A Posterior Predictive Approach to Multiple Response Surface Optimization

JOHN J. PETERSON

GlaxoSmithKline Pharmaceuticals, King of Prussia, PA 19406-0939

In this paper we present an approach to multiple response surface optimization that not only provides optimal operating conditions, but also measures the reliability of an acceptable quality result for any set of operating conditions. The most utilized multiple response optimization approaches of “overlapping mean responses” or the desirability function do not take into account the variance-covariance structure of the data nor the model parameter uncertainty. Some of the quadratic loss function approaches take into account the variance-covariance structure of the predicted means, but they do not take into account the model parameter uncertainty associated with the variance-covariance matrix of the error terms. For the optimal conditions obtained by these approaches, the probability that they provide a good multivariate response, as measured by that optimization criterion, can be unacceptably low. Furthermore, it is shown that ignoring the model parameter uncertainty can lead to reliability estimates that are too large. The proposed approach can be used with most of the current multiresponse optimization procedures to assess the reliability of a good future response. This approach takes into account the correlation structure of the data, the variability of the process distribution, and the model parameter uncertainty. The utility of this method is illustrated with two examples.

KEY WORDS: Bayesian Analysis; Desirability Function; Mixture Experiments; Preposterior Analysis; Proportion Conforming; Quadratic Loss Function; Reliability; Ridge Analysis; Simulation.

Introduction

Multiple response surface methods, based on mean response values, have proven useful for improving the quality of products and processes. Many examples of multiple response experiments can be found in the literature, including books by Khuri and Cornell (1996) and Myers and Montgomery (2002). In addition, popular statistical packages, such as SAS JMP and Design Expert, have multiple response optimization procedures based on mean responses. However, very little has been published about assessing the reliability that “optimal” operating conditions will produce good quality results, especially in multiple response situations.

One simple way of dealing with multiple response surfaces when there are only a small number of factors is to create overlapping contour plots of the response surfaces. This may then indicate a region in the factor space where each of the mean responses satisfies the experimenter’s requirements or where a compromise can perhaps be obtained. For a listing of articles providing examples or discussion of this approach, see Montgomery and Bettencourt (1977).

A somewhat more sophisticated way of dealing with multiple responses is to construct some type of “desirability” function that maps the individual influence of each of the response types into a single scalar response. This way, optimization can be achieved by treating the desirability function as an objective function for optimization with respect to the experimental factor levels. This idea was first put forth by Harrington (1965) and later refined by Derringer and Suich (1980) and Del Castillo, Montgomery, and McCarville (1996). These desirability function approaches employ a metric on [0, 1], which is the geometric mean of individual metrics on [0, 1], one for each response type. In addition, the individual metrics (i.e., individual desirability functions) can be weighted so that more emphasis is
placed on certain response types. This approach allows an engineer or scientist to incorporate desirability information relating to economic or specific performance issues. Harrington (1965) also created an absolute scale to accompany his desirability function, while other desirability functions only provide relative measures.

A major drawback of the “overlapping mean response” and desirability function methods is that they ignore the correlations among the responses and the variability of the predictions. Understanding the variability of the predictions has been stressed by Myers (1999) as a critical issue for practitioners. It is certainly a critical issue for quality assessment. In addition, these approaches do not take into account the uncertainty of the estimates of the model parameters.

As an alternative to the product [0, 1]-metric approaches above, Khuri and Conlon (1981), Pignatiello (1993), Ames et al. (1997), and Vining (1998) proposed various types of quadratic loss function procedures. Except for that of Ames et al. (1997), the rest of these quadratic loss function approaches take into consideration the correlation structure of the response types. While this is an advantage over the overlapping-mean and desirability-function approaches, these quadratic loss function methods still have two serious drawbacks. First, they do not take into account the uncertainty of the variance-covariance matrix of the regression model errors. Second, a quadratic loss function, especially one based on a multivariate quadratic form, may be difficult for some investigators to intuitively grasp, although Vining (1998) has stated that squared-error loss is familiar to many engineers.

Del Castillo (1996) proposed a constrained confidence region approach to multiresponse optimization that addresses response uncertainty and model parameter uncertainty to some degree, and it appears easier to understand than a multivariate quadratic form. However, it requires the type of problem that can be clearly and appropriately reformulated as having a primary response variable and various secondary response variables. In addition, the correlations among the response types are not addressed.

Recently, Chiasso and Hamada (2001) took an important step in addressing the assessment and optimization of the quality of a multiple response experiment. They provided a procedure to estimate the probability that a multivariate normal response will satisfy user-specified conditions. Their method is good in that it takes into account the variance-covariance structure of the data, can accommodate heteroscedastic and noise variable regression models, and is easy to interpret. A serious drawback of this approach, however, is that it does not take into account the uncertainty of the model parameters.

In this paper, a Bayesian reliability approach is proposed that takes into account the correlation structure of the data, the variability of the process distribution, and the model parameter uncertainty. This Bayesian approach utilizes the posterior predictive distribution of the multivariate response to compute the probability that a future multivariate response will satisfy specified quality conditions. Bayesian predictive distributions are a natural tool for computing such reliabilities (Bayarri and Mayoral (2002)). These reliabilities can involve proportions of conforming units, desirability functions, or any loss functions. This approach also allows an investigator to perform a “preposterior analysis” to see how reduction of model parameter uncertainty through acquiring additional data can modify the reliability of a multivariate response. The utility of this method is illustrated with two examples. The first example illustrates how reducing process variation and increasing sample size can improve poor reliabilities. The second example shows that this approach is able to assert that additional experimentation to reduce process variation or increase sample size is not necessary. This example also demonstrates how to perform a ridge analysis with the Bayesian reliability procedure of this paper.

