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

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

John J. Peterson
GlaxoSmithKline Pharmaceuticals,
709, Swedeland Road, King of Prussia, PA 19406

May 2004

Abstract

The uncertainty of the model form is typically neglected in process optimization studies. In addition, not taking into account the existence of noise factors and non-normal errors may invalidate the conclusions of such studies. In this paper, a Bayesian approach to model-robust process optimization in the presence of noise factors and non-normal error terms is presented. 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. This paper extends the idea of model-robustness of [9] using the Bayesian posterior predictive density of the process response to cases where there is uncertainty due to noise factors and due to non-normal error terms. Two examples taken from the literature, one based on a factorial experiment and another based on a mixture experiment are used to illustrate the proposed approach.

1 Introduction

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. This can be achieved using a Bayesian predictive approach (see [8]). 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 [8] 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 [6] extended this approach to include the presence of noise factors. Rajagopal and Del Castillo [9] took this idea one step further 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 [9] that there are no noise factors present in the system, and that the error terms in all the competing models are normally distributed.

The presence of noise factors or non-normal errors may invalidate the conclusions obtained with model-robust optimization methods. It is the purpose of the present paper to extend the MAP approach of [9] to cases where there are noise factors and when errors are t-distributed. The approach presented here thus provides robustness or resistance to uncertainty in the model form, in the parameters of the competing models, in the noise factors and in the distribution of the error terms.

The remainder of the paper is organized as follows. Section 2 reviews model-robust optimization approaches based on the Bayesian model-averaging technique. In section 3 we extend this technique to noise factors, and in section 4 to non-normal errors. Each of sections 3 and 4 contains an example that illustrates the extensions. The paper concludes with a summary of the findings in section 5.

2 A Review of the Model-robust Approach for Process Optimization

In this section, we briefly review the MAP approach of [9], which will be extended to include noise factors in section 3 and non-normal errors in section 4. The true process model in [9] is assumed to be of the form

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \tag{1}$$

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 is formulated as

$$\max_{x_1,...,x_k} \quad P(L \leq Y^* \leq U) = \sum_i \left[\int_L^U P(y^*|M_i, \mathbf{x}^*, \mathbf{y}) dy^* \right] P(M_i|\mathbf{y}), \tag{2}$$

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 is carried out with respect to the k control factors $(x_1, x_2, ..., x_k)$ that $\mathbf{x}^*$ depends on, over all competing models under consideration, each model denoted by M_i, where $\{i = 1, 2, 3, ...\}$. $P(M_i|\mathbf{y})$ 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, $\mathbf{x}^*$. The prior on the models assumed in [9] were

$$P(M_i) = \pi^{f_i}(1-\pi)^{k-f_i}, \tag{3}$$

where f_i out of the total k factors are present in model M_i, and the probability of a factor to be active in any of the models is π. In addition, the priors for the parameters for each model M_i were

$$\boldsymbol{\beta}_i \sim N(0, \Sigma_i\sigma^2), \tag{4}$$

$$P(\sigma^2) = \frac{1}{\sigma^2}, \tag{5}$$

and

$$P(\boldsymbol{\beta}_i, \sigma^2) = P(\sigma^2)P(\boldsymbol{\beta}_i). \tag{6}$$

Here $\boldsymbol{\Sigma}^{-1} = (\mathbf{X}_i'\mathbf{X}_i)\mathbf{V_i}$, where $\mathbf{V_i} = \frac{1}{g}\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{t_i} \end{pmatrix}$ if the model includes a constant term, or $\boldsymbol{\Sigma}_i^{-1} = (1/g)(\mathbf{X}_i'\mathbf{X}_i)$ if the model does not have any constant term. In other words, the parameter for the constant term is assumed to have a non-informative prior, and the remaining parameters are assumed to have a Zellner's g-prior [14]. The hyper-parameter g needs to be chosen by the user. Denote by r_i the number of terms in model M_i and by t_i 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. Using the above priors for the case where there are no noise factors and when the errors are normally distributed, it is shown in [9] that the model posteriors are given by

$$P(M_i|\mathbf{y}) \propto \pi^{f_i}(1-\pi)^{k-f_i}\gamma^{-t_i}|\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i|^{-\frac{1}{2}} S_i^{-\frac{(n-1)}{2}}, \tag{7}$$

where

$$\frac{g}{\gamma^2}\mathbf{V}_i = \boldsymbol{\Sigma}^{-1}, \tag{8}$$

$$\begin{aligned} S_i &= (\mathbf{y} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i)'(\mathbf{y} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i{}'\boldsymbol{\Sigma}_i^{-1}\hat{\boldsymbol{\beta}}_i & (9) \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}_i(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}, & (10) \end{aligned}$$

and

$$\hat{\boldsymbol{\beta}}_i = (\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}. \tag{11}$$

