A Bayesian Approach to the Analysis of Split-Plot Product Arrays and Optimization in Robust Parameter Design

Abstract

Many robust parameter design (RPD) studies involve a split-plot randomization structure and to obtain valid inferences in the analysis, it is essential to account for the design induced correlation structure. Bayesian methods are appealing for these studies since they naturally accommodate a general class of models, can account for parameter uncertainty in process optimization, and offer the necessary flexibility when one is interested in non-standard performance criteria. In this article, we present a Bayesian approach to process optimization for a general class of RPD models in the split-plot context using an empirical approximation to the posterior distribution of an objective function of interest. Two examples from the literature are used for illustration.

KEY WORDS: Bayesian Predictive Density; Generalized Linear Mixed Models; Hard-to-Change Factor; Markov Chain Monte Carlo; Process Optimization; Response Surface; Restricted Randomization.

Introduction

In many robust parameter design (RPD) applications, the response of interest follows a non-normal distribution. For instance, Lee and Nelder (2003) utilize Poisson regression to estimate the relationship between the number of solder defects and eight design variables (5 control factors and 3 noise factors). Myers, Brenneman and Myers (2005) demonstrate the use of gamma regression in modeling the relationship between resistivity and four design factors (3 control and 1 noise factor) using data from a wafer etching process in semiconductor manufacturing.

Quite often in RPD, the designed experiment is a crossed or product array (i.e., a design for the noise factors crossed with a design in the control factors). Although there are times when a completely randomized design (CRD) is appropriate for these experiments, a split-plot design (SPD) can be easier and more cost efficient to implement when hard/difficult-to-change factors exist. Even when hard-to-change factors do not exist, Box and Jones (1992) point out that SPD's are often more efficient than CRD's especially when the noise factors are in the outer array. Goos et al. (2006), Wolfinger and Tobias (1998), and Ganju and Lucas (1997) note that the induced correlation structure from a SPD must be accounted for in the analysis to obtain valid statistical inferences. Wolfinger and Tobias (1998) suggest the use of linear mixed models for RPD SPD's with normal responses but do not discuss prediction. Robinson et al. (2004) advocate the use of generalized linear mixed models (GLMMs) for industrial SPD's with non-normal responses. Robinson et al. (2009) propose Bayesian methods for the analysis of SPD's with non-normal responses.

Several uses of Bayesian methods for RPD have appeared in the literature but all assume complete randomization of the experimental run order. Chipman (1998) propose a Bayesian approach for product arrays. Miro-Quesada, Del Castillo, and Peterson (2004) present a Bayesian approach using the posterior predictive distribution for multiple response optimization. Rajagopal, Del Castillo and Peterson (2005) extend the work of Miro-Quesada et al. (2004) to cases in which the practitioner seeks an RPD solution robust to uncertainty in the process model by using Bayesian model averaging. Bayesian approaches are appealing in RPD since (a) they allow for uncertainty in parameter estimation to be accounted for when obtaining an optimal solution in the control factor space; (b) process quality can be optimized in terms of a variety of performance criteria such as conformance probabilities and other more useful characteristics of the response distribution than being restricted to the process mean and variance; and (c) the approach can handle a wide class of error distributions in a straightforward manner.

In this article, we define a general class of models for product array robust design experiments with a split-plot randomization structure. We then develop a Bayesian approach for modeling the data along with an optimization approach using an empirical approximation to the posterior distribution of a user-defined objective function. The methodology is illustrated with two examples from the literature and we consider some non-standard performance criteria in characterizing the underlying process.

Model Formulation and Examples

For an exponential family member response observed from a split-plot experiment, we consider an extended class of generalized linear models which include random effects. Let $\mathbf{y}$ be an $N{\times}1$ vector of responses (N denoting the total number of subplot runs), $\mathbf{X}$ an $N{\times}p$ matrix of whole plot and subplot model terms, $\boldsymbol{\beta}$ denote the associated $p{\times}1$ vector of fixed-effects regression parameters, $\mathbf{Z}$ an $N{\times}w$ model matrix for the random effects and $\boldsymbol{\delta}$ the corresponding w-vector of random effects. Models under consideration contain two parts:

1. Conditional on the random effect $\boldsymbol{\delta}$, the response distribution is a member of the exponential family of distributions with

$$E(\mathbf{y}|\boldsymbol{\delta}) = \boldsymbol{\mu},\ Var(\mathbf{y}|\boldsymbol{\delta}) = \phi V(\boldsymbol{\mu}),$$

where ϕ is the dispersion parameter and $V(\cdot)$ is the variance function. The linear predictor is given by

$$\boldsymbol{\eta} = \boldsymbol{g}(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\delta}.$$

Here, the link function, $\boldsymbol{g}(\cdot)$, is chosen to be monotonic and differentiable.

2. The random effects in $\boldsymbol{\delta}$ are assumed i.i.d. according to some specified probability distribution.

Normal Response Example

Box and Jones (1992) consider an example where the manufacturer seeks the best recipe for a box cake mix. In the example there are three control factors: X_1 = flour, X_2 = shortening and X_3 = egg powder, coded in the ranges -1 to +1. The manufacturer is concerned that random fluctuations in cooking time, Z_1, and cooking temperature, Z_2, in home ovens (both treated as noise factors with 2 levels codes as -1 and +1) may result in cake flavor that is too variable. In order to develop a robust cake recipe, a 2^2 factorial design in the noise factors is crossed with a 2^3 factorial design in the control factors. For a fixed level combination of Z_1 and Z_2, eight cakes are baked, each with a different combination of X_1, X_2 and X_3 levels and the run order of the control factor combinations are randomized. Note that in this experiment, the noise factors are in the whole plots, while the control factors are changed within the whole plots.

Consistent with Box and Jones (2002), we assume $y_{ij} \mid \mu_{ij}, \sigma_\varepsilon^2 \sim N\left(\mu_{ij}, \sigma_\varepsilon^2\right)$, where y_{ij} denotes the average cake quality (i.e., averaged across a panel of judges who each ranked taste from 1-7) for the j^{th} cake baked at the i^{th} level combination of Z_1 and Z_2 $\left(i=1,\ldots,4; j=1,\ldots,8\right)$. Also consistent with Box and Jones, we treat the response as continuous since ordinal scores are averaged over a panel of judges. While Box and Jones focused upon treatment comparisons we consider response optimization and fit a model with an identity link $\left[\text{i.e., } g\left(\mu_{ij} \mid \delta_i\right) = \mu_{ij} \mid \delta\right]$. Thus, the linear predictor is given by

$$\eta_{ij} \mid \delta_i = \mu_{ij} \mid \delta = \mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \delta_i, \qquad (1)$$

where each $\mathbf{x}_{ij}^{'}$ denotes a $1{\times}15$ vector of model terms consisting of the intercept, the two noise main effects, six control factor only effects (i.e., 3 main effects and 3 two-factor interactions), and the six control-by-noise two-factor interactions. The β's are the associated fixed-effects model parameters while the δ_i represent the whole-plot error terms and are assumed to be i.i.d. Normal$\left(0, \sigma_\delta^2\right)$. Relating the model in (0) to the general formulation, we have $g\left(\cdot\right) = \left(\cdot\right)$, $\mathbf{y}$ denotes the $32{\times}1$ vector of responses starting with the eight observations in the first whole plot and so on, $\mathbf{Z}$ is a $32{\times}4$ classification matrix of ones and zeros where the kl^{th} entry is a one if the k^{th} observation (*k=1,…,32*) belongs to the l^{th} whole plot ($l=1,\ldots,4$). The δ_i comprise the 4×1 vector of random effects, $\boldsymbol{\delta}$, and the rows of the 32×15 model matrix $\mathbf{X}$ are formed by $\mathbf{x}'_{ij}$.

Gamma Response Example

Robinson et al. (2004) consider data from a film manufacturer in which the investigator wishes to study the effects of six factors (three mixture components [X_1, X_2, X_3] and three process variables [p_1, p_2, and p_3]) on film quality. We assume for this application that the process variables represent environmental variables whose fluctuations in the process cause unwanted variation in the response. A product array with the mixture design in the control factors serving as the outer array and a 2^{3-1} design in the process variables as the inner array was conducted. Note that there is potential for confusion with the terminology of the experiment, since we need to clarify both the RPD and the split-plot pieces of the design. In RPD, the control factors are traditionally labeled as being in the inner array and noise factors in the outer array, and yet for this experiment, control factor levels were randomized at the whole plot level and the noise factor levels were randomized at the subplot level. Note that this experimental set-up differs from the first example where the noise factor levels were varied within the whole plots. The goals of this experiment were to: (a) examine the impact of the mixture variables on response quality; (b) study the contribution of the noise factors to process variance; and (c) determine optimal settings for the mixture variables. In the experiment, five distinct formulations of the mixture variables were used to produce a total of 13 batches (one batch=one roll). Upon production of a roll of film with a given setting of the mixture variables, four pieces of the roll were randomly assigned according to a 2^{3-1} design in the process variables.

Consistent with Robinson et al. (2004, 2009) and Lee, Nelder and Park (2010), we assume $y_{ij} \mid \alpha, \theta_{ij} \sim Gamma\left(\alpha, \theta_{ij}\right)$, where y_{ij} denotes the film quality for the j^{th} piece of film from the i^{th} roll. The following parameterization of the gamma density is assumed,

$$f_y\left(y\right) = \frac{\theta^{\alpha} y^{\alpha-1} e^{-\theta y}}{\Gamma\left(\alpha\right)}. \tag{2}$$

Using the model of Robinson et al. (2004, 2009), a log link $\left[\text{i.e., } g\left(\mu_{ij} \mid \delta_i\right) = ln\left(\mu_{ij} \mid \delta_i\right)\right]$, with linear predictor

$$\eta_{ij} \mid \delta_i = ln\left(\mu_{ij} \mid \delta_i\right) = \mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \delta_i$$
$$= \sum_{b=1}^{3}\beta_b x_{b,ij} + \beta_{12}x_{1,ij}x_{2,ij} + \sum_{c=1}^{3}\gamma_{c,ij}p_{c,ij} + \sum_{b=1}^{2}\sum_{c=1}^{3}\psi_{bc}x_{b,ij}p_{c,ij} + \delta_i, \quad (3)$$

is fit. Note that the linear predictor in (0) does not include an intercept since we are fitting a Scheffé model for mixture experiments and only a subset of the total number of possible two factor interactions were of interest. In (3), the β's are for the first and second order mixture/control factor terms, the γ's are for the linear process/noise terms, and the ψ's represent the mixture-by-process/control-by-noise interactions. Relating the model in (0) to the general formulation, we have $g(\cdot) = \ln(\cdot)$, $\mathbf{y}$ denotes the 52×1 vector of observations starting with the four observations from the first roll and so on, $\mathbf{Z}$ is a 52×13 classification matrix of ones and zeros, where the kl^{th} entry is a one if the k^{th} observation ($k=1,\ldots,52$) belongs to the l^{th} roll ($l=1,\ldots,13$). The rows of the 52×13 model matrix $\mathbf{X}$ are formed by $\mathbf{x}_{ij}^{'}$ for each observation.