A Posterior Predictive Approach

The Concept

We let $\mathbf{Y} = (Y_1, Y_2, ..., Y_p)'$ be the multivariate $(p \times 1)$ response vector and let $\mathbf{x} = (x_1, x_2, ..., x_k)'$ be the $(k \times 1)$ vector of factor variables. If the experimenter is simply interested in $\mathbf{Y}$ being in some desirable subset of the response space, $A$, then he or she should consider the discrete “desirability” function, $I(\mathbf{y} \in A)$, where $I(\cdot)$ is the 0-1 indicator function. If one can sample from the posterior
predictive distribution of \( Y \) for a given \( x \) value, then one can construct the posterior predictive distribution of \( I(Y \in A) \) conditional on \( x \). From a quality perspective, this gives the experimenter a measure of the reliability of \( Y \) being in \( A \) for a given \( x \). This measure of course takes into account the variance-covariance structure of the data and the uncertainty of the model parameters through the posterior predictive distribution of \( Y \). A search of the \( x \)-space then provides the experimenter with information on conditions for optimizing the reliability of \( Y \) being in \( A \).

Likewise, if \( D(y) \) is a desirability function, or if \( Q(y) \) is a quadratic loss function, and one can sample from the posterior predictive distribution of \( Y \) for a given \( x \) value, then one can construct the posterior predictive distribution of \( D(Y) \) or \( Q(Y) \) for each \( x \) value. Whatever the form of the \( D \) or \( Q \) functions, the variance-covariance structure of the data and the uncertainty of the model parameters are accounted for through the posterior predictive distribution for \( Y \). The posterior distribution of \( D(Y) \) or \( Q(Y) \) for each \( x \) value then provides valuable information to the investigator on how reliably the future realizations of \( D(Y) \) or \( Q(Y) \) can meet his or her expectations.

The Statistical Model

The standard regression model for multiple response surface modeling is the classical multivariate multiple regression model,

\[
Y = Bz(x) + e, \tag{1}
\]

where \( B \) is a \( p \times q \) matrix of regression coefficients and \( z(x) \) is a \( q \times 1 \) vector-valued function of \( x \). The vector \( e \) has a multivariate normal distribution with mean vector \( 0 \) and variance-covariance matrix \( \Sigma \). In response surface analysis, the model in Equation (1) is typically composed of a quadratic model for each mean response, but we are allowing a more general covariate structure here. The typical multivariate regression assumption, that \( z(x) \) is the same for each response type, is assumed in this paper.

To account for the uncertainty in the model parameters, \( B \) and \( \Sigma \), the posterior predictive density \( f(y \mid x, \text{data}) \) can be used. By using the typical noninformative joint prior for \( B \) and \( \Sigma \) and the model in Equation (1), the Bayesian predictive density for \( Y \) given \( x \) and the data can be obtained in closed form. This joint prior for \( B \) and \( \Sigma \) is proportional to \( |\Sigma|^{-(n+1)/2} \). The associated Bayesian predictive density for a specified \( x \)-value has the multivariate-\( t \) form (Johnson (1987, Chapter 6)) with \( n \) degrees of freedom (df), where \( n = n - p - q + 1 \) and \( n \) is the sample size, is

\[
f(y \mid x, \text{data}) = c \left( 1 + \frac{1}{n} (y - Bz(x))^T H (y - Bz(x)) \right)^{-(n+p)/2}, \tag{2}
\]

where

\[
c = \frac{\Gamma \left( \frac{n+p+1}{2} \right) \sqrt{|H|}}{\Gamma \left( \frac{n+1}{2} \right) (\pi n)^{p/2}}, \quad H = \frac{\nu V^{-1}}{1 + z(x)^T D^{-1} z(x)},
\]

\[
D = \sum_{i=1}^{n} z(x_i) z(x_i)^T,
\]

\[
V = \left( y^* - (\hat{B}z) \right)^T \left( y^* - (\hat{B}z) \right)
\]

(Press (2003, Chapter 12)). Here, \( n \) is the sample size, \( Z \) is the \( q \times n \) matrix formed by the \( n \) \( z(x) \) covariate vectors, \( \hat{B} \) is the least squares estimate of \( B \), and \( y^* \) is the \( n \times p \) matrix formed by the \( n \) \( (1 \times p) \) \( y \) response data vectors, \( i = 1, \ldots, n \).

Sampling from the Posterior Predictive Distribution

Since Equation (2) is a multivariate \( t \)-distribution, it is easy to simulate \( Y \)-values from this predictive density. Following Johnson (1987, Chapter 6), one can simulate a multivariate \( t \) random variable (r.v.), \( Y \), by simulation of a multivariate normal r.v. and an independent chi-square r.v. For this particular problem the simulation is done as follows. We let \( W \) be a multivariate normal r.v. with zero mean vector and variance-covariance matrix equal to \( H^{-1} \). We let \( U \) be a chi-square r.v. with \( \nu \) df that is independent of \( W \). Next, define

\[
Y_j = \left( \sqrt{\nu W_j / \nu} \mu \right) + \hat{\mu}_j, \quad \text{for } j = 1, \ldots, p \tag{3}
\]

where \( Y_j \) is the \( j \) element of \( Y \), \( W_j \) is the \( j \) element of \( W \), and \( \hat{\mu}_j \) is the \( j \) element of \( \hat{\mu} = Bz(x) \). It follows that \( Y \) has a multivariate \( t \)-distribution with \( \nu \) df.
Multiresponse Optimization Using Bayesian Reliabilities

From a quality and reliability perspective, a natural way to optimize a multiresponse process is to maximize a probability, \( p(x) \), of the form \( \Pr(Y \in A | x), \Pr(D(Y) \geq D^* | x), \) or \( \Pr(Q(Y) \leq Q^* | x) \), where \( A, D^* \), and \( Q^* \) are chosen by the investigator. More generally, we define \( C(y) \) to be a multiresponse optimization criterion such as \( I(y \in A), D(y), \) or \( Q(y) \), and let \( S \) be a conformance set for \( C(y) \), i.e., interest is in \( C(y) \in S \) (e.g., \( D(y) \in S = \{D^*, 1\} \)). We let \( x_0 \) be such that it maximizes \( p(x) \) over the experimental region. If \( p(x_0) \) is sufficiently large, then the experimenter knows that \( x_0 \) will provide operating conditions such that future responses have a high probability of satisfying desired conditions as specified by \( C(y) \in S \).

