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

Ramkumar Rajagopal
Enrique del Castillo
Dept. of Industrial & Manufacturing Engineering
The Pennsylvania State University
University Park, PA 16802

May, 2004

Abstract

The usual approach to solving the Multiple Criteria Decision Making (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 a MCDM problem 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 advantages of using the Bayesian approach as opposed to traditional methods are highlighted. The methodology is applied to engineering design problems, providing a rigorous formulation to popular "Design for Six Sigma" approaches.

KEYWORDS: Regression, Quality Control, Multivariate Statistics, Multiple Criteria Analysis, Response Surface Methodology.

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 we consider, 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 paper proposes a new algorithm to solve MCDM problems based on a Bayesian methodology that adopts a probabilistic approach to solving the MCDM problem. This paper is organized as follows: section 2 introduces some useful results behind the Bayesian methodology, section 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 4 then extends the methodology to the case of multiple decision makers. Examples are provided in both section 3 and 4. A summary of the approach is discussed in section 5.

1.1 Merits of the Bayesian Methodology

In the traditional sense of optimization, response models are first fit to the data as functions of the controllable factors (independent variables), and then the models are optimized to find the setting of the factors that give the desired values of the responses. However, this optimal setting does not allow us to predict what fraction of future responses 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 responses). In other words, there can be no inference made about the conformance to specifications. A natural way to circumvent this problem is to maximize the probability of conformance of the predicted responses to their specification limits [8]. 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.

Although both approaches are probabilistic, the Bayesian 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 [11]). Other non-Bayesian 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 [10]. Information collected from the decision maker is then used to identify the operating point on the efficient frontier. More recently, there have been developments in the application of Bayesian methods to certain MCDM problems. Hahn [4] uses Bayesian inferencing to derive priorities in the Analytic Hierarchy

Process (AHP).

2 Useful results in Bayesian and Bayesian Predictive Inference

This section contains some preliminary results useful in the Bayesian MCDM methodology of later sections.

2.1 Bayes' Theorem

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 [1]:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{l(\boldsymbol{\theta}|\mathbf{y})p(\boldsymbol{\theta})}{p(\mathbf{y})}, \tag{1}$$

where $l(\boldsymbol{\theta}|\mathbf{y})$ is the likelihood function of $\boldsymbol{\theta}$ for given $\mathbf{y}$. Thus, Bayes' theorem says that the distribution of $\boldsymbol{\theta}$ posterior to observing 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.

2.2 Bayesian Predictive Density

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 we can fit the following model to the response:

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \tag{2}$$

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 [9]):

$$p(y^*|\mathbf{x}^*, \mathbf{y}) = \int_{\sigma^2}\int_{\boldsymbol{\beta}} p(y^*|\mathbf{x}^*, \mathbf{y}, \boldsymbol{\beta}, \sigma^2)p(\boldsymbol{\beta}, \sigma^2|\mathbf{y})\, d\boldsymbol{\beta}\, d\sigma^2, \tag{3}$$

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 as follows:

$$p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2)p(\boldsymbol{\beta}, \sigma^2), \tag{4}$$

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,

$$p(\boldsymbol{\beta}) \propto constant, \tag{5}$$

$$p(\sigma^2) \propto \frac{1}{\sigma^2}, \tag{6}$$

and,

$$p(\boldsymbol{\beta}, \sigma^2) = p(\boldsymbol{\beta})p(\sigma^2), \tag{7}$$

the posterior predictive density is given by a t-distribution (see [9]). That is,

$$y^*|\mathbf{x}^*, \mathbf{y} \sim t_\nu(\mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}, \hat{\sigma}\sqrt{1 + \mathbf{x}^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^*}), \tag{8}$$

where $\nu = n - p$,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{9}$$

and,

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{n - p}. \tag{10}$$

From equation (8), 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 of the response at $\mathbf{x}^*$ is given by

$$E(y^*|\mathbf{x}^*, \mathbf{y}) = \mu_{y^*} = \mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}, \tag{11}$$

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

$$Var(y^*|\mathbf{x}^*, \mathbf{y}) = \frac{\nu}{\nu - 2}\sigma^2_{y^*} = \frac{\nu}{\nu - 2}\hat{\sigma}^2\left(1 + \mathbf{x}^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^*\right). \tag{12}$$

