,\mathbf{x}^*,\mathbf{y},\sigma^2,\boldsymbol{\beta}_i)P(\boldsymbol{\beta}_i,\sigma^2|\mathbf{y},M_i)d\boldsymbol{\beta}_i d\sigma^2, \qquad \text{(B.2)}$$

where $P(y^*|M_i, \mathbf{x}^*, \mathbf{y}, \sigma^2, \boldsymbol{\beta}_i)$ is the likelihood function, and $P(\boldsymbol{\beta}_i, \sigma^2|\mathbf{y}, M_i)$ is the joint posterior of the model parameters. Assuming normally distributed errors, the likelihood is:

$$P(y^*|M_i, \mathbf{x}^*, \mathbf{y}, \sigma^2, \boldsymbol{\beta}_i) \propto \sigma^{-1} exp\left[\frac{-1}{2\sigma^2}(y^* - \mathbf{x}^{*\prime}\boldsymbol{\beta}_i)'(y^* - \mathbf{x}^{*\prime}\boldsymbol{\beta}_i)\right], \tag{B.3}$$

and

$$P(\boldsymbol{\beta}_i, \sigma^2|\mathbf{y}, M_i) \propto P(\mathbf{y}|M_i, \boldsymbol{\beta}_i, \sigma^2)P(\boldsymbol{\beta}_i|M_i, \sigma^2)P(\sigma^2|M_i), \tag{B.4}$$

where $P(\mathbf{y}|M_i, \boldsymbol{\beta}_i, \sigma^2)$ is the likelihood function given by equation (3.7). $P(\boldsymbol{\beta}_i|M_i, \sigma^2)$ and $P(\sigma^2|M_i)$ are the priors on the model parameters assumed to be of the form:

$$p(\sigma^2|M_i) \propto \frac{1}{\sigma^2}, \tag{B.5}$$

and,

$$P(\boldsymbol{\beta}_i|M_i, \sigma^2) \propto \gamma^{-t_i}\sigma^{-t_i} \exp\left[\frac{-1}{2\sigma^2}\boldsymbol{\beta}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}_i\right], \tag{B.6}$$

where,

$$\frac{g}{\gamma^2}\mathbf{V}_i = \boldsymbol{\Sigma}^{-1}. \tag{B.7}$$

Let $k_{1,i} = \gamma^{-t_i}$. Then,

$$P(\boldsymbol{\beta}_i, \sigma^2|\mathbf{y}, M_i) \propto k_{1,i}(\sigma^2)^{-\frac{n+t_i+2}{2}} exp\left[\frac{-1}{2\sigma^2}\left\{(\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta}_i)'(\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta}_i) + \boldsymbol{\beta}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\beta}_i\right\}\right]. \tag{B.8}$$

Let $\Omega_i = (y^* - \mathbf{x}^{*\prime}\boldsymbol{\beta}_i)'(y^* - \mathbf{x}^{*\prime}\boldsymbol{\beta}_i) + (\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta}_i)'(\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta}_i) + \boldsymbol{\beta}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\beta}_i$. This gives,

$$P(y^*|M_i, \mathbf{x}^*, \mathbf{y}) \propto k_{1,i}\int_{\sigma^2}\int_{\boldsymbol{\beta}_i}(\sigma^2)^{-\frac{n+t_i+3}{2}} exp\left[-\frac{\Omega_i}{2\sigma^2}\right] d\boldsymbol{\beta}_i d\sigma^2. \tag{B.9}$$

By making a substitution $u = \frac{\Omega_i}{2\sigma^2}$, the above equation can be rewritten as,

$$P(y^*|M_i, \mathbf{x}^*, \mathbf{y}) \propto k_{1,i}\int_{\boldsymbol{\beta}_i}\left(\frac{\Omega_i}{2}\right)^{\frac{n+t_i+1}{2}}\left[\int_0^\infty u^{\frac{n+t_i-1}{2}} exp(-u)du\right] d\boldsymbol{\beta}_i. \tag{B.10}$$