For a desirability function, the choice of \( D^* \) can be based on a specified percent change from \( D_{opt} \), where \( D_{opt} \) is the optimal value of \( D(\hat{y}) \) over the experimental region and \( \hat{y} = Bz(x) \). A similar approach can be used for \( Q^* \). Furthermore, from a process improvement perspective the conformance set \( S \) can be constructed based upon current process conditions. Reliabilities relative to product specifications can also be computed. For example, for the Derringer-Suich desirability function, \( D = 0 \) can be chosen if a zero desirability means that one or more responses are outside of product specifications (the desirability function of Harrington has, by definition, cut-off values relating to product quality). For quadratic loss, one can use \( Q^* = Q(\Gamma) \), where \( \Gamma = \left( \Gamma_1, ..., \Gamma_p \right) \) are values that represent the outer limits of acceptable departure from response target values. Of course from a more free-form perspective, histograms, or quantiles, of the responses \( Y, D(Y), \) or \( Q(Y) \) can also be constructed.

A more penetrating strategy, however, can be to have a product expert, or team of experts, develop \( C(y) \) and \( S \) such that \( C(Y) \in S \) corresponds to a "good" product. In this case, if \( p(x) \) is sufficiently large for a set of \( x \)'s in an operational neighborhood of \( x_0 \), then a reliable and robust product is achieved. The importance of having an expert panel construct a desirability function to link values of \( D(y) \) to product quality is stressed by Derringer (1994).

Prior to computing these probabilities, the experimenter invariably wants to do a preliminary optimization corresponding to \( C(\hat{y}) \in S \). Once this "mean-optimal" \( x \)-point is obtained using one of the procedures discussed in the Introduction, the Bayesian reliability can be computed for that point. In addition, Bayesian reliabilities can be computed for a variety of \( x \)-points to see if further improvements in reliability can be obtained.

If the Bayesian reliability is sufficiently large, then the experimenter can feel confident that the optimized process will produce satisfactory responses with a high degree of likelihood. Of course, validation runs should be done to make sure that the statistical model still holds for the optimal factor conditions. If, however, the optimized reliability is not sufficiently large, then remedial work will be necessary.

If a preliminary optimization based on the mean response surfaces indicates that, on the average, satisfactory conditions are obtainable, then poor reliabilities must be due to the variance and covariance values of the posterior predictive distribution of \( Y \). If the posterior predictive distribution of \( Y \) is too dispersive, then the resulting Bayesian reliabilities will not be near one. The dispersion of this distribution depends on both the natural variability of data and the model parameter uncertainty, which of course is intimately related to the sample size and experimental design used in the study. If the process response variances can be reduced, then this may increase the resulting reliabilities. If the experimenter's Bayesian reliability cannot be increased sufficiently due to limited ability to reduce process variation, then some increase in reliability can be obtained by decreasing the model parameter uncertainty. This can be done by increasing the sample size, e.g., by replicating the study design.

Before taking such remedial measures, this posterior predictive approach allows the experimenter to forecast how such measures might work out if the data sampled are, in certain ways, representative of future data to be taken. For example, if it is possible to reduce the variation of the process responses, one can make the following transformation to forecast what might happen when process variation is reduced. The experimenter can use the pseudo-data, \( y^*_j = \hat{y}_j + (1 - \lambda_j) \epsilon_j, j = 1, ..., p \) (where \( 100\lambda_j \) represents a percent reduction in residual size for the \( j \)-th response type) to artificially generate new process data having the same mean values but reduced variances. This helps in understanding how the process variation affects the reliability.

An additional approach is to modify the posterior predictive distribution in such a way that one can simulate responses from this distribution as if more
data had been acquired. For the posterior sampling model in Equation (3) this is easily done as follows. Note that the multivariate t posterior predictive density depends upon the data only through the sufficient statistics, $\Sigma$ and $B$, the degrees of freedom, and the design matrix $Z$. By increasing the rows of the design matrix (to add new data points), and augmenting the df accordingly, one can simulate new data. This corresponds to adding artificial data in such a way that the sufficient statistics, $\Sigma$ and $B$, remain the same. This gives the experimenter an idea of how much the reliability can be increased by reducing model uncertainty. For example, the experimenter can forecast the effects of replicating the experiment a certain number of times. This idea is similar in spirit to the notion of a “pre-posterior” analysis as described by Raffin and Schlaiffer (2000).

**Computing and Optimizing the Bayesian Reliabilities**

Using Monte Carlo simulation from Equation (3), one can approximate the reliability $p(x)$ for various $x$-values in the experimental region using

$$p(x) \approx \frac{1}{N} \sum_{i=1}^{N} I(C(Y_i) \in S),$$

(4)

where $N$ is the number of simulations, and the $Y_i$, r.v. are simulated conditional on $x$ and the data. For a small number of factors it is computationally reasonable to grid over the experimental region to compute values of $p(x)$ for purposes of optimization. However, even for three or more factors, it may be preferable to have a more efficient approach to optimizing $p(x)$.

One approach is to maximize $p(x)$ using general optimization methods, such as those discussed in Nelder-Mead (1964), Price (1977), or Chatterjee, Landato, and Lynch (1996).

Another, possibly more revealing, approach is to compute $p(x)$ for $x$-points in some response surface experimental design and then fit a closed-form response surface model to obtain an approximate reliability surface, $\tilde{p}(x)$. Since values of $p(x)$ can be computed much more quickly than $p(x)$, approximate optimization of $p(x)$ can be done. For example, a ridge analysis can be done on $p(x)$ to explore, in an approximate fashion, how $p(x)$ changes as $x$ moves out from the center of the experimental region in an optimal way.

In a ridge analysis, an objective function (e.g., $p(x)$) is optimized over hyperspheres of radius $r$, $\{x : x'x = r^2\}$, for various values of $r$. It is typically assumed that the $x$-points have been centered at zero and given a common scale, such as $-1 \leq x_i \leq 1$, $i = 1, \ldots, k$. An optimal response plot of $p(x_i)$ vs. $x_i$ is then constructed, where

$$p(x_i) = \max_{x'^2 = x_i} p(x).$$

Also, an overlay plot of $x_{i*}$ ($i = 1, \ldots, k$) vs. $r$ is typically made to see how the $x_{i*}$'s change as $r$ varies. For an excellent overview of ridge analysis see Hoerl (1985).