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

$$y^*|M_i, \mathbf{x}^*, \mathbf{y} \;\propto\; t_\nu\left(\mathbf{x}^*\hat{\boldsymbol{\beta}}_i, \hat{\sigma}_i^2\left[1 + \mathbf{x}^{*\prime}(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^*\right]\right). \tag{12}$$

Therefore, the cumulative posterior predictive density for model M_i is computed by

$$P\left(\frac{y^* - \mathbf{x}^{*\prime}\hat{\boldsymbol{\beta}}_i}{\hat{\sigma}_i\sqrt{1 + \mathbf{x}^{*\prime}(\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{x}^*}} < t|M_i, \mathbf{x}^*, \mathbf{y}\right) = \frac{1}{2}\left[1 + I_{\frac{t^2}{\nu+t^2}}\left(\frac{1}{2}, \frac{\nu}{2}\right)\right], \tag{13}$$

where $I_z(a,b)$ is the incomplete beta function, $\nu = n-1$ and $\hat{\sigma}_i^2 = S_i/(n-1)$ for models that include a constant term, and $\nu = n$ and $\hat{\sigma}_i^2 = S_i/n$ otherwise. The above results were applied to two examples in [9] that demonstrated the advantages of model-averaging using the Bayesian predictive approach.

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 [11], [12], [13]). 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 [1],[7]. 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

$$\hat{y}(\mathbf{x_c}, \mathbf{x_n}) = b_o + \mathbf{x_c'b} + \mathbf{x_c'Bx_c} + \mathbf{x_n'c} + \mathbf{x_c'\Delta x_c}, \tag{14}$$

to the response as a function of both the controllable and noise factors. In equation (14), $\mathbf{x_c}$ 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 [2]. 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 [6] 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, we present an extension to the Bayesian predictive approach that also accounts for uncertainty in the true process model using the MAP approach reviewed in section 2.

We 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. We assume 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

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \tag{15}$$

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 (7).

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 (13). 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,

$$P(L \leq Y^* \leq U|M_i, \mathbf{x}^*, \mathbf{y}) = \int_{\mathbf{x}^*_\mathbf{n}} P(L \leq Y^* \leq U|M_i, \mathbf{x}^*_\mathbf{c}, \mathbf{x}^*_\mathbf{n}, \mathbf{y}) f_{\mathbf{x_n}}(\mathbf{x}^*_\mathbf{n})\, d\mathbf{x}^*_\mathbf{n} \quad (16)$$

$$= E_{\mathbf{x}^*_\mathbf{n}} \left[P(L \leq Y^* \leq U|M_i, \mathbf{x}^*_\mathbf{c}, \mathbf{x}^*_\mathbf{n}, \mathbf{y}) \right]. \quad (17)$$

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

$$\max_{\mathbf{x}^*_\mathbf{c}} \; E_{\mathbf{x}^*_\mathbf{n}} \left[P(L \leq Y^* \leq U) \right] = E_{\mathbf{x}^*_\mathbf{n}} \left[\sum_i \left(\int_L^U P(y^*|M_i, \mathbf{x}^*, \mathbf{y}) dy^* \right) P(M_i|\mathbf{y}) \right] \quad (18)$$

$$= \sum_i E_{\mathbf{x}^*_\mathbf{n}} \left[\left(\int_L^U P(y^*|M_i, \mathbf{x}^*, \mathbf{y}) dy^* \right) \right] P(M_i|\mathbf{y}). \quad (19)$$

The objective function above can be computed using equations (7) and (13). 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 $P(L \leq Y^* \leq U|M_i, \mathbf{x}_\mathbf{c}^*, \mathbf{x}_\mathbf{n}^*(count), \mathbf{y})$ for the sample using equation (13).

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),

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[P(L \leq Y^* \leq U|M_i, \mathbf{x}_\mathbf{c}^*, \mathbf{x}_\mathbf{n}^*(i)\mathbf{y})\right] = E_{\mathbf{x}_\mathbf{n}^*}\left[P(L \leq Y^* \leq U|M_i, \mathbf{x}_\mathbf{c}^*, \mathbf{x}_\mathbf{n}^*\mathbf{y})\right]. \tag{20}$$

The example below illustrates the proposed method.