The inner integral inside the square brackets is a constant given by the gamma function, $\Gamma\left(\frac{n+t_i+1}{2}\right)$. Let $k_{2,i} = k_{1,i}\Gamma\left(\frac{n+t_i+1}{2}\right)$. Then,

$$P(y^*|M_i, \mathbf{x}^*, \mathbf{y}) \propto k_{2,i} \int_{\boldsymbol{\beta}_i} \left(\frac{\Omega_i}{2}\right)^{\frac{n+t_i+1}{2}} d\boldsymbol{\beta}_i. \tag{B.11}$$

It can be shown that,

$$(\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta}_i)'(\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta}_i) + \boldsymbol{\beta}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\beta}_i = S_i + (\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)'(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i), \tag{B.12}$$

where S_i is defined in equation (3.15), and $\hat{\boldsymbol{\beta}}_i$ is defined in equation (3.16). From the above equation, we can rewrite Ω_i as,

$$\Omega_i = (y^* - \mathbf{x}^{*\prime}\boldsymbol{\beta}_i)'(y^* - \mathbf{x}^{*\prime}\boldsymbol{\beta}_i) + S_i + (\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)'(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i). \tag{B.13}$$

Define the $(r_i \times 1)$ vector $\mathbf{Q}_i$ and the scalar w_i as,

$$\mathbf{Q_i} = (\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i + \mathbf{x}^*\mathbf{x}^{*\prime})^{-1}(\hat{\boldsymbol{\beta}}_i'(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i) + y^*\mathbf{x}^{*\prime})', \tag{B.14}$$

and

$$\begin{aligned} w_i &= y^{*\prime}y^* + \hat{\boldsymbol{\beta}}_i'(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i + \mathbf{x}^*\mathbf{x}^{*\prime})\hat{\boldsymbol{\beta}}_i + S_i \\ &- (\hat{\boldsymbol{\beta}}_i'(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i) + y^*\mathbf{x}^{*\prime})(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i + \mathbf{x}^*\mathbf{x}^{*\prime})^{-1}(\hat{\boldsymbol{\beta}}_i'(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i) + y^*\mathbf{x}^{*\prime})'. \end{aligned} \tag{B.15}$$

Then by completing the squares, we get

$$\Omega_i = w_i + (\boldsymbol{\beta}_i - \mathbf{Q_i})'(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i + \mathbf{x}^*\mathbf{x}^{*\prime})(\boldsymbol{\beta}_i - \mathbf{Q_i}). \tag{B.16}$$

Thus, we have that the posterior predictive density is of the form:

$$P(y^*|M_i, \mathbf{x}^*, \mathbf{y}) \propto k_{2,i} \int_{\boldsymbol{\beta}_i} \frac{d\boldsymbol{\beta}_i}{[w_i + (\boldsymbol{\beta}_i - \mathbf{Q_i})'(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i + \mathbf{x}^*\mathbf{x}^{*\prime})(\boldsymbol{\beta}_i - \mathbf{Q_i})]^{\frac{n+t_i+1}{2}}}. \tag{B.17}$$

The integral in the above equation is a matrix T-distribution (see, e.g., [57]), and thus integrates to a constant which is a function of y^*, $\mathbf{x}^*$ and M_i. We include all the constant terms that are independent of y^* in constant $k_{3,i}$, and rewrite the above equation by including only the constant term that includes y^* as,

$$P(y^*|M_i, \mathbf{x}^*, \mathbf{y}) \propto \frac{k_{3,i}}{w_i^{n/2}}. \tag{B.18}$$

Using a well-known matrix identity (see [57]), we rewrite w_i as

$$w_i = S_i + \frac{(y^* - \mathbf{x}^*\hat{\boldsymbol{\beta}}_i)^2}{1 + \mathbf{x}^{*\prime}(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^*}. \tag{B.19}$$

We can integrate the joint posterior, $P(\boldsymbol{\beta}_i, \sigma^2|\mathbf{y}, M_i)$, over σ^2, to obtain the marginal posterior distribution of $\boldsymbol{\beta}_i$. This gives

$$\boldsymbol{\beta}_i|\sigma^2, \mathbf{y}, M_i \sim N\left(\hat{\boldsymbol{\beta}}_i, \sigma^2(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\right). \tag{B.20}$$