Bayesian Inference

Let $\boldsymbol{\Theta}$ denote the vector of model parameters. The Bayesian inferential approach combines prior information about $\boldsymbol{\Theta}$ with the information contained in the data. The prior information is described by a prior density, $\pi(\boldsymbol{\Theta})$, and summarizes what is known about the model parameters before data are observed. The information provided by data is captured by the data sampling model, $f_{\mathrm{y}}(\mathbf{y} \mid \boldsymbol{\Theta})$, known as the likelihood. The combined information is described by the posterior density, $\pi(\boldsymbol{\Theta} \mid \mathbf{y})$. We evaluate the posterior density using Bayes' Theorem [Degroot (1970) p. 28] as

$$\pi(\boldsymbol{\Theta} \mid \mathbf{y}) \propto f(\mathbf{y} \mid \boldsymbol{\Theta})\pi(\boldsymbol{\Theta}).$$

For the cake mix example, $\boldsymbol{\Theta}^{'} = \left(\boldsymbol{\beta}^{'}, \boldsymbol{\delta}^{'}, \sigma_\delta, \sigma_\varepsilon\right)$, where $\boldsymbol{\beta}^{'}$ is the 1×15 vector of regression coefficients, $\boldsymbol{\delta}^{'}$ is the 1×4 vector of random effects used to model the outer array error term, σ_δ is the standard deviation of the random effects distribution, and σ_ε is the normal error shape parameter. The likelihood $f_{\mathrm{y}}(\mathbf{y} \mid \boldsymbol{\Theta})$ has the form

$$f_{\mathbf{y}}(\mathbf{y}\,|\,\mathbf{\Theta}) = \prod_{i=1}^{4}\prod_{j=1}^{8} f_y(y_{ij}\,|\,\mu_{ij},\sigma_\varepsilon)\,,$$

with μ_{ij} given in (0), where $f_y(y_{ij}\,|\,\mu_{ij},\sigma_\varepsilon)$ is a normal density. We see from (0) that μ_{ij} also depends on $\mathbf{x}'_{ij}$, $\boldsymbol{\beta}'$, and δ_i. The prior density, $\pi(\mathbf{\Theta})$, has the form

$$\pi(\mathbf{\Theta}) \propto \left(\prod_{i=1}^{4} f_{\delta_i}(\delta_i\,|\,\sigma_\delta^2)\right)\left(\prod_{j=1}^{15} f_{\beta_j}(\beta_j)\right) f_{\sigma_\varepsilon}(\sigma_\varepsilon) f_{\sigma_\delta}(\sigma_\delta)\,,$$

where $f_\delta(\cdot)$ is the Normal$(0,\sigma_\delta^2)$ density assumed for the random effects, δ_i. For the regression coefficients (the β_j's), we use the following diffuse but proper prior distributions: Normal$(0,1000^2)$. For σ_ε, we use the diffuse prior distribution Uniform(0, 10). For σ_δ, we use the informative prior distribution Gamma(625,10000). An informative prior is needed here since there is no replication of whole plots, and hence the observed data do not contain information about both the interaction between time and temperature and the whole plot variability.

For the film manufacturing example, $\mathbf{\Theta}' = (\alpha, \boldsymbol{\beta}', \boldsymbol{\delta}', \sigma_\delta)$, where α is the gamma shape parameter, $\boldsymbol{\beta}'$ is the *1×13* vector of regression coefficients given in (3), $\boldsymbol{\delta}'$ is the *1×13* vector of random roll effects, and σ_δ is the standard deviation of the random effects distribution. The likelihood $f_{\mathbf{y}}(\mathbf{y}\,|\,\mathbf{\Theta})$ has the form

$$f_{\mathbf{y}}(\mathbf{y}\,|\,\mathbf{\Theta}) = \prod_{i=1}^{13}\prod_{j=1}^{4} f_y(y_{ij}\,|\,\alpha, \theta_{ij} = \alpha/\mu_{ij}),$$

with $f_y(y_{ij}\,|\,\alpha,\theta_{ij})$ from (0) and μ_{ij} given in (0). From (0), it is evident that μ_{ij} depends on $\mathbf{x}'_{ij}$, $\boldsymbol{\beta}$, and δ_i. The prior density, $\pi(\mathbf{\Theta})$, has the form

$$\pi(\mathbf{\Theta}) \propto \left(\prod_{i=1}^{13} f_{\delta_i}(\delta_i\,|\,\sigma_\delta^2)\right)\left(\prod_{j=1}^{13} f_{\beta_j}(\beta_j)\right) f_\alpha(\alpha) f_{\sigma_\delta}(\sigma_\delta)\,,$$

where $f_\delta(\cdot)$ is the Normal$(0,\sigma_\delta^2)$ density assumed for the random roll effects δ_i. For the regression coefficients (the β_j's) and α, we use the diffuse but proper prior distributions: Normal(0,1000^2) and Uniform(0,100), respectively. For σ_δ, we use the diffuse prior Uniform(0,100) distribution as suggested by Gelman (2006).

When the form of the posterior density is well known, the posterior distribution can be obtained in closed form. For more general forms of the posterior density, we can approximate the posterior distribution via Markov chain Monte Carlo (MCMC) [Gelfand and Smith (1990), Casella and George (1992), Chib and Greenberg (1995)]. The MCMC algorithm produces samples from the joint posterior distribution of $\mathbf{\Theta}$ by sequentially updating each model parameter conditional on the current values of the other parameters.

To analyze both examples, we used WinBUGS [Spiegelhalter, Thomas, Best, and Lunn (2004)]. For the cake mix data, summaries of the joint posterior draws of $\mathbf{\Theta}$ are given in Table 1, and are based on two chains of length 100,000 thinned to every 20th draw, i.e., totaling 10,000 draws. A similar summary for the film example is given in Table 2. Note that in the cake mix example there are no whole plot replicates, so the marginal posterior distribution of the whole plot error variance is essentially the chosen informative prior. For the film example, there are several whole plot replicates, so that the whole plot error variance can be estimated; thus, the data can update the prior.

To assess convergence of the MCMC algorithms and goodness-of-fit of the models to the data, several diagnostics were considered. We begin with diagnostics for the cake mix example. To assess the convergence of the MCMC algorithm, trace plots of the MCMC chains, such as those in Figure 1, work well. Figure 1 contains the trace plots of the MCMC chains for the first four regression coefficients in the cake mix example. Here, only one of the two chains is shown, and the pictured chain is thinned to every 100th iteration instead of every 20th. Figure 1 shows that convergence for the four parameters has occurred by iteration 10,000, and that they are mixing well. Trace plots for the other model parameters (not shown) all lead to similar conclusions. See Gelman et al. (2004, Section 11.6 pp. 294-299) for more discussion of MCMC convergence diagnostics.

The overall fit of the model to the data can be examined by the Bayesian χ^2 goodness-of-fit test (Johnson, 2004). The test proceeds as follows:

1. Divide the interval [0, 1] into $N^{0.4}$=32$^{0.4}$=4 bins (as recommended by Johnson (2004)), and set each bin count to zero.

2. For a given MCMC draw of $\mathbf{\Theta}$, calculate the cumulative probability of y_{ij}, $p_{ij} = F(y_{ij}|\mathbf{\Theta})$, $\forall\, i = 1, 2, \cdots, 4$ and $j = 1, 2, \cdots, 8$. Here, F is a normal cumulative distribution function.
3. If $0 \leq p_{ij} < 0.25$, the count of bin 1 is increased by 1. If $0.25 \leq p_{ij} < 0.5$, the count of bin 2 is increased by 1, etc.
4. Compute the standard χ^2 statistic for comparing the observed bin counts to the expected bin counts, $N/4=32/4=8$.
5. Compare the χ^2 statistic to the 95th percentile of the χ_3^2 distribution where the degrees of freedom are computed as 3=4-1.
6. Repeat 1-5 for all 10,000 (2 chains × 5,000 draws per chain) draws of $\mathbf{\Theta}$.

If approximately 95% of the tests fail to reject that the observed bin counts are consistent with the expected bin counts, there is no evidence of lack-of-fit. That proportion for the cake mix example is 94.14%.

To check the assumption of normal random effects, for each posterior draw of δ_i, $i = 1, 2, \cdots, 4$, an envelope based the corresponding draw of σ_δ is created. The number of δ_i draws, $i = 1, 2, \cdots, 4$, that falls within its corresponding envelope summarizes the appropriateness of the normal assumption for the random effects. An envelope is created by the following procedure:

1. Generate 1,000 sets of 4 $N(0, \sigma_\delta)$ deviates.
2. Order each set.
3. Calculate the 0.013/2 and 1-0.013/2 percentiles of each order statistic since $0.987^4 \approx 0.949$.

A draw of δ_i, $i = 1, 2, \cdots, 4$, is said to fall within its envelope if its order statistics fall within their corresponding intervals. If approximately 95% of the δ_i, $i = 1, 2, \cdots, 4$, draws fall within their corresponding envelope, the normal assumption for the random effects is tenable. The percentage for the cake mix example is 95.12%.

Lastly, we consider the conditional normality of the response, y_{ij}, by examining the residuals, $y_{ij} - \mu_{ij}$. Using the posterior distribution of $y_{ij} - \mu_{ij}$, we create a normal q-q plot from the medians of those distributions. The plot is given in Figure 2, and it suggessts the conditional normality of the response is reasonable.

Similar diagnostics were considered in the film manufacturing experiment, and they imply no problems with the convergence or mixing of the MCMC chains or the fit of the model to the data. One difference between the diagnostics for the cake mix experiment and the film manufacturing experiment is that deviance residuals, instead of raw residuals, are used for the film manufacturing experiment because the response is assumed to be conditionally gamma instead of normal. Also for the film manufacturing experiment, we may consider a different diagnostic for assessing the assumption of normal random effects. Specifically, we create a normal q-q plot from the posterior means of δ_i, $i = 1, 2, \cdots, 13$, which are given in Table 2. The plot is shown in Figure 3, and it implies the normal assumption for the random effects is reasonable. Such a plot in the cake mix experiment is not as useful since there are only four random effects.