3.1 Example

The data for this example is taken from Derringer and Suich [3], and is given in table 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, we only consider response y_3, and assume that it is desired to obtain y_3 within the specification limits $[400, 600]$. It is also assumed that factor x_1 is a noise factor. Using the data in table 1, and the priors mentioned earlier, we compute the model posteriors for all subsets belonging to the model class shown in equation (14). Here, we use hyper-parameters $\pi = 0.5$, and $\gamma = 0.6$. The choice of these hyper-parameters is discussed in Rajagopal and Del Castillo [9]. 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 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, we consider model numbers 1-7 from table 2 that account for over 95% of the total probability (see [9]).

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 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 (7) and (13) 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 we instead consider x_1 to be a noise factor and perform the optimization only with respect to x_2 and x_3 using equation (19), 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 we consider x_1 as a noise factor in the optimization, we notice 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 (17) 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 1: Data for example in section 3.1 from Derringer and Suich [3]

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

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 (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 [10],[4], 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, we present a methodology to extend the idea of Bayesian model-averaged process optimization to obtain a solution that is also robust to t_ν-distributed error terms. Here also, the objective we consider is to maximize the posterior predictive probability that the response lies within the specification limits.

We 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. We assume 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

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \tag{21}$$

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, we can

consider equation (21) as a weighted regression model,

$$y = \mathbf{x}'\boldsymbol{\beta} + e/\sqrt{w} \tag{22}$$

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

$$\tilde{y} = \tilde{\mathbf{x}}'\boldsymbol{\beta} + e. \tag{23}$$

Since the error term in equation (23) has a standard normal distribution, the model posteriors in this case can be obtained from equation (7). If $\mathbf{X}_i$ is the design matrix corresponding to model M_i, then using the notations and the priors in equations (3), (4), (5), and (6), the model posteriors conditional on $\mathbf{W}_i$ are given by,

$$\begin{aligned} P(M_i|\mathbf{y}, w_1, ...w_n) &\propto \pi^{f_i}(1-\pi)^{k-f_i}\gamma^{-t_i}|\boldsymbol{\Sigma}_i^{-1} + \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i|^{-\frac{1}{2}}\tilde{S}_i^{-\frac{(n-1)}{2}} &(24)\\ &= \pi^{f_i}(1-\pi)^{k-f_i}\gamma^{-t_i}|\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i|^{-\frac{1}{2}}\tilde{S}_i^{-\frac{(n-1)}{2}}, &(25)\end{aligned}$$

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 $\mathbf{W}_i = \begin{bmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{bmatrix}$,

$$\begin{aligned} \tilde{S}_i &= (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}}_i)'(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i{}'\boldsymbol{\Sigma}_i^{-1}\hat{\boldsymbol{\beta}}_i &(26)\\ &= (\mathbf{y} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i)'\mathbf{W}(\mathbf{y} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i{}'\boldsymbol{\Sigma}_i^{-1}\hat{\boldsymbol{\beta}}_i, &(27)\end{aligned}$$

and,

$$\begin{aligned} \hat{\boldsymbol{\beta}}_i &= (\boldsymbol{\Sigma}_i^{-1} + \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i)^{-1}\tilde{\mathbf{X}}_i'\tilde{\mathbf{y}} &(28)\\ &= (\boldsymbol{\Sigma}_i^{-1} + \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{W}_i\mathbf{y}. &(29)\end{aligned}$$

The model posteriors, $P(M_i|\mathbf{y})$ are then obtained by taking the expected value in equation (25) with respect to $(w_1, ..., w_n)$, that is

$$P(M_i|\mathbf{y}) = E_{w_1,...,w_n}\left[P(M_i|\mathbf{y}, w_1, ..., w_n)\right]. \tag{30}$$

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 (25).