Thus, we have that

$$Z_i = \frac{y^* - \mathbf{x}^*\hat{\boldsymbol{\beta}}_i}{\sigma\sqrt{1 + \mathbf{x}^{*\prime}(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^*}} \sim N(0,1). \tag{B.21}$$

Similarly, by integrating the joint posterior $P(\boldsymbol{\beta}_i, \sigma^2|\mathbf{y}, M_i)$ over $\boldsymbol{\beta}_i$, it can be shown that the marginal posterior distribution of σ^2 is given by

$$\frac{\sigma^2}{S_i} \sim inv{-}\chi^2_{n-1}, \tag{B.22}$$

(an inverse chi-square distribution) or in other words,

$$\frac{S_i}{\sigma^2} \sim \chi^2_{n-1}. \tag{B.23}$$

If $\hat{\sigma}_i^2 = S_i/(n-1)$,

$$W_i = \frac{(n-1)\hat{\sigma}_i^2}{\sigma^2} \sim \chi^2_{n-1}. \tag{B.24}$$

Thus, if $\mu_y^* = \mathbf{x}^*\hat{\boldsymbol{\beta}}_i$, and $\sigma_y^{*2} = \hat{\sigma}_i^2\left[1 + \mathbf{x}^{*\prime}(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^*\right]$, then equation (B.19) can be written as,

$$w_i = S_i\left[1 + \frac{1}{n-1}\frac{(y^* - \mu_y^*)^2}{\sigma_y^{*2}}\right]. \tag{B.25}$$

Thus, from equation (B.18),

$$P(y^*|M_i, \mathbf{x}^*, \mathbf{y}) \propto \left[1 + \frac{1}{n-1}\frac{(y^* - \mu_y^*)^2}{\sigma_y^{*2}}\right]^{-n/2} \tag{B.26}$$

The density above is a Student t with mean μ_y^*, and scale parameter σ_y^{*2}, with $(n-1)$ degrees of freedom. That is, the posterior predictive density is

$$y^*|M_i, \mathbf{x}^*, \mathbf{y} \propto t_{n-1}\left(\mathbf{x}^*\hat{\boldsymbol{\beta}}_i, \hat{\sigma}_i^2\left[1 + \mathbf{x}^{*\prime}(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^*\right]\right). \tag{B.27}$$

The cumulative posterior predictive density of the response, given the model and the data, at a given level of control factors can be computed using the c.d.f. of a t_ν distribution with $\nu = n - 1$. This can easily be computed using the following identity (see [34]):

$$P\left(\frac{y^* - \mu_y^*}{\sigma_y^*} < t|M_i, \mathbf{x}^*, \mathbf{y}\right) = \frac{1}{2}\left[1 + I_{\frac{t^2}{\nu+t^2}}\left(\frac{1}{2}, \frac{\nu}{2}\right)\right]. \tag{B.28}$$

Corollary: For mixture models (as in example 1 of section 3.3.1), or models without intercept, the models are fitted without a constant term. In this case, we assume a Zellner's g-prior [74] on all the β_i in the models. Thus, for a single response process with k controllable factors, under the process model of the form given in equation (3.1) with normally distributed error terms, and under the priors on the factors and the parameters given by equations (3.4), (3.8), (3.9) and (3.10) with $\boldsymbol{\Sigma}^{-1} = (\mathbf{X}_i'\mathbf{X}_i)\mathbf{V_i}$, where $\mathbf{V_i} = \frac{1}{g}\boldsymbol{I}_{t_i}$:

1. The posterior predictive density is

$$y^*|M_i, \mathbf{x}^*, \mathbf{y} \propto t_n\left(\mathbf{x}^*\hat{\boldsymbol{\beta}}_i, \hat{\sigma}_i^2\left[1 + \mathbf{x}^{*\prime}(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^*\right]\right), \tag{B.29}$$

where

$$\hat{\sigma}_i^2 = \frac{S_i}{n}, \tag{B.30}$$

and

$$\frac{n\hat{\sigma}_i^2}{\sigma^2} = \frac{S_i}{\sigma^2} \sim \chi_n^2. \tag{B.31}$$

2. The cumulative posterior predictive density for the new response y^* at a new observation $\mathbf{x}^*$ for a given model, M_i, is

$$P\left(\frac{y^* - \mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}_i}{\hat{\sigma}_i\sqrt{1 + \mathbf{x}^{*\prime}(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^*}} < t | M_i, \mathbf{x}^*, \mathbf{y}\right) = \frac{1}{2}\left[1 + I_{\frac{t^2}{\nu+t^2}}\left(\frac{1}{2}, \frac{\nu}{2}\right)\right], \tag{B.32}$$

where $I_z(a, b)$ is the incomplete beta function, $\nu = n$, and $\hat{\sigma}_i^2 = S_i/n$.