By computing $p(x)$ for $x$-values over a coarse-to-moderately dense grid over the experimental region, one can map out a “response surface” for this objective function. If there is a need to interpolate between grid points, a logistic regression model such as

$$\tilde{p}(x) = \frac{\exp(u(x)/\gamma)}{1 + \exp(u(x)/\gamma)},$$

(5)

can be fit to the binary data produced by the above Monte Carlo simulations to produce an approximating function over the experimental region. Basic response surface methods can then be applied to $u(x)/\gamma$ in Equation (5) to assess optimal conditions. For example, a canonical analysis or classical ridge analysis can be done if one can adequately model $p(x)$ by a quadratic polynomial function for $u(x)/\gamma$. If the standard quadratic model does not give an adequate fit, a ridge analysis using higher order polynomials or even nonparametric regression methods can be used, although a general nonlinear optimization procedure is required for the ridge analysis.

Since there are many (simulated) replications for each $x$-point, an easy way to check the fit of the model in Equation (5) is to plot $\tilde{p}(x)$ vs. $\hat{p}(x)$, where $\hat{p}(x)$ equals the right-hand-side of Equation (4) and $\hat{p}(x)$ is the estimated value of $\tilde{p}(x)$ from Equation (5). Across the various values of $x$ in an experimental region, the $\hat{p}(x)$ vs. $\tilde{p}(x)$ plot should draw out roughly a 45 degree line on the unit square. Of course, the final, optimized $\tilde{p}(x_0)$ value should be compared with $\hat{p}(x_0)$ evaluated using Monte Carlo simulations and then with $\hat{p}(x_0)$ obtained from actual process runs at $x_0$.

If a logistic regression model does not give a good fit over the entire experimental region, it may still be possible to obtain a good fit over a sub-region of the experimental region. Here, the Monte Carlo process described above can be re-run with a finer grid over this smaller region. After generating a refined and
smaller response surface over the sub-region of x-values, another logistic regression model can be fit. This time we can expect a somewhat better fit, as the model is approximating a smaller response surface. This process can be continued to explore and improve estimates of optimal conditions using, for example, only the standard quadratic model for which most response surface procedures have been developed.

Since p(x) is a joint probability, it turns out that the marginal probabilities corresponding to marginal events about some of the Y_i values can also be easily computed as a by-product of the Monte Carlo simulation process for p(x). If we have

\[ p_i(x) = \Pr(Y_i \in A_i \mid x, \text{data}), \]

where \( A_i \) is an interval (possibly one or two sided), then we can assess the \( p_i(x) \) simultaneously along with \( p(x) \). Modification of \( A \) and the \( A_i \) can be used to allow an investigator to directly observe economic or performance issues through the \( p(x) \) and \( p_i(x) \) respectively, which can be easily related to how often the process is involved. Of course, other marginal (joint) probabilities concerning the \( Y_i \) can be computed in a similar fashion.

For process ruggedness assessment, a Bayesian credible region can be computed. For example, this can be done by mapping out all of the x-values for which \( p(x) \) (\( p_i(x) \)) is at least 0.95.

**Examples**

**A Mixture Experiment**

This example involves a mixture experiment to study the surfactant and emulsification variables involved in pseudolatex formation for controlled-release, drug-containing beads. (Frisbee and McGinity (1994)). An extreme vertices design is used to study the influence of surfactant blends on the size of the particles in the pseudolatex and the glass transition temperature of films cast from those pseudolatexes. The factors chosen are: \( x_1=1\% \) of Pluronic F68; \( x_2=1\% \) of polyoxyethylene 40 monostearate; and \( x_3=1\% \) of polyoxyethylene sorbitan fatty acid ester NF. The experimental design used is a modified McLean-Anderson design (McLean and Anderson (1966)) with two centroid points, resulting in a sample size of eleven. The response variables measured were “particle size” and “glass transition temperature,” which are denoted here as \( Y_1 \) and \( Y_2 \), respectively. The goal of the study was to find values of \( x_1 \), \( x_2 \), and \( x_3 \) that minimize, as best as possible, both \( Y_1 \) and \( Y_2 \). In particular, it is assumed here that one desires \( Y_1 \) to be at most 234 and \( Y_2 \) to be at most 18.5. Anderson and Whitcomb (1998) also analyzed this data set to illustrate Design Expert’s capability to map out overlapping mean response surfaces.

Following Frisbee and McGinity (1994) and Anderson and Whitcomb (1998), quadratic mixture models are fit to the bivariate response data. For this example, however, a severe outlier run is deleted. The resulting regression models obtained are

\[
\begin{align*}
\hat{y}_1 &= 248.8 x_1 + 272.0 x_2 + 533.3 x_3 - 485.0 x_1 x_2 - 424.4 x_1 x_3 \\
\hat{y}_2 &= 18.7 x_1 + 14.1 x_2 + 35.4 x_3 - 36.7 x_1 x_2 + 18.0 x_2 x_3.
\end{align*}
\]

The Wilks-Shapiro test for normality of the residuals for each regression model yields p-values greater than 0.05. Tests for multivariate normality via skewness and kurtosis (Mardia (1974)) are not significant at the 5% level. However, it should be recognized that due to the small sample size in this (and the next) example, the tests for normality are not very powerful.

In this example, the \( I(y \in A), D(y) \), and \( Q(y) \) functions are used to compare the results one obtains with each. For \( I(y \in A), A = \{ y_1, y_2 \} : y_1 \leq 234 \) and \( y_2 \leq 18.5 \}. \) Here, \( D(y) \) corresponds to the Derringer-Suich (1980) desirability function and \( Q(y) \) corresponds to the quadratic loss function of Vining (1998).