Prediction, Performance Criteria and Optimization

Posterior Predictive Density and Performance Criterion

The primary goal of RPD is to find levels of the control factors which lead to desirable process quality results. Thus, we wish to optimize an appropriate characteristic of the response distribution under future production conditions. Inference regarding unobserved values of the response generally focuses upon characteristics of a posterior predictive density where the density incorporates all potential sources of uncertainty. However, we take a slightly different approach. Let y_{new} denote a new, observed value of the response from the process, δ_{new} an unobservable random effect associated with y_{new}, $\mathbf{x}_c$ a combination of control factors, and $\mathbf{x}_n$ a combination of noise factor levels. We consider the following density

$$\begin{aligned} f\left(y_{new}|\boldsymbol{\Theta},\mathbf{x}_c\right) &= \int_{\delta_{new}}\int_{\mathbf{x}_n} f\left(y_{new},\delta_{new},\mathbf{x}_n \mid \boldsymbol{\Theta},\mathbf{x}_c\right)\partial\mathbf{x}_n\partial\delta_{new} \\ &= \int_{\delta_{new}}\int_{\mathbf{x}_n} f\left(y_{new} \mid \boldsymbol{\Theta},\delta_{new},\mathbf{x}_c,\mathbf{x}_n\right) f\left(\delta_{new} \mid \boldsymbol{\Theta}\right) f\left(\mathbf{x}_n\right)\partial\mathbf{x}_n\partial\delta_{new}, \end{aligned}$$

where $f\left(y_{new} \mid \boldsymbol{\Theta},\delta_{new},\mathbf{x}_c,\mathbf{x}_n\right)$ is a normal density in the first example and a gamma density in the second example, $f\left(\delta_{new} \mid \boldsymbol{\Theta}\right)$ is a univariate normal density, and $f\left(\mathbf{x}_n\right)$ is a user-specified multivariate normal density representing anticipated random fluctuations of the noise factors during the process. Note that $f\left(y_{new}|\boldsymbol{\Theta},\mathbf{x}_c\right)$ incorporates the variability associated with random fluctuations of the noise factors (through $f\left(\mathbf{x}_n\right)$), the variability associated with a new whole plot effect (through $f\left(\delta_{new}\,|\,\boldsymbol{\Theta}\right)$) and the variability associated with the new observed responses (through $f\left(y_{new}\,|\,\boldsymbol{\Theta},\delta_{new},\mathbf{x}_c,\mathbf{x}_n\right)$). Also, $f\left(y_{new}|\boldsymbol{\Theta},\mathbf{x}_c\right)$ is related to the posterior predictive density, $f\left(y_{new}|\mathbf{y}_{old},\mathbf{x}_c\right)$, since

$$f\left(y_{new}|\mathbf{y}_{old},\mathbf{x}_c\right)=\int_{\boldsymbol{\Theta}} f\left(y_{new}|\boldsymbol{\Theta},\mathbf{x}_c\right)f\left(\boldsymbol{\Theta}\,|\,\mathbf{y}_{old}\right)\partial\boldsymbol{\Theta},$$

and $\mathbf{y}_{old}$ denotes the vector of observed data. The posterior predictive density, which integrates over the uncertainty from estimating the model parameters, is the focus of Miro-Quesada, Del Castillo and Peterson (2004) who discuss process optimization for completely randomized designs with normal responses. We choose to focus on the distribution of $f\left(y_{new}|\boldsymbol{\Theta},\mathbf{x}_c\right)$ because it summarizes what we expect to see under future regular process conditions, and allows for greater flexibility in specifying performance criteria or objective functions for the optimization. Specifically, let the function $h\left(\boldsymbol{\Theta},\mathbf{x}_c\right)$ denote a characteristic of the density in (4). This function is chosen to represent an important aspect of the response upon which process optimization should be based. Given that we have variability from many sources as well as uncertainty from the parameter estimation, we wish to examine what range of values of $h\left(\boldsymbol{\Theta},\mathbf{x}_c\right)$ are possible across the probable parameter values. For example, in the film example, guarding against a worst case scenario for the proportion of high quality film could be more important to the practitioner than estimating the proportion of high quality film using the posterior predictive density. Quite often in RPD, optimization focuses on a criterion such as squared error loss. In the cake example, high taste scores (close to the maximum of 7) are most desirable. Thus, one might wish to find $\mathbf{x}_c$ such that a property of the posterior distribution of

$$h\left(\boldsymbol{\Theta},\mathbf{x}_c\right)=E_{y_{new}|\boldsymbol{\Theta},\mathbf{x}_c}\left[\left\{y_{new}-7\right\}^2 * I\left(y_{new}\le 7\right)\right]$$

is minimized, where

$$E_{y_{new}|\mathbf{\Theta},\mathbf{x}_c}\left[\{y_{new}-7\}^2 * I(y_{new}\le 7)\right] = \int_{y_{new}} \{y_{new}-7\}^2 * I(y_{new}\le 7) f(y_{new}|\mathbf{\Theta},\mathbf{x}_c)\, dy_{new}.$$

In the definition of $h(\mathbf{\Theta},\mathbf{x}_c)$ above, we include the indicator function so as not to penalize any predicted response greater than 7, which is the maximum possible average score for taste.

For the film example, Robinson et al. (2009) suggest that film pieces whose quality exceeds 150 represent premium grade film. As such, it might be desirable to find $\mathbf{x}_c$ such that the exceedence probability $h(\mathbf{\Theta},\mathbf{x}_c) = \Pr(y_{new} > 150|\mathbf{\Theta},\mathbf{x}_c)$ is maximized, using

$$\Pr(y_{new} > 150|\mathbf{\Theta},\mathbf{x}_c) = \int_{y_{new}=150}^{\infty} f(y_{new}|\mathbf{\Theta},\mathbf{x}_c)\, dy_{new}.$$

General Description of Optimization Algorithm

To investigate robust values of $\mathbf{x}_c$, we evaluate properties of the posterior distribution of $h(\mathbf{\Theta},\mathbf{x}_c)$ over a grid of $\mathbf{x}_c$ points. Consider a single point in the grid of $\mathbf{x}_c$ values, say $\mathbf{x}_c^0$. To evaluate properties of the posterior distribution of $h(\mathbf{\Theta},\mathbf{x}_c^0)$, we first obtain a sample of N_{MCMC} $\mathbf{\Theta}'s$ from $f(\mathbf{\Theta}|\mathbf{y}_{old})$ using WinBUGS, and a sample of N_{Noise} observations of $\mathbf{x}_n$ from the user-specified $f(\mathbf{x}_n)$. Then, for each of the $N_{MCMC}\times N_{Noise}$ $(\mathbf{\Theta},\mathbf{x}_n)$ pairs, we sample a value of δ_{new} from the appropriate distribution. For our examples, this is the univariate normal. For each of these $N_{MCMC}\times N_{Noise}$ $(\mathbf{\Theta},\mathbf{x}_n,\delta_{new})$ triples, we sample a y_{new} value from the appropriate normal (first example) or gamma (second example) distribution. Since N_{noise} of these y_{new} values are generated from the same posterior draw, say $\mathbf{\Theta}^0$, we can estimate $h(\mathbf{\Theta}^0,\mathbf{x}_c^0)$. For example, if $h(\mathbf{\Theta}^0,\mathbf{x}_c^0) = \Pr(y_{new} > 150|\mathbf{\Theta}^0,\mathbf{x}_c^0)$,

$$\Pr(y_{new} > 150|\mathbf{\Theta}^0,\mathbf{x}_c^0) \approx \frac{1}{N_{noise}}\sum_{i=1}^{N_{noise}} I\left[y_{new}^i(\mathbf{\Theta}^0) > 150\right],$$

where $y_{new}^{i}\left(\mathbf{\Theta}^{0}\right)$ is the i^{th} value of y_{new} generated according to $\mathbf{\Theta}^{0}$ for a particular choice of $\mathbf{x}_c$. This process gives N_{MCMC} posterior samples of $h\left(\mathbf{\Theta},\mathbf{x}_c^0\right)$, thereby enabling the estimation of quantities such as $E_{\mathbf{\Theta}|\mathbf{y}_{old}}\left[h\left(\mathbf{\Theta},\mathbf{x}_c^0\right)\right]$, $Var_{\mathbf{\Theta}|\mathbf{y}_{old}}\left[h\left(\mathbf{\Theta},\mathbf{x}_c^0\right)\right]$, and the 5th, 10th, and 50th posterior percentiles of the posterior distribution of $h\left(\mathbf{\Theta},\mathbf{x}_c^0\right)$, for instance. The median gives us the center of the distribution while the 5th and 10th percentiles give lower bounds on $h\left(\mathbf{\Theta},\mathbf{x}_c^0\right)$. If we had focused on the posterior predictive distribution, then only a single summary, such as the mean or an exceedance probability is possible. With the proposed approach, understanding the characteristics of the distribution of $h\left(\mathbf{\Theta},\mathbf{x}_c^0\right)$ induced by the uncertainty in $\mathbf{\Theta}$ is possible, and allows for greater flexibility to select a most relevant attribute of the distribution.

Examining these quantities over a grid of $\mathbf{x}_c$ values provides the opportunity to choose $\mathbf{x}_c$ candidates that are robust to changes in $\mathbf{x}_n$. Note that maximizing $E_{\mathbf{\Theta}|\mathbf{y}_{old}}\left[\Pr\left(y_{new}>150|\mathbf{\Theta},\mathbf{x}_c^0\right)\right]$ over values of $\mathbf{x}_c$, is equivalent to maximizing

$\Pr(y_{new}>150\,|\,\mathbf{y}_{old},\mathbf{x}_c)$ over values of $\mathbf{x}_c$, since

$$E_{\mathbf{\Theta}|\mathbf{y}_{old}}\left[\Pr\left(y_{new}>150|\mathbf{\Theta},\mathbf{x}_c^0\right)\right]=\int_{\mathbf{\Theta}}\left[\int_{y_{new}=150}^{\infty} f\left(y_{new}\,|\,\mathbf{\Theta},\mathbf{x}_c^0\right)\partial y_{new}\right]f\left(\mathbf{\Theta}\,|\,\mathbf{y}_{old},\mathbf{x}_c^0\right)\partial\mathbf{\Theta}$$

$$=\int_{y_{new}=150}^{\infty}\int f(y_{new},\mathbf{\Theta}\,|\,\mathbf{y}_{old},\mathbf{x}_c^0)\partial\mathbf{\Theta}\partial y_{new}$$

$$=\Pr\left(y_{new}>150|\mathbf{y}_{old},\mathbf{x}_c^0\right).$$

Also note that optimizing $\Pr\left(y_{new}\in A|\mathbf{y}_{old},\mathbf{x}_c\right)$ over values of $\mathbf{x}_c$ was the approach of Miro-Quesada, Del Castillo and Peterson (2004) but they did so for situations where $f\left(y_{new}|\mathbf{y}_{old},\mathbf{x}_c\right)$ exists in closed form.

To facilitate a fair comparison for each $\mathbf{x}_c$ in the grid, we use the same sample of $N_{MCMC}xN_{Noise}$ $\left(\mathbf{\Theta},\mathbf{x}_n\right)$ pairs for each $\mathbf{x}_c$ in the grid. Further, when the analyses in both examples were re-done, using different samples of the $N_{MCMC} \times N_{Noise} (\boldsymbol{\Theta}, \mathbf{x}_n)$ pairs, the results were essentially unchanged.