Here the only difference in the computation of the cumulative posterior predictive density is in the degrees of freedom of the t distribution.

Appendix C

Derivation of Posterior Predictive Density: Multiple Response Case

Theorem: For a multiple response process with k controllable factors, under the standard multivariate regression (SMR)process model shown in equation (5.5) with normally distributed error terms, and under the priors on the parameters given by equations (5.6), (5.7) and (5.8), the posterior predictive density for the new vector of responses $\mathbf{y}^*$ at a new set of regressors $\mathbf{x}^*$ for a given model M_i is given by:

$$\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}, M_i \propto \mathbf{T}_q(\hat{\boldsymbol{\beta}}_i'\mathbf{x}^*, \mathbf{H}_i^{-1}) \tag{C.1}$$

where,

$$\mathbf{H}_i = \frac{\nu_i \mathbf{S_i}^{-1}}{1 + \mathbf{x}^{*\prime}(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^*},$$

and

$$\nu_i = n + t_i + 1 - p_i - q.$$

where t_i,p_i,q,n,$\mathbf{S}_i$, $\mathbf{Y}$, $\mathbf{X}_i$, $\hat{\boldsymbol{\beta}}_i$, $\boldsymbol{\Sigma}_i^{-1}$ are as defined in chapter 5.

Proof: The proof for the multiple response case follows very similar steps as in the single response case. In this proof also, the results are for a given model form M_i. However, for the sake of notational convenience the subscript i that was used in the proof for single response systems has been omitted and it will be assumed that all the conditional probabilities are for the given model.

It is assumed that the process follows a Standard Multivariate Regression (SMR) model given by

$$\mathbf{y} = \boldsymbol{\beta}'\mathbf{x} + \epsilon, \tag{C.2}$$

where $\mathbf{y}$ is the $(q \times 1)$ vector of responses, $\boldsymbol{\beta}$ is the $(p \times q)$ matrix of model parameters, $\mathbf{x}$ is the $(p \times 1)$ vector of regressors and $\boldsymbol{\epsilon}$ is the correlated error term with distribution $N(\mathbf{0}, \boldsymbol{\Sigma}_e)$. Assuming $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_e^{-1}$, we have $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Lambda}^{-1})$. As in the single response case it is assumed that there is data from an experiment with n runs. Denote the $(n \times p)$ design matrix by $\mathbf{X}$, and the $(n \times q)$ matrix of responses from the experiment by $\mathbf{Y}$.

Here, for a future set of responses $\mathbf{y}^*$ at a given future setting of the regressors $\mathbf{x}^*$, the posterior predictive density is defined as,

$$P(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}) = \int_{\boldsymbol{\Lambda}} \int_{\boldsymbol{\beta}} P(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\beta}) P(\boldsymbol{\beta}, \boldsymbol{\Lambda}|\mathbf{Y}) d\boldsymbol{\beta} d\boldsymbol{\Lambda}, \tag{C.3}$$

where $P(\boldsymbol{\beta}, \boldsymbol{\Lambda}|\mathbf{Y})$ is the joint posterior of the model parameters, and $P(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\beta})$ is the likelihood function given by

$$P(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\beta}) \propto |\boldsymbol{\Lambda}|^{1/2} \mathrm{e}xp\left[-\frac{1}{2}(\mathbf{y}^* - \boldsymbol{\beta}'\mathbf{x}^*)'\boldsymbol{\Lambda}(\mathbf{y}^* - \boldsymbol{\beta}'\mathbf{x}^*)\right]. \tag{C.4}$$