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),

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[P(M_i|\mathbf{y}, w_1, ...w_n)\right] = E_{w_1,...w_n}\left[P(M_i|\mathbf{y}, w_1, ...w_n)\right]. \tag{31}$$

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 (12) as,

$$\tilde{y^*}|M_i, \mathbf{x}^*, w^*, w_1, ..., w_n, \mathbf{y} \;\propto\; t_\nu\left(\tilde{\mathbf{x}^*}\hat{\boldsymbol{\beta}}_i, \hat{\sigma}_i^2\left[1 + \tilde{\mathbf{x}^*}'(\boldsymbol{\Sigma}_i^{-1} + \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i)^{-1}\tilde{\mathbf{x}^*}\right]\right), \tag{32}$$

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,

$$y^*|M_i, \mathbf{x}^*, w^*, w_1, ..., w_n, \mathbf{y} \;\propto\; t_\nu\left(\sqrt{w^*}\tilde{\mathbf{x}^*}\hat{\boldsymbol{\beta}}_i, w^*\hat{\sigma}_i^2\left[1 + \tilde{\mathbf{x}^*}'(\boldsymbol{\Sigma}_i^{-1} + \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i)^{-1}\tilde{\mathbf{x}^*}\right]\right). \tag{33}$$

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 $(w_1, ..., w_n)$:

$$P(L < y^* < U|M_i, \mathbf{x}^*, \mathbf{Y}) = E_{w^*}\left[E_{w_1,...,w_n}\left\{P(L < Y^* < U|M_i, \mathbf{x}^*, w^*, w_1, ..., w_n, \mathbf{Y})\right\}\right]. \tag{34}$$

The expected values in equation (34) can also be computed using simulation. The objective function in equation (2) can be computed using equations (30) and (34). 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.

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

Table 3: Sample data for example in section 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: Model posteriors for example in section 4.1

4.1 Example 1

To illustrate the methodology, we consider the simulated data given in table 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 (11). The fitted models are plotted in figure 1, using a value of $\pi = 0.5$ and $\gamma = 2.4$. The choice of these hyper-parameters is discussed in [9]. The posterior probabilities for the models under consideration are computed by using equation (12) for the case of normal errors and by using equation (30) for the case of t-distributed errors with 5 degrees of freedom. These are shown in table 4. From table 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 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, the probability $P(L < y^* < U|x^*)$ can be much different depending

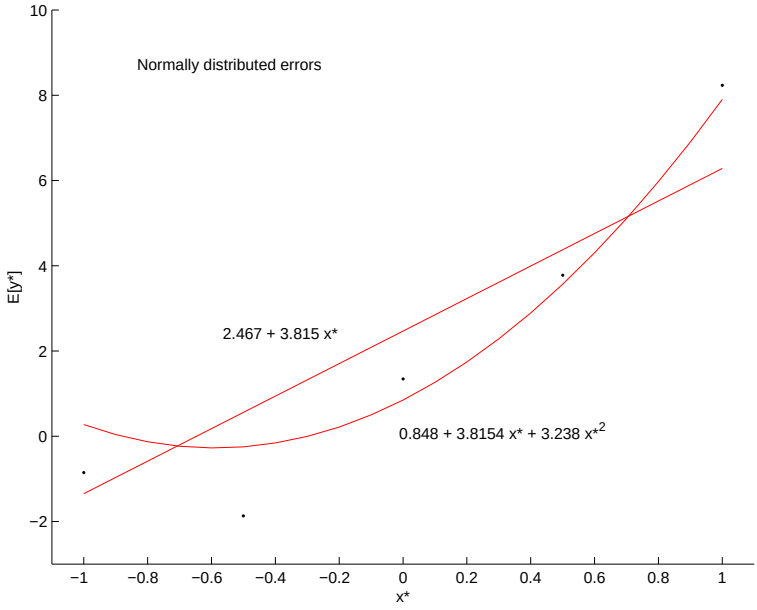

Figure 1: Parameter estimates for linear and quadratic model under normal errors assumption

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 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 $L = -\infty, U = 2$.

4.2 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. [5], and is given in table 5. 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