2.2.1 Multiple Responses

For multiple responses, it is of interest to maximize the posterior probability of conformance, $p(\mathbf{y}^* \in \mathbf{A})$, where $\mathbf{A}$ is a region of interest defined by the individual specifications, $[l_i, u_i]$, on each response. Here, each of the q responses is modelled as

$$y_i = \mathbf{x}_i'\boldsymbol{\beta}_i + \epsilon_i. \tag{13}$$

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. Thus, given the data $\mathbf{Y}$,

$$p(\mathbf{y}^* \in \mathbf{A}|x_1^*...x_k^*, \mathbf{Y}) \equiv p(y_1^* \in [l_1, u_1], y_2^* \in [l_2, u_2]...y_q^* \in [l_q, u_q]|x_1^*...x_k^*, \mathbf{Y}) \quad (14)$$

$$= \prod_{i=1}^{q} p(y_i^* \in [l_i, u_i]|x_1^*...x_k^*, \mathbf{Y}), \quad (15)$$

where each of the marginal posteriors is given by the t-distribution shown in equation (8). 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 $\mathbf{T}$-distribution [9]. 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

$$\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y} \sim T_\nu^q(\mathbf{B}'\mathbf{x}^*, \mathbf{H}^{-1}), \quad (16)$$

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

$$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \quad (17)$$

$$\mathbf{H} = \frac{\nu \mathbf{S^{-1}}}{1 + \mathbf{x}^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^*}, \quad (18)$$

and

$$\mathbf{S} = (\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB}). \quad (19)$$

For the SUR case, the posterior predictive density has no closed form but can be computed numerically. In particular, Percy [6] shows how a the posterior predictive distribution of a new observation $\mathbf{y}^*$ can be approximated using Gibbs sampling.

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 we can fit a model to the responses of the form,

$$y_j = \mathbf{x}_j'\boldsymbol{\beta}_j + \epsilon_j, \quad \forall j \in \{1, ..., q\} \quad (20)$$

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:

$$u = \mathbf{y}_f'\boldsymbol{\gamma} + \epsilon, \tag{21}$$

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 2.2), $\boldsymbol{\gamma}$ 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 (21). 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 (21). 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 (21) given by,

$$p(\boldsymbol{\gamma}) \propto constant, \tag{22}$$

$$p(\sigma_f^2) \propto \frac{1}{\sigma_f^2}, \tag{23}$$

and,

$$p(\boldsymbol{\gamma}, \sigma_f^2) = p(\boldsymbol{\gamma})p(\sigma_f^2), \tag{24}$$

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 (8) and is given by:

$$u^*|\mathbf{y}_f^*, \mathbf{u} \sim t_{\nu_f}(\mu_{u^*}, \sigma_{u^*}^2), \tag{25}$$

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,

$$\hat{\boldsymbol{\gamma}} = (\mathbf{Y}_f'\mathbf{Y}_f)^{-1}\mathbf{Y}_f'\mathbf{u}, \tag{26}$$

and,

$$\hat{\sigma}_f^2 = \frac{(\mathbf{u} - \mathbf{Y}_f\boldsymbol{\gamma})'(\mathbf{u} - \mathbf{Y}_f\boldsymbol{\gamma})}{m - p_f}. \tag{27}$$