The \( D(y) \) function for this example is defined as \( D(y) = (d_1(y_1) d_2(y_2))^{1/2} \), where, for \( i = 1, 2, \)

\[
d_i(y) = \begin{cases} 0 & \text{if } y_i > U_i \\ \frac{y_i - l_i}{U_i - l_i} & \text{if } l_i \leq y_i \leq U_i \\ 1 & \text{if } y_i < l_i \end{cases}
\]

and \( L_1 = 227.27, U_1 = 234, L_2 = 14.09 \), and \( U_2 = 18.5 \). Here, \( L_1 \) and \( L_2 \) are the minimum predicted response values over the mixture simplex corresponding to \( y_1 \) and \( y_2 \), respectively.

The \( Q(y) \) function is defined as \( Q(y) = (y - T)^T \Sigma(x(x))^{-1} (y - T) \), where \( T \) is a 2 x 1 vector of target values, and

\[
\Sigma(x(x)) = x(x)^T Z^T Z^{-1} x(x) \Sigma
\]

is the estimate of the variance-covariance matrix of

**Journal of Quality Technology**

Vol. 36, No. 2, April 2004

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\( \hat{y}(x) \), the predicted mean response at \( x \). The value chosen for \( T \) is \( (L_1, L_2)^T \).

In this example the posterior predictive distribution of \( Y \) is used to assess optimization for this mixture experiment using the three multivariate optimization approaches defined by \( I(y \in A) \), \( D(y) \), and \( Q(y) \), as specified above. Using these functions, two different strategies of assessment are employed. The first strategy is to use each of the optimization functions applied to the mean response surfaces to find an “optimal” \( x \)-point in the mixture simplex. Then the Bayesian reliability, \( p(x) \), is computed for that \( x \)-point. In this example, \( D^* = (1/2)D_{opt} \), where \( D_{opt} \) is the optimized value of \( D(\hat{y}) \) and \( Q^* = 2Q_{opt} \), where \( Q_{opt} \) is the optimized value of \( Q(\hat{y}) \). In addition, as discussed earlier, two modifications are made to see how reducing process variability and increasing sample size can improve these reliabilities.

The second optimization assessment strategy is to maximize \( p(x) \) over the mixture simplex to determine the optimal \( x \) value. As with the first strategy, modifications can be made to see how reducing process variability and increasing sample size improve these Bayesian reliabilities. For all of the simulations done in this section, 1000 multivariate responses were generated for a given factor configuration.

The contour plots for the overlapping mean response surface approach are shown in Figures 1 and 2. Figure 1 shows the contour plots of the mean response surfaces for responses \( y_1 \) and \( y_2 \). In Figure 2 the region in gray is the set of \( x \)-points such that \( \hat{y}_1 \leq 234 \) and \( \hat{y}_2 \leq 18.5 \). From viewing Figures 1 and 2, it appears that the point \( x = (0.75, 0.25)^T \) is a good choice. In addition, the point \( x = (0.78, 0.22)^T \) maximizes \( D(\hat{y}) \), while \( x = (0.74, 0, 0.26)^T \) minimizes \( Q(\hat{y}) \).

For Tables 1–6, the “percent reduction” represents the shaded gray area represents the factor combinations such that the mean particle size \( (y_1) \leq 234 \) and the mean glass transition temperature \( (y_2) \leq 18.5 \).
TABLE 1. $\Pr(Y \in A | x = (0.75, 0, 0.25)^T)$ for Increasing Replications and Decreasing Residual Error

<table>
<thead>
<tr>
<th>Percent reduction</th>
<th>Rep=1</th>
<th>Rep=2</th>
<th>Rep=3</th>
<th>Rep=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.683</td>
<td>0.883</td>
<td>0.924</td>
<td>0.932</td>
</tr>
<tr>
<td>25</td>
<td>0.802</td>
<td>0.947</td>
<td>0.971</td>
<td>0.981</td>
</tr>
<tr>
<td>50</td>
<td>0.919</td>
<td>0.995</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

TABLE 2. $\Pr(D(Y) \geq (1/2)D_{opt} | x = (0.78, 0, 0.22)^T)$ for Increasing Replications and Decreasing Residual Error

<table>
<thead>
<tr>
<th>Percent reduction</th>
<th>Rep=1</th>
<th>Rep=2</th>
<th>Rep=3</th>
<th>Rep=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.601</td>
<td>0.763</td>
<td>0.795</td>
<td>0.805</td>
</tr>
<tr>
<td>25</td>
<td>0.712</td>
<td>0.865</td>
<td>0.903</td>
<td>0.917</td>
</tr>
<tr>
<td>50</td>
<td>0.858</td>
<td>0.960</td>
<td>0.977</td>
<td>0.985</td>
</tr>
</tbody>
</table>

TABLE 3. $\Pr(Q(Y) \leq 2Q_{opt} | x = (0.76, 0, 0.24)^T)$ for Increasing Replications and Decreasing Residual Error

<table>
<thead>
<tr>
<th>Percent reduction</th>
<th>Rep=1</th>
<th>Rep=2</th>
<th>Rep=3</th>
<th>Rep=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.383</td>
<td>0.551</td>
<td>0.590</td>
<td>0.597</td>
</tr>
<tr>
<td>25</td>
<td>0.565</td>
<td>0.659</td>
<td>0.696</td>
<td>0.706</td>
</tr>
<tr>
<td>50</td>
<td>0.677</td>
<td>0.809</td>
<td>0.842</td>
<td>0.851</td>
</tr>
</tbody>
</table>

sent the percent reduction in residual variation as given by the data transformation $y^* = \tilde{y} + (1 - \lambda)\xi$. Here, $100\lambda$ represents a common percent reduction in residual variation. For columns 2–4 of Tables 1–6 a “preposterior” analysis as described earlier is displayed. Here, data is simulated from a posterior, predictive t-distribution with the same $\Sigma$ and $\bar{B}$ values as before, but using a design matrix, $Z$, that has been augmented by using 2, 3, or 4 design replications (for columns 2, 3, and 4, respectively). The degrees of freedom associated with the $t$-distribution are also increased accordingly. For the $Q(y)$ function, this preposterior analysis also requires that the variance-covariance matrix in Equation (7) use the augmented design matrix, $Z$, since $Q(y)$ is a function of $Z$. The table entries in the upper left hand corner (with 0% reduction and 1 replication) contain the results of the analysis of the original experimental data with no preposterior modifications.