Results for the Examples

Cake Mix Experiment Results

Using the sample of N_{MCMC} posterior $\theta's$ obtained from WinBUGS, we now seek to find the optimal combination of control factors (Flour, Shortening and Eggs) to minimize the expected squared loss from the target taste value of 7. These settings are chosen to be robust to variations in the cooking time and temperature which are not as tightly controlled when consumers bake the cake compared to the laboratory environment. Similar to other RPD applications, we assume that the distribution of the coded noise variables (cooking time and temperature) in the less controlled consumer environment is appropriately modeled with a bivariate normal distribution with mean (0, 0) and variance I_2.

When we examine the optimizations based on different criteria from the posterior distribution of $h(\boldsymbol{\Theta}, \mathbf{x}_c) = E\left[\{y_{new} - 7\}^2 * I(y_{new} \leq 7)\right]$ over the design region, we see that different settings are identified as best depending on the criterion considered. Table 3 shows various locations of optima along with their respective optimum values depending upon which characteristic of the posterior of $h(\boldsymbol{\Theta}, \mathbf{x}_c)$ one is interested in optimizing with respect to. Note that all optimal settings have Flour and Eggs set to their maximum amount (1 in the coded factors), but the amount of Shortening to be included varies by criterion. The optimization actually proceeded in two stages. In the first stage, $\boldsymbol{x}_c = (F, S, E)$ is optimized over the grid $\{-1, -0.8, \cdots, 0.8, 1\}^3$. In the second stage, optimization is performed over the finer grid $\{0.5, 0.55, \cdots, 0.95, 1\}^3$. By considering several properties of the posterior of $h(\boldsymbol{\Theta}, \mathbf{x}_c)$, rather than just a single summary, we are able to gain a more detailed understanding of how to best optimize the process. For example, the smallest mean and median of the posterior for $h(\boldsymbol{\Theta}, \mathbf{x}_c)$ occurs at the control factor setting of (1, 1, 1) with corresponding values 2.93 and 2.83, respectively. The optimal control factor setting shifts more towards middle values of Shortening to optimize the tail of the posterior distribution of $h(\boldsymbol{\Theta}, \mathbf{x}_c)$. Specifically, the 90th percentile is minimized at the control factor setting (1, 0.8, 1) with a value of 4.08 and the minimum 95th percentile is 4.44 at the control factor setting (1, 0.6, 1). Hence if the manufacturer wishes to optimize the typical taste characteristic for the consumer, then the optimal setting should be chosen as (1, 1, 1). However, if it is a priority for the company to make sure that very few consumers experience poor taste, then the setting of (1, 0.8, 1) or (1, 0.6, 1) should be considered as best.

Figure 4 shows the contour plot of the mean of the posterior distribution of $h(\boldsymbol{\Theta}, \mathbf{x}_c)$ across the design region. The first three sub-figures show contour plots for different fixed values of Flour across the range of design Egg and Shortening factors. The final sub-figure shows an enlargement of the optimal corner of the design region with more detailed contours.

Figure 5 shows contour plots for different attributes of the posterior distribution of $h(\boldsymbol{\Theta}, \mathbf{x}_c)$ with Flour fixed at a value of 1. Examining the plots, we see that the shape of the contour lines changes for the different percentiles of the posterior, with the median optimized in the corner of the design space, while the optimum control factor settings for tail percentiles favors smaller amounts of Shortening. While there are differences in the global optimum, we can see that the optimal RPD operation setting does not change the amounts of Flour and Eggs. Given the flat contour lines in the Shortening direction for Flour and Eggs at their maximums, these differences in the optimal settings represent some alternatives which perform quite well across a range of different distributional characteristics.

Finally, we compare the results of our optimization to those that would have been obtained if we had used a criterion of maximizing the proportion of the posterior predictive distribution above a certain threshold using methodology similar to what was outlined in Miro-Quesada et al. (2004). In this case, if one maximizes the proportion of responses above the taste score thresholds of 6 and 6.5 according to the posterior predictive distribution, similar but not identical results are found compared to those above. Optimal control factor settings for maximizing $\Pr(y_{new} > 6 \mid \mathbf{y}_{old})$ and $\Pr(y_{new} > 6.5 \mid \mathbf{y}_{old})$ along with the corresponding values are found in the last two rows of

Table 3. In the case of the posterior predictive distribution, the proportions of Flour and Eggs are maintained at their maximum values, but the optimal setting of Shortening is at its low level. The global best control factor setting is (1, -1, 1), with 37% and 24% of the distribution predicted to fall above each threshold, respectively. Based on these results, it is clear that the Flour and Eggs settings should be maximized, while the criterion for optimizing determines the value of the Shortening.

Film Manufacturing Experiment Results

We now consider RPD optimization for the film manufacturing example using the sample of N_{MCMC} posterior $\theta's$ obtained from a Bayesian analysis of the data in WinBUGS. The goal of this analysis is to determine the control/mixture factor level combination [X_1, X_2, X_3] which maximizes the proportion of high quality film given the unwanted variation from the process variables [p_1, p_2, p_3] in the less controlled production environment. As with other RPD applications, we assume that the coded noise factors are from a trivariate normal distribution with mean (0, 0, 0) and variance, I_3. The optimal settings and corresponding values based on different criteria from the posterior distribution of $h(\boldsymbol{\Theta}, \mathbf{x}_c) = \Pr(y_{new} > 150 | \boldsymbol{\Theta}, \mathbf{x}_c)$ are summarized in Table 4. For all summaries of the posterior of $h(\boldsymbol{\Theta}, \mathbf{x}_c) = \Pr(y_{new} > 150 | \boldsymbol{\Theta}, \mathbf{x}_c)$, the goal is to find the control factor settings resulting in maximization. We utilized a grid search to find the optimal mean, median, 5th and 10th percentiles of the posterior distribution of $h(\boldsymbol{\Theta}, \mathbf{x}_c)$. As in the cake mix example, the optimization proceeds in two stages. In the first stage, $\boldsymbol{x}_c = (x_2, x_3)$ is optimized over the grid $\{0, 0.02, \cdots, 0.78, 0.8\} \times \{0.2, 0.21, \cdots, 0.39, 0.4\}$. In the second stage, optimization is over the finer grid $\{0.1, 0.1075, \cdots, 0.3825, 0.4\} \times \{0.2, 0.2025, \cdots, 0.2475, 0.25\}$. The best mixture combination based on the mean of $\Pr(y_{new} > 150 | \boldsymbol{\Theta}, \mathbf{x}_c)$ is at $(X_1, X_2, X_3) = (0.5425, 0.2575, 0.2000)$. Figure 6 provides the contour plots of the mean of $\Pr(y_{new} > 150 | \boldsymbol{\Theta}, \mathbf{x}_c)$ across the control factor space. Since the control factors are mixture variables in this experiment (constrained to sum to one), the entire design region for the outer array is shown in two dimensions (X_2 and X_3) with $X_1 = 1 - X_2 - X_3$

inferred. Figure 6(a) shows the control factor design space, while Figure 6(b) shows an enlargement of the region with the optimum, shown with the dashed line in Figure 6(a). The surface corresponding to the mean of the posterior distribution of $h(\boldsymbol{\Theta}, \mathbf{x}_c)$ is quite smooth with a single maximum on the boundary of the space. We note that this solution is similar to the approach of Miro-Quesada et al. (2004) as it maximizes the proportion of the posterior predictive distribution above 150. For our implementation, we have estimated the value using the MCMC draws, rather than finding a closed form expression for the quantity.

Alternately, we consider different percentiles of the posterior of $h(\boldsymbol{\Theta}, \mathbf{x}_c)$ for selecting the best production recipe for film. The settings of the optimum based on the 50th, 5th and 10th percentiles of the posterior distribution of $h(\boldsymbol{\Theta}, \mathbf{x}_c)$ are also given in Table 4. All of the identified optima have $X_3 = 0.2$, but the proportions of X_1 and X_2 differ depending on the criterion considered. For the median, the estimated proportion of high quality film is 0.886, with two distinct settings, $(X_1, X_2, X_3) = (0.535, 0.265, 0.200)$ or $(0.52, 0.28, 0.20)$, producing this value. The median or mean of $h(\boldsymbol{\Theta}, \mathbf{x}_c)$'s posterior distribution represent "typical" proportions of high quality film under expected noise condition variation, while the lower percentiles (5th and 10th) monitor the "worst case" proportion of high quality film under the uncertainty in the model parameters. By considering the proportion of high quality film in the tails of the posterior distribution (0.623 and 0.719 for the 5th and 10th percentiles, respectively), we gain a more realistic assessment of what proportions might be observed during regular production. The choice of which metric to select depends on whether we want to optimize typical proportions of high quality film or determine a setting where we can be very confident of attaining at least a certain proportion. Figures 4, 5 and 6 shows contour plots for the median, 5th and 10th percentiles, respectively, across the design region. The second sub-figure 10 in each plot shows an enlargement of the contour plot around the optimum for the area denoted by the dashed line in the first sub-figure.

As we protect more against the tail of the posterior distribution, the X_2 proportion decreases to 0.2050 (with X_1 increasing to 0.5950). Quantifying the trade-off in both settings and the proportion of high quality film expected can help managers make more informed decisions about what to expect during regular production. Table 5 shows the values of the different criteria associated with $h(\boldsymbol{\Theta}, \mathbf{x}_c)$ for two settings which correspond to the optimal settings for the median and 5th percentile criteria. If we choose based on the median or "typical proportion", then an optimal setting is $(X_1, X_2, X_3) = (0.535, 0.265, 0.200)$. However, if we used the 5th percentile as a conservative measure to almost guarantee at least a certain proportion of high quality film, then the optimal setting is $(X_1, X_2, X_3) = (0.595, 0.205, 0.200)$. When we consider the results in Table 5, we see that while the optimum setting shifts slightly depending on the chosen criterion, none of the competing criteria perform too badly at the optima of another. By considering the entire posterior distribution of $\Pr(y_{new} > 150 | \boldsymbol{\Theta}, \mathbf{x}_c)$, and matching its characteristics to aspect of production on which we wish to focus, we can make a more informed choice for how to optimize the process.

Conclusions and Discussion

In this paper we present a method for finding optimally robust settings in the control factor space for different characteristics of the posterior distribution of an objective function. The Bayesian approach allows flexible objective functions for a very general class of models. In the two examples, the responses were normal and gamma distributed; the proposed methodology easily handles other response distributions such as the Weibull and lognormal distributions for continuous responses and the binomial and Poisson distributions for discrete responses. The methods presented were compared to the optimization based on the posterior predictive distribution described in Quesada et al. (2005) and shows the two approaches are related if the mean of the posterior distribution is selected. In other cases, the more general new approach allows different functions of the model parameters to be considered, and different attributes of the posterior distribution to be utilized. This flexibility allows the user to specify an attribute of the distribution that most realistic summarizes what is of primary interest in the experiment. Assessing the robustness of the optimization results across different potential attributes is straightforward and allows direct comparison of trade-offs.