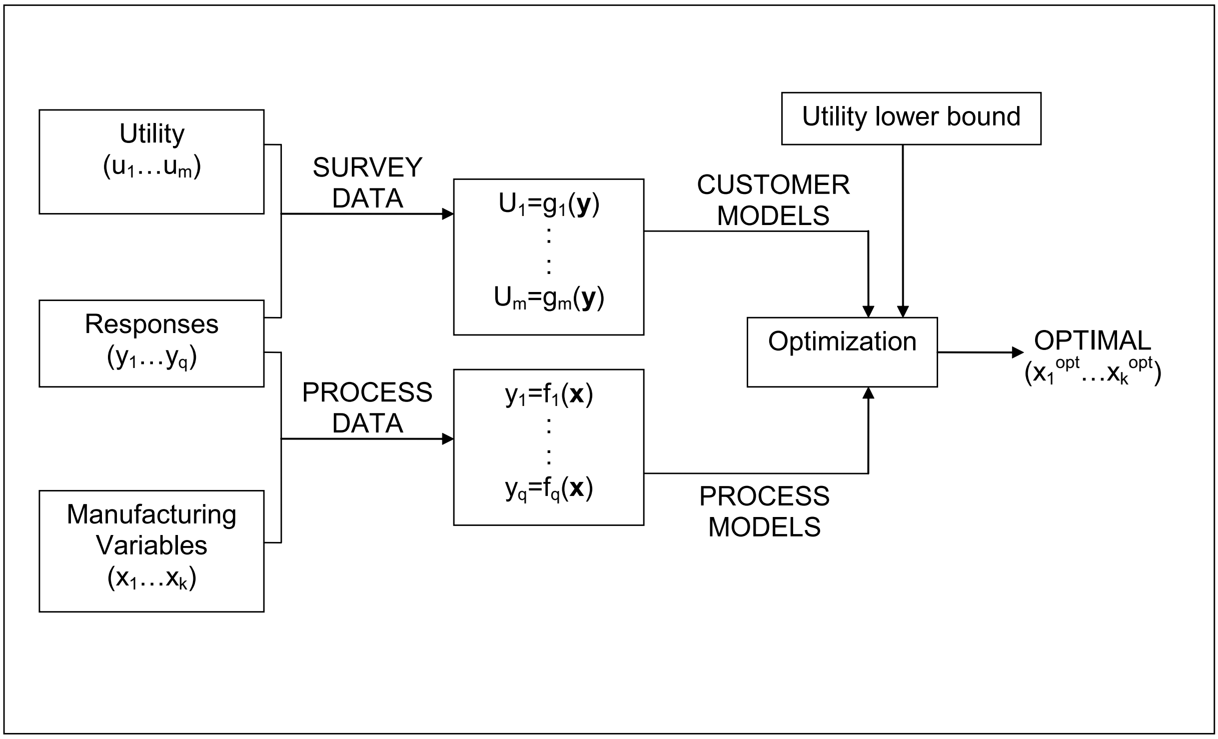

Figure 1: Block diagram of the proposed Bayesian MCDM method

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

$$E(u^*|\mathbf{y}_f^*, \mathbf{u}) = \mu_{u^*} = \mathbf{y}_f^{*\prime}\hat{\boldsymbol{\gamma}}, \tag{28}$$

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

$$Var(u^*|\mathbf{y}_f^*, \mathbf{u}) = \frac{\nu_f}{\nu_f - 2}\sigma_{u^*}^2 = \frac{\nu_f}{\nu_f - 2}\hat{\sigma_f}^2 \left(1 + \mathbf{y}_f^{*\prime}(\mathbf{Y}_f^\prime \mathbf{Y}_f)^{-1}\mathbf{y}_f^*\right). \tag{29}$$

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 1).

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 responses. For each possible outcome of the responses, the distribution of the customer's utility u follows the distribution shown in equation (25). 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 (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), that is

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[p(u^* \geq l_u|\mathbf{y}^*(count), \mathbf{u})\right] = E_{\mathbf{y}^*}\left[p(u^* \geq l_u|\mathbf{y}^*, \mathbf{u})\right]. \tag{30}$$

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 below illustrates the proposed methodology.

3.2 Application to Six Sigma Manufacturing

Six Sigma methodologies are gaining popularity in industries for managing quality [2]. 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.

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 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 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.

3.3 Example: CVD Process

This example uses data from Czitrom and Spagon [3] 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 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