As can be seen from Table 1, the entry with 0% reduction and 1 replication shows that $p(x) = \Pr(Y \in A | x) = 0.68$ for $x = (0.75, 0, 0.25)^T$, which is disappointingly small, even though both response surface means satisfy the conditions described by the set $A$. However, as the variation of the data is reduced it is evident that the reliability increases. Also, as the number of replications of the experimental design increases, the reliability increases as the parameter uncertainty is decreased. Using the strategy of Chiao and Hamada (2001), this proportion-conforming reliability is computed assuming a multivariate normal distribution using the estimated prediction Equations in (6) along with the estimated variances ($\hat{\sigma}^2 = 13.87$ and $\hat{\sigma}^2_2 = 2.97$) of $e_1$ and $e_2$, respectively, and estimated correlation of the residuals ($\hat{\rho} = -0.412$). These estimates are obtained using the unbiased estimator of $\Sigma$, which is the sum-of-squares-cross-product matrix divided by $(n - q)$. The maximum likelihood estimate (MLE) is obtained by dividing by $n$ instead of $(n - q)$. This reliability measure, computed using 1000 simulations from this multivariate normal distribution, yields $p(x) = 0.87$. If the MLE is used instead, then $p(x) = 0.954$. In any case, this shows that ignoring the model parameter uncertainty can cause the estimated reliability to be too large.

However, the decrease in parameter uncertainty due to increasing the sample size may not always drive the reliability to near one, because the natural variation in the process at hand may keep the reliability less than one, so that, as the parameter uncertainty decreases, the reliability converges to some value less than one. By decreasing the process variation, however, the reliability can be driven towards one, provided that the optimized mean response surfaces are satisfactory.

For the desirability function described earlier in this section, the maximized $D(\tilde{y})$ value is $D_{opt} = 0.74$ at $x = (0.78, 0, 0.22)^T$. Assuming this is an adequate desirability value, one might ask, from a quality perspective, “what is the probability that, for $x = (0.78, 0, 0.22)^T$, any future $D(y)$ value will be at least as large as some lower value, $D^*$?” Table 2 shows values of $\Pr(D(Y) \geq (1/2)D_{opt} | x = (0.78, 0, 0.22)^T)$. These reliabilities show the same pattern of increasing values as the error variation and the parameter uncertainty are reduced. So, from Table 2, it is appears from the data at hand that confirming a high degree of reliability for this process operating at $x = (0.78, 0, 0.22)^T$ requires a substantial reduction in

\[\Pr(Q(Y) \leq 2Q_{opt} | x = (0.76, 0, 0.24)^T)\]
TABLE 4. max Pr(Y ∈ A | x) and Associated Optimal x-point, for Increasing Replications and Decreasing Residual Error

<table>
<thead>
<tr>
<th>Percent reduction</th>
<th>Rep=1</th>
<th>Rep=2</th>
<th>Rep=3</th>
<th>Rep=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.741</td>
<td>0.902</td>
<td>0.936</td>
<td>0.950</td>
</tr>
<tr>
<td>(0.78, 0.0, 0.22)</td>
<td>(0.78, 0.0, 0.22)</td>
<td>(0.76, 0.0, 0.24)</td>
<td>(0.78, 0.0, 0.22)</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.840</td>
<td>0.963</td>
<td>0.987</td>
<td>0.989</td>
</tr>
<tr>
<td>(0.78, 0.01, 0.21)</td>
<td>(0.78, 0.0, 0.22)</td>
<td>(0.75, 0.0, 0.25)</td>
<td>(0.75, 0.0, 0.25)</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.937</td>
<td>0.999</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(0.76, 0.0, 0.24)</td>
<td>(0.79, 0.0, 0.21)</td>
<td>(0.75, 0.0, 0.25)</td>
<td>(0.73, 0.0, 0.27)</td>
<td></td>
</tr>
</tbody>
</table>

process variation as well as additional replications of the experiment.

For the quadratic loss function, Q(y), described above, the minimized value of Q(y) is Q_{opt} = 7.79 at x = (0.76, 0.0, 0.24)'. Table 3 shows values of \Pr(Q(Y) ≤ 2Q_{opt} | x = (0.76, 0.0, 0.24)') as with Tables 1 and 2, Table 3 shows very similar behavior for the Q(y) loss function at x = (0.76, 0.0, 0.24)'.

A similar pattern is evident from Tables 4–6, where the Bayesian reliabilities are maximized. Here, the optimal x-points do not change much from their respective values obtained by way of the mean response surfaces.

While the more commonly used Derringer-Suich (1980) desirability function does not provide an absolute quality criterion, the Harrington (1965) desirability function does. Harrington provides an absolute quality scale: 0-0.37 ("very poor") to "poor"); 0.37-0.60 ("fair"); 0.60-0.80 ("good"); and 0.80-1 ("excellent") for his desirability function. Here, it is interesting to note that, for the example in this section, D_H(Y) is maximized to 0.96 at x = (0.76, 0.0, 0.24)', where D_H is the Harrington desirability function. However, p_H(x) = Pr(D_H(Y) ≥ 0.60 | x) = 0.72. This shows that the probability of obtaining a future response that is at least borderline "good" is only 0.72, despite the fact that the optimal desirability was estimated to be 0.96, which is "excellent" on the Harrington scale.

As discussed previously, it is also interesting to consider "bottom-line" reliabilities such as Pr(D(Y) > 0 | x) and Pr(Q(Y) ≤ Q(I) | x) where I = {l1, ..., l_n} relates to the outer limits of departures from target specifications. For this example, D(Y) = 0 corresponds to y_1 ≤ l_1 = 234 or y_2 ≥ l_2 = 18.5. At x = (0.78, 0.0, 0.22)', Pr(D(Y) > 0 | x) is only 0.696. This situation is better for Q(I) at x = (0.76, 0.0, 0.24)' with Pr(Q(Y) ≤ Q(I) | x) = 0.883.