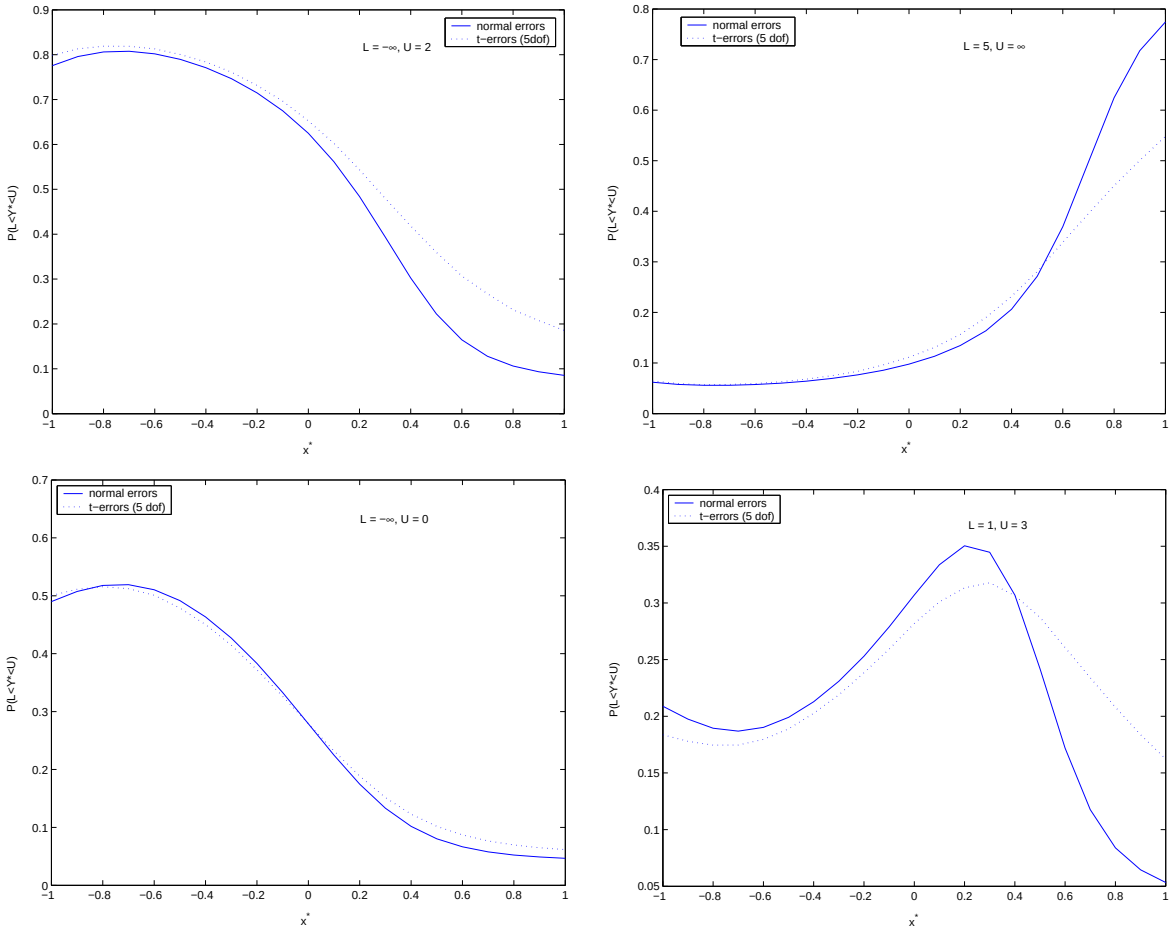

Figure 2: Cumulative model-averaged posterior probabilities for different specifications L, U

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 5: Mixture data from [5] where the response is Glass Transition temperature

factors that minimizes the glass transition temperature. We consider models belonging to the following two classes of models:

$$y = b_1x_1 + b_2x_2 + b_3x_3 + b_4(x_1x_2) + b_5(x_1x_3) + b_6(x_2x_3), \tag{35}$$

$$y = b_1x_1 + b_2x_2 + b_3x_3 + b_4\, min(x_1, x_2) + b_5\, min(x_1, x_3) + b_6\, min(x_2, x_3). \tag{36}$$

The results for model-robust process optimization for these classes of models under normal errors are given in Rajagopal and Del Castillo [9]. However, as there are only 11 runs, it is difficult to verify the distribution of the error terms. Tables 6, 7, and 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 [9]. 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 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 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 differ 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 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 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 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 9: Optimization results for example 3.2

5 Conclusion

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, especially 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.

Acknowledgements. This work was partially funded by the National Science Foundation (grant DMII 9988563) and by a grant from General Motors R&D center.

References

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

[2] 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).

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

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

[5] 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).

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

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

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

[9] Rajagopal, R., and Del Castillo, E., "Model-Robust Process Optimization using Bayesian Model Averaging", working paper, Engineering Statistics Laboratory, IME

department, Penn state University, (submitted for publication), (2003). Available at http://www.ie.psu.edu/researchlabs/Engineeringstatistics/publications.htm.

[10] Rousseeuw, P.J., and Leroy, A.M., *Robust regression and outlier detection*, John Wiley & Sons, New York, NY, (1987).

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

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

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

[14] 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).