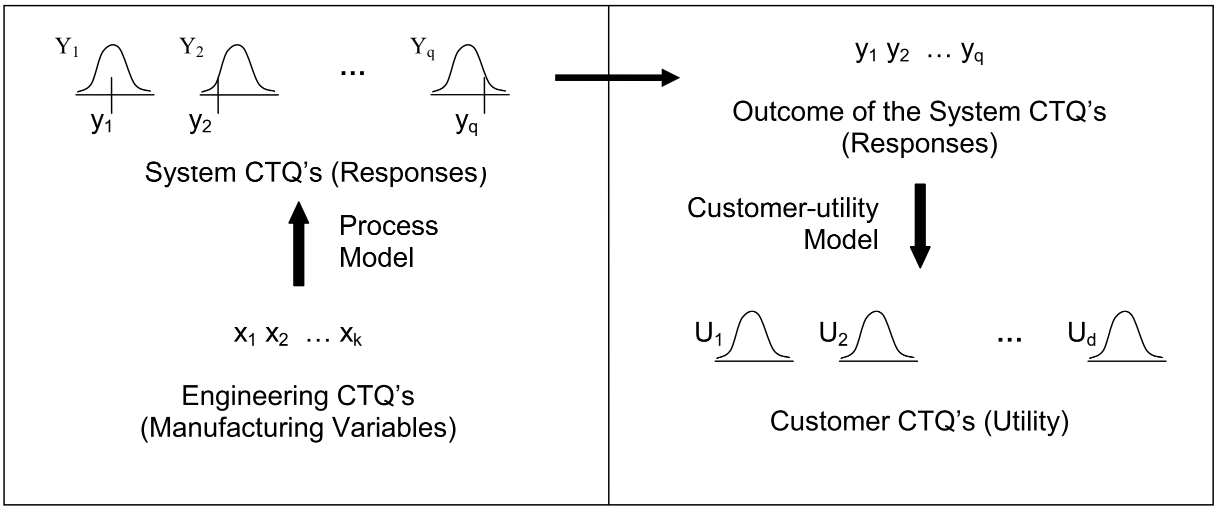

Figure 2: Block diagram of the Design for Six Sigma approach

for modelling. Based on the coded data, the model obtained for the responses using the ordinary least square estimates shown in equation (9) are given by:

$$\hat{y}_1 = 5.8661 - 1.9097x_1 - 0.2241x_2 + 1.6862x_1x_2 + 0.1337x_1^2 + 0.0337x_2^2, \tag{31}$$

$$\hat{y}_2 = 7.7900 + 0.7359x_1 + 0.4969x_2 + 0.0694x_1x_2 - 0.5287x_1^2 - 0.1187x_2^2. \tag{32}$$

Table 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 1 as -1, and the largest observed value as 1. Table 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 survey by comparing the results between the different surveys.

For each survey, i, a quadratic model of the form shown in equation (21) is fitted to the utility u_i from the data:

$$\hat{u}_1 = 6.00 - 2.61y_1 - 1.58y_2 + 0.25y_1y_2 - 0.25y_1^2 - 0.50y_2^2, \tag{33}$$

$$\hat{u}_2 = 4.00 - 2.36y_1 - 0.72y_2 + 0.25y_1y_2 + 0.37y_1^2 + 0.12y_2^2, \tag{34}$$

$$\hat{u}_3 = 2.00 - 0.85y_1 - 2.81y_2 + 0.87y_1^2 + 1.87y_2^2, \tag{35}$$

$$\hat{u}_4 = 2.00 - 0.42y_1 - 3.51y_2 + 1.06y_1^2 + 1.31y_2^2. \tag{36}$$

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 1: Data for CVD process example [3]

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 2: Data for 4 different surveys for CVD example