The priors used on the parameters are similar to those used in the single response case. Vague priors are assumed for $\boldsymbol{\Lambda}$ and for the $\boldsymbol{\beta}$ for the constant term, and the Zellner's

g-prior [74] is assumed for the remaining $\boldsymbol{\beta}$. Therefore, the prior probabilities are given by

$$P(\boldsymbol{\Lambda}) \propto \frac{1}{|\boldsymbol{\Lambda}|^{(q+1)/2}}, \tag{C.5}$$

$$P(\boldsymbol{\beta}) \propto |\boldsymbol{\Lambda}|^{t/2} exp\left[-\frac{1}{2}\mathrm{t}r\, \boldsymbol{\Lambda}(\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})\right], \tag{C.6}$$

and

$$P(\boldsymbol{\beta}, \boldsymbol{\Lambda}) \propto P(\boldsymbol{\beta})P(\boldsymbol{\Lambda}), \tag{C.7}$$

where t is the number of terms in the model excluding the constant term (i.e., if the model includes a constant term, then $t = p - 1$ and if the model excludes a constant term then $t = p$) and $\boldsymbol{\Sigma}^{-1} = (\mathbf{X}'\mathbf{X})\mathbf{V}$, where $\mathbf{V} = \frac{1}{g}\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_t \end{pmatrix}$. The joint posterior of the model parameters is given by the Bayes' theorem:

$$P(\boldsymbol{\beta}, \boldsymbol{\Lambda}|\mathbf{Y}) \propto P(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\Lambda})P(\boldsymbol{\beta}, \boldsymbol{\Lambda}), \tag{C.8}$$

where $P(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\Lambda})$ is the likelihood given by

$$P(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\Lambda}) \propto |\boldsymbol{\Lambda}|^{n/2} exp\left[-\frac{1}{2}\mathrm{t}r\, \boldsymbol{\Lambda}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right]. \tag{C.9}$$

The joint posterior of the model parameters can thus be written as

$$P(\boldsymbol{\beta}, \boldsymbol{\Lambda}|\mathbf{Y}) \propto |\boldsymbol{\Lambda}|^{(n+t-q-1)/2} exp\left[-\frac{1}{2}\mathrm{t}r\, \boldsymbol{\Lambda}\mathbf{W}\right], \tag{C.10}$$

where the $(q \times q)$ matrix $\mathbf{W}$ is given by

$$\mathbf{W} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + (\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}).$$

Substituting equations (C.4) and (C.10) in equation (C.3), the posterior predictive density can be written as

$$P(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}) = \int_{\boldsymbol{\Lambda}} \int_{\boldsymbol{\beta}} |\boldsymbol{\Lambda}|^{(n+t-q)/2} exp\left[-\frac{1}{2}\mathrm{t}r\, \boldsymbol{\Lambda}\mathbf{A}\right] d\boldsymbol{\beta} d\boldsymbol{\Lambda}, \tag{C.11}$$

where $(q \times q)$ matrix $\mathbf{A}$ is given by

$$\begin{aligned}\mathbf{A} &= W + (\mathbf{y}^* - \boldsymbol{\beta}'\mathbf{x}^*)(\mathbf{y}^* - \boldsymbol{\beta}'\mathbf{x}^*)' \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + (\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}) + (\mathbf{y}^* - \boldsymbol{\beta}'\mathbf{x}^*)(\mathbf{y}^* - \boldsymbol{\beta}'\mathbf{x}^*)'.\end{aligned}$$

In equation (C.11) the integral over $\boldsymbol{\Lambda}$ is in the form of the Wishart distribution (see [57]), and thus

$$P(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}) \propto \int_{\boldsymbol{\beta}} \frac{d\boldsymbol{\beta}}{|\mathbf{A}|^{(n+t+1)/2}}. \tag{C.12}$$

Substituting for $\mathbf{A}$,

$$P(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}) \propto \int_{\boldsymbol{\beta}} \frac{d\boldsymbol{\beta}}{|W + (\mathbf{y}^* - \boldsymbol{\beta}'\mathbf{x}^*)(\mathbf{y}^* - \boldsymbol{\beta}'\mathbf{x}^*)'|^{(n+t+1)/2}}. \tag{C.13}$$