Using the optimization algorithm, it is straightforward to substitute more complicated noise variable structures, including correlations, into the optimization. For example, in the film experiment, the distribution of the coded noise factors in the production environment could have been replaced with a multivariate normal distribution with mean (0,0,0) and variance $\begin{pmatrix} 1 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & 1 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & 1 \end{pmatrix}$, with no additional complication other than how to sample the N_{noise} values.

In the cake example, there were no whole plot replicates to allow estimation the variance of σ_δ^2 from the data. Hence the prior distribution provided the needed information to complete the analysis. In general, the authors feel that whenever possible designs should be selected with adequate whole plot replication to provide some confirmation of the range of values for σ_δ^2. However, when no replication is possible, the Bayesian approach allows incorporating external knowledge of uncertainty into the analysis.

Acknowledgements

We thank Derek Bingham of Simon Fraser University and Alyson Wilson of Iowa State University for helpful discussions.

References

Box, G.E.P. and Jones, S. (1992). "Split-Plot Designs for Robust Product Experimentation". *Journal of Applied Statistics*, 19, pp. 3-26.

Casella, G. and George, E. (1992). "Explaining the Gibbs Sampler". *The American Statistician*, 46, pp. 167-174.

Chib, S. and Greenberg, E. (1995). "Understanding the Metropolis-Hastings Algorithm". *The American Statistician*, 49, pp. 327-335.

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

DeGroot, M.H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

Ganju, J. and Lucas, J.M. (1997). "Bias in Test Statistics when Restrictions in Randomization are Caused by Factors". *Communications in Statistics: Theory and Methods*, 26, pp. 47-63.

Gelfand, A.E. and Smith, A.F.M. (1990). "Sampling-Based Approaches to Calculating

Marginal Densities". *Journal of the American Statistical Association*, 85, pp. 398-409.

Gelman, A. (2006). "Prior Distributions for Variance Parameters in Hierarchical Models". *Bayesian Statistics*, 1, pp. 515-533.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004), *Bayesian Data Analysis, Second Edition*, Chapman & Hall/CRC, Boca Raton, FL.

Goos, P., Langhans, I. and Vandebroek, M. (2006). "Practical Inference from Industrial Split-Plot Designs". *Journal of Quality Technology*, 38, pp. 162-179.

Johnson, V.E. (2004), "A Bayesian Chi-Square Test for Goodness-of-Fit," *Annals of Statistics*, 32, 2361-2384.

Lee, Y. and Nelder, J.A. (2003). "Robust Design Via Generalized Linear Models". *Journal of Quality Technology*, 35, pp. 2-12.

Lee, Y., Nelder, J.A., and Park, H. (2010), "HGLMs for Quality Improvement". *Applied Stochastic Models in Business and Industry*. DOI: 10.1002/asmb.840

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

Myers, R.H., Brenneman, W.A and Myers, W.R. (2005). "A Dual Response Approach to Robust Parameter Design for a Generalized Linear Model". *Journal of Quality Technology*, 37, pp. 130-138.

Rajagopal, R., Del Castillo, E. and Peterson, J.J. (2005). "Model and Distribution-Robust Process Optimization with Noise Factors". *Journal of Quality Technology* 37, pp. 210-222.

Robinson, T.J., Myers, R.H. and Montgomery, D.C. (2004). "Analysis Considerations in Industrial Split-Plot Experiments with Non-Normal Responses". *Journal of Quality Technology*, 36, pp. 180-192.

Robinson, T.J., Anderson-Cook, C.M. and Hamada, M.S. (2009). "Bayesian Analysis of Split-Plot Experiments with Non-Normal Responses for Evaluating Non-Standard Performance Criteria". *Technometrics*, 51, pp. 56-65.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2004). *WinBUGS Version 1.4 User Manual.*

Wolfinger, R.D. and Tobias R.D. (1998). "Joint Estimation of Location, Dispersion, and Random Effects in Robust Design". *Technometrics*, 40, pp. 62-71.

TABLE 1. Posterior Means, Standard Deviations, and Selected Quantiles for Regression Coefficients $(\beta's)$, Random Effect Variance $\left(\sigma_\delta^2\right)$ and Subplot Variance $\left(\sigma_\varepsilon^2\right)$ for the Cake Mix Experiment

			Quantiles		
Parameter	**Mean**	**Std. Dev.**	**0.025**	**0.5**	**0.975**
Intercept	3.444	0.162	3.127	3.444	3.760
Time (*t*)	0.268	0.159	-0.048	0.269	0.575
Temp (*T*)	0.553	0.158	0.240	0.552	0.862
$t \times T$	0.034	0.158	-0.278	0.034	0.338
Flour (*F*)	1.336	0.098	1.142	1.334	1.525
Shortening (*S*)	-0.167	0.100	-0.366	-0.167	0.034
Egg (*E*)	0.603	0.098	0.409	0.604	0.797
$F \times S$	0.098	0.099	-0.097	0.098	0.292
$F \times E$	0.103	0.099	-0.092	0.103	0.302
$S \times E$	0.065	0.099	-0.131	0.065	×0.263
$F \times T$	0.154	0.098	-0.039	0.155	0.347
$F \times t$	0.004	0.098	-0.191	0.003	0.197
$S \times T$	-0.461	0.098	-0.656	-0.461	-0.268
$S \times t$	-0.110	0.099	-0.307	-0.108	0.084
$E \times T$	-0.015	0.099	-0.216	-0.015	0.182
$E \times t$	-0.043	0.098	-0.238	-0.043	0.151
σ_ε^2	0.310	0.133	0.145	0.280	0.644
σ_δ^2	0.062	0.002	0.058	0.062	0.067

TABLE 2. Posterior Means, Standard Deviations, and Selected Quantiles for Regression Coefficients $(\beta's)$, Gamma Shape Parameter (α), Random Effects $(\delta_i's)$ and Random Effect Variance (σ_δ) for the Film Manufacturing Experiment

Effect	Estimate	Standard Dev.	Quantile 0.025	0.5	0.975
$x1$	6.169	0.4533	5.284	6.166	7.104
$x2$	4.035	0.4661	3.116	4.031	4.980
$x3$	1.265	0.9708	-0.7040	1.2780	3.213
$x1 \times x2$	10.550	2.8080	4.896	10.52	16.20
$p1$	-0.7074	0.5317	-1.7530	-0.7132	0.3595
$p2$	0.3728	0.5351	-0.6979	0.3743	1.417
$p3$	0.2952	0.5323	-0.7399	0.2886	1.344
$x1 \times p1$	1.217	0.7490	-0.2781	1.224	2.668
$x2 \times p1$	1.992	0.7506	0.5145	1.997	3.467
$x1 \times p2$	-0.2336	0.7551	-1.6920	-0.2372	1.283
$x2 \times p2$	-1.040	0.7582	-2.521	-1.039	0.4820
$x1 \times p3$	-0.4783	0.7477	-1.9690	-0.4745	0.9840
$x2 \times p3$	-0.1303	0.7524	-1.6230	-0.1255	1.333
α	4.835	1.167	2.861	4.717	7.412
δ_1	0.1236	0.2550	-0.3563	0.1084	0.6605
δ_2	-0.0301	0.2536	-0.5550	-0.0258	0.4871
δ_3	0.1519	0.2558	-0.3296	0.1310	0.6943
δ_4	0.0718	0.2560	-0.4311	0.05894	0.6122
δ_5	-0.05651	0.2552	-0.6006	-0.04433	0.4486
δ_6	-0.2264	0.2611	-0.7915	-0.2066	0.2395
δ_7	-0.04031	0.2986	-0.6639	-0.0308	0.5665
δ_8	-0.3423	0.2814	-0.9395	-0.3209	0.1273
δ_9	0.2259	0.2643	-0.2499	0.2045	0.7945
δ_{10}	0.1295	0.2582	-0.3637	0.1103	0.6803
δ_{11}	0.2589	0.2579	-0.1818	0.2358	0.8286
δ_{12}	0.04056	0.2991	-0.5474	0.03139	0.6818
δ_{13}	-0.3369	0.2880	-0.9675	-0.3139	0.1369
σ_δ	0.3372	0.1606	0.06390	0.3212	0.7082

TABLE 3. Optimal Control Factor Settings and Values for Different Summaries of $h(\boldsymbol{\Theta}, \mathbf{x}_c) = E\left[\{y_{new} - 7\}^2 * I(y_{new} \le 7)\right]$ and the Posterior Predictive Distribution for the Cake Mix Experiment

Minimizing summaries of $h(\boldsymbol{\Theta}, \mathbf{x}_c) = E\left[\{y_{new} - 7\}^2 * I(y_{new} \le 7)\right]$				
	Setting			Value
	Flour	Shortening	Eggs	
Mean	1.0	1.0	1.0	2.93
Median	1.0	1.0	1.0	2.83
90th Percentile	1.0	0.8	1.0	4.08
95th Percentile	1.0	0.6	1.0	4.44
Maximization Using Posterior Predictive Distribution				
$\Pr(y_{new} > 6 \mid \mathbf{y}_{old})$	1.0	-1.0	1.0	0.37
$\Pr(y_{new} > 6.5 \mid \mathbf{y}_{old})$	1.0	-1.0	1.0	0.24

TABLE 4: Optimal Control Factor Settings and Values for Different Summaries of $h(\boldsymbol{\Theta}, \mathbf{x}_c) = \Pr(y_{new} > 150 | \boldsymbol{\Theta}, \mathbf{x}_c)$ for the Film Manufacturing Experiment

	Setting			Value
	X1	**X2**	**X3**	
Mean*	0.5425	0.2575	0.2000	0.854 (maximum)
Median	0.5350 0.5200	0.2650 0.2800	0.2000 0.2000	0.886 (maximum)
5th Percentile	0.5950	0.2050	0.2000	0.623 (maximum)
10th Percentile	0.5650 0.5500	0.2350 0.2500	0.2000 0.2000	0.719 (maximum)

* equivalent to the posterior predictive optimization of Quesada et al. (2005)

TABLE 5: Posterior Percentiles of $h(\boldsymbol{\Theta}, \mathbf{x}_c) = \Pr(y_{new} > 150 | \boldsymbol{\Theta}, \mathbf{x}_c)$ at Two Optimal Control Factor Settings for the Film Manufacturing Experiment