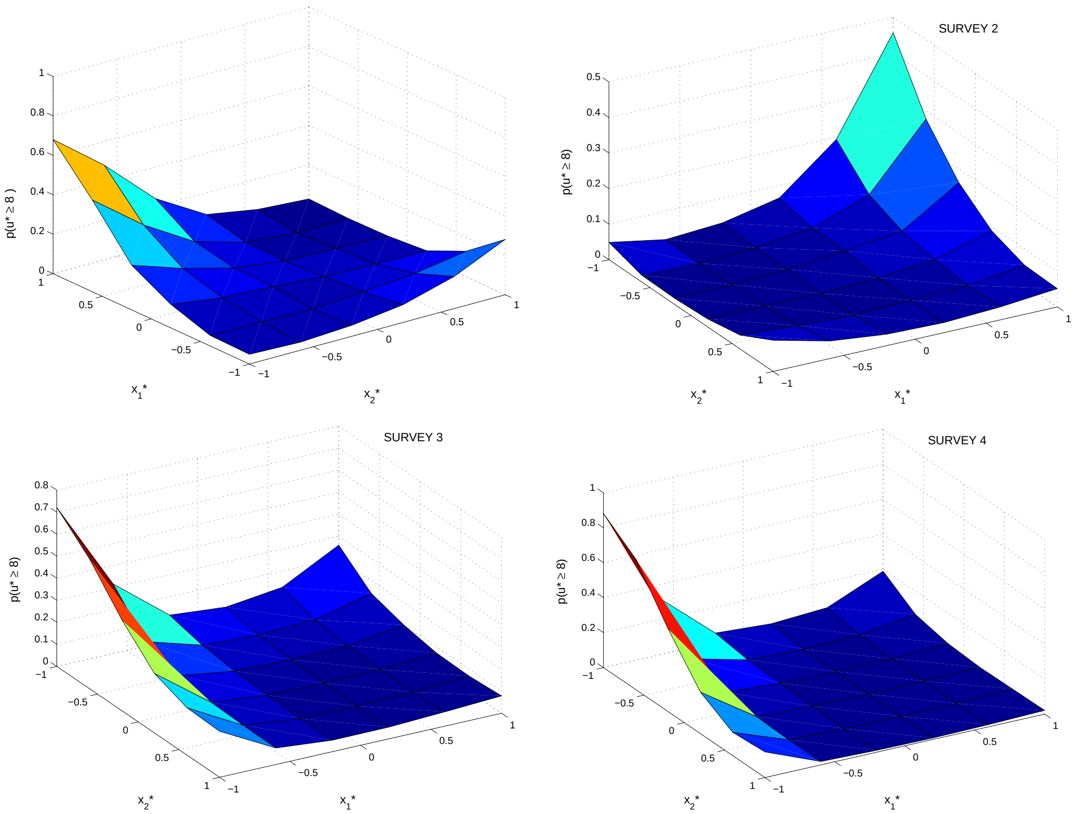

Figure 3: Surface plot of $p(u^* \geq 8)$ for different values of $\mathbf{x}^*$ for CVD example

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 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 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.

3.4 Example: HPLC Process

The data for this example is taken from [7] and is presented in table 4. There are four responses in the high performance liquid chromatography (HPLC) system namely, the critical

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 3: Optimization results for the 4 surveys

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, we assume 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 (17 - 19). As the model is SMR, the joint posterior distribution of the responses is given by a multivariate **T**-distribution as shown in equation (16), from which the responses are sampled in step 3 of the optimization process described in section 3.1.

Table 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 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,

$$\hat{u} = 14 + y_1 - 4.25y_2 + 6.50y_3 + 0.75y_4 - y_2y_3 - 2.75y_1^2 + 1.25y_2^2 - 0.25y_3^2 - 6.25y_4^2. \quad (37)$$

Figure 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 4: Data for HPLC process example [7]

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
1	2.39	19	350.5	0.8425	0.5	0.5	0.5	0.5	14

Table 5: Data for 4 different surveys for HPLC example

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 posterior predictive density of the responses, as described in section 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 (21). The model for the utility function of the i^{th} DM

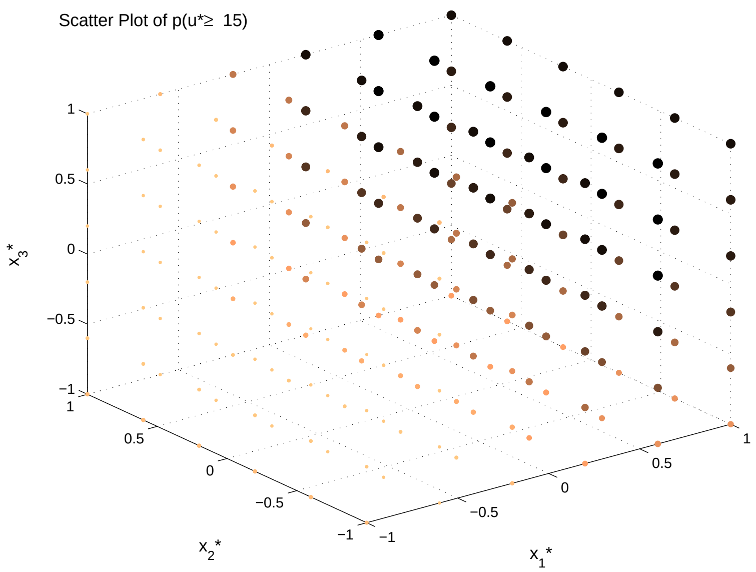

Figure 4: Surface plot of $p(u^* \geq 15)$ for different values of $\mathbf{x}^*$ for HPLC example

is given by:

$$u_i = \mathbf{y}'_{f_i}\boldsymbol{\gamma}_i + \epsilon_i, \tag{38}$$

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 (21) for all $i \in \{1...d\}$ as shown in equations (5-7), the posterior predictive density of the i^{th} DM follows a t-distribution, i.e.,

$$u^*_i|\mathbf{y}^*_{f_i}, \mathbf{u}_i \sim t_{\nu_{f_i}}(\mu_{u^*_i}, \sigma^2_{u^*_i}), \tag{39}$$