As in the proof of the single response systems, $\mathbf{W}$ can be rewritten as,

$$\begin{aligned}W &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + (\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}), \\ &= \mathbf{S} + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'(\boldsymbol{\Sigma}^{-1} + \mathbf{X}'\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}),\end{aligned}$$

where

$$\mathbf{S} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\beta}}), \tag{C.14}$$

and

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{\Sigma}^{-1} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \tag{C.15}$$

By completing the squares on $\boldsymbol{\beta}$ and using well-known matrix identities [57], equation (C.13) can be written as

$$P(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}) \propto \int_{\boldsymbol{\beta}} \frac{d\boldsymbol{\beta}}{|\mathbf{C}_3 + (\boldsymbol{\beta}' - \mathbf{C}_1)\mathbf{C}_2(\boldsymbol{\beta}' - \mathbf{C}_1)'|^{(n+t+1)/2}}, \tag{C.16}$$

where

$$\mathbf{C}_1 = (\hat{\boldsymbol{\beta}}'\mathbf{D} + \mathbf{y}^*\mathbf{x}^{*\prime})(\mathbf{D} + \mathbf{x}^*\mathbf{x}^{*\prime})^{-1},$$

$$\mathbf{C}_2 = \mathbf{D} + \mathbf{x}^*\mathbf{x}^{*\prime},$$

$$\mathbf{C}_3 = \mathbf{S} + \frac{(\mathbf{y}^* - \hat{\boldsymbol{\beta}}'\mathbf{x}^*)(\mathbf{y}^* - \hat{\boldsymbol{\beta}}'\mathbf{x}^*)'}{1 + \mathbf{x}^{*\prime}\mathbf{D}^{-1}\mathbf{x}^*},$$

and

$$\mathbf{D} = \boldsymbol{\Sigma}^{-1} + \mathbf{X}'\mathbf{X}.$$

The integrand in equation (C.16) is in the form of a matrix-$\mathbf{T}$ distribution, and therefore

$$P(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}) \propto \frac{1}{|\mathbf{C}_3|^{(n+t+1-p)/2}}, \tag{C.17}$$

which is the density of a multivariate student $\mathbf{T}$-distribution of dimension q. Thus,

$$\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y} \propto \mathbf{T}_q(\hat{\boldsymbol{\beta}}'\mathbf{x}^*, \mathbf{H}^{-1}) \tag{C.18}$$

where,

$$\mathbf{H} = \frac{\nu\mathbf{S}^{-1}}{1 + \mathbf{x}^{*\prime}\mathbf{D}^{-1}\mathbf{x}^*}, \tag{C.19}$$

and

$$\nu = n + t + 1 - p - q. \tag{C.20}$$

Corollary 1: For models that include a constant term, the number of terms in the model excluding the constant term denoted by t is one less than the total number of parameters p. Thus $t = p - 1$. In this case, the posterior predictive density is given by equations (C.18-C.20), where $\nu = n + t + 1 - p - q = n - q$.

Corollary 2: For models that are fitted without a constant term such as mixture models, we assume a Zellner's g-prior [74] on all the parameters $\boldsymbol{\beta}$. In this case, the number of terms in the model excluding the constant term is the same as the total number of parameters, i.e., $t = p$. In this case, the posterior predictive density is given by equations (C.18-C.20), where $\nu = n + t + 1 - p - q = n + 1 - q$.

Appendix D

Review of Non-linear Optimization Techniques Used in this Thesis

The nonlinear optimization problems in the examples in this dissertation were solved using MATLAB's *fmincon* routine, which can handle both linear and nonlinear constraints. The routine uses a sequential quadratic programming (SQP) algorithm [42] for medium-scale optimization. The basic idea of the algorithm is to approximate the Hessian of the Lagrangian function at each iteration using a Quasi-Newton updating method. The subroutine also estimates the gradient of the function at each iteration when the gradient information is not readily available such as in the examples in this dissertation. The gradient is typically estimated using a finite difference method [42]. This generates a quadratic programming (QP) subproblem that is solved in order to find a search direction for a line search procedure.