X1	**X2**	**X3**	**Mean**	**Median**	**5th %ile**	**10th %ile**	**Variance**
0.5350	0.2650	0.2000	0.854	0.886*	0.614	0.716	0.015
0.5950	0.2050	0.2000	0.847	0.875	0.623*	0.714	0.014

* denotes best value across all control factor locations in design space

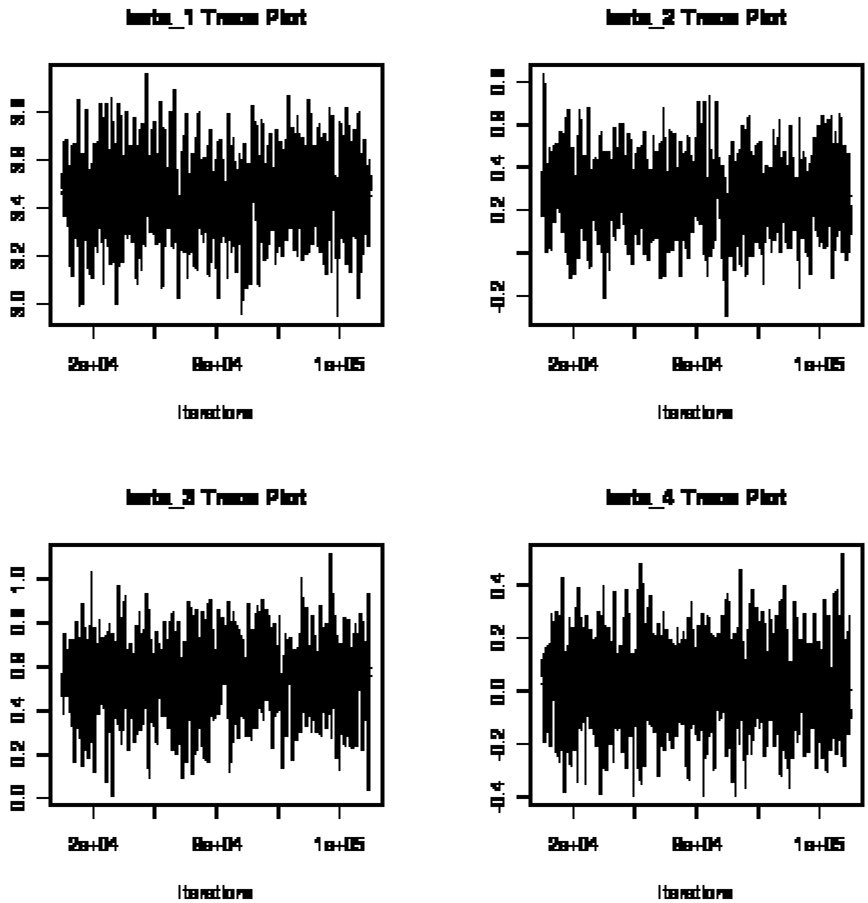


FIGURE 1: Trace Plots of the MCMC Samples for the First Four Regression Coefficients in the Cake Mix Experiment.

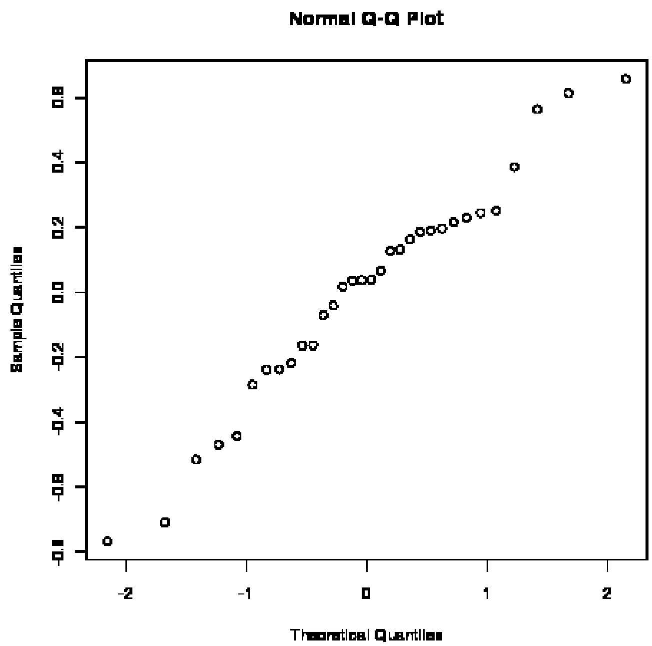


FIGURE 2: Q-Q plot of the Medians of the Posterior Distributions of the Raw Residuals for the Cake Mix Experiment.

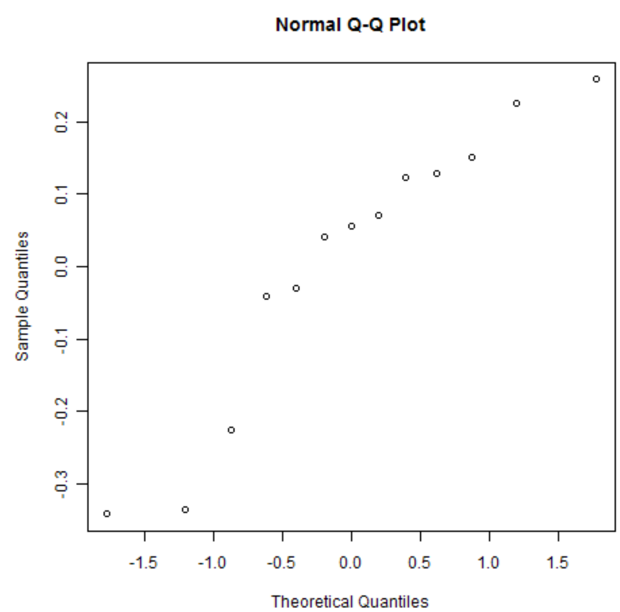


FIGURE 3: Q-Q Plot of the Posterior Means of the Random Effects in the Film Manufacturing Experiment.

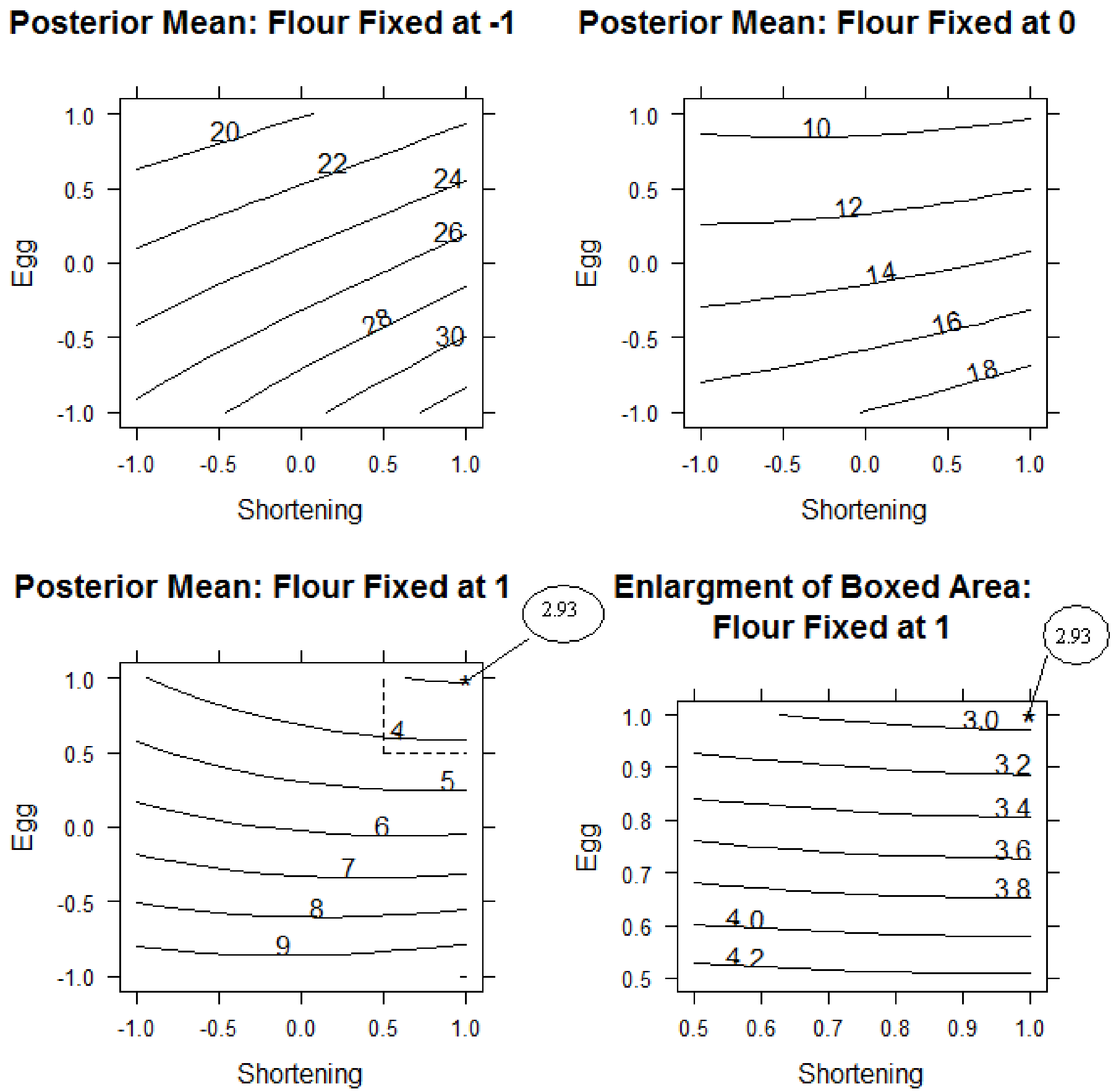


FIGURE 4: Contour Plots of Mean of Posterior Distribution for $h(\boldsymbol{\Theta}, \mathbf{x}_c) = E\left[\{y_{new} - 7\}^2 * I(y_{new} \le 7)\right]$ at Different Fixed Values of the Coded Flour Factor for the Cake Mix Experiment. Asterisk (*) denotes the optimal control factor setting.