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

$$\hat{\boldsymbol{\gamma}}_i = (\mathbf{Y}'_{f_i}\mathbf{Y}_{f_i})^{-1}\mathbf{Y}'_{f_i}\mathbf{u}_i, \tag{40}$$

and,

$$\hat{\sigma}^2_{f_i} = \frac{(\mathbf{u}_i - \mathbf{Y}_{f_i}\boldsymbol{\gamma}_i)'(\mathbf{u}_i - \mathbf{Y}_{f_i}\boldsymbol{\gamma}_i)}{m - p_{f_i}}. \tag{41}$$

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.,

$$p(u^*_1 \geq l_{u_1}, u^*_2 \geq l_{u_2}...u^*_d \geq l_{u_d}|\mathbf{y}^*_{f1}...\mathbf{y}^*_{fd}, \mathbf{u}_1...\mathbf{u}_d) = \prod_{i=1}^{d} p(u^*_i \geq l_{u_i}|\mathbf{y}^*_{f_i}, \mathbf{u}_i). \tag{42}$$

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 (39). 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 (39).
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),

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[\prod_{i=1}^{d} p(u_i^* \geq l_{u_i} | \mathbf{y}^*(count), \mathbf{u}_i) \right] = E_{\mathbf{y}^*} \left[\prod_{i=1}^{d} p(u_i^* \geq l_{u_i} | \mathbf{y}^*, \mathbf{u}_i) \right]. \tag{43}$$

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.

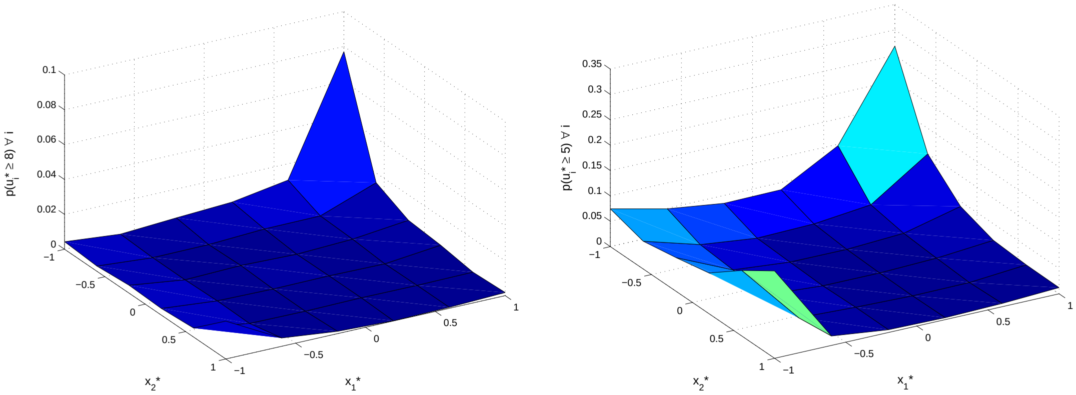

Figure 5: Surface plot of $p(u^* \geq 8)$ for different values of coded Pressure and H_2/WF_6

4.2 Example: Multiple DM's

To illustrate the method for multiple DM's, consider the data from the example in section 3.3. In that section, the MCDM problem was solved for each DM individually. Therefore, for each survey shown in table 2, the optimization problem gives a different solution as shown in table 3. Here, we consider the same data as shown in tables 1 and 2. However, we consider all the 4 surveys simultaneously. The models for the responses and the utility are the same as in section 3.3 as given by equations (31-32) and (33-36).

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

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".

Acknowledgements - We thank R. Jean Ruth and Sharon Zielinksi from General Motors R&D Vehicle Development Research Lab for helpful comments on the manuscript.

References

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

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

[3] 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).

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

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

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

[7] 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).

[8] Peterson, J.J., "A Posterior Predictive Approach to Multiple Response Surface Optimization ", *Journal of Quality Technology*, 36(2), 139-153, (2004).

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

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

[11] 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 and M.J. Todd, eds., Elsevier Science Publishers B.V., Amserdam, The Netherlands, (1989).