The routine solves optimization problems that can be formulated as shown below:

$$\min_{\mathbf{x}} \; f(\mathbf{x})$$

$$s.t.$$

$$\begin{aligned} g_i(\mathbf{x}) &= 0 \;\; i = 1...m_e \\ g_i(\mathbf{x}) &\leq 0 \;\; i = m_e + 1...m \\ \mathbf{x}_l &\leq \mathbf{x} \leq \mathbf{x}_u \end{aligned}$$

where g_i may be linear or nonlinear, $\mathbf{x}_l$ is the vector of lower bounds on $\mathbf{x}$ and $\mathbf{x}_u$ is the vector of upper bounds. The Lagrangian function is then given by

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}). \tag{D.1}$$

The algorithm starts with an initial solution x_0 set by the user and then iteratively converges to the optimum. It is important that this starting point is a feasible solution. As mentioned in this dissertation, for the kind of problems that are solved here, it is important that the optimization program is solved by using multiple starting points so that the solution obtained is not local. Two methods were used in the examples in this dissertation to choose these starting points. The first method was to generate the starting points by choosing randomly from a uniform distribution in the space of $\mathbf{x}$. The second method was to select starting points that form an evenly-spaced grid in the space of $\mathbf{x}$.

If the value of $\mathbf{x}$ at the k^{th} iteration of the optimization algorithm is $\mathbf{x}_k$, then the new iterate is given by

$$x_{k+1} = x_k + \mathbf{d}_k, \tag{D.2}$$

where, $\mathbf{d}_k$ is the search direction. Here, $\mathbf{d}_k$ is obtained at each iteration by solving the

following quadratic programming subproblem:

$$\min_{\mathbf{d}} \ \frac{1}{2}\,\mathbf{d}'\mathbf{H}_k\mathbf{d} + \nabla f(\mathbf{x}_k)'\mathbf{d}$$

$$s.t.$$

$$\begin{aligned} \nabla g_i(\mathbf{x}_k)'\mathbf{d} + g_i(\mathbf{x}_k) &= 0 \ \ i = 1...m_e \\ \nabla g_i(\mathbf{x}_k)'\mathbf{d} + g_i(\mathbf{x}_k) &\leq 0 \ \ i = m_e + 1...m, \end{aligned}$$

where the Hessian H_k is updated at each iteration using the BFGS method [52]. The BFGS method updates the Hessian using an estimate of the lagrangian multipliers $\boldsymbol{\lambda}_k$ at each iteration. After each iteration k, the value of $\boldsymbol{\lambda}_{k+1}$ is set to $\boldsymbol{\mu}_k$, where $(\mathbf{d}_k, \boldsymbol{\mu}_k)$ is the unique solution to the quadratic subproblem above. In other words, $\boldsymbol{\mu}_k$ is the value of the lagrange multiplier at the optimal solution of the quadratic programming subproblem. Thus the solution of the quadratic programming subproblem provides the direction $\mathbf{d}$ for the line search procedure used to update $\mathbf{x}$. The quadratic programming subproblem is also solved using an iterative method, where first a feasible point is calculated if the current point from the SQP is not feasible, and then an iterative sequence of feasible points is generated that converges to a solution [52, 42]. At each step in the QP subproblem, the active set or the set of binding constraints at the current solution point is used to form a basis for a search direction.

Vita

Ramkumar Rajagopal was born on September 23, 1976, in Madras, India. He received a Bachelor's degree in Chemical Engineering from the Indian Institute of Technology (I.I.T.) at Bombay, India, in 1998. He received his Masters in Industrial Engineering and Operations Research at The Pennsylvania State University, University Park PA, in the year 2000 under the guidance Dr. Enrique Del Castillo. His Masters thesis was titled "Analysis and Multivariate Extensions of Double EWMA Feedback Adjustment Scheme for Quality Control", resulting in two journal publications. After his Masters, he worked as an *Associate Product Manager* for a marketing optimization software in Virginia. He then resumed his graduate studies working towards a Ph.D. in Spring 2000. Four journal papers have been submitted based on the research presented in this dissertation. After graduation, Ramkumar Rajagopal will be joining Intel Corporation where he will be working on Process Control and Process Optimization problems.