A Ridge Analysis Example

This example illustrates the optimization of an event probability p(x) = Pr(Y ∈ A | x) for a high performance liquid chromatography assay. Here there are three factors (x_1 = percent of isopropl alcohol, x_2 = temperature, and x_3 = pH) and four responses (y_1 = resolution, y_2 = run time, y_3 = signal-

TABLE 5. max Pr(D(Y) ≥ (1/2)D_{opt} | x) and Associated Optimal x-point, for Increasing Replications and Decreasing Residual Error

<table>
<thead>
<tr>
<th>Percent reduction</th>
<th>Rep=1</th>
<th>Rep=2</th>
<th>Rep=3</th>
<th>Rep=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.631</td>
<td>0.761</td>
<td>0.797</td>
<td>0.825</td>
</tr>
<tr>
<td>(0.78, 0.0, 0.22)</td>
<td>(0.78, 0.0, 0.22)</td>
<td>(0.76, 0.0, 0.21)</td>
<td>(0.78, 0.0, 0.22)</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.734</td>
<td>0.872</td>
<td>0.900</td>
<td>0.914</td>
</tr>
<tr>
<td>(0.78, 0.0, 0.22)</td>
<td>(0.78, 0.0, 0.22)</td>
<td>(0.78, 0.0, 0.22)</td>
<td>(0.78, 0.0, 0.22)</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.869</td>
<td>0.969</td>
<td>0.986</td>
<td>0.988</td>
</tr>
<tr>
<td>(0.78, 0.0, 0.22)</td>
<td>(0.76, 0.0, 0.24)</td>
<td>(0.75, 0.0, 0.25)</td>
<td>(0.76, 0.0, 0.24)</td>
<td></td>
</tr>
</tbody>
</table>
to-noise ratio (s/n), and $y_i =$ tailing). For this assay, the chemist desires to have the event

$$A = \{ y : y_1 \geq 1.8, y_2 \leq 15, y_3 \geq 300, 
0.75 \leq y_4 \leq 0.85 \}$$

(8)

occur with high probability. As such, it is desirable to maximize $\Pr(Y \in A \mid x)$ as a function of $x$.

A Box-Behnken experimental design is run, with three center points, to gather data to fit four quadratic response surfaces. The data used are given in Table 7. Quadratic regression models with the covariate vector, $z(x) = (1, x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1 x_2, x_1 x_3)$, are used.

All of the response surface models fit well, with $R^2$ values above 99%. As in example 1, the Wilks-Shapiro test for normality of the residuals for each regression model yields $p$-values greater than 0.05, and the Mardia tests for multivariate skewness and kurtosis are not significant at the 5% level. The factor levels are coded so that all values are between $-1$ and $+1$, with the center of the experimental region at the origin.

For this example, a ridge analysis is used to examine how the value of $\Pr(Y \in A \mid x)$ changes when maximized over spheres of varying radii centered at the origin of the experimental region. In order to create a computationally efficient model, which is needed to do the ridge analysis, a fourth-degree polynomial model is fit to a cubical grid of 729 points containing the Box-Behnken experimental design. This grid is formed by the Cartesian product of nine equally spaced points (increments of 0.25).

### Table 7. Factor Levels and Responses for HPLC Example

<table>
<thead>
<tr>
<th>run</th>
<th>pipa</th>
<th>temp</th>
<th>pH</th>
<th>rs</th>
<th>runtime</th>
<th>s/n</th>
<th>tailing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65</td>
<td>30</td>
<td>0.175</td>
<td>2.14</td>
<td>22</td>
<td>172</td>
<td>0.76</td>
</tr>
<tr>
<td>2</td>
<td>65</td>
<td>50</td>
<td>0.175</td>
<td>1.73</td>
<td>12</td>
<td>311</td>
<td>0.88</td>
</tr>
<tr>
<td>3</td>
<td>65</td>
<td>40</td>
<td>0.050</td>
<td>1.93</td>
<td>16</td>
<td>251</td>
<td>0.80</td>
</tr>
<tr>
<td>4</td>
<td>65</td>
<td>40</td>
<td>0.300</td>
<td>1.95</td>
<td>16</td>
<td>241</td>
<td>0.80</td>
</tr>
<tr>
<td>5</td>
<td>70</td>
<td>40</td>
<td>0.175</td>
<td>2.17</td>
<td>14</td>
<td>278</td>
<td>0.79</td>
</tr>
<tr>
<td>6</td>
<td>70</td>
<td>50</td>
<td>0.050</td>
<td>1.97</td>
<td>11</td>
<td>371</td>
<td>0.86</td>
</tr>
<tr>
<td>7</td>
<td>70</td>
<td>30</td>
<td>0.300</td>
<td>2.38</td>
<td>19</td>
<td>194</td>
<td>0.74</td>
</tr>
<tr>
<td>8</td>
<td>70</td>
<td>50</td>
<td>0.300</td>
<td>1.98</td>
<td>11</td>
<td>360</td>
<td>0.86</td>
</tr>
<tr>
<td>9</td>
<td>70</td>
<td>30</td>
<td>0.050</td>
<td>2.37</td>
<td>18</td>
<td>204</td>
<td>0.74</td>
</tr>
<tr>
<td>10</td>
<td>70</td>
<td>40</td>
<td>0.175</td>
<td>2.20</td>
<td>14</td>
<td>280</td>
<td>0.78</td>
</tr>
<tr>
<td>11</td>
<td>75</td>
<td>40</td>
<td>0.300</td>
<td>2.42</td>
<td>13</td>
<td>314</td>
<td>0.78</td>
</tr>
<tr>
<td>12</td>
<td>75</td>
<td>30</td>
<td>0.175</td>
<td>2.61</td>
<td>17</td>
<td>223</td>
<td>0.73</td>
</tr>
<tr>
<td>13</td>
<td>75</td>
<td>50</td>
<td>0.175</td>
<td>2.14</td>
<td>10</td>
<td>410</td>
<td>0.85</td>
</tr>
<tr>
<td>14</td>
<td>75</td>
<td>40</td>
<td>0.050</td>
<td>2.42</td>
<td>12</td>
<td>324</td>
<td>0.78</td>
</tr>
<tr>
<td>15</td>
<td>70</td>
<td>40</td>
<td>0.175</td>
<td>2.20</td>
<td>14</td>
<td>281</td>
<td>0.79</td>
</tr>
</tbody>
</table>
from \(-1\) to \(1\) for each factor. At each grid point, 1000 binary responses are generated corresponding to whether or not the event \(A\) in Equation (8) occurred. The \(u(x)^3\) term used for the logistic regression model in Equation (5) is a complete fourth-degree polynomial in all three factors. Fourth-degree polynomial logistic regression models are also created to model each of the marginal probabilities, \(\Pr(Y_i \in A_i \mid x), \ i = 1, \ldots, 4\), where \(A_1 = \{y_1 : y_1 \geq 1.8\}, \ A_2 = \{y_2 : y_2 \leq 15\}, \ A_3 = \{y_3 : y_3 \geq 300\},\) and \(A_4 = \{y_4 : 0.75 \leq y_4 \leq 0.85\}\). This results in approximate probability functions of \(\tilde{p}_i(x), \ i = 1, \ldots, 4\). The