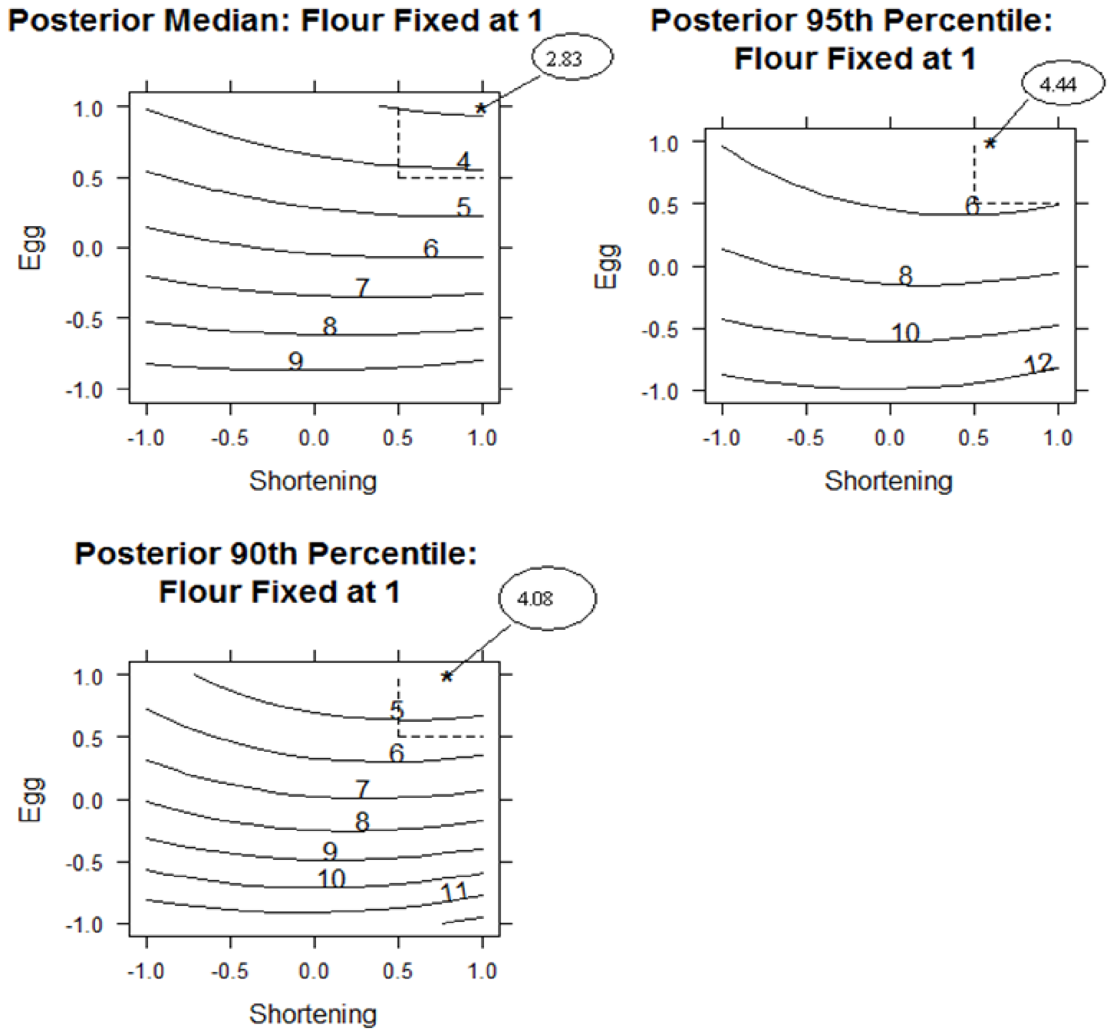
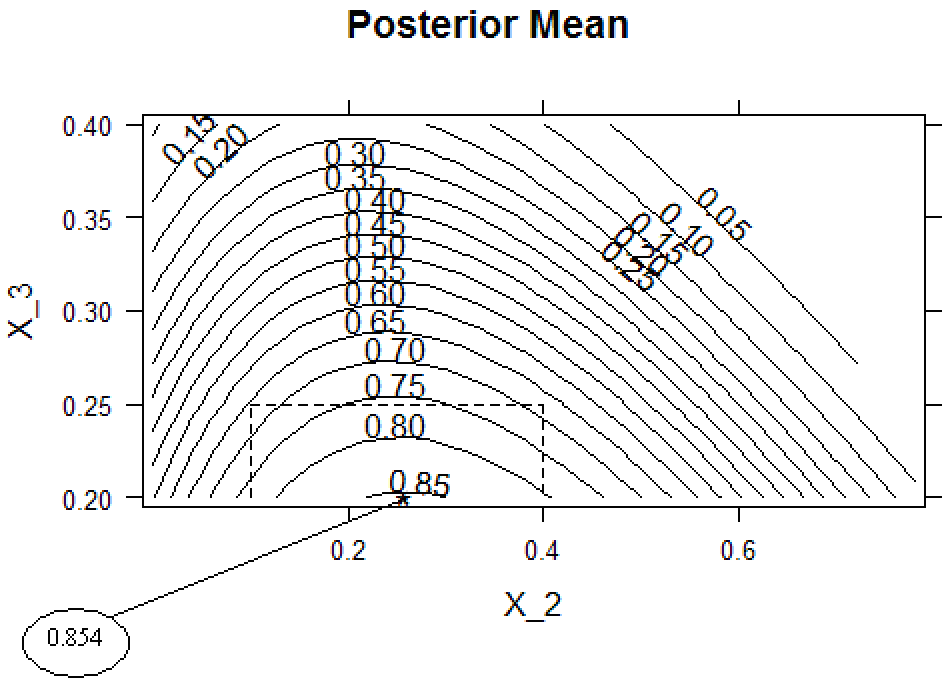


FIGURE 5: Contour Plots of Median, 90^{th} and 95^{th} Percentiles of the Posterior Distribution for $h(\boldsymbol{\Theta}, \mathbf{x}_c) = E\left[\{y_{new} - 7\}^2 * I(y_{new} \le 7)\right]$ Coded Flour factor Set at 1 for the Cake Mix Experiment. Asterisk (*) denotes the optimal control factor setting.

(a)



(b)

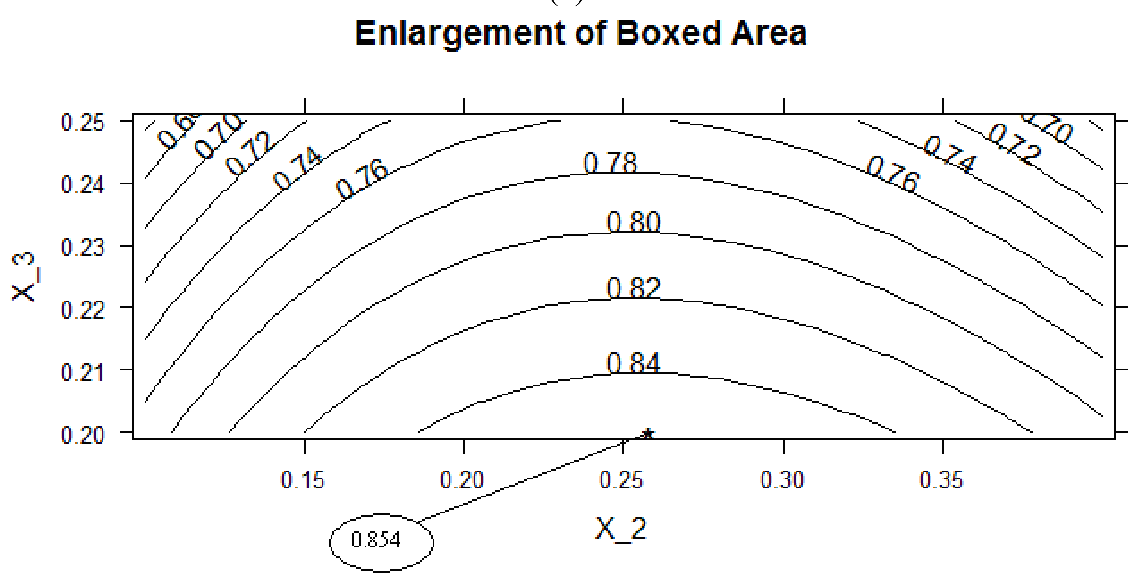
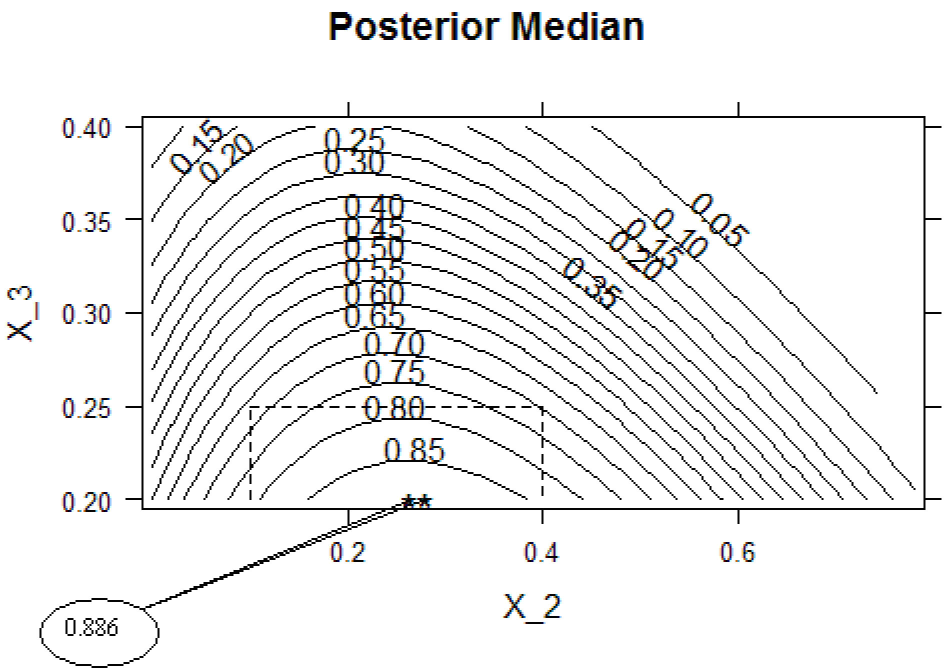


FIGURE 6: Contour Plots of Mean of Posterior Distribution for $h(\Theta, \mathbf{x}_c) = \Pr(y_{new} > 150 | \Theta, \mathbf{x}_c)$ for the Film Manufacturing Example: (a) Entire Design Region, (b) Enlargement of Region Close to Optimum Denoted by Dashed Line in (a). Asterisk (*) denotes the optimal control factor setting.

(a)



(b)

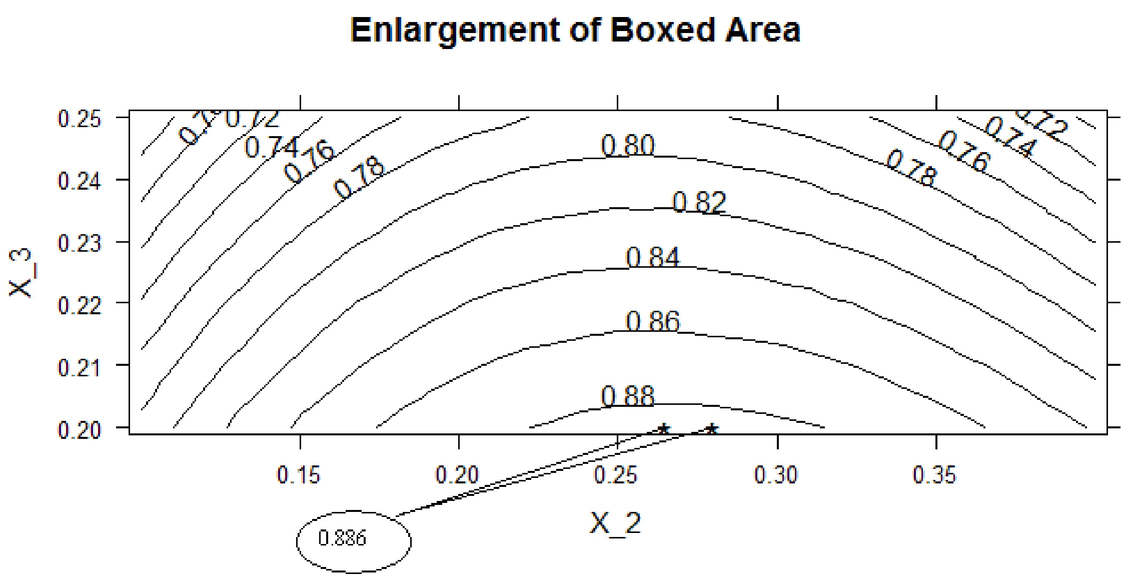
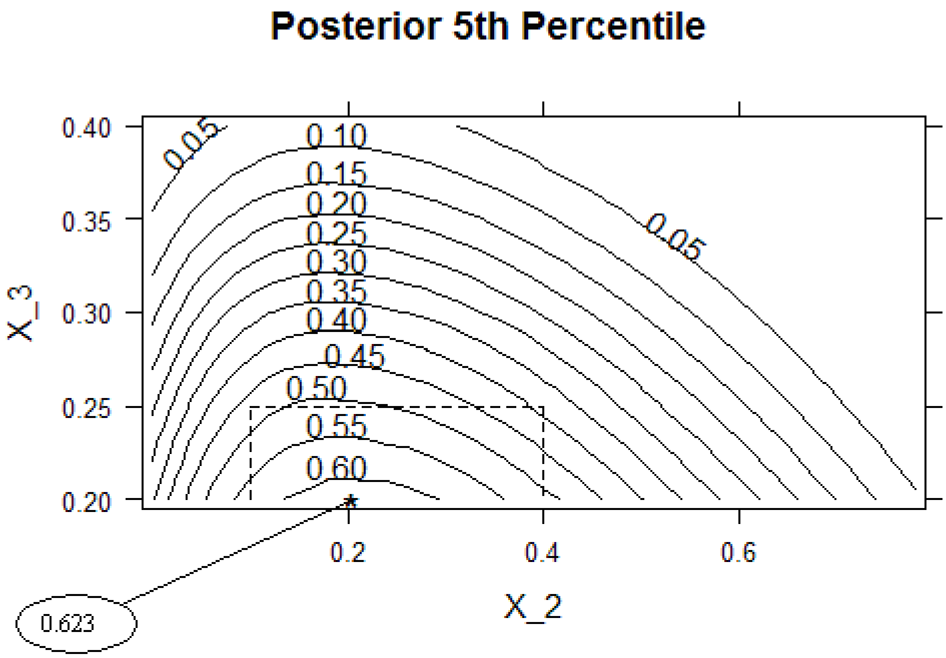


FIGURE 7: Contour Plots of Median of Posterior Distribution for $h(\Theta, \mathbf{x}_c) = \Pr(y_{new} > 150 | \Theta, \mathbf{x}_c)$ for the Design Space Region in the Film Manufacturing Experiment: (a) Entire Design Region, (b) Enlargement of Region Close to Optimum Denoted by Dashed Line in (a). Asterisk (*) denotes the optimal control factor setting.

(a)



(b)

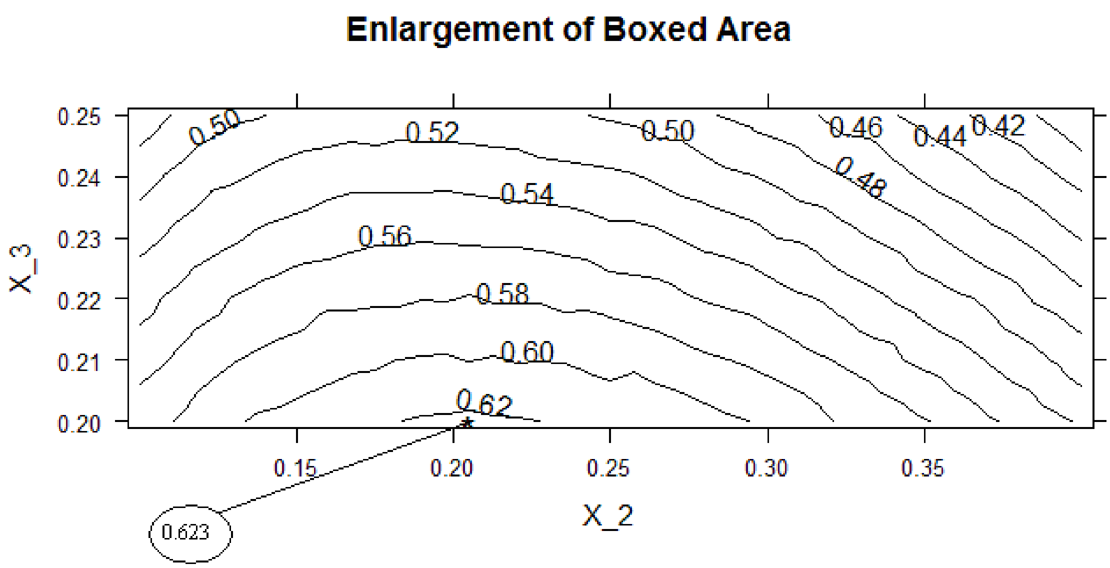
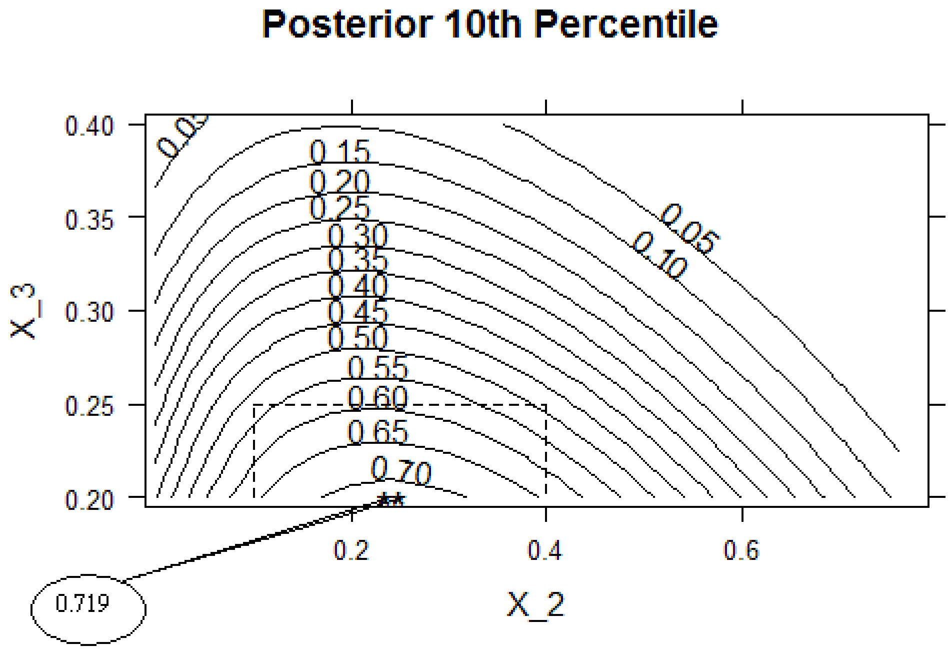


FIGURE 9: Contour Plots of 5th Percentile of Posterior Distribution for $h(\Theta, \mathbf{x}_c) = \Pr(y_{new} > 150 | \Theta, \mathbf{x}_c)$ for the Design Space Region in the Film Manufacturing Experiment: (a) Entire Design Region, (b) Enlargement of Region Close to Optimum Denoted by Dashed Line in (a). Asterisk (*) denotes the optimal control factor setting.

(a)



(b)

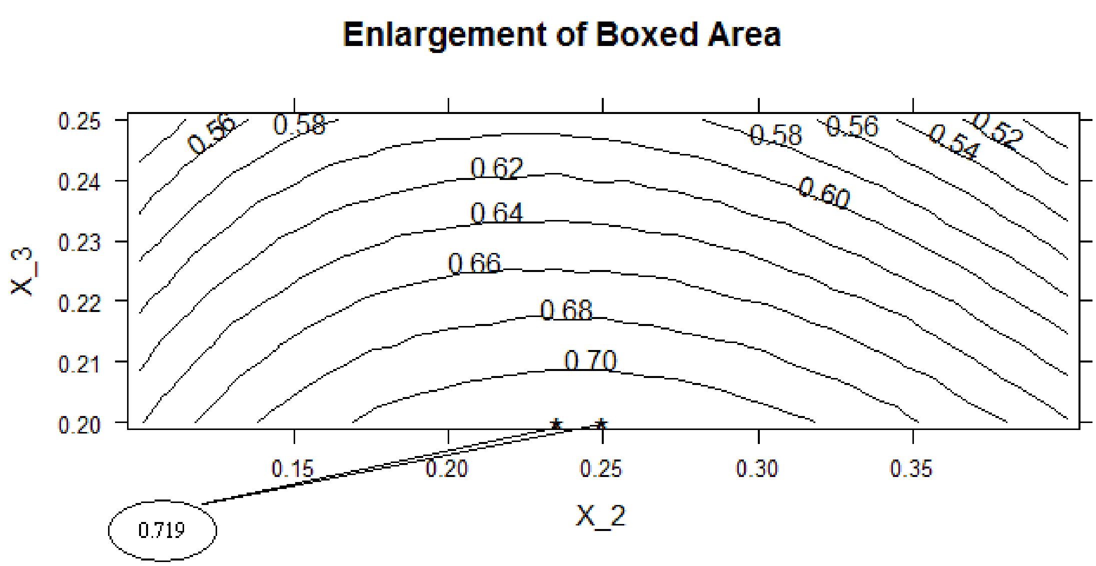


FIGURE 10: Contour Plots of 10th Percentile of Posterior Distribution for $h(\Theta, \mathbf{x}_c) = \Pr(y_{new} > 150 | \Theta, \mathbf{x}_c)$ for the Design Space Region in the Film Manufacturing Experiment: (a) Entire Design Region, (b) Enlargement of Region Close to Optimum Denoted by Dashed Line in (a). Asterisk (*) denotes the optimal control factor setting.