FIGURE 3. Goodness-of-Fit Plots: Observed Proportions from Simulations vs. Logistic Regression Model Predicted

Probabilities.

FIGURE 4. Optimal Response Plot for Ridge Analysis of \(p(x)\). Overlaid are the Values of \(p_i(x) - p_i(x)\) Corresponding to the Optimal Values of \(x\) for the Ridge Analysis of \(p(x)\). Here, the Ridge Analysis is Based Upon the \(p(x), p_1(x) - p_i(x)\) Functions Approximated by Logistic Regression Models.
logistic regression models corresponding to $\tilde{p}(x)$, $\tilde{p}_i(x)$, $i = 1, \ldots, 4$ have reasonably good predictive ability, as demonstrated by agreement between the observed and predicted proportions. See Figure 3 for scatter plots of observed vs. logistic model estimated probabilities.

A ridge analysis for the logistic regression model is conducted by gridding over the polar coordinate angle space of each sphere at nine equally spaced radius values from 0.1 to 1. The angle increments used are 1 degree. A detailed discussion of how to do such a ridge analysis is given by Peterson (1993). Figure 4 shows the optimal probability plot of $\tilde{p}(x_r)$ vs. $r$. Figure 5 shows the optimal coordinate plot of the $x_r$-factor levels vs. $r$. Using the ridge trace of optimal $x$-points for $\tilde{p}(x)$, plots of $\tilde{p}_i(x)$, $i = 1, \ldots, 4$, are overlaid on the plot in Figure 4. As a further check on the accuracy of the plots made by logistic regression, the $x$-values used to make the optimal coordinate plot were used to compute the $p(x)$, $p_1(x) - p_4(x)$ probabilities using 1000 Monte Carlo simulations. The resulting graph is virtually identical to Figure 4.
From Figure 4 it is clear that the marginal probabilities \( \tilde{p}_i(x) \), \( i = 1, 2, \) and 4 are close to one for all radius values between 0.1 and 1. However, a radius value of at least 0.3 is required for \( \tilde{p}_3(x) \) to get close to 1, and thereby allow the joint probability \( \tilde{p}(x) \) to get close to 1. Here, no pre-posterior or error variability reduction analyses are needed, since excellent event probabilities are obtainable with the data at hand. For this example, the Bayesian posterior predictive approach assures us that, if the modeling is done well, then the uncertainty due to the unknown model parameters is not a concern. However, the experimenter should, of course, still do some confirmatory experiments at the chosen factor configuration.

Figure 6 is a plot of factor points where \( p(x) \) is at least 0.95. This forms a 95% Bayesian credible region for the factor combination that maximizes \( \Pr(Y \in A \mid x) \). Here it can be seen that for higher levels of "% isopropyl alcohol" one has more latitude in varying pH and temperature while still keeping \( \Pr(Y \in A \mid x) \geq 0.95 \).

**Discussion**

We can see from the examples that the Bayesian reliability method described in this paper provides a complete way to assess the quality of an optimized multivariate response for the regression model of Equation (1). This approach is the only approach that takes into account the variance-covariance structure of the data as well as the uncertainty of all of the model parameters. In addition, this method easily allows the investigator to assess the effect of modifying the variance of the process being studied and the effect of additional data on the reliability.
estimate. This method also has the flexibility of computing reliabilities for many types of multivariable optimization criteria.

Computing a Bayesian reliability $p(x)$ by Monte Carlo methods can be done quite rapidly, so that optimizing $p(x)$ over several hundred $x$-points only takes a few minutes of computing time on a modern PC. A SAS/IML subroutine is available from the author for simulating posterior predicted values using the multivariate regression model in Equation (1) with Equation (3).

Future research will explore extending results to include noise variables, seemingly unrelated regression models (Zellner (1971)), residual error distributions with heavier-than-normal tails, and the multivariate regression model of Chiao and Hamada (2001). Sampling from the posterior predictive distributions for these models should be achievable by Markov Chain Monte Carlo methods. In addition to accounting for model parameter uncertainty, the uncertainty of the model form itself can be addressed by using Bayesian model averaging (Press (2003, Chapter 13)).

Appendix

In this appendix, we provide a flowchart (in Figure A1) as a general outline for using Bayesian reliability assessment for multiple response surface optimization as discussed in this paper. It is assumed throughout that good modeling practices are observed. See Johnson and Wichern (2002) for multivariate multiple regression modeling and Hosmer and Lemeshow (2002) for logistic regression modeling, if needed. Here, $y$-hat represents an estimated mean response vector for the multivariate regression model and $x$ represents the $x$-value that maximizes $p(x)$. See the section on “A Posterior Predictive Approach” for details about the statements in the flowchart.

Acknowledgments

I would like to acknowledge the discussions with Enrique del Castillo, Guillermo Miró Quesada, and Kwan Lee which were helpful in writing this manuscript. I would also like to thank the reviewers for their careful review and helpful comments. I would especially like to thank Lan Wang for her fine programming work on the second example in the Examples section.

References


